

# Fast MVC Prediction Structure Selection for Interactive Multiview Video Streaming

Ana De Abreu<sup>1,2</sup>, Pascal Frossard<sup>1</sup>, Fernando Pereira<sup>2</sup>

<sup>1</sup>Signal Processing Laboratory (LTS4), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

<sup>2</sup>Multimedia Signal Processing Group, Instituto Superior Técnico - Instituto de Telecomunicações (IST-IT), Lisbon, Portugal.  
e-mail: {ana.deabreu, pascal.frossard}@epfl.ch; fp@lx.it.pt

**Abstract** — *Multiview Video Coding (MVC)* has been developed to efficiently compress a set of camera views by exploiting the spatial, temporal and interview correlations among images of the same scene. However, the resulting compressed data has a lot of prediction coding dependencies, which may not suit interactive multiview video streaming (IMVS) systems, where only one view is requested at a time by the end-user. This paper proposes a fast selection mechanism for effective interview prediction structure (PS) in IMVS while minimizing the point-to-point transmission rate, given some storage and visual distortion constraints, and a user interactive behavior model. Simulation results show that our novel fast MVC PS selection algorithm has high efficiency with low computational complexity that is reduced by more than 40% in comparison to the exhaustive searching benchmark.

**Index Terms**— *Interactive Multiview Video Streaming (IMVS), Multiview Video Coding (MVC), optimal prediction structure, view popularity model.*

## I. INTRODUCTION

After developing several (monoview) video coding standards, the main international standardization bodies decided to jointly specify the Multiview Video Coding (MVC) standard as a backward compatible extension of the very popular H.264/AVC (Advanced Video Coding) standard. MVC targets the efficient compression of multiview video by exploiting the interview, spatial and temporal dependencies present in this type of content [1]. However, maximizing the redundancy reduction, notably across the views, is not necessarily optimal for interactive multiview video streaming (IMVS) applications where only one view is requested at a time by the end-user. Indeed, the numerous coding dependencies may bring significant penalties as the request of one view typically implies the transmission of data from many other views that the target view depends on. IMVS systems challenges have not been much addressed in the literature, where the focus has mainly been the overall compression efficiency [2, 3]. However, more recently, some prediction structure (PS) selection mechanisms have been proposed for IMVS systems. In [4, 5], the authors have studied the PSs that facilitate a continuous view-switching [4] or reduce the transmission rate [5]. However, the authors have considered a coding system with redundant P- and DSC-frames (distributed source coding), which is not standard compliant, notably not compatible with the MVC standard. In [6], a PS selection mechanism has been proposed for distortion-minimized IMVS systems fulfilling some rate constraints. However, even when the set of possible PSs is limited, the complexity of this PS selection mechanism exponentially grows with the number of views in the multiview set, which is a major disadvantage.

This paper proposes a fast selection mechanism for the optimal interview PS in the context of IMVS systems considering scarce point-

to-point transmission rate and storage and visual distortion constraints. This fast PS selection considers view popularity models to characterize the end-user interaction behavior [6, 7]. The proposed PS selection mechanism is able to identify the optimal or a close to the optimal PS in the sense of minimizing the point-to-point transmission rate while fulfilling other relevant system constraints. At the same time, our PS selection algorithm leads to a complexity reduction larger than 40% compared to PS exhaustive searching. Contrarily to our previous work in [6], we propose now a constructive algorithm that speeds up the PS search with low performance penalty. This paper is organized as follows: Section II outlines the main characteristics of the IMVS system under consideration. Section III describes the PS optimization problem and the PS selection mechanism. Section IV presents performance results demonstrating the benefits of the proposed solution and, finally, the conclusions are presented in Section V.

## II. IMVS SYSTEM CHARACTERIZATION

In this work, we consider an IMVS system, where  $K$  views are simultaneously captured by a set of equally spaced cameras. Then, these views are jointly encoded and stored in a server able to provide an interactive view selection service to multiple end-users. The most relevant IMVS system characteristics are described in the following.

