

# Cross-layer optimized Authentication and Error Control for Wireless 3D Medical Video Streaming over LTE

Yanwei Liu<sup>a</sup>, Jinxia Liu<sup>b,\*</sup>, Antonios Argyriou<sup>c</sup>, and Song Ci<sup>d</sup>

<sup>a</sup>*State Key Laboratory of Information Security, Institute of Information Engineering,  
Chinese Academy of Sciences, 100093 Beijing, China*

<sup>b</sup>*Zhejiang Wanli University, Ningbo, China*

<sup>c</sup>*University of Thessaly, Volos, Greece*

<sup>d</sup>*University of Nebraska-Lincoln, Omaha, USA*

---

## Abstract

3D video for tele-medicine applications is gradually gaining momentum since the 3D technology can provide precise location information. However, the weak link for 3D video streaming is the necessary wireless link of the communication system. Neglecting the wireless impairments can severely degrade the performance of 3D video streaming that communicates complex critical medical data. In this paper, we propose systematic methodology for ensuring high performance of the 3D medical video streaming system. First, we present a recursive end-to-end distortion estimation approach for MVC(multiview video coding)-based 3D video streaming over error-prone networks by considering the 3D inter-view prediction. Then, based on the previous model, we develop a cross-layer optimization scheme that considers the LTE wireless physical layer (PHY). In this optimization, the authentication requirements of 3D medical video are also taken into account. The proposed cross-layer optimization approach jointly controls and manages the authentication, video coding quantization of 3D video, and the modulation and channel coding scheme (MCS) of the LTE wireless PHY to minimize the end-to-end video distortion. Experimental results show that the proposed approach can provide superior 3D medical video streaming performance in terms of peak signal-to-noise ratio (PSNR) when compared to state-of-the-art approaches that include joint source-channel optimized streaming with multi-path hash-chaining based-authentication, and also conventional video streaming with single path hash-chaining-based authentication.

*Keywords:* Cross-layer optimization, 3D medical video, Authentication, LTE, Error control.

---

\*Corresponding author. E-mail addresses: liujinxia1969@126.com, liuyanwei@iie.ac.cn

## 1. Introduction

Recent advances in video coding and wireless communication technologies have boosted the use of mobile health systems [1]. Supported by a significant amount of mature deployments of 4G networks, intelligent remote medicine is gradually becoming a reality. The goal is to provide people a “networked healthcare” at any time or any place via remote medical image or video processing [2][3]. One particular video technology can be key for medical applications. This is 3D video that has brought people the experience of depth in 3D cinema or 3DTV [4][5]. Because 3D video captures the third dimension, it can bring humans a more natural perception and precise location information. Consequently, besides 3D films or 3DTV [6], it can also be used for many other applications, such as 3D video surveillance, 3D medical video [7][8] and 3D games. For example in recent years, the advancement in 3D medical displays has made tele-surgical system a common practice [9]. Despite the great promise of 3D video for medical applications, several problems have to be carefully addressed.

A first problem is that 3D medical video generally includes more than one viewpoint. The stereoscopic/multiview video coding introduces the inter-view prediction that makes the inter-frame prediction dependency chain longer than before. Once the packet loss occurs, the distortion drifting will be significant. Hence, error control with rate adaptation at the application layer is necessary. Since the visual quality is very important, e.g. for the doctor at the remote side to accurately diagnose the illness of the patient, the inter-view prediction chain should be fully considered in determining the end-to-end video distortion. In previous research works, the end-to-end distortion estimation for 2D video has been studied thoroughly [10][11]. However, for MVC-based 3D video streaming, the existing end-to-end distortion estimation approaches typically utilize 2D video distortion estimation techniques [12][13] and neglect the particular inter-view/inter-frame-combined prediction-dependency problem. To the best of our knowledge, there is currently no work dealing with the hybrid inter-view and inter-frame prediction dependency problem for recursive 3D video end-to-end distortion estimation.

A second problem for 3D medical video communication is security [14]. Generally, the communication network can provide a level of security through certain protocols. However, besides protecting the video data during the transmission over the network, the confidentiality requirement for the private medical information needs to be met initially at the application layer during the medical video distribution. Currently, medical video security is ensured by watermarking and authentication [15] at the video source. Either watermarking or authentication will inject bits in the video to provide security functionality [16]. This type of protection for medical video will naturally introduce a redundant bitstream to the transmitted video. For video authentication, the added redundant authentication bits will occupy channel bandwidth and further affect the quality of video

received by the user [17]. So, the amount of authentication bits plays an important role for improving the streaming performance. Recent research works have proposed joint source-channel-coding-based authentication redundancy selection schemes in [18][19][20]. These approaches considered source rate adaptation to match the channel bandwidth [21], while they were not concerned with other parameters at the physical or other protocol layers. Overall, the optimal level of redundancy in terms of bit-rate that is acceptable to guarantee the end-user 3D video quality and security is still an open issue.

The final problem is that in such a remotely operated system, the 3D medical video data communication over the wireless network is the weak link in ensuring reliable end-to-end system performance. Usually, the dynamic and error-prone nature of the wireless network results in the degraded video quality. The video quality will degrade considerably since the limited bandwidth of the wireless network cannot always satisfy the demand of transmitting 3D medical video. To guarantee communication reliability, different techniques at different layers of the protocol stack have been adopted in LTE. For example, Adaptive Modulation and Channel Coding (AMC) at the physical layer, and Automatic Repeat Request (ARQ) at the data link layer. Although these techniques can meet the communication quality requirements of 3D medical video, they usually fail to achieve the optimal performance for wireless 3D medical video streaming since they neglect the impact of cross-layer interactions on the overall system performance. Now the physical layer of the LTE network provides different modulation and channel coding schemes (MCS) which can bring different transmission capacities and different error protection strengths. Thus, a balance between source coding and channel protection is needed. This is typically exercised through cross-layer optimization [22][23]. However, existing works focus mainly on the cross-layer optimization of adjacent layers, such as link layer and physical layer. For 3D medical video streaming, besides the system parameters present at lower protocol layers in the network, the particular feature of authentication in 3D medical video at the application layer needs also to be considered in the cross-layer optimization.

Based on the above discussion, in this paper we propose a cross-layer optimized authentication and error control approach for 3D medical video streaming. The contributions of the paper are summarized as follows.

1) A recursive per-pixel end-to-end distortion estimation approach is proposed for MVC-based 3D video streaming. For multiview video coding in the H.264/AVC, the inter-view prediction, especially the weighted inter-view and inter-frame prediction brings new problems for the recursive process. By utilizing the prediction-dependency relationships among frames, the recursion process for distortion estimation of inter-view predicted pixel is derived analytically.

2) A cross-layer optimization scheme that considers the authentication for 3D medical video streaming over LTE is designed. In the proposed scheme, the authentication hash-chaining paths, the quantization parameter (QP) of the video coder, and the MCS at the

physical layer are jointly selected to maximize the end-to-end quality of the 3D medical video. The resource block error rate of the LTE physical layer is used to estimate the application layer packet loss rate. The authentication success probability is incorporated to the packet loss rate for estimating the end-to-end distortion. Hence, the effect of authentication on the video transmission performance is fully considered in the proposed cross-layer optimization scheme. Hash-chaining-based authentication redundancy can be un-equally allocated among the different views under different channel conditions through the end-to-end distortion control.

The rest of the paper is organized as follows. The proposed cross-layer optimized error control and authentication scheme for 3D medical video streaming is presented in Section 2. Specifically, the end-to-end distortion computation for 3D medical video coding is introduced and the joint cross-layer optimization of authentication, video coding QP and MCS is described. Experimental details and results are shown in Section 3. Finally, Section 4 concludes the paper.

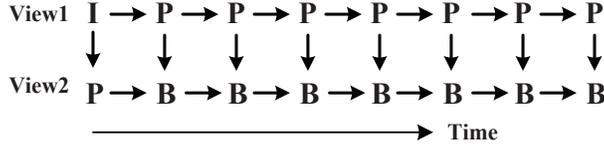
## 2. Cross-layer optimized 3D medical video streaming scheme

To improve the end-to-end quality of wireless 3D medical video, we design a cross-layer optimization scheme for 3D medical video streaming. After introducing the necessary background information, this section explains the proposed scheme in detail.

