# Flow Optimization for Scalable Multi-view Video over Wireless Network

Yang Yang, Junni Zou
Department of Communication Engineering
Shanghai University, China
{Yang1989,zoujn}@shu.edu.cn

Hongkai Xiong
Department of Electrical Engineering
Shanghai Jiao Tong University, China
xionghongkai@sjtu.edu.cn

## ABSTRACT

This paper studies network performance optimization to maximize the overall user utility for scalable multi-view video (SMVV) streaming over wireless networks. We utilize a hybrid temporal-spatial content rate-distortion metric to measure the temporal and spatial quality of each view received at the user and define the corresponding user utility. The flow allocation problem is formulated as an utility maximization model for efficient SMVV streaming, then a distributed and layered solution based on Lagrangian dual decomposition is proposed, in which the rate allocation in each video layer can be solved by the shortest path algorithm and the sub-gradient algorithm. Finally, the performance of the proposed algorithm is verified by the simulation results.

## Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Abstracting methods*

## General Terms

Theory, Algorithms, Experimentation

## Keywords

Scalable multi-view video, flow optimization, temporal-spatial content distortion, rate allocation

## 1. INTRODUCTION

With the rapid deployment of three-dimensional (3-D) techniques, 3-D video is likely to be introduced to home and mobile platforms via broadcast or on-demand in the near future. As one popular 3-D video format, multi-view video (MVV) technique has attracted considerable attention recently. MVV is implemented by simultaneously capturing the same scene from different viewpoints. As MVV contains a large amount of inter-view dependencies, multi-view video coding (MVC) [1] is proposed to provide high compression efficiency by exploiting both temporal and inter-view redundancies. However, MVC requires to transmit the entire multi-view sequence to the users, which not only is infeasible when the view number is very large, but is unnecessary for the sake that, relying on the head position, the user in a period might be interested in a subset of views. Interactive MVV streaming is designed to provide MVV service efficiently and flexibly by only transmitting the views demanded by the users.

Interactive multi-view video streaming systems are studied in [2]-[5]. The MVV in [2] is encoded with simulcast coding method, in which each view is encoded and transmitted independently. Therefore, each client can select the number of views required according to the available bandwidth. The authors in [3] proposed a successive view motion model, which discriminates all frames into potential frames and redundant frames in terms of user's motion, and then those potential frames are encoded and transmitted to the user. Shi etc. in [4] presented a client-driven selective streaming system for multi-view video transmission. For minimizing the total video distortion of all clients, they proposed an optimal rate allocation algorithm in which the views are delivered based on client selections as well as network conditions.

As an annex of the advanced video coding standard, scalable video coding (SVC) [6] provides various operating points in spatial resolution, temporal frame rate, and video reconstruction quality. The incorporation of SVC into MVV coding produces a promising scalable 3-D video, where each view is encoded independently with simulcast coding using the SVC standard. The authors in [5] presented a client-driven selective streaming for multi-view video, in which SVC is combined with MVC to improve compression efficiency and provide adaptive bandwidth allocation to the selected views. However, they only discuss the case with two views and do not consider how to distribute the selected views. Zou etc. in [7] studied flow optimization for SVC streaming over multicast networks, where multi-path video streaming, network coding based routing, and network flow control are jointly optimized to maximize the network utility. The flow optimization solution in [7] is nevertheless specifically designed for single-view view transmission.

This paper is motivated to study network performance optimization to maximize the overall user utility for scalable multi-view video (SMVV) streaming over wireless networks. We utilize a hybrid temporal-spatial content rate-distortion metric [8] to measure the temporal and spatial quality of each view received at the user and define the corresponding


*ICIMCS'14,* July 10-12, 2014, Xiamen, Fujian, China.

user utility. The flow allocation problem is formulated as an utility maximization model for efficient SMVV streaming, then a distributed and layered solution based on Lagrangian dual decomposition is proposed, in which the rate allocation in each video layer can be solved by the shortest path algorithm and the sub-gradient algorithm. Finally, the performance of the proposed algorithm is verified by the simulation results.

