

Robust and Multiresolution Video Delivery: From H.26x to Matching Pursuit Based Technologies.

THÈSE N°2322 (2000)

PRÉSENTÉE À LA SECTION DE D'ÉLECTRICITÉ

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES TECHNIQUES

PAR

PASCAL FROSSARD

Ingénieur EPF, EPFL, Switzerland
de nationalité suisse

Acceptée sur proposition du jury:

Prof. Philippe Robert, EPFL, président du jury
Prof. Murat Kunt, EPFL, directeur de thèse
Prof. Patrick Thiran, EPFL, corapporteur
Dr. Michel Benard, Hewlett-Packard, corapporteur
Dr. Olivier Verscheure, IBM, corapporteur
Prof. Pierre Vandergheynst, EPFL, corapporteur

Lausanne, EPFL
2000

To my beloved Chantal and Melvin.

A problem worthy of attack proves its worth by fighting back.

E.G. Hardy.

Acknowledgments

Writing a thesis can be considered as a solitary work, but research is clearly collaborative. Therefore I first would like to thank all the people that contribute in a variety of ways to make the completion of this work possible.

My deepest thanks go to my advisor Professor Murat Kunt for his support and encouragement. He guided me through this work with clearheadedness, drove me into very fruitful discussions and definitely contributed to the success of this work.

Nothing could have been done without the essential support of Hewlett-Packard and more particularly Michel Benard who offered me the great opportunity to work in a very enriching industrial partnership.

I would like also to express my greatest and sincere gratefulness to Dr Olivier Verscheure (IBM, US) and Prof. Pierre Vanderghynst (EPFL,LTS) for their invaluable contribution to this work. They undoubtedly helped a lot to the success of this thesis.

I express a special thanks to the members of the HPP group, and more particularly to Dr Eric Debes, Dr Julien Reichel and Dr Fulvio Moschetti for very fruitful discussions and collaborative work. A special prize moreover to Eric who had to bear me all day long in his fresh office.

I have a special gratitude to all the LTS members and especially to the LTS staff for their help at many stages of this work: Gilles Auric, Rizwan Khan, Isabelle Bezzi, Corinne Degott and Marianne Marion.

Many thanks also to the members of the ICA and previously TCOM group who welcomed me for my pre-thesis work and initiated me on this research topic. In particular Prof. Jean-Pierre Hubaux (EPFL, ICA), Dr. Xavier Garcia (Adventis), Dr Maher Hamdi (Cita), Dr Raffaele Noro (Vitesse, USA), Bruno Dufresne (Goolean) and Danielle Alvarez.

I would like to thank also several persons who have contributed in different ways to this work: Prof. Patrick Thiran (EPFL, ICA), Prof. Jean-Yves Le Boudec (EPFL, ICA), Dr Christian van den Branden Lambrecht (EMC, USA), Dr Stefan Winkler (Genimedia) and Mr. Nenad Buncic (EPFL, LTHI).

Finally, my gratitude goes to my wife Chantal who patiently encouraged and accompanied me during all the years of this thesis. This work undoubtedly owes a lot to her love, presence and support. A small thank also to my son Melvin for his sunbeam during the last year of this work.

Abstract

With the joint development of networking and digital coding technologies multimedia and more particularly video services are clearly becoming one of the major consumers of the new information networks. The rapid growth of the Internet and computer industry however results in a very heterogeneous infrastructure commonly overloaded. Video service providers have nevertheless to offer to their clients the best possible quality according to their respective capabilities and communication channel status. The Quality of Service is not only influenced by the compression artifacts, but also by unavoidable packet losses. Hence, the packet video stream has clearly to fulfill possibly contradictory requirements, that are coding efficiency and robustness to data loss.

The first contribution of this thesis is the complete modeling of the video Quality of Service (QoS) in standard and more particularly MPEG-2 applications. The performance of Forward Error Control (FEC) algorithms is first analyzed in terms of the Packet Loss Ratio and Average Burst Length of video data after error recovery. This study leads to a mapping of the network packet loss into video degradation. Along with the analysis of the source rate-distortion characteristic of the video coder, it results in a set of equations that completely model the video quality of service as a function of the scene content and the transmission conditions. An optimal rate distribution between source and channel coding is finally issued from this complete QoS model.

Even with an optimal channel rate, remaining losses may still generate severe damage to the decoded video. All current video standards are based on motion estimation techniques and differential coding which make them really vulnerable to data loss. To avoid error propagation within the displayed sequence, a new adaptive structuring and protection algorithm is applied to MPEG-2 video coding. This scheme basically triggers the insertion of syntactical redundancy (i.e., Intra-coded macroblocks and slice headers) to stop the spreading of data loss within the hierarchical video stream. Redundancy is thus inserted where most needed trading-off compression efficiency against error resilience. The most important video packets are additionally protected by FEC packets to avoid the loss of critical data. This complete error resilient coding scheme is shown to outperform other standard-based robust algorithms.

Although video standards are necessary to the deployment of video applications, they do not provide sufficient robustness to unavoidable packet losses. Moreover, they truly lack a characteristic considered as mandatory in nowadays heterogeneous framework, that is an efficient and fine granularity multi-resolution coding. To cope with these limitations, a new way of coding video based on Matching Pursuit is proposed in the second part of this work. The sequence is decomposed into a series of spatio-temporal functions from a redundant structured dictionary. A new formulation of the structural redundancy is proposed to evaluate the performance of non-orthogonal bases in signal decompositions. Several new spatio-temporal dictionaries are then proposed and compared in terms of compression efficiency and decoding complexity which is an important issue for heterogeneous decoding devices. A tentative comparison with wavelets and video standard coding shows that Matching Pursuit coding is at least as efficient in terms of compression efficiency and even more in low rates. Additionally Matching Pursuit offers intrinsic robustness and multi-resolution properties.

The properties of Matching Pursuit streams are then used in the design of a specific a posteriori quantization scheme. This novel algorithm only depends on the dictionary characteristics and

offers significant improvements over classical coefficient quantization methods. It is shown also that index quantization should be preferably avoided, hence emphasizing the need for a careful design of the dictionary as a compromise between compression efficiency and coding rate. Finally, the lossy and heterogeneous Matching Pursuit streams delivery is studied from the viewpoint of transcoding and error control. A new transcoding scheme is proposed as a simple algorithm to adapt structured Matching Pursuit streams to the resolution of the receivers. An optimal Unequal Error Protection scheme is lastly proposed to fight against packet losses. The end-to-end error control may however become relatively complex in very disparate environments, so that a novel Equivalent Error Packetization is designed to ease the video distribution management. All these different building blocks show that video Matching Pursuit coding constitutes a very promising alternative to classical schemes in lossy and heterogeneous environments.

Version Abrégée

Avec le développement conjoint des technologies réseaux et de codage numérique, le multimédia et plus particulièrement les services vidéo deviennent clairement un des principaux utilisateurs des nouveaux réseaux de l'information. La croissance rapide de l'Internet et de l'industrie informatique a abouti à une infrastructure très hétérogène et souvent surchargée. Dans ce contexte les fournisseurs de services vidéo doivent offrir à leurs clients la meilleure qualité possible en fonction de leurs caractéristiques respectives et de l'état des canaux de communication. La Qualité de Service est non seulement influencée par les artefacts due à la compression de la vidéo, mais aussi par les inévitables pertes de paquets. Le flux vidéo doit donc répondre à des contraintes éventuellement contradictoires que sont la compression et la robustesse aux pertes de paquets.

La première contribution de cette thèse est une modélisation complète de la qualité de service vidéo dans les applications basées sur le standard MPEG-2. Les performances d'algorithmes de correction d'erreur (FEC) sont tout d'abord analysées du point de vue du taux et de la longueur moyenne des pertes. Cette étude mène à une transcription des pertes réseau en termes de dégradation vidéo. Avec l'analyse des caractéristiques de débit-distorsion du codeur vidéo, ceci conduit à un ensemble d'équations qui modélisent complètement la qualité de service vidéo en fonction du contenu des scènes et des conditions de transmission. Un algorithme optimal de partage de débit entre le débit de source et le débit de canal est finalement proposé à l'aide du modèle de la qualité vidéo.

Même avec un débit de canal optimal, les pertes restantes peuvent générer d'importants dommages à la vidéo décodée. Tous les standards de compression vidéo sont basés sur des techniques d'estimation de mouvement et de codage différentiel qui les rendent très vulnérables aux pertes. Pour éviter que les erreurs ne se propagent dans les séquences, un nouvel algorithme de structuration et protection adaptative du flux vidéo est appliqué au codage MPEG-2. Cette méthode module l'insertion de redondance syntaxique (par exemple des blocs codés indépendamment ou des entêtes de slices) pour arrêter la diffusion des pertes à l'intérieur du flux vidéo. De la redondance est donc introduite là où elle est le plus nécessaire pour atteindre un compromis entre efficacité de compression et robustesse aux pertes. Les paquets les plus importants sont en plus protégés par des paquets de correction d'erreur pour éviter la perte de données critiques. Ce schéma complet de codage robuste aux pertes se montre beaucoup plus performant que les autres algorithmes basés sur les standards de compression.

Bien que les standards soient nécessaires au déploiement des applications vidéo, ils n'offrent pas suffisamment de robustesse contre les inévitables pertes de paquets. De plus, ils manquent clairement d'une caractéristique considérée comme essentielle dans le contexte hétérogène actuel, c'est-à-dire un codage à plusieurs résolutions qui soit efficace et de granularité fine. Pour lutter contre ces limitations, une nouvelle façon de coder la vidéo basée sur un algorithme de Matching Pursuit est proposée dans la seconde partie de ce travail. La séquence vidéo y est décomposée en une série de fonctions spatio-temporelles issues d'un dictionnaire structuré et redondant. Une nouvelle formulation de la redondance structurelle est établie pour évaluer les performances des bases non-orthogonales dans la décomposition de signaux. Plusieurs nouveaux dictionnaires spatio-temporels sont ensuite proposés et comparés en terme d'efficacité de codage, mais aussi de

complexité de décodage. Ce dernier critère devient d'ailleurs particulièrement important dans un contexte complètement hétérogène. Une comparaison est proposée entre des représentations basées sur le Matching Pursuit et sur les wavelets et les standards vidéo. Elle montre que le Matching Pursuit est tout aussi efficace, voire même meilleur à bas débit du point de vue de la compression. En même temps, le Matching Pursuit offrent des propriétés intrinsèques supplémentaires de robustesse et de généralité.

Les propriétés des flux Matching Pursuit sont ensuite utilisées pour la construction d'un algorithme spécifique de quantification à posteriori. Ce nouvel algorithme ne dépend que des caractéristiques du dictionnaire et offre des améliorations significatives par rapport aux méthodes de quantification classiques. La quantification des indexes doit par contre être de préférence évitée, ce qui était la nécessité d'une construction soignée du dictionnaire de codage en respectant un compromis entre efficacité de compression et taille des descriptions. Finalement, la distribution du flux Matching Pursuit dans un environnement générique et à pertes est étudiée du point de vue du transcodage et du contrôle d'erreur. Une nouvelle méthode de transcodage est présentée comme une façon simple d'adapter les flux structurés à la résolution dont dispose les clients. Un algorithme de protection inégale contre les erreurs est ensuite proposé pour lutter contre les pertes. Par contre, le contrôle d'erreur de bout en bout de la connexion peut dans ce cas devenir relativement complexe dans un environnement très disparate. C'est pourquoi un nouvel algorithme de paquetisation équivalente de l'énergie est finalement étudié pour faciliter le contrôle de la connexion vidéo. Tous les blocs de base proposés montrent que le codage vidéo basé sur le Matching Pursuit constitue une alternative très prometteuse aux algorithmes standards dans des environnements hétérogènes et à pertes.

Contents

1	Introduction	1
1.1	Motivations	1
1.1.1	General framework	1
1.1.2	Digital compression	2
1.1.3	Digital transmission and packet losses	3
1.2	Problem formulation	3
1.2.1	The need for joint source/channel coding	3
1.2.2	The need for scalable video delivery	4
1.3	Contribution of the thesis	5
1.4	Thesis structure	6
2	Video Communications Technologies	9
2.1	Introduction	9
2.2	Major video compression schemes	10
2.2.1	Video compression standards	10
2.2.2	MPEG-2 video standard	11
2.2.3	Other video coding schemes	12
2.3	Major networking technologies	13
2.3.1	ATM networks	13
2.3.2	IP-based networks	14
2.3.3	Access technologies	16
2.4	User-oriented QoS in video communications	17
2.4.1	What may affect the QoS ?	17
2.4.2	How to measure the video quality ?	18
2.5	Error resilience in video communications	19
2.5.1	Robust video encoding	19
2.5.2	Video error concealment techniques	19
2.5.3	Channel coding	20
2.5.4	Joint source channel coding	21
2.5.5	Multiple Descriptions Coding	21
2.6	Multiresolution coding and scalable video delivery	22
2.6.1	Multiresolution coding	22
2.6.2	Multicast and error control	23
2.6.3	Transcoding and filtering	23
2.7	Internet video: an overview	24
2.8	Conclusion	24

3	FEC Performance in Video Streaming	27
3.1	Introduction	27
3.2	Loss process after FEC recovery	28
3.3	FEC in a Gilbert-model loss process	31
3.4	Conclusions	33
4	Joint MPEG-2/FEC Rate Selection	35
4.1	Introduction	35
4.2	MPEG-2 perceptual distortion-rate function	36
4.3	Perceptual distortion under different loss patterns	38
4.4	Joint source/channel perceptual distortion	42
4.5	Experimental results	44
4.5.1	Optimal FEC parameters	44
4.5.2	End-to-end quality improvement	44
4.6	Conclusions	44
5	AMISP: Content-Based MPEG-2 Encoding	47
5.1	Introduction	47
5.2	Structuring and protection: problem formulation	48
5.2.1	MPEG-2 TM5 framework	48
5.2.2	Extra resynchronization points	48
5.2.3	Problem formulation	49
5.3	Adaptive MPEG-2 structuring	49
5.3.1	Loss probability matrices	49
5.3.2	Proposed algorithm: AMIS	52
5.3.3	Experimental results and comparisons	54
5.4	Adaptive MPEG-2 structuring and protection	55
5.4.1	FEC-based protection	55
5.4.2	The adaptive protection algorithm: AMISP	57
5.4.3	Experimental results and comparisons	58
5.5	Conclusions	59
6	From Standards to Matching Pursuit Codecs	61
6.1	Yet another codec ?	61
6.2	Matching Pursuit overview	63
6.3	Group theoretical design of dictionaries	65
6.3.1	Group representations	65
6.3.2	Group theoretical design of dictionaries	66
6.3.3	Semi-structured dictionaries	67
6.4	Structural redundancy in non-orthogonal transforms	67
6.4.1	New structural redundancy formulation	68
6.4.2	Experimental results	70
6.4.3	Uniformly distributed atoms	71
6.4.4	Structural redundancy in Matching Pursuit	72
6.5	Matching Pursuit complexity	73
6.5.1	Encoding complexity	73
6.5.2	Decoding complexity	74
6.6	Conclusion	74

7	3D Matching Pursuit Video Coding	75
7.1	3D video coding	75
7.2	3D dictionaries	77
7.2.1	Structured dictionaries	77
7.2.2	Separation of spatial and temporal functions	77
7.2.3	Complexity considerations	77
7.3	Spatial dictionaries	78
7.3.1	Gabor atoms	78
7.3.2	Anisotropically refined gaussians	79
7.3.3	B-spline functions	80
7.3.4	Rectangular functions	81
7.3.5	Comparison of spatial dictionaries	82
7.3.6	Comparison with wavelet coding	83
7.3.7	Discussion	85
7.4	Spatio-temporal atoms	85
7.4.1	Gaussian function	85
7.4.2	B-spline function	86
7.4.3	Rectangular function	86
7.4.4	Haar function	87
7.4.5	Comparison of temporal functions	88
7.4.6	Comparison with coding standards	90
7.5	Discussion and improvements	91
7.5.1	Size of the Group of Frames	91
7.5.2	Local motion estimation	93
7.5.3	Adaptive dictionaries	93
7.6	Conclusion	94
8	Redundancy-Driven Quantization of MP stream	95
8.1	Why quantization ?	95
8.2	Signal reconstruction error	96
8.3	Quantization of Matching Pursuit coefficients	97
8.3.1	A posteriori coefficients quantization	97
8.3.2	Coefficient norm decay	98
8.3.3	Redundancy-driven uniform coefficients quantization	98
8.3.4	Experimental results and discussion	101
8.4	Quantization of structured atoms indexes	104
8.4.1	A posteriori index quantization	104
8.4.2	Bounds on the atom selection error	104
8.4.3	Gabor atom selection error	105
8.4.4	Scalar quantization of structured atom indexes	108
8.4.5	Redundancy-driven Gabor indexes quantization	110
8.5	Conclusions	113
9	Multi-Resolution Lossy MP Delivery	115
9.1	Introduction	115
9.2	Matching Pursuit transcoding	116
9.2.1	Heterogeneous Matching Pursuit delivery	116
9.2.2	Transcoding of MP streams	117
9.2.3	Experimental results	118
9.2.4	Discussion	119
9.3	Error control	120
9.3.1	Unequal Error Protection	122
9.3.2	Equivalent Energy Packetization	125
9.3.3	Experimental comparisons	128

9.3.4 Discussion	130
9.4 Conclusion	131
10 Conclusions	133
A Optimal Number of MP Iterations	137
Resume	169
Personal Bibliography	171

List of Figures

1.1	End-to-end video transmission system.	1
1.2	Contradictory system design requirements of video services.	2
1.3	Multiresolution approach.	4
2.1	Convergence of multimedia enabling sectors.	9
2.2	Building blocks of generic compression schemes.	10
2.3	Block diagram of an encoder.	11
2.4	MPEG-2 video structure.	11
2.5	Multimedia services over ATM protocol stack.	14
2.6	H.222.1 overview.	15
2.7	IP datagram format.	15
2.8	RTP packet format.	16
2.9	Data loss propagation in MPEG-2 video streams.	18
2.10	Error concealment techniques.	20
2.11	Multiple Description Coding. General representation with two descriptions.	22
3.1	Media-independent FEC scheme.	28
3.2	Two-state Markov chain: Gilbert model.	31
3.3	Evolution of π_v versus the number of video packets k in a FEC block of length $n = 20$ in a Gilbert-model loss process.	32
3.4	Evolution of α_v versus the number of video packets k in a FEC block of length $n = 20$ in a Gilbert-model loss process.	32
4.1	Perceptual distortion as a function of the quantizer scale MQANT for ITU-R 601 test sequence.	36
4.2	Quality versus quantizer scale. (a) PSNR versus quantizer scale factor for 3 different scenes. (b) MPQM versus quantizer scale factor for 2 different scenes.	37
4.3	Average output encoding bit rate versus quantizer scale factor (MQANT) for ITU-R 601 test sequence. Exponential fitting curve.	38
4.4	Perceptual distortion as a function of the mean encoding bit rate. Fitting parameters of Eq. (4.2.1): $\chi_S = 14.59 \cdot 10^6$ and $\xi_S = -0.883$	38
4.5	Perceptual distortion versus percentage of spatially lost pixels for several loss patterns (source rate $R_S \in \{4.85, 2.35\}$ Mbps).	39
4.6	Spatial error propagation under $\alpha_v = \{1, 2\}$	39
4.7	Number of damaged pixels versus the impact byte in the stream.	40
4.8	Number of lost pixels versus the Average Burst Length α_v and the Packet Loss Ratio π_v (source rate $R_S = 4.75$ Mbit/s).	41
4.9	Perceptual distortion versus the video loss process parameters (source rate $R_S = 4.75$ Mbit/s)	41

4.10	(a) Q versus average bit rate and PLR: Simulations on the <i>Barcelona</i> sequence, without FEC protection. (b) Q versus average bit rate and PLR: Fitting function from Eq. (4.4.3) where $k = n$	43
4.11	Perceptual distortion (a) and PSNR (b) vs frame number for optimal rate distribution scheme and classical FEC schemes. The sequence is composed of 5 different scenes. The available rate R is set to 4 Mbps for the 204 first frames, and to 6 Mbps for the 196 last frames. The global loss process parameters are set to $(PLR, ABL) = (.1, 1)$	45
4.12	Perceptual distortion (a) and PSNR (b) vs frame number for optimal rate distribution scheme and classical FEC schemes. The sequence is composed of 5 different scenes. The available rate R is set to 4 Mbps for the 204 first frames, and to 6 Mbps for the 196 last frames. The global loss process parameters are set to $(PLR, ABL) = (.01, 1)$	46
5.1	End-to-end PSNR quality versus extra resynchronization points ratio for the Football scene (CBR encoding at 5 Mbps and $PLR = 10^{-2}$).	49
5.2	Illustration of $\mathbf{P}_n(i, j)$. The numbers in each macroblock represent $M_n + N_n$ values.	50
5.3	Mapping function $\mathcal{M}_n$ of loss probability matrices.	51
5.4	PSNR end-to-end quality versus PLR in comparison to random resynchronization points insertion, TM-5 encoding scheme and encoding algorithm proposed by Richardson (CBR encoding at 6 Mbps).	55
5.5	AMIS end-to-end quality versus PLR in comparison to random resynchronization points insertion, TM-5 encoding scheme and encoding algorithm proposed by Richardson (CBR encoding at 6 Mbps).	55
5.6	Sample frame of ITU-R test sequence, where AMISP additional slice headers are designed in black.	56
5.7	Sample frame of ITU-R test sequence, where AMISP additional Intra-macroblocks are designed in light gray.	56
5.8	End-to-end PSNR quality versus the packet loss ratio. Comparison of the AMISP algorithm ($k_{FEC} = 10$ and $n_{FEC} = 11$) with a TM-5 encoding with an adaptive and regular FEC scheme ($k = 10$ and $n = 11$). The total bit rate is about 6 Mbps.	58
5.9	End-to-end perceptual quality versus the packet loss ratio. Comparison of the AMISP algorithm ($k_{FEC} = 10$ and $n_{FEC} = 11$) with a TM-5 encoding with an adaptive and regular FEC scheme ($k = 10$ and $n = 11$). The total bit rate is about 6 Mbps.	59
6.1	Heterogeneous and error-prone video distribution environment. Each device D_i may have different resolution features and each path has its own characteristics in terms of rate R_i and loss L_i process.	61
6.2	Representation of the projection neighborhood $\mathcal{V}_{\gamma_i}$ in $\mathbb{R}^2$	68
6.3	Representation of the projection neighborhood $\mathcal{V}_{\gamma_i}$ in $\mathbb{R}^3$	69
6.4	The orthogonal projection of φ_i onto the intersection $H_1 \cap H_2$ is equivalent to the residue of the orthogonal projection of φ_i onto the orthogonal complement to the intersection $H_1 \cap H_2$. This representation in $\mathbb{R}^3$ generalizes to n-dimensional case.	70
6.5	Redundancy factor β as a function of the number of vectors N_v in a random dictionary. Exponential fitting function of the form: $\beta = 1 - A N_v^B$	71
6.6	Redundancy factor β as a function of the ratio between the number of vectors N_v in a random dictionary and the size N of the vectors. Exponential fitting function of the form: $\beta = 1 - A N_v^B$	71
6.7	Redundancy factor β and its approximation through the uniform vectors distribution as a function of the size of a random 5-dimensional vectors dictionary. The values are averaged on 100 different random dictionaries.	72
6.8	Residual energy decay for Matching Pursuit coding of two random signals (f_1, f_2) of length 10 with a dictionary formed of 50 random vectors.	73

6.9	Chromosomes cross-over in Genetic Algorithm.	74
7.1	Tri-dimensional Matching Pursuit scheme	76
7.2	Coordinates changes by translation, scaling and rotation.	78
7.3	Reconstructed version of <i>Lena</i> encoded with 500 Gabor atoms where $(\omega_0, \sigma) = (6, 2)$	79
7.4	Reconstructed version of <i>Lena</i> encoded with 500 anisotropically refined atoms.	80
7.5	Reconstructed version of <i>Lena</i> encoded with 500 B-spline atoms.	81
7.6	Reconstructed version of <i>Lena</i> encoded with 500 rectangular atoms.	82
7.7	Distortion of the reconstructed <i>Lena</i> image as a function of the number of MP iterations.	82
7.8	Distortion of the reconstructed <i>Autumn</i> image as a function of the number of MP iterations	83
7.9	Comparison of MP decomposition vs 4-level wavelet decomposition of the <i>Lena</i> image.	84
7.10	PSNR quality of MP decomposition with differential coefficient coding and SPIHT coding as a function of the compression ratio.	84
7.11	Reconstruction of the four first frames of <i>Foreman</i> encoded with 500 iterations of 2D B-splines and gaussian temporal function.	86
7.12	Reconstruction of the four first frames of <i>Foreman</i> encoded with 500 iterations of 3D B-splines function.	87
7.13	Reconstruction of the four first frames of <i>Foreman</i> encoded with 500 iterations of 2D B-splines and rectangular temporal function.	88
7.14	Reconstruction of the four first frames of <i>Foreman</i> encoded with 500 iterations of 2D B-splines and Haar temporal function.	89
7.15	Reconstruction distortion as a function of the number of iterations per group of four frames of <i>Foreman</i> . The atoms are made on 2D B-splines and four different temporal function.	89
7.16	PSNR quality of MP decomposition with differential coefficient coding and H.263+ and MPEG-2 as a function of the encoding rate for the <i>Foreman</i> sequence.	90
7.17	Reconstruction of the four first frames of <i>Foreman</i> encoded with 800 2D B-splines and temporal gaussian atoms.	91
7.18	Reconstruction of the four first frames of <i>Foreman</i> encoded with H.263+ at a quantizer scale of 31.	92
7.19	Reconstructed distortion of <i>Foreman</i> as of function of the coding rate for 3D splines atoms and GOF size of 4 and 8.	92
7.20	Comparison of the encoding quality for oriented and non-oriented 3D B-splines for the <i>Foreman</i> sequence.	93
8.1	Comparison of <i>a priori</i> and <i>a posteriori</i> uniform coefficient quantization (8 bits) for the <i>Lena</i> image.	97
8.2	Coefficient norm versus the iteration number. The MP coefficients using a 2D Gabor dictionary are represented for the <i>Lena</i> and <i>Autumn</i> images.	99
8.3	R-D curve for the decomposition of a random signal of length 10 over a random dictionary of 50 vectors. Comparison of exponential quantization and uniform quantization as a function of the total rate.	102
8.4	Rate-distortion curves for the new quantization scheme and uniform a posteriori quantization of <i>Lena</i> MP stream. Distortion is represented as a function of the total rate.	102
8.5	Reconstruction of the <i>Lena</i> image with Gabor atoms with exponentially upper-bounded quantization. PSNR = 26.7 dB, Rate = 16.4 kbit (without entropy coding).	103
8.6	Reconstruction of the <i>Lena</i> image with Gabor atoms with a posteriori coefficients uniform 16-bits quantization. PSNR = 25.8 dB, Rate = 16.4 kbit (without entropy coding).	103

8.7	Reconstruction of the <i>Lena</i> image with Gabor atoms with a priori coefficients uniform 16-bits quantization. PSNR = 25.7 dB, Rate = 16.4 kbit (without entropy coding).	103
8.8	Gabor atom selection error as a function of a and θ , respectively the scale and rotation part of the error on indexes.	107
8.9	Gabor atom selection error as a function of a and b , respectively the scale and position part of the error on indexes.	107
8.10	Gabor atom selection error as a function of b and θ , respectively the position and rotation part of the error on indexes.	107
8.11	Reconstruction of the <i>Lena</i> image with Gabor atoms after 1000 MP iterations (PSNR = 31.17 dB).	108
8.12	Reconstruction of the <i>Lena</i> image with Gabor atoms, where all positions indexes have been quantized (PSNR = 22.35 dB). The number of positions indexes has been divided by two compared to the encoding dictionary.	108
8.13	Reconstruction of the <i>Lena</i> image with Gabor atoms, where all rotation indexes have been quantized (PSNR = 27.72 dB). The number of rotation indexes has been divided by two compared to the encoding dictionary.	108
8.14	Index rate-distortion characteristic of uniform <i>a posteriori</i> rotation index quantization.	109
8.15	Index rate-distortion characteristic of uniform <i>a posteriori</i> position index quantization.	110
8.16	Evolution of the reconstruction error of <i>Lena</i> for Gabor sub-dictionaries and index <i>a posteriori</i> quantization. Curves taken by pairs have the same dictionary size.	110
8.17	Rate-distortion representation between size of the Gabor dictionary index and MSE distortion after 1000 iterations on the <i>Lena</i> 128×128 image.	111
8.18	Rate-distortion representation between size of the Gabor index after quantization and MSE distortion after 1000 iterations on the <i>Lena</i> 128×128 image.	111
8.19	Convergence of the reconstruction error of the <i>Lena</i> image for different Gabor dictionaries of equivalent coding size.	112
9.1	Multiresolution Streaming Application. The input stream may be delivered, through video gateways, to different resolutions devices.	116
9.2	Reconstruction of the <i>Foreman</i> sequence encoded with 500 3D-splines. The spatial resolution is divided by 2 ($\alpha = \frac{1}{2}$) and the temporal resolution is divided by $\frac{4}{3}$ ($\eta = \frac{3}{4}$) by comparison to the CIF coding format.	118
9.3	Wavelet coded <i>Lena</i> image of size 256×256 with the 1500 most significant coefficients decoded at a resolution of 128 by 128 pixels (PSNR = 24.98 dB).	120
9.4	Matching Pursuit coded <i>Lena</i> image of size 256×256 with 500 iterations decoded at a resolution of 128 by 128 pixels (PSNR = 25.42 dB).	120
9.5	Serial MP Packetization scheme.	122
9.6	Energy distribution between successive packets. 500 MP iterations of <i>Foreman</i> sequence have been serially packetized in 20 packets. The energy per packet is compared to the theoretical energy obtained from Eq. (9.3.1).	123
9.7	Unequal Energy Packetization scheme.	123
9.8	Generic UEP for N packets constituting B FEC blocks.	124
9.9	Equal Error Packetization scheme.	126
9.10	Energy distribution between successive packets. 500 MP iterations of <i>Foreman</i> sequence have been serially packetized in 20 packets. The energy per packet is compared to the theoretical energy obtained from Eq. (9.3.5).	127
9.11	Evolution of the expected energy as a function of the channel packet loss ratio ($ABL = 2$).	128
9.12	First four frames of <i>Foreman</i> after the loss of the first packet in UEP scheme.	129
9.13	First four frames of <i>Foreman</i> after the loss of the first packet in EEP scheme.	130

List of Tables

4.1	Optimal FEC parameters for several transmissions conditions and scenes.	44
9.1	PSNR quality of the decoded <i>Foreman</i> sequence after MP transcoding to lower spatio-temporal resolutions.	119
9.2	Optimal UEP parameters for the transmission of <i>Foreman</i> under different packet loss ratio and average burst length (π, α)	125
9.3	Optimal EEP FEC parameters for the transmission of <i>Foreman</i> under different packet loss ratio and average burst length (π, α)	128

Introduction

1.1 Motivations

1.1.1 General framework

Telecommunications are clearly becoming one of today's driving forces. Market deregulation and the strategical need for information have fueled the development of communication services and applications. Among these, multimedia applications are gaining momentum fostered by web technology and reach now various markets. The development of corporate intranets opens the door to video conferencing, distributed cooperative work and other multimedia-based services such as distant training and repairing. Residential access via technologies such as xDSL and cable TV is being pushed by the entertainment and the personal computer industries. The perspective of broadband access to the internet and a large spectrum of on-demand services lets foresee mass markets still to be exploited. Meanwhile, wireless communications market is growing tremendously, opening new perspectives to service providers. This new market clearly responds to the need for mobility even at the price of lower QoS. The creation of large companies resulting from the merging of communications and entertainment enterprises illustrates the strategic importance of business, residential and mobile customers.

Audiovisual applications (e.g., video conferencing, video on demand, teleteaching, WebTV, etc.) are foreseen as one of the major consumers of these new information networks. At the heart of this revolution is the digital compression of audio and video signals. The purpose of compression is data rate reduction and thus lower transmission costs. Since common networking technologies are packet-oriented, packet video streaming is becoming the main issue in these multimedia applications. However, today's Internet is not naturally suitable for this type of traffic. Delays and network congestion can indeed induce information losses and hence severe damage on user perception.

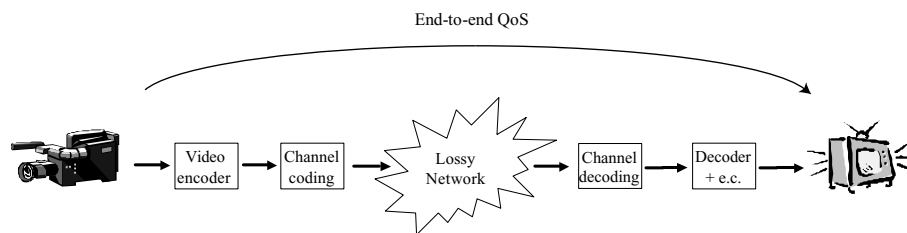


FIGURE 1.1: *End-to-end video transmission system.*

The Quality of Service (QoS) offered to the end-users therefore depends not only on distortion possibly caused by video compression but also on transmission losses and delay. A generic end-

to-end video transmission system is represented in Fig. 1.1. The video is first compressed to meet bandwidth constraints of the transmission channel. The video stream is then channel encoded to fight against transmission losses. This operation aims at recovering erasures or at least limiting loss impact in the reconstructed video. After transmission, the video stream is submitted to the inverse operations, which are channel decoding and source decoding. Error concealment can finally be performed to hide remaining losses before display.

Obviously, customers always ask for the best possible QoS at the lowest price. However, providers have to spare this expensive resource that is bandwidth and therefore highly compress video sequences to serve as many clients as possible. However, the higher the compression, the lower the quality and, in the same time, the larger the vulnerability to errors occurring during the video transmission. Some redundancy shall therefore be left intentionally in the bit stream, so that losses can be recovered or hidden to offer the end-user a reasonable QoS. The total bit rate should judiciously be split between source coding (i.e., video compression) and channel coding (i.e., protection against transmission errors) to provide the customer with the best possible QoS.

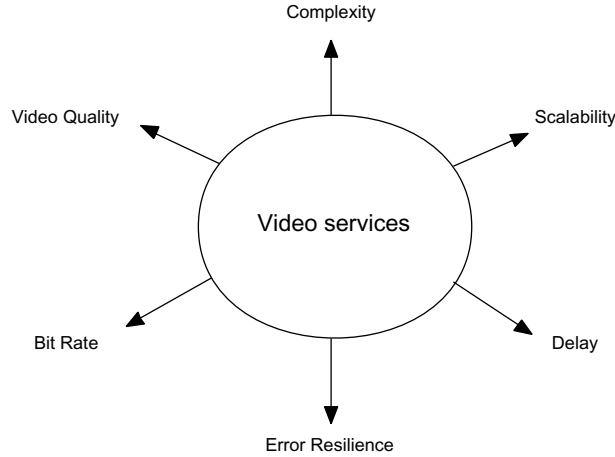


FIGURE 1.2: *Contradictory system design requirements of video services.*

Therefore, end-to-end system design has clearly to fulfill several possibly contradictory requirements (see Fig. 1.2). First, video compression reduces not only the source rate but generally also video quality and resilience to packet losses. Second, the most efficient but often complex coding schemes may induce coding delays or too expensive decoding operations depending on the receiver's capabilities. Finally, the most important issue in today's heterogeneous framework is certainly scalability. The same video stream has to serve different clients, which have their own characteristics in term of available rate, display resolution or device computing power. The need for scalability often deserves compression efficiency since the same stream may not be optimal simultaneously for all heterogeneous clients. A lot of work remains therefore to be done in order to optimize video end-to-end QoS for each particular user in a highly disparate and lossy infrastructure.

1.1.2 Digital compression

Since the available rates are truly limited, the source description has to be kept close to minimal. Many compression schemes have been studied, but standards are clearly necessary for widespread video communications. The choice of the compression algorithm mostly depends on the available bandwidth or storage capacity and the features required by the application. For broadband digital video, attention is now being focused on the MPEG-2 standard. The MPEG-2¹ standard [1, 2] is capable of compressing NTSC or PAL video into an average bit rate of 3 to 6 Mbits/s with

¹MPEG stands for Moving Picture Experts Group

a quality comparable to analog CATV² [3]. MPEG-2 is destined to diverse applications such as television broadcast over satellite, cable and other broadcast channels (DVB), and digital storage media (DVD). For lower rate applications, new standards have been recently defined, like H.263+ and MPEG-4. These standards propose additional features to limit packet loss degradation which was not really addressed in the design of MPEG-2. However, even with robust coding modes, video standards are vulnerable to packet losses due to their hierarchical structure.

Also, current coding standards clearly lack a feature that is becoming increasingly important in the emerging world of heterogeneous packet networks and wireless communications, which is fine scalability. They do not allow for an efficient multiresolution coding in terms of both rate and flexibility. Furthermore, they are all based on orthogonal transforms like DCT or wavelets which have shown their limitations in low rate compression and decoding complexity. Clearly the compression issue is not optimally solved in terms of generic scalability and error resilience. The source coding problem remains thus completely open.

1.1.3 Digital transmission and packet losses

The Internet and the IP protocol are becoming the main support for packet video communications. The current networking structure is however loosely defined and completely heterogeneous in terms of bandwidth and reliability. Intranets offer high transmission rates while at the opposite only very small rates may be expected on wireless networks. Nevertheless, both networks can be led to provide the same video content at however possibly different quality. Meanwhile, today's Internet is working in a best-effort policy. The freedom in Internet accessibility is quite laudable but it does not go without introducing management problems. Video streaming on the Internet can therefore encounter severe losses due mainly to overload in the network nodes.

The IP network can a priori drop any packet because of its best-effort policy. At the Internet beginning, the error control used to be performed by the transport layer. Communication link design mainly targeted error-free but possibly delayed data transmission and were clearly ill-designed for video applications. They did not take into account properties specific to video communications, like unequal bit importance, delay requirements or a certain tolerance to channel loss due to redundancy left in the compressed signal. New protocols like RTP have been proposed to enhance existing network capability to support multimedia applications. They provide timing and synchronization mechanisms but generally leave the error control to the application level. The video applications have therefore to be carefully designed to cope with the existing network infrastructure. Error control has to be incorporated in the upper-layers of the end-to-end system design, while timing control may be left to the transport or RTP protocols.

1.2 Problem formulation

In video streaming applications the problem is basically to provide the user with the best possible video quality for a given available bit rate. These applications have however to face two major problems on today's network and computers infrastructure. First, video streams encounter unavoidable packet loss during transmission. Second, it has to deal with a very heterogeneous set of decoding devices, which can make use of different effective bandwidth. The video applications has thus to meet possibly contradictory constraints to provide the client with the best possible quality.

1.2.1 The need for joint source/channel coding

The only way to apprehend the video transmission problem is to consider the end-to-end process as a whole, from encoding to display at destination [4, 5, 6]. The Quality of Service optimization has to be achieved by a judicious joint source and channel coding. Indeed, the separation theorem of Shannon [7, 8, 9] is justifiable only in the limit of an arbitrary encoding complexity and arbitrary

²Cable television

long code block lengths. However, these requirements are obviously not met in video communications. The latter indeed require timing constraints and low decoding complexity, especially in mobile applications. The source and channel encoders become therefore dependent on each other in practical cases.

Practical video coders indeed produce bits which are more important than others for the video reconstruction. This is mainly due to the non-stationary nature of video sequences and to the coding scheme properties. The unequal meaning of video data has obviously to be considered in the design of both the source and channel coding. Hence, a smart coder allocates most of the coding rate to the most significant video information. Similarly, an optimal error protection scheme tries to avoid the loss of the most important bits. A joint source/channel encoding scheme can benefit from the knowledge of the relative data importance to weight their protection. Similarly, it has to be aware of the channel status to efficiently distribute the overhead where most needed. Under fixed bit budget constraint, increasing the protection indeed imposes a decrease of the bit rate allocated to pure video information. Hence, there is a clear tradeoff between the video source rate and the amount of protection redundancy. Note that channel redundancy can also be incorporated within the video stream to increase its robustness to losses. Hence, optimal rate distribution necessitates the knowledge of the video content and the channel status. A joint source/channel coding scheme is clearly the only way to ensure the best possible quality for a given video sequence under delay and bandwidth constraints.

1.2.2 The need for scalable video delivery

The diversity of receivers capabilities and effective bit rates is surely one of the major problem in today's network infrastructure. This disparity is due to both the evolution of devices and network links and to the arrival of new services like wireless communications in addition to well-established applications. In this heterogeneous context the video service providers are facing the problem of guaranteeing the best possible QoS to each receiver, does it make use of a PDA or a workstation (see Fig. 1.3). Meanwhile they cannot afford to build a specific video stream for each customer under penalty of bandwidth explosion.

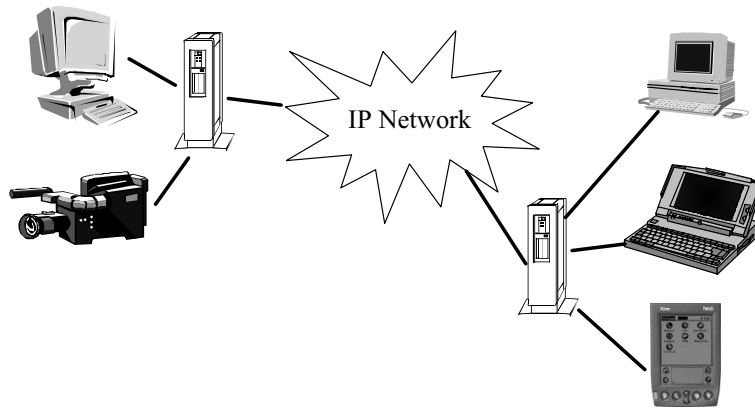


FIGURE 1.3: *Multiresolution approach.*

Scalability has to be offered at different levels in video applications. First, video services have to be proposed at different rates, adapting to the effective bandwidth available to the customer. The video coder should indeed adapt to the network state in terms of bit rate and loss process to guarantee the end-user the best possible video quality. Second, the end devices can also have completely different processing power or spatio-temporal display resolutions. The coder should thus take into account both the decoding device features (size of display, frame rate, etc.) and the video content.

In the same time, scalable coding alone is not sufficient to optimally serve all the customers.

Scalable coding has to be paired with a multi-resolution video distribution scheme capable to adapt the scalable video stream to the receivers requirements. This leads to a new approach in multimedia delivery that can be assimilated to proxy-based signal processing. Intermediate network nodes or gateways are asked to play an active role in the video delivery by performing basic transcoding or filtering operations. Gateways can also perform error control depending on the actual channel status. As the network processing is done locally and close to the end-user, it allows for an efficient utilization of resources in an heterogeneous framework.

1.3 Contribution of the thesis

This thesis mainly targets the joint source and channel coding parts of the end-to-end system presented in Fig. 1.1. In this framework the contributions of this work are manifolds. First, the MPEG-2 video coding process is adapted to the channel status and the video content. Under given rate constraints, bits are optimally distributed between source coding and channel redundancy which comprises MPEG-2 syntactic information and error control overhead. To this aim, this work proposes :

- a new formulation of the FEC performance for video packets recovery,
- a novel complete end-to-end model of the MPEG-2 quality of service, which allows to select the best pair source/channel rate,
- a new error resilient MPEG-2 compliant encoding technique which adapts to the video content and channel conditions to provide a robust video stream.

This joint source and channel coding of MPEG-2 video streams allows to adaptively spend bits where most needed, thus ensuring an optimal bandwidth utilization. The schemes can be easily adapted to other coding standards thanks to the general validity of their design principles.

New ways of coding video are then investigated even though coding standards are clearly mandatory for the development of the video industry. Block-based motion-compensation and DCT coding leads most video standards to a reasonable efficiency, but they lack a true scalability functionality now regarded as essential for emerging multimedia applications. In particular, resolution and fidelity or rate scalability are necessary to upcoming video services. The capability of progressive transmission by increasing resolution and fidelity is indeed considered as essential in the evolution of video applications. Moreover, due to its hierarchical structures, standard coding introduces memory in the coding process and is thus not truly robust to transmission losses. Hence, this work addresses in a second part the design of a new robust and multi-resolution video coder based on Matching Pursuit for lossy and heterogeneous video delivery. This work does however not provide a complete implementation of this new scheme, but mainly focuses on the joint source and channel coding part. To this aim, this second part proposes :

- a new formulation of the structural redundancy in non-orthogonal bases as used by Matching Pursuit. The structural redundancy then allows to quantify the coding efficiency of particular representations,
- a new way of decomposing video signals into a set of spatio-temporal functions through Matching Pursuit based on a Genetic Algorithm. New dictionaries are also proposed that outperform commonly used dictionaries in Matching Pursuit coding,
- a novel quantization scheme taking benefits of the Matching Pursuit stream properties to reach an optimal bit distribution between successive coefficients. A study of the atom selection error shows moreover that the index quantization should be avoided, thus emphasizing the need for a careful dictionary design,
- a new transcoding scheme based on the properties of the structured dictionaries used in coding process. This scheme permits an easy adaptation of the video stream to any spatio-temporal decoding resolution,

- an optimal error control by Unequal Error Protection of the Matching Pursuit stream,
- a novel Equivalent Energy Packetization scheme which permits to adapt the stream to the inherent structure of the Internet transmission and enables a very easy error control management.

From these building blocks, Matching Pursuit already shows to offer very promising perspectives in adaptive and robust multi-resolution coding.

1.4 Thesis structure

This thesis is organized as follows. Chapter 2 presents the state of the art in video communication technologies. A survey of common video coding schemes is proposed, emphasizing both video standards like MPEG-2 and non-standard schemes like wavelets-based coders. Major networking technologies (i.e., ATM and IP networks) are then overviewed, and the factors which influence the video quality of service are identified. Error resilient and multiresolution coding schemes are finally described, for standard and wavelets-based codecs. Finally current video coders implementations are briefly presented focusing on their limitations.

Chapter 3 studies the performance of Forward Error Correction (FEC), the preferred error control scheme to fight against losses in multicast or interactive applications. Performance is defined in terms of both packet loss ratio and average burst length after recovery. These parameters are very useful to correctly design video services and allow to draw a mapping between the Quality of Service at the network and video layers. In particular, along with models of the video quality as a function of the source rate, it allows to optimally distribute the available bandwidth between source and FEC rate, as shown in Chap. 4. We derive the global set of equations that leads to an optimal dynamic rate allocation dependent on the video information and the channel status. The optimal distribution is shown to outperform classical FEC schemes, thanks to its adaptivity features.

A new scene-complexity adaptive error resilient MPEG-2 encoding scheme is presented in Chap. 5, namely Addaptive MPEG-2 Information Structuring (AMIS). AMIS modulates the number of resynchronization points (i.e., slice headers and intra-coded macroblocks) in order to maximize the perceived video quality. It constantly determines the best compromise between the rate allocated to encode pure video information and the rate aiming at reducing the sensitivity to packet loss. Experimental results show that AMIS dramatically outperforms existing structuring techniques thanks to its efficient adaptivity. We then extend AMIS with a Forward Error Correction (FEC) based Protection algorithm to become AMISP. AMISP additionally triggers the insertion of FEC packets in the MPEG-2 video packet stream. The performances of the AMISP scheme in an MPEG-2 over RTP/UDP/IP scenario are also evaluated.

Due the lack of video standards in providing true robustness and fine granularity scalability, a new coding scheme based on Matching Pursuit is introduced in Chap. 6. Matching Pursuit is overviewed and complexity issues are discussed. Non-orthogonal decompositions as proposed by MP are then analyzed and a new formulation of the structural redundancy of dictionaries is proposed. Chap. 7 then presents a novel three-dimensional Matching Pursuit coding scheme as an alternative to classical motion estimation based coders. Several structured dictionaries are proposed and performance of the Matching Pursuit decomposition is studied. In Chap. 8, the properties of Matching Pursuit streams are used to design a specific quantization scheme, driven by the redundancy of the coding dictionary. This scheme is shown to outperform uniform quantization. The atom selection error is also analyzed, and indexes are shown to be highly sensitive to quantization error. The importance of an optimal dictionary design is thus emphasized and presented as a clear trade-off between compression efficiency and index transmission cost. Then Chap. 9 focuses on the heterogeneous video stream delivery. More particularly, a transcoding scheme is proposed to easily adapt the spatio-temporal video resolution to the receiver characteristics. Error control by Unequal Error Protection is also analyzed to cope with transmission losses. A new packetization scheme is also presented which builds equivalent energy packets to provide

more flexibility in the error control management. The robustness of Matching Pursuit stream in error-prone environment is also demonstrated in this context. Finally, concluding remarks are given in Chap. 10. Future works are also discussed with focus on the complete end-to-end design of Matching Pursuit based coding scheme.

Video Communications Technologies

2.1 Introduction

Since this work targets the end-to-end video quality of service optimization many topics are addressed within the thesis. Multimedia and more particularly video applications can be placed in the intersection of traditionally separated domains as depicted in Fig. 2.1. Each of these domains has its proper state of the art. For the sake of clarity, this chapter only presents the main parts of complete video transmission systems. More specific topics are further presented within the following chapters.

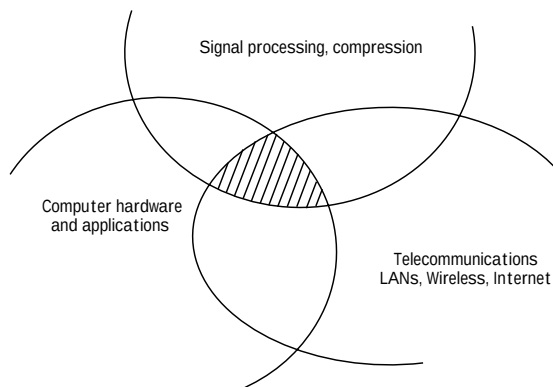


FIGURE 2.1: *Convergence of multimedia enabling sectors.*

This chapter first presents the major video coding methods, focusing on both standard compression algorithms and more recently defined schemes. IP and ATM networking technologies are then presented as the basis infrastructure of packet video transmission. The effects of unavoidable packet losses onto the video stream are further described. The most common error resilience schemes are then briefly presented as a way to fight against packet erasures. Finally, the major commercial video streaming applications are overviewed, showing the limitations of these implementations to offer robust and multiresolution video stream.

Parts of this work appeared in [10, 11].

2.2 Major video compression schemes

Video sequences naturally consist in very redundant information and the size of digital video signal representation is originally very large. For example, the data rate of an uncompressed NTSC¹ TV-resolution video stream is close to 170 MBits/s, which corresponds to less than 30 seconds of recording time on a regular compact disk (CD). Clearly, the video signal has thus to be compressed to reduce the needs for storage or bandwidth capacity. The compression basically consists in removing the inherent video redundancy and efficiently coding the remaining information.

The video compression schemes are generally composed of several building blocks as shown in Figure 2.2. First, the digital signal is transformed and represented as a sum of basis functions. In other words, the series of pixels values output from the recording camera are represented in a compact form to enable information compression. The transformed signal is then quantized [12] to limit the number of allowable values and thus the size of the description. Note that quantization is the only irreversible and lossy task in common compression schemes. The quantized signal is finally entropy-coded [7] to minimize the size of the description.

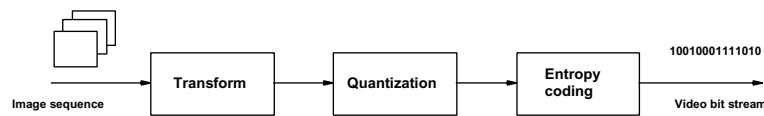


FIGURE 2.2: *Building blocks of generic compression schemes.*

The remaining of this section presents the video compression standards and focuses particularly on MPEG-2. All common video standards indeed proceeds similarly to MPEG-2 in their main principles. Other video compression technologies, and mainly wavelets coders are then presented as an alternative to DCT-based video coding standards.

2.2.1 Video compression standards

The choice of a compression standard mostly depends on the available transmission or storage capacity as well as the features required by the application. The most cited algorithms are the H.261, H.263, H.263+, MPEG-1, MPEG-2 and MPEG-4² [13] video standards. All these standards are based on discrete cosine transform (DCT) and motion prediction (see Fig. 2.3). The DCT allows for a more compact representation of the naturally redundant spatial information, while the motion compensation aims at reducing the temporal redundancy between successive images. The above standards however target different type of video applications (i.e. encoding rates and qualities). These applications range from desktop video-conferencing to TV channels broadcast over satellite, cable, and other channels [14]. The former typically uses H.261 or H.263 while MPEG-2 is the most appropriate compression standard for video broadcast applications [15]. The recent MPEG-4 standard proposes moreover additional features [16] which allows for object-oriented coding of video scenes.

As an example, the MPEG-2 standard is now described more in details. MPEG-2 is an audio-visual standard developed by the International Organization for Standards (ISO) together with the International Electrotechnical Commission (IEC) [1, 2, 17, 18]. The video part of MPEG-2 permits data rates up to 100 Mbps and proposes a number of advanced features, including those supporting HDTV³ coding. MPEG-2 is capable of compressing NTSC or PAL TV-resolution video into an average bit rate of 3 to 7 Mbps with a quality comparable to analog broadcast TV [19, 3].

¹NTSC stands for National Television Systems Committee

²MPEG stands for Moving Picture Experts Group

³HDTV stands for High-Definition TeleVision

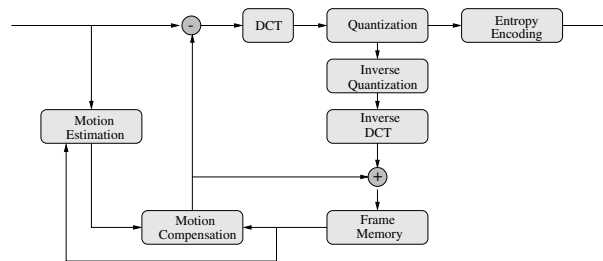


FIGURE 2.3: Block diagram of an encoder.

2.2.2 MPEG-2 video standard

An MPEG-2 video stream is hierarchically structured as illustrated in Fig. 2.4. The smallest entity defined by the standard is the *block*, which is an area of 8×8 pixels of luminance or chrominance. A *macroblock* (16×16 pixels) contains four blocks of luminance samples and two, four or eight blocks of chrominance samples, depending on the chrominance format. A variable number of macroblocks is encapsulated in an entity called a *slice*. A new slice always starts on each new line of macroblocks. Slices occur in the bitstream in the order in which they are encountered. Thus, each *picture* is composed of a variable number of slices.

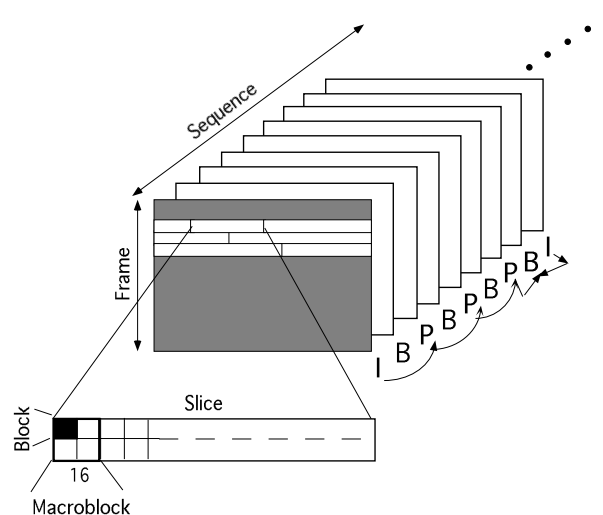


FIGURE 2.4: MPEG-2 video structure.

As represented in Fig. 2.4, the MPEG-2 video syntax defines three different types of pictures:

1. Intra-coded (or I-) pictures are coded without reference to preceding or upcoming pictures in the sequence. A picture is divided into 8×8 blocks of pixels and a two-dimensional discrete cosine transform (DCT) is applied to each block. The resulting DCT coefficients are quantized and variable-length coded. The quantization is performed by dividing each DCT coefficient by a quantizer and by rounding the result to an integer. The quantizer applied to a DCT coefficient comes from the multiplication of a quantizer scale, the so-called *MQANT*, and the corresponding quantizer matrix entry. A different *MQANT* value may be used for each macroblock. As the *MQANT* increases, the quality decreases in favor of the increased compression factor (see Sec. 5.2). Intra-coding provides a moderate compression rate while allowing random access into the compressed video data.
2. Predicted (or P-) pictures are coded with respect to the nearest previous I- or P-picture. The predictions of the best-matching macroblocks are indicated by motion vectors that describe

the displacement between them and the target macroblocks. The difference between the best-matching and the target macroblocks, called the *prediction error*, is encoded using the DCT-based intraframe technique summarized above. The motion vectors, as well as the quantized DCT coefficients, are variable-length coded. It is noteworthy that individual macroblocks may still be intra-coded in P-pictures (i.e. do not use motion estimation).

3. Bidirectional (or B-) pictures use both past and future I or P pictures as reference. Motion compensation is also applied here. However, both forward and backward motion vectors may be used for each macroblock since B-pictures are coded in relation to two reference pictures. Like in P-pictures, macroblocks in B-pictures may be intra-coded, and furthermore, some of the motion-estimated macroblocks might use only one motion vector (i.e. forward or backward motion vector). Bidirectional encoding provides the highest compression rate while introducing some delay.

The use of these three picture types allows MPEG-2 to be robust to packet loss (I-pictures provide stop points for the error propagation) and efficient in representing the video information (B- and P-pictures allow good compression). Furthermore, the MPEG-2 standard does not specify how I-, P- and B-pictures are mixed together. As mentioned, all coding modes can even be chosen per macroblock, which allows finely tuned tradeoff between robustness and efficiency.

Before being transmitted, a video stream goes through the MPEG-2 Transport Stream (TS) layer. Basically, the stream is first segmented into variable-length Packetized Elementary Stream (PES) packets and then subdivided into fixed-length TS packets. These packets are finally passed to the underlying layers for transmission. It is worth noting that a uncompressed header (i.e., syntactic information) is inserted before each of the following information elements: PES, TS, sequence, picture, and slice. In general, when a header is damaged, the underlying information is lost.

2.2.3 Other video coding schemes

The video coding standards are all block-based. DCT and motion estimation are indeed performed on a block of pixels basis. This type of coding can introduce blocking artifacts in the reconstructed video where block boundaries become visible at low rates. Other transform methods have been proposed to compact the spatial information without blocks limitations. Wavelets [20] and more generally subband coding [21, 22] constitutes certainly the major alternative to DCT coding [23]. Transform coding and subband coding are very similar, however subband decomposition provides additionally a multiresolution decomposition of images. In contrary to DCT for example, the wavelet representation provides a multiresolution and multifrequency expression of a signal with localization in both time and frequency. This property is particularly useful in image and video coding applications where the signal is nonstationary by nature. Wavelet-based coding decomposes, by filtering and subsampling, the nonstationary signal into a set of multiscaled wavelets where each component becomes relatively more stationary and thus easy to code. Another advantage of subband coding is the critical sampling of the input data. The transform signal has indeed the same size as the original, at the price however of constraints on the filter design, relatively poor low-pass approximation and the well-known edge-problem [24].

Several data organization and representation methods have been proposed to compact the wavelets data, exploiting the statistical properties of the wavelets coefficients [25]. Besides simple entropy coding schemes, more complex methods have been proposed in the literature: VQ [26], tree encoding like EZW [27] or SPIHT [28] algorithms, significance-linked connected components (SLCCA) [29, 30, 31] and morphological representation of wavelet data (MWRD) [32]. Arithmetic [33] or Huffman coding [34], as well as run-length coding [35, 36, 37] can then be performed to still improve the compression and reduce the coding rate of the wavelet representation.

To reduce temporal redundancy, the video coding standards generally use motion estimation methods. Motion estimation can also be used in video wavelet coding where the difference frame is coded by subband decomposition [38, 39, 40, 41, 42, 43, 44]. However, even if motion compensation has shown its efficiency [45], one can think about a temporal transform operation similar to the

spatial one. Instead of computing frame differences a filtering is performed along successive frames to reduce the temporal redundancy. This offers the advantage to eliminate the frame-to-frame dependency, thus simplifying the complexity for coding optimization. Temporal transform is generally performed through DCT [46, 47, 48, 49, 50] or wavelets [51, 52, 53, 54, 55, 56, 57], which are however different than the spatial wavelets. 3D transform compares favorably to standard coding [58, 59, 60, 49, 61, 62], but temporal filtering may introduce coding delays due to buffering at the encoder [63, 64]. This limitation restrain the temporal filtering efficiency for high motion scenes. A combined motion compensation and 3D wavelet transform along movement trajectories has been proposed to encompass this limitation [65, 66, 67, 68]. Lastly, the representation of tri-dimensional wavelets transform coefficients is mainly based on modifications of tree encoding schemes, like 3D SPIHT [60, 62, 69] or 3D EZW [70].

Finally, DCT or wavelet decompositions are based on general assumptions on the statistics of the input signal. They have been designed to offer a good compression efficiency for a large class of input signal. The transform design is fixed for any type of video sequence or scene while the signal characteristics can be however highly variable. To cope with this limitation, signal adaptive decompositions have recently been proposed to capture the real characteristics of the input signal, thus enable a more efficient compression. The signal is decomposed over a library of spatially and frequency localized basis elements, possibly redundant. The signal is expressed as a sum of basis function which are adaptively chosen accordingly to the signal features. The main adaptive compression algorithms group wavelets packets decomposition [71, 70, 72], quadtree [73, 74, 75, 76, 77], or double tree [78, 79] segmentation, JASF representation [80, 81], best basis pursuit [82, 83] and Matching Pursuit [84, 85, 86, 87, 88]. Matching Pursuit coding will be further developed in further chapters and presented as an interesting method for multi-resolution video coding.

2.3 Major networking technologies

The transmission of multimedia streams requires a network capable of handling different types of data, thus different types of constraints. For several years, the solution has been considered to be the *asynchronous transfer mode* (ATM) [89]. ATM is the network technology for the broadband integrated services digital network (B-ISDN). Now the role of ATM is being challenged by the success of the Internet and other IP-based networks due to the ongoing developments of integrated and differentiated services.

The asynchronous transfer mode combines the circuit-switched routing of telephony networks with the asynchronous time-division multiplexing of traditional packet switching. This is accomplished by establishing a virtual channel through the network before accepting any traffic. In ATM, data are sent in fixed size cells of 53 bytes. The network guarantees that all the cells of a call follow the same route and are delivered in the same order as sent.

The internet protocol differs in two major aspects from ATM. First, IP does not synchronize between the establishment of a route and the start of a session. Also, IP packets are of variable length (up to 65,535 octets). These packets may consequently arrive out of order if the routing decision has changed during the session.

In both ATM and IP networks transmission losses as well as delays are inevitable. The delay is chiefly caused by propagation and queuing. The queuing delay depends on the load of the links. Loss of information is mainly caused by a multiplexing overload of such magnitude and duration that the buffers in the nodes overflow. Loss may also be caused by misrouting due to bit-errors in the addresses, but this is less probable.

2.3.1 ATM networks

There are two types of ATM service classes suitable for transmission of audio/video streams: constant bit rate (CBR) service and variable bit rate (VBR) service (called deterministic bit rate and statistical bit rate transfer capabilities in the ITU Recommendations). CBR service means

that virtual channels are allocated portions of capacity that are at least equal to their declared peak rates. Note that CBR service only upper-bounds the bit rate which may vary within that bound. The ITU Recommendation does not state the associated quality of service but loss-free service with low maximum delay is possible. The VBR service means that a rate smaller than the declared peak rate is allocated for the connection [90]. Its specific value depends on the peak and sustainable rates, the maximum burst size that the connection initiator declares, and the quality level that the operator maintains for the service class. It is not clear what quality levels will be offered by network providers. The ATM Forum has separated the variable bit rate service into real-time VBR (rt-VBR) and non-realtime VBR (nrt-VBR). Delay limits are only guaranteed for rt-VBR. The ITU does not make this distinction. Among nrt-VBR services, one can further distinguish between ABR⁴ and UBR⁵. The ABR service class is defined for best-effort type applications and uses a feedback-based approach to control possible congestion. Finally, UBR does not offer any guarantee and are the lower priority services among all the defined classes.

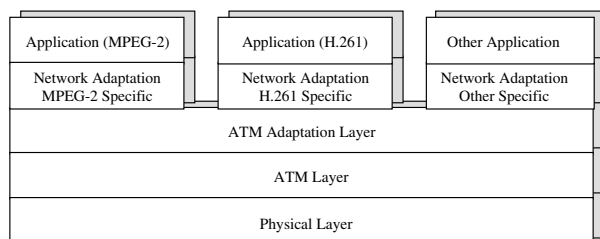


FIGURE 2.5: *Multimedia services over ATM protocol stack.*

The transport of compressed video over ATM [91, 92] can be done over two ATM adaptation layers (AAL) depending on the service required. If constant-rate video is to be transmitted, AAL1 may be used [93]. It provides a CBR service and features specific to real-time applications such as clock recovery and forward-error correction. Since an equivalent adaptation layer able to transport VBR data does not exist, AAL5 is considered as the second alternative for video transmission [94]. Even if AAL5 does not provide such real-time specific functions, it is generally accepted because it is simple and able to handle both CBR and VBR traffic [95].