### 2.1. Background

#### 2.1.1. 3D medical video encoding

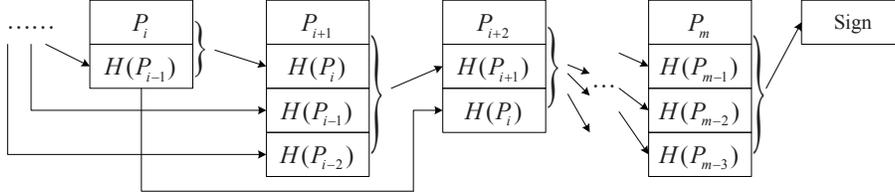
3D medical video is generally captured by two viewpoints to form the stereoscopic 3D video [24]. Thus, the stereoscopic encoding profile in the H.264/MVC standard is primarily used. To meet the decoding delay requirements, the low-delay encoding structure, as shown in Fig. 1, is also generally used in the practical applications [25]. In Fig. 1, I, P, B denote the intra-coded frame, directionally inter-coded frame and bi-directionally inter-coded frame, respectively. The inter-view prediction can be enabled to improve the 3D medical video coding efficiency with increased length of the dependent prediction chain. Due to additional prediction dependencies in the encoding process, the error drifting in the transmission will be significant once a packet loss occurs. Hence, the cross-layer optimized rate adaptation and error control algorithm should be designed to ensure a desired 3D medical video streaming quality. During encoding, a slice is the basic encoding unit which can be independently packed into one packet. The source packet size itself depends on the QP of video coding. A larger packet size is generated with small QP while the opposite happens for larger QP. In addition, the packet size will also affect its packet loss probability. Thus, the optimal QP should be carefully selected.



**Fig. 1.** 3D medical video encoding structure

### 2.1.2. Hash-chaining-based authentication

In the past, the hash-chaining-based authentication has been widely adopted for wireless video streaming [18][21][26]. The hash-chaining-based authentication organizes the packets into groups and signs only one packet in one group. The remaining packets in the group with the hash chain are connected to the signed packet. Since packet losses usually occur in wireless transmission, hash-chaining-based authentication can be disrupted. To ensure the error resilience of the hash-chaining-based authentication, the multi-path hash-chaining-based authentication was proposed in [21], and is illustrated in Fig. 2. In multi-path hash-chaining-based authentication, one packet can have many parent-chained packets. When one hash chain is corrupted due to packet loss, another hash chain can recover the authentication. However, the multiple hash chains increase the redundant hash bits and add overhead.



**Fig. 2.** Multi-path hash-chaining-based authentication

### 2.1.3. LTE downlink system

LTE is Long Term Evolution of the Universal Mobile Telecommunications System (UMTS) so that it is an advanced standard of cellular wireless communication. To maximize the use of the time/frequency resources, the LTE downlink transmission protocol adopts the Orthogonal Frequency Division Multiplexing (OFDM) technology. OFDM converts the wide-band frequency selective channel into a group of flat fading sub-channels with a subcarrier spacing of 15kHz. In LTE systems, the resource block (RB) is the smallest resource unit and consists of 12 adjacent sub-carriers and 6 or 7 consecutive OFDM symbols. To maintain reliable and efficient data transmission, LTE provides 15 candidate modulation and channel coding schemes (MCSs) to be selected depending on status of

transmission channel. The MCSs in the LTE downlink system have been investigated thoroughly recently [27][28].

## 2.2. Cross-layer Optimization

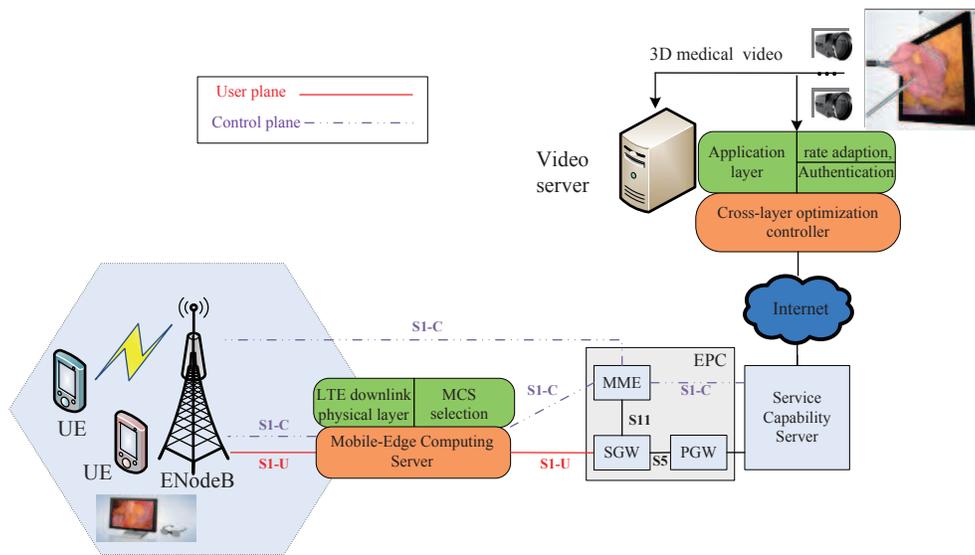
Even though several techniques in LTE systems can be utilized to improve the 3D medical video streaming performance, the parameters at different protocol layers typically counteract towards the goal of optimizing the overall system performance. For example, in the case of a bad channel state, the MCS modes with small constellation size and powerful channel codes are used to ensure transmission reliability. Accordingly, to adapt to the time-varying channel states and the selected MCS, the lower video bit-rate needs to be appropriately adjusted with the appropriate QP and the appropriate amount of authentication bits to ensure that the video packet is transmitted in a timely manner to the end user. In such a case, if the higher bit-rate (lower QP) is selected, the streaming system performance will be degraded due to the disharmony between the small size constellations with powerful channel codes and higher video bit-rate.

In the LTE system, the RB allocation across the multiple users is based on the ratio of user's minimum data rate requirement to the channel gain. The adopted RB allocation scheme considers the competing effect of multiple users by using proportional rate fairness among them [35]. After the RB allocation, the cross-layer optimization scheme operates for one particular user to further optimize the 3D medical video streaming performance from a global perspective. In the proposed scheme, the following cross-layer cooperations are investigated. First, the video coding QP at the application layer is taken into account to regulate the source bit-rate so that it matches the physical layer transmission capacity. Second, the multi-path hash-chaining-based authentication at the application layer is used to adjust the optimal authentication redundancy. Third, the MCS at the physical layer is dynamically selected to be in harmony with the dynamic source video bit-rate (the packet size under normal playback frame rate). These parameters at different protocol layers affect the end-to-end video quality and they can be jointly selected towards minimizing the end-to-end distortion.

Recently, mobile-edge computing (MEC) paradigm [29] has been proposed and implemented for service optimization over the radio access network (RAN). This provides a practical and feasible solution for cross-layer service optimization. The MEC server can communicate with an LTE base station (eNodeB) and the Mobility Management Entity (MME) in both the user plane and control plane [36]. Furthermore, the enhanced architecture models have been specified in 3GPP release 14 for LTE to provide 3GPP network service capability exposure to SCS (Service Capability Server) and AS (Application Server) [37]. Based on the models, the SCS can provide the AS with the control plane communication interface to the RAN. Thus, the external network entity (for example the

cross-layer optimization controller) can communicate with the RAN component in the control plane via the service capability exposure function.

Fig. 3 summarizes the proposed cross-layer optimization scheme for 3D medical video streaming over LTE. In this scheme, we use both the MEC server and the SCS component to realize the cross-layer optimized communication. The MEC server is placed near eNodeB. The SCS that is controlled by the mobile network operator is deployed near the Evolved Packet Core (EPC) network. Generally, the cross-layer controller takes charge of the end-to-end distortion estimation based on the information from the different protocol layers and the service provider takes care of the cross-layer optimization controller. The MEC server connects to the eNodeB of the LTE system for collecting mobile network feedback information from the radio access network. Then it transfers the collected information of different protocol layers to the controller. This is communicated in the data plane (user plane) via the S1-U interface. The cross-layer controller performs the necessary computations for the cross-layer optimization, and outputs the optimized parameters to the SCS. The SCS signals them to the MME and eNodeB in the control plane via S1-C interface to configure the optimized parameters of the different protocol layers for the given user. With the help of the MEC server and SCS, the cross-layer optimization can be practically implemented in LTE for 3D medical streaming.



**Fig. 3.** Cross-layer optimized authentication and error control scheme for 3D medical video streaming over LTE

### 2.2.1. Application layer video encoding and hash-chaining-based authentication

To ensure the video source rate adapts to the channel dynamics, a dynamic packet-based rate adjustment is implemented by the real-time video encoder. In the video encoder, one frame can be segmented into several slices which can be packed into independent packets. For real-time 3D medical video streaming, the packet size can be adjusted by regulating the encoding QP. The different QPs will result in different video distortions with different bit-rates.

Due to the multiple prediction dependencies among frames in 3D medical video encoding, the multi-path hash chain can be configured to increase the probability of authentication success during transmission. Thus, one packet in one hash group can select the hash parent from the previous decoding packets in the prediction chain pointing to the current packet. Hence, one packet may have multiple hash-chain parent selection scenarios. Based on the hash-chain knowledge of previous parent packets, the authentication probability of a current packet can be estimated and the corresponding redundancy in bits can also be computed. Different authentication hash-chains can result in different authentication success probabilities and different transmission overheads, and finally in different end-to-end video qualities.

