Optimized MVC Prediction Structures for Interactive Multiview Video Streaming

Ana De Abreu, Pascal Frossard, and Fernando Pereira

***Abstract*—The *Multiview Video Coding* (MVC) standard efficiently compresses multiview video by considering spatial, temporal and interview correlations. This letter studies the impact of the MVC interview prediction structure on both the transmission and the overall coding rates for an *interactive multiview video streaming* system, considering both unicast and multicast scenarios, with the user interactive behavior represented by some view-popularity model. We propose a method to identify the optimal prediction structure minimizing the visual distortion, given some storage and link capacities constraints. Simulation results confirm that the optimal prediction structure results from a non-trivial tradeoff between the system constraints, the transmission model and the views' popularity.**

***Index Terms*—Interactive multiview video streaming (IMVS), multicast, multiview video coding (MVC), popularity model, prediction structure, unicast.**

I. Introduction

THE MULTIVIEW video coding (MVC) standard is an extension of the H.264/AVC (Advanced Video Coding) standard, targeting the efficient compression of multiview video [1]. It exploits the interview, spatial and temporal dependencies to improve the rate-distortion (RD) performance in multiview scenarios. The exhaustive exploitation of these interview dependencies is well suited for applications where all the views have to be transmitted or stored together under limited channel and/or storage capacities. However, for interactive multiview video streaming (IMVS) applications, where users typically request only one view at a time, the increased coding dependencies may bring important penalties as the request of one view typically implies the transmission of data from many other views. Notwithstanding their importance, IMVS systems have not been much studied in the literature. In [2], an experimental analysis of the compression efficiency of MVC prediction structures (PSs) is presented, while the work in [3] proposes a new PS model for MVC to reduce the coding complexity while keeping a high compression rate. None of these works have considered the transmission aspects of the multiview data in interactive systems. Other works, like [4], [5], have studied the PSs that facilitate a continuous view-switching [4] or that reduce the transmission rate [5]. However, the authors have considered a more complex coding system with redundant P- and DSC-frames, which is not compatible with the MVC standard. In [6], the views' popularities of an IMVS system have been considered for solving a rate allocation problem; however, a fixed PS has been used, which is not appropriate for all the IMVS scenarios.

This letter studies the proper PS choice for MVC-based IMVS systems. More specifically, we study *the impact of interview coding dependencies on both the coding rate and the transmission rate per requested view, for distortion-minimized multiview streaming in resource-constrained systems*. To this end, a probabilistic model is proposed for the coding and transmission rate, considering a view-popularity model that accounts for the different characteristics of each view. The transmission rate is studied on the end-user and server sides, both for unicast and multicast transmission models. We further define a method able to identify the optimal PS that minimizes the visual distortion while fulfilling the relevant system constraints in different transmission settings. We show that MVC coding should be optimized as a tradeoff between compression and access flexibility in IMVS systems.

This letter is organized as follows: Section II outlines the assumptions and constraints adopted in this work while the coding and transmission rate probabilistic models are proposed in Section III. Section IV describes the prediction structure optimization problem for which a greedy algorithm is proposed and Section V presents the corresponding simulation results. Finally, the conclusions are presented in Section VI.

II. IMVS System Assumptions and Constraints

A. MVC Standard Related

In this work, MVC compliant bitstreams are considered fulfilling the following constraints:

- ***Fixed temporal PS***—The interview prediction dependencies are modified while the temporal dependencies are kept fixed. In the temporal domain, a *hierarchical B-frames* solution is assumed, since it has shown to improve the compression efficiency [2].
- ***Cascading quantization parameters (CQP)***—To control the quantization steps, and thus the distortion, a CQP strategy is used in the MVC encoder. The CQP model used for each view is similar to the one proposed in [7]: first, Q, the set of QPs for the *anchor pictures* (without temporal dependencies) [1] is selected; then, the QPs for the various temporal layers are assigned by increasing the QP of the previous layer with a pre-defined ΔQ.

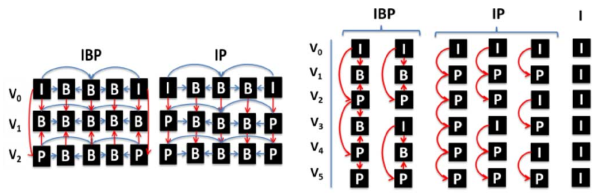
Manuscript received February 18, 2013; revised April 17, 2013; accepted April 20, 2013. Date of publication April 24, 2013; date of current version May 03, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yiannis Andreopoulos.