The remainder of this paper is organized as follows: Sec. 2 presents the system model and the basic assumptions. Sec. 3 describes the flow optimization model, and the distributed solution to allocate rate for SMVV multicast. Simulation results are discussed in Sec. 4. Finally, Sec. 5 concludes the paper.

## 2. SYSTEM MODELS

### 2.1 Network Model and Assumptions

A wireless network can be modeled as a directed graph $G = (V, E)$, where $V$ is the set of network nodes and $E$ is the set of directed wireless links between nodes. The set V can be further divided into three disjoint subsets $S$, $N$ and $R$, representing source nodes, relay nodes and receiver nodes, respectively. All the nodes have a maximum transmission range $d_{\max}$. A directed link $(i, j) \in E$ exists between node $i$ and node $j$ if their distance satisfies $d_{ij} \leq d_{\max}$.

We consider a SMVV sequence with $M$ views, where each view is independently encoded using the SVC standard. For each view, the video encoder provides $v_{\max}$ spatial scalable levels and $\eta_{\max}$ temporal scalable levels for users with different display resolutions and available bandwidths. Assume that each scalable layer $l \in L$, $L = v_{\max} + \eta_{\max} - 1$, of each view $m \in M$ is distributed over a multicast group at a maximum transmission rate $R_{m,l}$. Also, assume that there exists multiple alternative paths $P(r)$ from the source to receiver $r$. We use a matrix $Z^r$ to reflect the relationship between its paths and related links. The $(j, e)$ entry of $Z^r$ is defined as

$$Z^r_{j,e} = \begin{cases} 0, & \text{if link } e \text{ is included in path } j; \\ 1, & \text{otherwise.} \end{cases}$$

### 2.2 Hybrid Temporal-Spatial Utility

We adopt temporal-spatial content rate-distortion metric [8] to measure the video temporal and spatial quality of each view. For each spatial level $v$, the rate and distortion in temporal domain are given by:

$$\begin{cases} R_{\eta,v}(\eta) = \frac{1}{G} \sum\limits_{i \in \left(V_v - \Lambda_{\eta,v}\right)} r_v\left(f_i\right) \\ D_{\eta,v}(\eta) = \frac{1}{G} \sum\limits_{i \in V_v} d\left(f_i, f_i^*\right) \end{cases} \tag{1}$$

where $G$ is the GOP size, with a GOP containing $V = \{f_1, \cdots, f_G\}$ frames, $\Lambda_{\eta,v}$ is the dropped frame set in temporal level $\eta$, $d\left(f_i, f_i^*\right)$ is the distortion between two frames' summary, $r_v\left(f_i\right)$ is the rate of frame $f_i$ in spatial level $v$. Note that the content information is represented by the video summary, and detailed information can be found in [9].

For each temporal level $\eta$, the rate and distortion in spatial domain are given by:

$$\begin{cases} R_{\eta,v}(v) = \frac{1}{G} \sum\limits_{i \in \psi_{\eta,v}} r_\eta\left(f_i\right) \\ D_{\eta,v}(v) = \frac{1}{G} \sum\limits_{i \in \psi_{\eta,v}} \left[d\left(f_i\right) - \frac{f'_w}{f_w} \cdot \frac{f'_h}{f_h} \cdot d\left(f'_i\right)\right] \end{cases} \tag{2}$$

where $\psi_{\eta,v} = \{f'_1, \cdots, f'_G\}$ represents the playback frames in spatial level $v$ and temporal level $\eta$, and $d(\cdot)$ is computed as the principle component analysis distance. When the user received just the frames in spatial level $v$, the reconstructed video will be stretched from received frame size in term of height $f'_h$ and with $f'_w$, to original size $f_h$ and $f_w$.