### A. MVC Coding Model

As mentioned in the introduction, in MVC, temporal and interview correlations are considered to achieve a coding gain. Here, the MVC temporal and interview coding models, to be considered in the optimization, have the following characteristics:

- **Temporal coding model** – As it is common in the literature, we assume a fixed temporal PS for each view, with *hierarchical B-frames* [2]. To control the quantization steps in the temporal domain, and thus the distortion, a *cascading quantization parameters (CQP)* [8] strategy is used. In this model, the set of QPs for the *anchor pictures* (those coded without temporal dependencies) in each view is selected; then, the QPs for the other temporal layers are assigned by increasing the QP of the previous layer with a pre-defined  $\Delta Q$ .
- **Interview coding model** – The interview coding model considered here is based on the two most common interview PSs for the MVC standard, namely *IBP* and *IP*. In these PSs, hierarchical B-frames are used in the temporal domain and the IBP or IP modes, for both anchor and non-anchor pictures, are used in the interview domain [2]. To reduce the interview redundancy, the PSs usually have only one independent view (*key-view*), which is typically a lateral view. Since, in IMVS systems, compression efficiency is not the only objective, we allow here more than one *key-view* in the two basic PSs (IBP and IP). We consider also the simulcasting PS where all the views are key-views (*I PS*). Differently from [6], where only a limited set of PSs is considered, this paper allows PSs with any

number of key-views, and they can take any position in the multiview set.

### B. Interactivity Model

Regarding the user interactivity, the following models are considered:

- **Random access** – We consider an IMVS system with random access where view-switches occur from/to any viewpoint in the multiview set but only at anchor pictures.
- **View popularity** – To model the end-user interactive activity, a *view-popularity factor*,  $p_i$ , is considered, to express the probability that an end-user selects view  $V_i$  (at any switching time instant).

### C. Rate and Distortion Models

Similarly to [6], we define the following coding and transmission rate and distortion models:

- **Coding Rate (CR)** – CR is defined as the total number of bits per unit of time necessary to code a multiview sequence and it may be computed as:

$$CR = f \frac{\sum_{g=1}^G \sum_{i=1}^K \sum_{j=1}^N n_b(F_{i,j}^g)}{GKN} \quad (1)$$

where  $f$  is the frame rate,  $G$  the total number of GOPs,  $K$  the number of views,  $N$  the number of frames per view in each GOP (we assume that all the views have the same GOP size), and  $n_b(F_{i,j}^g)$  the number of bits used to code frame  $F_{i,j}^g$ , from view  $V_i$  at time instant  $T_j$ , from GOP  $g$ . Frame  $F_{i,j}^g$  can be an I-, a P- or a B-frame.

- **Transmission Rate (TR)** – To define TR, a point-to-point transmission rate where a dedicated video stream is transmitted between two network points is considered. Before defining TR, we need to define the so-called *frame- and GOP- dependency path size*,  $\phi(F_{i,j}^g)$  and  $\phi_{i,g}$ . The *frame-dependency path size*  $\phi(F_{i,j}^g)$ , whose concept is similar to the *transmission cost* defined in [5], is the number of bits per GOP  $g$  that have to be processed to be able to decode a particular frame,  $F_{i,j}^g$ , and may be recursively defined as:

$$\phi(F_{i,j}^g) = n_b(F_{i,j}^g) + \sum_{k \in \{i+1, i-1\}} \phi(F_{k,j}^g) + \sum_{t \in \{1, \dots, N\} \setminus j} \phi(F_{i,t}^g) \quad (2)$$

where  $F_{k,j}^g$  and  $F_{i,t}^g$  are the immediate spatial and temporal reference frames of  $F_{i,j}^g$ , respectively. Frame  $F_{k,j}^g$  corresponds to the reference frame of  $F_{i,j}^g$  from the same time instant but from one of the two neighboring views (depending on the interview PS considered), while frame  $F_{i,t}^g$  is a reference frame from the same view  $V_i$  and GOP  $g$ . As a consequence, 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:

$$\phi_{i,g} = \sum_{j=1}^N \phi(F_{i,j}^g) \quad (3)$$

We can finally compute the *expected point-to-point transmission rate*, TR, as:

$$TR = f \frac{\sum_{g=1}^G E\{\phi_{i,g}\}}{GKN} \quad (4)$$

where  $E\{\phi_{i,g}\}$  is defined as  $E\{\phi_{i,g}\} = \sum_{i=1}^K p_i \phi_{i,g}$ , considering also the view popularity model,  $p_i$ .

- **Distortion (D)** – The *average distortion per view*,  $D_i$ , corresponding to the coding noise associated to the quantization process, is taken as the temporal average of the distortion per frame,  $D_{i,j}^g$ . It

reads  $D_i = \sum_{g=1}^G \sum_{j=1}^N D_{i,j}^g / NG$ . The distortion perceived by the end-user takes the value  $D_i$  with probability  $p_i$  (view-popularity factor); then, the *expected distortion* for the overall multiview sequence is defined as:

$$D = \sum_{i=1}^K p_i D_i \quad (5)$$

Hereafter, we use the terms distortion and transmission rate when referring to the *expected distortion* and *expected point-to-point transmission rate*, respectively.

## III. PROPOSED FAST PS SELECTION ALGORITHM

After describing the main characteristics of our IMVS system, we first formulate the optimization problem addressed in this paper. Then, we propose a fast selection mechanism for selecting the optimal PS, where the main target is to reduce the overall computational complexity while fulfilling the system constraints.

### A. Optimization Problem Formulation

The problem addressed here is to find the optimal interview PS,  $PS^*$ , which minimizes the transmission rate (as defined in Section II.C), while considering the following constraints on storage and distortion:

- **Storage capacity constraint** – The coding rate of the multiview video sequence, CR, is directly proportional to the number of bits used to code the multiview sequence, as seen in (1). In order to simplify the notation, we express the storage capacity of the system in terms of total number of bits per unit of time, namely  $CR_{max}$ .
- **Distortion constraint** – The distortion,  $D$ , shall not exceed the maximum acceptable distortion, namely  $D_{max}$ .

In summary, the optimization problem is written as follows:

$$PS^* = \arg \min_{PS} TR(PS) \quad (6)$$

such that

$$\begin{aligned} CR(PS) &\leq CR_{max} && CR \text{ constraint} \\ D(PS) &\leq D_{max} && D \text{ constraint} \end{aligned}$$

where the metrics CR, TR and D are calculated as in (1), (4) and (5); respectively.

To solve the optimization problem in (6), we will first, in a pre-processing step, select the optimal QPs satisfying the distortion constraint to select after the optimal PS minimizing TR given the CR constraint. However, if the distortion constraint is directed expressed in terms of QP, as it happens often in the literature, the step described in the next subsection is not needed.

### B. Pre-processing for distortion constraint fulfillment

To fulfill the *distortion constraint* in (6), the QPs for the anchor pictures are determined first by adopting an exhaustive search approach. The QP is the same for all anchor pictures while the CQP strategy is used in the temporal domain (Section II.B). Given the QPs that satisfy the various distortion constraints for a particular PS, the same QPs are then used for all the other allowed PSs, for a specific sequence, since no significant distortion variations are typically experienced while using the same QP for other PS. The use of the same QP for all the PSs (for the same sequence) reduces complexity and separates the optimization of the QPs and PSs. However, this might lead to some sub-optimality, even if in practice this is rarely the case. Note that we do not consider this pre-processing step when we compare the algorithms performance later in the paper.