Several functions required by multimedia applications are however not provided by the ATM adaptation layers. Multimedia applications involve the transmission of synchronized audio, video and data flows. Recommendation H.222.1 [96] specifies a network adaptation layer (NAL) that provides multiplexing and synchronization functions (see Figure 2.5 for the MPEG-2 example). In particular, the NAL provides multiplexing, timebase recovery, error reporting and priorities on a packet basis. It also describes the mapping of access data units from the applications to the AALs. The NAL is not totally generic. It provides functions specific to MPEG-2 and embeds its systems layer as shown in Fig. 2.6.

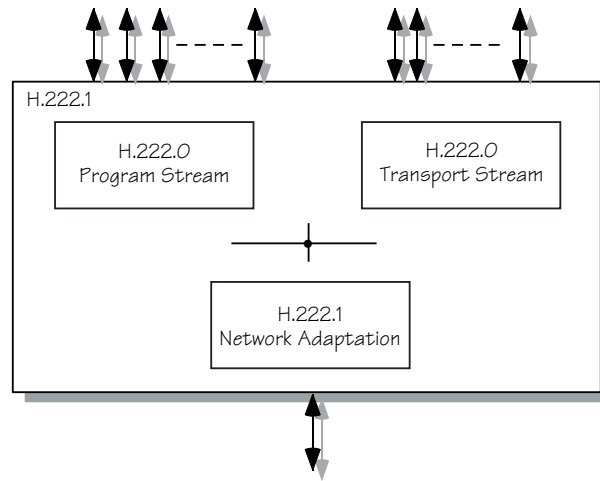
The encapsulation of MPEG-2 TS packets into AAL5-SDUs is defined in the ATM Forum's Video on Demand Specification 1.1. The proposed scheme encapsulates two single program transport stream (SPTS) packets, regardless of their information content (i.e. audio, video or timing data) into a single AAL5-SDU. The addition of the AAL5 8 byte trailer results in a AAL5-PDU that is segmented into exactly 8 ATM cells. For further readings on MPEG-2 over ATM, please refer to [97, 98, 99].

2.3.2 IP-based networks

The Internet is evolving to meet the needs of realtime traffic, but currently the requirements of realtime traffic are not truly supported. The Internet is not a new kind of physical network, but rather the technology that interconnects existing subnetworks. The Internetwork protocol (IP)

⁴ Available Bit Rate

⁵ Unspecified Bit Rate

**FIGURE 2.6:** *H.222.1 overview.*

resides in the OSI⁶ model network layer and is responsible for routing IP datagrams (see Fig. 2.7) through the Internet. It provides a connectionless datagram delivery service, where each datagram contains the full IP destination address.

0	4	8	16	19	24	31	
VERS		LEN		SERVICE TYPE		TOTAL LENGTH	
IDENT				FLAGS		FRAGMENT OFFSET	
TIME			PROTO		HEADER CHECKSUM		
SOURCE IP ADDRESS							
DESTINATION IP ADDRESS							
OPTIONS					PADDING		
DATA							
...							

FIGURE 2.7: *IP datagram format.*

IP is generally used jointly with an OSI transport layer protocol. The transport protocols employ the infrastructure provided by the network layer to establish an end-to-end service. On one hand, the user datagram protocol (UDP) extends the unreliable datagram service to the user. Each packet is treated as an independent unit, and error recovery is left to the application running on top of UDP. On the other hand, the transport control protocol TCP enhances the unreliable datagram service provided by the IP in order to produce a reliable end-to-end stream delivery of packets. TCP opens a connection between two devices by using a technique referred to as positive acknowledgment with retransmission to recover lost packets. With TCP the data stream is delivered lossless and in sequence to the receiver but with possibly high delays.

Basically, IP only offers a best-effort service. Therefore a number of changes are needed to make the Internet suitable for realtime traffic. It is necessary to provide a realtime traffic stream with a guaranteed quality of service (QoS) in the form of bandwidth, packet loss and end-to-end delay guarantees. To minimize the impact of such modifications, it is proposed to enhance rather than replace the basic service model that is currently in place. To this aim two service classes, referred to as guaranteed and controlled-load services [100, 101, 102] are defined in addition to the existing best effort service class. These two classes can be loosely compared to the CBR and VBR service classes of ATM :

⁶Open Systems Interconnection

1. Guaranteed service (GS) gives a lossless transfer with tight delay bounds.
2. Controlled load service (CLS) is supposed to yield a quality corresponding to a lightly loaded IP network at best effort; it is not expressed quantitatively. The admission control is based on the peak rate declared by a session initiator and on measurements of the load in the network. This could lead to higher network efficiency when compared to admission control based only on declared source descriptors.

Both GS and CLS connections can be established by RSVP⁷ [103] signaling. More recently, IPv6, the latest IP version, has been designed to convey a flow specification (*flowspec*) which is a description of the characteristics of the data in order to reserve resources and therefore guarantee QoS by interoperating with RSVP [104, 105]. RSVP is however not implemented in all today's routers which makes the usage of guaranteed services almost marginal.

In parallel the IETF has developed the real-time transport protocol (RTP) [106, 107, 108] for transmitting real-time data such as audio and video over multicast or unicast networks. RTP does however not address resource reservation for providing quality of service for real-time services. It provides payload-type identification, sequence numbers, time stamps and delivery reports (see Fig. 2.8). RTP relies on lower layers for multicast, timely data delivery and quality of service in general.

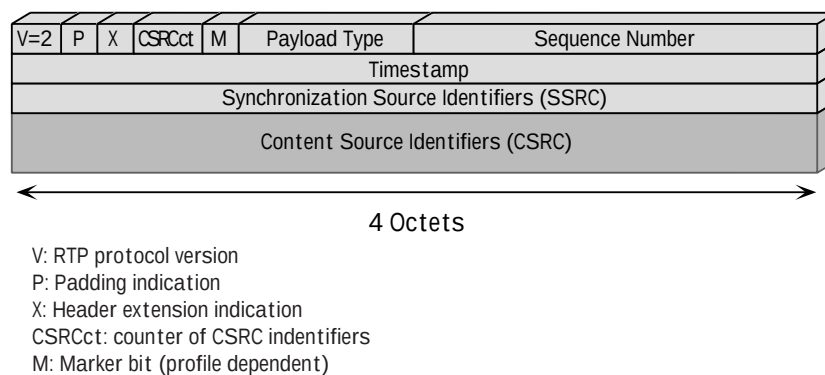


FIGURE 2.8: RTP packet format.

RTP is intended to be tailored to any particular application. To achieve the needed flexibility, the specification is deliberately incomplete. It defines a set of core functions but allows for extensions and customization depending on the target application. Companion documents or profiles have been written to specify the extensions needed for given applications. RFC 2250 defines the packet format for MPEG-1 and MPEG-2 audio and video [109]. It specifies payload identifier and encapsulation schemes for the different packet formats (i.e. TS or program stream).

Finally, the real-time control protocol (RTCP), part of the RTP specification, monitors the QoS and conveys information about the participants in a multicast session. RTCP is not intended to support all the requirements of applications control signaling. A higher level session control protocol as RTSP (Real-Time Streaming Protocol [110]) can fulfill these needs. RTCP is based on a periodic transmission of control packets to all participants in a session, using the same distribution mechanism as for data packets. The underlying protocol *must provide* multiplexing of the data and control packets.

2.3.3 Access technologies

Access to the Internet is nowadays highly heterogeneous. The best connections offers 10 Mbps in high-speed switched Ethernet, and even more with new technologies like Gigabit Ethernet. In parallel, residential access offers much smaller bandwidth to the end-user. Residential access

⁷Resource Reservation Protocol

includes 28 Kbps modem connections, 64 Kbps ISDN connections, xDSL connections [111, 112] with downstream rates between 100 Kbps and 6 Mbps and cable modems at 128 Kbps to 30 Mbps [113, 114, 115]. These rates are however only guaranteed on the last mile of the connection. Due to the load of the Internet, the available end-to-end rate is hardly predictable, and can be even lower than the values offered on the last mile.

The transmission rates are even smaller for wireless services [116]. The bandwidth on wireless connections is equivalent to 9.6 Kbps, values which further embraces a large amount of error protection. Even with the emerging wireless standards like UMTS [117] or the third generation wireless networks [118, 119], the rate offered to the users of wireless services lies around 2 Mbps, moreover shared between different users.

Clearly, the rates offered to the end-users make the task of video compression really challenging. The video services have to guarantee a good quality of service at very low effective bit rates. Moreover, the same application may be offered on possibly very different connections emphasizing the need for scalable video streams.

2.4 User-oriented QoS in video communications

2.4.1 What may affect the QoS ?

In digital video delivery, the video information flows through several subsystems (e.g. coder, network, decoder). Each of these subsystems may degrade the video quality, either by data loss or by introducing some delay. The video quality as perceived by the end-user [4] results thus from both the encoding artifacts and delays, as well as from the packet losses, delays and delay jitters caused by transmission.

On the one hand, all lossy compression schemes both distort and delay the signal. Degradation comes from the quantization which is the only irreversible process in common coding schemes. In general, the higher the quantization step, the smaller the source rate, but also the higher the degradation [120, 121]. Other coding decisions (e.g., coding modes) also influence the rate-distortion characteristic of video standard coders. Several studies have thus been proposed in this context to improve the quality for a given source rate. They propose to reduce the source coding distortion, either by adapting the coding mode [122, 123, 5, 124, 125, 126, 127] or the macroblock-based quantization process [128, 129, 130, 131, 132, 133, 134, 135].

The amount of delay introduced by the video coder is related to the size of the encoding buffer. This buffer is intended to limit the variations of the video rate output before transmission. The bigger the buffer, the smoother the bit rate may become, but it is at the expense of a higher delay. In Variable Bit Rate (VBR) encoding [136], a trade-off exists between the smoothing delay and bit rate variations in the output stream. Moreover, the regulation of the encoder to avoid overflow of the smoothing buffer causes quality variations in the decoded video stream. Video scenes are indeed variable by nature and thus not well adapted to Constant Bit Rate (CBR) encoding or to limitation of the rate variations. Therefore, adapting the source rate to the network constraints [137] by rate shaping for example [11, 138, 139, 140], is always a compromise between buffering delay and video quality.

On the other hand, data loss during the transmission may drastically reduce the user-QoS. The packet loss process has been widely studied on today's networks [141, 142, 143, 144, 145, 146] where actual loss traces are approximated with simple statistical models. The loss process is generally represented by the well-accepted Gilbert model [147] which can capture the bursty nature of packet losses [148, 149, 150, 151]. Such loss models are however not sufficient to quantify the loss of video information. The level of packet loss degradation on the video stream is indeed highly dependent on the importance of the lost information type, particularly in hierarchical coding schemes [152, 153, 154, 155]. In MPEG-2 video streams, for example, losses in headers result in more important degradation than losses of DCT coefficients and motion vectors. The transmission impairments depend also on the picture type of the lost video data because of the motion prediction used in MPEG-2. Figure 2.9 shows how network losses map into video information losses in different

types of pictures. Data loss spreads within a single picture up to the next resynchronization point (e.g. picture or slice headers) due to the variable-length [156] and differential coding. This is referred to as spatial propagation. When loss occurs in a reference picture (I- or P- picture), the lost macroblocks will affect the predicted macroblocks in subsequent frame(s). This is known as temporal propagation.

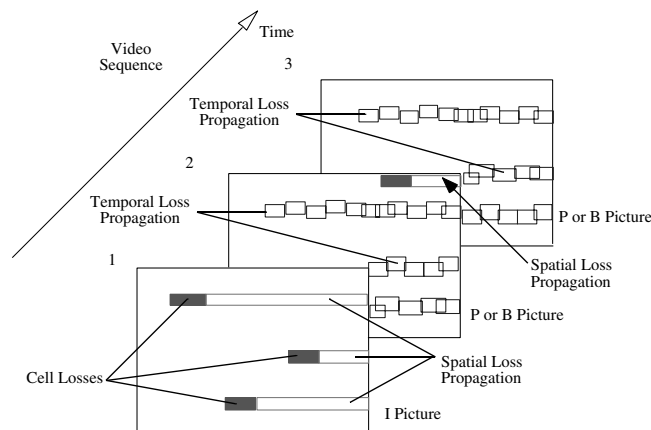


FIGURE 2.9: *Data loss propagation in MPEG-2 video streams.*

Finally, the impact of the loss of syntactic data is in general more important and more difficult to recover than the loss of semantic information. For instance when a frame header is lost, the entire frame is skipped since the decoder is not able to detect its beginning. If the skipped frame is a reference picture, the temporal error propagation may greatly reduce the perceptual quality. When a header is lost the whole underlying information is generally skipped. A classification of syntactic data can thus be operated. Similarly the semantic information can also be classified according to its importance to the end-user.

2.4.2 How to measure the video quality ?

Quality assessment of digitally coded pictures and image sequences nowadays becomes a very important issue and has received recently a large attention from the signal processing community. Indeed, in order to appreciate performances of new devices and algorithms, it is important to be able to quantize the quality of displayed images. Two classes of methods are available to measure the video quality: subjective tests, where subjects are asked to note the quality of images [157], and the objective ones, where a quality measure is generally computed by comparison of the original and distorted sequences. Objective methods consist in purely mathematical methods (i.e. Mean Squared Error, MSE) or more sophisticated methods which includes at some level the human visual system characteristics [158]. These last methods give naturally the most relevant results, since the visual information is finally destined to human user.

One of the simplest perceptual distortion metrics is the mean luminance difference between the original image and its reconstructed version. This elementary metric has the advantage to consider the human visual system most relevant information, that is the luminance. Hence, it gives results correlated with subjective testing under the assumption that the observer is far from the display device [159]. Distortion can also be measured in terms of contrast functions [160] or by spatio-temporal similarity metrics [161, 162, 163, 164]. More sophisticated metrics take into account, to some fidelity degree, the features of the human visual system [165, 166, 167, 168, 169]. One of the metric that provides the best correlation with human perception is certainly the Moving Picture Quality Metric (MPQM) [170, 171, 172, 173, 174]. The methods based on a human visual system model are however limited to degradations which are below a given level of distortion called the suprathreshold. If the degradations are too important, they generally behave similarly to pure mathematical metrics like PSNR. They are therefore generally not adapted to the evaluation of

sequences impaired by important packet losses. The notion of glitches has therefore been defined to measure the impact of losses on a video sequence [175, 176]. A glitch represents the impact of one loss on the video stream, characterized by a duration.

Finally, to be complete, two metrics will be used in this work: the HVS-based MPQM metric and the widespread purely mathematical Peak Signal-to-Noise Ratio (PSNR). The latter is simply defined by the ratio of the maximum pixel value to the mean squared distortion :

$$PSNR = 10 \log_{10} \frac{255^2}{MSE}.$$

2.5 Error resilience in video communications

Most coding algorithms are by nature not robust to transmission losses which may have dramatical effects onto the video quality. To improve the quality of service, error resilience algorithms [177, 178] can be implemented at both the application layer (encoder and decoder) and the network protocol stack. On one hand, the coding can be modified to generate a video stream robust to error by limiting the impact of packet losses. In parallel, lost areas can be hidden at decoder by means of concealment algorithms. These algorithms use remaining redundancy in correctly received information to approximate the missing data [179]. On the other hand, error control can also be deployed by the transmission protocols to recover lost packets. The most promising error resilience techniques are however built on the joint optimization of the mechanisms of each level.

Notice that we do not deal with delay in this section which rather focuses on the analysis of data loss. Delay jitter can mainly be solved via buffering techniques and the delay requirements can be translated into rate constraints [180]. However some delayed packets may fail to meet their respective decoding schedule and are therefore considered as lost.

2.5.1 Robust video encoding

In robust video coding redundancy is left in the source stream to limit the degradations due to losses. Several possibilities are offered to fight against propagation of losses in video standards [181, 182] where sensitivity to loss may be dramatically reduced by properly structuring video data and headers. For instance, slice headers [183, 184, 185] and intra-coded macroblocks [186] act as resynchronization points for spatial and temporal propagation of errors. Complete frame coding types can also be modified to limit the propagation of errors [187, 188]. Several error resilience coding modes have moreover been defined in the emerging coding standards like H.263+ [189, 190] and MPEG-4 [191, 192]. They basically consist in defining independently decodable video entities to avoid loss propagation within the reconstructed sequence.

Data partitioning or layered coding may also help to build more resilient video stream. Data is in this case at least partially available at decoder [193]. In wavelet-based coding scheme layered coding can help to reduce loss degradations [194, 195]. Additionally, in tree representation of wavelets coefficients (e.g., EZW), the division of the initial tree into independent subtrees may help to reduce loss impact onto this very vulnerable information structure [196, 197]. More generally, Variable Length Coding (VLC) and differential coding are very vulnerable to losses. Memory has to be kept minimal in coding operations to avoid loss propagation within the video stream. Resynchronization points should therefore be left in the stream to limit the impact of unavoidable errors. Hence, compression efficiency is reduced in comparison to encoding destined to error-free environments.

2.5.2 Video error concealment techniques

Error concealment is used at decoder to reduce the impact of data loss on the visual information. These algorithms include for example spatial interpolation, temporal interpolation and early resynchronization [198, 199] (see Fig. 2.10). Spatial methods generally uses features of the

surrounding macroblock to recover the missing data, by interpolation [200, 201, 202], MAP estimation [203, 204, 205] or projection [206, 207]. Temporal methods generally try to recover motion information from the surrounding blocks [208, 209, 210, 211, 212]. Hybrid concealment schemes are able to switch between spatial, temporal and even frequential concealment schemes [213]. Most of these methods are quite sophisticated in often too complex to be performed without influence on the decoding delay. The MPEG-2 standard proposes an elementary error concealment algorithm based on motion compensation, which will be used within the following chapters. Basically, it estimates the vectors of lost macroblocks by using the motion vectors of neighboring macroblocks in the same picture (provided these have not also been lost). This allows an efficient concealment of moving picture areas. There is however an obvious problem with lost macroblocks whose neighbors are intra-coded, because there have ordinarily no associated motion vectors. To get around this problem, the encoding can include motion vectors also for intra-coded macroblocks⁸.

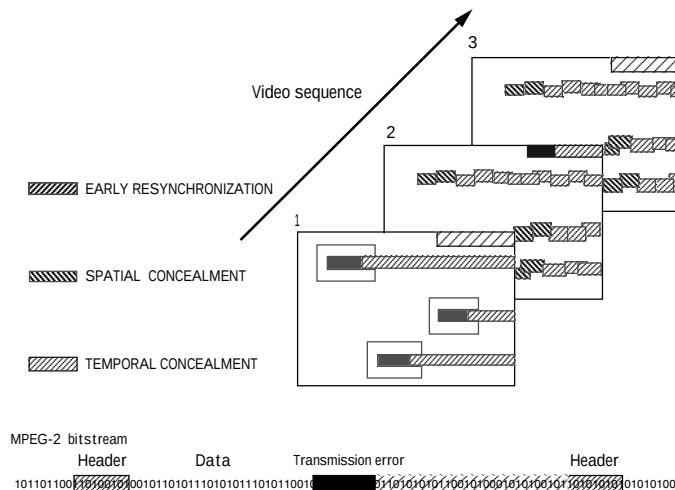


FIGURE 2.10: *Error concealment techniques.*

Error concealment may, in general, efficiently decrease the visibility of data loss. These methods alone are however not sufficient to provide a good quality of service when losses are important. Notice that most of these concealment algorithms are valid for block-based and motion compensated coding in general. In wavelet coding efficient data representations are generally quite different from block-based representation. In this context error concealment mainly consists in recovering the specific data structure [214]. Numerous post-processing methods can finally be used to smooth or attenuate the remaining artifacts before display.

2.5.3 Channel coding

From the transmission standpoint, video delivery can mainly be improved by providing error correction mechanisms [215, 216]. Two major techniques are proposed to recover from packet erasures: retransmission (ARQ) and forward error correction (FEC) [217, 218, 219, 220], or hybrid combination of both [221, 222]. Retransmission requires the receiver to inform the sender about the data that must be repeated. It has the advantage of guaranteeing error-free delivery, but at the price of a possibly large unpredictable delay [223]. FEC means that redundancy is added to the data so that the receiver can recover from losses or errors without any further intervention from the sender. Considering the delay requirements for interactive video and real-time applications in general, FEC is more appropriate than retransmission because it allows to respect the timing constraints. The Internet community is moreover considering the use of FEC for video transport over RTP [224]. However, FEC protection does not guarantee an error-free packet delivery.

⁸Some MPEG-2 encoder chips automatically produce concealment motion vectors for all macroblocks.

One of the cause of error propagation within video streams is the mismatch between video and network entities. Video coding defines entities like slices which are independently decodable. However, it may happen that such entities are segmented in different transmission packets, resulting in loss propagation if one of these packets is lost. Packetization algorithms can help to limit the effect of losses by providing means to encapsulate by nature varying length video entities into fixed-size packets [225, 226]. The packetization can even judiciously build independent packets, or packets which are decodable independently of the receipt of other packets [227, 228, 229]. The packetization process can also be modified by including resynchronization information to limit the degradation in case of loss [230].

Finally, for a network that supports separate service classes, it is possible to transfer different data layers with a service level commensurate with their importance [231, 232, 233]. Vital layers may thus be transferred in a class with guaranteed quality, while a signal layer destined to further enhance the quality of service could be sent in *best effort* mode.

2.5.4 Joint source channel coding

The most efficient error resilient schemes consider the entire system as a whole and target the global optimization of the communication channel, instead of optimizing individual components in isolation [234]. Since the separation principle of Shannon is not valid in practical systems, joint source and channel coding clearly allows to build the most efficient video stream [235].

In joint source and channel coding, the error protection presented in the previous section can be adapted to the encoded data. Knowledge of the video data is used to efficiently protect the most important information. Video packets can indeed have different importance considering their usefulness in the video decoding process. Therefore it is wasteful in terms of transmission bandwidth to apply a uniform protection for all packets. Unequal Error Protection (UEP) consists in varying the level of FEC protection to meet the user QoS requirements [236, 237, 238] while limiting the necessary bit rate. In the case of layered coding, for example, the FEC protection may be adapted to the relative importance of the different layers in standard-based [239, 240, 241, 242] as well as in wavelet-based coding schemes [243, 53, 56]. Error protection and quality provisioning could be selected individually for each layer rather than for the entire bitstream in once. The error control schemes can also be adapted either to the encoded sequence [244, 245], to the syntactic content of the packets [246, 247, 248, 10] or to the visual importance of the video content [249]. Finally unequal error protection is very helpful in the case of progressive video transmission, where the level of channel redundancy varies along the transmission [250, 251, 252].

Similarly, the video encoding can also be adapted to the transmission conditions [253, 254], taking benefits of information about the channel status. The amount of syntactic redundancy like Intra macroblocks [255, 256] or slice headers [257] can be adapted to the loss process. If the probability of loss becomes important, the syntactic overhead is increased to ensure an acceptable quality of service. The source or video coding rate can also be constrained to respect the appropriate balance between source and channel rate for a total bit rate [258, 259], in standard-based [260, 261, 262] as well as in wavelet-based coding schemes [263, 264]. Finally, the most complete joint source and channel coding schemes take into account hypothetical loss effects to globally optimize end-to-end quality [265, 266, 267, 268]. They take into account the error concealment performed at decoder to evaluate the visual importance of the data and thus protect in priority data that will be poorly concealed in case of loss.

Finally, layered or multiresolution coding can be also considered as a form of joint source and channel coding and are studied more in detail in the next chapters. The recently proposed Multiple Description Coding methods also represent a form of joint source and channel coding and provide a very promising alternative for error resilient video delivery.

2.5.5 Multiple Descriptions Coding

Multiple Descriptions Coding (see Fig. 2.11) is certainly one of the most promising coding approach in error-prone environments. This generic concept consists basically in representing the input

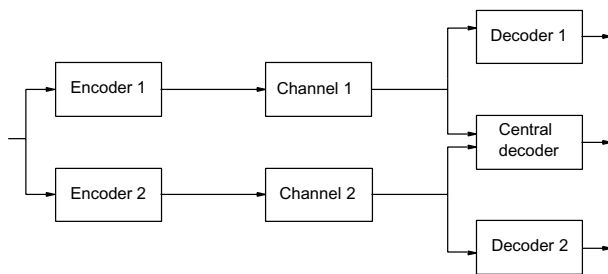


FIGURE 2.11: *Multiple Description Coding. General representation with two descriptions.*

signal by several descriptions simultaneously sent over the network. The best signal reconstruction is obtained when all descriptions are correctly received, while the correct reception of a single description should already provide a reasonable quality. The main problem lies in the construction of the signal descriptions which should neither be completely independent to ensure the delivery of the main content, nor completely redundant to limit the bit rate. Rate-distortions characteristics of Multiple Descriptions Coding (MDC) have been widely studied [269, 270, 271, 272]. In most implementations MDC uses different vector or scalar quantization levels [273] to build several coarsely quantized descriptions of the same signal. When taken together, all these descriptions result in a finely quantized source representation.

Due to the intrinsically redundant nature of video signals, MDC is particularly well suited for resilient video application. Multiple descriptions can easily be formed either by considering multiresolution structures like subband coding [274, 275] or the importance of video data [276]. Besides the methods based on classical orthogonal transforms [277], MDC schemes can also use non-orthogonal transforms to generate redundant descriptions of the input data [278, 279, 280, 281, 282]. Video multiple description coding can also be performed by judicious quantization [283] and/or packetization process [284]. All the above methods may obviously be combined in optimal MDC schemes. They only represent the main schemes from the literature. However MDC is a very generic concept which can be considered as a form of multiresolution coding, and also embrace the UEP schemes [285] presented above.

2.6 Multiresolution coding and scalable video delivery

2.6.1 Multiresolution coding

Multiresolution video coding [286] and scalable applications become primordial in nowadays highly heterogeneous framework. Indeed, service providers can obviously not afford to offer each individual client a different video stream. Layered coding is particularly useful when a specific target bit rate or quality level cannot be stated *a priori* for the transfer. By layering, the sender can provide a range of bit rates and qualities and particular values in that range can be dynamically chosen by the receiver. For instance, layered coding can be beneficial for stored programs and for multicast applications. The video sequence is in this case encoded once and decoded several times in different conditions. However, the number of allowable resolution is clearly limited to one of the possible combination of the available video layers.

The current video compression standards propose several scalable coding modes. MPEG-2, for example, defines temporal, spatial and SNR scalability and data partitioning profiles. These profiles offer the possibility to design scalable video applications [287, 288, 289] by judicious combination of the different video layers. Layers may be formed either by appropriate quantization [290] or by DCT coefficients partitioning [291, 292, 293]. The recently defined MPEG-4 standard moreover offers additional possibilities of scalable coding [294].

Multiresolution video decompositions go back to pyramid coding [295] and subband coding [296, 297]. In applied mathematics they are intimately related to wavelets coding [298, 299, 300] which

thus inherently offers the possibility to build multiresolution video signals. Several studies propose multiresolution video wavelet-based coding [56, 40, 301, 235, 41]. To provide different temporal resolutions the video coders either perform a hierarchical motion compensation [302, 303, 304, 305] or tri-dimensional wavelet decomposition [306]. Hybrid combination of MPEG and subband coding [307] or perceptual decompositions [308] also result in interesting solutions for scalable video coding.

To avoid costly bitstream adaptation within the gateways, it may be particularly useful to build progressive [309, 310, 311, 312] or embedded video streams. The video bit rate can thus be easily truncated according to the available bandwidth. Wavelets decompositions may offer interesting properties for efficient progressive transmission. 3D subband embedded video coding either are based on EZW extensions [313, 70] or even more advantageously on modifications of the SPIHT algorithm [60, 62, 69] to account for the temporal decomposition. They thus provide a hierarchical video stream [314, 315] and allow to serve both low and high bit rates applications with the same video stream.

While progressive coding offers a fine rate granularity, the number of achievable spatio-temporal resolutions is really limited in most scalable coding schemes. Standard video coding is highly constrained by the hierarchical structure of the video stream and wavelet-based schemes only provides with an octave-based granularity. Spatio-temporal subbands can however be combined to achieve a reasonable number of spatio-temporal resolutions [316].

2.6.2 Multicast and error control

The distribution of the scalable video streams to the receivers is generally performed by multicast applications. Their deployment on the Internet mainly lies on application-level protocols like RTP on top of the IP Multicast protocol [317]. The video distribution is generally initiated by the receiver. Receivers can subscribe to multicast groups and thus participate to the video session. The multicast session reacts to receivers requests and adapts the video layers coding rates to their requirements [318, 319]. The video is distributed along a multicast tree [320] and the stream may be adapted within the network nodes [321, 322, 323] according to the requirements of the receivers [324]. Besides layered coding, hierarchical or progressive streams can also greatly help to the management of the distribution tree, offering even a finer granularity for the stream adaptation.

Robust transmission in lossy multicast session can almost exclusively be obtained by FEC or hybrid ARQ/FEC mechanisms [325, 326]. Indeed, retransmission of lost packet may conduct to bandwidth explosion since all receivers may ask for a different packet. The receivers can however be divided into independent subgroups [327, 328] to limit the number of replicated packets. Finally, in case of hierarchical video transmission coding, efficient forward error correction has also to be hierarchical [329] and driven by the receiver according to the channel status [330].

2.6.3 Transcoding and filtering

Since receivers may be completely heterogeneous in terms of spatio-temporal resolutions or computational power proxy-based signal processing approaches become particularly useful. Some processing is thus asked to the network nodes or video gateways [331, 332] in order to adapt the video stream to the receivers. This resolution adaptation process or transcoding adapts the video stream to the resolution or the computational possibilities of the receivers [333]. Meanwhile, filtering operations adapt the rate to meet the bandwidth constraints of the slowest connections.

As discussed earlier hierarchical streams enable a very easy adaptation process within the network nodes. They facilitate filtering and fine adaptation of the video bit rate. The spatio-temporal resolution has however to be adapted through transcoding methods [334, 335] which have to stay simple enough so that it does not overload the network nodes. To this aim, methods that work in the transform domain [336, 337, 338, 339] are often preferred to the methods that necessitate an expensive decoding/re-encoding process in the gateways [340, 341]. Transcoding can also adapt the coding scheme to the capabilities of the receiver and even vary the resolution of the video objects when used with MPEG-4 for example [342]. Proxies can also select the appropriate

content to forward to the receiver [343] or even locally adapt the FEC protection to the channel status [344].

2.7 Internet video: an overview

This section briefly surveys related work in applications, systems and architectures for video transmission over the Internet. One of the first applications for Internet video transmission was the Network Video tool *nv* by Xerox PARC⁹. Its implementation used a Haar wavelet decomposition, but does not have a layered representation. Moreover, it was based on the unrealistic assumption that the network may provide the rate required by the user. Soon after the release of *nv*, INRIA proposed a video application called Video Conferencing System or *ivs* [345]. This system was based on H.261 and performed loss recovery by ARQ scheme for small sessions only. Cornell University then proposed an application called CU-SeeMe [346], which combines low-quality grayscale video coding scheme and a unicast-based transmission, which makes it very popular. In parallel, UCL developed certainly one of the most elaborated tool called *vic* [347]. This implementation proposed a compression based on H.261 on top of RTP, which conditionally replenishes Intra-MB to reach a trade-off between compression and error resilience. Finally, more recently, Microsoft released the videoconferencing tool called NetMeeting. It uses mainly H.263 standard video compression on top of the RTP/RTCP protocols and moreover offers the possibility to share documents during the audio-visual conference.

All the above applications are designed for real-time, interactive human-to-human communications. Another important application of video on the Internet is the playback of pre-stored video material, e.g., for video-on-demand applications. A number of companies provide streaming-video technologies, but the major players are certainly Apple's QuickTime, Microsoft's Windows Media and RealNetwork's Real Video. Users can access small-screen or low quality video from these products with 56 Kbps or even slower modems. And the best quality the users may require respectively 700 Kbps for QuickTime, 720 Kbps for Windows Media and 500 Kbps for RealVideo. However, RealVideo [348] offers the additional advantage of layered video representation, so the end-user may choose the best quality based on the available bandwidth, while in other applications the set of pre-stored resolutions is much smaller.

Finally, the spatio-temporal resolutions defined by the products here-above is generally fixed. The quality may be adapted to the available rate, but the resolution is fixed. Spatio-temporal scalability is implemented in CMT from Berkeley¹⁰, where the coder build a multiresolution video stream. Low-end devices like PDAs (e.g., 3Com PalmPilot, Compaq iPAQ, Yopy PDA) additionally need very low complexity decoding methods to be able to play video stream corresponding to their limited resolutions. The majors actors in video streaming on PDA or cellular phones are PacketVideo¹¹, whose product is based on the MPEG-4 standard, and ActiveSky¹².

2.8 Conclusion

In this chapter, we have surveyed the miscellaneous technologies implied in video communications. Compression schemes have been overviewed, with a special care on MPEG-2 as an example of standard coding schemes. The network technologies, mainly based on IP networks, have been shown to provide users with very different rates and network QoS. Several error resilience methods have then been presented to cope with packet loss inherent to IP networks. Joint source channel coding methods have been presented as the most efficient schemes to provide users with an optimal Quality of Service. Since the field of video communications is very broad, it has however not been possible to detail every addressed topic. Interested readers are referred to the cited publications for detailed schemes descriptions. Additional information will however be given in the following

⁹<ftp://ftp.parc.xerox.com/net-research>

¹⁰<http://www-video.eecs.berkeley.edu/~dtan/proj.html>

¹¹<http://www.packetvideo.com>

¹²<http://www.activesky.com>

chapters. Finally, commercial implementations have been briefly described. However, it has been shown that none of the today's applications completely fulfill the requirements of robust multiresolution and scalable video services.

FEC Performance in Video Streaming

3.1 Introduction

Forward Error Correction (FEC) techniques are the preferred error-control schemes for multicast or interactive streaming applications, as discussed in the previous chapter. In this work a very simple media-independent Forward Error Correction (FEC) mechanism is used. Due to the low bit error rates associated with modern communication media, the assumption is made that decoding is mainly impeded by packet loss. Either the packet is present and correct or it is lost. These losses are mainly caused by network congestion and the resultant buffer overflow and queuing delay. In this case, packet-level FEC schemes [351, 218, 352, 353] provide an efficient way to fight against losses, although perfect recovery cannot be guaranteed. The details of the FEC algorithms are outside the scope of this chapter but can be found in [354]. Recall however that common FEC schemes based on Reed-Solomon codes or X-OR functions [355] can generally correct as many losses as the number of redundancy packets.

The purpose of this section is to compute the video loss process parameters after FEC recovery. These parameters are the packet loss ratio π_v (i.e., the proportion of lost packets) and the average burst length α_v . They directly drive the final video quality. The packet loss ratio π_v simply represents the ratio between lost and sent video packets and the average burst length α_v is the average length of consecutively lost video packets.

Several studies have been performed to compute the FEC efficiency or the probability for data to be recovered in case of loss [356, 357, 358, 359, 360]. However this parameter do not bring enough information about the loss process after FEC recovery. To correctly model the video quality, for example, at least two parameters, namely the packet loss ratio π_v and the average burst length α_v , shall be computed [361]. The aim of this chapter is to compute these two parameters in the case of a renewal error process. The FEC efficiency depends on both the network performance and the media-independent FEC parameters k and n .

The chapter is organized as follows: Section 3.2 presents a brief overview of FEC error-control and develops the set of equations needed to compute π_v and α_v . Both parameters are then computed in Section 3.3 for a Gilbert-model loss process. The analytical values are validated against experimental data. Finally, concluding remarks are given in Section 3.4.

Parts of this work appeared in [349, 350].

3.2 Loss process after FEC recovery

Recall that common FEC schemes based on Reed-Solomon codes or X-OR functions can generally correct as many losses as the number of redundancy packets.

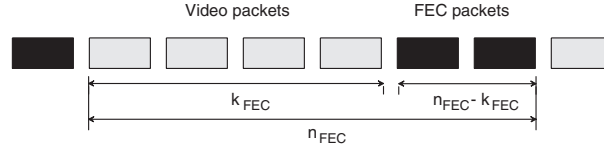


FIGURE 3.1: *Media-independent FEC scheme.*

Assume every block of k video packets are protected by $(n - k)$ FEC packets as represented in Figure 3.1 to form an n -packet FEC block. Each packet represents a different RTP packet, where a header field indicates if the packet is a video or FEC packet. This is referred to as *media-independent FEC*. If at least k out of n packets are correctly received, the underlying video information can be correctly decoded [362]. Otherwise, none of the lost packets can be recovered by the receiver. Hence the packet loss pattern experienced at the video level is quite different from the loss pattern observed on the global packet stream of the lossy channel.

Let us assume that the loss process can be modeled with a renewal error process. In other words, the lengths of consecutive inter-error intervals (also called gaps) are assumed to be independently and identically distributed. For the sake of clarity assume any packet is assigned a binary value 0 or 1 corresponding to correctly received and lost packets respectively.

Following the development of [356], let $p(i)$ denote the probability that a gap length is $i - 1$, i.e., $p(i) = \Pr(0^{i-1}1|1)$, where 0^{i-1} is a shorthand for $i - 1$ successive 0's. Similarly, let $P(i)$ denote the probability that at least $i - 1$ 0's follow a given error, i.e., $P(i) = \Pr(0^{i-1}|1)$. It has to be noted that $P(i)$ could be written as

$$P(i) = \sum_{m=i}^{\infty} p(m). \quad (3.2.1)$$

The probability of the sequence $1\ 0^{i-1}$ is thus given by

$$\Pr(1\ 0^{i-1}) = \pi\ P(i). \quad (3.2.2)$$

Order is irrelevant because of the independence among gap lengths of a renewal process. The events $1\ 0^{i-1}$ and $0^{i-1}1$ are therefore equiprobable. From this property, the probability $R(m, n)$ that $m - 1$ errors occur in the next $n - 1$ bits following an error can be easily computed by recurrence [356]. Thus,

$$R(m, n) = \begin{cases} P(n), & \text{for } m = 1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} p(i) R(m-1, n-i), & \text{for } 2 \leq m \leq n. \end{cases} \quad (3.2.3)$$

Similarly, we denote $r(m, n)$ the probability that $m - 1$ errors occur in the $n - 1$ bits between two errors. It can also be computed by recurrence as

$$r(m, n) = \begin{cases} p(n), & \text{for } m = 1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} p(i) r(m-1, n-i), & \text{for } 2 \leq m \leq n. \end{cases} \quad (3.2.4)$$

Finally, let $\bar{r}(m, n)$ represent the probability that $m - 1$ errors occur in the $n - 1$ following an error and preceding a 0:

$$\bar{r}(m, n) = R(m, n) - r(m, n). \quad (3.2.5)$$

Let moreover define the probability $q(i)$ that a burst is of length $i - 1$ and the probability $Q(i)$ that at least $i - 1$ 1's follow a zero. These probabilities are given by the loss process or can even be deduced from the above variables. The dual of $R(m, n)$, namely $S(m, n)$, represents the probability to have $m - 1$ 0's in the next $n - 1$ bits following a 0. This probability is obtained by recurrence from

$$S(m, n) = \begin{cases} Q(n), & \text{for } m = 1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} q(i) S(m-1, n-i), & \text{for } 2 \leq m \leq n. \end{cases} \quad (3.2.6)$$

The video packet loss rate after FEC recovery is now easy to compute. Two cases are considered with respect to the state of the last video packet of a FEC block. Its loss or its presence directly drives the loss process into the next FEC block. By the renewal process properties, π_v could thus be computed as

$$\begin{aligned} \pi_v = & \frac{\pi}{k} \sum_{i=1}^k i R(i, k) \sum_{j=\lfloor n-k+1-i \rfloor}^{n-k} R(j+1, n-k+1) \\ & + \frac{1-\pi}{k} \sum_{i=1}^{k-1} (k-i) S(i, k) \sum_{j=0}^{k-1-i} S(j+1, n-k+1), \end{aligned} \quad (3.2.7)$$

where the notation $\lfloor x \rfloor$ represents the positive part of x and π represents the global packet loss ratio.

The average video burst length after FEC recovery, α_v , can also be computed from the previous development. Since bursts of errors do not have the same probability to start on any packet of the FEC block, each position has first to be considered separately. The probability for a burst to start on the j^{th} packet of a FEC block (i.e., $1 \leq j \leq n$) is given by

$$P_j = \Pr(x_{j-1} = 0, x_j = 1), \quad (3.2.8)$$

where x_i denotes the binary state of the packet i (i.e., either lost or not after FEC recovery). From the previous development, P_j can be written as :

$$P_j = \begin{cases} \pi (P(2) + p(1) R_n) \sum_{i=n-k}^{n-1} R(i+1, n), & \text{if } j = 1 \\ \pi \sum_{i=0}^{n-j} \sum_{x=0}^{k-n+j+i-2} P(2) S(x+1, j-1) R(i+1, n-j+1), & \text{if } j > 1, \end{cases} \quad (3.2.9)$$

where R_n is the probability that all packets in a FEC block are recovered given that the first packet is missing. It can be written as

$$R_n = \sum_{i=0}^{n-k-1} R(i+1, n).$$

The average length of bursts of lost video packets, α_v , excludes redundancy packets. Let first T denote the number of FEC blocks transitions along the burst of length l :

$$T = \text{floor} \left(\frac{l+j-2}{k} \right).$$

Let v denote the position in a FEC block of the first video packet following a burst of length l :

$$v = \begin{cases} k, & \text{if } (l+j-1) \bmod k = 0 \\ (l+j-1) \bmod k, & \text{otherwise.} \end{cases}$$

Moreover let r_{n-k} and $\bar{r}_{n-k}$ denote the probability to loose all the video packets of a FEC block and that the first video packet of the next FEC block is respectively erased or not (before FEC recovery). These probabilities can be written as

$$r_{n-k} = \sum_{[n-2k+1]}^{n-k} r(i+1, n-k+1),$$

and

$$\bar{r}_{n-k} = \sum_{[n-2k+1]}^{n-k} \bar{r}(i+1, n-k+1).$$

Similarly r_{n-k}^j and $\bar{r}_{n-k}^j$ denote the probability that the last video packet (i.e., the packet k) of the first FEC block is lost and the first video packet of the second block is respectively erased or not (before FEC recovery). It is assumed that all video packets between packets j and k are lost. These probabilities are expressed as

$$r_{n-k}^j = \sum_{i=0}^{n-k} r(i+1, n-k+1) \sum_{x=0}^{2k-n+i-2} S(x+1, j-1),$$

and

$$\bar{r}_{n-k}^j = \sum_{i=0}^{n-k} \bar{r}(i+1, n-k+1) \sum_{x=0}^{2k-n+i-2} S(x+1, j-1).$$

Finally, the average length of a burst of lost video packets starting on the j^{th} video packet of a FEC block is given by :

$$L_j = \frac{\sum_{l=1}^{\infty} l \Pr(x_{j-1} = 0, x_j = 1, \dots, x_{j+l-1} = 1, x_{j+l} = 0)}{P_j}, \quad (3.2.10)$$

where only video packets are considered. Indeed FEC packets have to be skipped at the end of each FEC block. They obviously do not participate to the video burst length. Equivalently, the average length L_j can be written as :

$$L_j = \sum_{l=0}^{\infty} l p(1)^{l-T-1} \lambda_j(l), \quad (3.2.11)$$

where $\lambda_j(l)$ is the probability to have a video packet burst of length l starting at j . The conditional probability $\lambda_1(l)$ for bursts starting with the first video packet of the FEC block (i.e., $j = 1$) can be written as

$$\lambda_1(l) = \begin{cases} \frac{\pi (P(2)+p(1) R_n)}{P_1} (r_{n-k} R_n + \bar{r}_{n-k}) r_{n-k}^T, & \text{if } v = n \\ \frac{\pi (P(2)+p(1) R_n)}{P_1} P(2) \sum_{i=0}^{k-2} S(i+1, n-v) r_{n-k}^T, & \text{otherwise.} \end{cases}$$

For $2 \leq j \leq k$, λ_j is expressed as

$$\lambda_j(l) = \begin{cases} \frac{\pi p(1)}{P_j} \sum_{i=0}^{n-v-1} \sum_{x=0}^{k-3-i} S(x+1, j-1) p(1) S(i+1, n-v), & \text{if } l+j-1 < k \\ \frac{\pi p(1)}{P_j} (r_{n-k}^j R_n + \bar{r}_{n-k}^j), & \text{if } v = k \text{ and } T = 0 \\ \frac{\pi p(1)}{P_j} (r_{n-k} R_n + \bar{r}_{n-k}) r_{n-k}^j r_{n-k}^T, & \text{if } v = k \text{ and } T \neq 0 \\ \frac{\pi p(1)}{P_j} p(1) \sum_{i=0}^{k-2} S(i+1, n-v) r_{n-k}^j r_{n-k}^{\lfloor T-1 \rfloor}, & \text{otherwise.} \end{cases}$$

The video average burst length α_v is finally given by the following probability-weighted summation

$$\alpha_v = \frac{\sum_{j=1}^k P_j L_j}{\sum_{j=1}^k P_j}, \quad (3.2.12)$$

where P_j and L_j are given from Eqs (3.2.9) and (3.2.11).

3.3 FEC in a Gilbert-model loss process

Modeling end-to-end packet-loss behavior is a challenging problem for system designers. By making the natural assumption that buffer overflow and network congestion cause correlated packet loss, a two-state Markov chain may be employed to model packet-loss behavior for different priority classes. The Markov chain effectively captures the buffer overflow and underflow characteristics with minimal parameters [363]. Hence, assume now that the channel loss process can be characterized by the Gilbert model [364, 365, 148, 147]. Notice that our goal here is not to validate the Gilbert model, but rather to show how, given a channel model, the video loss pattern could be computed. The Gilbert model is a two-state Markov model [366] with geometrically distributed residence times (see Figure 3.2). States 0 and 1 correspond respectively to the correct reception or the loss of a packet. The transition rates p and q between the states control the lengths of the error bursts.

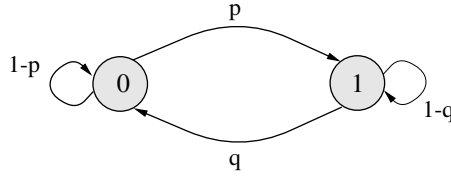


FIGURE 3.2: *Two-state Markov chain: Gilbert model.*

The global packet loss ratio π corresponds to the stationary probability to be in the loss state: $\pi = \frac{p}{p+q}$. The average error burst length α is given by the average residence time in the loss state: $\alpha = \frac{1}{q}$. These loss process parameters π and α can be sensed through control protocols (e.g., RTCP) or delay measurements [367]. From the previous development, the loss patterns π_v and α_v can be computed in the case of a Gilbert-model loss process. They obviously depend on both the model parameters p and q and the FEC parameters k and n .

For a Gilbert loss process the following relations hold

$$p(i) = \begin{cases} 1 - q, & \text{if } i = 1 \\ q (1 - p)^{i-2} p, & \text{otherwise,} \end{cases}$$

$$P(i) = \begin{cases} 1, & \text{if } i = 1 \\ q (1 - p)^{i-2}, & \text{otherwise,} \end{cases}$$

$$q(i) = \begin{cases} 1 - p, & \text{if } i = 1 \\ p (1 - q)^{i-2} q, & \text{otherwise,} \end{cases}$$

$$Q(i) = \begin{cases} 1, & \text{if } i = 1 \\ p (1 - q)^{i-2}, & \text{otherwise.} \end{cases}$$

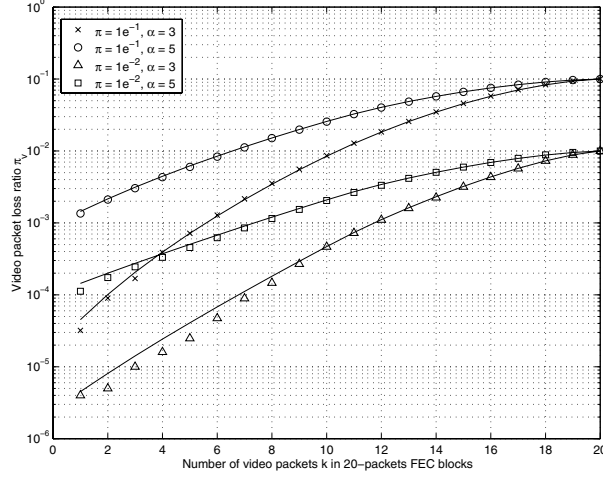


FIGURE 3.3: Evolution of π_v versus the number of video packets k in a FEC block of length $n = 20$ in a Gilbert-model loss process.

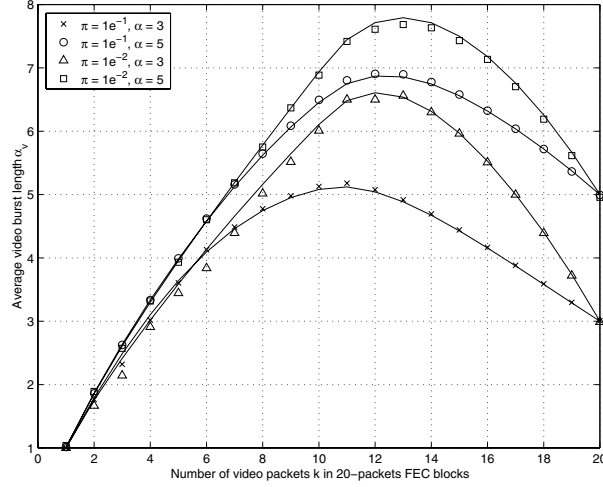


FIGURE 3.4: Evolution of α_v versus the number of video packets k in a FEC block of length $n = 20$ in a Gilbert-model loss process.

The probabilities $R(m, n)$, $r(m, n)$, $\bar{r}(m, n)$ and $S(m, n)$ can be computed by recurrence from Eqs. (3.2.3), (3.2.4), (3.2.5) and (3.2.6) respectively. The video packet loss ratio and average burst length are then computed from Eq. (3.2.7) and Eq. (3.2.12) respectively.

Figures 3.3 and 3.4 represent the evolution of the video loss parameters π_v and α_v for different network loss patterns. The analytical values perfectly fit the experimental data. Moreover, π_v obviously increases with k as the amount of protection decreases for a given n . FEC protection becomes also less efficient for bursty loss traffic (i.e., large α values) for a given π . Moreover the average length of lost video packets clearly exhibits a maximum. This can be explained as follows. When the amount of protection is very large α_v stays close to k . When the amount of protection decreases, the video loss pattern get closer to the channel loss pattern. In between there is a maximum which is less pronounced for bursty process. These behaviors still hold for different FEC block lengths.

3.4 Conclusions

In this chapter the performance of a media-independent FEC scheme for video streaming applications is analyzed. Both the packet loss ratio and average burst length as perceived by the multimedia application are computed for a Gilbert-model loss process. Experimental simulations validate the analytical results. Our main results can be used to compute the optimal FEC parameters k and n for a given packet loss process as seen by the network. Clearly, the results in this chapter hold for any type of multimedia data for which FEC protection is suitable.

Joint MPEG-2/FEC Rate Selection

4.1 Introduction

In general, the distortion the end-user perceives results from compression artifacts, packet losses, delays and delay jitters. All lossy compression schemes distort and delay the signal. Degradation mainly comes from the quantization, which is the only irreversible process in a coding scheme. Moreover, delays and packet losses are inevitable during transfers across today's networks. The delay is generally due to propagation and queuing. Information loss is mainly caused by multiplexing overloads of high magnitude and duration that lead to buffer overflow in the nodes. Data loss is particularly annoying in video streaming applications due to the predictive structure of MPEG-2 compression, as described in the Sec. 2.4.1.

Interactive video delivery can significantly be improved by providing sender-side mechanisms. These include (i) structuring techniques [184, 371, 183] and scalable coding [372, 319] to reduce data loss sensitivity, and (ii) forward error correction (FEC) mechanisms to lower the probability of loss at the application layer. FEC means that redundancy is added to the data so that the receiver can recover from losses or errors without any further intervention from the sender. As we have seen in the previous chapter, k video packets are protected by $n - k$ FEC packets to form an n -packet FEC block.

Clearly, under a given channel rate, the addition of FEC packets reduces the available rate for source coding. Rates allocated to source coding and FEC results thus from a trade-off whose optimization is the topic of this chapter, based on the results of Chap. 3. Also, this optimal distribution has to be dynamically adjusted according to varying video and network parameters.

The problem is therefore stated in the following manner. Let $R(t)$ denote the channel rate available for transmission at time t for a given time slot. Let $R_S(t)$ and $R_{FEC}(t)$ further denote the MPEG-2 source rate and the rate of FEC packets. The problem is then to *find* the optimal $R_S(t)$ and $R_{FEC}(t)$ at time t that *minimize* the end-to-end video distortion¹ *under* the constraint $R_S(t) + R_{FEC}(t) \leq R(t)$ and a given set of video and network parameters. The modelization of the end-to-end quality and the optimal rate distribution is very useful in traffic management of video services.

The chapter is organized as follows: The source perceptual distortion-rate function is empirically computed in Section 4.2, based on the results of the previous chapter. Section 4.3 then analyzes the error propagation and derives the degradation due to loss from the video loss patterns. The optimal rate distribution is the topic of Section 4.4. Section 4.5 shows how our algorithm behaves in different conditions, and compares the end-to-end distortion to classical FEC schemes. Finally, concluding remarks are given in Section 4.6.

Parts of this work appeared in [368, 369, 370, 350, 361].

¹We call "end-to-end video distortion" the distortion as perceived by the end-user.

4.2 MPEG-2 perceptual distortion-rate function

Several studies have already been conducted on the analysis of the rate-distortion curve for MPEG codecs [290, 373, 374]. They lead to the conclusion that the distortion evolves somehow exponentially with the decreasing source rate [124]. However, video distortion was measured by means of pure mathematical metrics such as the MSE.

Figure 4.1 shows that the perceptual video distortion varies exponentially with the quantizer scale factor (i.e., $MQQUANT$). The TV-resolution video sequence was encoded in open-loop VBR mode (OL-VBR) using a TM-5 MPEG-2 video coder [375]. The video distortion was measured by means of the PDM tool [174], which proved to behave consistently with human judgments. This tool relies on a model of the human visual system (HVS) [171]. Similar relations are however observed for MSE distortion measures [369, 370, 376].

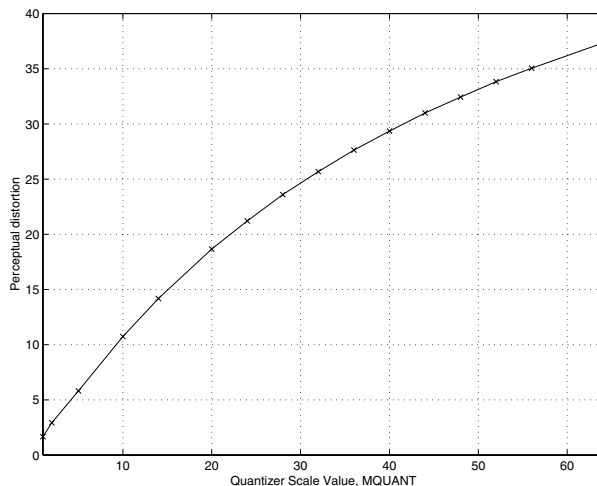


FIGURE 4.1: Perceptual distortion as a function of the quantizer scale $MQQUANT$ for ITU-R 601 test sequence.

In terms of quality, it can be shown that the quality obviously decreases with the $MQQUANT$. While the PSNR versus $MQQUANT$ curve may be represented by a decreasing exponential [377], it is to be noted that the MPQM metric exhibits a linear relationship with $MQQUANT$. This behavior is due to the scaling function of the perceptual distortion into perceptual quality [171]. We verified this important behaviour for all the GOPs of the five sequences constituting our testbed. The same characteristic has recently been observed through user's subjective evaluation [378, 379]. Computer simulation results as well as the corresponding fits are represented on Fig. 4.2 for both the *Barcelona* and *News* sequences. It has also to be noted that the slope of the quality decay is directly related to the complexity of the sequence: the higher the encoding complexity, the faster the quality decay. The video sequence *News* is a *Head and Shoulder* type of sequence and does not contain any high spatio-temporal complexities. The absolute value of the slope is thus smaller than for the *Barcelona* video sequence(see Fig. 4.2 b).

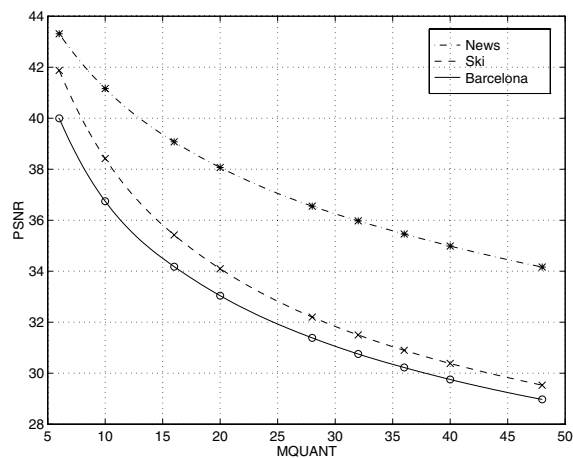
This linear relation between the video quality and the quantizer scale factor may have several impacts on the design of, for instance, perceptual rate controllers or consistent quality regulators operating in realtime [380].

Moreover, the average source rate R_S is also evolving exponentially with the $MQQUANT$ [377, 374, 373, 381], as can be shown in Fig. 4.3.

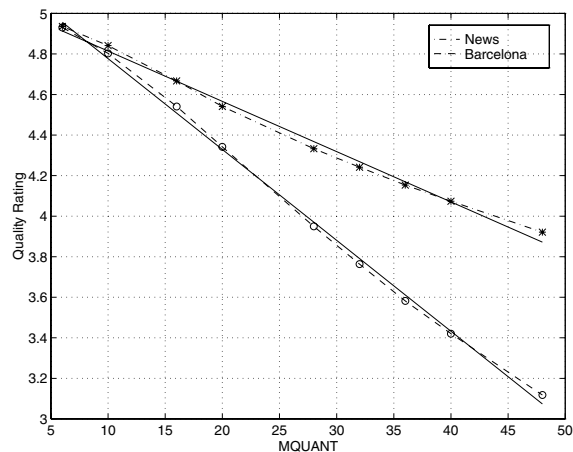
Therefore, the source perceptual distortion-rate function can be expressed as (see Figure 4.4):

$$D_S(R_S) = \chi_S R_S^{\xi_S}, \quad (4.2.1)$$

where the parameters χ_R and ξ_R are related to the encoding complexity of the set of frames under



(a)



(b)

FIGURE 4.2: *Quality versus quantizer scale. (a) PSNR versus quantizer scale factor for 3 different scenes. (b) MPQM versus quantizer scale factor for 2 different scenes.*

consideration. Details of the interpretation of Eq. (4.2.1) can be found in [370]. This relation holds at the sequence, group of pictures and even frame level [376].

Figure 4.4 exhibits an important, though trivial, behavior: for MPEG-2 source rates below 10 Mbps, a small increase in source rate leads to a great decrease in perceptual distortion. Distortion may however still be lowered by means of adaptive quantization [128, 133] or other optimization techniques [122]. In other words, equivalent distortion level can be reached at a lower average bit rate. Such optimization algorithms are however out of the scope of this work and standard MPEG-2 encoding will be used. The presented results remains completely valid in optimized encoding schemes, provided that the parameters ξ_S and χ_S are adjusted correspondingly.

A tentative of model parameterization is provided in [376]. Image activity metrics are also proposed in [382] and bit rate modelization as a function of the sequence content is proposed in [383]. However, for non-interactive video services, the parameters ξ_S and χ_S can be measured off-line on the encoded sequence, or even chosen for categorized sequence types.

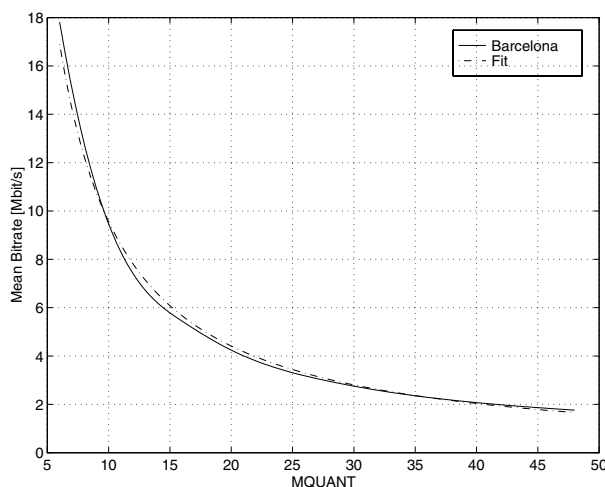


FIGURE 4.3: Average output encoding bit rate versus quantizer scale factor ($MQQUANT$) for ITU-R 601 test sequence. Exponential fitting curve.

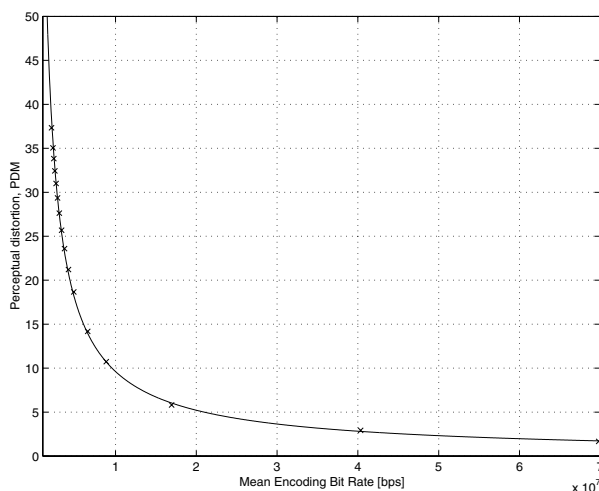


FIGURE 4.4: Perceptual distortion as a function of the mean encoding bit rate. Fitting parameters of Eq. (4.2.1): $\chi_S = 14.59 \cdot 10^6$ and $\xi_S = -0.883$.

4.3 Perceptual distortion under different loss patterns

We now investigate the impact of different packet loss patterns onto video distortion. To this aim, the influence of the loss patterns onto the number of lost pixels is first analyzed. Then the distortion is deduced from the number of lost pixels, and thus the packet loss patterns.

In MPEG-2 streaming over IP networks, a macroblock may be damaged in any of the three following cases:

1. it belongs to an RTP packet that has been lost during transmission,
2. it belongs to a slice that has been affected by a packet loss (*spatial propagation*),
3. it is temporally dependent on a damaged area of a previous reference frame (*temporal propagation*).

Intuitively speaking the perceptual distortion of the received video is in average proportional

to the number of lost macroblocks, hence to the number of lost pixels. To emphasize this, an MPEG-2 transmission system has been simulated using a 400 frames long sequence and a very large set of network loss patterns. Figure 4.5 shows the video distortion as a function of the number of lost pixels. The linear evolution is captured by a correlation factor lying around 0.992. Video distortion is thus directly related to the number of spatially lost macroblocks. Equivalently, the distortion is directly proportional to the number of spatially lost pixels, since macroblocks can only be entirely lost.

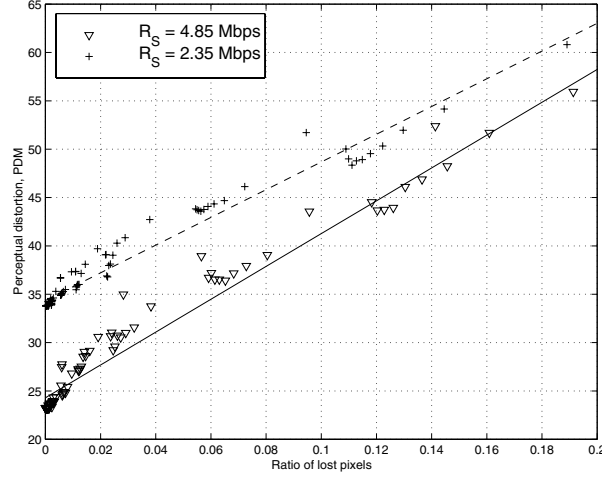


FIGURE 4.5: *Perceptual distortion versus percentage of spatially lost pixels for several loss patterns (source rate $R_S \in \{4.85, 2.35\}$ Mbps).*

Hence the perceptual distortion is driven by spatial error propagation. Temporal error propagation is clearly a direct consequence of the lost macroblocks. It is thus directly dependent on the number of lost pixels. The number of lost pixels has now to be computed from the network transmission parameters (π_v, α_v) . Notice first that the loss of a single packet could be enough to lose the entire corresponding slice. As an example, Figure 4.6 shows how data loss propagates spatially for two different loss patterns (i.e., $\alpha_v = 1$ and $\alpha_v = 2$). Since an entire slice can be fully lost with the loss of a single packet, it is desirable that a burst of lost packets damages the same slice.