Let $(\eta^{m,r}, v^{m,r})$ represent the received maximum temporal and spatial level of view $m$ at user $r$, and all the dependent levels in $\{(i, j)|i \in [1, \eta^{m,r}], j \in [1, v^{m,r}]\}$ are supposed to be successfully received by user $r$. Then, the utility of view $m$ achieved at user $r$ is:

$$\begin{aligned} U^r_m = & \alpha \left[D_{\eta_{\max}} - D_{\eta^{m,r},v^{m,r}}\left(\eta^{m,r}\right)\right] \\ & + \beta \left[D_{v_{\max}} - D_{\eta^{m,r},v^{m,r}}\left(v^{m,r}\right)\right] \end{aligned} \tag{3}$$

Where $\alpha$ and $\beta$ are the respective influence parameters from temporal and spatial domains.

### 2.3 Wireless Channel Capacity Model

In wireless networks, the capacity of a wireless link is interrelated with other adjacent wireless links. Suppose any link originating from node $k$ will interfere with link $(i, j)$ if $d_{kj} < (1 + \Delta)d_{ij}$ or $d_{ki} < (1 + \Delta)d_{ij}$. Here, $\Delta \geq 0$ specifies the interference range. Also define $\Psi(i, j)$ for each link $(i, j) \in E$ as the cluster of links that cannot transmit as long as link $(i, j)$ is active, then the wireless network channel interference constraint can be defined as [10]:

$$\sum_{m \in M} \sum_{l \in L} f^e_{m,l} + \sum_{m \in M} \sum_{l \in L} \sum_{e' \in \Psi(e)} f^{e'}_{m,l} \leq C \tag{4}$$

where $f^e_{m,l}$ is the transmission rate of view $m$'s layer $l$ over link $e$, and $C$ represents the maximum rate supported by the wireless shared-medium.

## 3. OPTIMIZATION PROBLEM FORMULATION

### 3.1 Optimization Problem

In this paper, we seek optimal flow allocation for SMVV streaming over wireless multicast networks for maximizing the overall user utility in terms of the received spatial and temporal quality. Mathematically, it can be formulated as:

$$\textbf{P0:} \quad \sum_{r \in R} \sum_{m \in M} \sum_{l \in L} U^r_{m,l} \tag{5}$$

s. t.

1) $\sum\limits_{j \in P(r)} x^{r,j}_{m,l} \geq R_{m,l}, \ \forall r \in R, m \in M, l \in L.$

2) $\sum\limits_{j \in P(r)} z^r_{j,e} \cdot x^{r,j}_{m,l} \leq f^e_{m,l}, \ \forall r \in R, m \in M, l \in L, e \in E.$

3) $\sum\limits_{m \in M} \sum\limits_{l \in L} f^e_{m,l} + \sum\limits_{m \in M} \sum\limits_{l \in L} \sum\limits_{e' \in \Phi(e)} f^{e'}_{m,l} \leq C, \ \forall e \in E.$

$U^r_{m,l}$ in the objective function represents the distortion reduction at user $r$ when receiving view $m$'s layer $l$, either temporal level or spatial level. Constraint 1) guarantees that the actual transmission rate of layer $l$ of view $m$ achieved by each user $r$ is no less than the maximum flow rate of that layer. Constraint 2) represents the relationship between information flow rate and physical flow rate on each link. Here, $\sum_{j\in P(r)} z^r_{j,e}\cdot x^{r,j}_{m,l}$ denotes the total bandwidth allocated on link $e$ for view $m$'s layer $l$ to user $r$, and $f^e_{m,l}$ represents the physical flow rate of that layer over link $e$. By using network coding, different users would not compete for link bandwidth within the same video layer.

## 3.2 Distributed Solution

In pursuit of low computational complexity and decentralized implementation, we propose a distributed solution for problem **P0**. For each view, it sequentially allocates the bandwidth for each layer from low to high in a distributed manner, until there are no enough bandwidth left for the next layer. When constructing the data distribution meshes for each layer, its objective is to minimize the total bandwidth consumed at the current layer, so as to reserve more available bandwidth for the following higher layers. The flow optimization model for layer $l$ of view $m$ can be defined as:

$$\textbf{P1:}\quad \sum_{e\in E} f^e_{m,l} \tag{6}$$

s. t.

