Multi-view video allows users to watch 3D video and select desired viewpoints. In order to reduce the switching delay, many researches tend to transmit all the views, which brings much more increment in the bandwidth requirement. And the traffic increase as the increase of the number of the views in the multi-view video. In order to overcome these problems, we have proposed a user dependent scheme called UDMVT for the transmission of multi-view video in [1]. And we further improved this scheme to support the 3D multi-view video in the [2]. The evaluation results proved that UDMVT is efficient to reduce the bit-rate of the transmission of the multi-view video in the successive motion model especially when the number of views is larger. In this paper, we will simply introduce the UDMVT and discuss the application of the UDMVT in the Free Viewpoint TV.

Index Terms— Multi-view Video, User dependent, UDMVT, Encoding, Transmission, Application, Free Viewpoint TV

1. INTRODUCTION

The developments of camera and display technology make recording a single scene with multiple video sequences possible. These multi-view video sequences are taken by cameras array from different positions and angles. Each multi-view video sequence represents a unique viewpoint of this scene. Today, there are many applications for such multi-view video, such as 3DTV [3], Free Viewpoint TV (FVT) [4], Remote Surgery System and Wireless Multimedia Sensor Networks [5]. In order to provide 3D viewing experience, the multi-view video should at least consist of two views, each one for one of the two eyes. And other views allow the user to switch the viewpoints. As consisting of multiple views, multi-view video owns the size which is several times larger than traditional multimedia. In order to reduce the switching delay and continuously play the multi-view video without interrupt, many researches tend to transmit all the views, which brings the dramatic increase in the bandwidth. However, multi-view video contains a large amount of inter-view statistical dependencies since all cameras capture the same scene from different viewpoints. So encoding and transport technologies are especially important for multi-view video stream. The state of the art in multi-view representations includes single-view-plus-depth, Ray-Space and MVC.

However, the research on single-view-plus-depth sequences [6] suggests that with the addition of depth maps and other auxiliary information, the bandwidth requirements could increase as much as 20%. MVC (multi-view video coding) is issued as an amendment to H.264/MPEG-4 AVC [7]. It was reported that MVC can make significant compression gains than Simulcast coding in which each view is compressed with MPEG independently. However, even with the MVC, bit-rates for multi-view video are high: at about 5 Mbps is a common operating point for a 704 × 480, 30fps, and 8 camera sequences with MVC encoding [8].

In paper [1], we analyze the switching model and propose a user dependent scheme called UDMVT for the transmission of multi-view video transmission according to the analytical result of successive motion model. And we further improve this scheme to support the 3D multi-view video in paper [2]. The performance evaluation shows UDMVT can significantly reduce the bit-rate of multi-view video transmission and as the number of views increase, more bit-rate can be reduced. In this paper, we will simply introduce UDMVT and discuss the application of the UDMVT in the Free Viewpoint TV.

2. UDMVT

2.1. Switching models

According to the way with which users switch the viewpoint, the switching model can be basically classified into two types: 1) Random access model; 2) Successive motion model.
2.1.1. Random access model
In this model, users can switch to any views at any time. In this model, all frames at the same time instance have the same possibilities to be switched to. Which frames should be displayed in next period of time are unpredictable. All the frames should be transmitted to avoid the switching delay.

2.1.2. Successive motion model
User is just able to switch from current viewpoint to the neighboring views. In other word, for any view $j$, user is just able to switch from $j$ to the view $j'$, where $\max(1, j-1) \leq j' \leq \min(j+1, M)$. $M$ is the number of the view in the multi-view video. This switching model is used in the applications such as Free Viewpoint TV and Remote Surgery System in which system decides the viewpoint by tracking users' head.

In the successive motion model, the frames to be displayed are predictable. In this paper, our works mainly focus on the successive motion model and the motion of the user is assumed to be uniform motion.