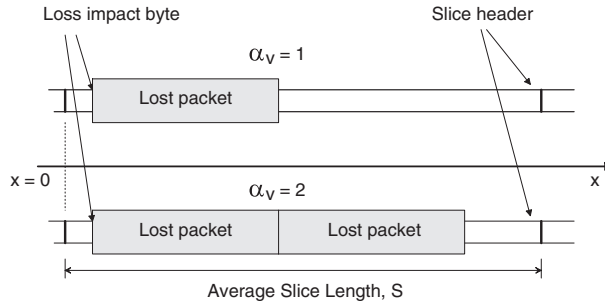


FIGURE 4.6: *Spatial error propagation under $\alpha_v = \{1, 2\}$.*

Let x denote the byte position where the loss starts. Let then σ represent the integer number of slices damaged by the loss. If P and S respectively represent the size of the transmission packet and the average slice length, σ has to verify

$$(\sigma - 1) S < \alpha_v P \leq \sigma S. \quad (4.3.1)$$

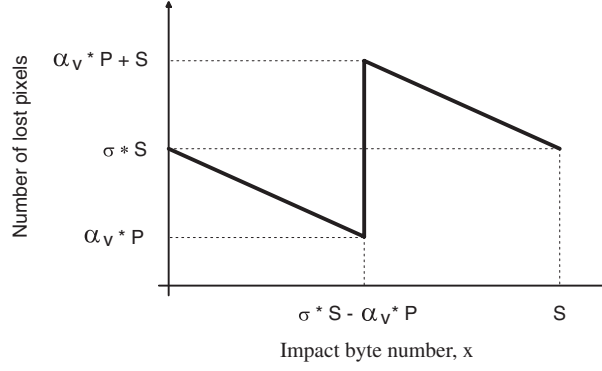


FIGURE 4.7: Number of damaged pixels versus the impact byte in the stream.

From Figure 4.6, it can be deduced that the number of lost pixels follows a piecewise linear function versus the impact byte, as represented in the Figure 4.7. The average number of lost video bytes per burst is therefore given by

$$\overline{B} = \alpha_v P + \frac{S}{2} = \alpha_v P + \frac{R_S}{2 N_S}, \quad (4.3.2)$$

where R_S represents the source rate. N_S denote the number of slices per time unit. The mean rate of packet loss bursts is then given by

$$\overline{R_B} = \frac{R_S \pi_v}{P \alpha_v}. \quad (4.3.3)$$

Finally, from Eq. (4.3.2) and (4.3.3), the rate of lost video bytes, $\overline{L}$, is simply given by

$$\overline{L} = \overline{B} \overline{R_B} = R_S \pi_v + \frac{R_S^2 \pi_v}{2 P N_S \alpha_v}. \quad (4.3.4)$$

Since in average the number of lost pixels is proportional to the number of lost video bytes, the ratio of lost pixels is equivalent to

$$P_l = \frac{\overline{L}}{R_S} = \pi_v + \frac{R_S \pi_v}{2 P N_S \alpha_v}, \quad (4.3.5)$$

where P represents the packet size (i.e., generally 184 bytes of MPEG-2 video data) and N_S represents the average number of MPEG-2 slices per second.

It has to be noticed that P_l represents also the probability for a pixel to be lost. To show the dependence of P_l on α_v , Figure 4.8 represents the number of spatially damaged pixels versus the video average burst length and the packet loss ratios π_v . The number of lost pixels is clearly decreasing in $1/\alpha_v$. Moreover, the phenomenon is more visible for low α_v values (i.e., typical values). Finally, it has to be noted that the influence of the Average Burst Length decreases when the encoding bit rate decreases, since the slice size decreases. Furthermore, it has to be noticed that the computed value of P_l is very close to experimental data. The correlation factor computed between simulation results and analytic value from Eq. (4.3.5) lies around .995 for a large set of source rate.

Accordingly to the previous development, under a given packet loss ratio, bursts of lost packets damage fewer pixels than individual lost packets. Hence, the average number of lost bytes (equivalently, the average number of spatially damaged pixels) decreases exponentially with α_v . We draw therefore the important conclusion that a uniform loss process corresponds to the worst case with respect to the amount of degradation due to data loss.

Now the effect of network losses has to be analyzed in terms of distortion. It has been shown that the perceptual distortion is proportional to the number of spatially lost pixels (see Fig. 4.5).

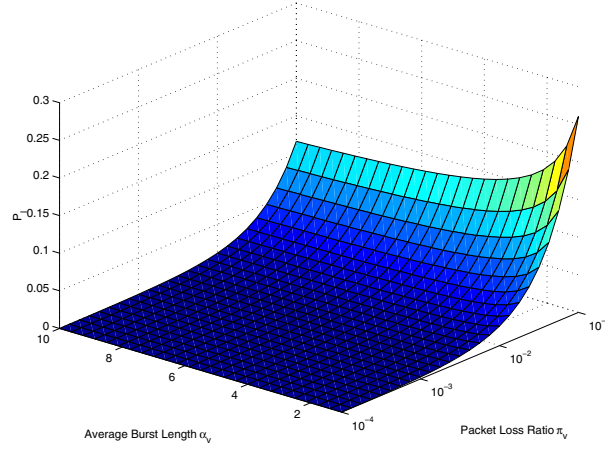


FIGURE 4.8: Number of lost pixels versus the Average Burst Length α_v and the Packet Loss Ratio π_v (source rate $R_S = 4.75$ Mbit/s).

Showing the dependence of π_v and α_v from k and n Eq. (4.3.5) leads to the distortion D_L due to data loss:

$$D_L(k, n) = \chi_L \left(\pi_v(k, n) + \frac{R_S \pi_v(k, n)}{2 P N_S \alpha_v(k, n)} \right), \quad (4.3.6)$$

where χ_L is a constant depending on the spatio-temporal complexity of the sequence and the error concealment scheme. The residual loss process parameters π_v and α_v directly depends on the FEC algorithm and the global loss process (See Chapter 3).

The variation of $D_L(k, n)$ on π_v and α_v as given by Eq. (4.3.6) is shown in Figure 4.9, for a given source rate. It can be seen that the distortion indeed increases linearly with π_v . Also, under a given π_v , the distortion decreases exponentially with α_v . This behavior has also been observed with the MSE distortion metric.

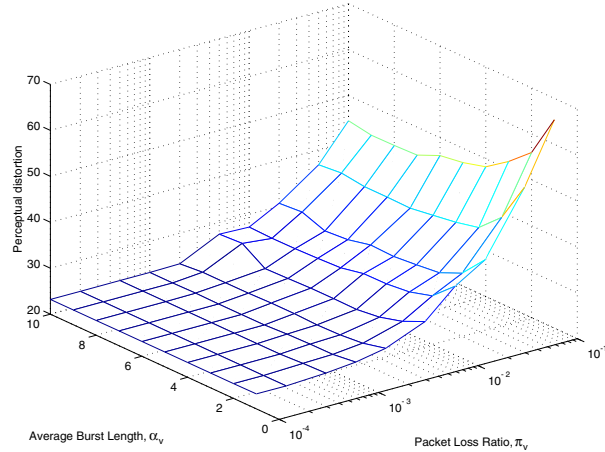


FIGURE 4.9: Perceptual distortion versus the video loss process parameters (source rate $R_S = 4.75$ Mbit/s)

Finally, the mapping between the distortion $D_L(k, n)$ and the number of spatially lost macroblocks $\overline{L_p}$ (i.e., the value of χ_L) is almost independent of the source rate [370]. In the range of MPEG-2 target rates (i.e., 3.5 to 15 Mbps), the relative distortion of concealed areas is similar

in low and high source rates. The quality of a reconstructed areas through error concealment depends on the accuracy of surrounding video elements, and thus on the source rate. For larger source rate, the reconstructed areas are better approximation of the original data. In the same time the quality of the lossless transmission also increases with the source rate. Hence the relative channel distortion is assumed to be independent of the source rate.

It is to be noted also that the Eq. (4.3.6) can be modified by error resilient encoding mechanisms as shown in Chap. 5. Equivalently, the number of lost pixels for a given PLR can be lowered by means of resynchronization mechanisms, which generally modifies the value of N_S . From Eq. (4.3.6), an important conclusion can also be drawn. That is it not efficient to simply increase the source rate in the presence of losses to improve the quality. In contrary, the distortion increases with R_S , for the same PLR. Therefore, clearly, the only way to improve the quality is either to add error control redundancy (i.e., modify the PLR experienced by the video), or to propose an error resilient coding scheme which decreases the sensitivity of the MPEG-2 stream to data loss.

4.4 Joint source/channel perceptual distortion

We now address the initial problem stated in Section 4.1 namely to find the optimal $R_S(t)$ and $R_{FEC}(t)$ at time t that *minimize* the end-to-end video distortion *under* the constraint $R_S(t) + R_{FEC}(t) \leq R(t)$ *given* a set of video and network parameters. The total bit rate $R(t)$ is one of the inputs to our system. It must be noted that this input may very well be adjusted to conform to any network policy (e.g., using the flow control algorithm of TCP [384]). Our rate allocation scheme then optimally uses the entire available rate $R(t)$.

By the FEC scheme structure, the source rate is expressed as $R_S(t) = \frac{k}{n} R(t)$ and the rate of FEC packets as $R_{FEC}(t) = \frac{n-k}{n} R(t)$ (see Figure 3.1). Recall that the distortion D_L as given by Eq. (4.3.6) represents the average distortion between a lossless and packet-lossy transmitted versions of the same MPEG-2 bit stream. We need however a total end-to-end distortion measure D between the original and received video. Consider now only the video elements (e.g., macroblocks) that are lost but replaced by error concealment at the receiver. Let $\widetilde{D}_L$ be the average distortion between these elements and their original version. The end-to-end average distortion can then be written as

$$D = D_S (1 - P_l) + \widetilde{D}_L P_l = D_S + P_l (\widetilde{D}_L - D_S), \quad (4.4.1)$$

where P_l represents the average probability for a video element (e.g macroblock) to be lost. Eq. (4.4.1) holds for MSE-like distortion metrics. Assume that it is also verified in average for perceptual distortion. In this case, we can interpret the second-term in the righthand side of Eq. (4.4.1) as the average distortion due to data loss

$$D_L = P_l (\widetilde{D}_L - D_S), \quad (4.4.2)$$

where D_L is directly related to the number of lost pixels from Eq. (4.3.6). Notice that Eq. (4.4.2) holds thanks to our definition of the distortion D_L .

Our problem becomes now the following: At time t , find the optimal FEC scheme parameters (k, n) that minimize the end-to-end distortion:

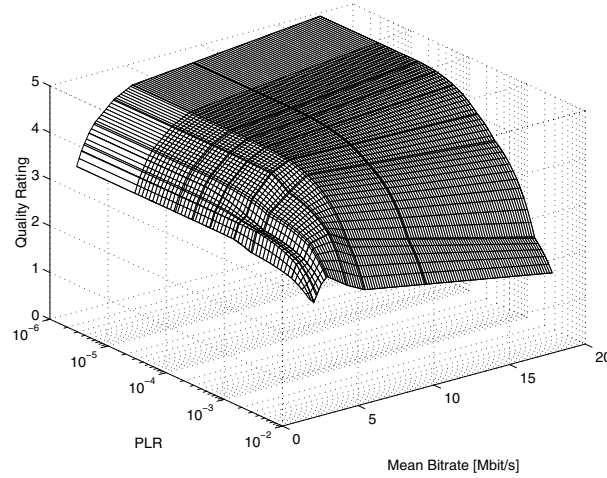
$$\min_{k \leq n \leq N} D(n, k) = \chi_S \left(\frac{k}{n} R(t) \right)^{\xi_S} + \chi_L \pi_v(k, n) \left(1 + \frac{k R(t)}{2 n P N_S \alpha_v(k, n)} \right) \quad (4.4.3)$$

given the total bandwidth $R(t)$, the channel state (i.e., p and q in the Gilbert model) and the scene-dependent parameters (χ_S, ξ_S, χ_L) . The constraint $n \leq N$ simply imposes a maximum reconstruction time for the FEC decoding.

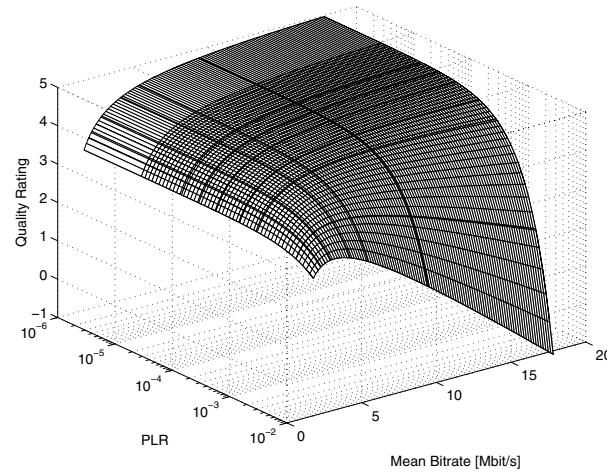
Since the parameters (k, n) only take integer values, Eq. (4.4.3) can be solved easily through numerical methods. Moreover the recursive relation of the loss parameters are not easily differentiable. The optimal values of k and n then lead to the optimal rate distribution between source

and FEC rate. In the same time, they define the optimal media-independent FEC algorithm in terms of the end-to-end distortion.

Finally we can turn the end-to-end distortion into end-to-end perceptual quality as represented in Fig. 4.10 through a simple scaling function of the form $Q = 5 (1 + 0.0625 * D)^{-1}$ [171]. In general, when considering video transmission over lossy networks, not only it is bandwidth consuming to increase the encoding bit rate above a certain threshold due to saturation of quality (which varies according to the scene complexity), it may also be quality consuming. In other words, when the user-oriented QoS is not high enough, an increase of the encoding bit rate at a fixed PLR may even degrade the quality, depending on the position of the working point on the 3D graph presented herebefore. There is an optimal bit rate to be determined that maximizes the end-user perception of the service under given network conditions.



(a)



(b)

FIGURE 4.10: (a) Q versus average bit rate and PLR: Simulations on the Barcelona sequence, without FEC protection. (b) Q versus average bit rate and PLR: Fitting function from Eq. (4.4.3) where $k = n$.

Such a conclusion is general enough to be applied to a different video encoding system. Indeed, most of the encoding standards (e.g., H.263, H.261, MPEG-1) are based on the techniques of discrete cosine transform (DCT) and motion prediction, even though they target different applications (i.e., encoding rates and qualities).

4.5 Experimental results

4.5.1 Optimal FEC parameters

The optimal FEC parameters are given in Table 4.1 for several network conditions and video scenes of different spatio-temporal complexities. The FEC reconstruction delay is set to approximately 6.5 ms (i.e., $N = 20$ at 4 Mbps and $N = 30$ at 6 Mbps). Optimal FEC parameters and hence rate distribution are then numerically computed from Eq. 4.4.3. It is shown that the FEC rate (i.e.,

PLR	ABL	$R = 4 \text{ Mbps}$				$R = 6 \text{ Mbps}$			
		Ski		Foot		Foot		News	
		k	n	k	n	k	n	k	n
0.1	1	14	18	16	20	23	29	22	29
	2	14	20	15	20	22	30	19	30
0.01	1	19	20	19	20	28	30	28	30
	2	1	1	1	1	29	30	25	30
0.001	1	1	1	1	1	1	1	29	30
	2	1	1	1	1	1	1	1	1

TABLE 4.1: Optimal FEC parameters for several transmissions conditions and scenes.

$\frac{n-k}{n}$) decreases when the global loss ratio (PLR) decreases. Moreover the required FEC rate is lower for the Foot sequence (i.e., high spatio-temporal complexity) than for the *News* video scene (i.e., low spatio-temporal complexity), at least for $PLR \leq 0.1$. This intuitive result clearly exhibits the joint role of both the source and FEC rates onto the end-to-end distortion. The evolution of the optimal parameters with the increasing global loss burstiness (ABL) is less straightforward. They indeed result from a tradeoff between the decreasing FEC efficiency and decreasing error propagation effect.

4.5.2 End-to-end quality improvement

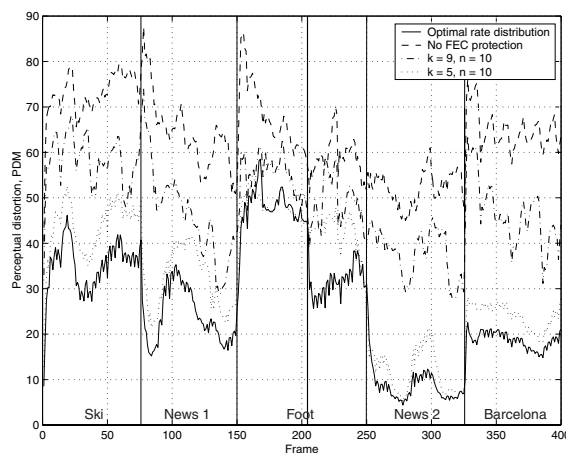
The temporal evolution of the distortion through a five-scene sequence is given in Figures 4.11(a) and 4.12(a). The distortion averaged over sliding windows through Minkowski summation is compared to the one obtained from classical FEC schemes. Thanks to its adaptivity features, our rate distribution algorithm outperforms the common schemes, as reported also by the PSNR evolution (see Figures 4.11(b) and 4.12(b)). The proposed algorithm indeed adapts to the scene complexity, to the available bandwidth and to the network conditions. It has to be noted that large loss ratio values have been chosen since the length of the sequence is relatively short. However, even if these values seem relatively high compared to usual mean ratios on today's network, they are likely to happen during small time intervals.

4.6 Conclusions

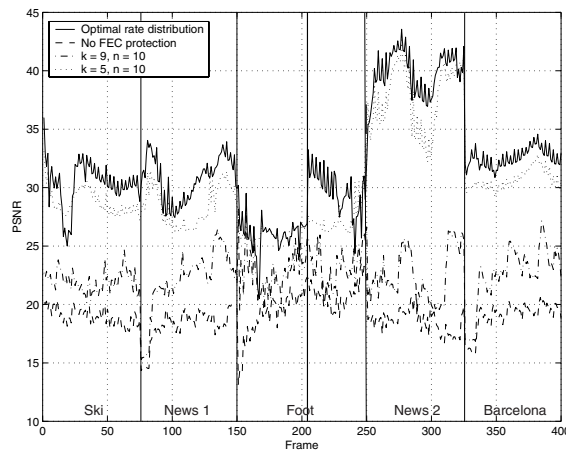
In this chapter we considered a joint source and channel coding problem. More specifically, we proposed a solution to the problem of optimal rate distribution between MPEG-2 and media-independent FEC. The optimality has been defined in terms of minimal end-to-end perceptual video distortion. The exponential source perceptual distortion-rate function has then be derived

from empirical results. The distortion due to loss has been shown to be directly proportional to the number of lost pixels. From this set of equations, the optimal rate distribution, as well as the optimal FEC scheme are obtained by solving a simple optimization problem. Finally, the proposed allocation scheme has been shown to outperform classical FEC schemes due to its adaptivity to the available bandwidth, to the network state [385] and to the video scene content.

The results presented in this chapter may greatly help in the design a video services over the Internet [386]. For example, they enable the design of network-aware rate controllers which adjust the encoding rate (MQUANT) in order to follow the optimal working point in a networking environment where the PLR may significantly vary over a video transmission duration (e.g.,



(a)

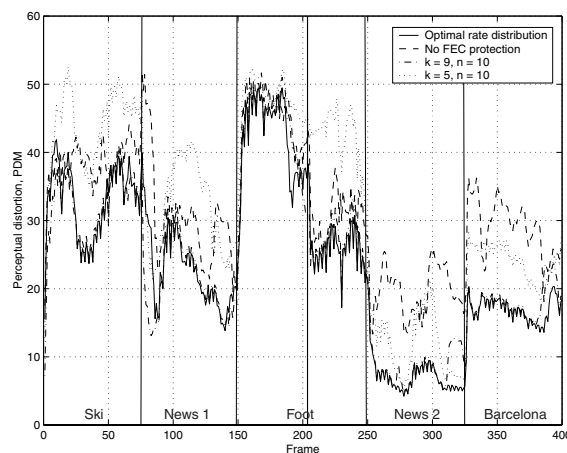


(b)

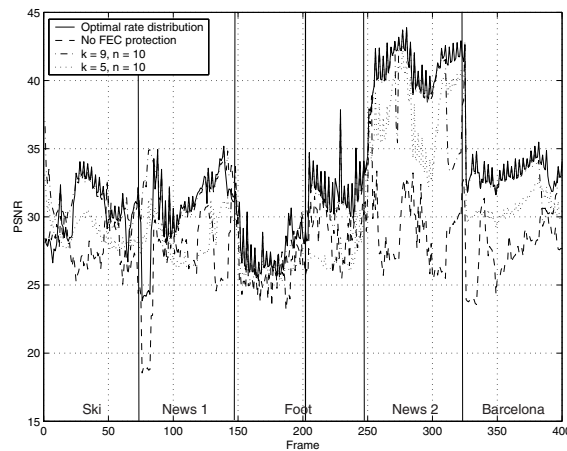
FIGURE 4.11: Perceptual distortion (a) and PSNR (b) vs frame number for optimal rate distribution scheme and classical FEC schemes. The sequence is composed of 5 different scenes. The available rate R is set to 4 Mbps for the 204 first frames, and to 6 Mbps for the 196 last frames. The global loss process parameters are set to $(PLR, ABL) = (.1, 1)$.

broadcasting via radio links, the Internet network). Network operators may therefore optimally allocate the bandwidth to the needs of their clients and offer them the best possible QoS at a given price.

Note that the same study could be applied to other encoding schemes (e.g., H.263, MPEG-4), although in this case the loss spatial propagation may be reduced. The development may also be used in more sophisticated coding schemes like the one presented in the next chapter, by simply adjusting the parameters (e.g., the mean slice length) to the coding algorithm.



(a)



(b)

FIGURE 4.12: Perceptual distortion (a) and PSNR (b) vs frame number for optimal rate distribution scheme and classical FEC schemes. The sequence is composed of 5 different scenes. The available rate R is set to 4 Mbps for the 204 first frames, and to 6 Mbps for the 196 last frames. The global loss process parameters are set to $(PLR, ABL) = (.01, 1)$.

AMISP: Content-Based MPEG-2 Encoding

5.1 Introduction

Like any other compressed data, compressed video is highly sensitive to data loss (see Section 2.2.2). Data loss propagates within the sequence and may thus become very annoying for the end-user [153]. We have seen in the previous chapter how the total bandwidth may be optimally shared between video source rate and FEC protection, for standard video coding methods. We will now see how loss impact may be reduced by appropriate joint source and channel coding.

Error resilience schemes have been introduced to limit the impairments caused by transmission losses [178]. These schemes could be roughly classified into three categories [50]. In the first category the concealment techniques try to estimate missing video data using information available at the receiver. The simplest methods would be to replace the missing block with a similar block. The motion vectors could also be approximated by those of the surrounding video blocks. However, even for the most sophisticated concealment techniques [155, 159, 201], important loss of data may lead to very annoying degradations. It becomes therefore mandatory to minimize missing information. In the second category, the resynchronization or error localization techniques aim at limiting spatial and/or temporal error propagation [185, 347, 183]. These techniques however do not take into account the local relevance of video data [247]. Finally, in the third category, unequal error protection schemes try to efficiently recover the missing video information [390, 391, 261]. They try to minimize degradations due to losses by providing different degrees of protection to each data classes. Similarly to the resynchronization techniques, the best results are however obtained only with a judicious packet prioritization process [232, 238]. In this category, layered coding [319, 372] and the Multiple Description Coding schemes [278] could be mentioned as the most promising algorithms.

Optimal error resilient schemes should not only however combine techniques of above three categories, but also exploit the local relevance of video data. Given bit budget constraints, such a combination is indeed the only way to provide the best video quality.

In this chapter, we propose an adaptive MPEG-2 information structuring and protection algorithm targeting interactive video applications. This algorithm fully exploits the nonstationary nature of the video signal exactly where it lies. It represents an ideal complement to any existing error concealment techniques available in the literature. The structuring part of the algorithm (AMIS) adaptively modulates the number of slice headers and intra-coded macroblocks in order to minimize the impact of data loss, and thus to maximize the perceived video quality. To do so, it measures the impact of an hypothetical packet loss. The protection part of the algorithm

Parts of this work appeared in [10, 387, 371, 388, 389].

(AMISP) adaptively triggers the underlying Network Adaptation Layer (NAL) to generate redundancy data (i.e., FEC packets), according to the structuring information. Data packets are thus protected whenever their loss would lead to annoying degradations in the reconstructed video. AMISP produces therefore a very low overhead in comparison to existing techniques.

5.2 Structuring and protection: problem formulation

The video error resilience could be improved in two different ways. Either the video source can not be throttled and the total bit rate simply increases with robustness overhead, or the video source rate is adapted to the robustness level to ensure a constant bit rate. In the first case the increased rate could have a very bad effect on network congestion and performances [217, 392]. In the following development the total bit rate is thus kept constant or at least constrained in a good citizen policy.

5.2.1 MPEG-2 TM5 framework

The Test Model served as a cook book for creating bit streams during the collaborative experimental phase of MPEG-2 video. Its last major update is the version 5 (TM5).

The MPEG-2 standard allows for building slices with a variable number of macroblocks. The only restriction is that a new slice shall start on every new line of macroblocks and that slices shall occur in the bit stream in the order in which they are produced. In the most widely accepted MPEG-2 TM5 implementation [375] the number of slices per frame is minimum with respect to the standard. In this scenario, every frame of a TV-resolution PAL sequence is composed of $576/16 = 36$ slices. Every slice further encapsulates $720/16 = 45$ macroblocks. In terms of transmissions packets, each slices corresponds roughly to 4 TS packets at 5 Mbps.

The MPEG-2 standard does not specify when a macroblock might be intra-coded in a non-intra picture¹. The MPEG-2 TM5 implementation decides whether or not to code a macroblock as intra, according to a per-macroblock activity metric. If the estimation error is too large for the attempted encoding by motion compensation, the macroblock is coded in Intra-mode.

5.2.2 Extra resynchronization points

MPEG-2 video streams are very vulnerable to transmission errors, mainly due to error propagation between resynchronization points (i.e., slice header or Intra-coded macroblocks). Resynchronization points aim at stopping error propagation but do not help improving video encoding quality. They constitute a pure syntactical overhead in a lossless environment but become really helpful in the presence of losses. The encoder has thus to trade off the available bandwidth resource between video encoding quality and robustness to data loss. To better understand this compromise, the impact of adding extra resynchronization points on the end-to-end quality is now briefly analyzed.

First the larger the number of slices, the bigger the overhead. Indeed, every new slice introduces a 5- to 6-byte length header which compose the major part of the overhead. It also resets the differential coding of the DC values and motion vectors. Second, the larger the number of intra-coded macroblocks, the higher the overhead. The amount of overhead generated in this case is however not easy to predict. Indeed, it depends on the encoding complexity of each extra macroblock encoded in the intra mode.

Extra overhead directly influences the encoded quality under a given bit budget. In an error-free environment the encoding quality decreases almost linearly with the percentage of slice headers. It also decreases linearly with the percentage of Intra-coded macroblocks in P-frames.

In a lossy transmission the end-to-end quality is not any more strictly decreasing with the amount of overhead. Indeed it results both from the encoding quality and the network degradation. As previously mentioned, under a given bit budget, extra resynchronization points increase the robustness of the system to data loss while reducing the bit rate allocated to encode pure

¹Macroblocks of an I-picture are obviously all intra-coded

video information. Therefore, an optimal number of slices headers and intra-coded macroblocks maximizes the end-to-end quality. This optimum is however dependent on the encoding bit rate and the packet loss ratio. Figure 5.1 illustrates this tradeoff, under a uniform and independent packet loss process assumption (i.e., $p = PLR$ and $q = 1 - PLR$ in the loss model of Fig. 3.2).

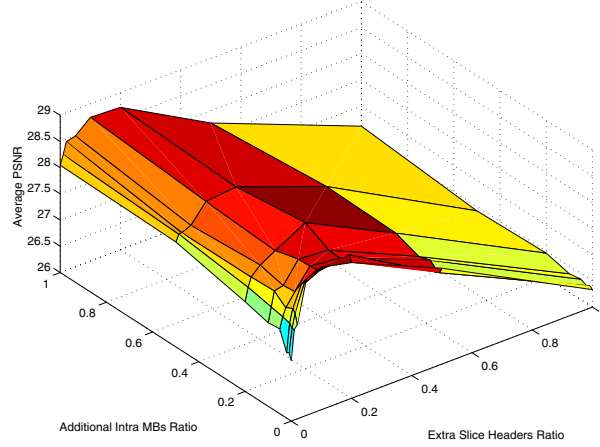


FIGURE 5.1: End-to-end PSNR quality versus extra resynchronization points ratio for the Football scene (CBR encoding at 5 Mbps and $PLR = 10^{-2}$).

5.2.3 Problem formulation

Let us assume first that the loss process is uniform and independent. This assumption greatly simplifies the error resilient scheme design. Furthermore, the uniform and independent loss process represents the worst case in MPEG-2 delivery in terms of video quality [153]. Therefore the amount of protection could only be slightly overestimated. This scenario is however preferable for the end-user to an underestimated protection.

The problem addressed in this chapter consists in finding the optimal trade-off between video information and error protection. In the same time, it is clear that a random insertion of extra resynchronization points in the bit stream or regular FEC protection would not be optimal. Indeed, the efficiency strongly depends on the content of the corresponding video area. There is no need to insert resynchronization points where the impact of data loss would not affect the video quality (under a given error concealment technique). Moreover the protection level has also to be adapted to the network performances or the expected loss probabilities.

Given (i) the expected PLR (or dynamically measured through RTCP²), (ii) the error concealment technique implemented at the decoder and (iii) a distortion metric, two related problems are considered:

1. *MPEG-2 Structuring*: determine the most appropriate MPEG-2 structure in terms of resynchronization points location.
2. *MPEG-2 Protection*: derive a content-based FEC scheme to protect areas where structuring is not sufficient.

5.3 Adaptive MPEG-2 structuring

5.3.1 Loss probability matrices

It has been noticed that a macroblock may be damaged in any of the three following cases:

²RTCP stands for Real-Time Control Protocol

1. it belongs to an RTP packet that has been lost during transmission
2. it belongs to a slice that has been affected by a packet loss (spatial propagation)
3. it is temporally dependent on a damaged macroblock of a previous reference frame (temporal propagation).

which are discussed below.

The first factor that may affect a macroblock is the transmission error. If we assume a uniform loss pattern, the probability θ for an RTP packet to be lost is given by the PLR experienced on the network. Therefore, without any other information about the packet loss process, every RTP packet has the same average probability to be lost, $\theta = PLR$. Let us now call $B_n(i, j)$, the macroblock at the i^{th} column and the j^{th} row of a given frame n . Under the assumption that a macroblock is lost as soon as part of it is missing, the probability $\lambda_n(i, j)$ for the macroblock $B_n(i, j)$ to be lost is given by:

$$\lambda_n(i, j) = \theta \mathbf{N}_n(i, j), \quad \forall (i, j) \mid 1 \leq i \leq B_{row} \text{ and } 1 \leq j \leq B_{column}. \quad (5.3.1)$$

where $\mathbf{N}_n(i, j)$ is the number of RTP packets containing the macroblock $B_n(i, j)$. B_{row} and B_{column} are respectively the number of macroblocks per frame row and column.

Even at high encoding rates, loss entities (i.e., roughly multiples of 188 bytes) are much larger than the macroblock size. Thus, in general, macroblocks belong to at most two RTP packets (i.e., $\mathbf{N}_n(i, j) \in \{1, 2\}$).

The second factor that may affect a macroblock is *spatial propagation*. In case of transmission error, an MPEG-2 decoder skips all video information up to the next slice header, which acts as a spatial resynchronization point. Consequently, when a macroblock is lost within a slice, all subsequent macroblocks of the same slice are considered as being damaged, even if they do not belong to the lost RTP packet.

Thus, for a given frame n , the probability $\mathbf{P}_n(i, j)$ for a macroblock $B_n(i, j)$ not to be correctly decoded (transmission error + spatial propagation) is given by:

$$\begin{aligned} \mathbf{P}_n(i, j) &= \lambda_n(i, j) + \theta \mathbf{M}_n(i, j) \\ &= \theta [\mathbf{N}_n(i, j) + \mathbf{M}_n(i, j)], \\ &\quad \forall (i, j) \mid 1 \leq i \leq B_{row} \text{ and } 1 \leq j \leq B_{column}. \end{aligned} \quad (5.3.2)$$

where, $\mathbf{M}_n(i, j)$ represents the number of RTP packets within the same slice before the first packet related to $B_n(i, j)$ (see Figure 5.2).

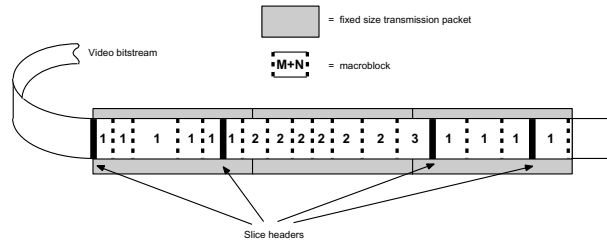


FIGURE 5.2: Illustration of $\mathbf{P}_n(i, j)$. The numbers in each macroblock represent $M_n + N_n$ values.

There is an exception to this rule. Indeed, according to the MPEG-2 syntax, every frame is preceded by a header. If a packet containing a frame header is lost, the entire frame is skipped, making Eq. (5.3.2) useless. We assume this case to be rare enough to be neglected. This assumption is enforced when these headers are protected via a specific FEC scheme [248].

The third factor that may affect a macroblock is *temporal propagation* [391]. Our objective is to derive the loss probability matrix $\mathcal{E}_n^{n-k}$ for every pixel of frame n due to temporal propagation of damaged pixels in frame $(n - k)$, with $k \leq n$.

Since motion estimation does not consider macroblock boundaries, but rather references areas of 16 by 16 pixels, the granularity of the loss probability matrix $\mathbf{P}_n$ must be refined to the pixel level. Indeed, in Eqs. (5.3.1) and (5.3.2) the entries of the matrix $\mathbf{P}_n$ refer to macroblocks whereas we need now reference to pixels. The *loss probability matrix* due to spatial propagation for every pixel of frame n is called $\mathcal{P}_n$. The mapping between $\mathbf{P}_n$ and $\mathcal{P}_n$ is straightforward. Indeed, every pixel of a given macroblock has the same probability to be lost. Hence $\mathcal{P}_n$ is obtained by the Kronecker product of a 16×16 unity matrix $\mathbf{I}_{16}$ by $\mathbf{P}_n$:

$$\mathcal{P}_n = \mathbf{I}_{16} \otimes \mathbf{P}_n. \quad (5.3.3)$$

The resulting matrix $\mathcal{P}_n$ has therefore the same size as the video frame (i.e., 720×576 in the ITU-R 601 format).

In the following development, the B-frames are not considered for additional intra-coded macroblock insertion. Indeed, B frames do not propagate degradation, since they are never referenced. Therefore, the impact of data loss in B-frames is barely visible (the temporal resolution of the human visual system is larger than a single frame duration [171]). These frames may also offer the highest compression ratio, and adding intra-coded macroblocks would result in the highest relative overhead. Finally, B-frames have generally the smallest number of bits so that losses have low chance to occur in these frames. Additional intra-coded macroblock in B-frames would therefore result in a waste of bandwidth.

Temporal propagation means that a pixel in the current frame is damaged because it refers to a badly decoded pixel from a previous reference frame. This badly decoded pixel may result from transmission error, spatial propagation, or/and temporal propagation, as explained before. Thus, $\mathcal{E}_n^{n-k}$ obviously depends on $\mathcal{P}_m$, with $m = n - k, n - k + 1, \dots, n$. Moreover, $\mathcal{E}_n^{n-k}$ needs to be computed in a recursive manner. Indeed, video areas each pixel refers to, have to be found by recursively following the successive motion vectors within the video sequence. Thus, within a macroblock, even though each pixel has the same probability to be lost, they do not have the same probability to be decoded into an erroneous value. They actually do not necessarily refer to the same macroblock in reference frames.

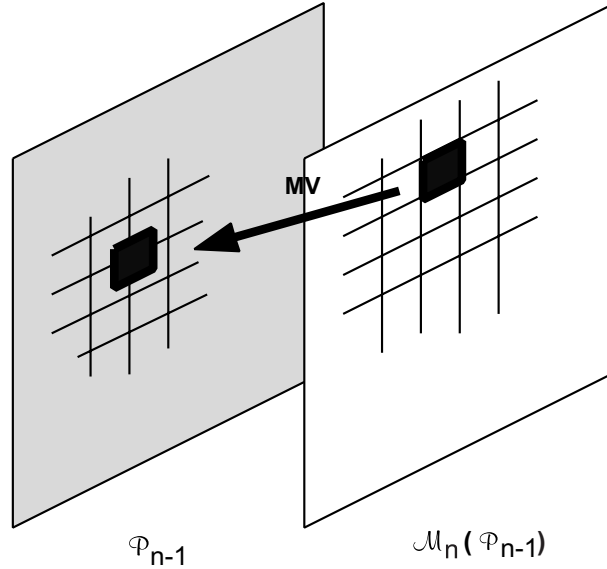


FIGURE 5.3: Mapping function $\mathcal{M}_n$ of loss probability matrices.

The loss probability matrix $\mathcal{E}_n^{n-1}$, derived from temporal propagation of errors occurring in reference frame $(n-1)$ and impacting frame n is first computed. The probability for all the pixels in frame n to be damaged by losses occurring in frame $(n-1)$ could be easily derived. First, the

motion vectors of frame n are used to reference matrix $\mathcal{P}_{n-1}$. Actually, the motion estimation performed by MPEG-2 is applied on the "frame" $\mathcal{P}_{n-1}$, on a macroblock basis. This mapping operation could be denoted by $\mathcal{M}_n(\mathcal{P}_{n-1})$ (see Figure 5.3). Finally, each element of $\mathcal{M}_n(\mathcal{P}_{n-1})$ should be multiplied by the probability for the corresponding pixel not to be lost in frame n . Indeed, there is no need to compute the probability for a pixel to be damaged in a previous frame if it is lost in the current frame n (it would make the consideration of temporal propagation useless). Therefore, $\mathcal{E}_n^{n-1}$ can be written as follows:

$$\begin{aligned} \mathcal{E}_n^{n-1}(i, j) &= \mathcal{M}_n(\mathcal{P}_{n-1})(i, j) (1 - \mathcal{P}_n(i, j)), \\ \forall (i, j) \mid 1 \leq i \leq W \text{ and } 1 \leq j \leq H. \end{aligned} \quad (5.3.4)$$

where $\mathcal{E}_n^{n-1}(i, j)$ represents the probability for pixel at i^{th} column and j^{th} line in frame n to be damaged by losses in reference frame $n-1$. The variables W and H correspond respectively to the number of pixels per row and column. Obviously, if the pixel given by (i', j') in frame n does not reference any video area of frame $(n-1)$, or if it belongs to an intra-coded macroblock, $\mathcal{E}_n^{n-1}(i', j') = 0$.

The probability matrix for pixels not to be spatially lost could be written as:

$$\overline{\mathcal{P}}_n = \mathbf{I}_{W,H} - \mathcal{P}_n, \quad (5.3.5)$$

where $\mathbf{I}_{W,H}$ is a $W \times H$ unity matrix. Eq. (5.3.4) could now be generalized, taking into account losses in any of the k previous reference frames, with $k \leq n$. The generic loss probability matrix due to temporal propagation, $\mathcal{E}_n^{n-k}$, can be obtained via recursion. Indeed, similarly to Eq. (5.3.4), $\mathcal{E}_{n-k+1}^{n-k}$ is given by:

$$\begin{aligned} \mathcal{E}_{n-k+1}^{n-k}(i, j) &= \mathcal{M}_{n-k+1}(\mathcal{E}_{n-k}^{n-k})(i, j) \overline{\mathcal{P}}_{n-k+1}(i, j), \\ \forall (i, j) \mid 1 \leq i \leq W \text{ and } 1 \leq j \leq H. \end{aligned} \quad (5.3.6)$$

with,

$$\mathcal{E}_{n-k}^{n-k}(i, j) = \mathcal{P}_{n-k}(i, j). \quad (5.3.7)$$

The process can then be generalized starting with the condition hereabove (Eq. (5.3.7)). It becomes:

$$\mathcal{E}_{n-k+m}^{n-k}(i, j) = \mathcal{M}_{n-k+m}(\mathcal{E}_{n-k+m-1}^{n-k})(i, j) \overline{\mathcal{P}}_{n-k+m}(i, j), \quad m = 1, 2, 3, \dots, k. \quad (5.3.8)$$

Following the same notation, $\mathcal{M}_{n-k+m}$ uses the motion vectors of frame $(n-k+m)$. Moreover, when a pixel given by (i', j') in one of the reference frames $(n-k+m)$ belongs to an intra-coded macroblock, or has no correspondence in its direct reference frame (according to $\mathcal{M}_{n-k+m}$), then:

$$\mathcal{E}_{n-k+m}^{n-k}(i', j') = 0. \quad (5.3.9)$$

Finally, $\mathcal{E}_n^{n-k}$ is obtained when $m = k$ in Eq. (5.3.8).

$\mathcal{E}_n^{n-k}$ represents the generic loss probability matrix for frame n due to temporal propagation of data loss in frame $(n-k)$. For sake of simplicity, the impact of data loss in each reference frames $(n-k')$ on $\mathcal{E}_n^{n-k'}$ is considered independently.

5.3.2 Proposed algorithm: AMIS

The proposed algorithm for adaptively inserting resynchronization points in an MPEG-2 bit stream, namely the Adaptive MPEG-2 Information Structuring (AMIS) algorithm, is now presented. AMIS strongly relies on the study presented in the previous subsection. Intuitively, it works as follows: an extra resynchronization point is inserted in the bit stream whenever hypothetical data loss, following a uniform loss process, would lead to video degradation above a desired threshold, after error concealment.

The *mean luminance difference* (MLD) has first been chosen as distortion measure. It corresponds to the simplest metric correlated with human perception [159] (under the assumption that the viewer stands far enough from the monitor). The distortion is computed between the current macroblock after encoding (generally given by the encoding scheme) and the same macroblock impaired by loss. This one is obtained by simulating losses and concealment in the encoder. The real effects of hypothetical losses are thus captured by the encoding scheme. The MLD for $B(i, j)$ is defined as follows:

$$\delta(i, j) = \left| \frac{1}{256} \sum_{p=1}^{256} B^p(i, j) - \frac{1}{256} \sum_{p=1}^{256} \tilde{B}^p(i, j) \right|, \quad (5.3.10)$$

where, (i, j) is the macroblock position in the frame and p is the pixel position in the corresponding macroblock. $B(i, j)$ and $\tilde{B}(i, j)$ represent respectively a correctly (error-free) decoded macroblock and the corresponding damaged macroblock.

The error concealment technique implemented at the decoder should also be specified to build the optimal structuring. However, if the error concealment technique is not known *a-priori*, the structuring algorithm would still produce good results, since major error concealment schemes have similar features. To be specific, the motion compensated concealment technique has been chosen for its simplicity.

It has to be noticed that the AMIS algorithm would not need any major modification if a different distortion measure and/or error concealment technique was imposed.

AMIS is divided in two distinct parts: (i) the *spatial* part, which deals with slice headers insertion, and (ii) the *temporal* part, which is in charge of deciding when a macroblock should be intra-coded. Indeed, inserting extra slice headers has no effect on temporal error propagation. Also, adding intra-coded macroblock does not help in limiting the spatial error propagation. Therefore, these two parts are considered independently. However, it is clear that the slice structure of reference frames may influence the insertion decision of intra-coded macroblocks.

AMIS-Spatial The *spatial* part of AMIS aims at limiting the spatial error propagation, or at least its visible degradation. It introduces an extra slice header as soon as the distortion due to hypothetical loss reaches a given threshold, Δ_s . Clearly a new slice is inserted as soon as:

$$\sum_{B_n(i, j) \in S} \delta_n^s(i, j) \mathbf{P}_n(i, j) \geq \Delta_s, \quad (5.3.11)$$

where, $B_n(i, j)$ is the current macroblock belonging to slice S and, $\delta_n^s(i, j)$ corresponds to the expected MLD in case $B_n(i, j)$ was damaged. $\mathbf{P}_n(i, j)$, defined in Eq. (5.3.2), represents the probability for $B(i, j)$ to be spatially damaged, by packet loss or spatial propagation.

Actually, the expected distortion is weighted by its likelihood to occur. There is indeed no need to protect an area not likely to be lost, even if the corresponding distortion would be high.

The spatial threshold Δ_s regulates the acceptable level of distortion. The smaller the threshold, the higher the number of slices.

AMIS-Spatial also takes the packetization process into account: no more than one slice header is encapsulated in the same network loss entity [183].

AMIS-Temporal The *temporal* part of AMIS is more complex. First, let us assume that losses in different reference frames can be considered independently in regard to their impact on the current frame. Even though not completely correct, this assumption places the encoding process in the worst case from the distortion point of view. It will tend to generate more protection than effectively needed, but greatly simplifies the AMIS mechanism.

AMIS-Temporal analyzes every single macroblock and decides whether or not to intra-code it. Again, this decision depends on the macroblock distortion due to temporal propagation of data loss.

The decision may be expressed as follows. The distortion due to temporal error propagation is weighted by the corresponding loss probability matrix and compared to a threshold Δ_t . This weighted distortion is obtained by summing the effects of uniformly-distributed packet losses in every single previous reference frame, up to the last intra-coded picture ($n - I$). Finally, the condition for a macroblock $B_n(i, j)$ to be intra-coded in frame n is given by:

$$\sum_{k=1}^I \left(\frac{1}{256} \sum_{p \in B_n(i, j)} \mathcal{E}_n^{n-k}(p) \delta_{\mathbf{n}, \mathbf{k}}^t(i, j) \right) \geq \Delta_t, \quad (5.3.12)$$

where $\mathcal{E}_n^{n-k}$ is given by Equation (5.3.8). The expected MLD between the current MB correctly decoded and its substitute in case of loss in the reference frame k is given by $\delta_{\mathbf{n}, \mathbf{k}}^t(i, j)$. Again, the temporal threshold Δ_t regulates the acceptable level of distortion. The smaller the threshold, the higher the number of intra-coded macroblocks.

Finally, a maximum refresh period, T_{max} , is also imposed. This period corresponds to the maximum number of frames a pixel may subsist without any intra-reference. When a pixel has no intra-reference for a period longer than T_{max} , the macroblock shall be intra-coded. This consideration is particularly useful in large GOP encoding schemes, or in case of large intervals between consecutive I-frames.

5.3.3 Experimental results and comparisons

The AMIS algorithm is now evaluated and compared to other encoding schemes in terms of final video quality. Figures 5.4 and 5.5 compare the behavior of (i) AMIS, (ii) a random resynchronization points insertion scheme (iii) the common TM-5 model [375] and (iv) the algorithm proposed by Richardson and Riley [185]. The comparison is performed in terms of final video quality under several packet loss ratios and a given bit budget. The PLR has been allowed to vary between 10^{-1} and 10^{-4} . These values could seem quite larger than commonly accepted networking performances. However the latter are generally average values computed over entire sessions. They do not reflect the short-term characteristics of the losses. Such RTP packet loss ratio values could likely be met during small periods of time (e.g., network congestion, atmospheric conditions). Finally, the video streams have been encoded at a constant bit rate of 6 Mbps. CBR encoding mode imposes the most stringent constraint on bit rate allocation. However, similar (and, certainly better) results could easily be obtained in OL-VBR encoding mode.

AMIS is obviously dramatically better than the TM-5 algorithm under medium to high PLRs. Also, under low PLRs, it is comparable to the TM5. Indeed, AMIS judiciously shares the total bit budget between pure video information and additional resynchronization points, according to the expected PLR. Moreover, the comparison of AMIS versus the random insertion of extra resynchronization points scheme, for the same overhead, clearly shows the relevance of the content-based structuring. Finally, AMIS offers better results than the algorithm proposed by Richardson and al., especially in bad transmission conditions. The latter algorithm indeed basically uses a static slice length for each frame type (i.e, I-, P- or B-frame). This length stays valid for the whole sequence, without PLR values considerations. All these results highlight the need for adaptivity to both video-content and transmission quality.

Fig. 5.6 and 5.7 finally shows a typical frame encoded with the AMIS algorithm. The additional slice headers and Intra-macroblocks are designed on the decoded frame. It can be seen that the slice header and Intra-macroblock insertion follows the high activities regions within the frame, as expected. Indeed, the concealment is not efficient in these regions, and the error propagation would be highly visible in case of losses. In the contrary, in uniform regions, the concealment provides a sufficient reconstruction quality, so that additional overhead is not necessary.

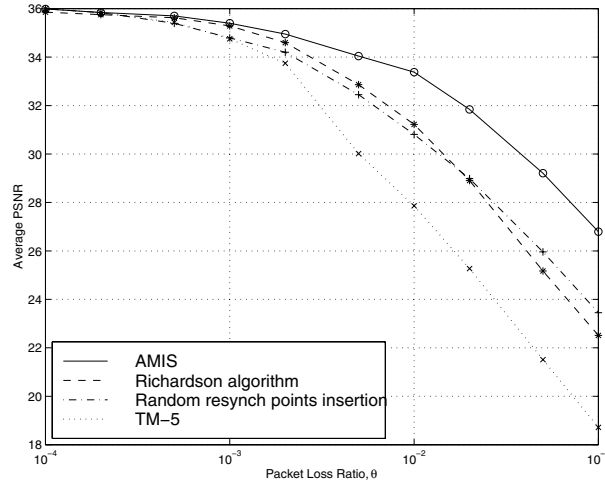


FIGURE 5.4: PSNR end-to-end quality versus PLR in comparison to random resynchronization points insertion, TM-5 encoding scheme and encoding algorithm proposed by Richardson (CBR encoding at 6 Mbps).

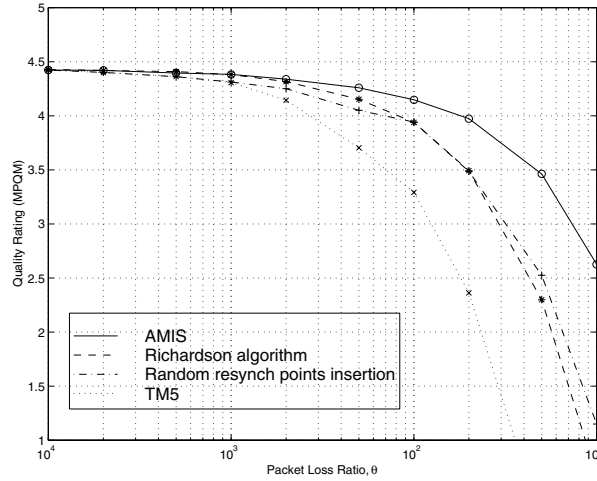


FIGURE 5.5: AMIS end-to-end quality versus PLR in comparison to random resynchronization points insertion, TM-5 encoding scheme and encoding algorithm proposed by Richardson (CBR encoding at 6 Mbps).

5.4 Adaptive MPEG-2 structuring and protection

5.4.1 FEC-based protection

In Section 5, AMIS has been presented. It smartly structures the MPEG-2 bit stream to make it more robust against data loss. However, it is clear that data loss may still induce unacceptable degradation in the reconstructed video. Indeed, some video areas may be highly sensitive to loss (e.g., fast-moving areas). In this section, AMIS is extended with a FEC-based protection scheme, to become AMISP.

Forward Error Correction (FEC) means that redundancy is added to the data so that the receiver can recover from losses or errors without any further intervention from the sender. As we have seen in Chap. 3, usual FEC schemes build n_{FEC} packets blocks where k_{FEC} video packets are protected by means of $n_{FEC} - k_{FEC}$ redundancy packets (see Figure 3.1). Recall that such



FIGURE 5.6: Sample frame of ITU-R test sequence, where AMISP additional slice headers are designed in black.



FIGURE 5.7: Sample frame of ITU-R test sequence, where AMISP additional Intra-macroblocks are designed in light gray.

schemes are able to recover up to $n_{FEC} - k_{FEC}$ lost packets in a block of n_{FEC} packets [362]. The video packet loss process is then modified, and the resulting Packet Loss Ratio for FEC-protected packets becomes θ_{FEC} . Under the assumption of independent losses, θ_{FEC} , is given by [325]:

$$\theta_{FEC} = \theta \left[1 - \sum_{i=0}^{n_{FEC} - k_{FEC} - 1} \binom{n_{FEC} - 1}{i} \theta^i (1 - \theta)^{n_{FEC} - i - 1} \right], \quad (5.4.1)$$

Several criteria have then to be considered in the choice of the FEC parameters n_{FEC} and k_{FEC} . First, the overhead $\frac{n_{FEC} - k_{FEC}}{n_{FEC}}$ has to be kept as small as possible and to be adapted to the expected loss ratio. However, this ratio does not need to be very large to ensure a good recovery probability. It has been shown indeed that, even for a small $\frac{n_{FEC} - k_{FEC}}{n_{FEC}}$ ratio, FEC can be very effective and reduces the loss probability by several orders of magnitude [392]. The ratio efficiency/overhead is moreover larger for large k_{FEC} values, assuming that losses occur independently [350].

Second, the FEC scheme has to satisfy strict delay constraints in interactive applications. The

delay introduced by FEC reconstruction³ should not be much larger than one frame, since other delays are also introduced along the transmission path. Since one TS packets represents already a delay of about 5.6 ms in a 6 Mbps connection, n_{FEC} should not be larger than 10 to 15 packets. This value is however directly dependent on the bit rate.

Third, it has been shown that for a given overhead, large values of k_{FEC} lead to the best reconstruction probabilities [390]. On the other hand, small k_{FEC} values ensure a more efficient protection of elected packets in an adaptive algorithm.

All the previous statements suggests that the value of n_{FEC} should be chosen as large as possible, given some delay constraints. Then k_{FEC} value should be computed accordingly to offer a sufficient protection, but also to minimize the overhead.

Finally, FEC parameters could vary dynamically according to loss patterns (i.e., PLR and ABL). On-going work is currently trying to optimize these parameters according to network conditions and the degree of protection accuracy. For sake of simplicity, $n_{FEC} - k_{FEC} = 1$ in the following experiments. This allows moreover a very simple and rapid exclusive-OR based FEC scheme.

5.4.2 The adaptive protection algorithm: AMISP

The proposed protection algorithm is the following. During the encoding process, a packet p is marked to be protected whenever its hypothetical loss would introduce an unacceptable degradation. Similarly to Eq. (5.3.11), the loss probability weighted distortion is compared to a third threshold Δ_{FEC} :

$$\sum_{B_n(i,j) \in p} \delta_s(i,j) \theta \geq \Delta_{FEC}, \quad (5.4.2)$$

Whenever AMISP decides to protect a packet, it triggers the underlying network adaptation layer (NAL). The NAL starts counting k_{FEC} video packets and then inserts $n_{FEC} - k_{FEC}$ FEC packets in the MPEG-2 bit stream. Of course, if the elected packet already belongs to a FEC block, no additional overhead is inserted. Like in the structuring scheme, the amount of redundancy is driven by the threshold Δ_{FEC} which represents the Quality of Service desired at the receiver. Finally, the adaptive FEC algorithm is easily implemented on RTP protocols, thanks to the support for FEC protection [393].

The structuring part of AMISP still works in the same manner. However, the macroblock loss probability $\mathbf{P}_n$ (see Eq. (5.3.2)) becomes $\widetilde{\mathbf{P}}_n$ and is now given by:

$$\widetilde{\mathbf{P}}_n(i,j) = \sum_{p=1}^{N_n(i,j)} \theta_p + \sum_{p=1}^{M_n(i,j)} \theta_p, \text{ with } \theta_p \in \{\theta, \theta_{FEC}\}, \quad (5.4.3)$$

where $M_n(i,j)$ still represents the number of RTP packets within the same slice before and excluding $B_n(i,j)$. $N_n(i,j)$ represents the number of packets containing part of the macroblock $B_n(i,j)$. The packet loss probability θ_p is either equal to θ_{FEC} or θ , depending on whether the packets are FEC-protected or not.

It has to be noticed that packets are FEC-protected in regard to their influence onto spatial distortion. These packets very likely contain a slice header due to the similarity between relations (5.3.11) and (5.4.2). However, the temporal propagation phenomenon is not taken into account by the protection decision process. The reasons of this choice are twofold. First, the temporal propagation of an error in the current frame cannot be predicted in one-pass encoding. Second, it can be assumed that the most relevant packets (i.e., FEC-protected packets) are the packets that would also cause the highest temporal distortion.

Moreover, the major MPEG-2 headers (i.e., sequence and picture headers) are also FEC protected [248]. Their loss would indeed cause a really important degradation. Each packet containing

³The FEC block construction at the sender does not introduce any queuing delay

such crucial information is therefore protected by a FEC packet. Finally, the rate control part of the encoding algorithm has been slightly modified. Indeed, the video encoding rate has to be adapted to the protection overhead to ensure a constant total bit rate. Basically, the modifications simply includes the number of bits used for protection in the loop of the TM-5 rate control algorithm [375].

5.4.3 Experimental results and comparisons

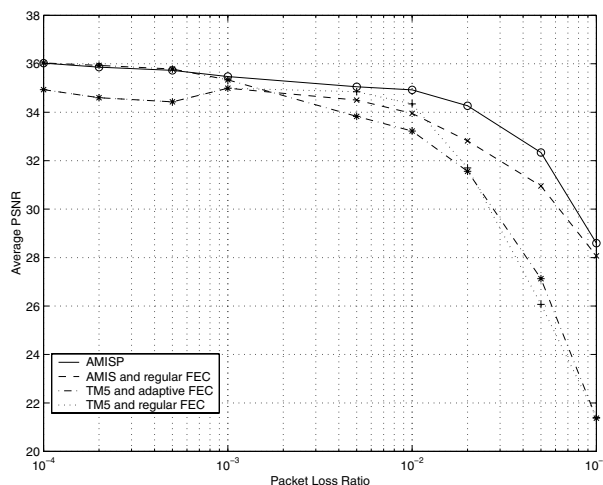


FIGURE 5.8: End-to-end PSNR quality versus the packet loss ratio. Comparison of the AMISP algorithm ($k_{FEC} = 10$ and $n_{FEC} = 11$) with a TM-5 encoding with an adaptive and regular FEC scheme ($k = 10$ and $n = 11$). The total bit rate is about 6 Mbps.

As stated before, the algorithm inserts only a single packet per FEC block (i.e., $n_{FEC} - k_{FEC} = 1$). The length of the FEC blocks (i.e., k_{FEC}) could be determined through simulations of different AMISP schemes [350]. As stated before, $k_{FEC} = 10$ seems therefore to fit both delay and robustness requirements, at least in the most common θ range (i.e., between 10^{-4} and 10^{-1}). The total bit rate (i.e., video information and FEC overhead) is fixed to 6 Mbps for each transmission. Similar results are presented in terms of both PSNR and perceptual quality.

Figures 5.8 and 5.9 compare AMISP with several protection schemes. First a basic TM-5 video encoding protected by a regular (by opposition to adaptive) FEC scheme is proposed. It generates one redundancy packet every ten video packets. It is clearly visible that AMISP provides a better end-to-end quality over the complete packet loss ratio range. At low θ values, the improvement in quality is mainly due to the adaptivity feature of AMISP: it generates less redundancy, and thus provides more accurate video information. At high loss rates, both schemes perform similarly in terms of protection. However, the quality offered by AMISP is much higher thanks to the underlying structuring scheme (i.e., AMIS). This scheme indeed greatly limits the residual error propagation within the decoded sequence.

Figures 5.8 and 5.9 also emphasize the useful adaptivity feature to network conditions. The end-to-end quality of AMISP is compared to the same AMIS video bit rate but protected by a regular FEC scheme. At low loss ratio AMISP provides a better quality since it does not generate useless overhead. However the difference is not very large. The relatively small FEC overhead only slightly decreases the encoding quality at medium encoding rate. Meanwhile, both algorithms are equivalent at high loss ratios. The number of packets AMISP has to protect becomes very large in these conditions. Hence, the adaptive protection becomes very close to a regular protection scheme.

Figures 5.8 and 5.9 finally demonstrate the advantages of the underlying structuring scheme. AMISP is compared to a TM-5 encoding protected by the same adaptive FEC scheme used in

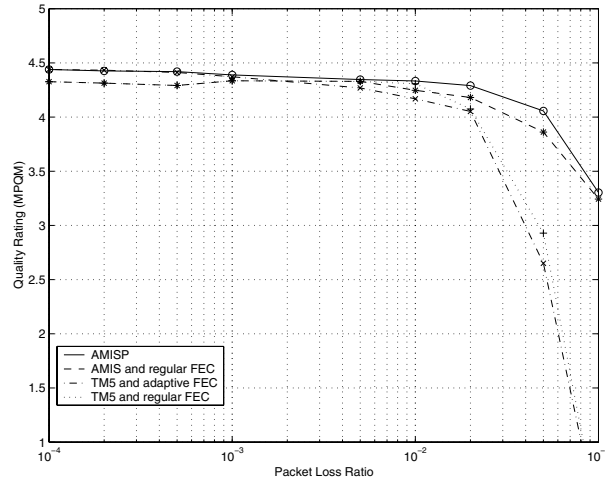


FIGURE 5.9: *End-to-end perceptual quality versus the packet loss ratio. Comparison of the AMISP algorithm ($k_{FEC} = 10$ and $n_{FEC} = 11$) with a TM-5 encoding with an adaptive and regular FEC scheme ($k = 10$ and $n = 11$). The total bit rate is about 6 Mbps.*

AMISP. It is clear that both algorithms leads to the same quality at low loss rates. Indeed, losses that would cause important degradations are recovered by the FEC algorithm. However, the gap between both schemes grows rapidly with the loss ratio. Indeed, the protection algorithm loses some of its efficiency. The errors propagate within the TM-5 sequence, while they are kept to an acceptable level in AMIS.

5.5 Conclusions

We have presented a new adaptive error resilient scheme for TV-resolution MPEG-2 video streams interactive delivery, namely AMISP. It includes a media-dependent FEC algorithm relying on an MPEG-2 syntactic structuring technique. A judicious combination of protection redundancy, MPEG syntactic data and pure video information showed to greatly improve the final quality under a given bit budget. Experimental results have shown that AMISP dramatically outperforms existing techniques, thanks to its efficient adaptivity. Major improvements are due to adaptivity of both FEC protection and structuring at respectively low and high loss ratios. Moreover, it must be noted that AMISP does not significantly increase the MPEG-2 encoding complexity. However, the protection part of AMISP requires the underlying layer (NAL) to provide FEC capabilities. If this is not the case, only the structuring part can be used.

In this work, retransmission of missing packets was assumed not to be feasible due to stringent timing constraints. However, we believe that retransmission might lead to an improved scenario in one-way real-time MPEG-2 delivery. Finally, AMISP could also be applied to other video standards at the cost of a few modifications. The mechanism is also flexible enough to still work properly with a different error concealment technique and a different macroblock-based quality degradation metric.

From Standards to Matching Pursuit Codecs

6.1 Yet another codec ?

Streaming systems tailored for IP applications have to find a good balance between several different and possibly contradictory requirements in today's heterogeneous and error-prone environments (see Fig. 6.1). First, the compressed video has to be scalable to serve different end-users D_i with the same initial stream. The coding efficiency has thus to be maintained across a wide range of bit rates. The transcoding or management operations in the gateways G_i have to be kept minimal in order not to overload the proxies. Meanwhile, handling of lost packets and end-to-end flow control have to fulfill the overall system scalability requirements and adapt to local transmission rate and reliability (R_i, L_i) .

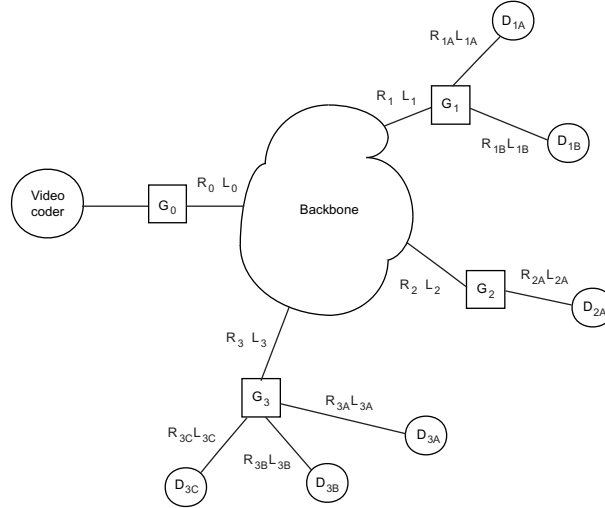


FIGURE 6.1: *Heterogeneous and error-prone video distribution environment. Each device D_i may have different resolution features and each path has its own characteristics in terms of rate R_i and loss L_i process.*

It has been shown in the previous chapters how the current video coding standards can be improved in terms of error resilience [181]. However, even with the improvements proposed in the

Parts of this work appeared in [394, 395].

recently defined MPEG-4 [191] or H.263+ [190], they remain vulnerable to losses. This vulnerability is mainly due to the hierarchical structure and to motion compensation. Moreover video standards clearly lack a feature that is becoming increasingly important in the emerging world of heterogeneous packet networks and wireless communication, which is fine scalability. Experiments with MPEG-2, MPEG-4 and H.263 scalability modes have indeed shown clear quality loss compared to a monolithic coding [3, 396]. Other layered coding schemes [319, 322, 56] are either greedy in terms of bandwidth or vulnerable to loss due to hierarchical data representations. Furthermore, they only provide a limited set of quality resolutions. Embedded coding [329, 65, 307] offers an efficient solution to the rate scalability problem but permits only octave-based spatio-temporal resolutions since it generally relies on subband coding.