1) $\sum_{j\in P(r)} x^{r,j}_{m,l} \ge R_{m,l},\ \forall r\in R.$

2) $\sum_{j\in P(r)} z^r_{j,e}\cdot x^{r,j}_{m,l} \le f^e_{m,l},\ \forall r\in R, e\in E.$

3) $f^e_{m,l} + \sum_{e'\in\Phi(e)} f^{e'}_{m,l} \le C_{\text{res}},\ \forall e\in E.$

where $C_{\text{res}}$ is the capacity of the residual network, or the surplus capacity after finishing the data distribution meshes for previous $l-1$ layers.

In the achieved $l-1$ sequential distribution meshes for each view $m$, assume that there are $\eta^m$ temporal scalable levels and $v^m = l-1-\eta^m$ spatial scalable levels. Then, the goal of problem **P1** is to find an appropriate scalable level between $(\eta^m+1, v^m)$ and $(\eta^m, v^m+1)$, and the corresponding flow allocation, such that the aggregate utility of all the users achieved at layer $l$ is maximized, meanwhile, the bandwidth is reserved as much as possible for higher layers.

**P1** can be solved with Lagrangian relaxation and subgradient algorithm. By relaxing constraint 2) with Lagrange multiplier $u^r_e$, we obtain the following Lagrangian dual as:

$$\max_{\mathbf{u}\ge 0} L(\mathbf{u}) \tag{7}$$

where

$$\begin{aligned} L(\mathbf{u}) = &\min_P \sum_{e\in E} f^e_{m,l} \\ &+ \sum_{r\in R}\sum_{e\in E} u^r_e \left( \sum_{j\in P(r)} z^r_{j,e}\cdot x^{r,j}_{m,1} - f^e_{m,l} \right) \end{aligned} \tag{8}$$

$$\begin{aligned} &= \min_P \sum_{e\in E} f^e_{m,l}\left(1-\sum_{r\in R} u^r_{m,e}\right) \\ &+ \sum_{r\in R}\sum_{e\in E} u^r_e \left( \sum_{j\in P(r)} z^r_{j,e}\cdot x^{r,j}_{m,l}\right) \end{aligned} \tag{9}$$

and polytope $P$ is constraints 1) and 3) in problem **P1**.

It is observed that the Lagrangian problem in (9) can be further decomposed into two sub-problems. One is $|R|$ shortest path problems:

$$\textbf{P1a:}\quad \min \sum_{e\in E} u^r_e \left( \sum_{j\in P(r)} z^r_{j,e}\cdot x^{r,j}_{m,l}\right) \tag{10}$$

s. t. $\sum_{j\in P(r)} x^{r,j}_{m,l} \ge R_{m,l}(\eta,\nu),\ \forall r\in R.$

The other is a minimization problem:

$$\textbf{P1b:}\quad \min \sum_{e\in E} f^e_{m,l}\left(1-\sum_{r\in R} u^r_e\right) \tag{11}$$

s. t.

$$f^e_{m,l} + \sum_{e'\in\Phi(e)} f^{e'}_{m,l} \le C_{\text{res}}, \forall e\in E.$$

The shortest path problems in **P1a** can be tackled with the distributed Bellman-Ford algorithm. The minimization problem in **P1b** can be solved as follows:

$$f^e_{m,l} = \begin{cases} 0 & \text{if } \sum_{r\in R} u^r_e \le 1, \\ C_{res} - \sum_{e'\in\Phi(e)} f^{e'}_{m,l} & \text{if } \sum_{r\in R} u^r_e > 1. \end{cases} \tag{12}$$