2.2. Analysis of Successive Motion Model

In successive motion model, which frames should be displayed when the user starts to switch to next view is decided by the both frame rate (frame/sec) of the video and the switching speed (view/sec) of the user. Let $k$ be the floor of the frame rate divided by switching speed: $k = \lfloor \frac{f}{s} \rfloor$ in which $f$ denotes frame rate while $s$ denotes the switching speed. So $k$ presents the number of the frames should be displayed in the current view between the user starts to switch and the display changes to the next view. Fig. 2 shows the display of frames when $k$ is 2 and 1.

2.3. UDMVT

By three-tuples $N(p, f, s)$, it is able to predict a range in which the frames may be displayed in the next a period time. $p$ is the initial position $[F_{i0,j0}, F_{i0,j0-1}]$ and $f$ is the frame rate. $s$ represents the switching speed of the user. $R(t)$ is the set of the frames pair $[F_{i,j}, F_{i,j-1}]$ that may be displayed at time $t$, in which:

$$i = i_0 + \lfloor f \times t \rfloor$$

$$j' \in \left[ \max(2, j_0 - \lfloor s \times t \rfloor), \min(j_0 + \lfloor s \times t \rfloor, M) \right]$$

The ranges of the frames are shown in Fig. 3 when $k$ is 1 and 2, respectively. The frames in the range are called potential frames (PFs), which means these frames are possible to be displayed. Those frames outside the range are the redundant frames (RFs). It can reduce the bit-rate of the multi-view video transmission if these RFs are not transmitted.

![Fig. 3. The ranges of the frames when $k$ = 1 (a) and $k$ = 2 (b)](image)

Fig. 4 shows the time sequence chart of the UDMVT. The $N(p, f, s)$ should be fed back periodically to encode and transmit next range of frames. In order to continuously play the video without interrupt, the the $N(p, f, s)$ should be fed back before playing to the end of current range because of the delay. This delay contains four components: feedback delay ($t_1$), encoding delay ($t_2$), transmission delay ($t_3$) and decoding delay ($t_4$). The $N(p, f, s)$ should be detected and fed back ($t_1+t_2+t_3+t_4$) before the end of current range. So when the video is played to the end of the current range, frames of next range have been received and decoded. It is able to keep on playing the video without interrupt. This process will be repeated with the feedback period $\tau$ to the end of this video.

![Fig. 4. Time sequence chart](image)
2.4. Evaluation

The evaluation results have been obtained by the multi-view video test sequences “ballroom” with 8 views and 640 \times 480 resolution. These test sequences are provided by MERL [9]. Two reference schemes are 1) Simulcast and 2) MVC.

![Average Bit Rate of Each Scheme](image)

Fig. 5. Average bit rate of each scheme

Fig. 5 depicts the average bit rate (ABR) in comparison to reference schemes. It can be seen that the ABRs of Simulcast and MVC increase as the increase of the number of the view. On the other hand, in the UDMVT, only those frames that are possible to be displayed are encoded and transmitted according to the value of k. Only part of frames is encoded and transmitted no matter how many views are in the multi-view video. So when the number of the views is bigger enough, the ABR will not increase any more. So as the increase of the number of views, more ABR are decreased.

3. APPLICATION

3.1. Free Viewpoint TV (FVT)

Conventional TV only provides a single view of a real 3D world. The users just can watch a fixed view decided by the program producer and cannot control their viewpoint. It is quite different from the experience in a real world. 3DTV attracts much attention today. There are already many productions of 3DTV in the market. It provides users the three dimension perception. However, viewing zone of 3DTV is still very limited. A FVT have many views that taken by many cameras. It allows user to switch their viewpoint and provides a real world experience to users.

The proposed user dependent scheme UDMVT is bandwidth efficient. It can reduce the bit-rate for the transmission of multi-view video. The requirement of bandwidth is smaller than other schemes. To make use of the UDMVT to the FVT, the following several problems still should be concerned:

- **User Interface**: The user interface should agree to the successive motion model.
- **Storage**: The captured raw data of FTV is huge. The prediction structure of UDMVT is not suitable for the storage.
- **Encoder**: Special encoder should be used to encode frames of different ranges when multiple users are watching the same program at same time.
- **Networks**: UDMVT reduce the traffic and is suitable for the limited bandwidth network such as the wireless network. But it is not suitable for the P2P networks and multicast.

