MSE cross-layer optimization criteria: What else?

Laura Toni, Italian Institute of Technology, Genova, Italy
Pascal Frossard, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
laura.toni@iit.it, pascal.frossard@epfl.ch

1. Introduction

Recently, video streaming over wireless networks has become quite popular with the development of novel exciting mobile media applications. It brings interesting research challenges with a deep change in design paradigms and research emphases compared to traditional streaming applications. In particular, cross-layer optimization techniques have been under intense research as primary strategies for adaptation to dynamic channel conditions, since they permit to improve the quality of experience by optimizing the network architectures across traditional OSI stack layers.

The main issue to be faced when adopting a cross-layer approach is the definition of a reliable metric to formulate the quality optimization problem. Ideally, a good metric should be straightforward to understand and immediate to calculate, encompassing, at the same time, the different components that influence end-user video distortion in wireless streaming. At the source coder, the distortion typically decreases with increasing rate, but the actual rate-distortion characteristic depends on the video content. For example, in a variable rate multiuser scenario, an accurate quality assessment can help the video senders in optimizing their transmission settings (packet dropping prior the transmission or packet scheduling) that, in turn, results in efficient usage of the network resources towards minimal average distortion. Additionally, a good distortion metric should also be able to capture the effects of channel impairments on the received video, or equivalently the distortion degradation due to packet losses and error concealment. Typically, an increase in the error protection tends to reduce the impact of losses, but this distortion characteristic is also influenced by the video content. In general, one faces an optimization problem where bits have to be traded off between source coding rate and channel protection for optimized video quality. An accurate measure of video distortion becomes key in the selection of the appropriate optimization techniques.

The video quality assessment might be considered from two different angles: as a mathematical evaluation (*mathematical metric*) or as a visual-perception test (*visual perception metric*). While the former is simple to assess and easy to reproduce, the latter provides results that are often more complex to evaluate but also more closely related to the video quality experienced by the final user, *i.e.*, the human eye. The compelling question is “Which is the best distortion metric to be considered for cross-layer design?” [2, 3]. In this letter, we aim at providing an insight on this topic, highlighting the advantages and limits of both mathematical and visual perception metrics.

2. Mathematical Metrics

Objective tests evaluate the video quality through criteria that can be rigorously measured and automatically computed. Among these mathematical metrics, the most common is the mean square error (MSE) metric. The MSE criterion has many attractive features. First, it is straightforward to evaluate. Being x_i and y_i the i-th pixel of the original and reconstructed frames of the video sequence, respectively, the MSE is calculated as the mean energy of the error signal $e=(x_i-y_i)$. In addition, MSE is able to capture the effects of both source and channel distortions in the video degradation. Finally, minimum-MSE optimization problem usually leads to closed-form analytical solutions or iterative numerical procedures that are easy to implement. For all these reasons, MSE is widely used as a parameter to be minimized in cross-layer design. [5-8].

For example, the authors propose a cross-layer algorithm for distributed video rate allocation over wireless systems in [5, 6]. The MSE distortion metric is justified by previous studies on subjective test [9], which have demonstrated that the rate allocation algorithm aimed at minimizing the overall MSE agrees sufficiently well with subjective opinions. MSE distortion is minimized also in [8], where a Rate-Distortion (RD) based video packet selection for distributed streaming over a shared communication channel is proposed. Although optimizing MSE does not necessarily correspond to optimizing subjective quality, the authors demonstrate that a video-aware cross-layer technique achieves a lower received video distortion in terms of MSE, compared to baseline algorithms.

A comprehensive RD analysis of video quality due to source and channel distortion has been reported in [10-12]. MSE typically decreases non linearly with the increase of the received rate. It can be modeled with parametric functions, where the parameters of the model can be evaluated through non linear regression techniques or least-square methods. Even if complete MSE models might be constructed, the proper handling of the temporal evolution of the video quality remains unsolved. Two different video streams with the same average distortion might not be perceived equally from an end-user point of view. This is mainly caused by the fact that average MSE distortion cannot distinguish the frame-by-frame quality variation and its effect on the final user. These effects are difficult to capture in cross-layer streaming solutions that usually try to minimize the average distortion.

