특집논문-07-12-4-05

Integrating Multi-view Stereoscopic Transmission System into MPEG-21 DIA (Digital Item Adaptation)

Seungwon Lee[a], Ilkwon Park[a]‡, Manbae Kim[b], and Hyeran Byun[a]

ABSTRACT

In general multi-view system, all the view sequences acquired at the server are transmitted to the client. However, this kind of system requires high processing power of the server as well as the client, thus it is posing a difficulty in practical applications. To overcome this problem, a relatively simple method is to transmit only two view-sequences requested by the client in order to deliver a stereoscopic video. In this system, effective communication between the server and the client is one of important aspects.

Therefore, we propose an efficient multi-view system that transmits two view-sequences according to user's request. The view selection process is integrated into MPEG-21 DIA (Digital Item Adaptation) so that our system is compatible to MPEG-21 multimedia framework. Furthermore, multi-view descriptors related to multi-view camera and systems are newly introduced. The syntax of the descriptions and their elements is represented in XML (eXtensible Markup Language) schema. Intermediate view reconstruction (IVR) is used to reduce such discomfort with excessive disparity. Furthermore, IVR is useful for smooth transition between two stereoscopic view sequences.

Finally, through the implementation of testbed, we can show the valuables and possibilities of our system.

Keyword : MPEG-21 DIA, Multi-view descriptors, Intermediate view reconstruction, Multi-view transmission system

Ⅰ. Introduction

Due to remarkable progress of digital technology, numerous digital contents were created and various multimedia services became available[1]. These progresses arouse viewer's desire for more reality. Stereoscopic can effectively bring the realistic feeling to viewers. However, viewers receive only video data at one viewpoint from server passively. To provide only one viewpoint is insufficient and restricted for viewers' demands. So we easily conjecture that multi-view system will be one of the next generation's multimedia services because it is able to offer more realistic scenes by "continuous look-around" from different viewing angles.

Multi-view system can provide a free viewpoint (FV) which is able to look-around because it uses multi-view video sequences captured by multiple cameras at the same time but at the different positions. If server sends the whole video data and client receives all of them, both server and client should requires high processing power. In addition, wide bandwidth is required between server and client to transfer the whole video sequences. However, the processing power of server and client is limited. network capability is also limited. To overcome these limitations, we

a) Dept. of Computer Science, Yonsei Univ., Seoul 120-749, Republic of Korea
b) Dept. of Computer, Information, and Telecommunication, Kangwon National University Chunchon 200-701, Republic of Korea
‡ 교신저자 : 박일권(ikheart@cs.yonsei.ac.kr)
※ "This research was supported by the Ministry of Information and Communications (MIC), Korea, under the Information Technology Research Center (ITRC) support program supervised by the IITA "(IITA-2006-C1090-0603-0017) "This paper was published by IWAIT 2007."

need to reduce the amount of data for transmit. In this paper, we propose Digital Item Adaptation method. To reduce the amount of data, we select and transmit some view-sequences according to user preferences instead of send the whole multi-view data. MPEG-21 Digital Item Adaptation (DIA) is used to sending user preferences to server.

The vision for MPEG-21 is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities [2]. And Digital Item Adaptation (DIA) is one of main MPEG-21 parts. The goal of the DIA is to achieve interoperable transparent access to multimedia contents by shielding users from network and terminal installation, management and implementation issues. This will enable the provision of network and terminal resources on demand to form user communities where multimedia content can be created and shared [3]. There are various types of digital item and its adaptation. The description which contains user preferences is written by XML. The descriptions are delivered using TCP/IP while multi-view stereoscopic data are delivered using RTP (Real-time Transport Protocol)/RTSP (Real-time Transport Streaming Protocol) for real-time streaming [4,5].

This paper is structured as follows. Section 2, we start with description of overall architecture of our proposed server-client system. Then in section 3, we will introduce the DIA description which represents user preference, view selection, and transmission of multi-view stereoscopic data. Section 4 describes the Intermediate View Reconstruction (IVR) and its use. The last section has the conclusion.

II. Overview of Multi-view Stereoscopic Transmission System

Digital Item (DI) is the basic unit of transaction in the MPEG-21 framework. They are structured digital objects, including a standard representation. More concretely, a Digital Item is a combination of resources such as videos, audio tracks or images and structure for describing the relationships between the resources. MPEG-21 Digital Item Adaptation (DIA) specifies the syntax and semantics of tools that may be used to assist the adaptation of Digital Items, for example, the Digital Item Declaration, metadata and resources referenced by the declaration[2].