### C. TR Minimization and CR Constraint: Selecting the Optimal PS

Solving the combinatorial optimization problem defined in (6) can

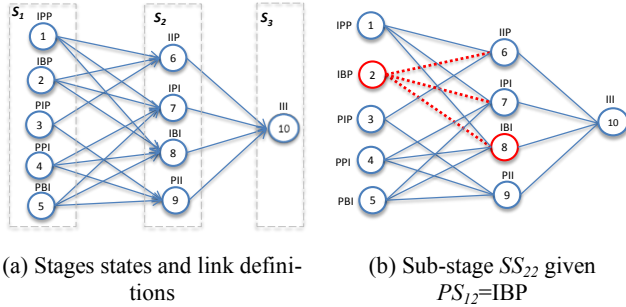


Fig. 1. Example of three stages graph definition and PS selection.

be very computationally intensive, notably if exhaustive search (ES) is applied. Indeed, the number of possible interview PSs exponentially grows with the number of views in the multiview set. To reduce the overall complexity regarding the ES approach, we propose a dynamic programming-based (DP-based) PS selection solution. With this approach, the problem in (6) is solved by breaking it down into a series of *stages*,  $S_i$ , which are solved successively one at a time. To better understand these different stages and how these stages depend on each other, we adopt a graph to embody all this information. In particular, the novel proposed optimal PS selection mechanism proceeds as follows:

1) **Stages graph creation** – The stages graph defines the various phases of the problem solution. Each stage has a set of associated states that represent the possible solutions at each phase of the problem. The states of consecutive stages are linked if certain conditions are fulfilled. In the following, we describe the two main steps in the stages graph creation process:

a) **Stage states definition** – We define the states in our graph in terms of the number of key-views in the interview PS, where the states in a particular stage correspond to the PSs with the same number of key-views, e.g., 1, 2, ...,  $K$ . We start by including in the first stage  $S_1$ , all possible PSs (for the considered IBP and IP basic PSs) with only one key-view, maximum number of interview coding dependencies – associated to *maximum coding efficiency and maximum transmission rate* – and we gradually increase the number of key-views in the PSs as we move towards the following stages, until the last stage,  $S_K$ , where all the  $K$  views are independently encoded, absence of interview coding dependencies – associated to *minimum coding efficiency and minimum transmission rate*.

b) **Links definition** – To link the states of two consecutive stages, we assume that the optimal PS in a particular stage  $S_i$ ,  $PS_i^*$ , determines the optimal position of the  $i$  associated key-views in the final optimal PS. Therefore, a link is defined between two states, defined by  $PS_{i-1,j}$  and  $PS_{i,l}$ , from stages  $S_{i-1}$  and  $S_i$  and states  $j \in S_{i-1}$  and  $l \in S_i$ , if the  $i-1$  key-views in  $PS_{i-1,j}$  preserve the same position in  $PS_{i,l}$ . Figure 1.a illustrates an example including the different states represented by circles, labeled with a unique number, and the links for the three stages case. The set of PSs in stage  $S_i$  linked to a same PS in stage  $S_{i-1}$ ,  $PS_{i-1,j}$ , is called sub-stage  $S_i$  and denoted as  $SS_{ij}$ ,  $\forall$  state  $j \in S_{i-1}$ . For instance, in Fig. 1.b, the IIP, IPI, IBI PSs define the sub-stage  $SS_{22}$  given  $PS_{12}=IBP$ , corresponding to state  $j=2$  of  $S_1$ .

2) **Iterative PS selection** – All stages are then successively processed, starting with stage  $S_1$ , until the adopted stopping criterion is fulfilled, meaning that the best PS solution has been found.

a) **Optimal PS selection at stage  $S_i$  ( $PS_i^*$ )** – The optimal PS for stage  $S_i$ ,  $PS_i^*$ , corresponds to the PS from  $S_i$  sub-stage,  $SS_{ij}$ , minimizing the cost function:

$$PS_i^* = \arg \min_{PS_i} TR_i(SS_{ij}, PS_i) \quad (7)$$

such that

$$CR_i(SS_{ij}, PS_i) \leq CR_{\max}$$

where  $SS_{ij}$  is defined by the optimal PS in stage  $S_{i-1}$ ,  $PS_{S_{i-1}}^* = PS_{i-1,j}$ .