3.2. User Interfaces

Paper [4] has developed two types of user interface for the Free Viewpoint TV. One of these two types of user interface only shows one view according the viewpoints given by the user. By this type of user interface, viewpoint of user can be switched by an eye/head-tracking system, moving the mouse of a PC or sliding the finger on touch panel of mobile device.

A real-time interactive multi-view video system was developed in [10]. In this system, user can switch the viewpoint by dragging the scroll bar to different position.

In the first type of user interfaces of [4] and the user interface of the system in [10], the changing of the position of user, moving of mouse, sliding of finger and dragging of scroll bar are all successive motion. The switching models of these user interfaces are all successive motion model.

By these user interfaces, it will take some time to switch from current view to the neighboring view. Such as in the head-tracking system, the user needs to take some time to move from current position to next position for the new viewpoint. For the FTV on a PC with mouse interface, mouse need to move enough distance to reach the next viewpoint, which will cost some time. If it is the FTV on a mobile player, finger need to slide some distance after press down to switch to next viewpoint. And the scroll bar also needs some time to be dragged to next position. The switching models of these user interfaces are all successive motion model.

The switching model of these user interfaces is used to the transmission of multi-view video in the FVT when these user interfaces are applied. The value of k depends on the switching speed of user by these user interfaces.

3.3 Storage

As the multi-view video contains so many video sequences in it, if the multi-view video is stored in raw data, the storage space requirement is large especially the number of views is large. UDMVT is a scheme for live encoding and transmitting. Which viewpoint the user will switch to is unknown beforehand. And maybe there are more than one user watching the same multi-view program in FVT at same time. All users have their own viewpoints. Different ranges of frames should be encoded for different users. It cannot just compress some ranges beforehand.
In order to reduce the store space for the multi-view video, other compression technologies of multi-view video such as Simulcast or MVC can be used to encode all the views. If the UDMVT is used for the FVT, the frames encoded by Simulcast or MVC need to be decoded and then encoded the necessary frames again before transmission. This process will decrease the efficiency of the encoder. So some efficient methods should be proposed to exchange MVC prediction structure to the prediction structure of UDMVT directly to improve the efficiency of the storage and encoding.

3.4. Encoder

Maybe there is more than one user watching the same multi-view program in FVT at same time. All users have their own viewpoints. So the encoder has to encode different ranges of frames for different users at same time. There are two methods can be used to overcome this problem.

The first method is that only an encoder is used to encode all the frames at the same time. This method requests the encoder is powerful enough to process the complex algorithm of encoding multiple ranges of frames at same time.

The second one is using the encoder pool. Encoders are allocated to different users to encode the different ranges of frames at the same time.

3.5. Networks

The UDMVT reduces the bit-rate of transmission of multi-view video and the traffic would not increase as the increase of the number of views. The requirement of the bandwidth is small. It can works well even the condition of the networks is not well. The UDMVT can also works well for the transmission of multi-view video on the networks which is limited at bandwidth, such as wireless networks.

However, UDMVT is not suitable for some networks for, such as the peer-to-peer (P2P) network proposed in [11] to stream multi-view video. In [12], this idea is used to a multicast scenario where each view is streamed to a different IP-multicast address. In FVT, the users are watching the same program of multi-view video with different viewpoint. Different ranges of frames should be transmitted to different users at same time in UDMVT. So UDMVT doesn’t support the transmitting the multi-view video over the P2P network or by multicast.

However, some methods also can be proposed to solve this problem. Such as the frames in the comment part of several ranges can be encoded once. Encoded frames can be shared or streamed by multicast to all users. Encoded frames of the overlap of different ranges can be shared by several users.

4. CONCLUSION

We have proposed a user dependent scheme called UDMVT for the transmission of multi-view video in [1]. And we further improved this scheme to support the 3D multi-view video in the [2]. The evaluation results proved that the UDMVT is efficient to reduce the bit-rate of the transmission of the multi-view video in the successive motion model especially when the number of view is larger. In this paper, we introduce the UDMVT and further discuss the application of the UDMVT in the Free Viewpoint TV.

5. REFERENCES