Our proposed system contains a DIA server and a client. The DIA server consists of Resource Adaptation Engine and Description Adaptation Engine. Resource Adaptation Engine is divided into two parts; the one is view selection part which select a specific viewpoint among multiple views and the other is Intermediate View Reconstruction (IVR) part which generates virtual camera shot. There are three main modules of the client; the digital item player to play stereoscopic video, the image analysis, and the user preference generator which generates a description written by XML format.

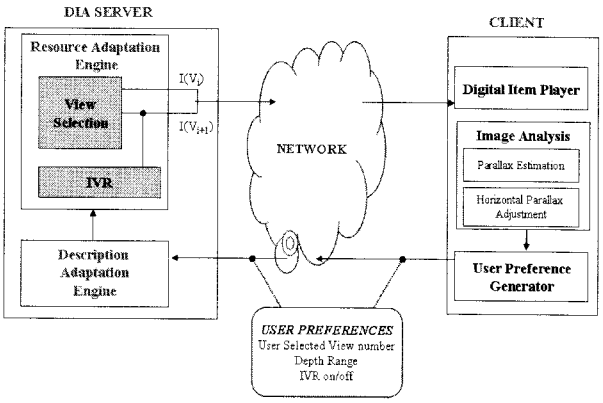


Fig. 1. The whole architecture for multi-view server-client transmission system

In Fig. 1, the description which contains user preferences is parsed and then modified to the adapted description in the Description Adaptation block. The resource is also converted to the adapted resource which is selected and dis-

parity is adjusted for stable stereoscopic video in the Resource Adaptation Engine.

The adapted digital items are transmitted through network. TCP/IP is used for exchanging description XML data between DIA server and client. On the other hand, adapted resource data is delivered in RTP and RTSP.

Ⅲ. View Selection and Transmission of Multi-view

1. Description of User Preference

In this section, we introduce how client write out user preferences about multi-view stereoscopic adaptation. There are some kinds of user preferences; display type, view point, whether use IVR or not and number of IVR etc. To apply these user requirements to server, client use XML formatted description. XML is standard description format of MPEG-21 DIA. MPEG-21 DIA has developed a description tool called MultiViewSelection. The following definition represents the information for Multi-view selection.

- DisplayType decides stereoscopic or not
- ViewNumber is the camera number which is chosen by user.
- MaximumParallax indicates the parallax of user allowable.
- ViewChangeIVR represents user preference about intermediate view during view change.
- NumberOfIntermediateView is the number of reconstructed intermediate view during view change.
- IntervalDistance shows the position of virtual camera.

Fig. 2 shows an instance of the MPEG-21 DIA MultiViewSelection that describes select 5th view. And it also describes some other user preferences; the user wants to watch stereoscopic video not mono one, and also watch

```
- <DIA xmlns:xsi="urn:mpeg:mpeg21:01-DIA-NS">
  - <Description xsi:type="UsageEnvironmentType">
    - <UsageEnvironment xsi:type="UserCharacteristicsType">
      - <UserCharacteristics xsi:type="PresentationPreferencesType">
        - <Display>
          - <MultiViewVideoSelection>
            - <MultiViewVideoSelectionType>
                <DisplayType>Stereoscopic</DisplayType>
                <ViewNumber>5</ViewNumber>
                <MaximumParallax>20</MaximumParallax>
              </MultiViewVideoSelectionType>
            - <IntermediateView>
                <ViewChangeIVR>ON</ViewChangeIVR>
                <NumberOfIntermediateView>3</NumberOfIntermediateView>
              </IntermediateView>
            - <ParrallaxAdjustment>
                <IntervalDistance>3</IntervalDistance>
              </ParrallaxAdjustment>
            </MultiViewVideoSelection>
          </Display>
        </UserCharacteristics>
      </UsageEnvironment>
    </Description>
</DIA>
```

Fig. 2.Example of MPEG-21 DIA multi-view stereoscopic description

the generated intermediate views.

Three intermediate views are generated during view change. Additionally the location of virtual camera is 1/3 point between left real camera and right one.

```
- <element name="CameraParameterType" minOccurs="0">
  - <complexType>
    - <sequence>
      - <element name="CameraSetUpInfo">
        - <simpleType>
          - <restriction base="string">
              <enumeration value="Arc" />
              <enumeration value="Parallel" />
            </restriction>
          </simpleType>
        </element>
      - <element name="QuantityOfCamera">
        - <simpleType>
          - <restriction base="integer">
              <minInclusive value="0" />
            </restriction>
          </simpleType>
        </element>
      - <element name="DistanceBetweenCameras">
        - <simpleType>
          - <restriction base="float">
              <minInclusive value="0.0" />
            </restriction>
          </simpleType>
        </element>
      </sequence>
    </complexType>
  </element>
```

Fig. 3.Example of camera setup information description schema

Fig. 3 illustrates a description schema for camera setup

information. A DIA server sends the description of camera setup information to a client for the purpose of announcement; the type of camera arrangement, the total number of the cameras and the distance between the cameras. When client connect server, these information are sent only one time. A client can find out the number of the selectable views and decide whether parallax adjustment is required or not by parsing the description of the camera setup information. However, if a server doesn't have the camera setup parameters then the description of the camera setup information will not be transmitted.