To simplify the discrete optimization problem in (7), we apply a *Lagrangian relaxation*, where according to [9] the constraints are first relaxed; in our case, we move the  $CR$  constraint to the objective function with an associated multiplier  $\lambda \geq 0$ , and we then minimize:

$$L_i(\lambda) = \min \left\{ TR_i(SS_{ij}, PS_i) - \lambda (CR_{\max} - CR_i(SS_{ij}, PS_i)) \right\} \quad (8)$$

We solve the *Lagrangian Dual* problem [9], where the value of  $L_i(\lambda)$  is maximized by updating  $\lambda$ :

$$L_i^* = \arg \max_{\lambda} L_i(\lambda) \quad (9)$$

Finally, the optimal PS for stage  $S_i$ ,  $PS_i^*$ , is the one, out of the PSs in substage  $SS_{ij}$ , minimizing (8) for the  $\lambda$  value found in (9).

b) **Stopping criterion evaluation** – The decision to process the next stage or stop the PS selection mechanism at the current stage depends on the fulfillment of the following stopping criterion. If  $L_i^*$  is larger than  $L_{i-1}^*$ , then  $PS_{i-1}^*$  is a locally optimum solution as moving to the next stage will increase the Lagrangian cost, which is not desirable. For instance, using the example in Fig. 1.b, if  $L_3^*$  ( $PS_3^*=III$ ) is larger than  $L_2^*$  ( $PS_2^*=IBI$ ), IBI is selected as the final optimal PS as it was already selected in the previous step as the optimal  $S_2$  PS solution.

Note finally that, although the DP-based solution determines the optimal PS at each stage, the final PS may not be the global optimal one. However, as will be shown later with the performance results, the sub-optimal PS solutions are always very close to the global optimal ones, while the computational complexity is considerably reduced.

#### IV. PERFORMANCE RESULTS

This section presents the test conditions and performance results obtained in different scenarios when the PS search is performed with our novel algorithm.

##### A. Content and Coding Test Conditions

The MVC reference software JMVC v8.2 [10] has been used to encode the following three sequences: 1) *Ballet* [11], 5 views, 1024×768 pixels, 15 fps; 2) *Akko & Kayo* [12], 5 views, 640×480 pixels, 30 fps; 3) *Pantomime* [12], 10 views, 280×960 pixels, 30 fps. For these three sequences, a GOP size of 15 frames was used. In the temporal domain, the *CQP* strategy has been used, with a fixed  $\Delta Q$  equal to 0, 3 and 1 when the temporal layer is equal to 0, 1 and larger than 1, respectively. The adopted popularity distributions were uniform, exponential, Gaussian and U-quadratic [6]. Given the different sequence characteristics, the best PS has been found for various scenarios defined in terms of storage capacity and distortion (using the PSNR as metric):

- *Ballet scenario*:  $CR_{\max} = 260$  kbps;  $PSNR_{\min} = 30$ dB.
- *Akko & Kayo scenario*:  $CR_{\max} = 550$  kbps;  $PSNR_{\min} = 30$ dB.
- *Pantomime scenario*:  $CR_{\max} = 1900$  kbps;  $PSNR_{\min} = 30$ dB.

##### B. Results and Analysis

The results reported in Table I, with the optimal PSs in bold, compare the performance of the proposed DP-based algorithm with the exhaustive search (ES) approach, that always ensures finding the global optimal PS. Note that a different optimization problem was addressed in our previous work [6], so that this work can unfortunately not be used as benchmark in the performance evaluation of our novel solution, proposed in this paper.