A. De Abreu is with the Signal Processing Laboratory (LTS4), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, and also with the Instituto Superior Técnico—Instituto de Telecomunicações (IST-IT), 1049-001 Lisbon, Portugal (email: ana.deabreu@epfl.ch).

P. Frossard are with Signal Processing Laboratory (LTS4), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (email: pascal.frossard@epfl.ch).

F. Pereira is with Instituto Superior Técnico—Instituto de Telecomunicações (IST-IT), 1049-001 Lisbon, Portugal. (email: fp@lx.it.pt).

Digital Object Identifier 10.1109/LSP.2013.2259815



(a) (b)

Fig. 1. PS examples. (a) PSs for a 3-view system: IBP(3,1) and IP(3,1); (b) PSs for a 6-view system IBP(6,1), IBP(6,2), IP(6,1), IP(6,2), IP(6,3) and I(6,6).

- ***Limited set of PSs***—To reduce the candidate set of inter-view PSs, we focus on the most relevant PSs that are built on the two most used PSs for the MVC standard, namely *IBP* and *IP*. For compression reasons, these two PSs usually have only one independent view (key-view) out of the total number of views, K, which is usually selected to be a lateral one (see Fig. 1(a)). Since, in IMVS systems, compression efficiency is not the only requirement, we allow here more than one key-view in these two basic PSs. Moreover, we also consider the simulcasting PS (I PS) with only key-views. Overall, this leads to three different basic PSs that are referred as $IBP(K,n), IP(K,n)$ and $I(K,K)$, where n stands for the number of key-views out of the total of K views. Fig. 1(b) illustrates a few PSs under consideration in our work.

B. Random Access

We consider an IMVS system with random access to different views, where view-switches occur from/to any viewpoint in a multiview set. The random access constraints are:

- ***Limited view-switches***—View-switches only occur at *instantaneous decoding refresh (IDR)* pictures [7] and anchor pictures[1].
- ***Fixed GOP size***—To facilitate random access, the GOP size is fixed and identical for all the views.
- ***View-popularity model***—A *view-popularity factor*, p_i, is considered, defined as the probability that an end-user decides at an anchor picture to switch to viewV_i, where $\sum_{i=1}^{K} p_i = 1$. Three view-popularity distributions are considered in our multi-camera setting, namely exponential, Gaussian and U-quadratic probability distributions, where the most popular views correspond to one of the lateral, central and both lateral views, respectively. At encoding, we adapt the PSs to the popularity distribution such that at least one key-view is located at one of the most popular viewpoints. Hence, the PSs are further labeled according to the position of the key-view(s). Structures labeled with A are those with key-views regularly spaced along the viewpoints starting with one of the most lateral viewpoints. Structures labeled with B and C have the key-view(s) at the middle and at both lateral extremes of the viewpoints, respectively. Table I shows the PS adaptation to the view-popularity model for the IBP PS with six views. IBP(6,1)-A corresponds to the IBP(6,1) PS adjusted to the exponential and U-quadratic distributions, while IBP(6,1)-B to the Gaussian distribution. In the IBP(6,2) PS case, structures labeled with A, B and C, are the PS adaptations to the exponential, Gaussian and U-quadratic distributions, respectively. The same key-views layout is adopted for the IP PS. The I(6,6) and IP(6,3) PS do not need this labeling as the same key-views layout is considered for all popularity models.

[1]The term anchor picture is used here to refer to both IDR and anchor pictures as, for the purposes of this letter, they are equivalent.