To rigorously describe our system let us denote with  $N$  the maximum hash chain number of parents that a packet has. Each parent will incur a cost of  $b$  bits, and the probability that the packet is successfully authenticated is  $\varphi$ . The application layer packet loss probability before authentication in LTE system is denoted with  $\rho$ . Since each packet can have several candidate hash parents, the authentication probabilities for different hash chains are different. Let us assume the available candidate hash parents for the current packet are  $M$ , and then the authentication success probability for the  $i$ th packet in the current group of pictures (GOP) is obtained by

$$\varphi_i = (1 - \rho_i)(\varphi_0\varphi_1 \cdots \varphi_{i-1}), \quad (1)$$

where  $\varphi_i$  denotes authentication success probability of the  $i$ th packet in the current GOP and  $\rho_i$  denotes the packet loss probability of the  $i$ th packet in the current encoding GOP. Because stereoscopic video encoding is enabled at the 3D video encoder, the GOP for two views is considered as a processing unit for authentication.

If  $m$  ( $m < M$ ) packets are selected as the parent hash-chain packets by the  $i$ th packet and the  $n$ th parent packet has a distance  $d_n$  with the current  $i$ th packet, the authentication success probability of the current packet can be re-written as

$$\varphi_i = (1 - \rho_i)(\varphi_{i-d_0}\varphi_{i-d_1} \cdots \varphi_{i-d_n} \cdots \varphi_{i-d_m}) \quad (2)$$

Assume now that the added hash bit overhead is  $m \cdot b$  and the source packet size is  $S$ . The packet size will be  $S + m \cdot b$  after adding the hash bits. When the packet

passes through the physical layer, the packet might occupy different number of resource blocks (RB) depending on the amount of the added hash bits. At the physical layer, the MCS for each RB can be changed depending on the signal to noise ratio (SNR) of the channel. When the amount of hash bits is increased, the number of needed RBs and the application layer packet loss probability will increase accordingly. Hence, the added amount of hash bits directly affects the authentication success probability and packet loss rate of the application layer.

Besides the source coding induced distortion, packets with failed authentication also introduce the video distortion at the receiver. In addition, the authentication bits will increase the transmission overhead and reduce the transmitted source data, and thus indirectly increase the video distortion. Hence, the authentication is amenable to cross-layer optimization.

### 2.2.2. MCS at the physical layer of LTE

At the physical layer of LTE downlink, there are many candidate MCS modes. Different MCS modes will result in different physical layer transmission capacities with different packet loss rates. To estimate the packet loss rates for the transmitted packets, the mutual information effective SNR mapping (MIESM) is used to measure the LTE downlink quality. For one candidate MCS mode  $z$ , the effective SNR  $\gamma_{miesm}(z)$  based on mutual information is calculated as [27]:

$$\gamma_{miesm}(z) = \kappa(z) \left[ F^{-1} \left( \frac{1}{N_{sb}} \sum_{j=1}^{N_{sb}} F \left( \sqrt{\frac{\gamma_j}{\kappa(z)}} \right) \right) \right]^2 \quad (3)$$

where  $N_{sb}$  is the number of the subcarrier,  $\gamma_j$  is the SINR at the  $j$ th subcarrier, and  $\kappa(z)$  is the calibration factor for the MCS mode  $z$ . The functions  $F(x)$  and  $F^{-1}(y)$  are defined as

$$F(x) \approx \begin{cases} -0.04210610x^3 + 0.209252x^2 - 0.00640081x, & 0 < x < 1.6363 \\ 1 - \exp(0.00181491x^3 - 0.142675x^2 - 0.08220540x + 0.0549608), & x \geq 1.6363 \end{cases} \quad (4)$$

$$F^{-1}(y) \approx \begin{cases} 1.09542y^2 + 0.214217y + 2.33727\sqrt{y}, & 0 < y < 0.3646 \\ -0.706692 \log(-0.386013(y-1)) + 1.75017y, & 0.3646 \leq y \leq 1 \end{cases} \quad (5)$$

Based on the MIESM  $\gamma_{miesm}(z)$ , the block error rate (BLER)  $\xi(\gamma_{miesm}(z))$  for one RB with the MCS mode  $z$  can be accurately predicted as [27]

$$\xi(\gamma_{miesm}(z)) = \frac{1}{2} \left[ 1 - \operatorname{erfc} \left( \frac{\gamma_{miesm}(z) - b(z)}{\sqrt{2} \cdot c(z)} \right) \right] \quad (6)$$

where  $\operatorname{erfc}(\cdot)$  is the complementary error function,  $b(z)$  and  $c(z)$  are the transition center and transition width, respectively. The values of  $b(z)$  and  $c(z)$  can be obtained via

fitting (6) to the exact BLER in the specific communication system. In this work, the MIMO $2\times 1$  AWGN LTE downlink channel is simulated using the LTE link-level simulator [28]. BLER-SNR curves for the 15 MCS modes are used in the LTE downlink physical layer. The specific 15 MCS modes and the associated parameters are shown in Table 1.

**Table 1.** The candidate LTE downlink MCS modes

MCS mode index $z$	Modulation order	Rate (bits/symbol)	$\kappa(z)$	$b(z)$	$c(z)$
1	QPSK	0.1523	3.07	-7.758	0.6003
2	QPSK	0.2344	4.41	-5.724	0.5182
3	QPSK	0.3770	0.60	-3.652	0.4032
4	QPSK	0.6010	1.16	-1.593	0.3588
5	QPSK	0.8770	1.06	0.3501	0.2910
6	QPSK	1.1758	1.06	2.348	0.2563
7	16QAM	1.4766	0.87	4.297	0.2549
8	16QAM	1.9141	1.01	6.214	0.2293
9	16QAM	2.4063	1.04	8.242	0.2253
10	64QAM	2.7305	1.03	10.13	0.2248
11	64QAM	3.3223	1.11	12.06	0.2028
12	64QAM	3.9023	1.01	13.89	0.1962
13	64QAM	4.5234	1.07	15.72	0.1958
14	64QAM	5.1152	1.00	17.50	0.2134
15	64QAM	5.5547	1.05	19.59	0.2592

For video streaming over LTE, one packet may occupy several RBs at the physical layer. Thus, for each candidate MCS mode  $z$ , the packet loss probability  $\rho_{n,i}(z)$  of the slice  $s_{n,i}$  is related to the BLERs for all the RBs that the packet occupies,

$$\rho_{n,i}(z) = 1 - \prod_{k=1}^{B_{num}} (1 - BLER_k(\gamma_{meff}(z))) \quad (7)$$

where  $B_{num}$  is the RB number that the packet occupies, and  $BLER_k(\gamma_{meff}(z))$  is the BLER for the  $k$ th RB corresponding to the packet  $s_{n,i}$ .

### 2.2.3. End-to-end Distortion Estimation

During 3D medical video transmission, there are many parameters that will introduce video distortion. First is the source video coding. The lossy nature of video quantization introduces a non-recoverable video signal loss. During the transmission, the bit errors or packet losses will occur and will introduce video distortion. Packet loss will introduce error drifting in the video decoder due to the prediction dependency. In 3D medical video coding, the inter-view prediction increases the dependency in the motion compensation

procedure. To estimate accurately the end-to-end distortion of 3D medical video streaming, the recursive per-pixel estimation approach [30] is utilized to take into account the multi-prediction problem in the MVC. For easy understanding, the basic mathematical symbols and notations throughout the paper are listed in Table 2.

**Table 2.** Basic symbols and notations

Notations	Definitions
$f_n^i$	The $i$ th pixel value in the $n$ th frame
$\tilde{f}_n^i$	The $i$ th decoded pixel value in the $n$ th frame
$E[\cdot]$	The expectation function
$\tilde{f}_{n,l}^i$	The $i$ th decoded pixel value in the $n$ th frame of the left view
$\tilde{f}_{n-1,l}^i$	$i$ th decoded pixel value in the $(n-1)$ th frame of the left view
$\tilde{f}_{n-1,l}^k$	The $k$ th decoded pixel value for concealing the error of $i$ th pixel in the $(n-1)$ th frame of the left view
$\widehat{f}_{n,l}^i$	The $i$ th reconstructed pixel value at the encoder in the $n$ th frame of the left view
$\theta_{n,l}^i$	The authentication success probability of pixel $\tilde{f}_{n,l}^i$
$\theta_{n-1,l}^i$	The authentication success probability of pixel $\tilde{f}_{n-1,l}^i$
$\hat{e}_{n,l}^i$	The encoding prediction residue of pixel $\tilde{f}_{n,l}^i$
$\tilde{f}_{n-1,l}^j$	The encoding prediction value of $\tilde{f}_{n-1,l}^i$
$\tilde{f}_{n,r}^i$	The $i$ th decoded pixel value in the $n$ th frame of the right view
$\tilde{f}_{n-1,r}^i$	The $i$ th decoded pixel value in the $(n-1)$ th frame of the right view
$\tilde{f}_{n-1,r}^k$	The $k$ th decoded pixel value for concealing the error of $i$ th pixel in the $(n-1)$ th frame of the right view
$\widehat{f}_{n,r}^i$	The $i$ th reconstructed pixel value at the encoder in the $n$ th frame of the right view
$\theta_{n,r}^i$	The authentication success probability of pixel $\tilde{f}_{n,r}^i$
$\theta_{n-1,r}^i$	The authentication success probability of pixel $\tilde{f}_{n-1,r}^i$
$\hat{e}_{n,r}^i$	The encoding prediction residue of pixel $\tilde{f}_{n,r}^i$
$\tilde{f}_{n-1,r}^j$	The encoding prediction value of $\tilde{f}_{n,r}^i$ in the $(n-1)$ th frame
$\tilde{f}_{n,l}^o$	The encoding prediction value of $\tilde{f}_{n,r}^i$ in the left view