Table I. DP-based (DP) and ES performances

Sequence	View Pop. Distribution	ES	DP ( $\Delta L$ [%], $\Delta T$ [%])
Ballet (5 views)	Uniform	<b>PIPII</b>	<b>PIPII</b> (0, 49)
	Exponential	<b>IIPIP</b>	<b>IIPIP</b> (0, 47)
	Gaussian	<b>PIIIP</b>	<b>PIIIP</b> (0, 48)
	U-Quadratic	<b>IPPII</b>	<b>IPPII</b> (0, 47)
Akko & Kayo (5 views)	Uniform	<b>PIPIP</b>	<b>PIPIP</b> (0, 41)
	Exponential	<b>PIPIP</b>	<b>PIPIP</b> (0, 41)
	Gaussian	<b>PIIIP</b>	<b>PIIIP</b> (0, 40)
	U-Quadratic	<b>IPPII</b>	<b>IPPII</b> (0, 51)
Pantomime (10 views)	Uniform	<b>PIPIPIPIP</b>	PIPIPIPIP (2.1, 85)
	Exponential	<b>IIIBPBPP</b>	IIIBPBPP (3.7, 85)
	Gaussian	<b>PIIPIPIP</b>	<b>PIIPIPIP</b> (0, 85)
	U-Quadratic	<b>PPBIIBPP</b>	IIIBPBPP (1.4, 65)

The comparison between the proposed DP-based algorithm and the ES approach is done here in terms of the Lagrangian cost (8), and computational complexity measured as CPU execution time. With this purpose, we use the normalized difference of the Lagrangian,  $\Delta L$ , and the execution time,  $\Delta T$ , both in percentage: the closer  $\Delta L$  is to zero, the closer the obtained PS solution is to the optimal solution. Moreover, the closer  $\Delta T$  is to 100%, the larger is the complexity reduction obtained with the proposed algorithm regarding ES. As it can be seen, in almost all the cases, the DP-based algorithm is able to identify the global optimal PS ( $\Delta L=0\%$ ) with a complexity reduction larger than 40% in comparison with the ES algorithm. The lowest accuracy is obtained by the *Pantomime* sequence, where a higher number of views tends to reduce the performance of the DP-based algorithm, given that the possible number of PSs is higher and the algorithm considers just a part of them as the associated complexity is highly reduced (above 60%).

Due to the reduction of accuracy of our algorithm with the *Pantomime* sequence, an additional analysis is performed to compare the *DP-based PSs* and all possible PSs. In Fig. 2, the solution obtained with the DP-based algorithm for the exponential popularity distribution and the *Pantomime* sequence is compared with the optimal PS and all other possible PSs with 1, 2 and 3 key-views found with the ES approach. Given the previous observations, the ideal working points in this chart for a PS would be the top left part of the chart, close to  $\Delta L=0\%$  and  $\Delta T=100\%$ . It can be seen that the DP-based solution is among the best ones (located in the top left part of the chart), even when Table I shows the lowest Lagrangian accuracy ( $\Delta L=3.7\%$ ). Moreover, the optimal PS obtained with the *ES algorithm* is located at the bottom left of the graph with both  $\Delta L$  and  $\Delta T$  equal to zero. In the ES approach, the PSs are sequentially processed in terms of the number of key-views, starting by one key-view. Therefore, in Fig. 2, PSs with one key-view, found with ES, have the highest complexity reduction ( $\Delta T=94\%$ ) and the ones with three key-views the lowest complexity reduction ( $\Delta T=0\%$ ).

In general, we can observe in Table I a *PS alignment with the popularity models* considered, where for both DP-based and ES algorithms, the chosen PSs accommodate the key-views on the most popular views. Also, the IP PSs are generally preferred over the IBP PSs. As the ratio of key-views increases in the multiview set, the transmission rate for an IP PS is lower than those for an IBP PS, since less interview dependencies are involved.