In the $k^{\text{th}}$ iteration of the sub-gradient algorithm, the Lagrangian multiplier is updated by:

$$\begin{aligned} &u^r_e[k] = \max(0, u^r_e[k-1] + \\ &\theta(k)\cdot[\sum_{j\in P(r)} z^r_{j,e}\cdot x^{r,j}_{m,l} - f^e_{m,l}]), \forall r\in R, e\in E. \end{aligned} \tag{13}$$

where $\theta$ is a prescribed sequence of step sizes satisfying:

$$\theta[k]>0,\ \lim_{k\to\infty}\theta[k]=0, \sum_{k=1}^{\infty}\theta[k]=\infty.$$

Since the primal values in the optimal solution of the Lagrangian dual are not necessarily optimal to the primal problem **P1**, the algorithm proposed by [11] is applied further to recover the optimal values. For convenient description, we let $y^{r,e}_{m,l} \triangleq \sum_{j\in P(r)} z^r_{j,e}\cdot x^{r,j}_{m,l}(k)$. Then, at the $k^{\text{th}}$ iteration, we aggregate a primal $\tilde{y}^{r,e}_{m,l}[k]$ via

$$\tilde{y}^{r,e}_{m,l}[k] = \sum_{h=1}^{k} \lambda^k_h\, y^{r,e}_{m,l}[h], \tag{14}$$

where $\sum_{h=1}^{k}\lambda^k_h = 1$ and $\lambda^k_h \ge 0$ for $h=1,\cdots,k$.

The proposed distributed optimization algorithm for layer $l$ of view $m$ is showed in Algorithm 1.

**Algorithm 1** Distributed Optimization Algorithm for layer $l$

1. Initialize Lagrangian multipliers $u_e^r[0], \forall r \in R,\ e \in E$, to non-negative values.
2. Repeat the following iteration until sequence $\{u_e^r[k]\}$ converges to $u_e^{r*}$, $\left\{y_{m,l}^{r,e}[k]\right\}$ converges to $y_{m,l}^{r,e\,*}$: $\forall r \in R,\ e \in E$
   (1) Compute $y_{m,l}^{r,e}[k]$ by distributed Bellman-Ford algorithm;
   (2) Compute $f_{m,l}^{e}[k]$ by Eqn. (12);
   (3) Compute $\tilde{y}_{m,l}^{r,e}[k] = \sum_{h=1}^{k} \frac{1}{k} y_{m,l}^{r,e}[h] = \frac{k-1}{k} \tilde{y}_{m,l}^{r,e}[k-1] + \frac{1}{k} y_{m,l}^{r,e}[k]$;
   (4) Update Lagrangian multiplier $u_e^r[k] = \max(0, u_{m,e}^r[k-1] + \theta[k-1] \cdot [y_{m,l}^{r,e}[k-1] - f_{m,l}^{e}[k-1]])$.
3. Compute optimal rate $f_{m,l}^{e\,*} = \max_{r \in R} y_{m,l}^{r,e\,*}$, $\forall e \in E$.

# 4. SIMULATION RESULTS

In this section, we present simulation results to demonstrate the overall performance of the proposed optimization algorithm. The simulation experiments are conducted on a 3-view "Race1" sequence. Each view is encoded using H.264 extended SVC to generate scalable video stream with 3 temporal levels: 7.5, 15 and 30 fps and 3 spatial levels: QCIF, CIF and D1 formats. The wireless network topology for the simulation is showed in Fig. 1, where $S$ is the source node, $R_1 - R_5$ are relay nodes, $U_1 - U_3$ are user terminals, and $d_{\max}$=45m.