By incorporating authentication in the video transmission, the probability of successful authentication is

$$\theta_i = \varphi_i \cdot (1 - \rho_i) \quad (8)$$

When the original pixel value and the decoded pixel value are available, the end-to-end

distortion for one pixel can be computed as

$$\begin{aligned} D_n^i &= E[(f_n^i - \tilde{f}_n^i)^2] \\ &= (f_n^i)^2 - 2 \cdot f_n^i \cdot E[(\tilde{f}_n^i)] + E[(\tilde{f}_n^i)^2] \end{aligned} \quad (9)$$

Generally, the end-to-end distortion for one slice can be computed for all pixels in terms of mean squared error (MSE) [31]. Since the 3D medical video generally includes two views [18], the distortions for one packet with total pixels  $\aleph_l$  and  $\aleph_r$  for left view and right view can be computed respectively as

$$\begin{aligned} D_{n,l} &= \frac{1}{\aleph_l} \sum_{i \in \aleph_l} D_{n,l}^i \\ &= \frac{1}{\aleph_l} \sum_{i \in \aleph_l} [(f_{n,l}^i)^2 - 2 \cdot f_{n,l}^i \cdot E[(\tilde{f}_{n,l}^i)] + E[(\tilde{f}_{n,l}^i)^2]] \end{aligned} \quad (10)$$

$$\begin{aligned} D_{n,r} &= \frac{1}{\aleph_r} \sum_{i \in \aleph_r} D_{n,r}^i \\ &= \frac{1}{\aleph_r} \sum_{i \in \aleph_r} [(f_{n,r}^i)^2 - 2 \cdot f_{n,r}^i \cdot E[(\tilde{f}_{n,r}^i)] + E[(\tilde{f}_{n,r}^i)^2]] \end{aligned} \quad (11)$$

One notes in the above that the MSE is completely determined by the first and second moments of the decoder reconstructed pixels, and so the distortion per pixel in Eq. (10) and Eq. (11) can be recursively computed by considering the video encoding prediction and compensation. Note that the recursive way of the end-to-end distortion estimation already accounts for all parameters that affect it, like quantization, packet loss, error propagation, and error concealment at the decoder. In this paper, we assume that the decoder simply uses the previous frame for error concealment. Thus, the decoder reconstruction process provides the derivation of per-pixel distortion. Based on the classification of video coding prediction modes, the distortion can be estimated as the following cases.

### (1) Left view frame

If the current frame belongs to left view (intra view) and the current pixel belongs to an intra encoded macroblock (MB),  $E[(\tilde{f}_{n,l}^i)]$  and  $E[(\tilde{f}_{n,l}^i)^2]$  will be computed based on the specific authentication success probability as follows

$$\begin{aligned} E[\tilde{f}_{n,l}^i] &= \theta_{n,l}^i \cdot \left(\tilde{f}_{n,l}^i\right) + (1 - \theta_{n,l}^i) \cdot \theta_{n-1,l}^i \cdot E[\tilde{f}_{n-1,l}^k] \\ &\quad + (1 - \theta_{n,l}^i) \cdot (1 - \theta_{n-1,l}^i) \cdot E[\tilde{f}_{n-1,l}^i] \end{aligned} \quad (12)$$

$$\begin{aligned} E\left[\left(\tilde{f}_{n,l}^i\right)^2\right] &= \theta_{n,l}^i \cdot \left(\tilde{f}_{n,l}^i\right)^2 + (1 - \theta_{n,l}^i) \cdot \theta_{n-1,l}^i \cdot E\left[\left(\tilde{f}_{n-1,l}^k\right)^2\right] \\ &\quad + (1 - \theta_{n,l}^i) \cdot (1 - \theta_{n-1,l}^i) \cdot E\left[\left(\tilde{f}_{n-1,l}^i\right)^2\right] \end{aligned} \quad (13)$$

When the MB is encoded in inter-prediction mode, the motion compensation will be involved in the pixel reconstruction process. The reconstructed pixel value consists of the predicted pixel expectation and the prediction residue. Consequently,  $E[(\tilde{f}_{n,l}^i)]$  and  $E[(\tilde{f}_{n,l}^i)^2]$  will be computed by mimicking the decoding image reconstruction process,

$$E[\tilde{f}_{n,l}^i] = \theta_{n,l}^i \cdot (\hat{e}_{n,l}^i + E[\tilde{f}_{n-1,l}^j]) + (1 - \theta_{n,l}^i) \cdot \theta_{n-1,l}^i \cdot E[\tilde{f}_{n-1,l}^k] \\ + (1 - \theta_{n,l}^i) \cdot (1 - \theta_{n-1,l}^i) \cdot E[\tilde{f}_{n-1,l}^i] \quad (14)$$

$$E[(\tilde{f}_{n,l}^i)^2] = \theta_{n,l}^i \cdot \{(\hat{e}_{n,l}^i)^2 + 2\hat{e}_{n,l}^i \cdot E[\tilde{f}_{n-1,l}^j] + E[(\tilde{f}_{n-1,l}^j)^2]\} \\ + (1 - \theta_{n,l}^i) \cdot \theta_{n-1,l}^i \cdot E[(\tilde{f}_{n-1,l}^k)^2] \\ + (1 - \theta_{n,l}^i) \cdot (1 - \theta_{n-1,l}^i) \cdot E[(\tilde{f}_{n-1,l}^i)^2] \quad (15)$$

Based on the above equations,  $E[(\tilde{f}_{n,l}^i)]$  and  $E[(\tilde{f}_{n,l}^i)^2]$  for left view can be recursively computed from  $E[(\tilde{f}_{n-1,l}^i)]$  and  $E[(\tilde{f}_{n-1,l}^i)^2]$  until to the acknowledged frame. If one frame is acknowledged (no matter received or not), the acknowledged frame will be decoded and updated by the distortion estimation module for performing the above end-to-end distortion estimation process.

## (2) Right view frame

When the current frame is inter-view coded frame, more complex prediction relationship is involved in the video encoding. Consequently, the computation of  $E\{(\tilde{f}_{n,r}^i)\}$  and  $E[(\tilde{f}_{n,r}^i)^2]$  need to be discussed in different cases. The computation of  $E\{(\tilde{f}_{n,r}^i)\}$  can be classified into four categories:

1) If the current pixel is intra-frame predicted,  $E\{(\tilde{f}_{n,r}^i)\}$  will be computed as

$$E[\tilde{f}_{n,r}^i] = \theta_{n,r}^i \cdot (\tilde{f}_{n,r}^i) + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[\tilde{f}_{n-1,r}^k] \\ + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[\tilde{f}_{n-1,r}^i] \quad (16)$$

2) If the current pixel is inter-frame predicted, the motion compensation needs to be carefully considered. Thus,  $E\{(\tilde{f}_{n,r}^i)\}$  will be computed as

$$E[\tilde{f}_{n,r}^i] = \theta_{n,r}^i \cdot (\hat{e}_{n,r}^i + E[\tilde{f}_{n-1,r}^j]) + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[\tilde{f}_{n-1,r}^k] \\ + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[\tilde{f}_{n-1,r}^i] \quad (17)$$

where  $\hat{e}_{n,r}^i$  denotes the encoding prediction residue of the current encoding block.

3) If the current pixel is inter-view predicted, the inter-view compensation needs to be utilized in the pixel reconstruction.  $E\{(\tilde{f}_{n,r}^i)\}$  will be computed as

$$E[\tilde{f}_{n,r}^i] = \theta_{n,r}^i \cdot (\hat{e}_{n,r}^i + E[\tilde{f}_{n,l}^o]) + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[\tilde{f}_{n-1,r}^k] \\ + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[\tilde{f}_{n-1,r}^i] \quad (18)$$

4) If the current pixel is predicted by the weighting of inter-view and inter-frame predictions, more complicated compensation relations are involved into the pixel reconstruction.  $E\{\tilde{f}_{n,r}^i\}$  will be computed as

$$E[\tilde{f}_{n,r}^i] = \theta_{n,r}^i \cdot (\hat{e}_{n,r}^i + w_1 \cdot E[\tilde{f}_{n-1,r}^j] + w_2 \cdot E[\tilde{f}_{n,l}^o]) + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[\tilde{f}_{n-1,r}^i] \\ + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[\tilde{f}_{n-1,r}^k] \quad (19)$$

where  $\tilde{f}_{n,l}^o$  denotes the encoding prediction value of the right view pixel  $\tilde{f}_{n,r}^i$  in the left view, and  $w_1$  and  $w_2$  denote the weights of inter-view and inter-frame predictions that satisfy  $w_1 + w_2 = 1$ . From Eqs. (16) to (19), the pixel reconstruction at the decoder side is described based on the expectation of pixel value in the previous frame or the time-synchronized frame in adjacent view.