Meanwhile, orthogonal transforms like DCT or wavelets have been intensively investigated in the last decade to become the basic block of most coding standards (i.e., JPEG, JPEG2000, MPEG-x, H.26x). These transforms have however shown their limitations in terms of compression especially at low rates. Wavelet decomposition, for example, is optimal for 1D signal compression. For higher order signals, wavelets are not optimal anymore [397] since they can not efficiently capture oriented contours for example. The DCT or wavelet transforms are moreover relatively complex to invert for low-complexity end-devices, although studies are carried on to optimize the decoding algorithms.

To cope with all these limitations, we propose now a completely different system based on Matching Pursuit coding. The goal is not to define yet another codec, but rather to propose a solution to the lack of current schemes in facing the problems inherent to today's streaming infrastructure. Matching Pursuit is basically a non-linear algorithm that iteratively decomposes the video stream into its most important components chosen in an overcomplete set of dictionary functions. The redundancy of the dictionary enables efficient low rate coding. In the same time, it provides a kind of intrinsic Multiple Description of the video sequence and thus improves robustness to losses. Due to the dictionary redundancy and the non-linear coding, components can indeed be represented by multiple atoms. Matching Pursuit offers also an efficient multiresolution coding method which compares favorably to Short Time Fourier or Wavelet Transforms because of their limitations relative to the uncertainty principle [398]. Depending on the dictionary construction Matching Pursuit offers a great design flexibility, generally much larger than the freedom proposed in frame-based coding [72, 20, 21]. Moreover it provides large scalability in term of both spatio-temporal resolution and decoding complexity. Finally, Matching Pursuit builds progressive and embedded streams which enable a fine rate granularity in the video content distribution. Notice that the matching of successive approximations of the source to different transmission rates depending on the quality of the channel may be connected to joint source-channel coding.

Several video coder based on Matching Pursuit coding of the displaced frame difference [86, 88, 399, 400] or hybrid wavelet(or DCT)-MP coding schemes [401, 87] have already been proposed. They showed that Matching Pursuit performances are better than the performances of MPEG-4 coders [402], especially at low rates. However, none of the proposed schemes is complete and really exploits the power of Matching Pursuit in terms of flexibility and multiresolution. The main limitations of Matching Pursuit may however be its encoding complexity which depends on the coder implementation [400]. Many new applications nowadays support an asymmetric coding model whereby the constraints at the encoder and the decoder are totally different. In digital video library, for example, the data is typically encoded once and stored, so that the encoding does not need to be realtime. The decompression however should be simple and fast. Matching Pursuit perfectly matches the requirements of asymmetric applications since its decoding complexity is very low. Furthermore, we may expect that efficient coding schemes will soon enable the use of Matching Pursuit in applications with more stringent delay.

This chapter gives an overview of Matching Pursuit, dictionary design and non-orthogonal decompositions properties. It also discusses complexity issues in MP-based systems. In the remaining of this thesis, we will not build a complete end-to-end optimal solution, but rather focus on three main parts of the system. The design of the MP dictionary for video coding is discussed in the next chapter, and a quantization scheme is proposed in the following chapter. Finally, in the last chapter, the heterogeneous and lossy MP stream delivery is studied. New transcoding

and error control schemes are proposed in taking benefit of the inherent features of MP streams.

6.2 Matching Pursuit overview

Non-orthogonal signal transforms have been recently investigated as new ways of signal coding. Decomposing a signal over an overcomplete set of functions may indeed improve the compression efficiency, especially a low bit rate where most of the signal energy is captured by only few elements. However, in contrary to orthogonal transforms, overcomplete expansions of signals are obviously not unique. The number of feasible decompositions is infinite, and finding the best solution under a given criteria is a NP-complete problem. In compression, one is interested in representing the signal with the fewest number of elements, that is to find the solution with most of the energy on only a few functions. Let define a vector b and a matrix A whose range is larger than the size of b . The problem is thus equivalent to solve $\|Ax - b\| < \epsilon$, such that x has the fewest number of non-zero values [403]. Matching Pursuit [84] is one of the sub-optimal approaches that greedily approximates the solution to this NP-complete problem in a reasonable time. Frame expansions [404, 20, 21] is another alternative to this problem. Although it presents the advantage of being linear, it allows much less flexibility and imposes a larger decoding complexity than Matching Pursuit. MP is moreover an adaptive signal expansion, while frames are non-adaptive to signal.

Matching Pursuit (MP) is an adaptive algorithm that iteratively decomposes any function f in the Hilbert space $\mathcal{H}$ in a possibly redundant dictionary of functions called *atoms* [84]. Let $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$ be such a dictionary with $\|g_\gamma\| = 1$ and Γ represents the set of possible indexes. The function f is first decomposed as follows :

$$f = \langle g_{\gamma_0} | f \rangle g_{\gamma_0} + \mathcal{R}f, \quad (6.2.1)$$

where $\langle g_{\gamma_0} | f \rangle g_{\gamma_0}$ represents the projection of f onto g_{γ_0} and $\mathcal{R}f$ is a residual component. Since all elements in $\mathcal{D}$ have by definition a unit norm, it is easy to see from Eq. (6.2.1) that g_{γ_0} is orthogonal to $\mathcal{R}f$, and this leads to

$$\|f\|^2 = |\langle g_{\gamma_0} | f \rangle|^2 + \|\mathcal{R}f\|^2. \quad (6.2.2)$$

To minimize $\|\mathcal{R}f\|$, one must choose g_{γ_0} such that the projection coefficient $|\langle g_{\gamma_0} | f \rangle|$ is maximum. The pursuit is carried out by applying iteratively the same strategy to the residual component. After N iterations, one has the following decomposition for f :

$$f = \sum_{n=0}^{N-1} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle g_{\gamma_n} + \mathcal{R}^N f, \quad (6.2.3)$$

where $\mathcal{R}^N$ is the residual of the N^{th} step with $\mathcal{R}^0 f = f$. Similarly, the energy $\|f\|^2$ is decomposed into :

$$\|f\|^2 = \sum_{n=0}^{N-1} |\langle g_{\gamma_n} | \mathcal{R}^n f \rangle|^2 + \|\mathcal{R}^N f\|^2. \quad (6.2.4)$$

It can be shown now that the Matching Pursuit (MP) scheme recovers the projection of f onto the space defined by the dictionary atoms $\{g_\gamma\}_{\gamma \in \Gamma}$. Let $\mathbf{V}$ be the closed linear span of $\mathcal{D}$. The following theorem, proved in [84], shows that MP recovers the projection of f onto $\mathbf{V}$:

Theorem 6.2.1. *The residue $\mathcal{R}f$ defined by Eq. (6.2.3) satisfies*

$$\lim_{n \rightarrow +\infty} \|\mathcal{R}^n f - \mathbf{P}_W f\| = 0, \quad (6.2.5)$$

where $\mathbf{W}$ is the orthogonal complement of $\mathbf{V}$ and $\mathbf{P}_W$ is the associated projection operator.

This also means that we can write

$$\mathbf{P}_V f = \sum_{n=0}^{\infty} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle g_{\gamma_n} \quad (6.2.6)$$

and

$$\|\mathbf{P}_V f\|^2 = \sum_{n=0}^{\infty} |\langle g_{\gamma_n} | \mathcal{R}^n f \rangle|^2. \quad (6.2.7)$$

This is an important issue. Indeed, when the dictionary is complete (i.e., $\mathbf{V} = L^2(\mathbb{R})$), f can be exactly decomposed: we have $\mathbf{P}_V f = f$, and thus $\mathbf{P}_W f = 0$. In this case, the original function can be expressed as :

$$f = \sum_{n=0}^{+\infty} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle g_{\gamma_n} \quad (6.2.8)$$

Designing complete dictionaries is thus an important issue that we address in the next section using an original geometrical technique. Although Matching Pursuit places very few restrictions on the dictionary, the latter is strongly related to convergence speed and thus to coding efficiency. In this chapter, the convergence speed has to be understood as the ability for the MP to capture most of the input signal energy in just a few iterations. Moreover, the decoder needs the dictionary to reconstruct the signal. Dictionaries built on elementary functions are thus preferred, since the encoder transmits only the parameters or indexes of these functions instead of the complete functions. Finally, any collection of arbitrarily sized and shaped functions can be used as dictionary, as long as completeness is respected.

The convergence speed of Matching Pursuit corresponds to its ability to extract the maximum signal energy in a few iterations. In other words, it corresponds to the decay rate of the residue and thus the coding efficiency of Matching Pursuit. The convergence speed depends directly on the dictionary set. The decay rate of the residual energy can thus be bounded once the dictionary is known, even without a priori information about the input signal.

The approximation error decay rate in Matching Pursuit have been shown to be bounded by an exponential [84, 405]. In other words, the decay of the residue norm is faster than an exponential decay curve whose rate depends on the dictionary only. From [72], there exists $\lambda > 0$ such that for all $N \geq 0$

$$\|\mathcal{R}^N f\| \leq 2^{-\lambda N} \|f\|, \quad (6.2.9)$$

or, equivalently,

$$\|\mathcal{R}^{m+1} f\| \leq 2^{-\lambda} \|\mathcal{R}^m f\|, \forall m. \quad (6.2.10)$$

The decay rate¹ can be written as [72]

$$2^{-\lambda} = (1 - \alpha^2 \beta^2)^{\frac{1}{2}}, \quad (6.2.11)$$

where $\alpha \in (0, 1]$ is an optimality factor. This factor depends on the algorithm that, at each iteration, searches for the best atom in the dictionary. The optimality factor α is set to one when Matching Pursuit browses the complete dictionary at each iteration. The parameter β depends on the dictionary construction. It represents the ability of the dictionary functions to capture features of any input function f . The upper-bound on the coefficient norm is reached in the worst case where the input function is the farthest from any dictionary vector. Hence, β is defined by :

$$\sup_{\gamma} |\langle f, g_{\gamma_n} \rangle| \geq \beta \|f\|. \quad (6.2.12)$$

This relationship is further developed in the next sections.

¹The decay parameter λ decreases when the size of the signal space increases. However, at the limit of infinite dimensional spaces, the convergence is not exponential any more [406].

6.3 Group theoretical design of dictionaries

6.3.1 Group representations

Clearly the dictionary has to be known at both encoder and decoders. To avoid huge memory requirements for storage, and expensive search within possibly very large dictionaries, we will use structured dictionary. Structured dictionaries obey a group law, and are mainly formed upon basic operations onto generating functions.

We collect first some definitions and properties of groups and group representations that will be used in this work. Interested readers can find further details in [407].

Definition 6.3.1. *Let G be a group and $\mathcal{H}$ be a Hilbert space. A unitary representation $\mathcal{U}$ of G in $\mathcal{H}$ is a homomorphism between G and the set of unitary operators on $\mathcal{H}$ with composition as the group law :*

$$\begin{aligned} \mathcal{U} : G &\mapsto U(\mathcal{H}) \\ \gamma &\mapsto \mathcal{U}(\gamma) \end{aligned}$$

The following properties, straightly derived from this definition, are valid for any $\gamma, \gamma_1, \gamma_2 \in G$ and used throughout the chapter :

$$\begin{aligned} \mathcal{U}(\gamma_1 \circ \gamma_2) &= \mathcal{U}(\gamma_1)\mathcal{U}(\gamma_2) \\ \mathcal{U}(\gamma^{-1}) &= \mathcal{U}^{-1}(\gamma) \\ \mathcal{U}(e) &= \mathbb{I} \end{aligned}$$

where $\circ$ denotes the group law, e is the neutral in G and $\mathbb{I}$ is the identity operator. An example used in the next chapters is the *similitude group of the plane* :

$$\text{SIM}(2) = \mathbb{R}^2 \rtimes (\mathbb{R}_*^+ \times \text{SO}(2)) .$$

This group is composed of dilations, rotations (e.g., $\text{SO}(2)$) and 2-D translations. Rotations and dilations commute but have a non-trivial action on translations. Indeed, writing a generic element $\gamma = (\vec{b}, a, \theta) \in \text{SIM}(2)$, the group law is given by :

$$(\vec{b}, a, \theta) \circ (\vec{b}', a', \theta') = (\vec{b} + ar_\theta \vec{b}', a \cdot a', \theta + \theta') , \quad (6.3.1)$$

where r_θ is a rotation matrix. Accordingly, the neutral and inverse elements read :

$$e = (\vec{0}, 1, 0) , \quad (\vec{b}, a, \theta)^{-1} = (-a^{-1}r_{-\theta} \vec{b}, a^{-1}, -\theta) .$$

Eq. (6.3.1) shows that $\mathbb{R}^2$ is endowed with a structure of G -space which enables us to compute an action σ of $\text{SIM}(2)$ on the plane :

$$\begin{aligned} \sigma[\gamma] : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ \sigma[(\vec{b}, a, \theta)]\vec{x} &= a \cdot r_\theta \vec{x} + \vec{b} . \end{aligned} \quad (6.3.2)$$

Finally a unitary representation of $\text{SIM}(2)$ in the Hilbert space $L^2(\mathbb{R}^2)$ is given by :

$$\mathcal{U}(\gamma)f(\vec{x}) = J(\sigma[\gamma])^{-1/2} f(\sigma[\gamma^{-1}]\vec{x}) ,$$

where $J(\sigma)$ is the Jacobian of the transformation (6.3.2). Now that the basics of group representation have been drawn, the remaining of the section presents the design of structured dictionaries.

6.3.2 Group theoretical design of dictionaries

Let G be a locally compact group and $\mathcal{U}$ a unitary representation of G in a given Hilbert space $\mathcal{H}$. Let $\mathcal{U}(\gamma)f$ denote the action of the unitary operator $\mathcal{U}(\gamma)$ on the function f . The unitary representation is said to be *irreducible* if it does not admit any non trivial invariant subspaces. That is if $\mathcal{S} \subset \mathcal{H}$ is such that

$$\mathcal{U}(\gamma)f \in \mathcal{S}, \quad \forall f \in \mathcal{S} \text{ and } \forall \gamma \in G$$

then either we have $\mathcal{S} = \mathcal{H}$ or $\mathcal{S} = \{0\}$. In particular, this means that the orbit of any $f \in \mathcal{H}$ under G is a dense subspace of $\mathcal{H}$:

$$\mathcal{O}_f = \{\mathcal{U}(\gamma)f \mid \forall \gamma \in G\}, \quad \overline{\mathcal{O}_f} = \mathcal{H}. \quad (6.3.3)$$

This shows that we can use unitary irreducible group representations (UIR) to design complete dictionaries of a Hilbert space. Such constructions are well-known to mathematical physicists as they are the first step in designing *coherent state* systems [408]. Such a construction induces covariance of the orbit, and thus of the dictionary, with respect to group action. Using the notation $f_\gamma = \mathcal{U}(\gamma)f$, this property is summarized by the following relation :

$$\mathcal{U}(\gamma_0)f_\gamma = \mathcal{U}(\gamma_0 \circ \gamma)f \quad (6.3.4)$$

where $\circ$ denotes the group law. The projection of a target signal f onto such a dictionary using the unitarity of $\mathcal{U}$ yields to the following equation :

$$\begin{aligned} \langle g | f_\gamma \rangle &= \langle g | \mathcal{U}(\gamma)f \rangle \\ &= \langle \mathcal{U}(\gamma^{-1})g | f \rangle. \end{aligned}$$

Hence, the projection on the dictionary is covariant² with respect to G . This also means that f_γ can be reconstructed from the coefficients of f by applying the group law to the dictionary elements selected by Matching Pursuit :

$$f_\gamma = \sum_{n=0}^{+\infty} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle g_{\gamma \circ \gamma_n}. \quad (6.3.5)$$

If we truncate this expansion, the norm of the residual remains unchanged since the representation is unitary. These properties will be particularly useful in transcoding operations, as will be shown in Chap. 9.

Example 1 (2-D wavelet dictionary). *Let us consider the similitude group of $\mathbb{R}^2$,*

$$SIM(2) = \mathbb{R}^2 \rtimes (\mathbb{R}_*^+ \times SO(2)).$$

This group is composed of translations, dilations and rotations in the plane. It can be checked [408] that the following representation is a UIR of $SIM(2)$:

$$\mathcal{U}(\vec{b}, a, \theta)g(\vec{x}) = \frac{1}{a}g(a^{-1}r_{-\theta}(\vec{x} - \vec{b})). \quad (6.3.6)$$

This representation is also the basic ingredient in the construction of the 2-D Continuous Wavelet Transform [409].

²Here covariant means that the projection coefficients are transformed according to the group law.

6.3.3 Semi-structured dictionaries

In some cases the operations we would like to apply to a given atom do not form a group. Nevertheless there are interesting cases where a subset of these operations do actually form a genuine group and some properties of the previous section still hold, though in a weaker form. Suppose our dictionary is spanned by a family of unitary operators U_α :

$$\mathcal{D} = \{U_\alpha g, \alpha \in A\}, \quad (6.3.7)$$

for a given set of indices A . Then suppose we have a unique decomposition :

$$U_\alpha = \mathcal{U}(\gamma_\alpha) \bullet R_\alpha, \quad (6.3.8)$$

where $\bullet$ denote the composition of operators, $\mathcal{U}$ is a unitary representation of a given group G and R_α is a unitary operator belonging to $U(\mathcal{H})$. The dictionary $\mathcal{D}$ is then invariant under G and MP coefficients are covariant under the action of G onto $\mathcal{H}$. We call such a dictionary *semi-structured*.

Example 2 (Anisotropic Refinement). *Let the index set A contain translations, rotations and two dilations, one for each principal direction :*

$$A = \left\{ \vec{b} \in \mathbb{R}^2, \theta \in SO(2), (a_1, a_2) \in \mathbb{R}_*^+ \times \mathbb{R}_*^+ \right\}. \quad (6.3.9)$$

The action of A on an atom g is defined by :

$$U_\alpha g = \mathcal{U}(\vec{b}, \theta) D(a_1, a_2) g, \quad (6.3.10)$$

where $\mathcal{U}$ is a representation of the Euclidean group :

$$\mathcal{U}(\vec{b}, \theta) g(\vec{x}) = g(r_{-\theta}(\vec{x} - \vec{b})), \quad (6.3.11)$$

and D acts as a dilation operator :

$$D(a_1, a_2) g(\vec{x}) = \frac{1}{\sqrt{a_1 a_2}} g\left(\frac{x_1}{a_1}, \frac{x_2}{a_2}\right), \quad (6.3.12)$$

where $\vec{x} = (x_1, x_2)$. When $a_1 = a_2$ one obviously gets back to the similitude group of the plane and, avoiding rotations, we end up with a group studied by Bernier and Taylor [410]. In general the set A can not be endowed with a group structure, but the factorization property of Eq. (6.3.10) guarantees covariance with respect to the Euclidean group. This dictionary offers a simple way to implement an anisotropic refinement process on atoms. This allows to efficiently capture the geometry of edges (and other elongated structures) in images.

6.4 Structural redundancy in non-orthogonal transforms

Compression efficiency is mainly driven by redundancy of the overcomplete set of functions chosen for the signal decomposition. In Matching Pursuit algorithms [84] the redundancy of the dictionary influences the convergence of the residual energy [72, 85, 405], as shown by Eq. (6.2.11). The set of functions or dictionary plays a crucial role in the non-orthogonal transform properties, and more particularly in the ability of this transform to compact the signal energy. Structural redundancy provides an important criteria in the design of dictionaries and quantifies the power of the transform to capture signal features. The size of the dictionary provides a first indicator of the dictionary properties [411, 412], but it does not take into account the distribution of the atoms. We propose here a formulation for the structural redundancy of an overcomplete set of functions, which can be applied in the context of Matching Pursuit or frame expansions.

Clearly, the size of the dictionary³ is not sufficient to represent the ability to catch the signal features. Indeed, in the case of non-uniform atom distribution for example, the number of atoms

³The term dictionary is used in this chapter to represent the overcomplete set of functions used in the signal transform.

does not give any information about the ability of the dictionary to represent any function f in the Hilbert space $\mathcal{H}$. The concept of redundancy is obviously useless in the context of orthogonal transforms, by definition of the basis functions. However, the redundancy factor directly drives the sparsity of non-orthogonal transforms in frame expansion [404] or Matching Pursuit [84].

Signal transforms are based on inner products to compute the contribution of each basis function or atoms into the signal reconstruction. The redundancy β can therefore be interpreted as the cosine of the maximum possible angle between any direction in $\mathcal{H}$ and the closest direction of any atom of the dictionary [84]. It characterizes the redundancy of the dictionary and tends to one when the size of the complete dictionary increases.

6.4.1 New structural redundancy formulation

Assume first that atoms or functions of the dictionary have a size L , equivalent to the size of the input signal f in $l^2(\mathbb{Z}_L)$. Let moreover $\{g_{\gamma_i}, i \in [1..S]\}$ define the set of functions in a dictionary of size $S \geq L$.

For each dictionary vector g_{γ_i} , one can define its projection neighborhood as the subspace of $\mathcal{H}$ which admits g_{γ_i} as closest direction.

Definition 6.4.1. *The projection neighborhood $\mathcal{V}_{\gamma_i}$ of the vector g_{γ_i} is the subspace of $\mathcal{H}$ defined by*

$$\mathcal{V}_{\gamma_i} = \{x \in \mathcal{H} \mid |\langle x | g_{\gamma_i} \rangle| \geq |\langle x | g_{\gamma_j} \rangle|, \forall j \neq i\} . \quad (6.4.1)$$

The projection neighborhood $\mathcal{V}_{\gamma_i}$, as represented in Fig. 6.2 and 6.3, corresponds to the intersection of couples of infinite convex polyhedral cones situated symmetrically with their apexes at the origin⁴. The polyhedral cones are specified by a pair of half-space constraints for each element of the dictionary. The half-space constraints can be expressed as

$$\langle x | g_{\gamma_i} + g_{\gamma_j} \rangle \langle x | g_{\gamma_i} - g_{\gamma_j} \rangle > 0, \forall j \neq i . \quad (6.4.2)$$

The parameter β corresponds thus to the cosine of the maximum possible angle, over all dictionary

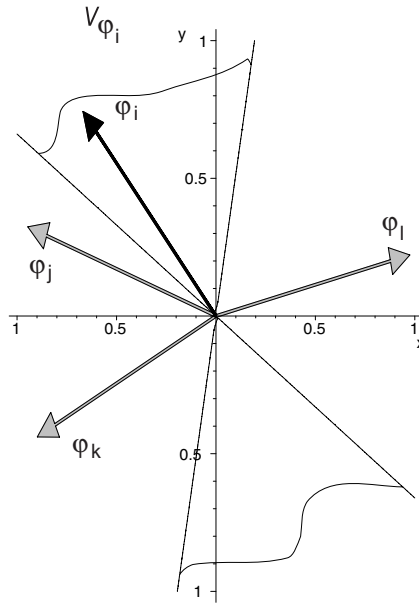


FIGURE 6.2: Representation of the projection neighborhood $\mathcal{V}_{\gamma_i}$ in $\mathbb{R}^2$.

⁴A similar geometrical representation is proposed in [404] for consistent reconstruction.

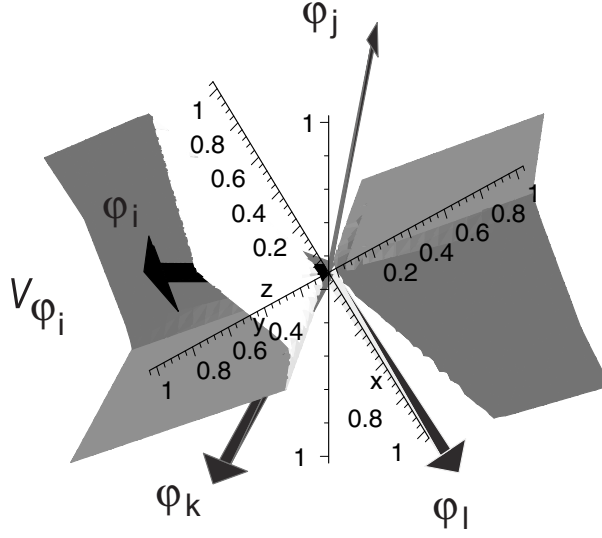


FIGURE 6.3: Representation of the projection neighborhood $\mathcal{V}_{\gamma_i}$ in $\mathbb{R}^3$.

vectors, between g_{γ_i} and any direction in its projection neighborhood $\mathcal{V}_{\gamma_i}$. It can be written as

$$\beta = \min_i \inf_{x \in \mathcal{V}_{\gamma_i}} \langle x | g_{\gamma_i} \rangle. \quad (6.4.3)$$

The direction which has the maximum angle with g_{γ_i} is the intersection of two $(L-1)$ -dimensional hyperplanes specified by the half-space constraints of Eq. (6.4.2). Instead of browsing the complete $\mathcal{V}_{\gamma_i}$ region, one can look for such an intersection with the maximum angle to g_{γ_i} , but within $\mathcal{V}_{\gamma_i}$.

The procedure can be formulated as follows. First pick a vector g_{γ_j} different from g_{γ_i} . Let $g_{\gamma_{i+j}}$ and $g_{\gamma_{i-j}}$ respectively represent the normalized sum and difference of g_{γ_i} and g_{γ_j} :

$$g_{\gamma_{i+j}} = \frac{g_{\gamma_i} + g_{\gamma_j}}{\|g_{\gamma_i} + g_{\gamma_j}\|} \quad (6.4.4)$$

and

$$g_{\gamma_{i-j}} = \frac{g_{\gamma_i} - g_{\gamma_j}}{\|g_{\gamma_i} - g_{\gamma_j}\|} \quad (6.4.5)$$

It has to be noted that $g_{\gamma_{i+j}}$ and $g_{\gamma_{i-j}}$ are orthogonal, i.e., $\langle g_{\gamma_{i+j}} | g_{\gamma_{i-j}} \rangle = 0$. We call $\mathcal{H}_{i+j}$ the hyperplane of dimension $(L-1)$ defined by $g_{\gamma_{i+j}}$ and orthogonal to $g_{\gamma_{i-j}}$. Similarly let $\mathcal{H}_{i-j}$ be the hyperplane defined by $g_{\gamma_{i-j}}$ and orthogonal to $g_{\gamma_{i+j}}$. Both hyper-planes define frontiers of $\mathcal{V}_{\gamma_i}$ since they satisfy

$$|\langle x | g_{\gamma_i} \rangle| = |\langle x | g_{\gamma_j} \rangle|, \quad (6.4.6)$$

or equivalently

$$\begin{cases} \langle x | g_{\gamma_i} - g_{\gamma_j} \rangle = 0 & \text{if } \langle x | g_{\gamma_i} \rangle \langle x | g_{\gamma_j} \rangle \geq 0 \\ \langle x | g_{\gamma_i} + g_{\gamma_j} \rangle = 0 & \text{otherwise.} \end{cases} \quad (6.4.7)$$

Now define $\mathcal{I}_\alpha$ as the direction defined by the intersection $\cap_{l \in \alpha} \mathcal{H}_{i \pm l}$, where α is the $(L-1)$ -tuple $\{k_1 \dots k_{L-1}\}$, with $k_l \neq i$. Let $\overline{\mathcal{I}}_\alpha$ be the orthogonal complement to $\mathcal{I}_\alpha$. Hence, the $(L-1)$ -dimensional hyperplane $\overline{\mathcal{I}}_\alpha$ can be written as

$$\overline{\mathcal{I}}_\alpha = \text{span} (g_{\gamma_{i+k_1}} \dots g_{\gamma_{i+k_{L-1}}}). \quad (6.4.8)$$

The orthogonal projection P_α^i of g_{γ_i} onto $\mathcal{I}_\alpha$ is equivalent to the residue of its orthogonal projection onto $\overline{\mathcal{I}_\alpha}$. Therefore,

$$P_\alpha^i = g_{\gamma_i} - \sum_{l=1}^{L-1} \langle g_{\gamma_i} | \widehat{g_{\gamma_{i+k_l}}} \rangle \widehat{g_{\gamma_{i+k_l}}} , \quad (6.4.9)$$

where $g_{\gamma_{i+k_l}}$ have been orthogonalized using the Gram-Schmidt procedure. The cosine of the angle between $\mathcal{I}_\alpha$ and g_{γ_i} is given by

$$\beta_\alpha^i = \frac{|\langle g_{\gamma_i} | P_\alpha^i \rangle|}{\|P_\alpha^i\|} . \quad (6.4.10)$$

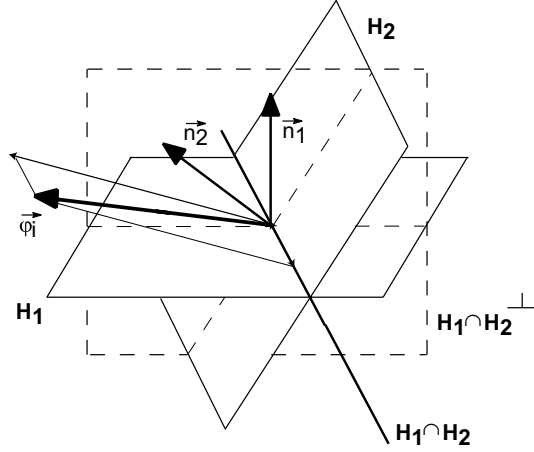


FIGURE 6.4: The orthogonal projection of φ_i onto the intersection $H_1 \cap H_2$ is equivalent to the residue of the orthogonal projection of φ_i onto the orthogonal complement to the intersection $H_1 \cap H_2$. This representation in $\mathbb{R}^3$ generalizes to n -dimensional case.

Finally, the structural redundancy factor β is given by searching, for each vector g_{γ_i} of the dictionary, the intersection $\mathcal{I}_\alpha$ belonging to the projection neighborhood $\mathcal{V}_{\gamma_i}$ and forming the maximum angle with g_{γ_i} . In other words,

$$\beta = \min_i \min_{\mathcal{I}_\alpha \in \mathcal{V}_{\gamma_i}} \beta_\alpha^i . \quad (6.4.11)$$

6.4.2 Experimental results

Fig. 6.5 represents the evolution of the redundancy factor β as a function of the size of the dictionary. The dictionary is formed of random atoms of dimension 5, 15 and 25 respectively. We can see that the redundancy factor obviously increases with the size of the dictionary. Finally, we can conjecture that the evolution of β is exponential with the number of vectors N_v . In other words,

$$\beta = 1 - A N_v^B . \quad (6.4.12)$$

Fig. 6.6 depicts the redundancy factor as a function of the ratio of the number of dictionary vectors N_{vec} and the size N of the vectors. It shows that the slope of the exponential fitting curves are quite similar. Therefore, for large signals, the number of vectors needed to reach a high redundancy factor explodes and becomes much bigger than the simple ratio $\frac{N_{vec}}{N}$. The redundancy definition generally used in the context of frame expansions provides therefore only a coarse qualitative approximation of the redundancy of MP dictionaries.

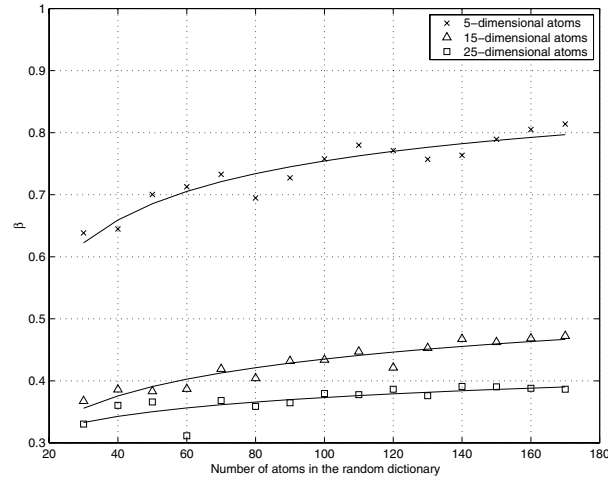


FIGURE 6.5: Redundancy factor β as a function of the number of vectors N_v in a random dictionary. Exponential fitting function of the form: $\beta = 1 - A N_v^B$.

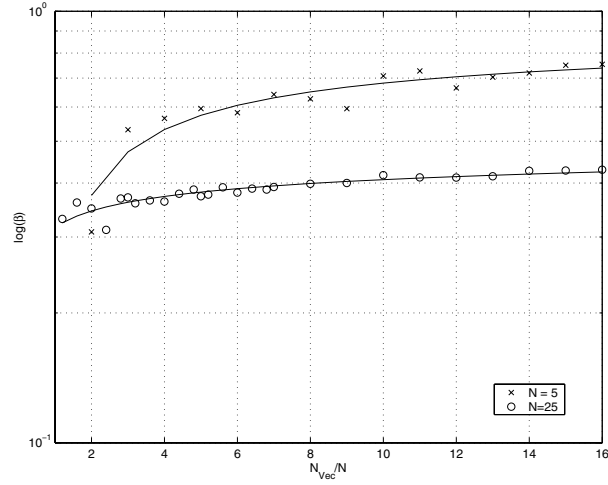


FIGURE 6.6: Redundancy factor β as a function of the ratio between the number of vectors N_v in a random dictionary and the size N of the vectors. Exponential fitting function of the form: $\beta = 1 - A N_v^B$.

6.4.3 Uniformly distributed atoms

The exact computation of β proposed in the previous section becomes heavy for a very large dictionary or a high dimensional signal, like, for example, an image. Therefore, in the case of uniformly distributed atoms on the L -dimensional unit sphere [413, 414] one can roughly approximate β as

$$\tilde{\beta} = \min_j \max_{i \neq j} |\langle g_i, g_j \rangle|. \quad (6.4.13)$$

The assumption of uniformly distributed atoms is also valid for very redundant dictionaries. The so-computed redundancy factor is however larger than the exact bound given by Eq. 6.4.11. The upper-bound onto the coefficient norm may thus be under-estimated in this case. However the difference between both β values decreases rapidly with the size of the dictionary. For very redundant dictionary, this difference becomes small enough for $\tilde{\beta}$ to be used in practical cases. Moreover the upper-bound onto the coefficient decay is generally very loose since it is computed

in the worst case where the input signal is at each step the farthest possible from any dictionary vector.

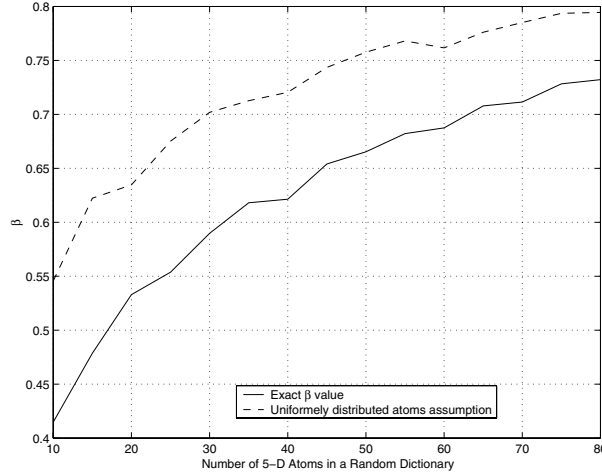


FIGURE 6.7: Redundancy factor β and its approximation through the uniform vectors distribution as a function of the size of a random 5-dimensional vectors dictionary. The values are averaged on 100 different random dictionaries.

Fig. 6.7 represents the evolution of β and its approximation of a function of the size of the random dictionary. The curves have been averaged on 100 different dictionaries. They clearly show that the approximated β values are always above the exact value, and that the difference between both values decreases with the size of the dictionary, as expected.

Finally, the β parameter characterizes the redundancy of the dictionary for uniformly distributed atoms, independently from the function f to approximate. A value of β close to one implies a highly redundant dictionary, while a zero value indicates that the dictionary is not complete. In the same time, β characterizes the importance of the atom index and thus the coefficient in carrying information. Indeed, the importance of the atom index increases with the redundancy of the dictionary (i.e., the β value), while the relative importance of the coefficient decreases. In contrast, most of the information is carried on by the coefficients in orthogonal transforms.

6.4.4 Structural redundancy in Matching Pursuit

In Matching Pursuit, the structural redundancy factor β directly drives the decay rate of the residue and thus the coding efficiency of the algorithm, as shown in Eq. (6.2.11). Fig. 6.8 shows that the residual energy is clearly upper-bounded by the exponential curve computed from the redundancy factor. Moreover, this upper-bound does not depend on the input function f , but, as such, may become one of its major weaknesses. It does not permit to estimate the true convergence of the Matching Pursuit, since the latter does not only depend on the dictionary, but also on the input signal. In other words, the convergence will indeed be faster for an input signal whose main characteristics are represented by single dictionary vectors. However one can easily include probability-weighting in the computation of β , provided that the statistics about the input signal are known *a priori*.

It has to be noted that very similar computation can be applied to the computation of the bounds of frame expansions. Frame bounds have been computed in the particular case of 2D Gabor wavelets [415]. However, the hereabove method is very generic and may be applied to any frame construction. The efficiency of the frame design can also be evaluated by similar development.

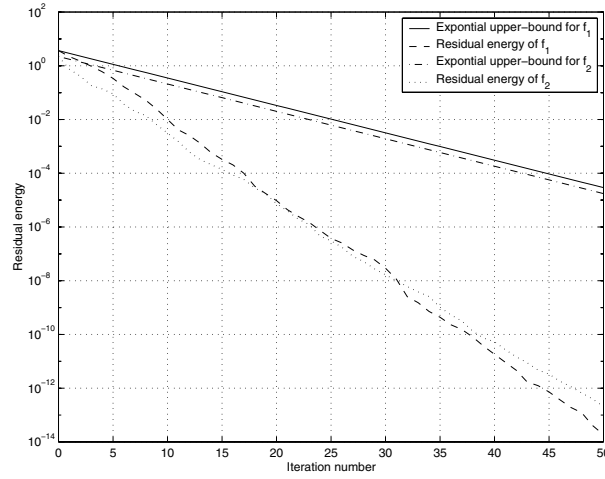


FIGURE 6.8: Residual energy decay for Matching Pursuit coding of two random signals (f_1, f_2) of length 10 with a dictionary formed of 50 random vectors.

6.5 Matching Pursuit complexity

6.5.1 Encoding complexity

The main drawback in a Matching Pursuit system is the encoding complexity. Indeed, for each iteration, the encoder has to browse the complete dictionary to find the best matching functions. The complexity of the search therefore grows with both the signal and the dictionary size.

Several methods may be proposed to reduce the encoding complexity. First, the dictionary can be designed upon functions enabling efficient inner product computations [416, 417] by filter-based implementation. Dictionaries may also be built with functions corresponding to combination of elementary functions, easily implemented [418]. Second, the encoder can be modified to first localize highest energy regions, and then apply Matching Pursuit on small areas within the so-defined regions [419, 88]. It can also use statistic information about the previously elected dictionary functions to focus the research around similar atoms [86]. Finally, Matching Pursuit may be orthogonalized to decrease the complexity, hence sacrificing the coding efficiency at low rates. Parallel implementation have been proposed to decrease the computation time [420].

In this work, we use a sub-optimal search scheme based on Genetic Algorithm. This sub-optimal search is general enough to be applied to any structured dictionary while providing an efficient solution in terms of encoding complexity. For each Matching Pursuit iteration, a set of chromosomes is first defined. The chromosomes are composed of genes, one gene for each atom index. For example, a gene may represent the horizontal position of the atom, another the scale parameter. At initialization, genes are randomly chosen within the range of allowed values. The algorithm then searches for the chromosome corresponding to the best matching function, or equivalently the function that best matches the signal characteristics.

Mutations and cross-over are then operated on the chromosomes. The best chromosomes is retained and several mutations of its genes define new chromosomes. Cross-over between chromosomes as represented in Fig. 6.9 serves to generate a new population of chromosomes. Respective genes of successive chromosomes are paired together and give birth to a child whose genes are issued from the genes of its two parents. This process is iterated and rapidly converges to a near-optimal solution to the Matching Pursuit search. Moreover, to avoid the process to be trapped into a local minimum, the complete set of chromosomes, besides the best one, is reinitialized when the mutations and cross-over become confined to a very small set of genes combinations. Details of the Genetic Algorithm may be found in [421].

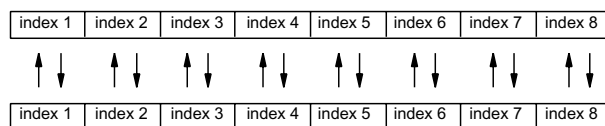


FIGURE 6.9: *Chromosomes cross-over in Genetic Algorithm.*

6.5.2 Decoding complexity

Matching Pursuit system has the advantage of a very low decoding complexity. The decoder does not need to perform expensive inverse transforms as in wavelets or DCT-based codecs. Matching Pursuit decoder only need to build dictionary functions from the indexes sent by the encoder. It then simply superposes MP iterations to generate the signal reconstruction.

It has been shown that Matching Pursuit decoding of the displaced frame difference is less complex than the video decoding process of MPEG-4 [402]. However the decoding complexity is highly dependent on the dictionary functions, although it remains small even for complex functions. Finally, the decoding complexity depends also on the number of MP iterations used in the reconstruction. Since MP naturally generates a progressive stream, it offers a great flexibility in the system design. Scalability in the decoding complexity may be reached by adapting either the dictionary functions or the number of iterations. Hence Matching Pursuit streams can easily be decoded on high-end workstations as well as on low power PDAs, with different quality however.

6.6 Conclusion

This chapter has presented the basics of a new video coding system based on Matching Pursuit. This new system is issued from the lack of the classical coding schemes in fulfilling both scalable and error-resilient constraints imposed by today's network infrastructure. Matching Pursuit has been presented as an effective solution to the multiresolution and robust video delivery. It decomposes the video signal over a overcomplete set of dictionary functions, whose redundancy directly drives the compression efficiency. Finally, even though the MP encoder can have a large complexity, the decoding process is very simple and can moreover be adapted to the receiver possibilities. This makes MP very attractive for asymmetric applications. In the next chapters, we will see how the Matching Pursuit can be more particularly applied to video coding.

3D Matching Pursuit Video Coding

7.1 3D video coding

Most of the video coding schemes (e.g., MPEG and H.26x families) are based on motion estimation [423, 45]. Motion compensation proved to efficiently reduce the temporal redundancy between frames. However, it introduces a hierarchy between frames and hence between transmission packets. Transmission losses thus propagate within the reconstructed sequence. The packets containing the reference frames are indeed mandatory to decode predicted frames. The video degradation becomes therefore more important than the loss ratio on the network since some information can be useless even though it has been correctly received (see Chap. 4). Furthermore, the temporal multiresolution properties are very limited in motion-compensated sequences, although some hierarchical or multiresolution motion estimation schemes have been proposed to alleviate this constraint [41, 304]. The reference frames are obviously mandatory for the decoding, so that the number of achievable temporal resolutions is truly limited in the absence of complex transcoding methods.

3D transforms offer an interesting alternative to motion compensation. In this case, the temporal redundancy reduction is performed in a similar way to the spatial redundancy reduction, that is through an efficient transform based either on wavelets [56, 316, 69, 62, 52, 51] or DCT [49]. The source coding efficiency of a temporal transform is equivalent and even better than the efficiency of motion compensation [48, 61]. Algorithms based on 3D transforms have moreover proven to provide better performances than coding standards like MPEG-2 [69, 62] or H.261 [59] and a comparable efficiency to H.263 [60]. The type of artifacts are however different in both approaches. While the motion-compensated and block-based DCT approach tends to produce blocky types of artifacts at low rates, 3D transforms coding introduces generally distortions in the form of lost high-frequency details or motion blur. This limitation is mainly due to the buffering constraints imposed by the allowable coding delay. It can be improved by local motion estimation before temporal transform, thus resulting in a hybrid 3D-motion compensated coding scheme [65, 66, 67, 68]. However, 3D transforms often increases decoding complexity compared to schemes based on motion compensation since the inverse transform is often more demanding than motion compensation. In the same time, the encoding complexity is however reduced [48] in comparison to the very expensive motion estimation. This property is interesting in the context of real-time applications. The complexity seems however not to be the principal bottleneck regarding to the evolution of today's processors capabilities.

Not only the coding efficiency of spatio-temporal transform is at least comparable to motion compensated scheme, but the error resilience in three-dimensional coding is also improved. There is indeed no temporal error propagation, since coefficient can be decoded independently of any

Parts of this work appeared in [422].

reference frame. The loss patterns on the reconstructed sequence are different from the degradation observed in motion compensation based schemes. They obviously may still last for several frames depending on the size of the frame buffer in 3D transform. The losses are however reduced since data are independently decodable and completely predictable from the networking conditions. Hence, the temporal error propagation phenomenon is avoided in 3D coding.

Taking advantage of the properties of three-dimensional decompositions, we will now focus on non-linear 3D transforms based on Matching Pursuit, as introduced in Chap. 6. As described before, MP presents several interesting advantages over usual orthogonal transforms like an improved robustness to errors, and a reduced decoding complexity. It has to be noted that common video Matching Pursuit coders [86, 88, 417, 399, 400] exclusively work on the residual error frame. They perform traditional motion compensation and then code the displaced difference frame through Matching Pursuit. Matching Pursuit has shown to provide better results than classical DCT-based coding, especially at low rates [402]. To take benefit of 3D transform, we propose here a novel 3D Matching Pursuit video coder, which transforms entire blocks of frames into sets of spatio-temporal functions from a 3D dictionary. The decoder in this case simply reconstructs the video sequence by superposing spatio-temporal functions. The decoding complexity is thus even simpler than traditional DCT/Motion compensated and the common MP/Motion compensated video decoders [402]. The encoding complexity is however larger since the dictionary becomes wider, thus increasing the cost of the MP search algorithm. The 3D Matching Pursuit scheme is therefore mainly destined to highly asymmetric applications.

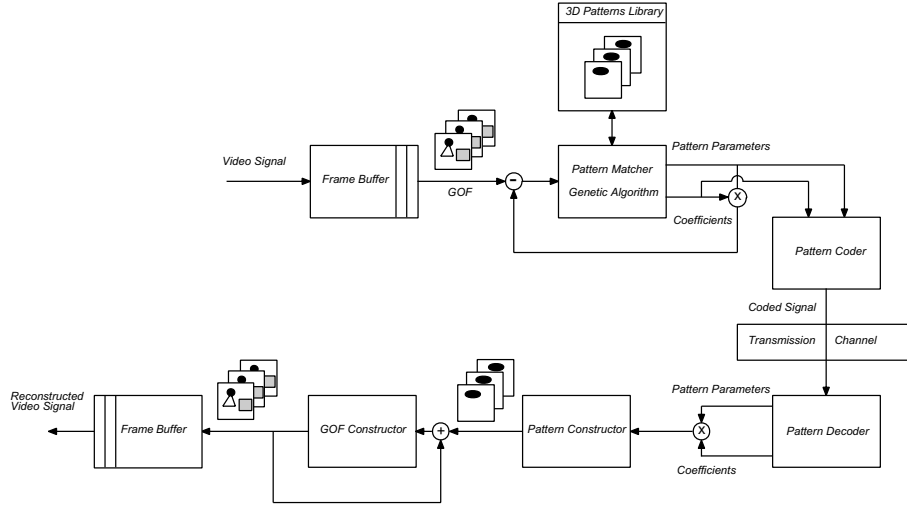


FIGURE 7.1: *Tri-dimensional Matching Pursuit scheme*

3D Matching Pursuit may be represented as in Fig. 7.1. The video sequence is first divided into group of frames (GOF). The GOF is compared by means of a Genetic Algorithm to the spatio-temporal patterns in the 3D dictionary. The best matching pattern is removed from the GOF, and the same operation is repeated on the residual GOF until a predefined number of iterations is reached. The 3D pattern parameters are then encoded and packetized along with their respective correlation coefficient. At decoder, the GOF is simply reconstructed by superposition of 3D functions multiplied by their respective coefficients.

This chapter focuses on the transform block of the encoder (see Fig. 2.2). It has been shown that Matching Pursuit may be as efficient as classical DCT-based coding [402] with Gabor dictionaries. We first present new dictionaries which offer improved performances by comparison to Gabor dictionaries. A special attention is put on the scalability requirements with respect to decoding complexity. 3D Matching Pursuit video coding is finally analyzed in terms of both decoding complexity and compression efficiency.

7.2 3D dictionaries

7.2.1 Structured dictionaries

The main problem in Matching Pursuit applications is the size of the dictionary. To reach a good compression efficiency, the dictionary has to be highly redundant, as shown in Sec. 6.4, and thus generally of large size. In the same time, dictionary atoms have to be known at the decoder, but transmission and storage of large sized dictionary and atoms consume a lot of resources. The problem of the dictionary transmission can be avoided by designing structured dictionaries. A structured dictionary is built by applying meaningful transformations (e.g., rotation and translation and dilatation) to a generating function, as seen in Chap. 6. In this case, the decoder has only to know the generating function and the parameters of the transformations (i.e., indexes) to reconstruct the coded atoms.

In many cases the transformation of the generating function will combine into a group G (see Sec. 6.3), the similitude group $SIM(2) \times SIM(1)$ in our case. The dictionary is then constructed by applying transformations to basic functions, so that it forms a complete and preferably redundant set of functions spanning the Hilbert space $\mathcal{H}$ of the input signal. As we will see in the next chapter, the index quantization has preferably to be avoided. The key-point in the dictionary design becomes to find the optimal number of possible indexes (i.e., translations, dilations and rotations of the generating function), so that the dictionary is sufficiently redundant to allow a good compression, while staying reasonably large to optimize the transmission cost. Note that the complete optimization of the dictionaries is however not targeted in this chapter, but rather the design principles leading to efficient MP dictionaries.

7.2.2 Separation of spatial and temporal functions

Spatial and temporal features in natural video sequences are completely different. We propose therefore to build 3D atoms by separable spatial and temporal functions. By separating spatial and temporal dimensions, significant reduction in computation may be reached with little penalty, especially in view of the nature of video signals. These are indeed formed of image sequences smoothly moving over time. Note that such a separation is often implemented in 3D transform coding, where the decomposition is generally performed along three 1D wavelets, where the temporal wavelet is different than the spatial ones.

The 3D atoms are formed by multiplying 2D spatial atoms by a temporal function, possibly different from the spatial generating functions. The spatial functions are designed to capture specific spatial features of natural video images. The temporal function, generally simpler, multiplies the spatial function to form 3D atoms. Intuitively, the temporal function has to manage the visibility of the spatial atoms within the images of the Group of Frames under consideration. Basically, the atom at pixel of coordinates $(\tilde{x}, \tilde{y}, \tilde{z})$ may be expressed as :

$$g_{\gamma_i}(\tilde{x}, \tilde{y}, \tilde{z}) = s_{\gamma_i}(x, y) t_{\gamma_i}(z) , \quad (7.2.1)$$

where s_{γ_i} represents spatial two-dimensional functions and t_{γ_i} represents the temporal function. The structured dictionaries are generated by translation, dilations and rotations of basic spatio-temporal functions. The atoms are thus generally represented by seven indexes, namely the spatio-temporal position, three indexes of scale and a rotation parameters for the spatial atoms.

7.2.3 Complexity considerations

Decoding complexity is a very important criteria in the design of 3D atoms. Not only the compression efficiency has to be accounted, but also the possibility to decode the atoms on heterogeneous receivers. The atoms have to capture most of the signal energy in only few iterations, but also offer simplicity of decoding for the lowest end devices. We will present in the next sections atoms of different computing complexity, trading off compression efficiency against decoding complexity. It has to be noted that the decoding process may also be adapted to the decoder possibilities by

adapting the number of iterations in the reconstruction. Finally, one could imagine a transcoding process where complex atoms are approximated by more simple atoms to be decodable onto low complexity receivers. However, this point is not treated here. The transcoding process which will be discussed in Chap. 9 only deals with resolution adaptation, and not atom approximation. Another interesting possibility would be to code the first iterations with simple atoms and then refine the representation with more complex atoms. The low power decoder would in this case only decode the first iterations according to their possibilities. This strategy will not be implemented either in this work but may provide a very flexible bit stream in an heterogeneous environment as the one we are targeting.

7.3 Spatial dictionaries

Any function may a priori serve as generating function for 2D atoms, like for example wavelets or cosine functions. The generating functions have to meet a trade-off between compression efficiency and decoding complexity. This section presents four types of 2D atoms of different complexity and compression efficiency. These dictionaries are mainly issued from heuristics considerations however driven by two main design principles: an optimal spatial and frequency localization and the efficiency to code oriented contours. The spatial atoms are thus characterized by two indexes of positions, p_x and p_y , an orientation index ϕ and possibly two scales indexes s_x and s_y .

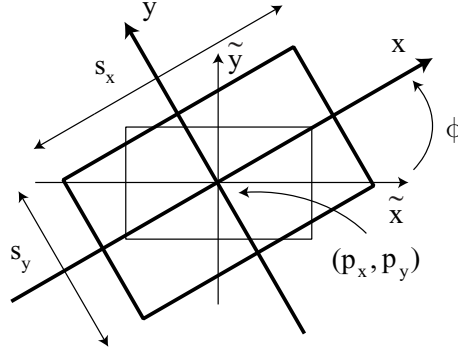


FIGURE 7.2: *Coordinates changes by translation, scaling and rotation.*

A coordinate change (See Fig. 7.2) is first operated before applying the generating function. If the coordinates of pixels in images are given by $(\tilde{x}, \tilde{y})$, the coordinates (x, y) after translation, scaling and rotation are given by

$$x = \frac{\cos(\phi) (\tilde{x} - p_x) + \sin(\phi) (\tilde{y} - p_y)}{s_x}, \quad (7.3.1)$$

and

$$y = \frac{\cos(\phi) (\tilde{y} - p_y) - \sin(\phi) (\tilde{x} - p_x)}{s_y}. \quad (7.3.2)$$

Hence, the coordinate change is equivalent to dilate, translate or rotate the generating functions. It permits to build complete dictionaries upon transformation of basic functions which are chosen according to efficiency and complexity criteria.

7.3.1 Gabor atoms

Image representation through Matching Pursuit almost exclusively uses 2D Gabor atoms [88, 424]. A spatial and frequency invariant dictionary is obtained by scaling, translating and modulating

a gaussian window [84, 425]. Gaussian windows are chosen because of their optimal spatial (or time) and frequency energy concentration, proven by the uncertainty theorem [426, 427]. It has to be noted that Gabor functions were initially obtained by modulations of a fixed size gaussian function. They have been modified in the wavelet context by fixing the modulation of the gaussian function, while varying its size. These functions form 2D Gabor atoms [415, 428], whose real form is given by :

$$s_{\gamma_i}(x, y) = \frac{1}{2 \pi \sigma^2} \cos(\omega_0 x) e^{-\frac{x^2 + y^2}{4 \sigma^2}}, \quad (7.3.3)$$

where ω_0 , generally even, is the oscillation parameter and σ is a scale parameter. This form of Gabor atoms have a isotropic scaling which is identical in both directions (i.e., $s_x = s_y$ in Eqs (7.3.1) and (7.3.2)). This means that a scale index may be spared but also that Gabor atoms are not well suited to approximate edges.



FIGURE 7.3: *Reconstructed version of Lena encoded with 500 Gabor atoms where $(\omega_0, \sigma) = (6, 2)$.*

Fig. 7.3 represents the *Lena* image reconstructed with 500 MP iterations of 2D Gabor atoms. We can see that Gabor atoms are indeed not optimal to approximate edges, like for example the shoulders line. Indeed, several small atoms are necessary to code it completely because of the isotropic scaling of Gabor atoms.

7.3.2 Anisotropically refined gaussians

The main limitation of Gabor atoms is their isotropic scaling. This feature does not allow to efficiently catch edges and contours. We therefore propose to use another dictionary able to efficiently code high frequency patterns. The new dictionary is built with anisotropically refined gaussians [422]. We keep gaussian functions due to their optimal property of space-frequency localization. Furthermore, to efficiently catch edges, one direction of the two-dimensional atoms correspond to a gaussian function while the orthogonal direction represents its second derivative. Also, to offer more flexibility to the coding process, the scaling parameters can be different in both directions. The coding cost of an additional index is largely compensated by the improved compression efficiency. Anisotropically refined gaussian atoms can be written as

$$s_{\gamma_i}(x, y) = (4 x^2 - 2) \exp(-(x^2 + y^2)), \quad (7.3.4)$$

where x and y results from the coordinate change defined in Eqs. (7.3.1) and (7.3.2).

These atoms have similar properties to wavelets in terms of compression efficiency. However, in contrary to subband decompositions, they allow to efficiently catch contours which are not oriented along horizontal or vertical axis. Finally, it can be shown that they are more efficient than wavelets in catching two-dimensional contours where wavelets efficiency is not optimal [397].



FIGURE 7.4: *Reconstructed version of Lena encoded with 500 anisotropically refined atoms.*

Fig. 7.4 represents the reconstruction of the *Lena* image with 500 anisotropically refined gaussian atoms. It can be seen clearly that the quality is much better than in the case of the Gabor dictionary presented above. This improvement is mainly due to the anisotropic property of the new atoms and their gradient-like construction which allows to efficiently catch edges of any direction.

7.3.3 B-spline functions

The anisotropically refined gaussian atoms are optimum considering spatial and frequency localization. The main problem with gaussian functions is however their infinite support. Indeed, the atoms take a non-zero value for all pixels of the image. It is not possible to a priori limit the support since the pixels that do not significantly contribute to the correlation product are not known.

We propose therefore to replace gaussian functions with B-splines which are known to have a limited support [429, 430] and therefore a reduced computational complexity. A B-spline of order n is given by the following equation :

$$\beta^n(x) = \frac{1}{n!} \sum_{k=0}^{n+1} \binom{n+1}{k} (-1)^k \left[\tilde{x} - k + \frac{n+1}{2} \right]_+^n, \quad (7.3.5)$$

where $[x]_+^n$ represents the positive part of x^n . It can be shown that the support region of the B-spline of n^{th} order is equivalent to $-\frac{n+1}{2} \leq x < \frac{n+1}{2}$. For x values outside of this interval, the B-spline takes null value since the terms of Eq. (7.3.5) sum up to zero.

B-splines are sub-optimal by comparison to gaussians which minimizes the uncertainty principle, even if their localization performance are very close [430]. The splines have moreover the advantage to be widely used in computer graphics, hence offering very fast implementations. Compression efficiency is thus sacrificed to lower the decoding complexity.

The same design method as for the gaussian dictionary is used for the B-splines dictionary, which are anisotropic refinement and gradient-like design. The order of the B-spline has to be sufficiently large to offer good compression efficiency, but not too large for computational reasons. We propose to use 3^{rd} order B-splines due to their good localization properties in one direction, and their first derivative in the orthogonal direction. Even if larger derivative orders are expected to improve compression efficiency, only the first derivative is implemented for decoding complexity reasons. The first derivative of a n^{th} order B-spline may be computed by the difference of two B-splines of order $(n-1)$, as :

$$\frac{\partial \beta^n}{\partial x} = \beta^{n-1} \left(x + \frac{1}{2} \right) - \beta^{n-1} \left(x - \frac{1}{2} \right). \quad (7.3.6)$$

Rotation, translation and anisotropic dilatation of the B-spline generates an overcomplete dictionary. The anisotropic refinement permits to use different scaling in orthogonal dimensions, hence efficiently capturing edges for example. The 2D B-spline atoms can finally be written as :

$$s_{\gamma_i}(x, y) = \beta^3(x) \left(\beta^2 \left(y + \frac{1}{2} \right) - \beta^2 \left(y - \frac{1}{2} \right) \right) , \quad (7.3.7)$$

where x and y are obtained from Eq. (7.3.1) and (7.3.2).



FIGURE 7.5: *Reconstructed version of Lena encoded with 500 B-spline atoms.*

Fig. 7.5 shows the reconstruction of the *Lena* image with 500 iterations of MP with B-spline atoms. It can be seen clearly that the quality is much better than in the case of the Gabor dictionary presented above, for similar reasons as in the case of the gaussian dictionary. However, the efficiency of the B-splines is slightly lower than the gaussian ones, since gaussian are optimal in terms of time-frequency localization. However, B-splines are much less complex to compute for two main reasons. They only require to compute polynomial functions (the factorial may be tabulated), and the support is bounded, which is particularly important for small atoms. Therefore, B-splines offer a reasonable trade-off between compression efficiency and computational load.

7.3.4 Rectangular functions

For very low power end-devices, the B-splines are still too computationally demanding. In this case, we may use one of the simplest possible functions, the 2D rectangular function. The rectangular function is still anisotropic, since different scaling may be used in orthogonal directions. They are thus able to capture oriented features, but are clearly not optimal in terms of compression efficiency. For the least powerful decoders, we may even suppress the orientation index, thus computing only horizontal rectangles.

The rectangular functions are simply given by :

$$s_{\gamma_i}(x, y) = \begin{cases} 1 & \text{if } x \in [-1, 1] \text{ and } y \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} , \quad (7.3.8)$$

where x and y are obtained from Eq. (7.3.1) and (7.3.2).

Fig. 7.6 shows the reconstruction of the *Lena* image with 500 rectangular atoms. Clearly, the quality is the worst among all above dictionaries. However, the computational load is very small. Moreover, this dictionary is destined to low end-devices which have often very low display resolution. In this case, artifacts are obviously less visible than in Fig. 7.6.



FIGURE 7.6: Reconstructed version of *Lena* encoded with 500 rectangular atoms.

7.3.5 Comparison of spatial dictionaries

We have presented in this section three new dictionaries for image decomposition through Matching Pursuit. These dictionaries cover the trade-off between compression efficiency and decoding complexity. They are now compared in terms of compression efficiency in Fig. 7.7 and 7.8 for two different images.

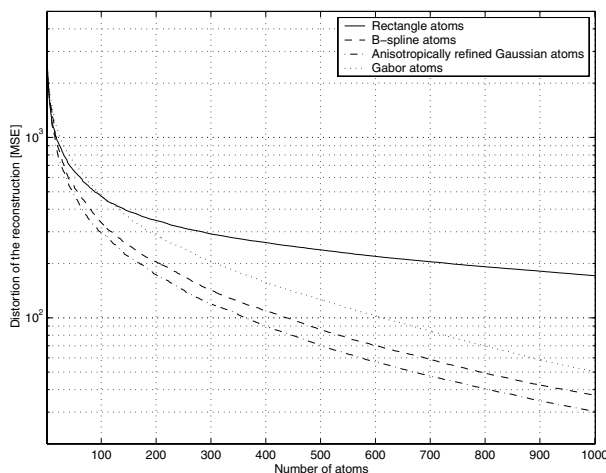


FIGURE 7.7: Distortion of the reconstructed *Lena* image as a function of the number of MP iterations.

All dictionaries integer indexes cover the same range, i.e, 128 positions, 70 scales and 128 orientations values. As expected, the best dictionary in terms of distortion of the reconstructed image is the gaussian dictionary. Gaussians minimize the uncertainty principle and anisotropic refinement guarantees an efficient coding of both edges and smooth regions. The worst dictionary, but also the most simple one, is clearly the one composed of rectangles functions. Finally, we can see that anisotropic refinement of B-splines makes them more efficient than oriented but isotropically scaled Gabor atoms, even though the latter are also built on optimal gaussian functions. Since Gabor functions have proven to enable a more efficient coding than DCT-based scheme [402, 431], we can expect that anisotropically refined gaussian or B-spline functions can still improve the compression gain provided that the coding is accurate.

Interestingly, even the worst dictionary (i.e., the rectangles functions) is able to efficiently catch

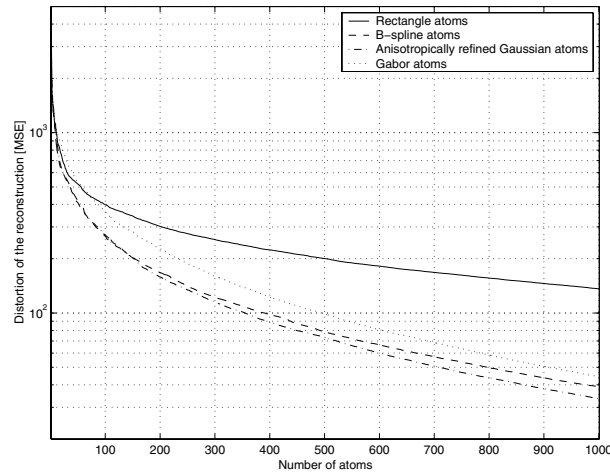


FIGURE 7.8: *Distortion of the reconstructed Autumn image as a function of the number of MP iterations*

energy of the signal in the first iterations. This property may lead to a dynamical coding scheme, where the dictionary would evolve with the iteration number. At decoder, the less powerful devices would decode only the first and simplest iterations, and the more powerful the decoder, the larger the number of decoded iterations and the larger their complexity (and thus accuracy).