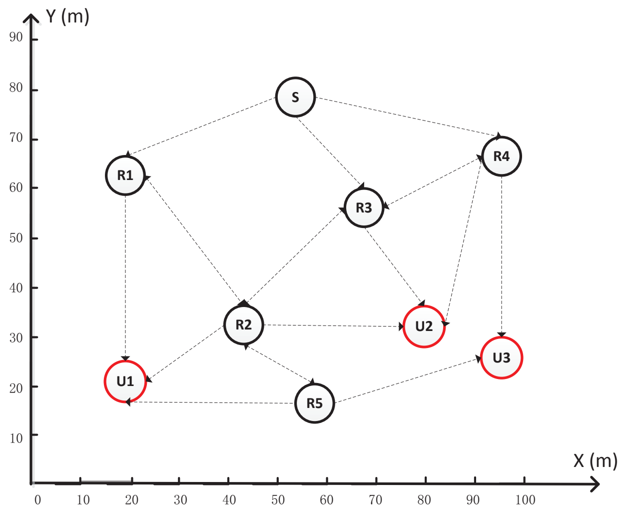



**Figure 1: Simulation network topology**

Fig. 2 shows the optimal layer structure for "Race1" sequence with hybrid temporal and spatial domain scalability support. For each view, the color line represents its layer structure achieved by the proposed flow optimization algorithm. For example, the green line of view 1 tells that view 1 will be distributed in an order of QCIF@7.5→ CIF@7.5→ D1@7.5→D1@15→D1@30. Namely, the bandwidth will be first allocated to the basic layer flow with QCIF resolution and temporal level of 7.5 fps, then sequentially to the enhancement layer flow with CIF resolution, D1 resolution, temporal level of 15 fps, and 30 fps.

Fig. 3 shows the allocated flow rate for each user at each layer. It can be seen that all flow rates reach the optimal value after 300 iterations. In practice, the convergence speed can be controlled by the step size. A large step size leads to a fast convergence, while a small step size yields to a smooth video quality.

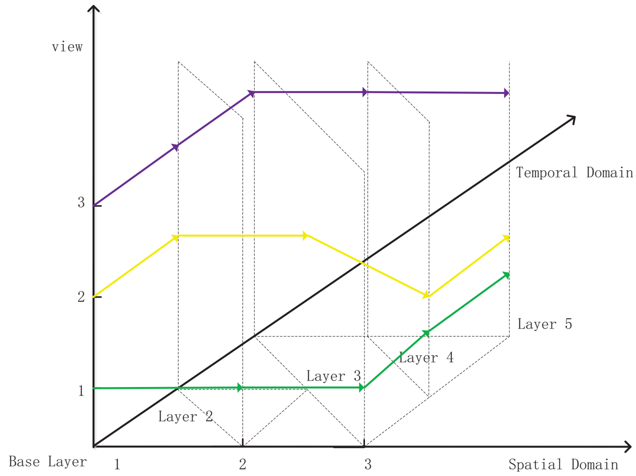



**Figure 2: The optimal layer structure for "Race1" sequence**

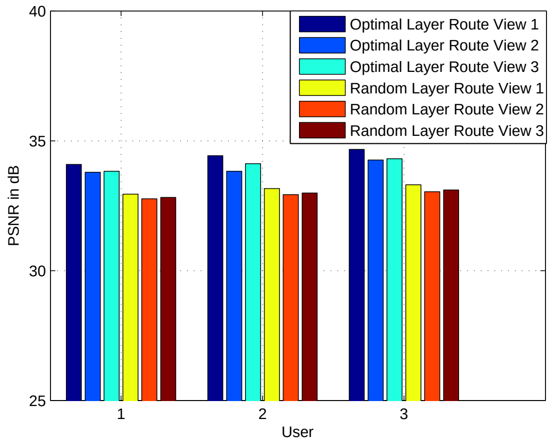
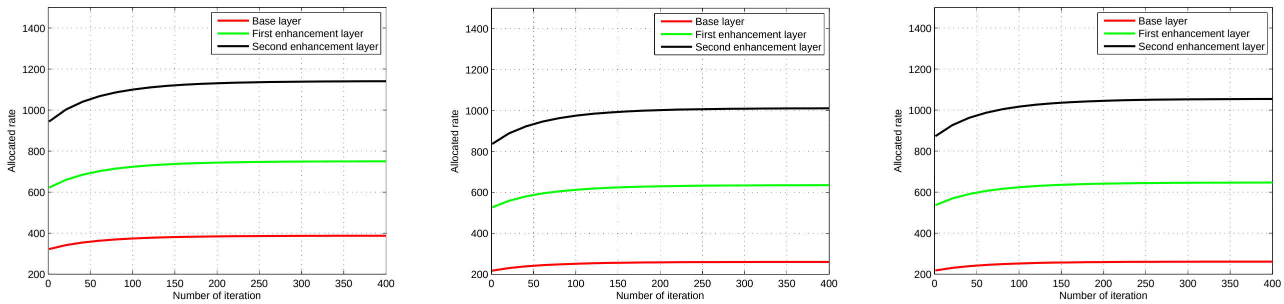