Similarly, the computation of  $E[(\tilde{f}_n^i)^2]$  can also be classified into four categories:

1) If the current pixel is intra-frame predicted,  $E[(\tilde{f}_n^i)^2]$  will be computed as

$$E\left[\left(\tilde{f}_{n,r}^i\right)^2\right] = \theta_{n,r}^i \cdot \left(\hat{f}_{n,r}^i\right)^2 + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E\left[\left(\tilde{f}_{n-1,r}^k\right)^2\right] \\ + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E\left[\left(\tilde{f}_{n-1,r}^i\right)^2\right] \quad (20)$$

2) If the current pixel is inter-frame predicted,  $E[(\tilde{f}_n^i)^2]$  will be computed as

$$E[(\tilde{f}_{n,r}^i)^2] = \theta_{n,r}^i \cdot \{(\hat{e}_{n,r}^i)^2 + 2\hat{e}_{n,r}^i \cdot E[\tilde{f}_{n-1,r}^j] + E[(\tilde{f}_{n-1,r}^j)^2]\} \\ + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[(\tilde{f}_{n-1,r}^k)^2] \\ + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[(\tilde{f}_{n-1,r}^i)^2] \quad (21)$$

3) If the current pixel is inter-view predicted,  $E[(\tilde{f}_n^i)^2]$  will be computed as

$$E[(\tilde{f}_{n,r}^i)^2] = \theta_{n,r}^i \cdot \{(\hat{e}_{n,r}^i)^2 + 2\hat{e}_{n,r}^i \cdot E[\tilde{f}_{n,l}^o] + E[(\tilde{f}_{n,l}^o)^2]\} \\ + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[(\tilde{f}_{n-1,r}^k)^2] \\ + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[(\tilde{f}_{n-1,r}^i)^2] \quad (22)$$

4) If the current pixel is predicted by weighting inter-view and inter-frame signals,  $E[(\tilde{f}_n^i)^2]$

will be computed as

$$\begin{aligned}
E[(\tilde{f}_{n,r}^i)^2] &= \theta_{n,r}^i \cdot \{(\hat{e}_{n,r}^i)^2 + 2\hat{e}_{n,r}^i \cdot (w_1 \cdot E[\tilde{f}_{n-1,r}^j] + w_2 \cdot E[\tilde{f}_{n,l}^o]) \\
&\quad + w_1^2 \cdot E[(\tilde{f}_{n-1,r}^j)^2] + w_2^2 E[(\tilde{f}_{n,l}^o)^2] + 2w_1w_2E[\tilde{f}_{n-1,r}^j \cdot \tilde{f}_{n,l}^o]\} \\
&\quad + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[(\tilde{f}_{n-1,r}^k)^2] \\
&\quad + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[(\tilde{f}_{n-1,r}^i)^2] \\
&\approx \theta_{n,r}^i \cdot \{(\hat{e}_{n,r}^i)^2 + 2\hat{e}_{n,r}^i \cdot (w_1 \cdot E[\tilde{f}_{n-1,r}^j] + w_2 \cdot E[\tilde{f}_{n,l}^o]) \\
&\quad + w_1^2 \cdot E[(\tilde{f}_{n-1,r}^j)^2] + w_2^2 E[(\tilde{f}_{n,l}^o)^2] + 2w_1w_2E[\tilde{f}_{n-1,r}^j] \cdot E[\tilde{f}_{n,l}^o]\} \\
&\quad + (1 - \theta_{n,r}^i) \cdot \theta_{n-1,r}^i \cdot E[(\tilde{f}_{n-1,r}^k)^2] \\
&\quad + (1 - \theta_{n,r}^i) \cdot (1 - \theta_{n-1,r}^i) \cdot E[(\tilde{f}_{n-1,r}^i)^2]
\end{aligned} \tag{23}$$

Based on the previous recursive computations of  $E\{(\tilde{f}_{n,r}^i)\}$  and  $E[(\tilde{f}_{n,r}^i)^2]$ , the end-to-end distortion for MVC-based 3D medical video streaming can be calculated.

#### 2.2.4. Cross-layer Optimization Formulation

Generally, to guarantee the smooth video playback at the receiver, the transmission time for the packet needs to be constrained by the normal playback frame rate. Let us denote the packet number for one frame  $N_p$ , the normal playback frame rate  $F_p$ , and the maximum allowable transmission delay for one packet  $T_{\max}$ , and we can obtain  $T_{\max} = \frac{1}{N_p \cdot F_p}$ . By considering the transmission delay constraint we can formulate the cross-layer optimization problem by minimizing end-to-end distortion as,

$$\begin{aligned}
(m_{hash}(d_0, \dots, d_{m-1})^{opt}, QP^{opt}, z^{opt}) &= \arg \min_{m \in H, QP \in Q, z \in \mathbb{Z}} D(m, QP, z) \\
\text{subject to } t &< T_{\max},
\end{aligned} \tag{24}$$

where  $m_{hash}(d_0, \dots, d_{m-1})^{opt}$  denotes the optimal hash chain parents for the current packet,  $QP^{opt}$  denotes the optimal quantization parameter for the packet to adapt to the available channel capacity, and  $z^{opt}$  denotes the optimal MCSs. In Eq. (24), the  $H$ ,  $Q$  and  $\mathbb{Z}$  denotes the candidate hash chain parents selection scenarios, the candidate QPs and the candidate MCS modes, respectively,  $t$  denotes the transmission time for the current packet and  $D(m, QP, z)$  is the end-to-end distortion for the packet under the combination of parameters  $m$ ,  $QP$  and  $z$ . Specifically, the instantaneous transmission time  $t$  of one packet is equal to  $\frac{S(QP)}{C_z}$ , where  $S(QP)$  denotes the packet size which is determined by the selected encoding  $QP$ , and  $C_z$  denotes the transmission rate provided by the MCS mode  $z$ . To minimize the video quality fluctuations among the frames and keep the balanced quality between the views, the calculations of  $Q$  and  $\mathbb{Z}$  for the current packet are relative to those values of the co-located packets in previous frames and adjacent views.

To solve Eq. (24), one can recursively compute the end-to-end distortion for each combination of parameters and then select one combination of parameters that lead to minimal distortion. However, this exhaustive search of the optimal parameters is time-consuming. To reduce the complexity of the cross-layer optimization, the branch and bound algorithm in [32] is used to reduce the search space for each type of parameters. Specifically, we prune the search space by reducing the candidate parameter set of the application and physical layers. At the physical layer, a subset of the MCSs result in very high packet loss probability when the channel SNR is low. In such cases, they are pruned. We utilized a packet loss rate bound as a threshold for pruning the sub-optimal MCSs. In our implementation, the target packet loss rate bound is set to 2%. At the application layer, the initial candidate QP range is shown in Table 3, and the QP range estimation based on the rate-quantization-model [32] is utilized to reduce the QP range. For the authentication, the hash chain parents will be reduced from 5 to 2 according to the PLR from 30% to 0 every other 10%. Thus the cross-layer optimization was implemented in practice based on a reduced parameter search space. The specific flowchart of the low complexity algorithm can be found in [32].

### 3. Experimental Results

We implemented the proposed cross-layer optimization scheme for 3D medical video streaming over LTE downlink with the H.264/MVC software and matlab software [33]. Specifically, we simulated the hash-chaining-based authentication scheme in the proposed cross-layer optimized 3D medical video streaming approach. In the simulation, the medical 3D clips, surgery with spatial resolution of  $960 \times 540$  and laparoscope (navigate) with  $640 \times 480$  [34] were used to verify the performance of proposed authentication and streaming approach. The specific experimental parameters are shown in Table 3.

To verify the advancements of the proposed approach, the state-of-the-art approaches including the streaming approaches with single-path hash-chaining-based authentication [16] and joint source-channel optimized multi-path hash-chaining-based authentication [17] were utilized as the baseline schemes. For having a fair comparison, we used the adaptive modulation and channel coding scheme in LTE at the physical layer when we experimented with the joint source-channel multi-path hash-chaining-based authentication.

#### 3.1. Distortion estimation accuracy

For the proposed medical streaming approach, accurate end-to-end distortion is the basis for high performance cross-layer optimization. To demonstrate the accuracy of proposed end-to-end distortion estimation approach, we performed distortion estimation for 3D medical video streaming at different channel conditions. Fig. 4 shows the distortion

**Table 3.** The specific experimental parameters

Coding structure	IPPP with inter-view prediction
Error concealment	Temporal replacement
Maximal reference frames	4
Candidate MCS mode	Table 1
Candidate QP range	20-45
Authentication	Hash-chaining-based approach
Average SNR	(2,4,9,14,20)dB
SNR distribution	Rayleigh
Transmission rate	1M symbol/s

estimation accuracy in terms of peak signal to noise (PSNR) for 3D surgery sequence under different channel conditions. Fig.4 (a) and (b) show the distortion estimation results for different encoding bit-rates with 3% packet loss rate (PLR) and Fig.4 (c) and (d) show the distortion estimation results for Surgery and Laparoscope with rates of 2200kbps and 1000kbps (average rate for each view) under different PLRs of 1%, 3%, 5%, and 7%. From Fig. 4, it can be seen that the proposed distortion estimation approach can accurately obtain the end-to-end distortion. For good channel quality (low packet loss rate), the accuracy of distortion estimation is as expected higher than that of bad channel quality (high packet loss rate). This observation reflects that the packet loss events are very random and the packet loss probability is difficult to be accurately predicted.