TABLE I
PS ADAPTATION TO THE VIEW-POPULARITY MODELS

View index	IBP(6,1)		IBP(6,2)		
	(A)	(B)	(A)	(B)	(C)
V_0	I	P	I	P	I
V_1	B	B	B	B	B
V_2	P	I	P	I	P
V_3	B	B	I	I	P
V_4	P	P	B	B	B
V_5	P	P	P	P	I

III. CODING AND TRANSMISSION RATES MODELING

In this section, we propose models for the coding and expected end-user/server transmission rates. The *coding rate*, CR, is defined as the total number of bits per unit of time that is necessary to code a multiview sequence. The *expected end-user and server transmission rate*, TR^u and TR^s, are defined in the context of IMVS where the server has to transmit the frames for a requested view together with the frames from the associated reference views, until the next view-switching point. The server transmission rate depends on the transmission model, namely *unicast* or *multicast*. In the unicast model, every client is served with a separate stream. In the multicast model, batches of views are sent as a single stream to multiple synchronous clients. In this case, we consider that an end-point in the network eventually filters the streams and only sends to the client the views that are necessary to fulfill its request. The end-user transmission rate is independent of the transmission model, since in both cases the client receives the same streams. Before deriving the models for CR, TR^u and TR^s, some definitions are presented for better understanding.

We first define the *frame-dependency path size*, $\phi(F_{i,j}^g)$, whose concept is similar to the *transmission cost* defined in [5]. In particular, $\phi(F_{i,j}^g)$ is the number of bits per GOP g that have to be processed in order to decode a particular frame, $F_{i,j}$, from viewV_i at time instant T_j, where $F_{i,j}$ can be an I-, a P- or a B-frame. The size $\phi(F_{i,j}^g)$, is recursively defined as:

$$\phi\left(F_{i,j}^g\right) = n_b\left(F_{i,j}^g\right) + \sum_k \phi\left(F_{k,j}^g\right) + \sum_t \phi\left(F_{i,t}^g\right) \tag{1}$$

where $n_b(F_{i,j}^g)$ stands for the number of bits used to code $F_{i,j}$ from GOP g, and $F_{k,j}$ and $F_{i,t}$ are the spatial and temporal reference frames for $F_{i,j}$, respectively. Now, the number of bits required to decode a GOP g of view V_i, named *GOP-dependency path size* $(\phi_{i,g})$ is defined as:

$$\phi_{i,g} = \sum\nolimits_{j=1}^{L} \phi\left(F_{i,j}^g\right) \tag{2}$$

where each frame is considered only once.

The *coding rate*, CR, is computed based on the average of the coded frame size in bits, $n_b(F_{i,j}^g)$, over the entire multiview video sequence of K views and N frame instances per view:

$$\mathrm{CR} = f\frac{\sum_i \sum_j n_b\left(F_{i,j}^g\right)}{NK} \tag{3}$$

where f is the video frame rate. CR is directly proportional to the number of bits used to code the multiview sequence, which,

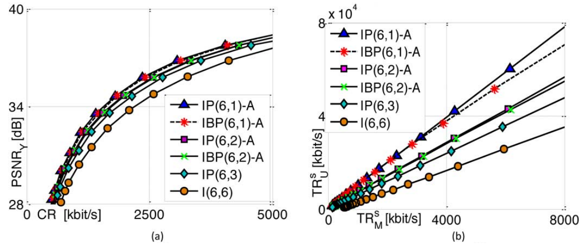


Fig. 2. (a) $\mathrm{PSNR_Y}$ vs. CR; (b) TR_U^s vs. TR_M^s for various PSs. Ballroom sequence [8] and exponential popularity distribution have been considered.

in this work, is limited by the *storage capacity* of the system. Then, since CR depends on the number of bits used to encode the sequence, its value increases with the number of views but decreases when the number of interview dependencies increases, since the redundancy across views is better exploited. Alternatively, for a given available bitrate, the encoder is able to produce a higher encoding quality when the number of interview dependencies increases (see Fig. 2(a)).

We now compute the *expected end-user transmission rate*, TR^u, when a dedicated video stream is transmitted between the end-user and the network end-point. The TR^u reads:

$$\mathrm{TR}^u = f\frac{\sum_{g=1}^{G} E\{\phi_{i,g}\}}{NK} \tag{4}$$

where $E\{\phi_{i,g}\}$ is the expectation of the size $\phi_{i,g}$, defined as $E\{\phi_{i,g}\} = \sum_{i=1}^{r} p_i\phi_{i,g}$. This means that, contrarily to CR, TR^u increases with the number of interview dependencies.