Another important aspect that should improve the fidelity of the distortion metric is the relation between MSE and packet losses. In [13], the authors analyze the evolution of expected distortion versus loss pattern, for video communication over error-prone channels. They verify that the packet loss pattern, and in particular the burst length, does have a significant effect on the resulting distortion; hence it should be taken into consideration in the optimization problems. Despite its importance in video cross-layer designs, the relation between distortion and channel impairment usually depends on many parameters, such as coding scheme, the bit rates and the network architectures. Even if MSE is a priori a simple metric, its appropriate use in the design of cross-layer optimization solutions is not as trivial as it appears.

3. Visual Perception Metrics

Since the ultimate receivers of wireless video streams are usually human observers, mathematical metrics, which measure the video distortion without considering perceptual factors, might not reflect the video quality perceived by the human eye. Visual perception metrics (e.g., visual quality metric, VQM, [14], or subjective tests) are able to capture the effects of source distortion as well as channel distortion (loss distortion, temporal error propagation, variable bit-rates) with high fidelity to the human perception, at a price of complexity in the evaluation. Much of the efforts in understanding the visual impact of channel impairments are focused on the evaluation of the overall perceived quality. In [14], a subjective study to assess the visual quality of compressed video affected by wireless channel impairments is provided. The main limit is that the evaluation of mean opinion score (MOS) is extremely complex (e.g., the quality scales are generally not equally interpreted by the users). It further requires fully decoded video streams and is highly dependent on the testing conditions. It is not easy to factor the effects of each system parameters in terms of overall subjective quality.

To the best of our knowledge, there exists no objective metric which is able to assess the visual quality of wireless video in terms of the relative importance of the different packets to be employed in cross-layer optimization techniques. An alternative consists in taking into account a set of simpler metrics, which can help the system designer in optimizing the transmission scheme in order to augment the visual quality. One of these metrics is the bitstream-based metric provided in [15, 16]. Bitstream-based metrics predict video quality looking at the packet header information and the encoded bitstream without decoding the video source. In these works, the metric aims at evaluating the effects of packet losses on the perceived quality. The authors conducted subjective tests in which the viewers' task is to indicate when an artifact, due to packet losses, becomes visible. The subjective tests are conducted for video streams in which isolated losses could be experienced. From these tests, a packet loss visibility metric (PLV) has been proposed with the goal of predicting whether a packet loss in the video stream becomes visible from a human viewer. The main advantages of the proposed technique are that the PLV is predictable with an objective metric, takes into account the temporal evolution of the video quality, and does not require to fully decode the video sequence. The authors demonstrate that the adoption of the PLV model as packet prioritization policy for video transmission leads to a better overall perceived quality compared to baseline cross-layer streaming solution.

The main drawback of the PLV model lies on a simplistic channel model with isolated losses. Extensions of visibility metrics to more realistic wireless channel models should however lead to improved cross-layer design as long as the computational complexity stays limited. Fidelity and complexity are therefore posed as a trade-off in the selection of the distortion metric in cross-layer design optimization.

4. Conclusions

In this letter, we provide an insight on the distortion metric that could be adopted in order to optimize the cross-layer video system design. Mathematical metrics, especially the MSE, are simple to evaluate and well accepted from the video processing community. Although reducing the MSE is not equivalent to minimizing the visual quality, the MSE metric provides a reasonable level of fidelity when dealing with cross-layer techniques, but only provides a coarse approximation of the actual visual distortion. Visual quality metrics, which estimate the actual quality perceived by the final user, are suitable to benchmark systems and algorithms in comparative studies, but their use in cross-layer design is not trivial. Recent works on visibility models offer promising perspectives in cross-layer design, as long as these models are able to adapt to realistic wireless channels with a controllable complexity.