### 3.2. Authentication verification performance

For the proposed hash-chaining-based authentication approach, the authentication bit-rate can be adaptively added into the transmission to obtain the optimal authentication probabilities under different channel conditions in terms of the rate-distortion trade-off. Fig. 5 shows the authentication verification probabilities with single-path hash-chaining-based authentication (Singlepath), joint source-channel optimized multi-path authentication (JSCC-Multipath) and the proposed cross-layer optimized multi-path authentication (CR-Multipath). It can be seen from Fig. 5 that under bad channel conditions, the authentication success probabilities are relatively lower than those of good channel qualities for all three approaches. CR-Multipath can always achieve the highest verification probability among the three approaches. This is because CR-Multipath can dynamically select the authentication hash chain parent, and at the same time the source coding bit-rate dynamically adapts to the channel so that the packet loss rate is reduced and the authentication success probability is increased. The Singlepath approach only adds one

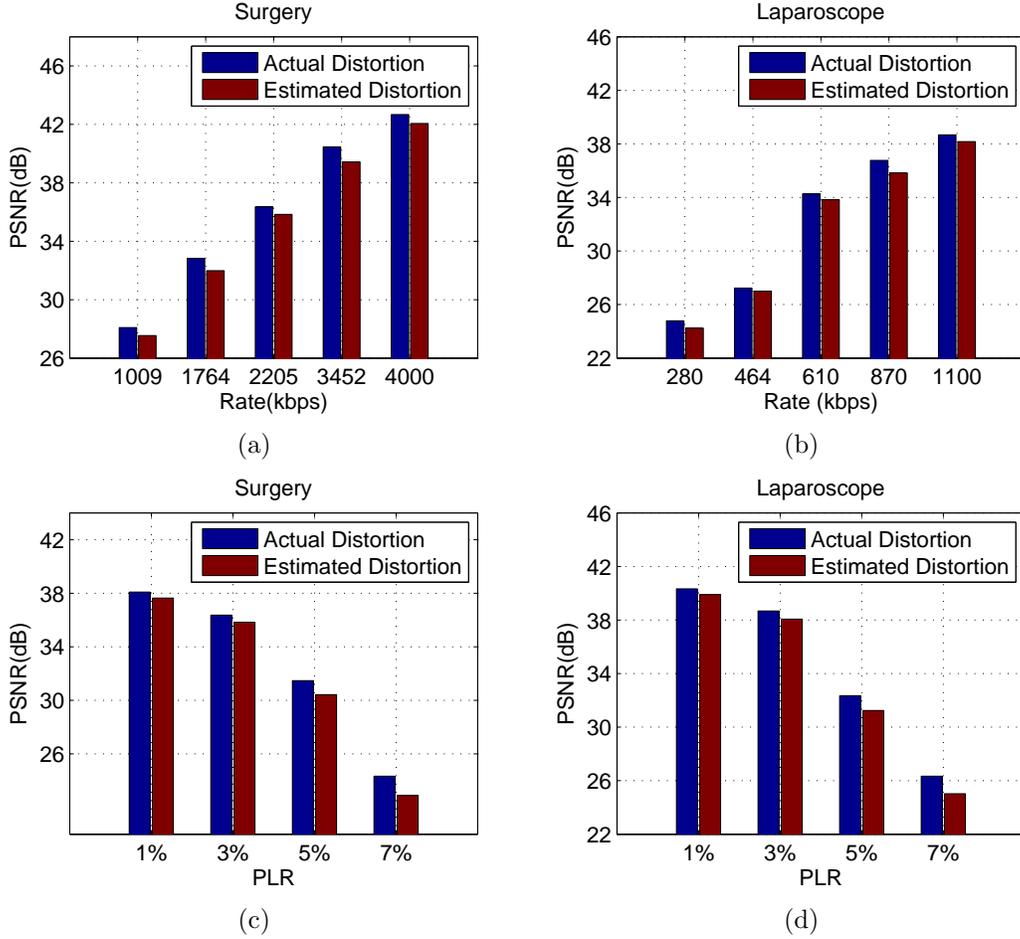
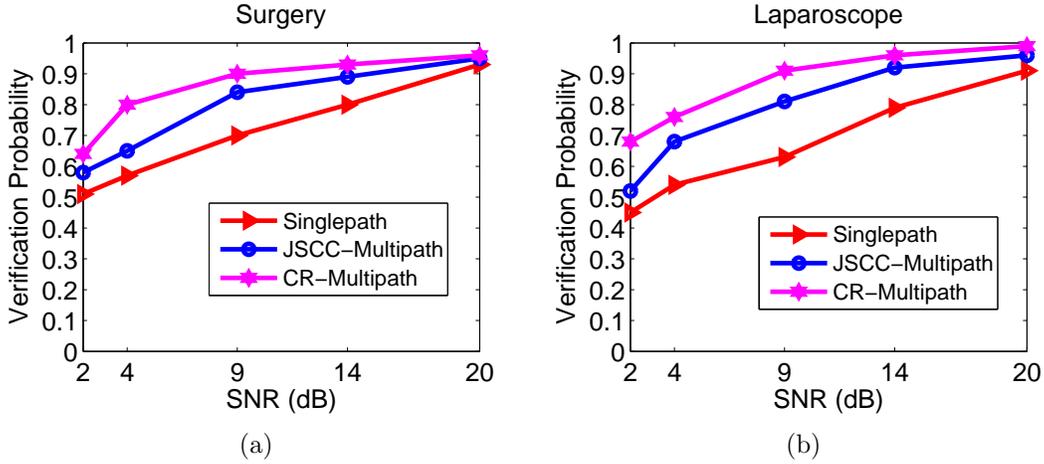


Fig. 4. Distortion estimation accuracy

authentication path in the transmission and it cannot successfully be authenticated when any packet in the path is lost. As for the JSCC-Multipath approach, it suffers from lower content-adaptive protection with a transmission time constraint for the packets and so it incurs a relatively higher delay-induced PLR and lower verification probability than the CR-multipath approach.

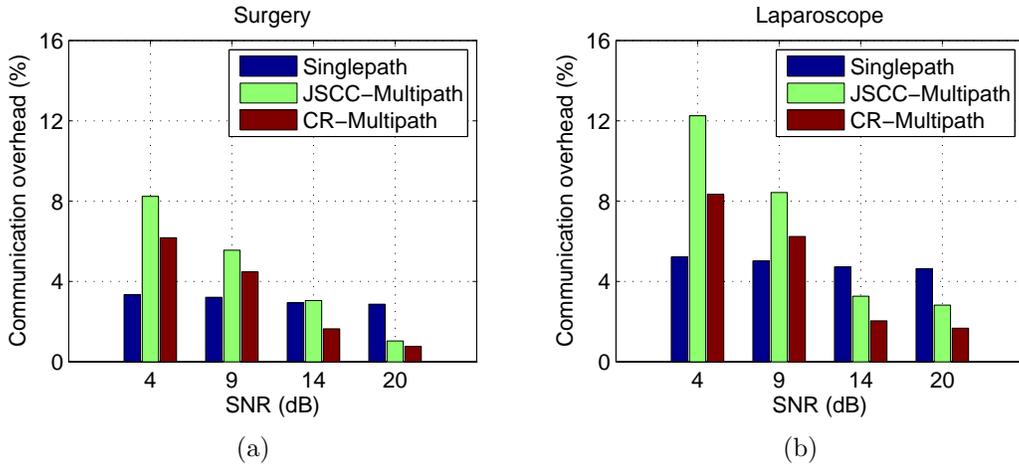
### 3.3. Communication overhead

We recorded the additional hash bits as the communication overhead to evaluate the proposed authentication approach. We compared the additional communication overheads for three authentication approaches under different channel conditions. Fig. 6 shows the overhead ratio over the total bit-rate for two views received by the end-user for different se-

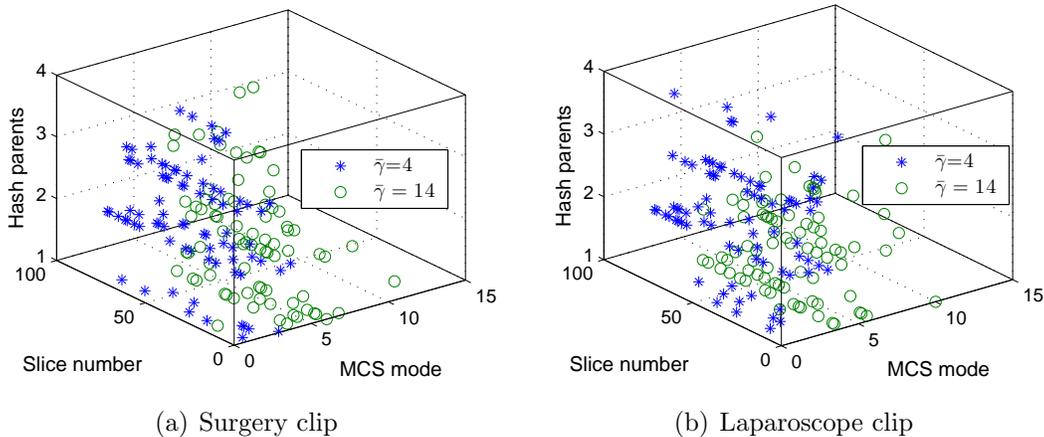


**Fig. 5.** Authentication verification probabilities under different channel conditions

quences. From Fig. 6, it can be seen that the overhead is gradually reduced with improved channel conditions for JSCC-Multipath and CR-Multipath. It indicates that the transmission bit-rate gradually increases with dynamic MCS and in the mean while the bit-rate received by the end-user increases with improved channel conditions. Consequently, the overhead ratio decreases with improved channel quality. The overhead for Singlepath approach is almost unaffected since a fixed MCS is used. For the overhead comparison between JSCC-Multipath and CR-Multipath, we note that the CR-Multipath adaptively regulates the authentication overheads with dynamic channel conditions so that it can reduce the overhead when compared to JSCC-Multipath.