Different from common PSs in multiview systems, the best PSs, shown in Table I, have *more than one key-view*. This solution results from the trade-off between minimizing the transmission rate (associated to PSs with less interview dependencies) and compression efficiency, given storage constraint (associated to PSs with more inter-

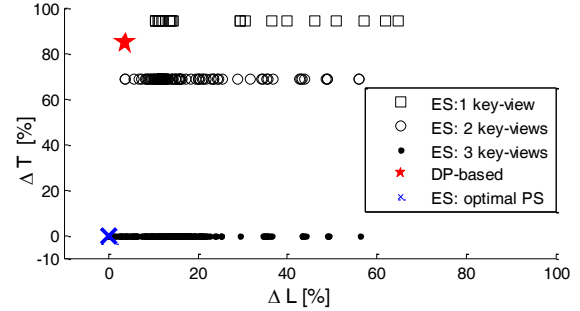


Fig. 2. Comparing PS selection mechanisms in terms of Lagrangian cost and execution time.

view dependencies). These results indicate that a *pure compression efficiency objective is not ideal in IMVS systems*.

## V. CONCLUSION

We have proposed an algorithm to efficiently select the MVC interview PS minimizing the transmission rate in IMVS systems, given some storage and visual distortion constraints. Simulation results have shown that even when the global optimal PS is not always obtained by the proposed algorithm, the associated complexity is considerably reduced while the sub-optimal solution is still closer to the optimal PS in Lagrangian cost. Future work will focus on the extension of the current optimization algorithm to systems where the synthesis of new views is possible by coding texture and depth for a small number of views.

## 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," in *Proc. of the IEEE*, vol. 99, n.º 4, pp. 626-642, Apr. 2011.
- [2] P. Merkle, A. Smolic, K. Muller and T. Wiegand, "Efficient Prediction Structures for Multi-view Video Coding," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, n.º 11, pp. 1461-1473, Nov. 2007.
- [3] Y. Liu, Q. Dai, Z. You and W. Xu, "Rate-Prediction Structure Complexity Analysis for Multi-view Video Coding Using Hybrid Genetic Algorithms," in *Proc of SPIE VCIP*, San Jose, CA, Jan. 2007.
- [4] X. Xiu, G. Cheung and J. Liang, "Delay-Cognizant Interactive Multi View Streaming Video with Free Viewpoint Synthesis," in *IEEE Trans. on Multimedia*, vol. 14, no. 4, pp. 1109-1126, Aug. 2012.
- [5] G. Cheung, A. Ortega and T. Sakamoto, "Coding Structure Optimization for Interactive Multiview Streaming in Virtual World Observation," in *IEEE Int. Workshop on MMSP*, Cairns, Queensland, Australia, Oct. 2008, pp 450-455.
- [6] A. De Abreu, P. Frossard, F. Pereira, "Optimized MVC Prediction Structures for Interactive Multiview Video Streaming," in *IEEE Signal Processing Letters*, vol.20, no.6, pp. 603-606, June 2013.
- [7] A. Fiandrotti, J. Chakareski and P. Frossard, "Popularity-Aware Rate Allocation in Multi-view Video Coding," in *Proc of SPIE VCIP*, Huangshan, China, Dec. 2010.
- [8] H. Schwarz, D. Marpe and T. Wiegand, "Analysis of Hierarchical B Pictures and MCTF," in *IEEE Int. Conf. on Multimedia and Expo*, Toronto, Ontario, Canada, July 2006, pp 1929-1932.
- [9] D. Bertsekas "Nonlinear Programming" 2nd Edition. Athena Scientific. 1995.
- [10] JMVC 8.2 software [Online]. Available: garcon.ient.rwth-aachen.de.
- [11] Microsoft Research 3D Video Download [Online]. Available: <http://research.microsoft.com/enus/um/people/sbkang/3dvideodownload/>
- [12] Tanimoto Laboratory test sequences for MVC-FTV [Online]. Available: <http://www.tanimoto.nuee.nagoya-u.ac.jp/>.