7.3.6 Comparison with wavelet coding

A tentative comparison between wavelet transform and Matching Pursuit representation is now proposed. Fig. 7.9 represents the *Lena* image reconstructed from 500 MP iterations, and the 500, 1000 and respectively 1500 most important coefficients of a wavelet decomposition.

Assume that the wavelet transform outputs 2 position indexes for each coefficient. For the same rate, two wavelets coefficients are thus approximately equivalent to one MP iterations, where three additional indexes (i.e., rotation and scaling) have to be coded. We see that even 1500 wavelet coefficients are not sufficient for an equivalent distortion to 500 MP iterations, which is a ratio of 3 to 1. Therefore, we may expect Matching Pursuit to provide an efficient signal transform alternative, even though the previous comparison is not completely exhaustive.

Entropy coding and quantization possibilities should also be studied to fairly compare both schemes. Indeed, for example, DCT and wavelets have been shown to be equivalent in terms of transform efficiency, but wavelets are more appropriate for entropy coding [432, 23]. The properties of MP stream let however expect the possibility for efficient entropy coding. Indeed, a clear hierarchy may be established between coefficients, which are upper-bounded by an exponentially decaying curve. In the same time, scales tend to be larger for the first iterations, thus opening possibility for an efficient coding. The MP scheme proposed here is furthermore very generic. We may think about modifications of the encoding to improve the coding. For example, an entropy-constrained atom selection may be implemented. For equivalent correlation coefficient the atom with the closest indexes to the previous ones will thus be chosen [433]. Finally, the MP overcomplete representation becomes less efficient when the coding rate increases. Indeed, due to its greediness, MP over redundant set of functions may take an infinite number of atoms to provide an exact signal representation, while orthogonal transforms converges in a finite number of steps. Here again we face a trade-off between the convergence rate of the Matching Pursuit which increases with the redundancy, and the behavior at high rates which is better for orthogonal decompositions [405].

Fig. 7.10 gives a tentative comparison between SPIHT wavelet based coding [28] and a very basic MP coder with 2D anisotropic gaussian atoms. The position, scale and orientation indexes of



FIGURE 7.9: Comparison of MP decomposition vs 4-level wavelet decomposition of the Lena image.

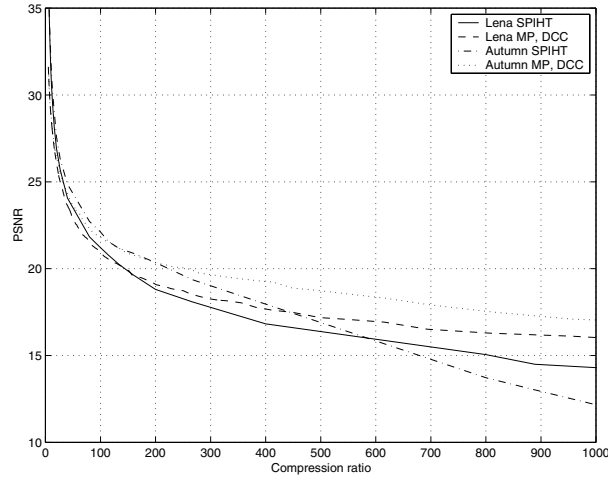


FIGURE 7.10: PSNR quality of MP decomposition with differential coefficient coding and SPIHT coding as a function of the compression ratio.

MP atoms may take 64, 35 and respectively 32 values within their respective range of $[0..128]$ for the positions and scales, and $[0..2\pi]$ for the orientation parameters. Each atom index is thus coded with 28 bits. The coefficient norms are coded differentially in this basic scheme. It can be shown that even with this basic coding scheme, MP provides results comparable with a very efficient

wavelet coder, which is the SPIHT algorithm. As expected, for compression ratios larger than approximately 150 (i.e., low coding rate), Matching Pursuit provides better results than SPIHT. Keep in mind however that these results are only indicative since work still remains to be done for an efficient coding of MP elements.

7.3.7 Discussion

The goal of this section was not to find the optimal dictionary, but rather to show the main design principles which lead to efficient Matching Pursuit encoding. The examples provided above moreover correspond voluntarily to very low bit rate streams, so that the coding artifacts are clearly visible and may offer comparison possibilities between the respective dictionaries. We think that there is no truly optimal dictionary atoms in the heterogeneous framework we are working in. The dictionary may have to be adapted to the input signal, or to a class of input signals, and to the set-up of the video distribution tree including the receivers capabilities. In this context dictionaries composed of several basis functions can be very helpful. Matching Pursuit makes possible such an adaptation with very few effort. For example, a single flag may indicate to the decoder which basis function is used so that it can correctly decode the stream.

To improve dictionaries, one may also take into account the features of the Human Visual System (HVS). For example, the relative ratio between scaling indexes can be limited to avoid long and narrow atoms which are very annoying to the HVS (see Fig. 7.5). Such criteria can be incorporated into the design of the dictionary. Finally, HVS considerations may also be incorporated into the Matching Pursuit algorithm itself. They could help to choose at each iteration the most visually important features rather than the one leading to the maximal inner product.

7.4 Spatio-temporal atoms

7.4.1 Gaussian function

The temporal function of three-dimensional atoms has to catch very different features than the spatial function. Basically, it has only to report the redundancy between successive frames. Due to their optimal time-frequency localization properties, gaussian functions can also be used in the temporal dimension. In the same time, they are relatively complex to compute because of their non-limited support. However, this complexity is much less important than in the spatial domain, since the size of the GOF is obviously not in the same range than the frame size. It has to be noted that modulation of the gaussian is not suited to catch temporal features in natural sequences. The generic Gabor functions can thus be eliminated from the candidates to temporal basic functions. The gaussian function can be expressed as :

$$t_{\gamma_i}(z) = e^{-z^2}, \quad (7.4.1)$$

where z is obtained by scaling and translating temporal coordinates $\tilde{z}$ as :

$$z = \frac{\tilde{z} - p_z}{s_z}. \quad (7.4.2)$$

The temporal indexes are the position and scaling parameters p_z and s_z . Note that the range of the temporal indexes is much smaller than the range of the spatial ones, thus clearly enabling a lower index rate. By the construction of three-dimensional atoms, the temporal gaussian functions allows to weight the contribution of the spatial features into successive frames. Fig. 7.11 shows the four first frames of the CIF *Foreman* encoded with 500 3D atoms built on 2D B-splines and gaussian temporal function. The reconstructed quality is quite impressive considering the very low coding rate. 500 iterations indeed correspond roughly to 22 kbit without coefficient quantization or any further coding. This emphasizes the efficiency of the temporal function and more particularly the spatio-temporal decomposition. Indeed, in the case of GOF of 4 frames, 2 bits suffice to code temporal position and scale indexes. That means that the addition of 4 bits to the spatial atoms indexes offers to possibility to efficiently extend the image decomposition to video coding.

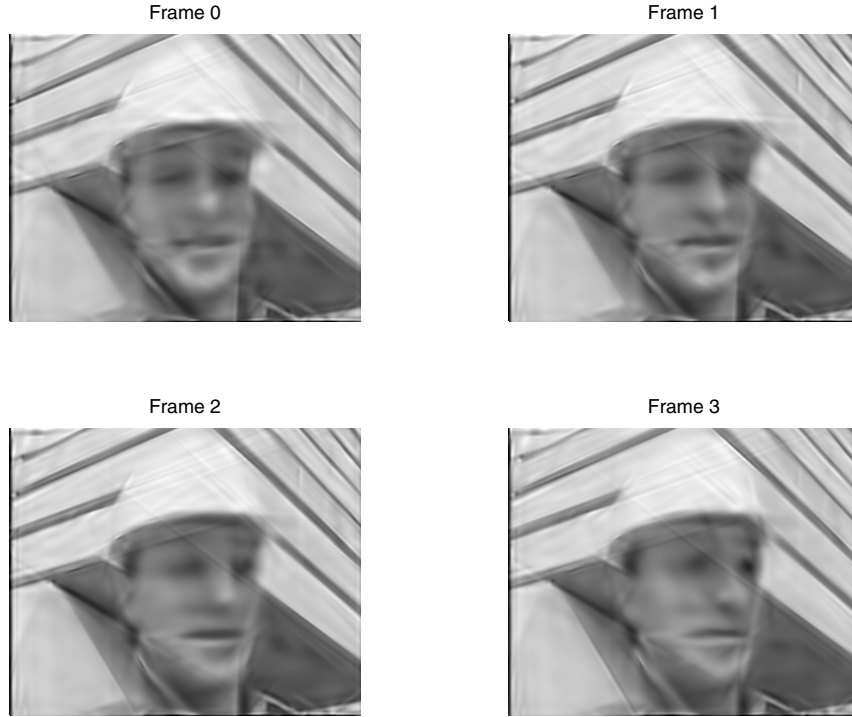


FIGURE 7.11: *Reconstruction of the four first frames of Foreman encoded with 500 iterations of 2D B-splines and gaussian temporal function.*

7.4.2 B-spline function

Similarly to the spatial atoms building functions, we can also use B-splines in the temporal domain. B-splines are less complex to decode than gaussian functions, mainly because their support is limited. We will choose a second order B-spline to limit the computation along the temporal dimension. The temporal function may therefore be expressed as :

$$t_{\gamma_i}(z) = \frac{1}{2} \sum_{j=0}^3 \binom{3}{j} (-1)^j \left[z - j + \frac{3}{2} \right]_+^2, \quad (7.4.3)$$

where z is defined by Eq. (7.4.2) and $[z]_+^n$ represents the positive part of z^n . B-splines allow similar flexibility as the gaussian in the design of the temporal function. Meanwhile, they are almost as efficient, since the size of the GOF generally do not permit to really take benefit of the optimal time-frequency localization property offered by the gaussian function. They permit to weight differently spatial features into successive frames. Fig. 7.12 shows the four first frames of *Foreman* encoded with 500 3D atoms built on 3D B-splines functions. It can be seen that the quality is similar to the quality of coding with gaussian functions. The complexity is however much smaller since the computation of a second order B-spline is cheaper than the calculation needed by gaussian functions.

7.4.3 Rectangular function

The most simple temporal function which may be proposed is again the rectangular function. In this case the rectangular function is however expected to be more efficient than in the spatial domain. Indeed, the features to capture are completely different and the rectangular function may efficiently catch temporal redundancy, particularly in sequences with low temporal motion.

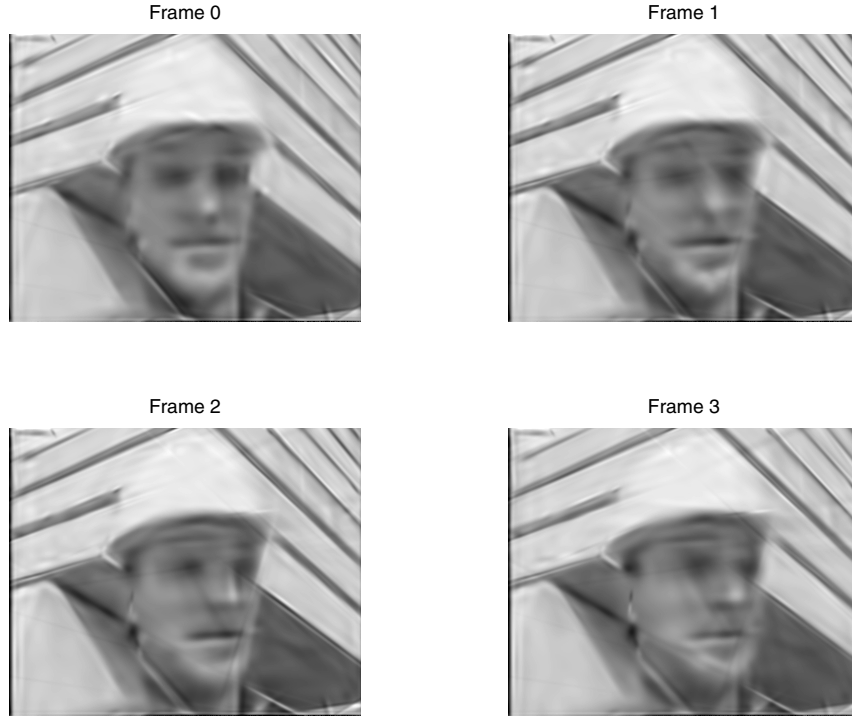


FIGURE 7.12: *Reconstruction of the four first frames of Foreman encoded with 500 iterations of 3D B-splines function.*

Basically, the rectangular temporal function only repeats spatial atoms, with the same amplitude, in successive frames. The temporal rectangular function can be written as :

$$t_{\gamma_i}(z) = \begin{cases} 1 & \text{if } 0 \leq z < 1, \\ 0 & \text{otherwise.} \end{cases}, \quad (7.4.4)$$

where z is defined by Eq. (7.4.2). Fig. 7.13 shows the four first frames of *Foreman* encoded with 500 3D atoms built on 2D B-splines and rectangular temporal function. It can be seen that the temporal redundancy is efficiently captured by temporal functions as simple as rectangles. The quality of the four frames is quite acceptable for the very low encoding rates. The quality is even comparable to 3D B-splines functions, which offer however much more flexibility which can be advantageous in high motion sequences. In this case, the rectangular function indeed only permits to code the presence or absence of spatial features but not to weight their contribution as B-splines do.

7.4.4 Haar function

Inspired from temporal wavelets function in three-dimensional subband coding [51], we can use a simple 2-tap Haar wavelet as temporal component in Matching Pursuit atom. This temporal function induces a computing complexity similar to rectangle functions. They however offer additional flexibility by their anti-symmetrical structure. This function is given by :

$$t_{\gamma_i}(z) = \begin{cases} 1 & \text{if } 0 \leq z < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} \leq z < \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}, \quad (7.4.5)$$

where z is defined by Eq. (7.4.2). Intuitively the Haar function which is relatively efficient in subband coding do not present real benefits in Matching Pursuit coding. It is indeed relatively

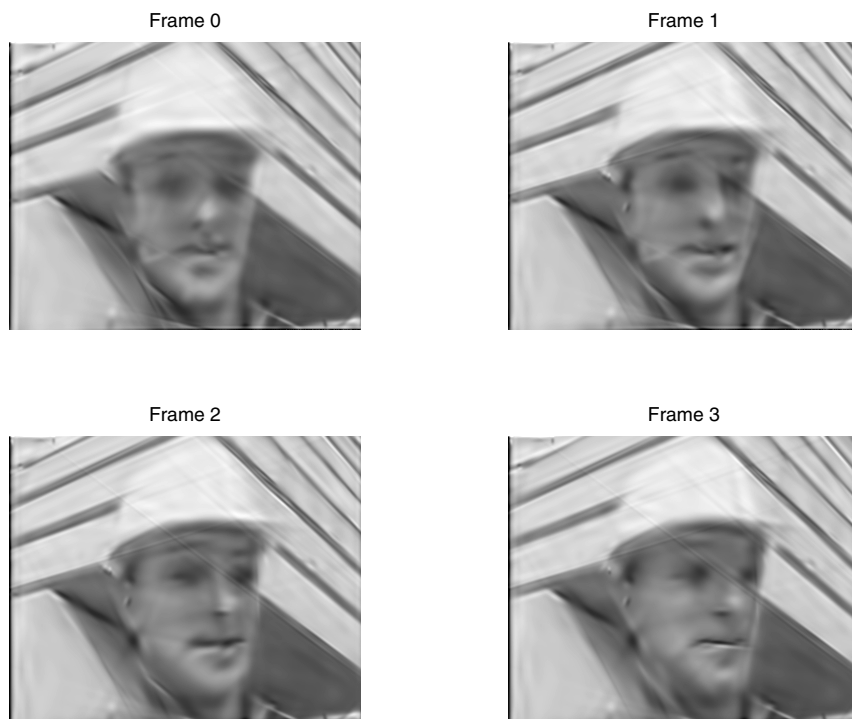


FIGURE 7.13: *Reconstruction of the four first frames of Foreman encoded with 500 iterations of 2D B-splines and rectangular temporal function.*

rare in natural scenes that spatial features present in a frame have a inverse amplitude in the following frame. In the same time, the anti-symmetrical structure of the Haar function can be very interesting in the highest order iterations to improve small details. It can be also useful to capture slight motions within video scenes. Fig. 7.14 shows the four first frames of *Foreman* encoded with 500 3D atoms built on 2D B-splines and Haar temporal function. It can be seen that the Haar function is, as expected, slightly less efficient than the simple rectangle functions for low rate coding. The properties of the Haar function are therefore not very useful in low rate three-dimensional MP coding.

7.4.5 Comparison of temporal functions

Fig. 7.15 compares the efficiency of the temporal functions proposed before in terms of distortion of the reconstruction as a function of the number of spatio-temporal atoms. The distortions have been averaged over 12 groups of four frames ($\text{GOF} = 4$). It can be seen clearly that no real winner can be designated among the temporal functions. All functions have almost the same efficiency in terms of MSE. Note that the spatial functions (i.e., 2D B-spline) are not necessarily exactly identical for all the temporal functions at a given iteration. This has however only a slight influence on the given results thanks to the averaging over several GOFs. The equivalence of the different temporal functions can be explained as follows. The temporal redundancy is mainly captured by rectangular functions, especially in low motion sequences. Spatial features are indeed present in one or several successive frames with the same intensity. In low rate coding most of the atoms capture low frequency components efficiently coded by rectangular functions. It is rare that the spatial features are weighted differently in successive frames, besides the case of high motion sequences or zooming effects, for example. Finally, the small GOF size due to encoding delay constraints does not allow to completely take benefit of the enhanced properties of gaussian

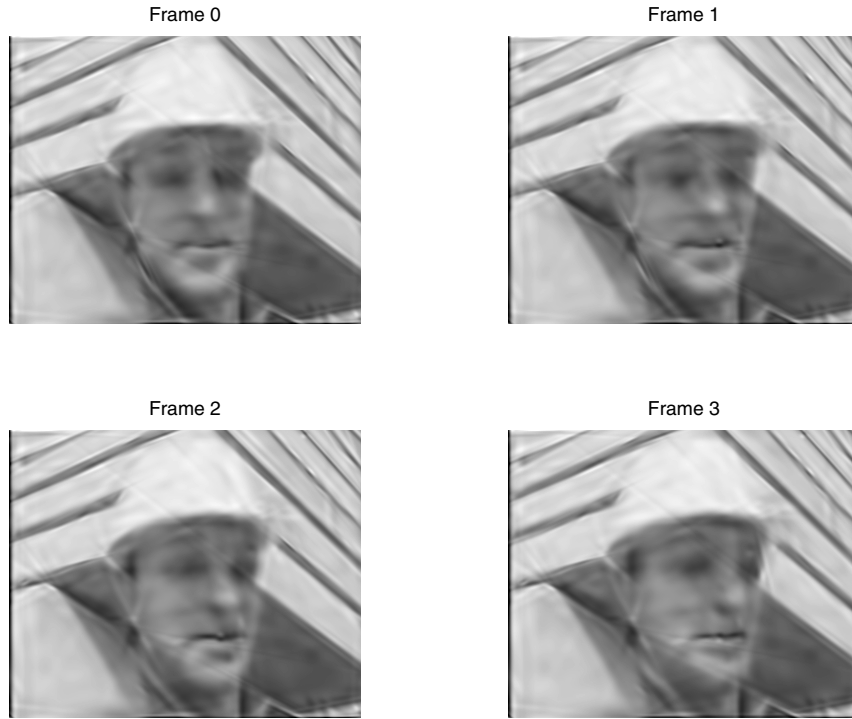


FIGURE 7.14: *Reconstruction of the four first frames of Foreman encoded with 500 iterations of 2D B-splines and Haar temporal function.*

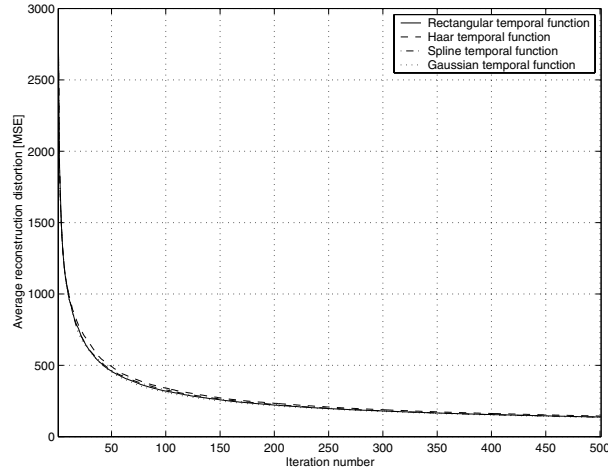


FIGURE 7.15: *Reconstruction distortion as a function of the number of iterations per group of four frames of Foreman. The atoms are made on 2D B-splines and four different temporal function.*

and even B-spline function.

Even if the MSE does not capture any sensitive difference between the temporal functions, preference may be visually given to gaussian or spline functions according to the Fig. 7.13 to 7.11. Clearly, the MSE is not sufficient to evaluate the reconstruction quality of the MP encoding. It can be worth investigating the insertion of perceptual criteria [434] in the encoding scheme [435]. The

topic will nevertheless not be covered here. In the meantime, for complexity reasons, the usage of simple rectangle functions in the temporal domain may often be sufficient. They indeed clearly offer the best trade-off between efficiency and decoding complexity. In this case this compromise is however less important than for the spatial atoms since the number of computations are in any case much smaller.

7.4.6 Comparison with coding standards

A tentative comparison is now proposed between 3D Matching Pursuit and video coding standards. Fig. 7.16 represents the evolution of the coding quality as a function of the coding rate. 40 frames of the CIF *Foreman* sequence have been encoded with H.263+ and MPEG-2 with a GOP size of 12. The encoding has been performed without rate control, and the quantizer scale factors have been chosen to cover the range of the smallest bit rate values. Matching Pursuit has used a spatio-temporal dictionary of 2D B-splines and gaussian temporal functions. The MP coefficients have been truncated to the closest integers and differentially encoded to reduce the coding rate. No entropy coding or even additional quantization has however been performed on the Matching Pursuit stream.

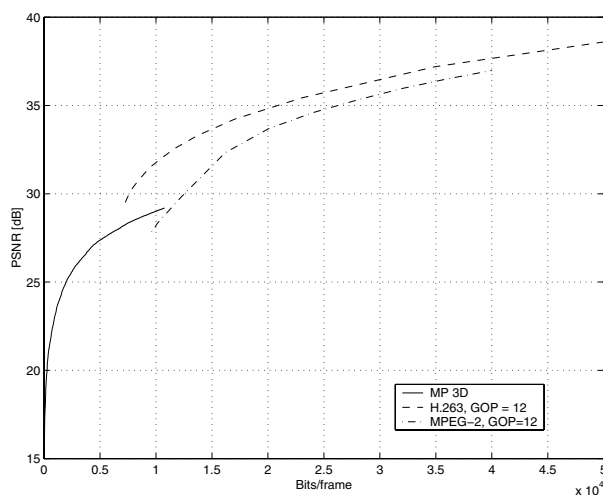


FIGURE 7.16: *PSNR quality of MP decomposition with differential coefficient coding and H.263+ and MPEG-2 as a function of the encoding rate for the Foreman sequence.*

Clearly, Matching Pursuit offers very similar performance to coding standards even in the absence of entropy coding and quantization. It moreover allows to reach much smaller coding rates thanks to the absence of complex coding hierarchy. It provides a slightly lower quality than H.263+, but a better quality than MPEG-2 which is however not optimized for such low rates. The Matching Pursuit as implemented here is however expected to become less efficient at very high rates due to the non-linear encoding on a redundant dictionary. Note that similar results are obtained with other sequences.

Figures 7.17 and 7.18 compare the four first frames of *Foreman* encoded with respectively 3D Matching Pursuit and H.263+ at a quantizer scale of 31. The coding rates are equivalent but the Matching Pursuit stream is not quantized or entropy coded. It can be seen that the coding artifacts are completely different due to the block-based structure of H.263+. The distortion is more smooth in Matching Pursuit coding. Even though the PSNR is slightly better for H.263+, the visual quality of Matching Pursuit is thus preferable.

The above comparison is however only qualitative. Indeed, no real coding is performed on the Matching Pursuit stream. Moreover the encoding process based on Genetic Algorithm and the Matching Pursuit dictionary are not truly optimized. The comparison however shows that Match-

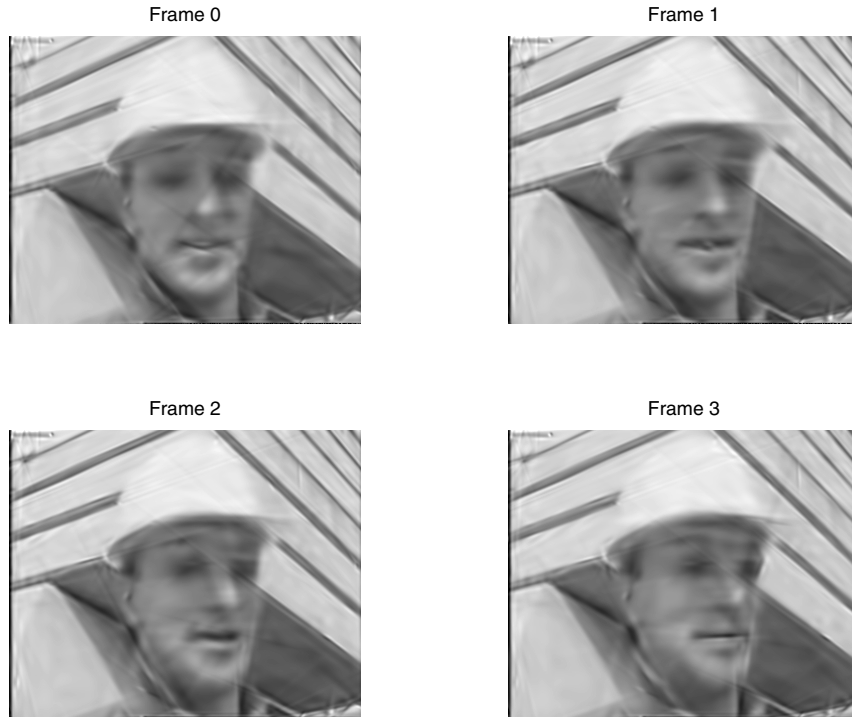


FIGURE 7.17: *Reconstruction of the four first frames of Foreman encoded with 800 2D B-splines and temporal gaussian atoms.*

ing Pursuit provides a very interesting alternative to coding standards in terms of compression, while it moreover brings additional flexibility as shown in the next chapters.

7.5 Discussion and improvements

7.5.1 Size of the Group of Frames

An important issue in three-dimensional video coding is the size of the Group of Frames (GOF) simultaneously encoded by the spatio-temporal Matching Pursuit. Large GOFs may improve the compression efficiency, since few atoms are needed to code low frequency components. This improvement however decreases when the temporal motion in the sequence becomes important. Note that in three dimensional MP coding we do not encounter the wavelet *edge* problem due to the limitation of the size of the signal [24]. Indeed, the Matching Pursuit decomposition does not translate the signal into the frequency domain, and thus do not need theoretically infinite length signals to correctly describe localized features. The Matching Pursuit can build a dictionary with very localized pattern and hence code them efficiently for any signal size.

However, the larger the GOF, the larger the encoding delay. This is not a critical issue in asymmetric applications like we are targeting in this work. The delay mainly postpone the display of the first frame of the video sequence. However, the delay has clearly to be limited for the comfort of the end-user and to permit a reasonable level of interactivity. We will thus limit the encoding delay to less than half a second, or equivalently to 4 to 8 frames, depending on the frame rate.

The Figure 7.19 compares the reconstruction distortion for 3D splines atoms computed on GOF of length 4 and 8 respectively. The coding rate is indicative since no optimization is performed in terms of entropy coding for example, but it takes into account the additional bit necessary to code

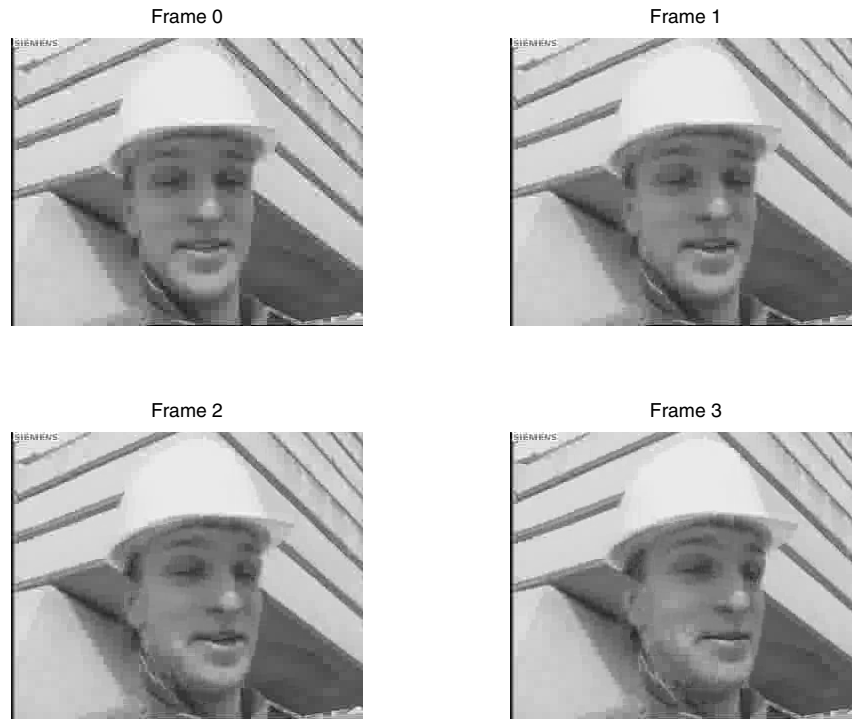


FIGURE 7.18: *Reconstruction of the four first frames of Foreman encoded with H.263+ at a quantizer scale of 31.*

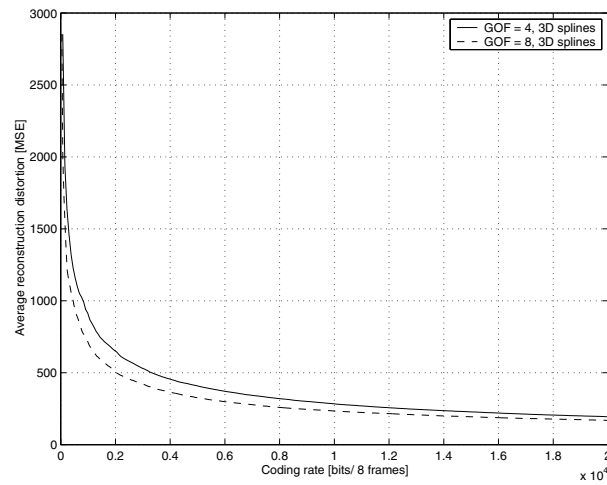


FIGURE 7.19: *Reconstructed distortion of Foreman as a function of the coding rate for 3D splines atoms and GOF size of 4 and 8.*

temporal position and scale indexes when the GOF is double. Despite this additional bits, the longer GOF clearly provides a lower distortion, especially for the first iterations where most of the energy is captured by large atoms. For higher iterations, or when the coding rate increases, the difference between both GOF sizes decreases, since coding of details is equivalent in both cases. Note however that *Foreman* is a low motion sequence clearly in favor of long GOFs. The influence of the GOF size is less pronounced in high motion sequences.

7.5.2 Local motion estimation

For highly moving sequences the 3D decompositions proposed hereabove may be not sufficiently efficient. Indeed, it can not catch the movement of objects or pattern but only report their presence or absence at a given location within the GOF. To cope with this limitation, we can build oriented temporal functions. The aim of these functions is to capture local motion of spatial features. Note that this kind of motion estimation does not generate error propagation since no reference frame is defined in contrary to the common motion compensation process used in schemes like MPEG.

Two additional temporal indexes are thus defined to capture the local motion of the spatial atoms. A motion vector (h_m, v_m) defines the horizontal and vertical displacement of the spatial atom between successive frames. In this case, the values of the three dimensional atom at pixel of coordinates $(\tilde{x}, \tilde{y}, \tilde{z})$ is expressed as :

$$g_{\gamma_i}(\tilde{x}, \tilde{y}, \tilde{z}) = s_{\gamma_i}(x + z h_m, y + z v_m) t_{\gamma_i}(z) . \quad (7.5.1)$$

Due to the temporal motion, the spatial atoms support may be wider than in the previous spatio-temporal dictionaries. Indeed some functions which are not entirely visible in the previous dictionaries can become completely visible in a frame due to the local motion. The phenomenon can slightly increase the decoding complexity, but the additional complexity is very small compared to the complexity of the whole process.

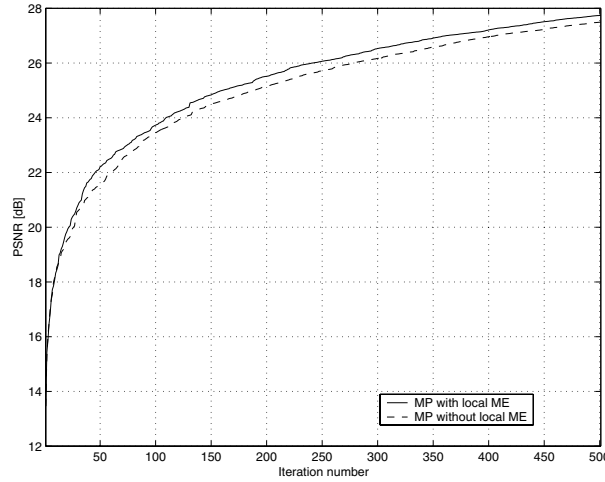


FIGURE 7.20: Comparison of the encoding quality for oriented and non-oriented 3D B-splines for the Foreman sequence.

Fig. 7.20 compares the encoding quality of the CIF *Foreman* sequence with and without local motion estimation on 3D B-splines. It can be seen that the quality is slightly better for oriented atoms. However, the *Foreman* sequence does not contain high temporal activity, so that the quality improvement is not really impressive. The main part of the MP atoms get indeed zero motion vectors. Furthermore, the addition of two indexes significantly reduces the convergence speed of the encoder based on Genetic Algorithm. Due to the additional complexity and coding cost, such a basic local motion estimation scheme should be used carefully for sequences with very low motion. Local motion estimation is finally expected to bring more benefits for large GOFs and sequences with high temporal activity. It can be also particularly useful for sequences where the frames are dropped to meet bandwidth requirements.

7.5.3 Adaptive dictionaries

As seen before, it is difficult to define the optimal encoding dictionary in the heterogeneous framework we are working with. Indeed the dictionary has rather to meet a compromise between

compression efficiency and the receiver decoding capabilities. Moreover, it has to adapt to the input signal or signal class, since the same dictionary obviously may efficiently capture the features of a particular signal and be totally inefficient for another signal. The most efficient coding in an heterogeneous environment is surely an adaptive or dynamic coding. The adaptivity may be thought in terms of both adaptivity to the decoder and to the input signal. The adaptivity to the input signal is not studied in this work. The adaptivity to the decoder concerns both the decoding complexity and the device resolution.

Adaptive encoding may be seen in terms of hierarchical encoding. The first iterations of Matching Pursuit can be computed on a sub-dictionary, of reduced complexity. The dictionary can then grow with more complex but also accurate functions destined to more powerful devices. Similarly, thanks to the group properties, the encoding can also follow a hierarchy in decoding spatio-temporal resolution. Indeed, the encoder can encode the first iterations onto sub-sampled GOFs, with a sub-sampled dictionary. Thanks to the group law, these atoms can be decoded on larger resolutions by simply adapting scales and positions indexes. Similar process is used for transcoding, and will be describe in more details in the next chapter. Primary results have shown that the hierarchical resolution encoding does not loose in terms of compression efficiency, while greatly lowering the coding complexity. The problem in hierarchical encoding is however to define the right number of iterations in each loop.

7.6 Conclusion

This chapter has presented a new way of coding video sequences with Matching Pursuit. Common MP-based video implementations code the displaced difference frame with Gabor dictionaries [399, 87, 88, 86]. The beauty of Matching Pursuit is its incomparable design flexibility. Any function may be used in the dictionary, as soon as the completeness constraint is respected. This flexibility is of greatest interest in the heterogeneous framework defined in the first chapter. Several 3D dictionaries have been proposed, with a special care on the trade-off between compression efficiency and decoding complexity. Matching Pursuit have been shown to be at least as efficient as classical transform coder (e.g., DCT or wavelets) [402], with additional advantages however. Matching Pursuit have intrinsic multiresolution properties which can be exploited in proxy-based transcoding, as it will be shown in the next chapter, or for hierarchical MP coding [424]. Matching Pursuit builds an intrinsically embedded stream, since iterations are of decreasing importance along the coding. The coding of colour images or sequences is furthermore expected to be very efficient with MP in comparison with standard-based coding. It can indeed be assumed that the same patterns are present in all colour components. In this case, the indexes are coded only once for the three components. Hence, colour coding need only two additional coefficients, while classical coding scheme independently codes the colour components, often down-sampled. Finally, this chapter does not provide with an optimal dictionary, which may moreover not exist in an heterogeneous context. The proposed Matching Pursuit coder is only one possible implementation whose complexity can be decreased by hierarchical coding, for example. Hierarchical coding can be particularly efficient with large input signals.

Redundancy-Driven Quantization of MP stream

8.1 Why quantization ?

The aim of this chapter is to study the effects of quantization onto reconstruction of Matching Pursuit streams. Since Matching Pursuit coefficients generally take on real values, quantization is necessary to reduce the bandwidth needed for transmission. Quantization errors have been studied in [437, 404, 438, 439] in the context of overcomplete frame expansions and consistent Matching Pursuit. Similarly, in the case of real or even very large dictionaries, atom indexes should also be quantized to limit the bit rate. This chapter studies the quantization error of the coefficients and atoms for a general Matching Pursuit decomposition. This study focuses particularly on the transmission of Matching Pursuit streams over heterogeneous networks.

The quantization error is thus studied for *a posteriori* quantization schemes, in contrast with usual schemes [431, 440, 433, 441] where the encoder uses the quantized coefficient to update the residual signal. In *a posteriori* quantization, the quantized version of the elements does not influence the decomposition. Hence the Matching Pursuit stream can be computed only once and then quantized several times to serve different constraints. The signal reconstruction error due to quantization is shown to depend on the error on both the coefficient and the atom. Moreover, the contribution of each Matching Pursuit element to the quantization error clearly depends on its position within the encoded stream, and hence on the dictionary redundancy, as computed in Sec. 6.4. Based on the characterization of the energy decay curve, an exponentially bounded quantization scheme is proposed for the Matching Pursuit coefficients. This scheme truly outperforms uniform quantization scheme, especially at low bit rates. The scheme is shown to achieve very good results in the practical case of image compression.

The set of functions that forms the dictionary obviously plays a crucial role in the Matching Pursuit coding. A very redundant dictionary generally allows to capture the main features of the signal with only a few dictionary functions. However, the coding rate of the function parameters obviously increases with the size of the dictionary. We will study the index quantization error in structured dictionaries, and more particularly 2D Gabor dictionaries. It will finally be shown that index quantization can not be performed in an optimal way. This conclusion yet emphasizes the importance of a correct design of the dictionary to avoid the need of index quantization.

This chapter is organized as follows. An upper-bound on the signal reconstruction error is given in Section 8.2 for quantization errors. Section 8.3 studies the *a posteriori* quantization of Matching Pursuit coefficients and proposes an exponential quantization which takes advantage of the properties of the encoding. Section 8.4 studies a structured dictionary index quantization

Parts of this work appeared in [395, 436].

scheme driven by the redundancy of the dictionary. It therefore focuses on the error due to wrong atom selection at reconstruction. Finally, concluding remarks are given in Section 8.5.

8.2 Signal reconstruction error

In practical cases an exact reconstruction of the signal is rarely possible. First, the number of iterations is generally limited due to computational cost and bandwidth constraints. Thus, the signal is approximated by the first N iterations of the Matching Pursuit stream. Note also that the exact signal reconstruction may need an infinite number of iterations in MP with redundant dictionaries. Second, the Matching Pursuit elements (i.e., coefficients and dictionary indexes) have often to be quantized to reduce the overall size of the signal representation. Indeed, the MP coder generally outputs real coefficients, which clearly need to be quantized to limit the dedicated rate. Moreover, if the dictionary is very large to ensure a good convergence, or even if it contains real parameters, the atoms should also be quantized before transmission. The framework that is considered here is an *a posteriori* quantization, which is very useful in a context where a stream is encoded once and then quantized several times to meet different constraints. Matching Pursuit is indeed the heaviest part of the encoding. Therefore, a different encoding for each client would be too demanding. Moreover, Matching Pursuit systems are particularly well suited for asymmetric applications, where the same stream serves several possibly heterogeneous devices. Third and finally, transmission noise can also alter the signal reconstruction. Transmission errors can be studied in the same way as quantization error, under certain conditions. All the results of this section are given for quantization error, though most of them can easily be translated into transmission error.

Let c_{γ_n} denote a MP coefficient, or equivalently the inner product $\langle g_{\gamma_n} | \mathcal{R}^n f \rangle$. The squared reconstruction error Δ between the signal approximation and its reconstructed version $\tilde{f}$ can be written as

$$\Delta = \|f_N - \tilde{f}\|^2 = \left\| \sum_{n=0}^{N-1} c_{\gamma_n} g_{\gamma_n} - \sum_{n=0}^{N-1} \tilde{c}_{\gamma_n} g_{\tilde{\gamma}_n} \right\|^2, \quad (8.2.1)$$

where $\tilde{c}_{\gamma_n}$ and $g_{\tilde{\gamma}_n}$ are respectively the distorted coefficient and atom. Let $\xi_n = c_{\gamma_n} - \tilde{c}_{\gamma_n}$ denote the error on the coefficient. Similarly, let E_n be the error on the n^{th} atom, with $E_n = g_{\gamma_n} - g_{\tilde{\gamma}_n}$. The reconstruction error can thus be decomposed further as

$$\begin{aligned} \Delta &= \left\| \sum_{n=0}^{N-1} (c_{\gamma_n} g_{\gamma_n} - (c_{\gamma_n} - \xi_n) g_{\tilde{\gamma}_n}) \right\|^2 \\ &= \left\| \sum_{n=0}^{N-1} (c_{\gamma_n} E_n + \xi_n g_{\tilde{\gamma}_n}) \right\|^2. \end{aligned} \quad (8.2.2)$$

By triangular inequality, Δ can be bounded by

$$\begin{aligned} \Delta &\leq \left\| \sum_{n=0}^{N-1} c_{\gamma_n} E_n \right\|^2 + \left\| \sum_{n=0}^{N-1} \xi_n g_{\tilde{\gamma}_n} \right\|^2 \\ &\leq \sum_{n=0}^{N-1} |c_{\gamma_n}|^2 \|E_n\|^2 + \sum_{n=0}^{N-1} |\xi_n|^2 \|g_{\tilde{\gamma}_n}\|^2. \end{aligned} \quad (8.2.3)$$

Finally, since $\|g_{\gamma}\| = 1$, the error bound can be written as

$$\Delta \leq \sum_{n=0}^{N-1} |c_{\gamma_n}|^2 \|E_n\|^2 + \sum_{n=0}^{N-1} |\xi_n|^2. \quad (8.2.4)$$

Eq. (8.2.4) provides an upper-bound to the reconstruction error, due to either quantization or transmission. The errors on the coefficients and indexes can thus be separated in finite dimensions. The reconstruction error is bounded by the sum of errors on both the coefficients and the atoms, weighted by the respective contributions. The first term of Eq. (8.2.4) corresponds to the atom selection error multiplied by the respective coefficients. For highly redundant dictionaries, only the first few atoms participate to the reconstruction error, since the coefficient norm rapidly decreases. Meanwhile, if the atoms are correctly decoded (i.e., the first term is zero), the reconstruction error can be bounded by the sum of the error on the coefficients. Due to the exponential convergence of MP, this error depends generally almost only on the first coefficients, i.e., the most important ones.

It has to be noted here that consistent reconstruction algorithms [404] can improve the quality of the reconstructed signal. Such concealment algorithms results in a reconstructed signal which would produce the same MP decomposition sequence as the original. However, such algorithms are not considered in this chapter due to complexity constraints on the decoder. In our case, the receiver simply reconstructs the signal from the available coefficients and atom indexes, without error detection or concealment algorithms.

In the next sections, errors on both coefficients and atom indexes are studied in the context of quantization. The atom selection error is analyzed in detail for common Gabor atoms.

8.3 Quantization of Matching Pursuit coefficients

8.3.1 A posteriori coefficients quantization

This section uses Matching Pursuit properties in the quantization of coefficients. Recall that the purpose of the *a posteriori* quantization scheme lies in the context of completely asymmetric applications. Due to the encoding complexity, the Matching Pursuit stream is computed only once and then differently quantized to match various receiver requirements. The coefficient quantization cannot be included in the Matching Pursuit decomposition in contrast to [437, 404, 438, 439, 431, 440], since this would imply dedicated encoding for each end-user. While *a posteriori* quantization provides flexibility for the quantization, its distortion is not compensated by further MP iterations. The quantization error propagation can however be kept small for sufficiently fine quantization. Fig. 8.1 shows that distortion resulting from a posteriori quantization is slightly higher than in the

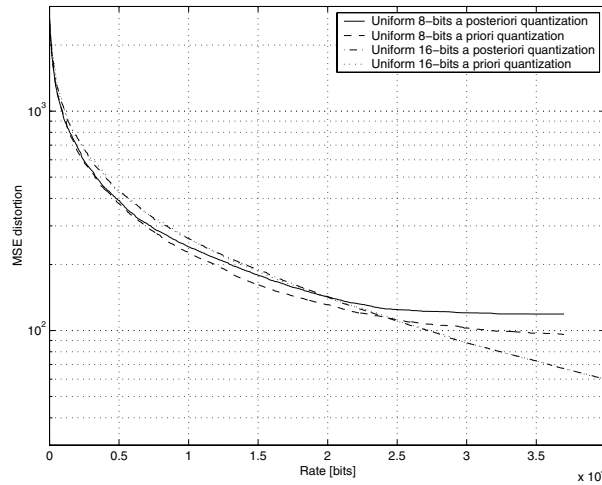


FIGURE 8.1: Comparison of *a priori* and *a posteriori* uniform coefficient quantization (8 bits) for the Lena image.

in-loop quantization case, depending on the quantization step size however. Indeed, in the latter

case, the MP adapts to the quantized coefficient, and the residue corresponds to the stream used for reconstruction. However, such a scheme is less flexible than the *a posteriori* quantization since the stream is targeted for a given rate. Moreover, it can be seen that the distortion increases mainly for the highest iteration orders, where coefficients become smaller. For the first few coefficients, the relative quantization error is small in both cases. Finally, for an overall fine quantization, the error is similar in both quantization schemes. Note also that in-loop quantization may slow down the convergence speed of the MP.

The aim of the quantization is to offer the best possible reconstruction quality for a given bit budget. The optimal solution minimizes the reconstruction error given by Eq. (8.2.1), where the atoms $\|E_n\| = 0$ are correctly decoded. We propose a sub-optimal solution using the exponential decay of the residual energy. This interesting property directly drives the quantization of the coefficients in two ways. Intuitively, the quantization error on a Matching Pursuit element depends first on the iteration. Therefore, the highest iteration elements can be more coarsely quantized than the first elements. Second, the number of Matching Pursuit elements can also be adapted to the available bandwidth, since the highest iterations carry very low energy.

8.3.2 Coefficient norm decay

Since the Matching Pursuit elects at each step only the most important atom from an overcomplete dictionary, the coefficients have a decreasing importance along the iteration number. Recall that the coefficient c_{γ_n} corresponds to the scalar product $|\langle g_{\gamma_n} | \mathcal{R}^n f \rangle|$. A Matching Pursuit iteration can be thus characterized as

$$\mathcal{R}^n f = c_{\gamma_n} g_{\gamma_n} + \mathcal{R}^{n+1} f. \quad (8.3.1)$$

Since $\mathcal{R}^{n+1} f$ is orthogonal to g_{γ_n} thanks to Matching Pursuit properties, an energy conservation relation could be written for each iteration:

$$\|\mathcal{R}^n f\|^2 = |c_{\gamma_n}|^2 + \|\mathcal{R}^{n+1} f\|^2, \quad (8.3.2)$$

where $\|g_{\gamma_n}\| = 1$. Eq. (8.3.2) means that the norm of the coefficient is strongly related to the energy of the residue. From Eq. (6.2.9), this norm is thus upper-bounded by an exponential function, which can be written as

$$|c_{\gamma_n}| \leq (1 - \alpha^2 \beta^2)^{\frac{n}{2}} \|f\|. \quad (8.3.3)$$

The upper-bound depends on both the energy of the input function and the construction of the dictionary. More particularly, as described in Sec. 6.4, highly redundant dictionaries induce a very rapid decay of the coefficients norm. Indeed, since the coefficient can obviously not bring more energy than the residual function, the norm of the coefficient is strongly related to the residual energy decay curve.

Figure 8.2 represents the energy of the Matching Pursuit coefficients for the coding of images with a Gabor dictionary. It clearly shows that the coefficient norm can be upper-bounded by an exponential decay curve. These qualitative result does not depend neither on the signal nor on the dictionary. However, the decay rate directly depends on the redundancy factor given by Eq. (6.4.11). Moreover, in this case, MP search is not optimal in contrast to the experiments presented for the random dictionaries. An exhaustive search would be too demanding in the case of images, for which a sub-optimal method has been provided using Genetic Algorithms [421]. This introduces a sub-optimality factor α (see Eq. (6.2.11)) smaller than one and influences the decay rate of the coefficients.

8.3.3 Redundancy-driven uniform coefficients quantization

The exponential upper-bound of the coefficients is now used to design an efficient quantization scheme. Indeed, there is clearly no need to quantize all coefficients on the same range, since their

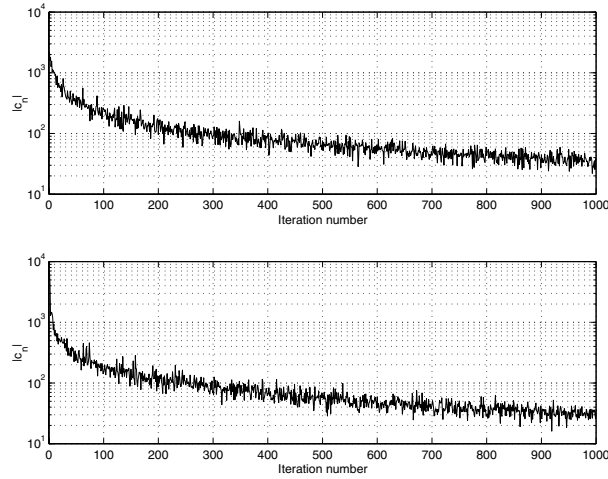


FIGURE 8.2: Coefficient norm versus the iteration number. The MP coefficients using a 2D Gabor dictionary are represented for the Lena and Autumn images.

value decreases exponentially. In other words, the quantization applied to the first coefficients is certainly not efficient on the last ones, as the range of the latter is much smaller. Bits can thus be saved by simply limiting the quantization region between 0 and the exponential decay curve given by the parameters (i.e., β and $\|f\|$). The accuracy of the quantization is also improved, especially for decompositions over highly redundant dictionaries. For the signal to be reconstructed, the decoder needs the parameters of the exponential upper-bound. An additional bit of sign finally suffices to completely characterize the coefficients¹.

The number of coefficients, as well as the number of bits per coefficient have now to be optimized in the context of exponentially upper-bounded quantization range. The distortion at decoding can be bounded by the sum of the quantization error and the approximation error due to the limit on the number of decoded iterations. In the case where atom indexes are correctly received, the total distortion can be written as using Eq. (8.2.4) :

$$\begin{aligned} D &\leq \sum_{n=0}^{N-1} |\xi_n|^2 + \|\mathcal{R}^N f\|^2 \\ &\leq \sum_{n=0}^{N-1} |\xi_n|^2 + (1 - \alpha^2 \beta^2)^N \|f\|^2, \end{aligned} \quad (8.3.4)$$

where the energy of the residue at iteration N is bounded thanks to Eq. (6.2.9). Assume now that the distribution of the coefficients norm is uniform between 0 and the exponential upper-bound given by Eq. (8.3.3). This distribution depends on the Matching Pursuit search algorithm. Keep in mind though that the assumption of a uniform distribution is valid only as a first approximation.

For complexity reasons, and under the previous assumptions, the coefficient c_j is uniformly quantized within the exponentially decaying quantization range $I_j = \nu^j \|f\|$, where $\nu = (1 - \alpha^2 \beta^2)^{\frac{1}{2}}$. Let n_j be the number of quantization steps within I_j for the quantization of the j^{th} coefficient. In the case of sufficiently fine uniform quantization, the distortion can therefore be written as :

$$D_Q = \sum_{n=0}^{N-1} |\xi_n|^2 = \sum_{j=0}^{N-1} \frac{\nu^{2j} \|f\|^2}{12 n_j^2}. \quad (8.3.5)$$

¹The sign bits can be reported on the rotation index in the case of structured dictionary built on anti-symmetric functions.

The optimal quantization now minimizes the distortion for a bit rate R , or equivalently, minimizes the bit rate for a given distortion. In other words, we have to find the optimal parameters $n_j \geq 1$ and N that minimizes the distortion for a given rate. The Lagrangian multiplier method [442, 443, 290, 128, 135] is well suited for this kind of constrained optimization problems. It defines a cost function $\mathcal{L}(\lambda)$ as the sum of the objective distortion function and the constraint on the rate, weighted by the Lagrangian multiplier λ . In our case the cost function may be written as :

$$\begin{aligned} \mathcal{L}(\lambda) &= D + \lambda R \\ &= \sum_{j=0}^{N-1} \frac{\nu^{2j} \|f\|^2}{12 n_j^2} + \nu^{2N} \|f\|^2 + \lambda \left(\sum_{j=0}^{N-1} \log_2(n_j) + N B \right), \end{aligned} \quad (8.3.6)$$

where B represents the number of bits needed to code the atom index. The index size is assumed to be constant through the iterations, even if it is possible to perform an entropy coding of the index parameters [433]. The Lagrangian formulation allows to solve the hard constrained problem of finding the optimal set of n_j and N by converting it to a set of unconstrained problems, parameterized by λ .

The optimal quantization is obtained by differentiating $\mathcal{L}(\lambda)$ with respect to both n_j and N . First, solving

$$\frac{\partial \mathcal{L}(\lambda)}{\partial n_j} = 0, \forall j \quad (8.3.7)$$

for n_j positive and finite, the optimal quantization is given by

$$n_j = \sqrt{\frac{\|f\|^2 \nu^{2j} \log 2}{6 \lambda}}. \quad (8.3.8)$$

The solution of Eq. (8.3.8) is indeed a minimum of the Lagrangian since the second derivative is positive at this point, regardless of the value of j . Hence, the optimal quantization imposes an exponential law to the number of quantization levels :

$$\frac{n_{j+1}}{n_j} = \frac{\nu^{j+1}}{\nu^j} = (1 - \alpha^2 \beta^2)^{\frac{1}{2}}. \quad (8.3.9)$$

Interestingly, this bit distribution leads to an equivalent participation of each iteration to the total distortion. Indeed, the distortion per coefficient is equal to :

$$\frac{\lambda}{2 \log(2)}, \forall n_j > 1, \quad (8.3.10)$$

independently of the iteration. Notice however that the Lagrangian formulation provides only an approximation to the optimal quantization. Indeed, in the practical case, n_j only takes values that are integer powers of two.

The appropriate number of MP iterations has now to be found to minimize the reconstruction error for a given rate. The rate can indeed be limited by transmitting only part of the stream. Moreover, it can be seen from Eq. (8.3.6) that indexes can be transmitted alone, even if no coefficient is coded ($n_j \leq 1$). Depending on the redundancy of the dictionary, indexes often carry more information than coefficients. One can therefore imagine a scheme where the coefficients are simply interpolated from the exponential decay curve, especially for high order iterations (i.e., small coefficients). The optimal number of iterations N is thus given by minimizing the Lagrangian cost function of Eq. (8.3.6) where the n_j have been replaced by their optimal values from Eq. (8.3.8) :

$$\mathcal{L}(\lambda) = \sum_{j=0}^{N-1} \mathcal{L}_j(\lambda) + \nu^{2N} \|f\|^2 \quad (8.3.11)$$

with

$$\mathcal{L}_j(\lambda) = \begin{cases} \frac{\nu^{2j} \|f\|^2}{12 n_j^2} + \lambda (\log_2 n_j + B) & \text{if } n_j \geq 1 \\ \frac{\nu^{2j} \|f\|^2}{12} + \lambda B & \text{otherwise.} \end{cases} \quad (8.3.12)$$

The Lagrangian is defined as a piecewise function since indexes can be transmitted without coefficients, as stated before. The break-point occurs at $j = N_B = \lfloor \frac{-\log(n_0)}{\log(\nu)} \rfloor$ (i.e., $n_j = 1$) where $n_0 = \sqrt{\frac{\|f\|^2 \log(2)}{6 \lambda}}$ from Eq (8.3.8). Notice that N_B is positive only if $\lambda \leq \frac{\|f\|^2 \log(2)}{6}$. Otherwise, the weight on the rate in the Lagrangian cost function becomes much more important than the distortion, and the best scheme would be not to transmit any coefficient. The optimal number of iterations, N_{opt} , is therefore obtained by the zeros of the first derivative on both parts of the Lagrangian cost function. Thus,

$$N_{opt} = \begin{cases} N_1 & \text{if } B > -\frac{23 \log(\nu) + 1}{2 \log(2)} \\ N_2 & \text{otherwise.} \end{cases} \quad (8.3.13)$$

where :

$$N_1 = -\frac{1 + 2 B \log(2) - \log(\nu) + \log\left(\frac{\|f\|^2 \log(2)}{6 \lambda}\right)}{2 \log(\nu)} - \frac{W_{-1}\left(3 \cdot 2^{3-2B} \nu \log(\nu) e^{-1}\right)}{2 \log(\nu)}, \quad (8.3.14)$$

$$N_2 = \frac{\log\left(-\frac{\lambda B}{2 \|f\|^2 \log(\nu) \left(1 + \frac{1}{12 \nu^2 - 12}\right)}\right)}{2 \log(\nu)} \quad (8.3.15)$$

are the optimum values of N on both sides of the piecewise function (i.e., for respectively $N \leq N_B$ and $N > N_B$). The choice between N_1 and N_2 depends only on the relation between B and ν . Indeed, the atom index size and the decay rate of the coefficient norm influence the decision of coding an atom without coefficient. In Eq. (8.3.14), $W_{-1}(x)$ represents the second branch of the Lambert W function [444]. Finally, these solutions are indeed a minimum of the Lagrangian cost function, since the second derivative is positive at $N = N_{opt}$. Details of the computation are given in Appendix A.

Now that the unconstrained problem of Eq. (8.3.6) has been solved optimally for an arbitrary λ , the next step is to find the optimal λ_{opt} that guarantees a bit budget R_{budget} . The solutions of the optimization problem forms a convex hull of the achievable rate-distortion curve, and the bit budget constraint imposes a λ_{opt} that represents the slope of the rate-distortion function at $R = R_{budget}$. Several bisection methods have been proposed to solve this second problem [445, 443, 71]. Since in our case we can express the bit rate in terms of the Lagrangian multiplier λ , the constraint on the bit budget directly imposes an approximated value for λ_{opt} from :

$$R_{budget} = \begin{cases} \sum_{j=0}^{N_1-1} \log_2(n_j) + N_1 B & \text{if } B > -\frac{23 \log(\nu) + 1}{2 \log(2)} \\ \sum_{j=0}^{N_B-1} \log_2(n_j) + N_2 B & \text{otherwise.} \end{cases} \quad (8.3.16)$$

where N_1 , N_2 and n_j are functions of λ from the above equations. The approximation is due to the constraint of integer number of bits for each coefficient. The values of λ_{opt} can finally easily be computed through numerical methods.

8.3.4 Experimental results and discussion

Fig. 8.3 shows the average rate-distortion characteristic for a decomposition of ten random signals of 10 samples over a random dictionary of 50 vectors. The redundancy-driven quantization

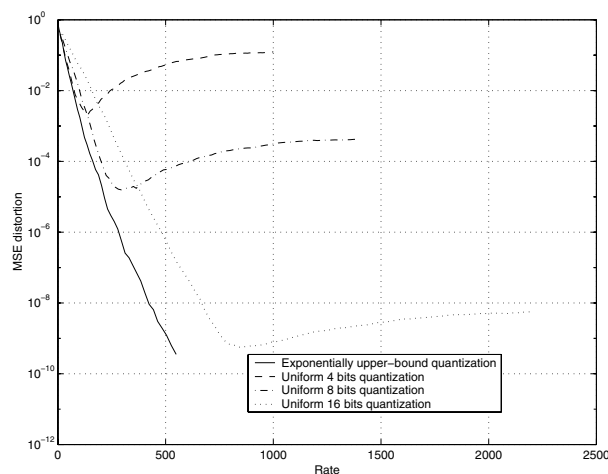


FIGURE 8.3: *R-D curve for the decomposition of a random signal of length 10 over a random dictionary of 50 vectors. Comparison of exponential quantization and uniform quantization as a function of the total rate.*

scheme is compared to different uniform mid-step quantization schemes. The latter quantize the coefficients over a range defined by the energy of the signal. The proposed exponentially upper-bounded quantization clearly outperforms the uniform quantization, since it adapts to the rate and to the range of coefficients to provide the best approximation for the available rate. Uniform quantization schemes provide good results for low rate, since the first coefficients are finely quantized. For high rates, however, the coefficients become too small compared to the quantization steps. The error therefore increases due to mid-step quantization, since the quantized coefficient is often larger than the true value computed by the MP.

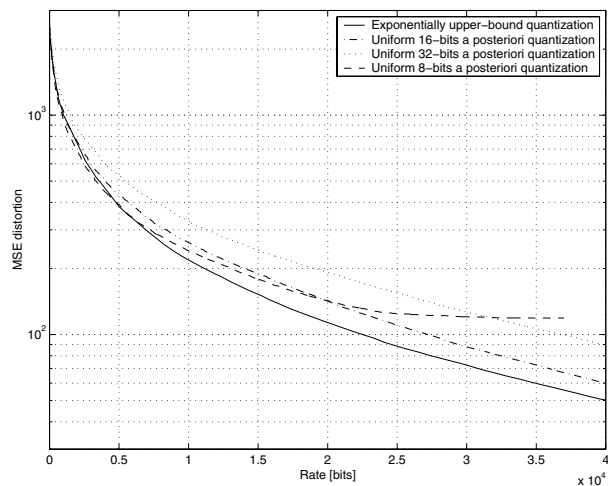


FIGURE 8.4: *Rate-distortion curves for the new quantization scheme and uniform a posteriori quantization of Lena MP stream. Distortion is represented as a function of the total rate.*

Fig. 8.4 shows the behavior of the redundancy-driven quantization scheme in the practical case of image coding. The *Lena* image has been decomposed onto a structured Gabor dictionary. The coefficients have then been quantized with respectively our scheme and common uniform quantization schemes. The distortion is smaller for the proposed quantization scheme. However, for very low rates, uniform quantization can be slightly better due to mismatch between the

coefficients decay rate and the values of the first coefficients (see Fig. 8.2). The mismatch is due to the sub-optimal search within the MP. The reconstructed image are represented in Fig. 8.5 to 8.7 for similar rates. The redundancy-driven quantization provides better reconstructed images compared to *a posteriori* and even *a priori* uniform quantization. The PSNR quality is slightly better, and the visual quality is also better since more components are used for the reconstruction, for the same transmission rate. Notice that the rates given in the previous figures represent only



FIGURE 8.5: *Reconstruction of the Lena image with Gabor atoms with exponentially upper-bounded quantization. PSNR = 26.7 dB, Rate = 16.4 kbit (without entropy coding).*



FIGURE 8.6: *Reconstruction of the Lena image with Gabor atoms with a posteriori coefficients uniform 16-bits quantization. PSNR = 25.8 dB, Rate = 16.4 kbit (without entropy coding).*



FIGURE 8.7: *Reconstruction of the Lena image with Gabor atoms with a priori coefficients uniform 16-bits quantization. PSNR = 25.7 dB, Rate = 16.4 kbit (without entropy coding).*

qualitative values, and can not be used to compare Matching Pursuit with other coding schemes. The efficiency of MP is described for example in [399, 402]. Indeed, in our application, the dictionary is not optimized. Bits can be saved on rotations, positions and scales for example without decreasing the convergence (as shown in the next Section). The bit rate can thus be reduced by 25-30 percent, before additional entropy coding.