**Figure 4: Comparison of the achievable multi-view video quality.**

Fig. 4 compares the achievable multi-view video quality at each user with the layer structure achieved by the proposed algorithm and a random layer structure. Within the two layers, although three users can receive the data of all three views under these two layer structures, the total quality of the layers received by three users with the proposed layer structure is clearly better than the random layer structure.

# 5. CONCLUSION

In this paper, we proposed a framework of scalable multi-view video multicasting over wireless networks. We jointly considered the video quality scalability and channel capacity to maximize an aggregate utility function over heterogenous users. Also, a hybrid temporal-spatial layer structure and rate allocation optimization model was established. For




(a) view1 (b) view2 (c) view3

**Figure 3: Allocated rate for user 1. (a) For view 1. (b) For view 2. (c) For view 3.**

low computational complexity and decentralized implementation, the proposed model is decomposed into a set of submodels with each corresponds to a separate video layer distribution problem. Experimental results demonstrated that the proposed algorithm has a fast convergence speed and provides better video quality with heterogeneous users.

## 6. ACKNOWLEDGMENTS

The work has been partially supported by the grants from NSFC (No. 61271211).

## 7. 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," *Proceedings of the IEEE*, vol. 99, no. 4, pp.626-642, Apr. 2011.

[2] K. Jongryool, C. Kiho, L. Hyunyong, and W. K. Jong, "Multi-view 3D video transport using application layer multicast with view switching delay constraints," *3DTV Conference*, pp. 1-4, 2007.

[3] P. Ziyuan, Y. Ikuta, M. Bandai, and T. Watanabe, "A User Dependent System for Multi-view Video Transmission," *IEEE International Conference on AINA*, pp. 732-739, 2011.

[4] Z. Shi and J. Zou, "A Client-Driven Selective Streaming System For Multi-view Video Transmission," *International Forum of Digital TV & Wireless Multimedia Communication (IFTC'2012)*, 2012.

[5] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, "Client-driven selective streaming of multiview video for interactive 3DTV," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1558-1565, Nov. 2007.

[6] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of H.264/AVC,ąś *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, Sep. 2007.

[7] J. Zou, H. Xiong, C. Li, and T. Chen, "Prioritized Flow Optimization with Multi-Path and Network Coding Based Routing for Scalable Multirate Multicasting," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 259-273, Mar. 2011.

[8] W. Ji and Z. Li, "Heterogeneous QoS video broadcasting with optimal joint layered video and digital fountain coding," *Proc. IEEE ICC*, Jun. 2011, pp. 1-6.

[9] Z. Li, G. M. Schuster, A. K. Katsaggelos, and B. Gandhi, "Rate-distortion optimal video summary generation," *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 1550-1560, Oct. 2005.

[10] K.Yuen, B. Liang and B. Li, "Distributed Framework for Correlated Data Gathering in Sensor Networks,ąś *IEEE Tran. on Vehicular Technology*, vol. 57, no. 1, pp. 578-593, Jan. 2008.

[11] H. Sherali and G. Choi, "Recovery of primal solutions when using subgradient optimization methods to solve lagrangian duals of linear programs," *Operations Research Letter*, pp. 105-113, 1996.