The *expected server transmission rate*, TR^s, depends on the transmission model. If the *unicast* model is adopted, the server needs to transmit, to each of the m end-users, one copy of the requested data, independently of the other requests. Hence, in the unicast case, $\mathrm{TR}^s = \mathrm{TR}_U^s$ increases with the number of interview dependencies and reads:

$$\mathrm{TR}_U^s = m\mathrm{TR}^u \tag{5}$$

If a multicast model is adopted for transmission to synchronous users, the dissemination of the requested views traverses each network link only once and no duplicate views have to be transmitted by the server. Before deriving the multicast server rate TR_M^s expression, some parameters have to be defined. First, the set of all K captured views is denoted by $S = \{V_1, V_2, \ldots, V_K\}$, from which the m end-users request a subset of r views at the same time instant, called s^r. Then, the number of bits required to decode a particular subset s^r for GOP g, denoted by $\phi_{s^r,g}$, is written as $\phi_{s^r,g} = \sum_{i=1}^{r}\phi_{i,g}$. The probability that a view V_i is not requested by any of the m end-users in a GOP g is defined as: $q_{i,g}^m = (1-p_{i,g})^m$, while the probability that at least one end-user requests a particular view V_i is $p_{i,g}^m = (1-(1-p_{i,g})^m) = 1 - q_{i,g}^m$. Then, the probability that r views are requested by an end-user, and the remaining $K-r$ views are not requested, comes $P_{s,g}^r = \prod_{i\in s^r} p_{i,g}^m \prod_{j\in S\setminus s^r} q_{j,g}^m$. Hence, the expected number of bits transmitted for each GOP g in a multicast scenario, is defined as $E\{\phi_{s^r,g}\} = \sum_{r=1}^{K}\sum_{s^r} P_{s^r,g}\phi_{s^r,g}$. Finally, TR_M^s reads:

$$\mathrm{TR}_M^s = f\frac{\sum_{g=1}^{G} E\{\phi_{s^r,g}\}}{NK} \tag{6}$$

In this case, in contrast to TR_U^s, interview dependencies reduce the redundancy across the s^r requested views and thus the transmission rate, TR_M^s. This may be confirmed in Fig. 2(b) where the relation between the server transmission rates for both transmission models, TR_U^s and TR_M^s, is illustrated.

We can also compute the expected distortion in our IMVS model, which corresponds to the coding noise associated to the quantization process. The *average distortion per view*, D_i, is taken as the temporal average of the distortion per frame, $D_{i,j}$, and is written as $D_i = \sum_{j=1}^{N} D_{i,j}/NK$. Then, the *expected distortion* for the multiview sequence is defined as:

$$\bar{D} = \sum_{i=1}^{K} p_i D_i \tag{7}$$

where p_i is the popularity for view V_i.

IV. Prediction Structure Optimization

Equipped with the previous definitions, we can now study the rate and distortion for different MVC PSs. In particular, we want to *find the interview prediction structure derived from a subset of preferred PSs, and the associated set of QPs*, Q, *that minimizes the expected distortion, while considering the following constraints:*

1. ***Storage capacity constraint***—The total number of bits to code the multiview video sequence per unit of time, CR, is restricted by the available storage capacity, named $\mathrm{CR_{max}}$.
2. ***Transmission constraint***—The expected transmission rates TR^u and TR^s $(\mathrm{TR}_U^s, \mathrm{TR}_M^s)$ cannot exceed the link capacity at the end-users, $\mathrm{TR}_{\max}^u$, and at the server, $\mathrm{TR}_{\max}^s$, respectively.

In summary, the optimization problem is written as follows:

$$\arg\min_{Q,\mathrm{PS}} \bar{D}(Q,\mathrm{PS}) \tag{8}$$

such that

$$\begin{aligned} \mathrm{CR}(Q,\mathrm{PS}) &\le \mathrm{CR_{max}} && \text{CR constraint}\\ \mathrm{TR}^s(Q,\mathrm{PS}) &\le \mathrm{TR}_{\max}^s && \mathrm{TR}^s \text{ constraint}\\ \mathrm{TR}^u(Q,\mathrm{PS}) &\le \mathrm{TR}_{\max}^u && \mathrm{TR}^u \text{ constraint}\end{aligned}$$

We propose a greedy approach to solve the optimization problem defined in (8). This algorithm has the following steps:

1. ***Local solution***—First, for each considered PS, the QPs of anchor views, Q, and the distortion D achieved for the maximum storage capacity, $\mathrm{CR_{max}}$, are determined by an exhaustive search. The same process is performed for $\mathrm{TR}_{\max}^u$ and $\mathrm{TR}_{\max}^s$. The transmission/coding rates are estimated from (3), (4), (5) and (6) with help of encoding the multiview sequences with different QPs. The QPs of the anchor views are modified while the CQP strategy is used in the temporal domain (Section II.A). After this step, for each PS, there is a set of three local (Q, D) solutions, one for each constraint.
2. ***Global feasibility***—Second, for each PS, it is determined which local solutions (Q, D) are globally feasible thus identifying the solutions that simultaneously fulfill all the constraints. This provides a set of feasible (Q, D) solutions for each PS. From these sets, the solution associated to the lowest distortion is selected. Then, for each PS, there is now an optimal Q corresponding to the (Q, D) solution fulfilling all the constraints with the minimum possible distortion D.

3. ***Optimal solution***—Finally, from the set of optimal solutions above, the PS with the minimum distortion is selected thus defining the global optimal PS.

V. Experimental Results

This section presents the test conditions and the experimental results obtained for the studied scenarios when the PS is optimized with the algorithm described in Section IV.

A. Test Conditions

The MVC reference software JMVC v8.2 [9] has been used to encode six views from three sequences: *Ballroom* [8] and *Akko & Kayo* [10] at a resolution of 640×480 with 25 fps captured with a parallel camera array with 20 and 5 cm of separation, respectively; *Ballet* [11] (1024×768 pixels at 15 fps) captured with a 1-D arc camera arrangement spanning about 30°. The CQP strategy has been used in the temporal domain, with a fixed ΔQ equal to 0, 3 and 1 when the temporal layer is equal to 0, 1 or larger than 1; respectively. For instance, the QPs for temporal layers 0, 1 and 3 are $Q, Q+3$ and $Q+3+1$; respectively. The basic PSs used were IP, IBP and simulcasting, adapted to the adopted popularity distributions, i.e., exponential, Gaussian and U-quadratic. The optimal PS has been found for various scenarios defined in terms of available server/end-user transmission rates and storage capacity, with the help of the greedy algorithm proposed in Section IV.

B. Results and Analysis

Two multi-constrained scenarios have been studied, one associated to unicast and another to multicast, using the following constraints:

- ***Unicast scenario***: $\mathrm{TR}^u_{\max} = 1$ Mbps, $\mathrm{TR}^s_{\max} = 10$ Mbps and $\mathrm{CR}_{\max} = 3.50$ Mbps.
- ***Multicast scenario***: $\mathrm{TR}^u_{\max} = 1$ Mbps, $\mathrm{TR}^s_{\max} = 2.40$ Mbps and $\mathrm{CR}_{\max} = 3.50$ Mbps.

The results are reported in Table II, where the limiting rate and the optimal PS for each case are shown in bold. Based on these results, we can make the following observations:

- ***Different minimum distortion solutions***—Given the same system constraints, the algorithm finds different quality levels for the optimal solutions of the three sequences. Ballroom solutions are around 5 and 4 dB lower than the Akko & Kayo and Ballet solutions. This is due to different sequence characteristics and camera spacing and arrangement.
- ***Coding efficiency versus random access trade-off***—In these scenarios, the optimal PSs have two or three key-views. This solution results from the tradeoff between compression efficiency (associated to the best PS under coding rate constraint) and random access capabilities (associated to the best PS under end-user and server unicast transmission rate constraints). This shows that a pure compression efficiency objective is not ideal in IMVS systems.
- ***PS alignment with popularity models***—For each popularity distribution, the optimal PS is the one that best adapts the key-views to the most popular views. For example, for a Gaussian popularity model, the PSs labeled with B, with central key-views, offer the best performance.