**Fig. 6.** Communication overheads under dynamic channel conditions

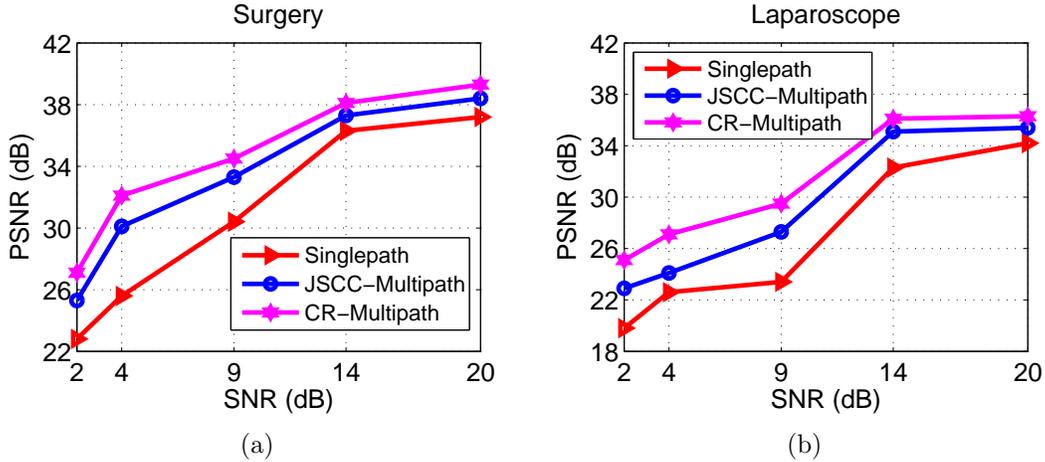


**Fig. 7.** The selected number of authentication hash parents and MCS modes for different slices

### 3.4. 3D Medical Video Streaming Performance

In the proposed system, the authentication redundancy bit-rate is adaptively added by cross-layer optimization. Hence, the received 3D medical video quality dynamically changes with the time-varying channel quality. The selected MCS modes and the number of authentication hash chain parents for a set of 3D video slices are shown in Fig. 7. From Fig. 7, it can be seen that the MCS modes with small size constellations and powerful channel codes (small MCS mode indexes), and large number of hash parents were selected to guarantee the transmission and authentication reliability under the transmission delay bound at the condition of  $\bar{\gamma}=4\text{dB}$ . For  $\bar{\gamma}=14\text{dB}$ , the MCS modes with larger size constellations (larger MCS mode indexes) and small numbers of hash parents were selected to increase the transmission capacity and decrease the authentication redundancy bit-rate. Thus, we see that the proposed cross-layer optimization scheme can properly select the MCS mode and authentication hash parent numbers to make the encoded video stream and the added authentication overheads dynamically adapt to the time-varying channel.

Fig. 8 shows the PSNR performance for 3D medical video streaming under different channel qualities. The PSNR value in Fig. 8 is the average PSNR for two views. At the conditions of  $\bar{\gamma} = 2\text{dB}$  and  $\bar{\gamma}=4\text{dB}$ , the proposed CR-Multipath approach improves the streaming performance significantly over the other two approaches. This is because the cross-layer optimized parameter selection makes the streamed content and the added authentication redundancy bit-rate adapt to the channel quality and further avoids possible packet losses. For  $\bar{\gamma}=14\text{dB}$  and  $\bar{\gamma}=20\text{dB}$ , the CR-multipath approach improve the 3D medical video streaming performance at a limited range over the other approaches since the packet loss is not very serious.

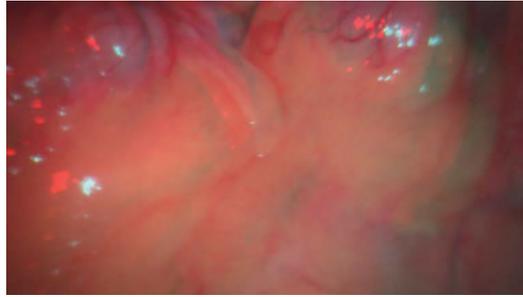


**Fig. 8.** The 3D video streaming performances under different channel conditions for different approaches

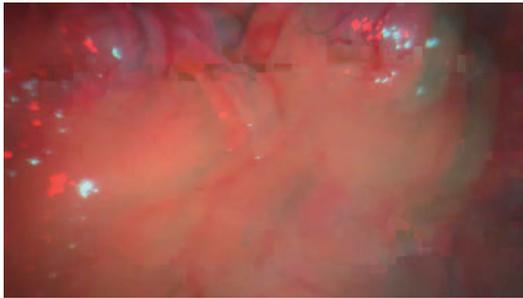
Five anaglyph 3D images (101th frame) of the sequence Surgery (one original frame, two decoded frames at the condition of  $\bar{\gamma}=4\text{dB}$  and two decoded frames at the condition of  $\bar{\gamma}=14\text{dB}$ ) are shown in Fig. 9. From the perceptual quality comparison of the images, it can be seen that the proposed cross-layer optimized 3D medical video transmission approach can provide superior subjective quality to the traditional streaming approach with joint source and channel coding optimized multi-path hash-chaining-based authentication and the streaming approach with the single path hash-chaining-based authentication. Especially, for SNR when the packet loss is serious, the proposed cross-layer optimization can take an active action on the suppression of packet loss. This can be verified by noticing the upper regions of images in Fig. 9.

#### 4. Conclusion

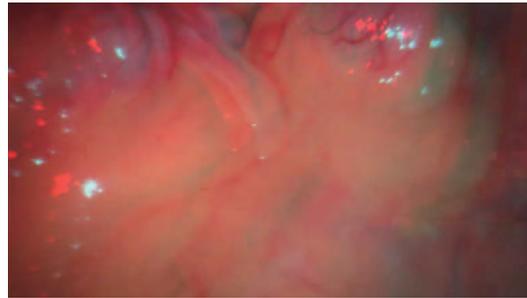
This paper presented a cross-layer optimized 3D medical video streaming and authentication approach for wireless telemedicine application over LTE. By considering the authentication requirements of 3D medical video, we performed a joint optimization of the parameters of different protocol layers including the application authentication redundancy and video coding bit-rate, and the physical layer MCS are performed to improve the error-resiliency capability of 3D medical video streaming. In the optimization, the MVC-based recursive end-to-end distortion estimation approach is derived by considering inter-view prediction. Extensive experimental results showed that the proposed cross-layer optimization approach can improve the 3D medical video streaming performance in terms of PSNR over a state-of-the-art joint source-channel optimized multi-path authentication



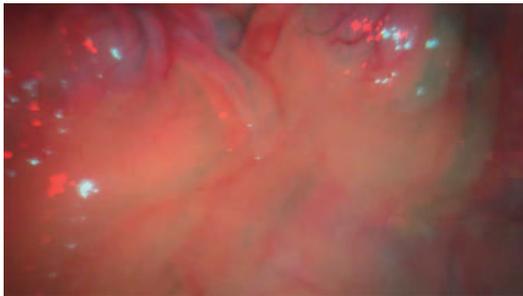
(a) Original 3D medical image



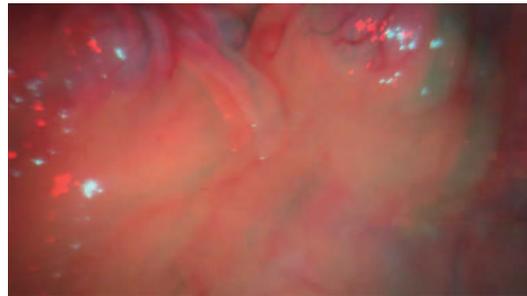
(b) JSCC-multipath decoded 3D medical image at the condition of  $\bar{\gamma} = 4\text{dB}$



(c) CR-multipath decoded 3D medical image at the condition of  $\bar{\gamma} = 4\text{dB}$



(d) JSCC-multipath decoded 3D medical image at the condition of  $\bar{\gamma} = 14\text{dB}$



(e) CR-multipath decoded 3D medical image at the condition of  $\bar{\gamma} = 14\text{dB}$

**Fig. 9.** Five anaglyph 3D images (101th frame) of Surgery ((a) one original frame, two decoded frames (b) and (c) at the condition of  $\bar{\gamma} = 4\text{dB}$ , and two decoded frames (d) and (e) at the condition of  $\bar{\gamma} = 14\text{dB}$ )

approach.