Note that the exponentially upper-bounded uniform quantization is equivalent to the division of the coefficients by an exponential quantization table factor and multiplied by a quantizer scale factor which is given by the bit budget. The optimal number of iterations depends only on the bit budget and the design of the input dictionary. Practically speaking, the only parameters to transmit for the decoder to perform the inverse quantization are ν , $\|f\|$ and n_0 , or equivalently the quantizer scale factor.

The redundancy-driven quantization becomes particularly interesting for highly redundant dictionaries, where the values of the coefficients decrease very rapidly. In this case, the distribution of bits among coefficients is particularly efficient compared to uniform quantization. Similarly, for small index size the proposed scheme is also more efficient since additional elements can be sent at low price.

Finally, the redundancy-driven quantization could be improved by a periodic refreshing of the exponential parameters, or even by describing a piecewise upper-bound to fit data similar to the previous Gabor decomposition. Moreover, a limit to the maximum number of bits per coefficients can be set to ensure that the quantization of the first coefficient is not too fine.

8.4 Quantization of structured atoms indexes

8.4.1 A posteriori index quantization

In the context of asymmetric applications, index quantization of a structured dictionary may also limit the rate to transmit the MP stream. Recall however that *a posteriori* index quantization is not equivalent to a Matching Pursuit encoding with a sub-dictionary. The coder works in this case with a large dictionary, thus ensuring optimal convergence of the MP. The quality of the reconstruction is then sacrificed to respect some rate constraints on the transmission of the MP stream. Hence, the convergence of MP is preserved [446], but obviously the reconstruction suffers from an inexact representation, since the decoded atoms do not necessarily correspond to the atoms used within the encoding process.

The Matching Pursuit properties can help to design an appropriate index quantization scheme. First, successive atom indexes are not related to each other. Even worse, due to the Matching Pursuit encoding, successive atoms are close to orthogonal. However, it has been shown heuristically that atoms with large scale are generally encoded first, since they often capture more energy than small atoms. Second, one can take advantage of the fact that the energy captured by atoms decreases with the iteration number. The quantization of the atom indexes should therefore be finer for the first terms. Under the assumption that each Matching Pursuit iteration should participate equivalently to the reconstruction error, the error on the atom is allowed to increase exponentially with the iteration number. Indeed the error on the atom is multiplied by the respective coefficient as shown in Eq. (8.2.4), while the coefficient norm decreases exponentially. The evolution of the atom selection error should therefore satisfy:

$$\frac{\|E_{n+1}\|}{\|E_n\|} = \frac{|c_{\gamma_n}|}{|c_{\gamma_{n+1}}|} = (1 - \alpha^2 \beta^2)^{\frac{1}{2}}, \quad (8.4.1)$$

in the worst case where the Matching Pursuit coefficient norm decay follows exactly the exponential upper-bound.

The effects of atom selection error on the reconstruction are first investigated in this section, that is the error arising when the reconstruction formula (6.2.8) is biased :

$$\tilde{f} = \sum_{n=0}^{+\infty} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle g_{\tilde{\gamma}_n}. \quad (8.4.2)$$

The decoder receives in this case an approximation $\tilde{\gamma}$ of the indexes chosen by the encoder. The atom selection error is shown to be bounded in the general case. This error is then analyzed in detail for structured dictionaries, and more particularly for Gabor atoms. Finally, the performance of a simple scalar quantization of Gabor atom indexes is studied in terms of rate-distortion characteristics.

8.4.2 Bounds on the atom selection error

In this section, the atom selection error is shown to be bounded in the case of structured dictionaries. Let Δ_E be the error between the original signal and its reconstructed version with possible error on indexes:

$$\Delta_E = \|f - \tilde{f}_E\|^2. \quad (8.4.3)$$

The reconstruction error Δ_E can be bounded. The following result shows that the reconstructed signal is still in $L^2(\mathbb{R})$ provided that there exists a uniform bound on each error².

Theorem 8.4.1. *Let $f \in L^2(\mathbb{R})$ and let $\mathcal{D}$ be a complete dictionary of atoms. Let $\{g_{\gamma_n}\}$ be the set of atoms selected by a Matching Pursuit of f and $\{g_{\tilde{\gamma}_n}\}$ be the associated biased set of atoms. The reconstruction error Δ_E is bounded if there exists a constant C such that*

$$\|g_{\tilde{\gamma}_n} - g_{\gamma_n}\|^2 < C, \quad \text{uniformly in } n. \quad (8.4.4)$$

²This assumption is perfectly realistic in the settings of quantization or channel errors.

Proof : From Eq. (8.4.3), the reconstruction error Δ_E can be written as :

$$\Delta_E \leq \sum_{n=0}^{+\infty} |c_{\gamma_n}|^2 \|E_n\|^2. \quad (8.4.5)$$

Defining $S_n = |c_{\gamma_n}|^2$, Eq. (8.4.5) becomes

$$\Delta_E \leq \sum_{n=0}^{+\infty} S_n \|E_n\|. \quad (8.4.6)$$

Now we immediately have $S \in \ell^1$ because

$$\sum_{n=0}^{+\infty} S_n = \|f\|^2, \quad (8.4.7)$$

and, by Köthe duality [447], the sum on the right-hand side of equation (8.4.6) converges if and only if $G \in \ell^\infty$, that is if and only if there exists $C > 0$ such that $\|E_n\| < C$, uniformly in n . Finally, the total error is thus bounded by :

$$\Delta_E \leq C \|f\|^2 \sup_n \|E_n\|. \quad (8.4.8)$$

■

Let us now focus on the last term in this equation. Suppose our dictionary has been built using the technique described in Section 6.3.2. By definition of E_n , we then have

$$E_n = \mathcal{U}(\gamma_n)g - \mathcal{U}(\tilde{\gamma}_n)g, \quad (8.4.9)$$

which can be computed as :

$$\|E_n\|^2 = \langle \mathcal{U}(\gamma_n)g - \mathcal{U}(\tilde{\gamma}_n)g | \mathcal{U}(\gamma_n)g - \mathcal{U}(\tilde{\gamma}_n)g \rangle \quad (8.4.10)$$

$$= 2(1 - \Re\{\langle \mathcal{U}(\gamma_n)g | \mathcal{U}(\tilde{\gamma}_n)g \rangle\}) \quad (8.4.11)$$

$$= 2(1 - \Re\{\mathcal{K}_g[\gamma_n; \tilde{\gamma}_n]\}), \quad (8.4.12)$$

where $\mathcal{K}_g$ stands for the reproducing kernel associated to the dictionary and $\Re\{x\}$ represents the real part of x . Using the unitarity of the representation $\mathcal{U}$, we can also write

$$\|E_n\|^2 = 2(1 - \Re\{\mathcal{K}_g[\tilde{\gamma}_n; \gamma_n]\}). \quad (8.4.13)$$

Eq. (8.4.13) shows that the kernel $\mathcal{K}$ endows the elements of the dictionary $\mathcal{D}$ with a notion of distance. Indeed, when $\tilde{\gamma}_n = \gamma_n$, the error vanishes, $\Delta = 0$. In the next section, the atom selection error is computed in the practical case of Gabor atoms.

8.4.3 Gabor atom selection error

Following Eq. (8.4.13), we now give explicit estimates of the error on the atoms in the settings of a Gabor dictionary [426]:

$$h(\vec{x}) = \frac{1}{2\pi\sigma^2} \cos(\vec{\omega}_0 \cdot \vec{x}) e^{-\frac{\|\vec{x}\|^2}{4\sigma^2}}, \quad (8.4.14)$$

where $\vec{\omega}_0$, generally equal to $(\omega_0, 0)$ with ω_0 even, is the oscillation parameter and σ is a scale parameter. The MP encoder transmits in this case translation, scaling and rotation parameters applied to the elementary Gabor function. The atom selection error corresponds therefore to the error arising from decoding the atoms with distorted translation, scaling and rotation indexes.

The formulation of Eq. (8.4.14) corresponds to the real part of the complex function:

$$\psi(\vec{x}) = \frac{1}{4\pi\sigma^2} e^{i(\omega_0 \cdot \vec{x})} e^{-\frac{\|\vec{x}\|^2}{4\sigma^2}}. \quad (8.4.15)$$

Making use of this complex form of the Gabor atom we can write :

$$g(\vec{x}) = \frac{\psi(\vec{x}) + \overline{\psi(\vec{x})}}{\|h(\vec{x})\|} \quad (8.4.16)$$

The atom selection error is given by re-writing the inner product :

$$\begin{aligned} \langle g_\gamma, g_{\tilde{\gamma}} \rangle &= \frac{\langle \psi_\gamma, \psi_{\tilde{\gamma}} \rangle + \overline{\langle \psi_\gamma, \psi_{\tilde{\gamma}} \rangle} + \langle \psi_\gamma, \overline{\psi_{\tilde{\gamma}}} \rangle + \overline{\langle \psi_\gamma, \overline{\psi_{\tilde{\gamma}}} \rangle}}{\|h(\vec{x})\|^2} \\ &= \frac{2(\Re \langle \psi_\gamma, \psi_{\tilde{\gamma}} \rangle + \Re \langle \psi_\gamma, \overline{\psi_{\tilde{\gamma}}} \rangle)}{\|h(\vec{x})\|^2}, \end{aligned} \quad (8.4.17)$$

since $\|h(\vec{x})\| = \|\tilde{h}(\vec{x})\|$ thanks to the properties of the representation $\mathcal{U}$. Moreover, the norm of h depends only on the oscillation and scale parameters (ω_0, σ) and is given by :

$$\|h(\vec{x})\|^2 = \frac{1 + e^{-2\sigma^2\omega_0^2}}{4\pi\sigma^2}. \quad (8.4.18)$$

Assume now that a , θ , b_1 and b_2 respectively represents the scale, rotation and positions parts of the error on the index $\tilde{\gamma}$. The index γ is given in a 2-dimensional Gabor dictionary by the quadruplet $(\vec{p}, s, \varphi)$ which respectively correspond to the position, scale and rotation parameters of the basic Gabor function. Thanks to the group properties of the representation $\mathcal{U}$, the index error is therefore written as [448]:

$$\begin{aligned} (\vec{b}, a, \theta) &= \tilde{\gamma}^{-1} \circ \gamma \\ &= \left(\tilde{s}^{-1} r_{-\tilde{\varphi}}(\vec{p} - \vec{p}), \frac{s}{\tilde{s}}, \varphi - \tilde{\varphi} \right). \end{aligned} \quad (8.4.19)$$

The inner product between two complex Gabor atoms given by the parameters σ and ω_0 can then be expressed as :

$$\langle \psi_\gamma, \psi_{\tilde{\gamma}} \rangle = \frac{a}{4\pi\sigma^2\alpha^2} e^{-\sigma^2\Gamma^2} e^{-\frac{b_1^2+b_2^2}{4\sigma^2\alpha^2}} e^{-i\frac{\beta_1 b_1 + \beta_2 b_2}{\alpha^2}}, \quad (8.4.20)$$

where

- $\alpha = \sqrt{a^2 + 1}$
- $\Gamma^2 = 2\omega_0^2 - \frac{\beta_1^2 + \beta_2^2}{a^2 + 1}$
- $(\kappa_0, \eta_0) = (\omega_0 \cos \theta, \omega_0 \sin \theta)$
- $\beta_1 = a\kappa_0 + \omega_0$, $\beta_2 = a\eta_0$.

Using the properties of complex numbers, the inner product can be rewritten as :

$$\begin{aligned} \langle g_\gamma, g_{\tilde{\gamma}} \rangle &= \frac{a e^{-\frac{b_1^2+b_2^2}{4\sigma^2\alpha^2}}}{2\pi \|h(\vec{x})\|^2 \sigma^2 \alpha^2} \left[e^{-\sigma^2\Gamma^2} \cos \frac{\beta_1 b_1 + \beta_2 b_2}{\alpha^2} \right. \\ &\quad \left. + e^{-\sigma^2\hat{\Gamma}^2} \cos \frac{\hat{\beta}_1 b_1 + \hat{\beta}_2 b_2}{\alpha^2} \right], \end{aligned} \quad (8.4.21)$$

where

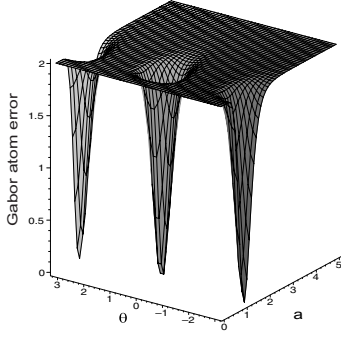


FIGURE 8.8: *Gabor atom selection error as a function of a and θ , respectively the scale and rotation part of the error on indexes.*

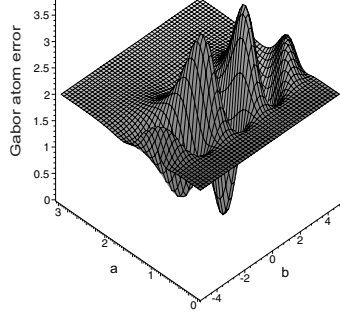


FIGURE 8.9: *Gabor atom selection error as a function of a and b , respectively the scale and position part of the error on indexes.*

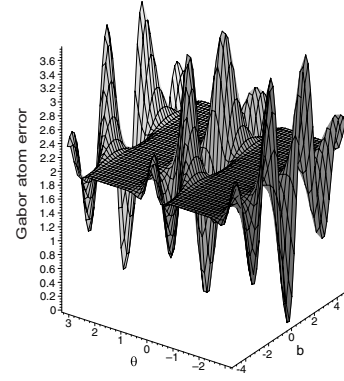


FIGURE 8.10: *Gabor atom selection error as a function of b and θ , respectively the position and rotation part of the error on indexes.*

- $(\hat{\kappa}_0, \hat{\eta}_0) = (\omega_0 \cos(\theta + \pi), \omega_0 \sin(\theta + \pi))$
- $\hat{\Gamma}^2 = 2\omega_0^2 - \frac{\hat{\beta}_1^2 + \hat{\beta}_2^2}{a^2 + 1}$
- $\hat{\beta}_1 = a\hat{\kappa}_0 + \omega_0, \hat{\beta}_2 = a\hat{\eta}_0.$

Finally, from Eq. (8.4.13), the atom selection error is expressed as

$$\|E_n\|^2 = 2 (1 - \langle g_\gamma, g_{\bar{\gamma}} \rangle). \quad (8.4.22)$$

The error on Gabor atoms is therefore clearly bounded since its value can not be larger than 4. Indeed, since atoms have a unit norm by definition, the error arising from a completely wrong atom selection can not grow indefinitely. The error forms moreover a convex function along the axis representing the error on the scale parameter. The minimum of the error is reached when $a = 1$ (i.e., when the error on the scale is null), and then increases when a goes both towards zero or infinity. Due to Gabor function properties, the error is clearly periodic with the error onto the rotation parameter.

Finally, it can be seen also from Eq. (8.4.19) that, for a given reconstruction error, the allowable error on position increases with the scale. This can be explained by the decrease of the relative importance of the position of large scale atoms. The evolution of the reconstruction error as a function of the error on the position index depends on both the scale and the rotation indexes (see Eq. (8.4.19)). It is therefore not straightforward to handle it, since it does form neither a convex nor a periodic function. Finally, Gabor atoms are particularly sensitive to position error due to the oscillations within the elementary function.

Fig. 8.8 to 8.10 shows the variation of the atom selection error as a function of the position, scale and rotation components of the error on the Gabor atom indexes. As expected from the last section, the reconstruction error is clearly bounded. It can be seen that the reconstruction error is a convex function of a , the error on the scale. Moreover, due to the modulation in the Gabor atoms, oscillations are also present along the axis of θ , the error on the rotation index. As expected, the reconstruction error is minimal when the error on the rotation parameter correspond to multiples of π , due to the symmetry in Gabor atoms.



FIGURE 8.11: *Reconstruction of the Lena image with Gabor atoms after 1000 MP iterations (PSNR = 31.17 dB).*



FIGURE 8.12: *Reconstruction of the Lena image with Gabor atoms, where all positions indexes have been quantized (PSNR = 22.35 dB). The number of positions indexes has been divided by two compared to the encoding dictionary.*



FIGURE 8.13: *Reconstruction of the Lena image with Gabor atoms, where all rotation indexes have been quantized (PSNR = 27.72 dB). The number of rotation indexes has been divided by two compared to the encoding dictionary.*

8.4.4 Scalar quantization of structured atom indexes

To avoid a complex decoding we restrain the atom quantization to scalar quantization of structured atom indexes. Atom index quantization is a typical candidate to apply Vector Quantization [449], where bits can be saved on the dictionary size while guaranteeing a uniform distribution of the atoms in the sub-dictionary. Vector Quantization is therefore expected to generate less degradation than scalar quantization of atom indexes for the same rate reduction. However, the quantization and inverse quantization are more complex. Indeed, to avoid the transmission of the complete dictionaries, the decoder should be able to build uniformly distributed dictionaries to perform the inverse quantization.

In the case of structured or semi-structured dictionaries (see Sec. 6.3.2 and 6.3.3) the parameters of the structured dictionary can be *a posteriori* quantized to cope with the decoding complexity problems. In other words, the atom index i is quantized as

$$i_q = i_0 + k \frac{i_N - i_0}{N} = i_0 + k \delta_i, \quad (8.4.23)$$

where i_0 and i_N are the bounds on the index value, N is the number of quantization steps and k is the transmitted bin ($k \in [0, N]$). A scalar quantization is performed separately on the position, scale and rotation parameters.

Figures 8.12 and 8.13 show the effect of *a posteriori* quantization of respectively the position and rotation indexes in a Gabor dictionary. Even though the relative error is not exactly the same in both cases, it is clear that the atom selection error is much more sensitive to position indexes, as explained in the previous section. Indexes quantization can generate annoying distortion effects if not handled carefully, as seen in the same figures.

Fig. 8.14 shows the rate distortion curves for uniform *a posteriori* quantization of Gabor rotation index. Compared to the index length, reduction of rate due to rotation quantization appears very small by opposition to the distortion increase. However, the initial dictionary is not very redundant, and the relative error on the rotation indexes becomes rapidly large. It can be seen that the distortion increase is smaller for a fine quantization (7 bits instead of 8 bits in the initial dictionary) than for coarse quantization. The quantization distortion increase follows approximately a logarithmic function of the number of spared bits. Similar results are obtained for the

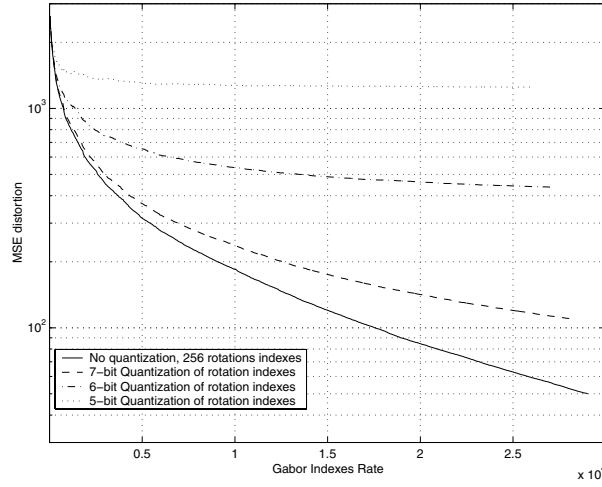


FIGURE 8.14: Index rate-distortion characteristic of uniform *a posteriori* rotation index quantization.

quantization of scale and position indexes (see Fig. 8.15), although the latter is less straightforward since the error is also dependent on the scale of the atoms. In this case, the error onto the first atoms is compensated by their large scale, as can be seen in Eq. (8.4.19). For higher order iteration however the error onto the position indexes is additionally increased by the small scale of the atoms.

Index scalar quantization may provide flexibility in the design of asymmetric applications, but obviously degrades the picture quality compared to a MP with the quantized sub-dictionary. Fig. 8.16 shows the reconstruction error as a function of the iteration number of Matching Pursuit, for *a posteriori* quantization of rotation and scale indexes, and for an encoding over the quantized sub-dictionaries. It clearly shows that *a posteriori* quantization, for the same dictionary size, causes more distortion than a MP encoding on the quantized dictionary. For the first iterations, the two schemes are very similar, since generally first atoms are very large and therefore less sensitive to errors. For the higher iterations however, the reconstruction error clearly converges more slowly in the case of *a posteriori* quantization.

Clearly there is a trade-off between convergence of the MP encoding (i.e., the size or redundancy of the dictionary) and the coding rate of atom indexes. We have seen that Gabor atoms are really sensitive to index *a posteriori* quantization. However, this sensitivity is driven by the error on the indexes. For the same rate reduction, this error decreases when the size of the dictionary increases. Indeed, saving one bit in a dictionary of R rotation indexes generates more distortion than in a double-sized dictionary, for example.

To illustrate the trade-off between the size of the initial dictionary and *a posteriori* index quantization, Figures 8.17 and 8.18 represents the reconstruction distortion in the case of the *Lena* 128×128 image with 1000 Gabor atoms as a function of the coding size of the quantized atom index. It is clear that, for the same index size, the index quantization is worse than an encoding on a smaller dictionary. In this case the approximation converges slower than for large dictionaries, but there is no error propagation, in contrary to the case of the *a posteriori* quantization. For very large dictionaries, the quantization error is small for fine quantization (i.e. gain of one or two bits per index). However, it can be seen from the Figure 8.17 that the convergence of the reconstruction error with the dictionary size saturates relatively early, for relatively small dictionaries. Hence encoding dictionary has clearly to satisfy a trade-off between its size to limit the transmission rate, and its properties of convergence. The need for a careful design of the encoding dictionary is illustrated in the Figure 8.19. It can be seen not only that the convergence speed becomes slower when the number of allowable index values decreases, but also that the indexes have a different influence for an similar size reduction. In sight of the previous results, only very fine scalar index

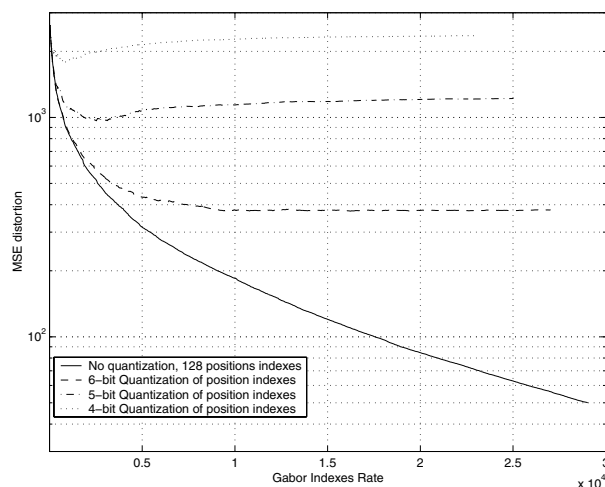


FIGURE 8.15: *Index rate-distortion characteristic of uniform a posteriori position index quantization.*

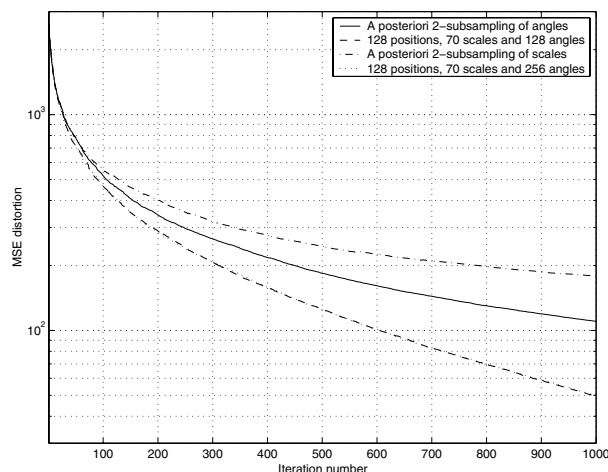


FIGURE 8.16: *Evolution of the reconstruction error of Lena for Gabor sub-dictionaries and index a posteriori quantization. Curves taken by pairs have the same dictionary size.*

quantization may be allowed, under penalty of very large reconstruction distortion.

Finally, the results presented here above are only valid for Gabor atoms, which are particularly sensitive to position error due to oscillations. A similar development can however be done for any structured dictionary as defined in Sec. 6.3.2. Anisotropic dictionaries, for example, are expected to be less sensitive to index quantization due to their inherent properties.

8.4.5 Redundancy-driven Gabor indexes quantization

The main design features of the scalar index quantization have been drawn in the previous paragraphs. However, the detailed scheme is closely related to the chosen dictionary. The problem becomes now to allocate bits between the different indexes (e.g., positions, scales, angle) so that the error on the atoms given by Eq. (8.4.1) is respected. In other words, the number of quantization steps of each index has to be properly determined, while respecting the complexity constraints at decoder. Finally, the optimal bit distribution, in a rate-distortion sense, between coefficients and the different atom indexes is obtained by a Lagrangian formulation similar to Eq. (8.3.6). In

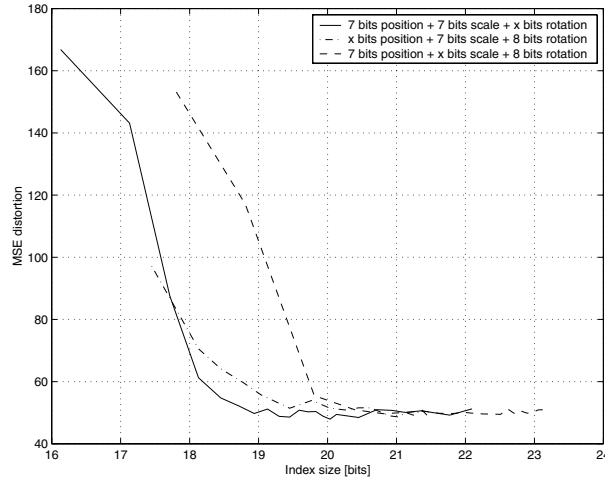


FIGURE 8.17: Rate-distortion representation between size of the Gabor dictionary index and MSE distortion after 1000 iterations on the Lena 128×128 image.

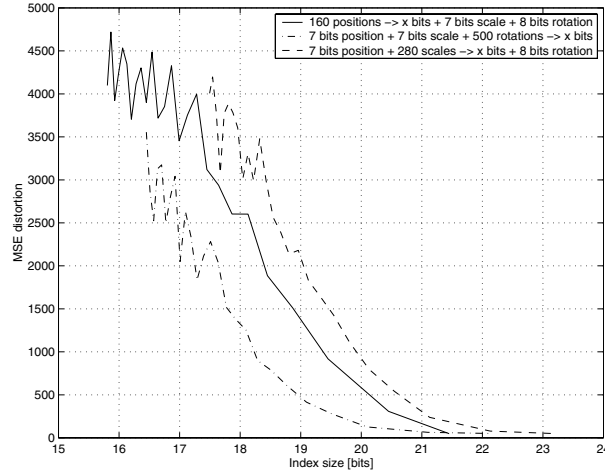


FIGURE 8.18: Rate-distortion representation between size of the Gabor index after quantization and MSE distortion after 1000 iterations on the Lena 128×128 image.

particular, the upper-bound on the reconstruction error given by Eq. (8.2.4) should be minimized for a given encoding rate. One has to minimize the following general Lagrangian formulation :

$$\begin{aligned}
 \mathcal{L}(\lambda) &= D + \lambda R \\
 &= \sum_{j=0}^{N-1} \nu^{2j} \|f\|^2 \left(\frac{1}{12 n_j^2} + \|E_j(p_j, s_j, r_j)\|^2 \right) + \nu^{2N} \|f\|^2 \\
 &\quad + \lambda \sum_{j=0}^{N-1} (\log_2(n_j) + \log_2(p_j) + \log_2(s_j) + \log_2(r_j)) ,
 \end{aligned} \tag{8.4.24}$$

where n_j , p_j , s_j and r_j represents the number of quantization steps, at iteration j , for respectively the coefficient and the position, scale and rotation indexes. The solution of the Lagrangian formulation is obtained by differentiating Eq. (8.4.24) with respect to n_j , p_j , s_j , r_j and N , as presented in the previous section.

Let us now focus on the scalar quantization of Gabor atom indexes. For simplicity reasons, let

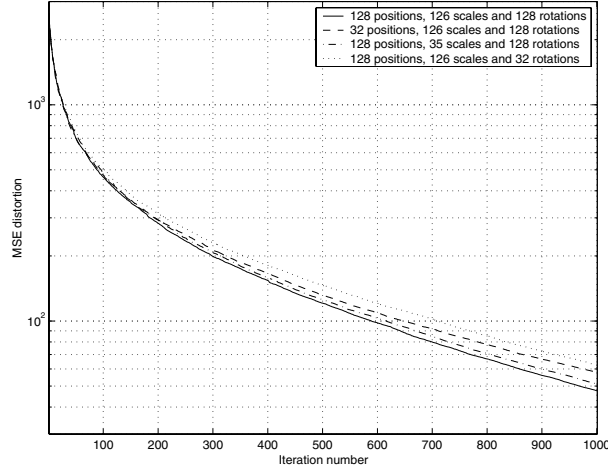


FIGURE 8.19: *Convergence of the reconstruction error of the Lena image for different Gabor dictionaries of equivalent coding size.*

assume moreover that no error is allowed on the position indexes (i.e., $b_1 = b_2 = 0$ in Eq. 8.4.22). The reconstruction error is therefore convex around $a = 0$ and $\theta = 0$. Moreover, let us assume a logarithmic distribution of scale indexes, since accuracy of scale indexes should be finer for small atoms. The scale index s_n is therefore given by :

$$s_n = s_0 \left(\frac{S}{s_0} \right)^{\frac{n}{K}} = s_0 \delta_s^{\frac{n}{K}}, \quad (8.4.25)$$

where s_0 is the smallest scale, S is the largest scale (generally the size of the image) and N is the number of allowed scales within these two bounds, or equivalently the number of quantization steps. The maximal quantization error onto the scale index is therefore written as :

$$2^{-\frac{S}{2 \cdot K}} \leq a \leq 2^{\frac{S}{2 \cdot K}}, \quad (8.4.26)$$

when scales are given by powers of two. The rotation parameter is uniformly quantized between 0 and 2π :

$$\varphi_m = m \frac{2 \pi}{L} = m \delta_\varphi. \quad (8.4.27)$$

The redundancy-driven quantization problem is now equivalent to find the optimal step size for each indexes (i.e., δ_s , and δ_φ), so that the reconstruction error is minimized for a given rate. The solution of this optimization problem is again given by a Lagrangian formulation of the form :

$$\begin{aligned} \mathcal{L}(\lambda) &= D + \lambda R \\ &= \sum_{j=0}^{N-1} \nu^{2j} \|f\|^2 \left(\frac{1}{12 n_j^2} + \|E_j(2^{\frac{s}{2 \cdot K_j}}, \frac{\pi}{L_j})\|^2 \right) \\ &\quad + \nu^{2N} \|f\|^2 + \lambda \sum_{j=0}^{N-1} (\log_2(n_j) + \log_2(K_j) + \log_2(L_j)), \end{aligned} \quad (8.4.28)$$

where the atom selection error $\|E_n\|$ is given by Eq. (8.4.22). The evolution of the coefficient step-size given by n_j has already been computed in the previous section. The optimal index quantization is given by the minimum of the Lagrangian, obtained by differentiating with respect to K_j and L_j . It can be shown that the minimum of the Lagrangian formulation is given for

both K_j and L_j as large as possible, or a quantization error as small as possible, independently of the iteration. Unless the constraint on the rate is very high (i.e., for large λ values), the index quantization has to be really fine and even avoided. When the constraint on the rate is very high, or the coefficient energy is very small, the iterations have to be skipped to reach to optimal distortion under a given rate constraint. The optimal scheme is therefore directly given by the exponentially upper-bounded coefficient quantization presented in the Sec. 8.3, where the number of iterations to be transmitted has been computed.

The interesting behavior of the Lagrangian is due to the very rapid increase of the reconstruction error in comparison to the decrease of coding rate while the quantization step size increases. The atom selection error is moreover bounded, so that coarse quantization do not increase the distortion anymore when the upper-bound has been reached. Similar results were presented in the previous section, where Gabor atoms were shown to be very sensitive to index quantization. The quantization of Gabor atoms must therefore be avoided, and the encoding dictionary should be carefully chosen, even dynamically, so that coding rate is kept minimal.

Furthermore, a careful position index quantization can however be worthy if quantization is based on the value of the scale index. Indeed, for large scale indexes, the position can be coarsely quantized without introducing a large reconstruction error (see Eq. (8.4.21)), and thus saving bits onto both position indexes.

The optimal index quantization scheme, in R-D sense, is certainly a Vector Quantization satisfying Eq.(8.4.1). Another less complex, but suboptimal, quantization scheme could however be proposed, taking benefit of MP properties. Indeed, as generally larger scale atoms are encoded first, one can think about a hierarchical encoding scheme, where the dictionary is updated periodically. The size of the dictionary is thus kept small for the first iterations, and then grows with the iteration number. Bits can thus be saved to code atom indexes, since the resolution of the dictionary evolves during the encoding. The scheme provides an efficient way to reduce the coding rate, while ensuring a good convergence of the MP and avoiding *a posteriori* index quantization.

Finally, the adaptivity to the rate constraints imposed by the decoder has more to be seen in terms of number of transmitted iterations, besides coefficients quantization. Indeed, as we have seen, the scalar index quantization must be avoided, since it generally induces too important degradation. However, the encoding dictionary has to be carefully designed to avoid the need of index quantization. Note that the results presented in the last section are valid only for index scalar quantization, and more particularly for Gabor atoms. The conclusions however are expected to be the same for other dictionaries, since quantization errors behaves similarly.

8.5 Conclusions

A new approach of Matching Pursuit coding has been proposed in this chapter. It describes an *a posteriori* quantization of the MP stream to meet the constraints of asymmetric applications targeting an heterogeneous set of decoders. The encoder takes advantage of the intrinsic properties of the MP stream, and more particularly of the exponential decay of the coefficients. The exponential decay has been shown to depend only on the dictionary structure. Depending on the parameters of this exponential upper-bound, a redundancy-driven coefficient quantization scheme has been proposed. This scheme clearly outperforms uniform quantization schemes since it adapts to the coefficient decay to provide the best possible approximation with the lowest coding rate.

Similarly to the coefficient quantization, an simple index quantization scheme was proposed to reduce the size of atom index. However, for sake of simplicity, the quantization was assumed to be uniform and independent for each atom indexes. Gabor atoms have been shown to be particularly sensitive to index quantization, and do not allow for a simple index quantization scheme. However, a index quantization scheme based on Vector Quantization, and thus ensuring a uniform error distribution, is expected to give much better results than scalar index quantization. The decoding complexity becomes in this case too important for the stream to be decoded on any device in the context of heterogeneous receivers. Similarly, a non uniform scalar quantization, taking part of perceptual characteristics for example, would surely improve the signal reconstruction by

distributing bits where they are most needed.

Entropy coding can be performed on top of the quantization to reduce coding rate [433]. One can also think about an entropy-constrained quantization, where the entropy is considered in the design of the quantization scheme. The design of optimal dictionaries was not considered in this chapter, although its importance has been emphasized. The structure and the size of the dictionary is one of the key point of Matching Pursuit applications. Note finally that, even if only 2D quantization examples were given in this chapter, the same coefficient quantization scheme can be directly applied to any Matching Pursuit stream and more particularly to video coding as presented in the last chapter.

Multi-Resolution Lossy MP Delivery

9.1 Introduction

New classes of pervasive computing devices such as personal digital assistants (PDAs), hand-held computers (HHC) and smart phones allow users more ubiquitous access to information than ever. As users are beginning to rely more heavily on pervasive computing devices, there is a growing need for applications to bring multimedia information to these devices. However, due to heterogeneous device capabilities (in terms of display size, storage, processing power, access bandwidth and reliability) there are new challenges for designing appropriate tools that allow these devices to effectively access, store and process multimedia and more particularly video information. Concurrent with the developments in pervasive computing, advances in storage, networking and authoring tools are driving the production of large amounts of rich multimedia content. The result is a growing mismatch between the available rich content and the capabilities of the client devices to access and process it.

Many researchers have been working on adapting multimedia content in so-called transcoding proxies [324, 343]. Transcoding proxies are generally located in access providers. Video transcoding includes spatial scaling (image size changes), temporal scaling (frame rate modifications) and rate reduction [450, 336, 338, 314, 306]. In a general sense, it may also include changes in coding format [332, 334] or object-based transcoding [342]. A good transcoding algorithm maximizes the trade-off between the quality of the signal after manipulation and related processing power. Processing power is indeed an issue in order for an access provider to scale to the demand, which is the reason why only compressed domain transcoding is viable. Common multimedia compression standards create bit streams that are delicate to manipulate in the compressed domain. Due to block-based and motion estimation coding, transcoding is very difficult without expensive decoding/recoding operations. In contrary, wavelet-based coding may offer easy transcoding possibilities due to the subband structure. Signal manipulations in the compressed domain are however limited to octave-based resolutions. In terms of rate scalability, wavelet schemes provide very fine granularity thanks to embedded coding through SPIHT [28] or EZW [27] algorithms. In this context, Matching Pursuit presents the advantage of being inherently progressive. The stream can be truncated anywhere while guaranteeing that the most energetic components are still present in the remaining stream. In the same time, Matching Pursuit streams built on structured or semi-structured dictionaries are intrinsically scalable, thus enabling a very simple spatio-temporal transcoding.

In parallel, channel reliability and loss process may also be very different from a receiver to another. The error control scheme has therefore also to be scalable and simple enough to be performed in the network nodes. Proxy-based error control, operated closely to the end-user, is indeed more efficient than an end-to-end error control in terms of both performance and bandwidth usage. Unfortunately, the combination of proxy-based transcoding and error control

is very expensive in terms of computational power for standard-based video streams. Indeed, the standard coders produce packets with different importance due to the hierarchical structure of the video stream. In contrary, Matching Pursuit permits for an efficient packetization thanks to the independence between iterations. Moreover, the coding size of independent iterations is a lot smaller than the independent entities (e.g., GOP) in standard based coding. They can therefore easily be encapsulated in different packets. Each iterations can thus be decoded independently, and order is completely irrelevant. This flexibility allows an independent and energy-equivalent packetization, hence an efficient error control between the gateways.

This chapter presents first the inherent multi-resolution features of Matching Pursuit streams, which allow for an efficient transcoding in the gateways. MP transcoding is shown to outperform classical schemes in terms of quality, complexity and flexibility. To cope with packet losses, an Unequal Error Protection scheme is proposed due to the structure of the MP stream. An optimal error control in a heterogeneous framework is however shown to be complex in this case. A new energy-equivalent MP stream packetization is then proposed as the basis for simple and efficient error control. Issues in heterogeneous and lossy Matching Pursuit Delivery are finally discussed.

9.2 Matching Pursuit transcoding

9.2.1 Heterogeneous Matching Pursuit delivery

Heterogeneous Matching Pursuit delivery is represented in Fig. 9.1. A sequence of spatial resolution $W \times H$ and of frame rate F is distributed through video gateways to an heterogeneous set of receivers, at possibly different spatio-temporal resolutions. Resolution adaptation is performed within the gateways for two main reasons. First, resolution adaptation in the receivers would be computationally too expensive for most low end devices. Second, transcoding within gateways permit to save bandwidth in the downlink connections, once the proxy is aware of the receiver capabilities.

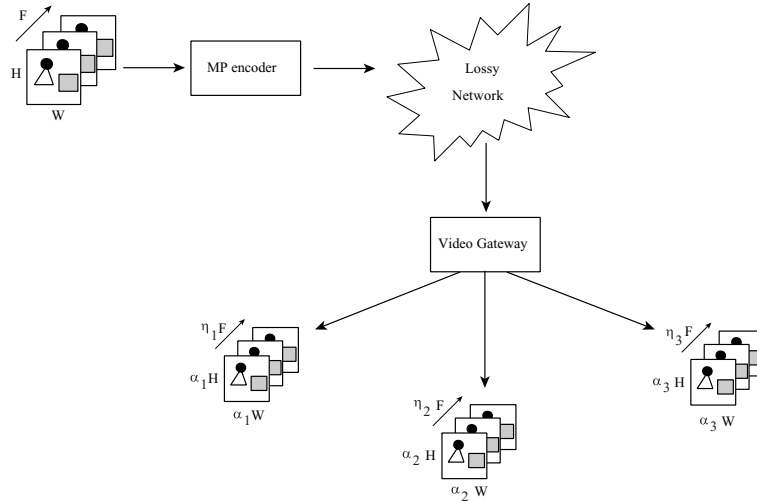


FIGURE 9.1: *Multiresolution Streaming Application. The input stream may be delivered, through video gateways, to different resolutions devices.*

Matching Pursuit transcoding may be performed in several ways. The MP stream can be adapted to the available rate, to the spatio-temporal resolution of the receiver, to the color depth of

the receiver display or to the computational power of the decoder by approximating the dictionary by simpler functions. The two latter methods are not treated in this work, although they are expected to lead to non negligible improvements. As explained in the Chap. 8, rate can be adapted by varying the number of MP iterations. Finally, the spatio-temporal resolutions can be adapted through simple transcoding thanks to the inherent properties of Matching Pursuit with structured or semi-structured dictionaries.

9.2.2 Transcoding of MP streams

Structured dictionaries (see Sec. 6.3) are basically built by applying meaningful transformations to a generating function. These transformations include rotation, translation and dilation. Generally, these transformations combine into a group G . In other words, using a unitary and irreducible representation $\mathcal{U}$ of G in our Hilbert space of signals, they are built by computing the orbit of any fixed atom [408] :

$$\mathcal{D} = \{\mathcal{U}(\gamma)g, g \in L^2(\mathbb{R}^2 \times \mathbb{R}), \gamma \in G\}. \quad (9.2.1)$$

In this context the law of the similitude group $SIM(2) \times SIM(1)$ applies [408], allowing for covariance of spatial rotation and spatio-temporal scaling and translation. The same group law applies also for the semi-structured dictionaries defined in Chap. 7 when both spatial scaling parameters coincide. One of the main advantages of this property is that it guarantees covariance of the Matching Pursuit reconstruction formula Eq. (6.2.8) with respect to the action of G :

$$\mathcal{U}(\gamma)f = \sum_{n=0}^{+\infty} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle g_{\gamma \circ \gamma_n}. \quad (9.2.2)$$

This property enables an efficient transcoding of MP stream, by appropriate modifications of the indexes within gateways. Notice that the validity of this property does not depend on the size of the dictionary. Dictionary can be formed with one-, two- or three-dimensional functions, as long as the group law holds. We will now concentrate on the latter in the context of video coding.

The transcoder can easily modify the spatio-temporal resolutions of the video stream by acting on the atom indexes sent by the Matching Pursuit coder. The images may be resized by a factor α by multiplying spatial positions and scales by α . The frame format (i.e., the ratio between picture width and height) has to be respected in the resizing process. Both spatial resizing parameters shall be identical to guarantee the respect of the similitude group law. This limitation is however not really restrictive since applications that need deformation of the video content are quite rare.

Similarly, the temporal resolution can also be modified very easily. The frame rate F can be changed into a new frame rate ηF by multiplying temporal position and scale indexes by the same factor η . Hence, the coded video signal f of size $W \times H$ and frame rate F can be transcoded into a video signal $\tilde{f}$ of spatial resolution $\alpha W \times \alpha H$ and frame rate ηF by :

$$\tilde{f} = \alpha \sqrt{\eta} \sum_{i=0}^N c_{\gamma_i} \widetilde{g_{\gamma_i}}, \quad (9.2.3)$$

where c_{γ_i} are the Matching Pursuit coefficients (see Eq. (6.2.3)) and $\widetilde{g_{\gamma_i}}$ corresponds to the atom g_{γ_i} after transcoding. In other words the index

$$(p_x, p_y, p_z, s_x, s_y, s_z)$$

becomes

$$(\alpha p_x, \alpha p_y, \eta p_z, \alpha s_x, \alpha s_y, \eta s_z)$$

after spatio-temporal resizing. Due to resizing, coefficients have finally to be multiplied by $\alpha \sqrt{\eta}$ since they have been initially computed with original sized atoms.

The resizing parameters α and η can take any real value. Values larger than one correspond to signal expansions, while values smaller than one correspond to lower spatio-temporal resolutions. Only the latter case is generally used in transcoding operations, where full resolution video have to be transmitted to devices with smaller display capabilities. Matching Pursuit streams transcoding offers much more flexibility than transcoding wavelets or standard-based streams. Compressed domain transcoding of wavelets stream is limited to octave multiples of the initial resolutions. Block-based coding still imposes more restrictive constraints. Similarly, since the only practical constraints on the value of η is imposed by GOF limits, Matching Pursuit transcoding offers a wider range of target frame rates than hierarchical scheme based on motion estimation.

To avoid prohibitive index coding costs, indexes values have however to be quantized or rounded to the nearest integer after transcoding. Indeed, even if coded indexes only take integer values, multiplication by resizing parameters does not guarantee any integer index value after transcoding. This approximation may introduce some distortion which remains however very small (see Sec. 8.4). The spatio-temporal resizing method may also be applied directly at receiver, where indexes can then take exact values. Similar pixel-based methods of image resizing by decompositions onto spline functions have also been proposed in [451, 452]. However benefits of proxy-based processing are lost in this case and bandwidth is wasted on the downlink connections.

9.2.3 Experimental results

Transcoding has been performed on a Matching Pursuit stream of 3D-splines atoms. The CIF format *Foreman* sequence has been encoded with 500 atoms whose indexes have been modified according to the Eq. (9.2.3). The indexes values are then rounded to the nearest integer to allow a reasonable coding cost after resizing. Fig. 9.2 shows the reconstruction of the *Foreman* sequence where spatial dimensions have been halved, and the frame rate corresponds to $\frac{3}{4}$ of the original frame rate. We see that spatial resizing does not perceptibly degrade image quality, by comparison to the Fig. 7.12 for example. Note that frame rate changes implies frame interpolation. The low computational cost interpolation is very efficient thanks to the spline temporal function. Similar results can be obtained with the other dictionaries proposed in the Chap. 7. The only restriction for the transcoding to apply is that the dictionary enables the respect of the similitude group law in the resizing process.



FIGURE 9.2: Reconstruction of the *Foreman* sequence encoded with 500 3D-splines. The spatial resolution is divided by 2 ($\alpha = \frac{1}{2}$) and the temporal resolution is divided by $\frac{4}{3}$ ($\eta = \frac{3}{4}$) by comparison to the CIF coding format.

Table 9.1 shows the effects of the Matching Pursuit stream transcoding onto the distortion of the reconstructed sequence. To this aim, the *Foreman* MP stream has been transformed into different spatio-temporal resolutions according to Eq. 9.2.3. The decoded sequence has been compared to downscaled version of the original sequence. To this aim original frames have been skipped to adapt to the temporal resolutions. The remaining frames have been resized through cubic interpolation of pixel values.

Clearly, we see that lowering the temporal resolutions does almost not impact the quality of the decoded video. In the same time, lowering the spatial resolution may impair the decoded video for very small resizing factors (e.g., $\alpha = 0.25$), while it does not have any sensible impact

onto the quality for larger scaling values. This phenomenon is however due to the imperfect encoding which does not extract all the components participating to the frames of low resolution. Transcoding may therefore hide some of the atoms for low spatial resolutions, since small scale coded atoms have been transcoded to non-visible atoms, thus degrading the reconstruction. The smallest spatial resolution sequences are generally destined to low resolutions displays (e.g., PDA or cellular phones), so that the impact of quality degradation due to transcoding may be limited.

To cope with this problem however, the encoding may be modified to ensure that energetic atoms of low resolution sequence are encoded in the first MP iterations. Hence benefits may be drawn from multi-resolution features of MP stream to build a hierarchical encoding scheme, where the first iterations are computed on a low resolution sequence and the following onto higher resolutions. Furthermore, such a hierarchical coding scheme would also greatly reduce the encoding complexity without loosing on compression efficiency. Generally the largest atoms are indeed encoded in the first iterations anyway.

		Temporal scaling, η		
		1	0.5	0.25
Spatial scaling α	1	26.76 dB	26.13 dB	26.45 dB
	0.5	26.19 dB	25.7 dB	25.88 dB
	0.25	22.86 dB	22.77 dB	22.97 dB

TABLE 9.1: *PSNR quality of the decoded Foreman sequence after MP transcoding to lower spatio-temporal resolutions.*

Finally, Fig. 9.3 and 9.4 propose a tentative comparison between transcoding performances on wavelet and Matching Pursuit stream. Similarly to the comparison proposed in Sec. 7.3.5, the number of wavelets coefficients is three times larger than the number of MP iterations to guarantee close coding rates. Transcoding of the wavelet stream consists in retaining only the lowest resolution subband. Degradation due to transcoding is similar in both cases (the quality of the Matching Pursuit representation is already slightly better than the wavelet one before resolution change). Both transcoding methods are close to optimum since they identically respect similitude group law in $\mathcal{R}^2$. Most of the wavelets coefficients are concentrated in the lowest subband. MP coding however does not ensure that most of the atoms corresponding to high energetic functions in low resolution signal are present in the MP stream, as discussed previously. Nevertheless, flexibility offered by Matching Pursuit is much larger, since wavelets domain resizing methods are limited to octave multiples of the original resolution.

9.2.4 Discussion

The transcoding method proposed hereabove only deals with lowering spatio-temporal resolution of the video sequence. However, in practical heterogeneous system, the rate on the downlink connections has also to be kept minimal. These connections often offers the lowest bandwidth of the video distribution network. Besides limiting the number of forwarded iterations, two main methods can be used to reduce the rate after transcoding. First the atoms whose scales after transcoding becomes too small can be skipped without influencing the quality of the decoded video. Second, the coefficients and atoms indexes can be re-quantized. Indeed, since transcoding generally produces reduced resolutions streams, the respective range of both coefficients and atoms index values decreases in the same proportions (see Eq. 9.2.3). Hence these can be coded with



FIGURE 9.3: Wavelet coded Lena image of size 256×256 with the 1500 most significant coefficients decoded at a resolution of 128 by 128 pixels (PSNR = 24.98 dB).



FIGURE 9.4: Matching Pursuit coded Lena image of size 256×256 with 500 iterations decoded at a resolution of 128 by 128 pixels (PSNR = 25.42 dB).

a smaller rate. Re-quantization process can be performed similarly to the method proposed in Chap. 8, where the optimal number of iteration to transmit for a given rate is also given. A smart transcoder may additionally limit the number of rotation index values for very small resolution, without introducing annoying distortion. The dictionary can thus be adapted to the targeted resolution. For similar structural redundancy values (see Sec. 6.4) the number of functions (i.e., the number of indexes) in the dictionary need obviously to be larger for high resolution signals than for small ones.

Interestingly, multiresolution capabilities depend only on the structure of the dictionary atoms, and they do not impose any strict additional constraint on the Matching Pursuit algorithm. Similarly, video editing functions like spatio-temporal translations, rotation or zooming are also straightforward with the same MP streams. One can easily rotate the video frames or generate special effects by changing the rotation index values, or even crop parts of the images by imposing limits to the signal area. The latter method can also be used to change the frame format (i.e., the ratio between height and width of images) according to the receiver display.

Finally, it was assumed in this section that atom indexes are easily accessible by the transcoder. The packetization method that will be studied in the next section has to enable simple localization of atom indexes. The de-packetization and eventually decoding of Matching Pursuit iterations shall therefore stay simple enough not to overload the proxies with expensive operations.

9.3 Error control

The Matching Pursuit stream has also to cope with heterogeneous packet loss processes when transmitted on IP networks. To this aim, Forward Error Correction (FEC) techniques are clearly the preferred error-control scheme for multi-resolution and point-to-multi-point applications. We will use here a very simple media-independent Forward Error Correction (FEC) mechanism as presented in Chap. 3. Recall basically that K_{FEC} video packets are protected by $N_{FEC} - K_{FEC}$ FEC packets in a N_{FEC} -packets FEC block. For sake of simplicity, assume that the FEC algorithm is able to correct up to $N_{FEC} - K_{FEC}$ packet loss within the FEC block. The total rate has therefore to be optimally shared between source information (i.e., MP iterations) and channel coding (i.e., FEC packets) to provide the best possible quality to the end-user.

The effect of packet losses are dependent on both the FEC parameters and the Matching Pursuit packetization scheme. As shown before, there is no data dependencies between Matching Pursuit iterations. Moreover, the order in which the MP iterations are decoded is irrelevant. The effect of packet losses is therefore easily deduced once one knows which elements have been lost.

The energy of the coefficients can indeed be approximated thanks to the exponential energy decay bound of Matching Pursuit iterations, as shown in Sec. 6.2. However, from the same exponential decay curve, it becomes obvious that not all MP iterations are of the same importance. The first iterations clearly bring more energy than the last ones.

IP networks does however not handle these differences since all IP packets have the same probability to be lost, independently of their content. This mismatch between network behavior and source data can be handled in two main ways. First, the degree of protection can be adapted to the packet content, so that the highest energy packets get the smallest loss probabilities. Second, the packetization can be performed so that each packet has approximately the same importance, which is measured in terms of contribution to the reconstruction energy. The former method, while efficient, is certainly more complex to handle within network nodes, where the protection may have to be adapted to the local channel status. The latter is less complex in terms of error control management, but implies a smarter packetization which may need to be easily reproduced where the rate may need to be adapted.

The packetization and error control problems have clearly to be solved together, thus leading again to a joint source/channel coding problem. Indeed, knowledge about video stream features shall be used to build the best packet stream, either by varying the error protection or the packetization. However, both schemes have to stay simple enough to be performed in gateways in an heterogeneous framework. Network nodes should be able to adapt the rate and/or error protection to the channel status on the downlink connections.

Moreover, to enable the transcoding operations presented in the previous section, MP iterations number need to be deduced from the packet numbers without explicit transmission. The packetization scheme has therefore to stay simple enough to allow for the gateway to reproduce it. In order to enable the quantization scheme proposed in the Chap. 8, the parameters ν and $\|f\|^2$ (i.e., the exponential coefficients decay curve parameters) may be repeated at the beginning of each transmitted packet without a large overhead penalty. These parameters which are expected to vary slowly along a video sequence can also be transmitted once at connection setup, and then managed similarly to the resolution and rate information along the video distribution network.

In this section, two algorithms are proposed to increase the robustness of an MP stream against data loss, that are Unequal and Equivalent Error Protection. Basically, the first algorithm dynamically adapts the level of redundancy according to the relevance of the MP atoms. The second one performs a smart packetization in order to minimize the visual relevancy difference between data packets and then performs a simple and uniform FEC scheme. While both schemes are expected to provide similar protection level, the latter leads to efficient adaptive FEC implementation within transcoding proxies thus facilitating the end-to-end error control management.

Both packetization schemes are based on the following assumptions. Similarly to the encoding, the packetization process is performed on a GOF basis. It is assumed that the total number of iterations M per GOF is given by the targeted reconstruction quality and available source rate. The number of MP packets N per group of frames is given by the source rate (i.e., M) and the average packet length L which depends on the networking type and the delay constraints. The packet loss probability is assumed to be independent of the packet size, provided all packets have approximately the same length. Also, an integer number of Matching Pursuit iterations (i.e., coefficient and index) are gathered into the same packet to avoid any kind of dependency between packets. The packet length is finally assumed to be driven by the number of iterations per packet, in a first approximation. While this hypothesis is valid in the absence of MP iterations coding, quantization and entropy coding may cause slight differences between the length of successive packets. However, this difference is expected to be negligible in regards to the packetization scheme performances.

Lastly, the coefficient energy is expected to be close to the exponential decay curve given by Eq. (6.2.11). This hypothesis is not always validated, and highly depends on the encoding scheme. Heuristics have however shown that the coefficient value generally follow a decreasing exponential curve, although its parameters may be slightly over-estimated by the structural redundancy factor computed in Sec. 6.4. The exponential decay of the coefficients is however the only available information for a priori packetization scheme design. Clearly, it is indeed not practically feasible

to dynamically adapt the packetization to current data even if this would surely lead to a better packetization.

9.3.1 Unequal Error Protection

Matching Pursuit stream consists in typical data for Unequal Error Protection. The energy indeed decreases along increasing iteration numbers and the relative importance of the coefficients thus similarly decreases. As discussed previously, the packetization scheme has to stay simple so that iterations number can be easily deduced without explicit transmission. The most simple packetization scheme is certainly the serial iterations encoding presented in Fig. 9.5. The first $I = \frac{M}{N}$ iterations are fed into the first packet, the next I iterations in the second packet and the packetization continues similarly till all MP iterations have been encapsulated.

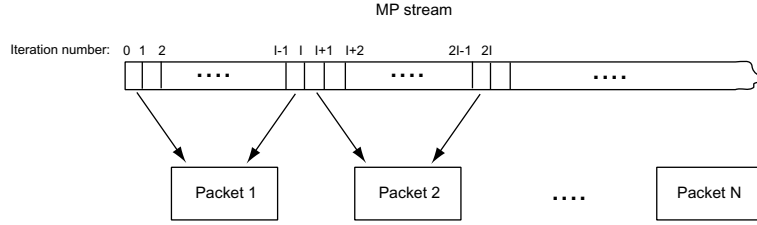


FIGURE 9.5: Serial MP Packetization scheme.

In serial packetization the energy of the packet n can be approximated as :

$$E_n = \sum_{i=(n-1)I+1}^{nI} \|f\|^2 \nu^{2i}, \quad (9.3.1)$$

for $1 \leq n \leq N$. The parameters $(\|f\|^2, \nu)$ correspond to the exponential upper-bound of the coefficients energy (see Eq. (6.2.11)). The serial packetization scheme clearly generates packets of different importance. The ratio between the energy in successive packets is equal to $\frac{E_n}{E_{n+1}} = \frac{1}{\nu^{2I}}$. Hence, the disparity between successive packets increases when ν decreases or I increases (ν smaller than one). The simple serial packetization scheme presents however the advantage of very small delay, thus allowing for efficient packet re-encapsulation within gateways, if needed.

Fig. 9.6 shows the distribution of signal energy in a serial packetization scheme. The CIF *Foreman* sequence have been encoded with 3D splines and encapsulated in 20 packets. The energy distribution is compared to the theoretical distribution given by Eq. (9.3.1). It is shown that, as expected, the energy distribution approximately follows a decaying exponential curve. However the experimental data energy decay is slightly faster than the theoretical decay in the first packets. This is mainly due to the limitations of the Genetic Algorithm encoder that only approximates the Matching Pursuit optimal decomposition. The first coefficients are close to optimal since their generally large respective atoms are not very sensitive to small index variations. For the following iterations, the coder remains often farther from optimal since atoms are very sensitive to small indexes variations. The residual energy decreases slower, resulting in a under-estimated convergence speed for the first iterations. This phenomenon results in under-estimation of the importance and then the protection level of the first packets. To cope with this problem, one can provide a piecewise coefficient upper-bound curve to offer more flexibility to the exponential fitting. Another solution would be to distribute the first coefficient on all packets. This solution is also interesting for rapidly decaying coefficient values (i.e., small ν , in the case of very redundant dictionary). The exponential decay of coefficient energy causes in this case a huge difference between successive packets energy, and thus necessitate a complex unequal error protection scheme.

Hence, the serial packetization may be modified in order to avoid a too important disparity between packet importance. MP iterations can rather be distributed sequentially to the different

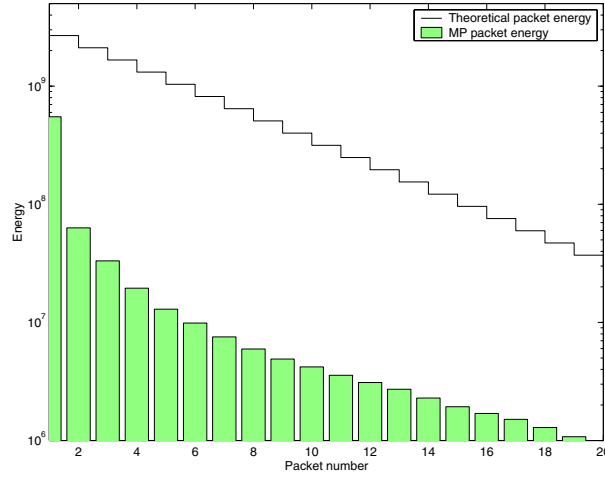


FIGURE 9.6: Energy distribution between successive packets. 500 MP iterations of Foreman sequence have been serially packetized in 20 packets. The energy per packet is compared to the theoretical energy obtained from Eq. (9.3.1).

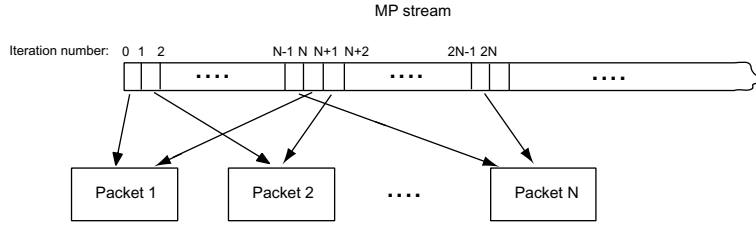


FIGURE 9.7: Unequal Energy Packetization scheme.

packets, i.e., the iterations number $i N + n$ are fed into the packet n . The iterations may even be distributed in a back and forth order among the N transmission packets as represented in Fig. 9.7. The first iteration is fed into the first packet, the second in the second one and similarly until the N^{th} packet. The packets are then scanned backwards, putting an MP iteration in each packet up to the end of the stream. This packetization scheme guarantees that the length of packets is approximately constant, while the energy disparity between packet is kept reasonable. The delay is however more important than in a serial iterations packetization, since no packet can be transmitted before the last iterations are encapsulated. However, this delay stays negligible by comparison to encoding delay.

The energy of each packet n becomes :

$$E_n = \sum_{i=0}^{\lfloor \frac{I-1}{2} \rfloor} \|f\|^2 \nu^{2(n+4+iN)} + \sum_{i=1}^{\lfloor \frac{I}{2} \rfloor} \|f\|^2 \nu^{2(N-n+1+4+iN)}, \quad (9.3.2)$$

and the ratio between the energy in successive packets is equal to $\frac{E_n}{E_{n+1}} = \frac{1}{\nu^2}$. Clearly the energy is better balanced than in the serial packetization scheme. However the packets are far from being equivalent for small ν values (i.e., in the case of an efficient MP encoding thanks to a very redundant dictionary), since the difference between extreme packets is equal to $\frac{E_1}{E_N} = \frac{1}{\nu^{2N}}$.

In both hereabove schemes, the packets have very different importance in terms of energy. The impact of loosing the first packet, for example, is clearly larger than the impact of dropping any other packet. A Unequal Error Protection scheme is proposed to fight against these inequalities and lower the loss probability of the most important packet. To this aim, the N -packets MP stream

is divided into B FEC blocks. Each of these FEC block may have different FEC rate (i.e., K_{FEC} and N_{FEC} parameters), as represented in Fig. 9.8. Varying the FEC block parameters allows to adapt the protection to the importance of the packets. It is assumed here that the transmission protocol like RTP provides mechanisms to distinguish between MP and FEC packets, as proposed in [224]. Notice that the proposed UEP scheme may ask for eventual byte stuffing to compensate the eventual difference in packet size within the same FEC block. The so-generated overhead is kept very low, since packets have very similar size. Other solutions can be proposed to cope with this problem, like EREC [225] or PET [232] encoding, but such schemes will increase the packetization delay.

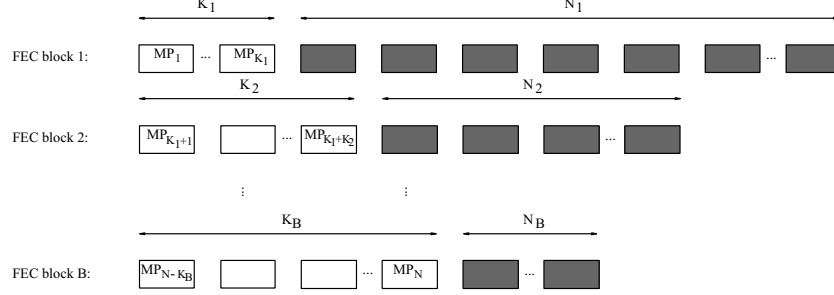


FIGURE 9.8: Generic UEP for N packets constituting B FEC blocks.