TABLE II
OPTIMAL PS FOR BALLROOM, AKKO & KAYO AND BALLET SEQUENCES FOR VARIOUS POPULARITY DISTRIBUTIONS AND BOTH UNICAST AND MULTICAST SCENARIOS

Scenario	Statistics	Exponential			Gaussian			U-quadratic		
		Ballroom	*Akko & Kayo*	*Ballet*	*Ballroom*	*Akko & Kayo*	*Ballet*	*Ballroom*	*Akko & Kayo*	*Ballet*
Unicast	Optimal PS	**IP(6,2)-A**	**IP(6,2)-A**	**IP(6,3)**	**IP(6,3)**	**IP(6,3)**	**IP(6,3)**	**IP(6,2)-C**	**IP(6,2)-C**	**IP(6,2)-C**
	$PSNR_y$ [dB]	36.78	42.01	40.91	36.73	42.07	40.91	36.67	42.01	40.92
	TR^u [$Mbps$]	**1**	**1**	0.86	0.91	0.91	0.91	**1**	**1**	**1**
	TR^s_U [$Mbps$]	**10**	**10**	8.6	9.05	9.10	9.10	**10**	**10**	**10**
	CR [$Mbps$]	3.30	3.23	**3.50**	**3.50**	**3.50**	**3.50**	3.21	3.29	3.36
Multicast	Optimal PS	**IP(6,2)-A**	**IBP(6,2)-A**	**IP(6,3)**	**IP(6,2)-B**	**IP(6,2)-B**	**IP(6,2)-B**	**IP(6,2)-C**	**IP(6,2)-C**	**IP(6,2)-C**
	$PSNR_y$ [dB]	36.21	41.43	40.70	36	41.23	40.61	35.98	41.13	40.65
	TR^u [$Mbps$]	0.88	0.88	0.75	0.85	0.86	0.83	0.85	0.81	0.83
	TR^s_U [$Mbps$]	**2.40**	**2.40**	**2.40**	**2.40**	**2.40**	**2.40**	**2.40**	**2.40**	**2.40**
	CR [$Mbps$]	2.88	2.84	3.00	2.74	2.73	2.74	2.72	2.72	2.78

- ***IP PS more efficient than IBP PS***—The IP PSs are genera-lly preferred over the IBP PSs. As the number of key-views increases, the expected transmission rates for the point-to-point communications (TR^u and TR^s_U) for an IP PS is lower than those for an IBP PS, under a similar distortion.
- ***Differences between unicast and multicast scenarios***—No significant differences are observed in Table II between the unicast and multicast scenarios. Still, PSs with a higher number of key-views tend to be preferred in the unicast scenario.

VI. Conclusion

We have studied the optimization of MVC prediction structures that minimize the visual distortion given storage and transmission rate constraints in IMVS systems. These rates have been analyzed, when interactivity is captured by a view-popularity distribution. Experimental results have shown that the optimal PS results from a tradeoff between interactivity and coding efficiency, and the optimal arrangement of the key-views depends on the view-popularity distribution. Future work will focus on the improvement of the current optimization algorithm to offer a low complexity solution allowing the online optimization of the PSs.

References

[1] A. Vetro, T. Wiegand, and G. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proc. IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[2] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[3] Y. Liu, Q. Dai, Z. You, and W. Xu, "Rate-prediction structure complexity for multi-view video coding using hybrid genetic algorithms," in *Proc. VCIP*, San Jose, CA, USA, 2007, p. 650804.

[4] X. Xiu, G. Cheung, and J. Liang, "Delay-cognizant interactive multi view streaming video with free viewpoint synthesis," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1109–1126, Aug. 2012.

[5] G. Cheun, A. Ortega, and T. Sakamoto, "Coding structure optimization for interactive multiview streaming in virtual world observation," in *IEEE Int. Workshop MMSP*, Cairms, Queensland, Australia, Oct. 2008, pp. 450–455.

[6] A. Fiandrotti, J. Chakareski, and P. Frossard, "Popularity-aware rate allocation in multi-view video coding," in *Proc. VCIP*, Huangshan, China, Dec. 2010.

[7] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in *IEEE Int. Conf. Multimedia and Expo*, Toronto, ON, Canada, Jul. 2006, pp. 1929–1932.

[8] "MERL test sequences for MVC," [Online]. Available: [Online]. Available: ftp://ftp.merl.com/pub/avetro/mvc-testseq

[9] "JMVC 8.2 software," [Online]. Available: [Online]. Available: garcon.ient.rwth-aachen.de

[10] "Tanimoto Laboratory test sequences for MVC-FTV," [Online]. Available: [Online]. Available: http://www.tanimoto.nuee.nagoya-u.ac.jp/

[11] Microsoft Research 3D Video Download [Online]. Available: [Online]. Available: http://research.microsoft.com/enus/um/people/sbkang/3dvideodownload/