## Acknowledgement

This work was supported in part by Zhejiang Provincial Natural Science Foundation of China under Grant LY17F010001, Public welfare project of Zhejiang Province under Grant 2014C31072, Ningbo Natural Science Foundation under Grant 2015A610133 and National Natural Science Foundation of China under Grant 61472388.

## References

- [1] L. Constantinescu, J. Kim, D. Feng, SparkMed: A Framework for Dynamic Integration of Multimedia Medical Data into Distributed m-Health Systems, *IEEE Transactions on Information Technology in Biomedicine*, 16(1) (2012) 40-52.
- [2] J. Li, Z. Chen, and W. Li, A multimedia telemedicine system, In: 2010 International Conference on Biomedical Engineering and Computer Science (ICBECS), 2010, pp. 1-3.
- [3] X. M. Zhang and N. Zhang, An open, secure and flexible platform based on internet of things and cloud computing for ambient aiding living and telemedicine, In: International Conference on Computer and Management (CAMAN), 2011, pp. 1-4.
- [4] G. Tech, Yi. Chen, K. Müller, J. Ohm, A. Vetro, Y. Wang, Overview of the Multiview and 3D Extensions of High Efficiency Video Coding, *IEEE Transactions on Circuits and Systems for Video Technology*, 26(1), (2015) 35-49.
- [5] G. Tech, K. Wegner, Y. Chen, and S. Yea, 3D-HEVC Draft Text 6, Doc. JCT3V-J1001, Strasbourg, FR, Oct. 2014.
- [6] [Online]. Available: <http://www.3dmovielist.com/list.html>. Accessed on: Sep 20, 2016.
- [7] G. Welch, D. H. Sonnenwald, K. Mayer-Patel, et al., Remote 3D Medical Consultation, *Virtual Realities*, Dagstuhl Seminar 2008, Springer, Ch. 8, pp.139-160.
- [8] H. Liao, T. Inomata, I. Sakuma, and T. Dohi, 3-D Augmented Reality for MRI-guided Surgery using Integral Videography Autostereoscopic Image Overlay, *IEEE Transactions on Biomedical Engineering*, 57(6) (2010) 1476-1486.
- [9] D. Zhao, L. Ma, C. Ma, J. Tang, H. Liao, Floating autostereoscopic 3D display with multidimensional images for telesurgical visualization, *International Journal of Computer Assisted Radiology and Surgery*, 2015.

- [10] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for H.264 video coding over errorprone networks", *IEEE Trans. Circuits Syst. Video Technol.*, 9(3)(2007) 445-454.
- [11] S. Zamani, T. Nanjundaswamy, and K. Rose, "Recursive End-To-End Distortion Estimation for Error-Resilient Adaptive Predictive Compression Systems", In: *IEEE Workshop on Statistical Signal Processing*, Jun. 2016, pp. 1-5.
- [12] A. S. Tan, A. Aksay, G. B. Akar, and E. Arikan, "Rate-distortion optimization for stereoscopic video streaming with unequal error protection," *Eurasip J. Adv. Signal Process.*, 2009, (2009) 1-14.
- [13] E. Masala, "Rate-Distortion Optimized Low-Delay 3D Video Communications", In: *IEEE Intl. Workshop on Multimedia Signal Processing (MMSP)*, October 2010, pp.1-5.
- [14] H. Mohaghegh, N. Karimi, S.M.R. Soroushmehr, S. Samavi, K. Najarian, Adaptive Stereo Medical Image Watermarking Using Non-Corresponding Blocks, In: *37th Annual IEEE International Conference of the Engineering in Medicine and Biology Society*, 2015.
- [15] L. Guo, C. Zhang, J. Sun, Y. Fang, A Privacy-Preserving Attribute-Based Authentication System for Mobile Health Networks, *IEEE Transactions on Mobile Computing*, 13(9) (2014) 1927-1941.
- [16] R. Gennaro and P. Rohatgi, How to sign digital streams, *Information and Computation*, 165(1) (2001) 100-116.
- [17] X. Zhu, C. Chen, A Joint Source-Channel Adaptive Scheme for Wireless H.264/AVC Video Authentication, *IEEE Transactions on Information Forensics and Security*, 11(1) (2016) 141-153.
- [18] Z. Li, Q. Sun, Y. Lian, and C. W. Chen, Joint source-channel authentication resource allocation and unequal authenticity protection for multimedia over wireless networks, *IEEE Trans. on multimedia*, 9(4)(2007) 837-850.
- [19] Y. Wang, H. Wang, C. Wang, Graph-Based Authentication Design for Color-Depth-Based 3D Video Transmission over Wireless Networks, *IEEE Transactions on Network and Service Management*, 10(3)(2013) 245-254.
- [20] A. Boho, G. Van Wallendael, A. Doooms, et al., End-To-End Security for Video Distribution: The Combination of Encryption, Watermarking, and Video Adaptation, *IEEE Signal Processing Magazine*, 30(2) (2013) 97-107.

- [21] Q. Sun, J. Apostolopoulos, C. Chen, and S. Chang, Quality-optimized and secure end-to-end authentication for media delivery, *Proceedings of the IEEE*, 96(1) (2008) 97-111.
- [22] E. Setton, Taesang Yoo, X. Zhu, A. Goldsmith, and B. Griod, Cross-layer design of ad hoc networks for real-time video streaming, *IEEE Wireless Communications*, 12(4) (2005) 59-65.
- [23] A. Argyriou, Cross-layer error control for multimedia streaming in wireless/wireline packet networks, *IEEE Transactions on Multimedia*, 10(6) (2008) 1121-1127.
- [24] M. H. P. H. van Beurden, W. A. IJsselsteijn and J. F. Juola, Effectiveness of stereoscopic displays in medicine: A review, *3D Display Research*, 3(3) (2012) 1-15.
- [25] A. Vetro, T. Wiegand, and G. J. Sullivan, Overview of the Stereo and Multiview Video Coding Extensions of the H.264/AVC Standard, *Proceedings of the IEEE*, 99(4) (2011) 626-642.
- [26] Z. Zhang, Q. Sun, W. Wong, J. Apostolopoulos, S. Wee, Rate-Distortion-Authentication Optimized Streaming of Authenticated Video, *IEEE Transactions on Circuits and Systems for Video Technology*, 17(5) (2007) 1661-1664.
- [27] T.L. Jensen, S. Kant, J. Wehinger, and B.H. Fleury, Fast link adaption for MIMO OFDM, *IEEE Trans. on Vehicular Technology*, 59(8) (2010) 3766-3778.
- [28] J. Fan, Q. Yin, G.Y. Li, B. Peng, and X. Zhu, MCS selection for throughput improvement in downlink LTE systems, In: *International Conference on Computer Communications and Networks*, 2011, pp. 1-5.
- [29] [Online]. Mobile-Edge Computing-Introductory Technical White Paper, [https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge\\_Computing\\_-\\_Introductory\\_Technical\\_White\\_Paper\\_V1%2018-09-14.pdf](https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf). Accessed on: Sep 20, 2016.
- [30] R. Zhang, S.L. Regunathan, and K. Rose, Video coding with optimal inter/intra-mode switching for packet loss resilience, *IEEE Journal on Selected Areas in Communications*, 18 (6) (2000) 966-976.
- [31] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Second Edition, John Wiley & Sons, Inc. Hoboken, New Jersey, 2006.

- [32] P. Zhao, Y. Liu, J. Liu, A. Argyriou, S. Ci, SSIM-based Error-resilient Cross-layer Optimization for Wireless Video Streaming, Elsevier Signal Processing: Image Communication, 40, (2016) 36-51.
- [33] [Online]. Available: <http://www.nt.tuwien.ac.at/ltesimulator/>. Accessed on: Sep 20, 2016.
- [34] [Online]. Available: <http://hamlyn.doc.ic.ac.uk/vision/>. Accessed on: Sep 20, 2016.
- [35] Z. Shen, J. G. Andrews, and B. L. Evans, Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints, IEEE Trans. Wireless Commun., 4 (6) (2005) 2726-2737.
- [36] 3GPP TS 23.682, Architecture enhancements to facilitate communications with packet data networks and applications, (release 14).
- [37] C. Chang, K. Alexandris, N. Nikaen, K. Katsalis and T. Spyropoulos, MEC architectural implications for LTE/LTE-A networks, Proceedings of the Workshop on Mobility in the Evolving Internet Architecture, MobiArch'16, October 03-07, 2016, New York City, NY, USA.