To adapt the protection level to the data, the FEC parameters (K_{FEC}^i, N_{FEC}^i) is chosen separately for each sub-groups of packets forming the B FEC blocks. In the serial packetization, the maximal average reconstruction energy after FEC recovery is obtained by optimizing the general equation :

$$\begin{aligned}
 E_{UEP} &= \max_{1 \leq B \leq R} \max_{\{(K_{FEC}^b, N_{FEC}^b)\}, \forall b \leq B} \sum_{k=1}^K E_k (1 - \pi_k) \\
 &= \max_{1 \leq B \leq R} \max_{\{(K_{FEC}^b, N_{FEC}^b)\}, \forall b \leq B} \sum_{k=1}^K (1 - \pi_k) \sum_{i=(k-1)I+1}^k \|c_{\gamma_i}\|^2,
 \end{aligned} \tag{9.3.3}$$

where $K = \sum_{i=b}^B K_{FEC}^b$ under the constraints that $\sum_{b=1}^B N_{FEC}^b$ corresponds to the total available rate R and that $N_{FEC}^b \leq N_{max}$ to limit the FEC reconstruction delay. The parameters π_k depends on both the respective FEC parameters (K_{FEC}^b, N_{FEC}^b) and the channel Packet Loss Ratio (π) and Average Burst Length (α). Under the assumption that the loss process can be represented by the Gilbert-model, the loss probability after FEC recovery is given by Eq. (3.2.7)). The number of iterations per packet, I , is kept constant with the channel status since the packet size remains independent of the transmission conditions. However, the total number of MP iterations has to adapt to the optimal source rate according to the Eq. (9.3.3) for given channel status (i.e., $M = I \sum_{b=1}^B K_{FEC}^b$). Hence the least significant iterations are traded off for channel rate when the network conditions degrade. Note that the Eq. (9.3.3) can be solved numerically or through the Viterbi algorithm. In practical cases, the optimal UEP parameters may be tabulated to reduce the packetization complexity. They indeed depend only on the static parameters ($\|f\|^2, \nu, I$) and the varying channel conditions (R, PLR, ABL).

Optimal UEP parameters are given in Table 9.2 for the transmission of the *Foreman* MP-coded sequence under different network conditions. The serial packetization encapsulates in this case 20 MP iterations in each packet. The packet length is kept relatively small (around 200 bytes) to emphasize the UEP effect. The total rate is set to the equivalent of 500 iterations per GOF. The maximal number of packets in a FEC blocks is set to 10 to limit the reconstruction delay. It can be seen that the channel rate $\frac{\sum_b K_{FEC}^b}{\sum_b N_{FEC}^b}$ obviously increases with the PLR experienced on the channel. Moreover, as expected, for high loss ratios, the protection overhead concentrates on the

first and most important packets. In this case, it is preferable not to transmit the least significant MP iterations to increase the chance for the first packets to be correctly received. Obviously the average energy thus decreases with the increase of the loss ratio. Finally, the small FEC block size does not permit to make any difference between the optimal parameters for different Average Burst Length.

	$\pi = 1 \ e^{-2}$		$\pi = 1 \ e^{-1}$		$\pi = 5 \ e^{-1}$	
	$\alpha = 2$	$\alpha = 3$	$\alpha = 2$	$\alpha = 3$	$\alpha = 2$	$\alpha = 3$
$\frac{\sum_b K_{FEC}^b}{\sum_b N_{FEC}^b}$	1	1	0.85	0.85	0.5	0.5
K_{FEC}^1	10	10	7	7	3	3
N_{FEC}^1	10	10	10	10	9	9
K_{FEC}^2	10	10	10	10	3	3
N_{FEC}^2	10	10	10	10	7	7
K_{FEC}^3	-	-	-	-	4	4
N_{FEC}^3	-	-	-	-	4	4
B	2	2	2	2	3	3
$\frac{E_{UEP}}{\sum_i \ c_{\gamma_i}\ ^2}$	0.99	0.99	0.9568	0.9370	0.8835	0.8162

TABLE 9.2: Optimal UEP parameters for the transmission of Foreman under different packet loss ratio and average burst length (π, α) .

In order to limit the complexity of the optimal UEP scheme, the number of FEC blocks B can be limited. One can even define a policy where either single packets are protected ($K_{FEC}^i = 1$), or packets are gathered in a limited number of FEC groups ($K_{FEC}^i > 1$) or packets are not protected. This three-level protection scheme would be a good compromise between optimal protection and error control complexity.

9.3.2 Equivalent Energy Packetization

The aim of the Equivalent Energy Packetization (EEP) algorithm is to build a stream of N equivalent packets of roughly the same length. All packets have in this case the same importance in terms of MSE distortion in case of loss. The error control scheme is thus facilitated in comparison to UEP scheme, since the same protection overhead is now applied to all packets. Construction of equivalently significant packets have already been proposed in the context of wavelet coders [453, 454] or for direct transmission of pixels values [455]. In the new scheme proposed here, we take benefit of the Matching Pursuit properties to build equivalent energy packets.

The best equivalent energy packetization can surely be reached through classical optimization methods [456], with constraints on both the energy and the packet length. However, as explained before, the packetization has to stay simple and follow a priori defined algorithms for the proxy to be able to repeat packetization operations. Moreover, the decoder has to know the iteration number without their explicit transmission. This requirement is imposed by the design of coding mechanisms based on the MP coefficients decay property (e.g., the quantization scheme proposed in Chap. 8). The only a priori information that can be used for the packetization scheme design is the approximately exponential coefficients decay, or equivalently the parameters $\|f\|^2$ and ν . Based on this knowledge, the equivalent energy packetization scheme proceeds as follows.

As the coefficient energy decays with the iteration number, an increasing number of iterations is fed into successive packets to guarantee equivalent packet importance. Let $\{k_i\}$ with $0 \leq i \leq N$

denote the series of boundaries of equivalent MP iterations sub-groups. This notation means that iteration numbers $k_i + 1$ to k_{i+1} are fed into the packet i . The packetization scheme first puts one MP iteration into the first packet ($k_1 = 0$ and $k_2 = 1$). The number of iterations that have to be put in the second packet is given by solving :

$$\|f\|^2 \nu^2 = \sum_{k=k_2+1}^{k_3} \|f\|^2 \nu^{2k}, \quad (9.3.4)$$

which does however not always have a solution, as discussed below. Equivalently, the iterations sub-groups are given by the following recursive equations :

$$k_{i+1} = \text{round} \left(\frac{\log(\nu^{2k_i} + \nu^2 - 1)}{\log(\nu^2)} \right), \text{ with } k_1 = 0. \quad (9.3.5)$$

The previous relation defines N energy equivalent sub-groups. However, the number of iterations in each sub-groups is clearly different. To build packet of similar size, the EEP scheme therefore scans the packets in back and forth order, as represented in Fig. 9.9. The N equivalent MP iterations sub-groups are first distributed among the N packets. New sub-groups are then formed with the following MP iterations from the rule defined in Eq. (9.3.5). These new sub-groups are then put into the N packets taken in backwards order. If there are remaining MP iterations to be packetized, the algorithm starts another loop in forward packet order, and continues until all iterations are packetized.

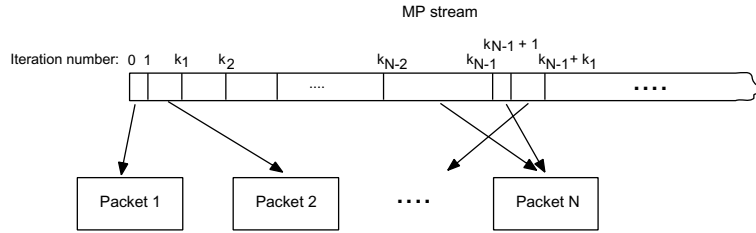


FIGURE 9.9: *Equal Error Packetization scheme.*

Note that the back and forth packet scanning allows to build packets of similar length, since in practical cases the exponential parameter ν is close to one. The constraints on the packet size is due to the packet-based FEC mechanism. FEC blocks are indeed generally composed of packets of identical size. For small ν values however, the first and last packets may be smaller than other packets. In this case, packets of similar length can be grouped into the same FEC blocks. Another solution would be to cyclically distribute the back and forth loops among the transmission packets. The first loop starts in this case on the first packet, the second on the second packet etc., so that the number of iterations in each packet is truly identical after N loops. Remember however that byte stuffing may also be used to correct small packet size differences.

Fig. 9.10 shows the distribution of the coefficients energy among the transmission packets. The CIF *Foreman* sequence has been encoded with 500 3D splines atoms and encapsulated in 20 packets. It can be clearly seen that the energy distribution is better balanced than in the UEP scheme represented in Fig. 9.6. The effects of mismatch between the exponential upper-bound in the first MP iterations are still perceptible in the first packets. To cope with this problem, the first pass of the EEP scheme could be performed with different parameters to better fit the energy of the first coefficients. Finally, the theoretical energy distribution is not perfectly distributed either because of the rounding operation in Eq. (9.3.5). This phenomenon is however negligible in regards to the efficiency of the packetization scheme in lossy transmission.

The EEP scheme suffers however from a severe constraint, namely that the solution may become unavailable depending on the number of packets N and the exponential parameter ν .

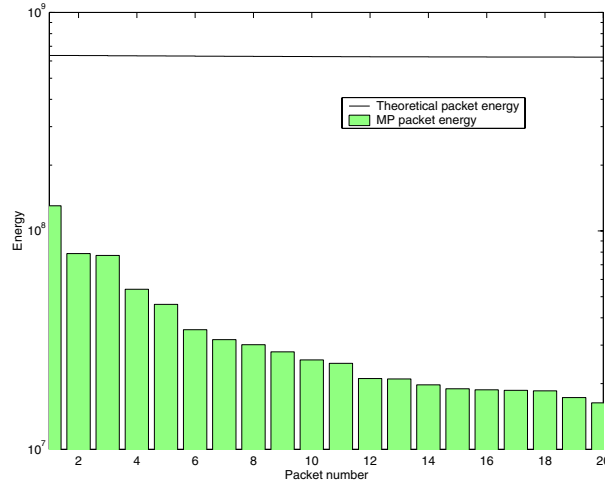


FIGURE 9.10: *Energy distribution between successive packets. 500 MP iterations of Foreman sequence have been serially packetized in 20 packets. The energy per packet is compared to the theoretical energy obtained from Eq. (9.3.5).*

Indeed, from Eq. (9.3.5), it appears that the following constraint has to be respected :

$$\nu^2 k_N > 1 - \nu^2 . \quad (9.3.6)$$

This constraint is due to the a finite integral of exponentially decaying functions. The parameter ν , which depends on both the dictionary and the MP algorithm [72], is generally very close to one in practical applications. The redundancy of the dictionary is generally small due to constraints on the size of the dictionary, and the search algorithm is not optimal due to timing constraints at encoder. Therefore, the exponential energy decay rate is often in the interval $[\frac{1}{2}, 1]$ so that the constraint of Eq (9.3.6) is respected for low rates, or small N values. Notice also that, the EEP scheme does not depend on the signal energy, but only on the parameter ν . Hence, the same packetization can interestingly be performed for different input signal, thus simplifying the video distribution implementation.

The EEP scheme presents the advantage of enabling a very easy error control by media-independent FEC algorithms. The optimal rate distribution between source and channel coding is indeed easily computed in this case. The degradation resulting from losses is directly dependent on the packet loss ratio since each independent packet contributes with the same energy to the signal reconstruction. The FEC performance can be computed from the channel status and the FEC parameters (i.e., the parameters K_{FEC} and N_{FEC}) as shown in Chap. 3 under the Gilbert-model loss process assumption. The optimal rate sharing consists in this case in trading source information for channel protection depending on the packet loss ratio experienced on the network to provide the end-user with the best possible quality.

The performance of the EEP scheme can be measured in terms of average reconstruction energy, which corresponds to the energy brought by MP packets weighted by the probability of being correctly received. Assuming the least significant coefficients are sacrificed for channel coding, the optimal reconstruction energy E_{EEP} is given by :

$$\begin{aligned} E_{EEP} &= \max_{(K_{FEC}, N_{FEC})} \sum_{p=1}^K E_p (1 - \pi_p) \\ &= \max_{(K_{FEC}, N_{FEC})} (1 - \pi) \sum_{i=1}^{\frac{K_{FEC}}{N_{FEC}} M} c_{\gamma_i}^2 , \end{aligned} \quad (9.3.7)$$

where M is the total number of iterations in the MP stream, and π is the packet loss ratio experienced by MP information packets, which depends on both the FEC parameters (K_{FEC}, N_{FEC}) and the channel status (PLR, ABL) from Eq. (3.2.7). The parameter c_{γ_i} represent the i^{th} MP coefficient.

The optimal FEC parameters are given in Table 9.3 for different channel conditions. The *Foreman* sequence has been encoded with 500 iterations, and each MP packet contains 25 iterations. The total rate corresponds to 20 packets, or equivalently two FEC blocks of length 10. It can be seen that the channel rate increases with the PLR, as expected. But the influence of the Average Burst Length is not perceptible onto the FEC parameters due to the relatively short FEC blocks.

	$\pi = 1 e^{-2}$		$\pi = 1 e^{-1}$		$\pi = 5 e^{-1}$	
	$\alpha = 2$	$\alpha = 3$	$\alpha = 2$	$\alpha = 3$	$\alpha = 2$	$\alpha = 3$
$\frac{K_{FEC}}{N_{FEC}}$	1	1	1	0.8	0.4	0.4
K_{FEC}	10	10	10	8	4	4
N_{FEC}	10	10	10	10	10	10
$\frac{E_{UEP}}{\sum_i \ c_{\gamma_i}\ ^2}$	0.99	0.99	0.9363	0.9	0.8384	0.7718

TABLE 9.3: Optimal EEP FEC parameters for the transmission of *Foreman* under different packet loss ratio and average burst length (π, α) .

9.3.3 Experimental comparisons

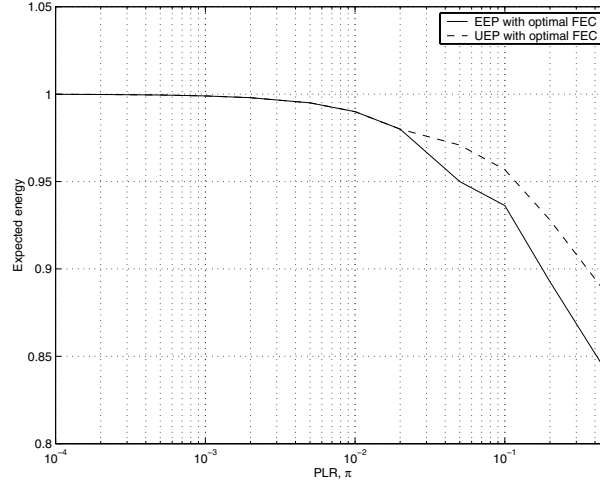


FIGURE 9.11: Evolution of the expected energy as a function of the channel packet loss ratio ($ABL = 2$).

This section now establishes a comparison between both packetization schemes. Fig. 9.11 compares the evolution of the expected energy of both UEP and EEP schemes as a function of the packet loss ratio. The expected energy represents the part of the signal energy available to the decoder weighted by its respective probability to be correctly received. The source and channel rates are optimally chosen in both cases, according to Eqs 9.3.3 or 9.3.7, and the size of the FEC block is limited to 10 to respect delay constraints.

Even if both schemes perform similarly at low loss ratio, the UEP algorithm behaves clearly better than the EEP method for the highest loss rates. Indeed, in this case, the protection is specifically applied to the highest energy iterations in UEP, while it is spread on all coefficients in EEP. UEP therefore offers a higher expected energy, since it provides much more design flexibility for the protection scheme. However, the end-to-end error control is more complex for the UEP scheme. Note nevertheless that the difference between both schemes can be reinforced by the mismatch between the theoretical design and the experimental data, as discussed previously. Indeed, the schemes are designed upon theoretical values of the exponential decay curve, while data does not truly fit this upper-bound, especially for the first MP elements. This phenomenon favors the UEP scheme where the protection is exaggerated for the highest energy coefficients.

Even if the UEP is better than the EEP scheme in average, degradation may be much more important in the first scheme. Fig. 9.12 and 9.13 compares the effect of the loss of the first packet in both cases. It can be clearly seen that the degradation is much more graceful in the EEP scheme, since the signal energy is distributed in all packets. In contrary the I first and most important iterations are lost in the UEP scheme, where $I = 25$ is the number of iterations per packet. The first packet gets however more protection in the UEP algorithm for high loss ratios, and hence a smaller probability to be lost. Nevertheless remaining losses may induce very severe damage.

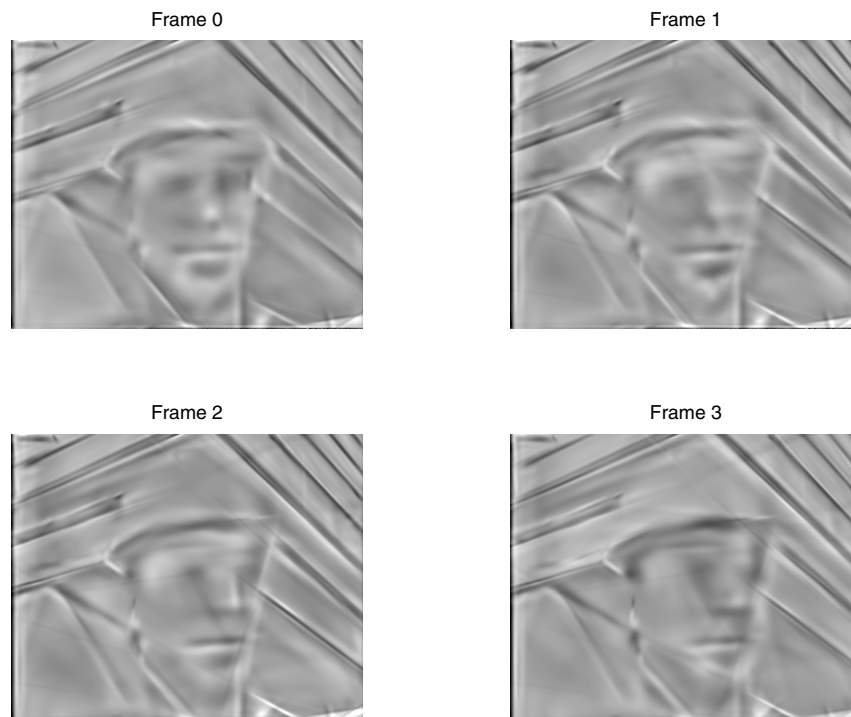


FIGURE 9.12: First four frames of Foreman after the loss of the first packet in UEP scheme.

Finally, the data issued from Matching Pursuit coding are particularly well suited for error protection. The decay in coefficients energy facilitates the bit stream truncation to allow for additional channel rate. Moreover the source energy and the channel rate can be efficiently distributed so that even for loss ratio as high as $5e^{-1}$ the major part of the coded signal energy (80 %) can be expected at receiver.

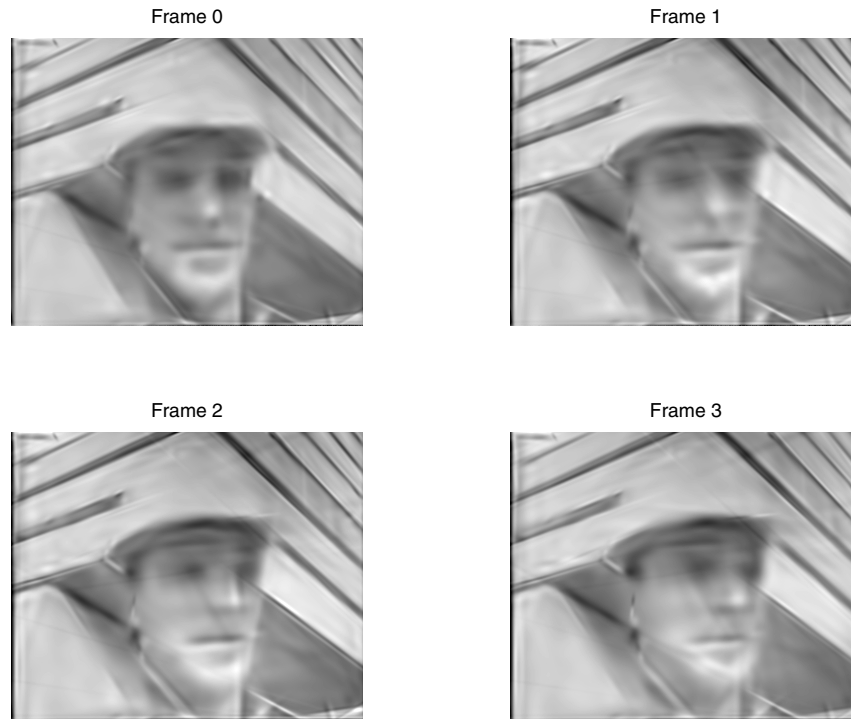


FIGURE 9.13: *First four frames of Foreman after the loss of the first packet in EEP scheme.*

9.3.4 Discussion

In an heterogeneous framework the packetization scheme has to meet a trade-off between efficiency and complexity. The MP stream may need to be de-packetized into the transcoders to form new packets adapting to rate and channel constraints on the outgoing connections. The error control is difficult to adapt in the UEP scheme, since the optimal FEC parameters are not straightforward for given channel status. However, the serial packetization in UEP allows a very rapid bit rate truncation, since packets can simply be dropped to reduce the rate.

In the EEP scheme, the error control management is simple to perform since the channel overhead is identical for each packet. The transcoder can thus easily adapt the protection to the outgoing channel status. However, operations like rate truncation are more difficult to perform than in the UEP scheme. The packets have indeed in this case to be completely decoded to remove the least significant MP elements. Rate could also be adapted by packet dropping, but this solution would induce the loss of the MP compression benefits. However, if transcoder has anyway to adapt the resolution of the MP stream, the computational overload generated by the EEP scheme is not significative.

Equal Energy Packetization can be seen as a form of Multiple Description Coding [457]. The scheme aims at defining several equivalent descriptions (i.e., packets) of the coded signal. The quality of the reconstruction thus increases linearly with the number of correctly received packets. Matching Pursuit however offers much more possibilities for efficient Multiple Description Coding, thanks to the redundant structure of the dictionary. The EEP scheme is only a first step in this promising direction, but modifications of the coding are necessary to build multiple optimal descriptions.

9.4 Conclusion

Matching Pursuit is particularly well appropriate for video transmission over lossy and heterogeneous networks. It offers larger flexibility than video standard-based schemes in terms of both scalability and error control. MP provides inherent multiresolution properties to adapt to the receiver capabilities. Furthermore, Matching Pursuit streams enable the construction of independent packets, and thus limits the error propagation. Two new packetization schemes were proposed by taking benefits of the MP properties. These algorithms constitutes the basis for efficient error control, which may have to be performed locally within the heterogeneous video distribution networks.

However, we did not address the management of the complete video distribution application, driven by the receivers [319]. When these ask for receiving the video stream, they indicate their capabilities to the proxies. The proxies have then to exchange information about receiver features and channel status to permit the design of an optimal video distribution application. Such an application has to offer each receiver the best possible video quality, according to their respective features and the effective available bandwidth. Another important issue that was only partially treated is the management of the end-to-end delay for video delivery. This delay depends on the coding, packetization and transmission delay, but also on the transcoding operations performed within gateways. This latter process can really delayed the video delivery if repeated along the transmission path. It has therefore also to be handled carefully by the end-to-end application.

Finally, iterations have been coded independently within packets. This independence facilitates the management of the Matching Pursuit stream within gateways. However, the compression can be improved by providing a packet-based coding. Indeed, as the transmission is packet-oriented, information within packets is either completely received or not. Entropy or differential coding may thus be performed within packets without sacrificing on error resilience. The MP stream has however to stay progressive and allows an easy transcoding within gateways (and hence de/recoding) to properly work in today's heterogeneous infrastructure.

Conclusions

This thesis has tackled a range of issues in scalable joint-source and channel coding based on video standards and non-orthogonal signal representations. We summarize here the key results of this work. We also briefly discuss some of the remaining issues in the design of a complete robust and scalable video scheme.

An end-to-end Quality of Service model has first been proposed. The FEC performance has been analyzed in terms of both Packet Loss Ratio and Average Burst Length of video data after error recovery. Both parameters are necessary to correctly translate the channel quality of service into video degradation. The MPEG-2 video encoding distortion has then been expressed as an exponential function of the source rate. Both the source rate-distortion and the channel loss-degradation functions finally establish the set of equations used in a complete modeling of the video Quality of Service. The end-to-end quality model helped to build an optimal transmission scheme by adaptive distribution of the available bandwidth between source and channel rate. This optimal rate distribution has been shown to outperform classical video transmission schemes thanks to its adaptivity to the video scene, the loss process and the available rate.

A robust MPEG-2 encoding scheme has then been proposed to limit the packet loss damage within the decoded video. Clearly packet losses can not be avoided in practical cases. Even if the available rate is correctly allocated between FEC protection and source coding, remaining losses can still induce severe quality degradation. To cope with this problem, an error-resilient MPEG-2 coding scheme intentionally introduces additional syntactic information within the video stream to limit error propagation. The scheme modulates the slice length and the number of Intra-coded macroblocks according to the packet loss ratio experienced on the network. It basically simulates the effects of hypothetical losses to optimally protect information where most needed. The complete AMISP encoding scheme additionally triggers the dynamic insertion of FEC packets to recover the most important information. This joint source and channel coding scheme has been shown to outperform other structuring schemes thanks to its improved adaptivity.

Even with adaptive structuring and protection, standard-based video streams can severely suffer from packet losses due to their highly hierarchical structures. Since they are not truly scalable either, a completely new video coding scheme has been proposed to cope with these limitations. The new algorithm is based on spatio-temporal Matching Pursuit decomposition of the video signal over a set of redundant basis functions. A new formulation of the structural redundancy of over-complete dictionaries is proposed to quantify the compression efficiency of Matching Pursuit coding. Several spatio-temporal dictionaries have then been proposed to cover the trade-off between compression efficiency and decoding complexity in an heterogeneous video distribution framework. The most efficient dictionaries have been shown to offer better coding results than wavelet coding in a tentative comparison at low coding rate.

Besides inherent multi-resolution properties, Matching Pursuit coding presents an increased robustness to losses thanks to the non-orthogonal transforms and an exponential decaying energy

distribution among successive coefficients. The latter property has been used to design an optimal quantization scheme which has been shown to outperform classical quantization algorithms. Meanwhile, it has been demonstrated that index quantization has preferably to be avoided. Clearly, the atom indexes bring possibly more information than coefficients in a non-orthogonal signal representation as performed by Matching Pursuit. This clearly emphasizes the need for a careful design of the dictionary which has to provide good compression features but also a low coding rate.

Finally the heterogeneous and lossy Matching Pursuit video delivery has been studied. More particularly the multi-resolution features of structured dictionaries have been used to design low-complexity and near-optimal transcoding schemes. Filtering operations are also straightforward on Matching Pursuit stream thanks to their intrinsically progressive nature. This latter property has been used to build an optimal Unequal Error Protection scheme. This scheme however may lead to complex error control management within disparate networks. Hence, a new Equivalent Energy Packetization scheme has lastly been proposed to ease the error control while not loosing in recovery performance.

Primary results have shown that Matching Pursuit video coding is very promising as an interesting alternative to classical motion-compensation and DCT based coders in the today's heterogeneous framework. Nevertheless a lot of work remains to be done to build a complete robust and multi-resolution video communication system.

For example, the proposed coding scheme voluntarily maintains independence between Matching Pursuit iterations to ensure a good robustness to losses. In this context entropy coding has not been studied since the encoder almost optimally decorrelates the components. However, the compression can be improved by packet-based entropy or differential coding. Indeed, packets represent the smallest independent entity in transmission. Differential coding within packets can not lead to error propagation since the packet is either completely received or lost. Hence error resilience benefits can be drawn from the fact that the transmission entity is smallest than video entities. In contrary video standard-based entities (e.g., slices or GOP) are much larger than transmission packets.

In parallel, the Matching Pursuit coding complexity has also to be reduced to spread the usage of this promising technologies. From the inherent multi-resolution properties of structured dictionaries a hierarchical encoding scheme can be proposed. Such a scheme consists in computing the first iterations onto sub-sampled video signals and sub-dictionaries. The size of both the signal and the dictionary is then increased along the coding process. Hierarchical coding leads to a tremendous complexity reduction since the inner product computation is greatly simplified. The dictionary can also be chosen to reduce the complexity of the Matching Pursuit algorithm by simplifying the inner product computation or the search for the best atoms. Similarly, the Matching Pursuit coding can be combined with quadtree-like signal decomposition to limit the size of the input signal. The video sequence can be first decomposed into smaller regions where Matching Pursuit is then focused on. In the same context, optimal dictionary design as a trade-off between compression efficiency and coding rate has also to be investigated. The efficiency of Matching Pursuit coding has also to be improved at high rates, either by dictionary orthogonalization or eventually hybrid coding.

Finally, the management of the video distribution tree has not been addressed in this work. The encoder however has to know the status of the distribution tree to build the best possible video stream. Gateways have also to exchange information about their respective receivers capabilities and outgoing channel status. All this information has to be managed by a streaming protocol capable to dynamically adapt the video distribution process and proxies processing.

APPENDIX

Optimal Number of MP Iterations

In this appendix the details of the computation of the optimal number of MP iterations N_{opt} are presented in the case of a redundancy-driven quantization of the coefficients.

Recall that the Lagrangian cost function of eq. (8.3.11) is piecewise defined, where the break-point corresponds to the first non-transmitted coefficient. Eq. (8.3.11) may be rewritten as :

$$\mathcal{L}(\lambda) = \begin{cases} \frac{\lambda N}{\log(2)} \left(\frac{1}{2} + \log \sqrt{\frac{\|f\|^2 \log(2)}{6 \lambda}} + \frac{N-1}{2} \log(\nu) + B \log(2) \right) + \nu^{2N} \|f\|^2 & \text{if } N \leq N_B \\ \frac{\lambda N_B}{\log(2)} \left(\frac{1}{2} + \log \sqrt{\frac{\|f\|^2 \log(2)}{6 \lambda}} + \frac{N_B-1}{2} \log(\nu) \right) + \lambda B N + \nu^{2N} \|f\|^2 + \frac{\|f\|^2}{12} \frac{\nu^{2N} - \nu^{2N_B}}{\nu^2 - 1} & \text{otherwise.} \end{cases} \quad (\text{A.0.1})$$

The minimum are now computed for the right and left sides of the piecewise Lagrangian cost function. The solutions are shown to depend on the atom index size and the convergence rate, respectively the parameters B and ν .

The first derivative of the right part of the Lagrangian (i.e., for $N > N_B$) can be written as :

$$\frac{\partial \mathcal{L}(\lambda)}{\partial N} = \lambda B + 2 \|f\|^2 \left(1 + \frac{1}{12 \nu^2 - 12} \right) \nu^{2N} \log \nu. \quad (\text{A.0.2})$$

The zero of the eq. (A.0.2) is given by :

$$N_2 = \frac{\log \left(-\frac{\lambda B}{2 \|f\|^2 \log(\nu) \left(1 + \frac{1}{12 \nu^2 - 12} \right)} \right)}{2 \log(\nu)}. \quad (\text{A.0.3})$$

However, this equation is valid only for $N > N_B$, by definition. The condition for N_2 to be larger than the value of N_B is given by :

$$\frac{\log \left(-\frac{\lambda B}{2 \|f\|^2 \log(\nu) \left(1 + \frac{1}{12 \nu^2 - 12} \right)} \right)}{2 \log(\nu)} > \frac{-\log \left(\sqrt{\frac{\|f\|^2 \log(2)}{6 \lambda}} \right)}{\log(\nu)}, \quad (\text{A.0.4})$$

or, for $\nu \in]0, 1[$ and B positive,

$$B < -\frac{12 \log(\nu) + \frac{\log(\nu)}{\nu^2 - 1}}{\log(2)}. \quad (\text{A.0.5})$$

The second derivative of the right part of the Lagrangian cost function (i.e., for $N > N_B$) is always positive. Indeed,

$$\frac{\partial^2 \mathcal{L}(\lambda)}{\partial N^2} = 4 \|f\|^2 \left(1 + \frac{1}{12 \nu^2 - 12} \right) \nu^{2N} \log^2 \nu \geq 0, \quad (\text{A.0.6})$$

for $\nu \in]0, 1[$. Hence N_2 is a minimum of the Lagrangian cost function provided that the constraint of eq. (A.0.5) is satisfied. The constraints indicates whether it is worth transmitting the indexes without coefficients and depends only on the size of indexes B and the decay rate of the energy of the coefficients ν .

The first derivative of the left side of the Lagrangian cost function (i.e., for $N < N_B$) can be written as :

$$\begin{aligned} \frac{\partial \mathcal{L}(\lambda)}{\partial N} &= \frac{\lambda \log(\nu)}{\log(2)} N + 2 \nu^{2N} \log(\nu) \|f\|^2 \\ &+ \frac{\lambda}{\log(2)} \left(\frac{1}{2} + \log \sqrt{\frac{\|f\|^2 \log(2)}{6 \lambda}} - \frac{\log(\nu)}{2} + B \log(2) \right). \end{aligned} \quad (\text{A.0.7})$$

For the sake of clarity, let rewrite eq. (A.0.7) as :

$$\frac{\partial \mathcal{L}(\lambda)}{\partial N} = a + b x + d e^x, \quad (\text{A.0.8})$$

with

- $a = \frac{\lambda}{2 \log(2)} + \lambda B - \frac{\lambda \log(\nu)}{2 \log(2)} + \frac{\lambda}{2} \log \left(\frac{\|f\|^2 \log(2)}{6 \lambda} \right)$,
- $b = \frac{\lambda}{2 \log(2)}$,
- $d = 2 \|f\|^2 \log(\nu)$,
- $x = 2 N \log(\nu)$.

The eq. (A.0.8) presents possibly several zeros given by :

$$x = -\frac{a + b W\left(\frac{d}{b} e^{-\frac{a}{b}}\right)}{b}, \quad (\text{A.0.9})$$

where $W(z)$ is the Lambert W function [444], which can present two real branches $W_0(z)$ and $W_{-1}(z)$ for $z \in [-e^{-1}, 0]$. First it can be shown that the argument of W is indeed in the region where two branches are defined. It can be shown that, for $B > 1$ ¹, the argument of W is negative and larger than $-e^{-1}$. Indeed,

$$-e^{-1} \leq (3 \cdot 2^{3-2B} \nu \log(\nu) e^{-1}) < 0, \quad (\text{A.0.10})$$

for $\nu \in]0, 1[$. This ensures the existence of two solutions for N_{opt} and hence that the Lagrangian cost function presents two optimum which are studied now. From the definition of the Lambert W function, $0 \geq W_0(z) \geq -1$ and $W_{-1}(z) < -1$. This means that $N_1^0 > N_{11}$ since $\log(\nu) < 0$, where $N_{10} = \frac{x_0}{2 \log(\nu)}$ and $N_{11} = \frac{x_{-1}}{2 \log(\nu)}$ are the solutions of eq. (A.0.9) for respectively the first and second branches of the Lambert W function.

The second derivative of the left part of the Lagrangian cost function is given by :

$$\frac{\partial^2 \mathcal{L}}{\partial N^2} = 2 b \log(\nu) + 2 d \log(\nu) e^{2 \log(\nu) N}, \quad (\text{A.0.11})$$

¹The atom length is larger than one bit in our application since the number of atoms is clearly larger than two to form over-complete dictionaries

where $b \geq 0$ and $d < 0$ for $\nu \in]0, 1[$ and $\lambda \geq 0$. The second derivative is therefore positive when

$$2 d \log(\nu) e^{2 \log(\nu) N} > -2 b \log(\nu) , \quad (\text{A.0.12})$$

or equivalently

$$e^{2 \log(\nu) N} > -\frac{b}{d} . \quad (\text{A.0.13})$$

From eq. (A.0.9),

$$e^{2 \log(\nu) N_1} = e^{-\frac{a}{b}} e^{-W\left(\frac{d}{b} e^{-\frac{a}{b}}\right)} . \quad (\text{A.0.14})$$

Since, by definition, $e^{-W(z)} = \frac{W(z)}{z}$ for $z \neq 0$, we have :

$$e^{2 \log(\nu) N_1} = \frac{b}{d} W\left(\frac{d}{b} e^{-\frac{a}{b}}\right) . \quad (\text{A.0.15})$$

Finally, from eq. (A.0.13),

$$e^{2 \log(\nu) N_{10}} = \frac{b}{d} W_0\left(\frac{d}{b} e^{-\frac{a}{b}}\right) < -\frac{b}{d} , \quad (\text{A.0.16})$$

and

$$e^{2 \log(\nu) N_{11}} = \frac{b}{d} W_{-1}\left(\frac{d}{b} e^{-\frac{a}{b}}\right) > -\frac{b}{d} , \quad (\text{A.0.17})$$

since by definition $0 \geq W_0(z) \geq -1$ and $W_{-1}(z) < -1$. The second derivative is therefore positive for $N = N_{11}$ and negative for $N = N_{10}$. Hence the Lagrangian cost function presents a minimum at $N = N_{11}$ and a maximum at $N = N_{10}$. Since $N_{11} < N_{10}$, the Lagrangian cost function is first a decreasing function of N , for N small. Moreover, the maximum of the left part of the function is larger than N_B , since

$$-W_0\left(\frac{d}{b} e^{-\frac{a}{b}}\right) < 1 + 2 B \log(2) - \log(\nu) , \quad (\text{A.0.18})$$

for $\nu \in]0, 1[$ and B positive.

Finally, it remains to show that the optimal solution for the left part of the Lagrangian function is positive and smaller than N_B . The minimum N_{11} is first positive if the following condition is respected :

$$\lambda < \frac{4 \|f\|^2 \log(2) \log(\nu)}{W_{-1}(3 \cdot 2^{3-2B} \nu \log(\nu) e^{-1})} . \quad (\text{A.0.19})$$

This means that, in the worst case for the Lambert function,

$$\lambda < 4 \|f\|^2 \log(2) \log\left(\frac{1}{\nu}\right) . \quad (\text{A.0.20})$$

This imposes a limit on the weight of the rate in the Lagrangian cost function. Second, N_{11} is smaller than N_B if :

$$-W_{-1}(3 \cdot 2^{3-2B} \nu \log(\nu) e^{-1}) > 1 - \log(\nu) + 2 \log(2) B . \quad (\text{A.0.21})$$

Since $W(z) = a$ is equivalent to $a e^a = z$, by definition, the previous condition is equivalent to :

$$B > -\frac{23 \log(\nu) + 1}{2 \log(2)} . \quad (\text{A.0.22})$$

It has to be noted that both conditions from Eqs (A.0.5) and (A.0.22) are mutually exclusive. If one constraint is respected, the other cannot be satisfied. Moreover, it can be shown that one of the constraint is always satisfied. This ensures the existence of one and only one minimum to the Lagrangian cost function. Indeed, the product of both conditions is always positive.

$$\left(-\frac{23 \log(\nu) + 1}{2 \log(2)} - B\right) \left(-\frac{12 \log(\nu) + \frac{\log(\nu)}{\nu^2 - 1}}{\log(2)} - B\right) \geq 0 . \quad (\text{A.0.23})$$

The previous relation is always positive since both terms have the same sign. Indeed, if

$$B + \frac{12 \log(\nu) + \frac{\log(\nu)}{\nu^2 - 1}}{\log(2)} \leq 0 , \quad (\text{A.0.24})$$

then

$$B + \frac{23 \log(\nu)}{2 \log(2)} \leq -\frac{\log(\nu)}{2 \log(2)} - \frac{2 \log(\nu)}{2 \log(2) (\nu^2 - 1)} . \quad (\text{A.0.25})$$

Furthermore, if eq. (A.0.24) is respected, then

$$B + \frac{23 \log(\nu) + 1}{2 \log(2)} \leq 0 . \quad (\text{A.0.26})$$

From eq. (A.0.25), the last relation can be rewritten as :

$$\frac{2 \log(\nu)}{\nu^2 - 1} \leq 1 - \log(\nu) . \quad (\text{A.0.27})$$

Setting $\nu = e^{-\frac{t}{2}}$, for $t \in]0, \infty[$, it is equivalent to prove that

$$t \geq (e^{-t} - 1) \left(1 + \frac{t}{2}\right) , \quad (\text{A.0.28})$$

where both left and right terms are null for $t = 0$. The left terms increases with t with a slope 1. The right term also increases with t , since its first derivative is always positive, but the increasing rate is always smaller or equal to 1. This proves the inequality of eq. (A.0.28).

It can be shown in the same way that if one of the factor of eq. (A.0.23) is negative, the other one is negative also. Developing this conditions ends up in proving again the relation of eq. (A.0.28). This means either the first conditions is satisfied and not the second one, or the reverse. This ensures the existence of one and only one solution to the optimal number of iterations to transmit for a given rate.

Bibliography

- [1] ISO International Standard, *Information technology - Generic coding of moving pictures and associated audio information - Part 1: Systems Specification*. ISO/IEC 13818-1, first edition 1996-05-15 ed.
- [2] ISO International Standard, *Information technology - Generic coding of moving pictures and associated audio information - Part 2 :Video*. ISO/IEC 13818-2, first edition 1996-05-15 ed.
- [3] Haskell B.G., Puri A. and Netravali A.N., *Digital Video: an Introduction to MPEG-2*. Digital Multimedia Standards Series, Chapman and Hall, 1997.
- [4] Huard J.-F. and Lazar A. A., "On End-to-End QoS Mapping," in *Proceedings of the International Workshop on QoS*, pp. 303–314, May 1997.
- [5] Kawashima M., Chen C.-T., Jeng F.-C. and Singhal S., "Adaptation of the MPEG Video-Coding Algorithm to Network Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 261–269, August 1993.
- [6] Appadwedula S., Jones D. L., Ramchandran K. and Konzintsev I., "Joint Source Channel Matching for a wirelss communications link," in *Proceedings of the International Conference on Communications*, (Atlanta, GA), June 7-11 1998.
- [7] Shannon C. E., *Collected Papers*. IEEE Press, 1993.
- [8] Shannon C.E., "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, July 1948.
- [9] Shannon C.E., "A Mathematical Theory of Communication - Part II," *The Bell System Technical Journal*, vol. 27, pp. 623–656, October 1948.
- [10] Garcia Adanez X., Verscheure O. and Frossard P., "Reliable Transmission of MPEG-2 VBR Video Streams over New Network Adaptation and ATM Adaptation Layers," in *IEEE ATM Workshop*, (Lisboa), pp. 158–167, May 1997.
- [11] Verscheure O., Frossard P. and Le Boudec J.-Y., "Joint Smoothing and Source Rate Selection for Guaranteed Service Networks," in *Proceedings of the IEEE INFOCOM Conference*, (Anchorage, Alaska), April 2001. accepted for publication.
- [12] Gray R.M. and Neuhoff D.L., "Quantization," *IEEE Transactions on Information Theory*, vol. 44, pp. 2325–2383, October 1998.
- [13] Haskell B.G., Howard P.G., LeCun Y.A., Puri A., Ostermann J., Civanlar M.R., Rabiner L., Bottou L. and Haffner P., "Image and Video Coding - Emerging Standards and Beyond," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 814–837, November 1998.

- [14] Reimers U., "Guideline for the Use of DVB Specifications and Standards," Tech. Rep. TR 101 200, The DVB Project, June 1998.
- [15] Chiariglione L., "Impact of MPEG Standards on Multimedia Industry," *Proceedings of the IEEE*, vol. 86, pp. 1222–1227, June 1998.
- [16] Avaro O., Chou P.A., Eleftheriadis A., Herpel C., Reader C. and Signes J., "The MPEG-4 systems and description languages: A way ahead in audio visual information representation," *Signal Processing: Image Communications*, vol. 9, pp. 385–431, May 1997.
- [17] Le Gall D., "MPEG: A Video Compression Standard for Multimedia Applications," *Communications of the ACM*, vol. 34, April 1991.
- [18] Le Gall D. J., "The MPEG video compression algorithm," *Signal Processing: Image Communications*, vol. 4, pp. 129–140, 1992.
- [19] Mitchell J.L., Pennebaker W.B., Fogg C.E. and LeGall D., *MPEG Video Compression Standard*. Chapman and Hall, 1997.
- [20] Daubechies I., "Ten Lectures on Wavelets," in *CBMS-NSF Lecture Notes*, no. 61, SIAM, 1992.
- [21] Vetterli M. and Kovacevic J., *Wavelets and Subband Coding*. Prentice Hall, 1995.
- [22] Vetterli M. and Herley C., "Wavelets and filter banks: Theory and design," *IEEE Transactions on Signal Processing*, vol. 40, pp. 1862–1875, August 1992.
- [23] Xiong Z., Ramchandran K., Orchard M.T. and Zhang Y.-Q., "A Comparative Study of DCT- and Wavelet-Based Image Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 692–695, August 1999.
- [24] Karlsson G. and Vetterli M., "Extension of finite length signals for subband coding," *Signal Processing*, vol. 17, pp. 161–166, June 1989.
- [25] Chai B.-B., Vass J. and Zhuang X., *Visual Information, Representation, Communication and Image Processing*, ch. Statistically Adaptive Wavelet Image Coding. 1999.
- [26] Antonini M., Barlaud M., Mathieu P. and Daubechies I., "Image Coding Using Wavelet Transform," *IEEE Transactions on Image Processing*, vol. 1, pp. 205–220, April 1992.
- [27] Shapiro J.M., "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445–3462, December 1993.
- [28] Said A. and Pearlman W.A., "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, June 1996.
- [29] Vass J., Chai B.-B., Palaniappan K. and Zhuang X., "Significance-Linked Connected Component Analysis for Very Low Bit-Rate Wavelet Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 630–647, June 1999.
- [30] Vass J. and Zhuang X., "Efficient Three-Dimensional Wavelet Codecs for Networked Video Communications," in *Proceedings of the IEEE International Conference on Image Processing*, 1999.
- [31] Chai B.-B., Vass J. and Zhuang X., "Significance-Linked Connected Component Analysis for Wavelet Image Coding," *IEEE Transactions on Image Processing*, vol. 8, pp. 774–784, June 1999.

- [32] Servetto S., Ramchandran K. and Orchard M.T., "Wavelet based image coding via morphological prediction of significance," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 530–533, October 1995.
- [33] Witten I.H., Neal R.M., Cleary J.G., "Arithmetic Coding for Data Compression," *Communications of the ACM*, vol. 30, pp. 520–540, June 1987.
- [34] Huffman D.A., "," *Proceedings of the Institute of Radio Engineers*, vol. 40, pp. 1098–1101, 1952.
- [35] Golomb S.W., "Run-Length Encodings," *IEEE Transactions on Information Theory*, vol. 12, pp. 399–401, July 1966.
- [36] Song I. and Ann S., "An Upper Bound on Run-Length Coding Entropy," *Proceedings of the IEEE*, vol. 73, pp. 1141–1142, June 1985.
- [37] Knockaert L., "Comment on An Upper Bound on Run-Length Coding Entropy," *Proceedings of the IEEE*, vol. 74, pp. 1284–1285, September 1986.
- [38] Yang X. and Ramchandran K., "Hierarchical Backward Motion Compensation for Wavelet Video Coding using Optimized Interpolation Filters," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, pp. 85–88, 1997.
- [39] Shen K. and Delp E.J., "Wavelet Based Rate Scalable Video Compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 109–122, February 1999.
- [40] Ohm J.R., "Three-Dimensional Subband Coding with Motion Compensation," *IEEE Transactions on Image Processing*, vol. 3, pp. 559–571, September 1994.
- [41] Uz K.M., Vetterli M. and LeGall D.J., "Interpolative Multiresolution Coding of Advanced Television with Compatible Subchannels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, pp. 86–99, March 1991.
- [42] Ohm J.R., "Advanced Packet Video Coding Based on Layered VQ and SBC Techniques," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 208–221, June 1993.
- [43] Chung W.C., Kossentini F. and Smith M.J.T., "A New Approach to Scalable Video Coding," in *Proceedings of the IEEE Data Compression Conference*, (Snowbird, Utah), March 1995.
- [44] Kim Y.H. and Modestino J.W., "Adaptive Entropy-Coded Subband Coding of Images Sequences," *IEEE Transactions on Communications*, vol. 41, pp. 975–987, June 1993.
- [45] Girod B., "The efficiency of motion-compensated prediction for hybrid coding of video sequences," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 1140–1154, August 1987.
- [46] Natarajan T.R. and Ahmed N., "On Interframe Transform Coding," *IEEE Transactions on Communications*, vol. 25, pp. 1323–1329, November 1977.
- [47] Westwater R. and Fuhr B., "Three-Dimensional DCT Video Compression Technique Based on Adaptive Quantizers," in *Proceedings of the International Conference on Engineering of Complex Computer Systems*, pp. 189–198, 1996.
- [48] Westwater R. and Fuhr B., "The XYZ Algorithm For Real-Time Compression of Full-Motion Video," *Journal of Real-Time Imaging*, vol. 2, pp. 19–34, 1996.
- [49] Fuhr B., Westwater R. and Ice J., "Multimedia Broadcasting over the Internet: Part II-Video Compression," *IEEE Multimedia Magazine*, vol. 6, pp. 85–89, January-March 1999.

- [50] Lee G.H., Song J.H. and Park R.-H., "Three-Dimensional DCT/WT Compression Using Motion Vector Segmentation for Low Bit-Rate Video Coding," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 456–459, 1997.
- [51] Podilchuk C.I., Jayant N.S. and Farvardin N., "Three-Dimensional Subband Coding of Video," *IEEE Transactions on Image Processing*, vol. 4, pp. 125–139, February 1995.
- [52] Karlsson G. and Vetterli M., "Three-dimensional sub-band coding of video," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, (New York), pp. 1100–1103, April 1988.
- [53] Karlsson G. and Vetterli, "Sub-band coding of video for packet networks," *Optical Engineering*, vol. 27, pp. 574–586, July 1988.
- [54] Andrew R.K., Stewart B.K., Langer S.G. and Stegbauer K.C., "A 3-D Wavelet-Based Codec for Lossy Compression of Pre-Scan-Converted Ultrasound Video," in *Proceedings of the SPIE Conference on Image Display*, vol. 3658, pp. 396–405, February 1999.
- [55] Chou C.-H. and Chen C.-W., "A Perceptually Optimized 3-D Subband Codec for Video Communication over Wireless Channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 143–156, April 1996.
- [56] Taubman D. and Zakhor A., "Multirate 3-D Subband Coding of Video," *IEEE Transactions on Image Processing*, vol. 3, pp. 572–588, September 1994.
- [57] Lladós-Bernaus R. and Stevenson R.L., "Robust Low Bit-Rate 3-D Subband Codec," in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 3024, pp. 610–621, 1997.
- [58] Roese J.A., Pratt W.K. and Robinson G., "Interframe Cosine Transform Image Coding," *IEEE Transactions on Communications*, vol. 25, pp. 1329–1339, November 1977.
- [59] Mainguy A. and Wang L., "Performance Analysis of 3-D Subband Coding for Low Bit Rate Video," in *Proceedings of the SPIE*, vol. 2952, pp. 373–379, 1996.
- [60] Pearlman W.A., Kim B.-J. and Xiong Z., *Wavelet Image and Video Compression*, ch. Embedded Video Subband Coding with 3D SPIHT, pp. 397–432. Kluwer Academic Publishers, 1998.
- [61] Hamdi M., Afifi H., Rolin P. and Rojas L.M., "Designing Point-to-Point Interactive Video Applications," in *Proceedings of COST-237 International Conference on Multimedia Transport and Teleservices*, (Copenhagen), November 1995.
- [62] Kim B.-J. and Pearlman W.A., "An Embedded Wavelet Video Coder Using Three-Dimensional Set Partitioning in Hierarchical Trees (SPIHT)," in *Proceedings of the IEEE Data Compression Conference*, pp. 251–260, 1997.
- [63] Kim Y.K., Kim R.C. and Lee S.U., "On the adaptive 3D subband video coding," in *Proceedings of the SPIE - Visual Conference on Image Processing*, vol. 2727, (Orlando, FL), pp. 123–132, March 1996.
- [64] Khalil H., Atiya A.F. and Shaheen S., "Three-Dimensional Video Compression Using Subband/Wavelet Transform with Lower Buffering Requirements," *IEEE Transactions on Image Processing*, vol. 8, pp. 762–773, June 1999.
- [65] Martucci S.A., Sodagar I., Chiang T. and Zhang Y.-Q., "A Zerotree Wavelet Video Coder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 109–118, February 1997.

- [66] Klock H., Polzer A. and Buhmann J.M., "Region-Based Motion Compensated 3D-Wavelet Transform Coding of Video," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, pp. 776–779, 1997.
- [67] Choi S.-J. and Woods J.W., "Motion-Compensated 3-D Subband Coding of Video," *IEEE Transactions on Image Processing*, vol. 8, pp. 155–167, February 1999.
- [68] Wang A., Xiong Z., Chou P.A. and Mehrotra S., "Three-Dimensional Wavelet Coding of Video with Global Motion Compensation," in *Proceedings of the IEEE Data Compression Conference*, (Snowbird, UT), March 1999.
- [69] Kim B.-J. and Pearlman W.A., "Low-Delay Embedded 3-D Wavelet Color Video Coding with SPIHT," in *Proceedings of the SPIE*, vol. 3309, pp. 955–964, 1997.
- [70] Tham J.Y., Ranganath S. and Kassim A.A., "Highly Scalable Wavelet-Based Video Codec for Very Low Bit-rate Environment," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 12–27, January 1998.
- [71] Ramchandran K. and Vetterli M., "Best Wavelet Packet Bases in a Rate-Distortion Sense," *IEEE Transactions on Image Processing*, vol. 2, pp. 160–175, April 1993.
- [72] Mallat S., *A Wavelet Tour of Signal Processing*. Academic Press, 2 ed., 1999.
- [73] Samet H., "The quadtree and related hierarchical data structures," *ACM Computing Surveys*, vol. 16, no. 2, pp. 187–260, 1984.
- [74] Shusterman E. and Feder M., "Image compression via improved quadtree decomposition algorithms," *IEEE Transactions on Image Processing*, vol. 3, no. 2, pp. 207–215, 1994.
- [75] Tamminen M., "Encoding pixel trees," *Computer Graphics Image Processing*, vol. 28, pp. 44–57, May 1984.
- [76] Samet H., "A top-down quadtree traversal algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1985.
- [77] Lee W.S., "Tiling and Adaptive Image Compression," *IEEE Transactions on Information Theory*, vol. 46, pp. 1789–1799, August 2000.
- [78] Herley C., Kovacevic J., Ramchandran K. and Vetterli M., "Tilings of the Time-Frequency Plane: Construction of Arbitrary Orthogonal Bases and Fast Tiling Algorithms," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3341–3359, December 1993.
- [79] Wien M. and Niehsen W., "Space-Frequency Adaptive Coding of Motion Compensated Frame Differences," in *Proceedings of the Picture Coding Symposium*, (Portland, OR, USA), pp. 65–68, April 1999.
- [80] Smith J.R. and Chang S.-F., "Joint Adaptive Space and Frequency Basis Selection," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 702–705, 1997.
- [81] Smith J.R., *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis*. PhD thesis, Columbia University, February 1997.
- [82] Chen S.S., *Basis Pursuits*. PhD thesis, Stanford University, November 1995.
- [83] Chen S. and Donoho D.L., "Examples of Basis Pursuits," in *Proceedings of the SPIE*, vol. 2569, pp. 564–574, 1995.
- [84] Mallat S.G. and Zhang Z., "Matching Pursuits With Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, December 1993.

- [85] Davis G., *Adaptive Nonlinear Approximations*. PhD thesis, New York University, September 1994.
- [86] Vetterli M. and Kalker T., "Matching Pursuit for Compression and Application to Motion Compensated Video Coding," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, pp. 725–729, 1994.
- [87] Banham M.R. and Brailean J.C., "A Selective Update Approach to Matching Pursuits Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 119–129, February 1997.
- [88] Neff R. and Zakhor A., "Very Low Bit-Rate Video Coding Based on Matching Pursuits," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 158–171, February 1997.
- [89] De Prycker M., *Asynchronous transfer mode : solution for broadband ISDN*. Prentice Hall, third ed., 1995.
- [90] Karlsson G. and Djuknic G., "On the Efficiency of Statistical Bit Rate Service for Video," in *IFIP International Conference on Performance of Information and Communications Systems*, (Lund, Sweden), May 25–28 1998.
- [91] Leduc J.-P., *Digital Moving Pictures - Coding and Transmission on ATM Networks*. Elsevier, 1994.
- [92] Leduc J.-P., "Digital Television on ATM Networks: optimum Chain for Coding and Transmission," *IEEE Transactions on Communications*, vol. 45, pp. 829–839, July 1997.
- [93] ITU-T Study Group 13, *Recommendation I.363.1 B-ISDN ATM Adaptation Layer Specification, Type 1*. Geneva: ITU-T, August 1996.
- [94] ITU-T Study Group 13, *Recommendation I.363.5 B-ISDN ATM Adaptation Layer Specification, Type 5*. Geneva: ITU-T, August 1996.
- [95] Ortega A., Garrett M.W. and Vetterli M., "Rate constraints for video transmission over ATM networks based on joint source/network criteria," *Annals of telecommunications*, vol. 50, pp. 603–616, July–August 1995.
- [96] ITU-T, *Recommendation H.222.1 Multimedia Multiplex and Synchronization for Audiovisual Communications in ATM Environments*. Geneva: ITU-T, March 1998.
- [97] Gringeri S., Khasnabish B., Lewis A., Shuaib K., Egorov R. and Basch B., "Transmission of MPEG-2 Video Streams over ATM," *IEEE Multimedia*, pp. 58–72, May 1998.
- [98] Orzessek M. and Sommer P., *ATM and MPEG-2 - Integrating Digital Video into Broadband Networks*. Prentice Hall, 1998.
- [99] Goldman M., "MPEG over ATM," *The ATM Forum Newsletter*, vol. 3, October 1995.
- [100] Almquist P., "Type of service in the internet protocol," Tech. Rep. RFC 1349, IETF, July 1992.
- [101] Clark D. and Wroclawski J., "An Approach to Service Allocation in the Internet," tech. rep., Internet Engineering Task Force, July 1997. Internet draft, expired 12/97.
- [102] Heinanen J., "Use of the IPv4 TOS Octet to Support Differentiated Services," tech. rep., Internet Engineering Task Force, Nov. 1997. Internet draft, expired 05/98.
- [103] Braden R., Zhang L., Berson S., Herzog S. and Jamin S., "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification," Tech. Rep. RFC 2205, Internet Engineering Task Force, September 1997.

- [104] Bernet Y., "The Complementary Roles of RSVP and Differentiated Services in the Full-Service QoS Network," *IEEE Communications Magazine*, vol. 38, pp. 154–162, February 2000.
- [105] Carpenter B. and Kandlur D.D., "Diversifying Internet Delivery," *IEEE Spectrum Magazine*, vol. 36, pp. 57–61, November 1999.
- [106] Schulzrinne H. and Casner S., Frederick R. and Jacobson V., "RTP: A Transport Protocol for Real-Time Applications," Tech. Rep. RFC 1889, Internet Engineering Task Force, January 1996.
- [107] IETF, Audio/Video Transport Working Group, "RTP: A Transport Protocol for Real-Time Applications," Standard Track RFC:1889, IETF, January 1996.
- [108] IETF, Audio/Video Transport Working Group, "RTP Payload Format for MPEG1/MPEG2 Video," Internet Draft draft-ietf-avt-mpeg-new-01.txt, IETF, June 1997.
- [109] Hoffman D., Fernando G., Goyal V. and Civanlar R., "RTP Payload Format for MPEG1/MPEG2 Video," Tech. Rep. RFC 2250, Internet Engineering Task Force, January 1998.
- [110] Schulzrinne H., Rao A. and Lanphier R., "Real Time Streaming Protocol (RTSP)," Tech. Rep. RFC 2326, IETF, 1998.
- [111] Zheng H. and Liu K.J.R., "Multimedia Services over Digital Subscriber Lines," *IEEE Signal Processing Magazine*, vol. 17, pp. 44–60, July 2000.
- [112] Cioffi J.M., Oksman V., Werner J.-J., Pollet T., Spruyt P.M.P., Chow J.S. and Jacobsen K.S., "Very-High-Speed Digital Subscriber Lines," *IEEE Communications Magazine*, vol. 37, pp. 72–79, April 1999.
- [113] Rabbat R. and Siu K.-Y., "QoS Support for Integrated Services over CATV," *IEEE Communications Magazine*, vol. 37, pp. 64–68, January 1999.
- [114] Fox B., "Digital TV comes down to earth," *IEEE Spectrum Magazine*, vol. 35, pp. 23–29, October 1998.
- [115] Dutta-Roy A., "Cable it's not just for TV," *IEEE Spectrum Magazine*, vol. 36, pp. 53–59, May 1999.
- [116] Black U., *Mobile and wireless network*. Prentice Hall, 1996.
- [117] Johnston W., "Europe's future mobile telephony system," *IEEE Spectrum Magazine*, vol. 35, pp. 49–53, October 1998.
- [118] Oliphant M.W., "The mobile phone meets the Internet," *IEEE Spectrum Magazine*, vol. 36, pp. 20–28, August 1999.
- [119] Goodman D.J., "The Wireless Internet: Promises and Challenges," *IEEE Computers Magazine*, vol. 33, pp. 36–41, July 2000.
- [120] Davisson L. D., "Rate-Distortion Theory and Application," *Proceedings of the IEEE*, vol. 60, pp. 800–808, July 1972.
- [121] Berger T. and Gibson J.D., "Lossy Source Coding," *IEEE Transactions on Information Theory*, vol. 44, pp. 2693–2723, October 1998.
- [122] Sullivan G.J. and Wiegand T., "Rate-Distortion Optimization for Video Compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, November 1998.

- [123] Lee S. and Madiseti V.K., "Parameter Optimization of Robust Low-Bit-Rate Video Coders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 849–855, September 1999.
- [124] Lin L.-J. and Ortega A., "Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 446–459, August 1998.
- [125] Sun H., Kwok W., Chien M. and John Ju C. H., "MPEG Coding Performance Improvement by Jointly Optimizing Coding Mode Decisions and Rate Control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 449–458, June 1997.
- [126] Richardson I.E.G. and Riley M.J., "Controlling the Rate of MPEG Video by Dynamic Variation of Sequence Structure," in *International Picture Coding Symposium*, (Melbourne), pp. 471–476, March 1996.
- [127] Smoot S.R., *Maximizing Perceived Quality Given Bit-rate Constraints in MPEG Encoding through Content-adaptivity*. PhD thesis, University of California, Berkeley, 1998.
- [128] Ortega A. and Ramchandran K., "Rate-Distortion Methods for Image and Video Compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 23–50, November 1998.
- [129] Cho N.I., Lee H. and Lee S.U., "An Adaptive Quantization Algorithm for Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 527–535, June 1999.
- [130] Favalli L. and Mecocci A., "Bandwidth Renegotiation for MPEG-2 Sources in ATM Networks," in *2nd Erlangen Symposium*, (Erlangen), pp. 25–33, April 1997.
- [131] Lee J. and Dickinson B. W., "Rate-Distortion Optimized Frame Type Selection for MPEG Encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 501–510, June 1997.
- [132] Lin F.-H. and Mersereau R.M., "Rate-quality tradeoff MPEG video encoder," *Signal Processing: Image Communications*, vol. 14, pp. 297–309, February 1999.
- [133] Verscheure O. and van den Branden Lambrecht C.J., "Adaptive Quantization Using a Perceptual Visibility Predictor," in *Proceedings of the International Conference on Image Processing*, (Santa Barbara), October 1997.
- [134] Saw Y.-S., *Rate-Quality Optimized Video Coding*. No. 0-7923-8376-1, Kluwer Academic Publishers, 1999.
- [135] Schuster G.M. and Katsaggelos A.K., *Rate-distortion based video compression: optimal video frame compression and object boundary encoding*. Kluwer Academics Publishers, 1997.
- [136] Lakshman T.V., Ortega A. and Reibman A.R., "VBR Video: Tradeoffs and Potentials," *Proceedings of the IEEE*, vol. 86, pp. 952–973, May 1998.
- [137] Floyd S. and Fall K., "Promoting the Use of End-to-End Congestion Control in the Internet," *IEEE/ACM Transactions on Networking*, vol. 7, pp. 458–472, August 1999.
- [138] Jacobs S. and Eleftheriadis A., "Streaming Video Using Dynamic Rate Shaping and TCP Congestion Control," *Journal of Visual Communication and Image Representation*, vol. 9, pp. 211–222, September 1998.
- [139] Kim Y.-G., Kim J. and Kuo C.-C. J., "Smooth and Fast Rate Adaptation and Network-Aware Error Control for TCP-friendly Internet Video Transmission," *IEEE Transactions on Circuits and Systems for Video Technology*, 2000. submitted paper.