References

[1] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, pp. 1071-1083, 2002.

[2] W. Zhou and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," IEEE Signal Processing Magazine, vol. 26, pp. 98-117, 2009.

[3] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," IEEE Transactions on Broadcasting, vol. 54, pp. 660-668, 2008.

[4] R. R. Pastrana-Vidal and J.-C. Gicquel, "Automatic quality assessment of video fluidity impairments using a no-reference metric" Int'l Workshop on Video Proc. and Quality Metrics, Jan. 2006.

[5] Z. Xiaoqing and B. Girod, "Distributed Media-Aware Rate Allocation for Wireless Video Streaming," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, pp. 1462-1474.

[6] Z. Liang, W. Xinbing, T. Wei, G. M. Muntean, and B. Geller, "Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks," IEEE Journal on Selected Areas in Communications, vol. 28, pp. 409-419.

[7] Y.S. Chan, P. C. Cosman, and L. B. Milstein, "A Multiple Description Coding and Delivery Scheme for Motion-Compensated Fine Granularity Scalable Video," IEEE Transactions on Image Proc., vol. 17, pp. 1353-1367, 2008.

[8] J. Chakareski and P. Frossard, "Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources," IEEE Transactions on Multimedia, vol. 8, pp. 207-218, 2006.

[9] Z. Xiaoqing and B. Girod, "Subjective evaluation of multi-user rate allocation for streaming heterogeneous video contents over wireless networks," in IEEE ICIP 2008. 15th 2008, pp. 3092-3095.

[10] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-Aware Resource Allocation and Packet Scheduling for Video Transmission over Wireless Networks," IEEE Journal on Selected Areas in Communications, vol. 25, pp. 749-759, 2007.

[11] C. Yim and A. C. Bovik, "Evaluation of temporal variation of video quality in packet loss networks," Signal Processing: Image Communication, vol. 26, pp. 24-38.

[12] M. Naccari, M. Tagliasacchi, and S. Tubaro, "No-Reference Video Quality Monitoring for H.264/AVC Coded Video," IEEE Transactions on Multimedia, vol. 11, pp. 932-946, 2009.

[13] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of Packet Loss for Compressed Video: Effect of Burst Losses and Correlation Between Error Frames," IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, pp. 861-874, 2008.

[14] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless Video Quality Assessment: A Study of Subjective Scores and Objective Algorithms," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, pp. 587-599.

[15] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling packet-loss visibility in MPEG-2 video," IEEE Transactions on Multimedia, vol. 8, pp. 341-355, 2006.

[16] L. Ting-Lan, S. Kanumuri, Z. Yuan, D. Poole, P. C. Cosman, and A. R. Reibman, "A Versatile Model for Packet Loss Visibility and its Application to Packet Prioritization," *IEEE Transactions on Image Processing,* vol. 19, pp. 722-735.

Laura Toni has a postdoctoral position at the TeleRobotics and Applications (TERA) department at the Italian Institute of Technology (IIT), Genova, Italy. In 2009, she received her Ph.D. degree in electronics, computer science and telecommunications from the University of Bologna, Italy. During 2007, she had been a visiting scholar at the University of California at San Diego, CA, working on video processing over wireless systems. Since 2009, she has been a frequent visitor to the UCSD, working on joint source and channel coding on wireless communication systems. Her research interests are in the areas of image and video processing, wireless communications, and underwater communications.

Pascal Frossard received the M.S. and Ph.D. degrees, both in electrical engineering, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997 and 2000, respectively. Between 2001 and 2003, he was a member of the research staff at the IBM T. J. Watson Research Center, Yorktown Heights, NY. Since 2003, he has been an assistant professor at EPFL, where he heads the Signal Processing Laboratory (LTS4). His research interests include image representation and coding, visual information analysis, distributed image processing and communications, and media streaming systems.