- [140] Rexford J. and Towsley D., "Smoothing Variable-Bit-Rate Video in an Internetwork," *IEEE/ACM Transactions on Networking*, vol. 7, pp. 202–215, April 1999.
- [141] Zhang L., "Statistics of Cell Loss and its Application for Forward Error Recovery in ATM Network," in *Proceedings of the IEEE ICC Conference*, vol. 2, (New York), pp. 694–698, 1992.
- [142] Cidon I., Khamisy A. and Sidi M., "Analysis of Packet Loss Processes in High-Speed Networks," *IEEE Transactions on Information Theory*, vol. 39, pp. 98–108, January 1993.
- [143] Bolot J.-C., "End-to-End Packet Delay and Loss Behavior in the Internet," *ACM SIGCOMM Computer Communication Review*, vol. 23, pp. 289–298, October 1993.
- [144] Bolot J.-C., "Characterizing End-to-End Packet Delay and Loss in the Internet," *Journal of High Speed Networks*, vol. 2, pp. 304–323, 1993.
- [145] Cohen D. M. and Heyman D. P. and Reibman A. M., "Comparisons among Models of Cell Losses for Video Conferences," in *Sixth International Workshop on Packet Video*, (Portland, USA), Sept. 1994.
- [146] Borella M.S., "Measurement and Interpretation of Internet Packet Loss," *Journal of Communications and Networking*, June 2000.
- [147] Jeruchim M.C., Balaban P. and Shanmugan K.S., *Simulation of communication systems*. Plenum Press, 1992.
- [148] Paxson V., "End-to-End Internet Packet Dynamics," *ACM SIGCOMM Computer Communication Review*, vol. 27, pp. 139–152, October 1997.
- [149] Paxson V., "End-to-End Routing Behavior in the Internet," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 601–615, October 1997.
- [150] Paxson V., "End-to-End Internet Packet Dynamics," *IATN*, vol. 7, pp. 277–292, June 1999.
- [151] Yajnik M., Kurose J. and Towsley D., "Packet Loss Correlation in the Mbone Multicast Network: Experimental Measurements and Markov Chain Models," in *Proceedings of the IEEE Global Internet Conference*, (London), November 1996.
- [152] Zamora J., Anastassiou D. and Chang S.-F., "Objective and Subjective Quality of Service Performance of Video-on-Demand in ATM-WAN," *Signal Processing: Image Communications*, vol. 14, pp. 635–654, May 1999.
- [153] Zhang Y.-Q. and Lee X., "Performance of MPEG Codecs in the Presence of Errors," *Journal of Visual Communication and Image Representation*, vol. 5, pp. 378–387, December 1994.
- [154] Chui D., Cai L.N., McCutcheon M., Ito M.R., Neufeld G.W., "Transport of MPEG-2 Video in a Routed IP Network: Effects of Network Congestion on Transport Stream Errors," in *Proceedings of the IDMS Conference*, (Toulouse, France), October 12–15 1999.
- [155] Chen C.-T., "Error Detection and Concealment with an Unsupervised MPEG2 Video Decoder," *Journal of Visual Communication and Image Representation*, vol. 6, pp. 265–279, September 1995.
- [156] Maxted J.C. and Robinson J.P., "Error Recovery for Variable Length Codes," *IEEE Transactions on Information Theory*, vol. 31, pp. 794–801, November 1985.
- [157] CCIR Recommendation 500-3, *Method for the Subjective Assessment of the Quality of Television Pictures*, 1990.

- [158] Wandell B.A., *Foundations of vision*. 0-87893-853-2, Sinauer Associates, Inc, Publishers, 1995.
- [159] Chen M.-J., Chen L.-G. and Weng R.-M., "Error Concealment of Lost Motion Vectors with Overlapped Motion Compensation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 560–563, June 1997.
- [160] Yu T.-H. and Mitra S.K., "Novel nonlinear predictive image coding algorithm," *Journal of Electronic Imaging*, vol. 6, pp. 181–188, April 1997.
- [161] Choi J.-G., Lee S.-W. and Kim S.-D., "Spatio-Temporal Video Segmentation Using a Joint Similarity Measure," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 279–285, April 1997.
- [162] Voran S., "The Development of Objective Video Quality Measures that emulates Human Perception," in *Proceedings of the IEEE GLOBECOM Conference*, pp. 1776–1781, 1991.
- [163] Webster A.A., Jones C.T., Pinson M.H., Voran S.D., Wolf S., "An objective video quality assessment system based on human perception," in *Proceedings of the SPIE - Human Vision, Visual Processing, and Digital Display IV*, vol. 1913, pp. 15–26, SPIE, 1993.
- [164] Wolf S., Pinson M.H., Voran S.D. and Webster A.A., "Objective Quality Assessment of Digitally Transmitted Video," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pp. 477–481, May 1991.
- [165] Taylor C. C., Pizlo Z., Allebach J. P., Bouman C. A., "Image Quality Assessment with a Gabor Pyramid Model of the Human Visual System," in *Proceedings of the SPIE: Human vision and electronic imaging 2*, (San Jose), pp. 58–69, February 1997.
- [166] Daly S., *Digital images and Human Vision*, ch. The visible differences predictor: An algorithm for the assessment of image fidelity, pp. 179–205. Cambridge, MA: MIT Press, 1993.
- [167] Lubin J. and Fibush D., *Sarnoff JND Vision Model*, August 5 1997.
- [168] Tan K.T., Ghanbari M. and Pearson D.E., "A Video Distortion Meter," in *Picture Coding Symposium*, (Berlin), pp. 119–122, September 1997.
- [169] Lubin J., Brill M.H. and Crane R.L., "Vision Model-Based Assessment of Distortion Magnitudes in Digital Video," tech. rep., David Sarnoff Research Center, 1996. <http://www.mpeg.org/MPEG/JND/index.html>.
- [170] van den Branden Lambrecht C.J. and Verscheure O., "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System," in *Proceedings of the SPIE*, vol. 2668, (San Jose, USA), pp. 450–461, February 1996.
- [171] van den Branden Lambrecht C. J., *Perceptual Models and Architectures for Video Coding Applications*. PhD thesis, Swiss Federal Institute of Technology, CH 1015 Lausanne, Switzerland, 1996.
- [172] Lindh P. and van den Branden Lambrecht C.J., "Efficient Spatio-Temporal Decomposition for Perceptual Processing of Video Sequences," in *Proceedings of the IEEE International Conference on Image Processing*, (Lausanne), September 1996.
- [173] van den Branden Lambrecht C.J., Costantini D.M., Sicuranza G.L. and Kunt M., "Quality Assessment of Motion Rendition in Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 766–782, August 1999.
- [174] Winkler S., "A Perceptual Distortion Metric for Digital Color Video," in *Human Vision and Electronic Imaging*, vol. 3644, (San Jose, CA), January 23–29 1999.

- [175] Dalgic I. and Tobagi F.A., "Glitches as a Measure of Video Quality Degradation Caused by Packet Loss," in *7th International Workshop on Packet Video*, (Brisbane, Australia), pp. 201–206, March 1996.
- [176] Dalgic I., *Performance of Ethernet and ATM Networks Carrying Video Traffic Based on Accurate Characterization of Video Sources*. PhD thesis, Stanford University, August 1996.
- [177] Wang Y., Wenger S., Wen J. and Katsaggelos A.K., "Error Resilient Video Coding Techniques," *IEEE Signal Processing Magazine*, vol. 17, pp. 61–82, July 2000.
- [178] Wang Y. and Zhu Q.-F., "Error Control and Concealment for Video Communication: A Review," *Proceedings of the IEEE*, vol. 86, pp. 974–997, May 1998.
- [179] Sayood K. and Borkenhagen J.C., "Use of Residual Redundancy in the Design of Joint Source/Channel Coders," *IEEE Transactions on Communications*, vol. 39, pp. 838–846, June 1991.
- [180] Hsu C.-Y., Ortega A. and Khansari M., "Rate Control for Robust Video Transmission over Burst-Error Wireless Channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 756–772, May 1999.
- [181] Villasenor J.D., Zhang Y.-Q. and Wen J., "Robust Video Coding Algorithms and Systems," *Proceedings of the IEEE*, vol. 87, pp. 1724–1733, October 1999.
- [182] Arnold J.F., Frater M.R. and Zhang J., "Error resilience in the MPEG-2 video coding standard for cell based networks - A review," *Signal Processing: Image Communications*, vol. 14, pp. 607–633, 1999.
- [183] Zhang J., Frater M.R., Arnold J.F. and Percival T.M., "MPEG 2 Video Services for Wireless ATM Networks," *Journal on Selected Areas in Communications*, vol. 15, pp. 119–128, January 1997.
- [184] Richardson I. E. G. and Riley M. J., "Improving the error tolerance of MPEG video by varying slice size," *Signal Processing*, vol. 46, pp. 369–72, 1995.
- [185] Richardson I.E.G. and Riley M.J., "Varying slice size to improve error tolerance of MPEG video," in *Proceedings the SPIE*, vol. 2668, pp. 365–371, 1996.
- [186] Willebeek-LeMair M. H., Shae Z.-Y. and Chang Y.-C., "Robust H.263 Video Coding for Transmission over the Internet," in *Proceedings of the IEEE INFOCOM Conference*, (San Francisco), April 1998.
- [187] Riley M.J. and Richardson I.E.G., *Digital Video Communications*. 0-89006-890-9, Artech House, 1997.
- [188] Wenger S., "Video Redundancy Coding in H.263+," in *International Workshop on Audio-Visual Services over Packet Networks*, (Aberdeen, UK), pp. 23–28, September 1997.
- [189] Cote G., Erol B., Gallant M. and Kossentini F., "H.263+: Video Coding at Low Bit Rates," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 849–866, November 1998.
- [190] Wenger S., Knorr G., Ott J. and Kossentini F., "Error Resilience Support in H.263+," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 867–877, November 1998.
- [191] Moccagatta I., Soudagar S., Liang J. and Chen H., "Error-Resilient Coding in JPEG-2000 and MPEG-4," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 899–914, June 2000.

- [192] Talluri R., Oehler K., Bannon T., Courtney J.D., Das A. and Liao J., "A Robust, Scalable, Object-Based Video Compression Technique for Very Low Bit-Rate Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 221–233, February 1997.
- [193] Aravind R., Civanlar M.R. and Reibman A.R., "Packet Loss Resilience of MPEG-2 Scalable Video Coding Algorithms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 426–435, October 1996.
- [194] Van Dyck R.E. and Ganti H.V., "Wavelet video transmission over wireless channels," *Signal Processing: Image Communications*, vol. 12, pp. 135–145, April 1998.
- [195] Man H., Kossentini F. and Smith M.J.T., "A Family of Efficient and Channel Error Resilient Wavelet/Subband Image Coders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 95–122, February 1999.
- [196] Man H., Kossentini F. and Smith M.J.T., "Robust EZW Image Coding for Noisy Channels," *IEEE Signal Processing Letters*, vol. 4, pp. 227–229, August 1997.
- [197] Creusere C.D., "A New Method of Robust Image Compression Based on the Embedded Zerotree Wavelet Algorithm," *IEEE Transactions on Image Processing*, vol. 6, pp. 1437–1442, October 1997.
- [198] Lopez Fernandez C., Basso A. and Hubaux J.-P., "Error concealment and early resynchronization techniques for MPEG-2 video streams damaged by transmission over ATM networks," in *Proceedings of the SPIE*, vol. 2668, pp. 372–383, 1996.
- [199] Lee W.S., Pickering M.R., Frater M.R. and Arnold J.F., "Error Resilience in Video and Multiplexing Layers for Very Low Bit-Rate Video Coding Systems," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1764–1774, December 1997.
- [200] Atzori L. and De Natale F.G.B., "Error concealment in video transmission over packet networks by a sketch-based approach," *Signal Processing: Image Communications*, vol. 15, pp. 57–76, 1999.
- [201] Shyu H.-C. and Leou J.-J., "Detection and Concealment of Transmission Errors in MPEG-2 Images - A Genetic Algorithm Approach," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 937–948, September 1999.
- [202] Zeng W. and Liu B., "Geometric-Structure-Based Error Concealment with Novel Applications in Block-Based Low-Bit-Rate Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 648–665, June 1999.
- [203] Zhang Y. and Ma K.-K., "Error Concealment for Image Transmission by Multiscale Markov Random Field Modeling," in *Proceedings of the IEEE ICME*, (New York, USA), July 2000.
- [204] Salama P., Shroff N.B. and Delp E.J., "Error Concealment in MPEG Video Streams over ATM Networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1129–1144, June 2000.
- [205] Shirani S., Kossentini F. and Ward R., "A Concealment Method for Video Communications in an Error-Prone Environment," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1122–1128, June 2000.
- [206] Chang J.-H., Kim J.-K. and Lee C.W., "A Projections onto the Overcomplete Basis Approach for Block Loss Recovery," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, pp. 93–96, 1997.

- [207] Yu G.-S., Lui M. M.-K. and Marcellin W., "POCS-Based Error Concealment for Packet Video Using Multiframe Overlap Information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 422–434, August 1998.
- [208] Aign S., Panis S., Stockhammer T., Jakob T. and Vlaicu R., "MPEG-4 Video Transmission via DAB: Error Detection and Error Concealment," in *Proceedings of the Picture Coding Symposium*, (Berlin), pp. 731–736, September 1997.
- [209] Aign S., "A Temporal Error Concealment Technique for I-Pictures in an MPEG-2 Video-Decoder," in *Proceedings of the SPIE - Visual Communication and Image Processing*, vol. 3309, pp. 405–416, January 1998.
- [210] Ghanbari M. and Seferidis V., "Cell-Loss Concealment in ATM Video Codecs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 238–247, June 1993.
- [211] Zhang J., Arnold J.F. and Frater M.R., "Improved Motion-Compensated Concealment for MPEG2 Coded Video," in *International Workshop on Audio-Visual Services over Packet Networks*, (Aberdeen, UK), pp. 11–16, September 1997.
- [212] Lim C.P., Tan E.A.W. and Ghanbari M., "Packetization and Cell Loss Concealment of Packet Video," in *International Workshop on Audio-Visual Services over Packet Networks*, (Aberdeen, UK), pp. 1–4, September 1997.
- [213] Cuenca P., Garrido A., Quiles F. and Orozco-Barbosa L., "An Efficient Protocol Architecture for Error-Resilient MPEG-2 Video Communications over ATM Networks," *IEEE Transactions on Broadcasting*, vol. 45, pp. 129–140, March 1999.
- [214] Salama P., Shroff N. and Delp E.J., "Error Concealment in Embedded Zerotree Wavelet Codecs," in *Proceedings of the International Workshop on Very Low Bit Rate Video Coding*, (Urbana, IL), pp. 200–203, October 1998.
- [215] Costello D.J., Hagenauer J., Imai H. and Wicker S.B., "Applications of Error-Control Coding," *IEEE Transactions on Information Theory*, vol. 44, pp. 2531–2559, October 1998.
- [216] Liu H., Ma H., El Zarki M. and Gupta S., "Error control schemes for networks: An overview," *Mobile Networks and Applications*, vol. 2, no. 2, pp. 167–182, 1997.
- [217] Ohta H. and Kitami T., "A Cell Loss Recovery Method Using FEC in ATM Networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 1471–1483, December 1991.
- [218] Rizzo L., "Effective Erasure Codes for Reliable Computer Communication Protocols," *ACM Computer Communication Review*, vol. 27, pp. 24–36, April 1997.
- [219] Rizzo L., "On the feasibility of software FEC," tech. rep., University of Pisa, 1997. <http://www.iet.unipi.it/~luigi/softfec.ps>.
- [220] Lei S.-M., "Forward Error Correction Codes for MPEG2 over ATM," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 200–203, April 1994.
- [221] Kousa M.A., Elhakeem A.K. and Yang H., "Performance of ATM Networks Under Hybrid ARQ/FEC Error Control Scheme," *IEEE/ACM Transactions on Networking*, vol. 7, pp. 917–925, December 1999.
- [222] Rice M. and Wicker S.B., "Adaptive Error Control For Slowly Varying Channels," *IEEE Transactions on Communications*, vol. 42, pp. 917–925, February/March/April 1994.
- [223] Butto M., Colombo G. and Toniatti A., "Delay distribution in a data communication system with error recovery," Vol. VII 3, CSELT Technical Report, September 1979.

- [224] IETF, Audio/Video Transport Working Group, "An A/V Profile Extension for Generic Forward Error Correction in RTP," Internet Draft draft-ietf-avt-fec-00.txt, IETF, July 1997.
- [225] Redmill D.W. and Kingsbury N.G., "The EREC: An Error-Resilient Technique for Coding Variable-Length Blocks of Data," *IEEE Transactions on Image Processing*, vol. 5, pp. 565–574, April 1996.
- [226] Swann R. and Kingsbury N., "Bandwidth efficient transmission of MPEG-2 video over noisy mobile links," *Signal Processing: Image Communications*, vol. 12, pp. 105–115, April 1998.
- [227] Boyce J.M. and Gaglianella R.D., "Packet Loss Effects on MPEG Video Sent Over the Public Internet," in *Proceedings of the ACM Multimedia Conference*, (Bristol, UK), pp. 181–190, 1998.
- [228] Wah B.W. and Lin D., "Transformation-Based Reconstruction for Real-Time Voice Transmissions over the Internet," *IEEE Transactions on Multimedia*, vol. 1, pp. 342–351, December 1999.
- [229] Tan W.-T. and Zakhori A., "Real-Time Internet Video Using Error Resilient Scalable Compression and TCP-Friendly Transport Protocol," *IEEE Transactions on Multimedia*, vol. 1, pp. 172–186, June 1999.
- [230] Cherriman P. and Hanzo L., "Robust H.263 Video Transmission over mobile channels in interference limited environments," in *IEEE Wireless Video Communications Workshop*, (Loughborough, UK), September 1996.
- [231] Garrett M.W. and Vetterli M., "Joint Source/Channel Coding of Statistically Multiplexed Real-Time Services on Packet Networks," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 71–80, February 1993.
- [232] Albanese A., Blomer J., Edmonds J., Luby M. and Sudan M., "Priority Encoding Transmission," *IEEE Transactions on Information Theory*, vol. 42, pp. 1737–1744, November 1996.
- [233] Shin J., Kim J. and Kuo C.-C. J., "Quality of Service Mapping Mechanism for Packet Video in Differentiated Service Networks," *IEEE Transactions on Multimedia*, 2000. submitted paper.
- [234] Verscheure O., Garcia Adanez X., Karlsson G. and Hubaux J.-P., "User-oriented qos in packet video delivery," *IEEE Network Magazine*, vol. 12, pp. 12–21, November/December 1998.
- [235] Ramchandran K. and Vetterli M., *Wireless Communications: A Signal Processing Perspective*, ch. Multiresolution Joint Source-Channel Coding for Wireless Channels, pp. 1–50. Prentice-Hall, 1998.
- [236] Park K. and Wang W., "AFEC: An Adaptive Forward Error Correction Protocol for End-to-End Transport of Real-Time Traffic," in *Proceedings of the IEEE IC3N*, (Los Alamitos, CA), pp. 196–205, 1998.
- [237] Park K. and Wang W., "QoS-Sensitive Transport of Real-Time MPEG Video using Adaptive Forward Error Correction," in *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, vol. 2, (Florence, Italy), pp. 426–432, June 1999.
- [238] Bystrom M., Parthasarathy V. and Modestino J.W., "Hybrid Error Concealment Schemes for Broadcast Video Transmission over ATM Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 868–881, September 1999.
- [239] Zhang T. and Xu Y., "Unequal Packet Loss Protection for Layered Video Transmission," *IEEE Transactions on Broadcasting*, vol. 45, pp. 243–252, June 1999.

- [240] Liu H. and El Zarki M., "Transmission of video telephony images over wireless channels," *Wireless Networks*, vol. 2, no. 2, pp. 219–228, 1996.
- [241] Stuhlmuller K., Link M., Girod B. and Horn U., "Scalable Internet Video Streaming With Unequal Error Protection," in *Proceedings of the Packet Video Workshop*, (New York), April 1999.
- [242] Ayanoglu E., Pancha P., Reibman A.R. and Talwar S., "Forward Error Control for MPEG-2 Video Transport in a Wireless ATM LAN," *Mobile Networks and Applications*, vol. 1, pp. 245–257, 1996.
- [243] Stedman R., Gharavi H., Hanzo L. and Steele R., "Transmission of Subband-Coded Images via Mobile Channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 15–26, February 1993.
- [244] Bystrom M. and Modestino J.W., "Combined Source-Channel Coding Schemes for Video Transmission over an Additive White Gaussian Noise Channel," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 880–890, June 2000.
- [245] Saporilla D. and Ross K.W., "Optimal Streaming of Layered Video," in *Proceedings of the IEEE INFOCOM Conference*, March 2000. submitted paper.
- [246] Lu J., Liou M.L., Letaief K.B. and Chuang J.C.-I., "Mobile image transmission using combined source and channel coding with low complexity concealment," *Signal Processing: Image Communications*, vol. 12, pp. 87–104, April 1998.
- [247] Batra P. and Chang S.-F., "Effective algorithms for video transmission over wireless channels," *Signal Processing: Image Communications*, vol. 12, pp. 147–166, April 1998.
- [248] Garcia Adanez X., *Designing New Network Adaptation and ATM Adaptation Layers for interactive Multimedia Applications*. PhD thesis, Swiss Federal Institute of Technology, 1998.
- [249] van der Schaar M., Radha H. and Dufour C., "Scalable MPEG-4 Video Coding with Graceful Packet-Loss Resilience over Bandwidth-Varying Networks," in *Proceedings of the IEEE ICME Conference*, (New York), July 2000.
- [250] Mohr A.E., Riskin E.A. and Ladner R.E., "Graceful Degradation over Packet Erasure Channels through Forward Error Correction," in *Proceedings of the IEEE Data Compression Conference*, March 1999.
- [251] Mohr A.E., Riskin E.A. and Ladner R.E., "Unequal Loss Protection: Graceful Degradation of Image Quality over Packet Erasure Channels Through Forward Error Correction," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 819–828, June 2000.
- [252] Chande V. and Farvardin N., "Progressive Transmission of Images over Memoryless Noisy Channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 850–860, June 2000.
- [253] Turletti T. and Huitema C., "Videoconferencing on the Internet," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 340–351, June 1996.
- [254] Servetto S.D. and Nahrstedt K., "Broadcast Quality Video over IP," *IEEE Journal on Selected Areas in Communications*, September 1999. submitted paper.
- [255] Wu D., Hou Y.T., Li B., Zhu W., Zhang Y.-Q. and Chao H.J., "An End-to-End Approach for Optimal Mode Selection in Internet Video Communication: Theory and Application," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 977–995, June 2000.

- [256] Stuhlmüller K., Farber N., Link M. and Girod B., "Analysis of Video Transmission over Lossy Channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1012–1032, June 2000.
- [257] Aho O.A. and Juhola J., "Error resilience techniques for MPEG-2 compressed video signal," in *IEE International Broadcasting Convention (IBC)*, pp. 327–332, January 1994.
- [258] Hochwald B. and Zeger K., "Tradeoff Between Source and Channel Coding," *IEEE Transactions on Information Theory*, vol. 43, pp. 1412–1424, September 1997.
- [259] Hochwald B., "Tradeoff Between Source and Channel Coding on a Gaussian Channel," *IEEE Transactions on Information Theory*, vol. 44, pp. 3044–3055, November 1998.
- [260] Salamati M.R., "Joint source-channel coding applied to multimedia transmission over lossy packet network," in *Proceedings of the Packet Video Workshop*, (New York), April 1999.
- [261] Parthasarathy V., Modestino J. W. and Vastola K. S., "Design of a Transport Coding Scheme for High-Quality Video over ATM Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 358–376, April 1997.
- [262] Cheung G. and Zakhori A., "Joint Source/Channel Coding of Scalable Video over Noisy Channels," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, (Lausanne, Switzerland), pp. 767–770, 1996.
- [263] Srinivasan M., "Adaptive Source-Channel Subband Video Coding for Wireless Channels," in *IEEE Workshop on Multimedia Signal Processing*, (Princeton, New Jersey, USA), pp. 407–412, June 1997.
- [264] Parthasarathy V., Modestino J.W. and Vastola K.S., "Reliable Transmission of High-Quality Video over ATM Networks," *IEEE Transactions on Image Processing*, vol. 8, pp. 361–374, March 1999.
- [265] Zhang R., Regunathan S.L. and Rose K., "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 966–976, June 2000.
- [266] Cote G., Shirani S. and Kossentini F., "Optimal Mode Selection and Synchronization for Robust Video Communications over Error-Prone Networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 952–965, June 2000.
- [267] Cote G. and Kossentini F., "Optimal intra coding of blocks for robust video communication over the Internet," *Signal Processing: Image Communications*, vol. 15, pp. 25–34, 1999.
- [268] Liao J.Y. and Villasenor J.D., "Adaptive Intra Update for Video Coding over Noisy Channels," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 763–766, 1996.
- [269] Zamir R., "Gaussian Codes and Shannon Bounds for Multiple Descriptions," in *Proceedings of the IEEE International Symposium on Information Theory*, (Cambridge, MA, USA), p. 429, August 1998.
- [270] Marutyan R.S., "Rates for Multiple Description of a Source with given Error Exponents and Distortion Levels," *Problems of Information Transmission*, vol. 26, pp. 68–74, January 1990. Translated from Problemy Peredachi Informatsii.
- [271] Zhang Z. and Berger T., "Multiple Description Source Coding with No Excess Marginal Rate," *IEEE Transactions on Information Theory*, vol. 41, pp. 349–357, March 1995.
- [272] Ahlswede R., "The Rate-Distortion Region for Multiple Descriptions Without Excess Rate," *IEEE Transactions on Information Theory*, vol. 31, pp. 721–726, November 1985.

- [273] Vaishampayan V.A., "Design of Multiple Description Scalar Quantizers," *IEEE Transactions on Information Theory*, vol. 39, May 1993.
- [274] Srinivasan M. and Chellappa R., "Multiple Description Subband Coding," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, pp. 684–688, 1998.
- [275] Servetto S.D., Ramchandran K., Vaishampayan V.A. and Nahrstedt K., "Multiple Description Coding Wavelet Based Image Coding," *IEEE Transactions on Image Processing*, vol. 9, pp. 813–826, May 1999.
- [276] Miguel A.C., Mohr A.E. and Riskin E.A., "SPIHT for Generalized Multiple Description Coding," in *Proceedings of the IEEE International Conference on Image Processing*, 1999.
- [277] Battlo J.-C. and Vaishampayan V.A., "Asymptotic Performance of Multiple Description Transform Codes," *IEEE Transactions on Information Theory*, vol. 43, pp. 703–706, March 1997.
- [278] Jiang W. and Ortega A., "Multiple Description Coding via Polyphase Transform and Selective Quantization," in *Proceedings of Visual Communications and Image Processing, VCIP'99*, (San Jose, USA), January 1999.
- [279] Ridge J., Ware F.W. and Gibson J.D., "Permuted Smoothed Descriptions and Refinement Coding of Images," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 915–926, June 2000.
- [280] Chung D.-M. and Wang Y., "Multiple Description Image Coding Using Signal Decomposition and Reconstruction Based on Lapped Orthogonal Transforms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 895–908, September 1999.
- [281] Arean R., Kovacevic J. and Goyal V.K., "Multiple Description Perceptual Audio Coding with Correlating Transforms," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 140–145, March 2000.
- [282] Jiang W. and Ortega A., "Multiple Description Coding via Scaling-Rotation Transform," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Phoenix, AZ), March 1999.
- [283] Servetto S.D., Vaishampayan V.A. and Sloane N.J.A., "Multiple Description Lattice Vector Quantization," in *Proceedings of the IEEE Data Compression Conference*, p. 1999, 1999.
- [284] Servetto S.D., Nahrstedt K., Ramchandran K., Vaishampayan V., "Video Streaming over the Public Internet: Multiple Description Codes and Adaptive Transport Protocols," in *Proceedings of the IEEE International Conference on Image Processing*, 1999.
- [285] Lee W.S., Frater M.R., Pickering M.R. and Arnold J.F., "A Diversity-based Scheme for Reducing Error Propagation in Video," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 582–585, 1997.
- [286] Vetterli M. and Uz K.M., "Multiresolution coding techniques for digital video: a review," *Special Issue on Multidimensional Processing of Video Signals, Multidimensional Systems and Signal Processing*, vol. 3, no. 3, pp. 161–187, 1992. No. 3.
- [287] Gonzales C. and Viscito E., "Flexibly Scalable Digital Video Coding," *Signal Processing: Image Communications*, vol. 5, no. 1-2, pp. 5–20, 1993.
- [288] Chiang T. and Anastassiou D., "Hierarchical Coding of Digital Television," *IEEE Communications Magazine*, vol. 32, pp. 38–45, May 1994.

- [289] Chiang T. and Anastassiou D., "Hierarchical HDTV/SDTV Compatible Coding Using Kalman Statistical Filtering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 424–437, April 1999.
- [290] Ramchandran K., Ortega A. and Vetterli M., "Bit Allocation for Dependent Quantization with Applications to Multiresolution and MPEG Video Coders," *IEEE Transactions on Image Processing*, vol. 3, pp. 533–545, September 1994.
- [291] Amir E., McCanne S. and Vetterli M., "A layered DCT coder for Internet Video," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, (Lausanne), pp. 13–16, September 1996.
- [292] Huang Y., Dreizen H.M. and Galatsanos N.P., "Prioritized DCT for Compression and Progressive Transmission of Images," *IEEE Transactions on Image Processing*, vol. 1, pp. 477–487, October 1992.
- [293] Eleftheriadis A. and Anastassiou D., "Optimal Data Partitioning of MPEG-2 Video," *IEEE Transactions on Circuits and Systems for Video Technology*, 1996. submitted paper.
- [294] Radha H., Chen Y., Parthasarathy K. and Cohen R., "Scalable Internet video using MPEG-4," *Signal Processing: Image Communications*, vol. 15, pp. 95–126, 1999.
- [295] Burt P.J. and Adelson E.H., "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions on Communications*, vol. 31, pp. 532–540, April 1983.
- [296] Vetterli M., "Multi-dimensional sub-band coding: Some theory and algorithms," *Signal Processing*, vol. 6, pp. 97–112, February 1984.
- [297] Woods J.W. and O'Neil S.D., "Subband Coding of Images," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 1278–1288, October 1986.
- [298] Mallat S.G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989.
- [299] DeVore R.A., Jawerth B. and Lucier B.J., "Image Compression Through Wavelet Transform Coding," *IEEE Transactions on Information Theory*, vol. 38, pp. 719–746, March 1992.
- [300] Ramchandran K., Vetterli M. and Herley C., "Wavelets, Subband Coding, and Best Bases," *Proceedings of the IEEE*, vol. 84, pp. 541–560, April 1996.
- [301] Ramchandran K., Ortega A., Uz M. and Vetterli M., "Multiresolution Broadcast for Digital HDTV Using Joint Source/Channel Coding," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 6–23, January 1993.
- [302] Kim S., Rhee S., Jeon J.G. and Park K.T., "Interframe Coding Using Two-Stage Variable Block-Size Multiresolution Motion Estimation and Wavelet Decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 399–410, August 1998.
- [303] Lee J.-Y., Kim T.-H. and Ko S.-J., "Layered video coding for multimedia applications," in *Proceedings of the SPIE, Multimedia Systems and Applications*, vol. 3528, (Boston, MA), pp. 509–517, November 1998.
- [304] Bierling M., "Displacement estimation by hierarchical block matching," in *Proceedings of the SPIE: Visual Communications and Image Processing*, vol. 1001, pp. 942–951, SPIE, 1988.
- [305] Zhang Y.-Q. and Zafar S., "Motion-Compensated Wavelet Transform Coding for Color Video Compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, pp. 285–296, September 1992.

- [306] Conklin G.J. and Hemami S.S., "A Comparison of Temporal Scalability Techniques," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 909–919, September 1999.
- [307] Wang Q. and Ghanbari M., "Scalable Coding of Very High Resolution Video Using the Virtual Zerotree," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 719–727, October 1997.
- [308] Ramos M.G. and Hemami S.S., "Perceptually based scalable image coding for packet networks," *SPIE Journal of Electronic Imaging*, vol. 7, pp. 453–463, July 1998.
- [309] Equitz W.H.R. and Cover T.M., "Successive Refinement of Information," *IEEE Transactions on Information Theory*, vol. 37, pp. 269–275, March 1991.
- [310] Dreizen H.M., "Content-Driven Progressive Transmission of Grey-Scale Images," *IEEE Transactions on Communications*, vol. 35, pp. 289–296, March 1987.
- [311] Simon S. and De Vriendt J., "Progressive Image Transmission with Run-Length Coding," *Journal on Communications*, vol. 45, pp. 45–50, May-June 1994.
- [312] Sherwood P.G. and Zeger K., "Progressive Image Coding on Noisy Channels," *IEEE Signal Processing Letters*, vol. 4, pp. 189–191, July 1997.
- [313] Chen Y. and Pearlman W.A., "Three-Dimensional Subband Coding of Video Using the Zero-Tree Method," in *Proceedings of the SPIE - Visual Communication and Image Processing*, vol. 2727, (Orlando), p. 1302, March 1996.
- [314] Vishwanath M. and Chou P., "An Efficient Algorithm for Hierarchical Compression of Video," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 275–279, 1994.
- [315] Tan W., Chang E. and Zakhor A., "Real Time Software Implementation of Scalable Video Codec," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, (Lausanne, Switzerland), pp. 17–20, 1996.
- [316] Smith J.R., "VideoZoom Spatio-Temporal Video Browser," *IEEE Transactions on Multimedia*, vol. 1, pp. 157–171, June 1999.
- [317] McCanne S., "Scalable Multimedia Communication with Internet Multicast, Light-weight Sessions, and the MBone," Tech. Rep. CSD-98-1002, Computer Science Department, UC Berkeley, July 1998.
- [318] Amir E., McCanne S. and Katz R., "Receiver-driven Bandwidth Adaptation for Light-weight Sessions," in *Proceedings of the ACM Multimedia Conference*, (Seattle, WA), pp. 415–426, November 1997.
- [319] McCanne S., Vetterli M. and Jacobson V., "Low-complexity Video Coding for Receiver-driven Layered Multicast," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 983–1001, August 1997.
- [320] Maxemchuk N.F., "Video Distribution on Multicast Networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 357–372, April 1997.
- [321] Sudan M. and Shacham N., "Gateway based Approach for Managing Multimedia Sessions over Heterogeneous Signaling Domains," in *Proceedings of the IEEE INFOCOM Conference*, 1997.
- [322] Rejaie R., Handley M. and Estrin D., "Quality Adaptation for Congestion Controlled Video Playback over the Internet," *Computer Communication Review*, vol. 29, October 1999.

- [323] Rejaie R., Handley M. and Estrin D., "RAP: An End-to-end Rate-based Congestion Control Mechanism for Realtime Streams in the Internet," in *Proceedings of the IEEE INFOCOM Conference*, (New York, USA), March 1999.
- [324] Yeaton N., Garcia F., Hutchinson D. and Shepberd D., "Continuous media filters heterogeneous internetworking," in *Proceedings of the SPIE - Multimedia computing and networking*, vol. 2667, pp. 246–257, 1996.
- [325] Nonnenmacher J., Biersack E.W. and Towsley D., "Parity-Based Loss Recovery for Reliable Multicast Transmission," *IEEE/ACM Transactions on Networking*, vol. 6, pp. 349–361, August 1998.
- [326] Chou P.A., Mohr A., Wang A. and Mehrotra S., "FEC and Pseudo-ARQ for Receiver-driven Layered Multicast of Audio and Video," in *Proceedings of the IEEE Data Compression Conference*, (Snowbird, UT), 2000. submitted paper.
- [327] Xu X.R., Myers A.C., Zhang H. and Yavatkar R., "Resilient Multicast Support for Continuous-Media Applications," in *Proceeding of the IEEE 7th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, pp. 183–194, 1997.
- [328] Paul S., Sabnani K.K., Lin J.C.-H. and Bhattacharyya, "Reliable Multicast Transport Protocol (RMTP)," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 407–421, April 1997.
- [329] Kim B.-J., Xiong Z., Pearlman W.A. and Kim Y.S., "Progressive Video Coding for Noisy Channels," *Journal of Visual Communication and Image Representation*, vol. 10, pp. 173–185, 1999.
- [330] Tan W.-T. and Zakhor A., "Multicast Transmission of Scalable Video using Receiver-driven Hierarchical FEC," in *Proceedings of the Packet Video Workshop*, (New York), April 26–27 1999.
- [331] Meggers J., Strang T. and Park A S.-B., "A Video Gateway to Support Video Streaming to Mobile Clients," in *ACTS Mobile Summit*, (Aalborg), October 1997.
- [332] Amir E., McCanne S. and Zhang H., "An Application Level Video Gateway," in *Proceedings of the ACM Multimedia Conference*, (San Francisco, CA), pp. 255–265, November 1995.
- [333] Lengwehasatit K., *Complexity-Distortion Tradeoffs in Image and Video Compression*. PhD thesis, University of Southern California, April 2000.
- [334] Shanableh T. and Ghanbari M., "Heterogeneous Video Transcoding to Lower Spatio-Temporal Resolutions and Different Encoding Formats," *IEEE Transactions on Multimedia*, vol. 2, pp. 101–110, June 2000.
- [335] Smith J.R., Mohan R. and Li C.-S., "Content-based Transcoding of Images in the Internet," in *Proceedings of IEEE International Conference on Image Processing*, 1998.
- [336] Shen B., Sethi I.K. and Vasudev B., "Adaptive Motion-Vector Resampling for Compressed Video Downscaling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 929–936, September 1999.
- [337] Kassler A., Neubeck A. and Schulthess P., "Filtering Wavelet based Video Streams for Wireless Interworking," in *Proceedings of the IEEE ICME Conference*, (New York), July 2000.
- [338] Assuncao P.A.A. and Ghanbari M., "A Frequency-Domain Video Transcoder for Dynamic Bit-Rate Reduction of MPEG-2 Bit Streams," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 953–967, December 1998.

- [339] Chang S.-F., "Compressed-domain techniques for image/video indexing and manipulation," in *Proceedings of the IEEE International Conference on Image Processing*, October 1995.
- [340] Youn J., Xin J. and Sun M.-T., "Fast Video Transcoding Architectures for Networked Multimedia Applications," in *Proceedings of the IEEE ISCAS*, vol. 4, (Geneva, Switzerland), pp. 25–28, May 2000.
- [341] Sorial H., Lynch W.E. and Vincent A., "Selective Requantization for Transcoding of MPEG Compressed Video," in *Proceedings of the IEEE ICME*, (New York, USA), July 2000.
- [342] Vetro A., Sun H. and Wang Y., "Object-Based Transcoding for Scalable Quality of Service," in *Proceedings of the IEEE ISCAS*, vol. 4, (Geneva, Switzerland), pp. 17–20, May 2000.
- [343] Mohan R., Smith J.R., Li C.-S., "Adapting Multimedia Internet Content for Universal Access," *IEEE Transactions on Multimedia*, vol. 1, pp. 104–114, March 1999.
- [344] Meggers J., Bautz G. and Park A. S.-B., "Providing Video Conferencing for the Mobile User," in *Proceedings of the 21st Annual Conference on Local Computer Networks*, (Minneapolis), October 1996.
- [345] Turletti T., "The INRIA Videoconferencing System (IVS)," *ConneXions - The Interoperability Report Journal*, vol. 8, pp. 20–24, October 1994.
- [346] Dorsey T., "CU-SeeMe desktop videoconferencing software," *ConneXions*, vol. 9, March 1995.
- [347] McCanne S and Jacobson V., "vic: A Flexible Framework for Packet Video," in *Proceedings of the ACM Multimedia Conference*, (San Francisco, CA), pp. 511–522, November 1995.
- [348] Real Networks, "SureStream - Delivering Superior Quality and Reliability." White paper, 1999. <http://www.real.com/devzone/library/whitepapers/surestrm.htm>.
- [349] Frossard P., "FEC Performances in Multimedia Streaming," *IEEE Communications Letters*, August 2000. accepted for publication.
- [350] Frossard P. and Verscheure O., "Joint Source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery," *IEEE Transactions on Image Processing*, 1999. accepted for publication.
- [351] Rabin M.O., "Efficient Dispersal of Information for Security, Load Balancing and Fault Tolerance," *Journal of the ACM*, vol. 36, pp. 335–348, April 1989.
- [352] Lin S. and Costello D.J.jr, *Error control coding: fundamentals and applications*. Prentice-Hall, 1983.
- [353] Shacham N. and McKenney P., "Packet recovery in High-Speed Networks Using Coding and Buffer Management," in *Proceedings of the IEEE INFOCOM Conference*, vol. 1, (Los Alamitos, CA), pp. 124–131, 1990.
- [354] Blahut R.E., *Theory and Practice of Error Control Codes*. Addison Wesley, 1983.
- [355] Luby M.G., Mitzenmacher M., Shokrollahi M.A., Spielman D.A. and Stemmann V., "Practical Loss-Resilient Codes," in *Proceedings of the ACM STOC Conference*, (El Paso, TX), pp. 150–159, 1997.
- [356] Elliott E.O., "A Model of the Switched Telephone Network for Data Communications," *The Bell System Technical Journal*, vol. 44, pp. 89–109, January 1965.
- [357] Adoul J.-P.A., "Error Intervals and Cluster Density in Channel Modeling," *IEEE Transactions on Information Theory*, vol. 20, pp. 125–129, January 1974.

- [358] Kanak L.N. and Sastry A.R.K., "Models for Channels with Memory and Their Applications to Error Control," *Proceedings of the IEEE*, vol. 66, pp. 724–744, July 1978.
- [359] Kawahara K., Kumazoe K., Takine T. and Oie Y., "Forward Error Correction in ATM Networks: An Analysis of Cell Loss Distribution in a Block," in *Proceedings of the IEEE INFOCOM Conference*, vol. 3, pp. 1150–1159, 1994.
- [360] Sakakibara K., "Performance Analysis of the Error-Forecasting Decoding for Interleaved Block Codes on Gilbert-Elliott Channels," *IEEE Transactions on Communications*, vol. 48, pp. 386–395, March 2000.
- [361] Frossard P. and Verscheure O., "Joint Source/FEC Rate Selection for Optimal MPEG-2 Video Delivery," in *International Conference on Multimedia and Expo*, vol. 3, (New York), pp. 1301–1304, July 2000.
- [362] McAuley A.J., "Reliable Broadband Communications Using a Burst Erasure Correcting Code," in *Proceedings of the ACM SIGCOMM Conference*, (Philadelphia, USA), pp. 297–306, September 1990.
- [363] Baiocchi A., Melazzi N.B., Listanti M., Roveri A. and Winkler R., "Loss Performance Analysis of an ATM Multiplexer Loaded with High-Speed ON-OFF Sources," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 388–393, April 1991.
- [364] Bolot J., Fosse-Parisis S. and Towsley D., "Adaptive FEC-Based Error Control for Internet Telephony," in *Proceedings of the IEEE INFOCOM*, vol. 3, pp. 1453–60, 1999.
- [365] Yajnik M., Kurose J. and Towsley D., "Packet Loss Correlation in the MBone Multicast Network: Experimental Measurements and Markov Chain Models," Tech. Rep. 95-115, Department of Computer Science, University of Massachusetts at Amherst, 1995.
- [366] Gilbert E.N., "Capacity of a Burst-Noise Channel," *The Bell System Technical Journal*, pp. 1253–1265, September 1960.
- [367] Paxson V., Almes G., Mahdavi J. and Mathis M., "Framework for IP Performance Metrics," Tech. Rep. RFC 2330, IETF, May 1998.
- [368] Verscheure O., Frossard P. and Hamdi M., "Joint Impact of MPEG-2 Encoding Rate and ATM Cell Losses on Video Quality," in *Proceedings of the IEEE GLOBECOM Conference*, vol. 1, (Sidney), pp. 71–76, November 1998.
- [369] Verscheure O., Frossard P. and Hamdi M., "MPEG-2 Video Services over Packet Networks: Joint Effect of Encoding Rate and Data Loss on User-Oriented QoS," in *Proceedings of the International Conference on Network and Operating System Support for Digital Audio and Video*, (Cambridge, UK), pp. 257–264, July 1998.
- [370] Verscheure O., Frossard P. and Hamdi M., "User-Oriented QoS Analysis in MPEG-2 Video Delivery," *Journal of Real-Time Imaging*, vol. 5, pp. 305–314, October 1999.
- [371] Frossard P. and Verscheure O., "MPEG-2 Video over Lossy Packet Networks: Perceptual Syntactic Information Coding," in *SPIE's International Symposium on Voice, Video, and Data Communications*, vol. 3528, (Boston), pp. 113–123, November 1998.
- [372] Horn U., Stuhlmüller K., Link M. and Girod B., "Robust Internet Video Transmission Based on Scalable Coding and Unequal Error Protection," *Signal Processing: Image Communications*, vol. 15, no. 1-2, pp. 77–94, 1999.
- [373] Ding W. and Liu B., "Rate Control of MPEG Video Coding and Recording by rate-Quantization Modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 12–20, February 1996.

- [374] Chiang T. and Zhang Y.-Q., "A New Rate Control Scheme Using Quadratic Rate Distortion Model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 246–250, February 1997.
- [375] Fogg C., "mpeg2encode/mpeg2decode," in *MPEG Software Simulation Group*, 1996.
- [376] Frossard P., "MPEG-2 over Lossy Packet Networks: QoS Analysis and Improvements," Graduate School Thesis, Swiss Federal Institute of Technology, Communications System Division, Lausanne, Switzerland, July 1998.
- [377] Sakazawa S., Takishima Y., Wada M. and Hatori Y., "Coding Control Scheme for a Multi-Encoder System," in *7th International Workshop on Packet Video*, pp. 83–88, March 1996.
- [378] Fukuda K., Wakamiya N., Murata M. and Miyahara H., "On Flow Aggregation for Multicast Video Transport," in *IWQoS*, (Napa Valley), May 1998.
- [379] Fukuda K., Wakamiya N., Murata M. and Miyahara H., "Effective Algorithms for Multicast Video Transport to Meet Various QoS Requirements," *IEICE Transactions on Communications*, vol. E00-A, pp. 1–8, January 1998.
- [380] Frossard P., "MPEG-2 streams transmission over ATM networks," Master's Thesis, Swiss Federal Institute of Technology, Electrical Engineering Department, Lausanne, Switzerland, March 1997.
- [381] Hamdi M., Roberts J. W. and Rolin P., "Rate Control for VBR Video Coders in Broadband Networks," *IEEE Journal on Selected Areas in Communication*, vol. 15, pp. 1040–1051, August 1997.
- [382] Saha S. and Vemuri R., "An Analysis on the Effect of Image Activity on Lossy Coding Performance," in *Proceedings of the IEEE ISCAS*, vol. 43, (Geneva, Switzerland), pp. 295–298, May 2000.
- [383] Bocheck P. and Chang S.-F., "Content-based Dynamic Resource Allocation for VBR Video in Bandwidth Limited Networks," in *Proceedings of the Sixth International Workshop on Quality of Service (IWQoS'98)*, (New York, USA), pp. 204–6, 1998.
- [384] Jacobson V., "Congestion Avoidance and Control," *Computer Communications Review*, vol. 18, pp. 314–329, August 1988.
- [385] Alagoz F., Walters D., Alrustamani A., Vojcic B. and Pickholtz R., "On the Effects of Adaptive Forward Error Correction Mechanism in Direct Broadcast Satellite Networks," in *Proceedings of the ACM MSWiM Conference*, (Seattle, USA), pp. 25–32, 1999.
- [386] Zhang Q., Zhang Y.-Q. and Zhu W., "Resource Allocation for Audio and Video Streaming over the Internet," in *Proceedings of the IEEE ISCAS*, vol. 4, (Geneva, Switzerland), pp. 21–24, May 2000.
- [387] Verscheure O. and Frossard P., "Perceptual MPEG-2 Syntactic Information Coding: A Performance Study based on Psychophysics," in *Picture Coding Symposium*, (Berlin), pp. 297–302, September 1997.
- [388] Frossard P. and Verscheure O., "Content-based MPEG-2 Structuring and Protection," in *SPIE International Symposium on Voice, Video, and Data Communications*, vol. 3845, (Boston, USA), pp. 350–361, September 1999.
- [389] Frossard P. and Verscheure O., "AMISP: A Complete Content-Based MPEG-2 Error-Resilient Scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, 1999. submitted paper.

- [390] Carle G. and Biersack E.W., "Survey of Error Recovery Techniques for IP-Based Audio-Visual Multicast Applications," *IEEE Network*, vol. 11, pp. 24–36, November/December 1997.
- [391] Wada M., "Selective Recovery of Video Packet Loss Using Error Concealment," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 807–814, June 1989.
- [392] Biersack E., "Performance Evaluation of Forward Error Correction in an ATM Environment," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 631–640, May 1993.
- [393] Rosenberg J. and Schulzrinne H., "An RTP Payload Format for Generic Forward Error Correction," internet draft, IETF, 1999.
- [394] Frossard P. and Vandergheynst P., "Redundancy in Non-Orthogonal Transforms," in *IEEE International Symposium on Information Theory*, (Washington D.C.), June 2001. submitted paper.
- [395] Frossard P., Vandergheynst P. and Kunt M., "Redundancy-Driven A Posteriori Matching Pursuit Quantization," *IEEE Transactions on Signal Processing*, November 2000. submitted paper.
- [396] Yang L., Martins F.C.M. and Gardos T.R., "Improving H.263+ scalability performance for very low bit rate applications," in *Proceedings of the SPIE - Visual Communication and Image Processing*, vol. 3653, pp. 768–779, January 1999.
- [397] DeVore R.A., *Acta Numerica*, vol. 7, ch. Nonlinear Approximations, pp. 51–150. 1998.
- [398] Sheybani E. and Sankar R., "Multiresolution Decomposition Techniques for Robust Signal Processing," in *Proceedings of the IEEE Southeastcon*, pp. 20–23, 1998.
- [399] Al-Shaykh O.K., Miloslavsky E., Nomura T., Neff R. and Zakhor A., "Video Compression Using Matching Pursuits," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 123–143, February 1999.
- [400] Czerepinski P., Davis C., Canagarajah N. and Bull D., "Matching Pursuits Video Coding: Dictionaries and Fast Implementation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 1103–1115, October 2000.
- [401] Rabiee H.R., Safavian S.R., Gardos T.R. and Mirani A.J., "Low bit-rate subband image coding with matching pursuits," in *Proceedings of the SPIE*, vol. 3309, pp. 875–880, 1997.
- [402] Neff R., Nomura T. and Zakhor A., "Decoder Complexity and Performance Comparison of Matching Pursuit and DCT-Based MPEG-4 Video Codecs," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, (Chicago, IL), pp. 783–787, October 1998.
- [403] Natarajan B.K., "Sparse Approximate Solutions to Linear Systems," *SIAM Journal on Computing*, vol. 24, pp. 227–234, April 1995.
- [404] Goyal V.K., Vetterli M. and Thao N.T., "Quantized Overcomplete Expansions in R^N : Analysis, Synthesis and Algorithms," *IEEE Transactions on Information Theory*, vol. 44, pp. 16–31, January 1998.
- [405] Davis G., Mallat S. and Avellaneda M., "Adaptive Greedy Approximations," *Journal of Constructive Approximations*, vol. 13, pp. 57–98, 1997.
- [406] Jones L.K., "On a Conjecture of Huber Concerning the Convergence of Projection Pursuit Regression," *The Annals of Statistics*, vol. 15, no. 2, pp. 880–882, 1987.

- [407] Barut A.O. and Raczka R., *Theory of Group Representations and Applications*. ISBN: 997150216X, World Scientific Pub Co, February 1987.
- [408] Ali S.T., Antoine J.-P. and Gazeau J.-P., *Coherent States, Wavelets and Their Generalizations*. Springer-Verlag, 2000.
- [409] Antoine J.-P., Vandergheynst P. and Murenzi R., "Two-dimensional directional wavelets in image processing," *International Journal on Imaging Systems and Technology*, vol. 7, pp. 152–165, 1996.
- [410] Bernier D. and Taylor K., "Wavelets from square-integrable representations," *SIAM Journal on Mathematical Analysis*, vol. 27, no. 2, pp. 594–608, 1996.
- [411] Cvetkovic Z. and Vetterli M., "Tight Weyl-Heisenberg Frames in $l^2(Z)$," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1256–1259, May 1998.
- [412] Cvetkovic Z., "Analysis of Errors in Quantization of Weyl-Heisenberg Frame Expansions and Oversampled A/D Conversion," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1435–1438, 1996.
- [413] Sloane N.J.A., Hardin R.H. and Smith W.D., "Nice arrangements of points on a sphere in various dimensions." <http://www.research.att.com/njas/packings/>.
- [414] Sloane N.J.A., "The Sphere Packing Problem," in *Proceedings of the International Congress on Mathematics*, vol. 3, (Berlin), pp. 387–396, 1998. Documenta Mathematica, III.
- [415] Lee T.S., "Image Representation Using 2D Gabor Wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 959–971, October 1996.
- [416] Goodwin M.M. and Vetterli M., "Matching Pursuit and Atomic Signal Models Based on Recursive Filter Banks," *IEEE Transactions on Signal Processing*, vol. 47, pp. 1890–1902, July 1999.
- [417] De Vleeschouwer C. and Macq B., "Subband Dictionaries for Low-Cost Matching Pursuits of Video Residues," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 984–993, October 1999.
- [418] Neff R. and Zakhor A., "Dictionary Approximation for Matching Pursuit Video Coding," in *Proceedings of the IEEE International Conference on Image Processing*, (Vancouver), September 2000.
- [419] Liu Q., Wang Q. and Wu L., "Improvement of encoder for matching-pursuit-based video coding," *Electronics Letters*, vol. 36, pp. 548–549, March 2000.
- [420] Feichtinger H.G., Turk A. and Strohmer T., "Hierarchical parallel matching pursuit," in *Proceedings of the SPIE*, vol. 2302, pp. 222–232, 1994.
- [421] Figueras R.M., "Image coding with Matching Pursuit," Master's thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, August 2000.
- [422] Vandergheynst P. and Frossard P., "Efficient image representation by anisotropic refinement in Matching Pursuit," in *IEEE ICASSP Conference*, (Salt Lake City, UT), May 2001. submitted paper.
- [423] Aggarwal J.K. and Nandhakumar N., "On the computation of motion from sequences of images - a review," *Proceedings of the IEEE*, vol. 76, pp. 917–935, August 1988.
- [424] Li H. and Wolff I., "Multiscale Matching Pursuit for Image Coding," in *Proceedings of the Fifth International Conference on Signal Processing and its Applications*, (Brisbane, Australia), pp. 805–808, August 1999.

- [425] Qian S. and Chen D., "Signal Representation via adaptive normalized Gaussian functions," *IEEE Transactions on Signal Processing*, 1994.
- [426] Gabor D., "Theory of Communication," *The Journal of the Institution of Electrical Engineers*, vol. 93, pp. 429–457, 1946.
- [427] Janssen A.J.E.M., "Gabor Representation and Wigner Distribution of Signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 41B.2.1–41B.2.4, 1984.
- [428] Li S., "Discrete Multi-Gabor Expansions," *IEEE Transactions on Information Theory*, vol. 45, pp. 1954–1967, September 1999.
- [429] Blu T., Thevenaz P. and Unser M., "Minimum Support Interpolators with Optimum Approximation Properties," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 242–245, 1998.
- [430] Unser M., "Splines - A Perfect Fit for Signal and Image Processing," *IEEE Signal Processing Magazine*, vol. 16, pp. 22–38, November 1999.
- [431] Goyal V.K. and Vetterli M., "Dependent coding in quantized matching pursuit," in *Proceedings of the SPIE - Visual Communication and Image Processing*, vol. 3024, pp. 2–12, 1997.
- [432] Jones P.W., Daly S.J., Gaborski R.S., Rabbani M., "Comparative study of wavelet and discrete cosine transform (DCT) decompositions with equivalent quantization and encoding strategies for medical images," in *Proceedings of the SPIE - Medical Imaging*, vol. 2431, pp. 571–582, 1995.
- [433] Gharavi-Aikhansari M., "A model for entropy coding in matching pursuit," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, pp. 778–782, 1998.
- [434] Jayant N., Johnston J. and Safranek R., "Signal Compression Based on Models of Human Perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, October 1993.
- [435] Horowitz M.J. and Neuhoﬀ D.L., "Image Coding by Matching Pursuit and Perceptual Pruning," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 654–657, 1997.
- [436] Frossard P. and Vandergheynst P., "A Posteriori Quantized Matching Pursuit," in *IEEE Data Compression Conference*, (Snowbird, UT), March 27-29 2001. accepted for publication.
- [437] Goyal V.K., Vetterli M. and Thao N.T., "Quantization of Overcomplete Expansions," in *Proceedings of the IEEE Data Compression Conference*, pp. 13–22, 1995.
- [438] Cvetkovic Z., "Source Coding with Quantized Redundant Expansions: Accuracy and Reconstruction," in *Proceedings of the IEEE Data Compression Conference*, pp. 344–358, 1999.
- [439] Cvetkovic Z. and Vetterli M., "Overcomplete Expansions and Robustness," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 325–328, 1996.
- [440] Neﬀ R. and Zakhor A., "Adaptive Modulus Quantizer Design for Matching Pursuit Video Coding," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, pp. 81–85, 1999.
- [441] Neﬀ R. and Zakhor A., "Modulus Quantization for Matching-Pursuit Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 895–912, September 2000.

- [442] Everett H., "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations Research*, vol. 11, pp. 399–417, 1963.
- [443] Shoham Y. and Gersho A., "Efficient Bit Allocation for an Arbitrary Set of Quantizers," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1445–1453, September 1988.
- [444] Corless R.M., Gonnet G.H., Hare D.E.G., Jeffrey D.J. and Knuth D.E., "On the Lambert W Function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.
- [445] Schuster G.M. and Katsaggelos A.K., "An Optimal Quadtree-Based Motion Estimation and Motion-Compensated Interpolation Scheme for Video Compression," *IEEE Transactions on Image Processing*, vol. 7, pp. 1505–1523, November 1998.
- [446] Gribonval R., "A counter-example to the convergence of greedy algorithms under very weak conditions," *Journal of Approximation Theory*, March 2000. submitted paper.
- [447] Kothe G., *Topological Vector Spaces I*. Berlin: Springer-Verlag, 1969.
- [448] Antoine J.-P., Murenzi R., Vandergheynst P. and Ali S.T., *2-D wavelets and their relatives*. Cambridge University Press, 2000. to appear.
- [449] Gray R.M., "Vector Quantization," *IEEE ASSP Magazine*, pp. 4–29, April 1984.
- [450] Lee J.-B. and Eleftheriadis A., "2-D Transform-Domain Resolution Translation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 704–714, August 2000.
- [451] Munoz A., Blu T. and Unser M., "Efficient Image Resizing Using Finite Differences," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, (Kobe, Japan), pp. 662–666, October 1999.
- [452] Horbelt S., Munoz A., Blu T. and Unser M., "Spline Kernels for Continuous-Space Image Processing," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Istanbul, Turkey), June 2000.
- [453] Tan W.-T. and Zakhor A., "Resilient Compression of Video for Transmission over the Internet," in *32nd Asilomar Conference on Signal Systems and Computers*, November 1998.
- [454] Tan W.-T. and Zakhor A., "Internet Video using Error Resilient Scalable Compression and Cooperative Transport Protocol," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, (Chicago, IL), pp. 458–462, October 1998.
- [455] Briceno H.M., Gortler S. and McMillan L., "NAIVE - Network Aware Internet Video Encoding," in *Proceedings of the ACM Multimedia Conference*, (Orlando, FL), October 30 - November 5 1999.
- [456] Guntzer M.M. and Jungnickel D., "Approximate minimization algorithms for the 0/1 Knapsack and Subset-Sum Problem," *Operations Research Letters*, vol. 26, pp. 55–66, 2000.
- [457] Goyal V., *Beyond Traditional Transform Coding*. PhD thesis, University of California, Berkeley, 1998.

Resume

Education

Swiss Federal Institute of Technology, Lausanne, Switzerland Jul 1998 - Dec 2000.
Department of Electrical Engineering
Signal Processing Laboratory
Ph.D. Program.

Swiss Federal Institute of Technology, Lausanne, Switzerland Oct 1997 - Jul 1998.
Department of Communications Systems
Graduate School Program
Specialized areas: information theory, stochastic processes, communication networks.

Swiss Federal Institute of Technology, Lausanne, Switzerland Sep 1996 - Apr 1997.
Department of Electrical Engineering
Telecommunications Laboratory
M. Sc. Degree.

Swiss Federal Institute of Technology, Lausanne, Switzerland Oct 1992 - Jul 1996.
Department of Electrical Engineering
Degree of Electrical Engineer
Specialized areas: telecommunications, signal processing, programmable systems and electronic.

Honors, Awards

Hewlett-Packard grant for PhD program.

Experiences

Swiss Federal Institute of Technology, Lausanne, Switzerland Jul 1998 - present.
Department of Electrical Engineering
Signal Processing Laboratory
Research and Teaching Assistant under Hewlett-Packard Partnership.

Swiss Federal Institute of Technology, Lausanne, Switzerland Apr 1997 - Oct 1997.
Department of Computer Science
Institute for Computer Communications and Applications
Research and teaching assistant.

Swisscom, Lausanne, Switzerland Jul 1995 - Oct 1995.
Computing service.

Personal Publications

- [J01] Frossard P., “FEC Performances in Multimedia Streaming,” *IEEE Communications Letters*, December 2000, accepted for publication.
- [J02] Frossard P., Vanderghelynst P. and Kunt M., “Redundancy-Driven A Posteriori Matching Pursuit Quantization,” *IEEE Transactions on Signal Processing*, November 2000, submitted paper.
- [J04] Frossard P. and Verscheure O., “Joint Source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery,” *IEEE Transactions on Image Processing*, 2000, accepted for publication.
- [J03] Frossard P. and Verscheure O., “AMISP: A Complete Content-Based MPEG-2 Error-Resilient Scheme,” *IEEE Transactions on Circuits and Systems for Video Technology*, 1999, submitted paper.
- [J05] Verscheure O., Frossard P. and Hamdi M., “User-Oriented QoS Analysis in MPEG-2 Video Delivery,” *Journal of Real-Time Imaging*, vol. 5, no. 5, pp. 305–314, October 1999.
- [C01] Vanderghelynst P. and Frossard P., “Efficient image representation by anisotropic refinement in Matching Pursuit,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001, submitted paper.
- [C02] Frossard P. and Vanderghelynst P., “A Posteriori Quantized Matching Pursuit,” in *IEEE Data Compression Conference*, Snowbird, UT, March 27-29 2001, accepted for publication.
- [C03] Frossard P. and Vanderghelynst P., “Redundancy in Non-Orthogonal Transforms,” in *IEEE International Symposium on Information Theory*, Washington D.C., June 2001, submitted paper.
- [C04] Verscheure O., Frossard P. and Le Boudec J.-Y., “Joint Smoothing and Source Rate Selection for Guaranteed Service Networks,” in *IEEE INFOCOM Conference*, Anchorage, Alaska, April 2001, accepted for publication.
- [C05] Frossard P. and Verscheure O., “Joint Source/FEC Rate Selection for Optimal MPEG-2 Video Delivery,” in *Proceedings of the IEEE International Conference on Multimedia and Expo*, New York, July 2000, vol.3, pp. 1301–1304.
- [C06] Frossard P. and Verscheure O., “Content-based MPEG-2 Structuring and Protection,” in *SPIE International Symposium on Voice, Video, and Data Communications*, Boston, USA, September 1999, vol. 3845, pp. 350–361.
- [C07] Frossard P. and Verscheure O., “MPEG-2 Video over Lossy Packet Networks: Perceptual Syntactic Information Coding,” in *SPIE’s International Symposium on Voice, Video, and Data Communications*, Boston, November 1998, vol. 3528, pp. 113–123.

- [C08] Verscheure O., Frossard P. and Hamdi M., "MPEG-2 Video Services over Packet Networks: Joint Effect of Encoding Rate and Data Loss on User-Oriented QoS," in *Proceedings of the NOSSDAV Conference*, Cambridge, UK, July 1998, pp. 257–264.
- [C09] Verscheure O., Frossard P. and Hamdi M., "Joint Impact of MPEG-2 Encoding Rate and ATM Cell Losses on Video Quality," in *Proceedings of the IEEE GLOBECOM Conference*, Sidney, November 1998, vol.1, pp. 71–76.
- [C10] Verscheure O. and Frossard P., "Perceptual MPEG-2 Syntactic Information Coding: A Performance Study based on Psychophysics," in *Proceedings of the Picture Coding Symposium*, Berlin, September 1997, pp. 297–302.
- [C11] Garcia Adanez X., Verscheure O. and Frossard P., "Reliable Transmission of MPEG-2 VBR Video Streams over New Network Adaptation and ATM Adaptation Layers," in *IEEE ATM Workshop*, Lisboa, May 1997, pp. 158–167.
- [M01] Frossard P., "MPEG-2 over Lossy Packet Networks - QoS Analysis and Improvements," Graduate School Thesis, Swiss Federal Institute of Technology - Communications System Division, Lausanne, Switzerland, July 1998.
- [M02] Frossard P., "MPEG-2 streams transmission over ATM networks," Master's Thesis, Swiss Federal Institute of Technology - Electrical Engineering Department, Lausanne, Switzerland, March 1997.