2. The Transmission of Multi-view Image and Description

The proposed system treats the transmission between single broadcasting server and many clients. In multi-view video that obtained by multiple cameras, only two-view images instead of multi-view images are transmitted from server to client for stereoscopic generation. These video data are transmitted by RTP (Real-time Transport Protocol) and RTSP (Real-time Transport Streaming Protocol) that are suitable for streaming broadcasting.

On the other hand, a description which includes user preferences is independently transmitted. That is, the transmission of description data and video data are separated. A description represents various parameters and user preferences by XML format. This description is transmitted by using TCP/IP.

3. View Selection

In multi-view transmission system, one of the main purposes is the selection of the view point by user preferences. For a view selection in this system, the descriptor that includes user preferences from the client is sent to DIA description adaptation engine in the server, and then DIA description adaptation engine parses this description to analyze the user's request. Finally, the server stops sending current view video and starts sending the view which just selected by the information from the parsed description.

On the other hand, when view selection happens, intermediate view image are generated and interpolated from current view point to designated view point for smoothed view change.

Fig. 4. Before view change(cam #6 - 26th frame)

Fig. 5. After view change(cam #3 - 27th frame)

Fig. 4 and Fig. 5 show the view change from 3rd camera to 6th camera at 26th frame. The difference which is

caused between the two viewpoints is remarkable.

Ⅳ. Using Intermediate View

1. Parallax Adjustment

Depth is the relative distance of objects from the observer within a scene. Perception of depth is achieved by coordination of our eyes and brain [6].

In this system, viewers on clients watch stereoscopic image by two-view images which consist of left image and right image. Our system uses multi-view sequences captured by multiple cameras, so we can easily obtain left image and right image. Also we can easily obtain their disparity and can make stereoscopic image. However, it is not sufficiently complete because of the disparity which originated from multi-view cameras is sometimes too large or too small. If the stereoscopic video has small disparity then the 3D realistic sense will be decrease. Or if there is excessive disparity, viewers will be felt with eye strain. By generate intermediate view, we can solve these problems. This stereoscopic is generated by a experimental decision.

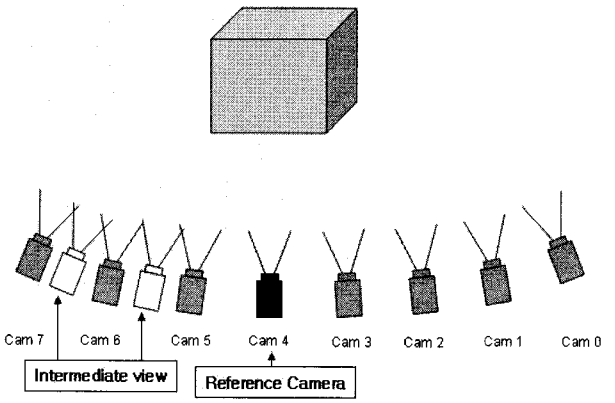


Fig. 6. Example of camera installation and virtual cameras for intermediate view

We can provide a high quality stereoscopic with less the tiring of viewer's eyes. That is, a disparity can be real-time adjusted by intermediate view image. As adjusting the gap of Multi-view cameras, we can adjust how the stereoscopic scenes are shown. However, this is difficult to apply a real time broadcasting system and it can not support the requests of many clients. Therefore, virtual view is generated by the moving of virtual camera.

In Fig. 6 we can find how the 8 real cameras were installed and at where the 2 virtual cameras were placed.

In this system, the generation of intermediate view uses a proposed method in C. L. Zitnick et al. [7]. At least a depth image for given images is required to generate intermediate view and acquired depth image is used to generate intermediate image at each view point. To generate depth image, first of all, basic depth image is generated by a correlation between images acquired from left and right cameras for a camera in the middle of given multi-view camera[8] A depth acquisition is not real-time processing but off-line processing. However, intermediate view image is generated in real-time by acquired depth image. The description for parallax adjustment in client is sent to DIA engine in server and a server sends modified view image after view selection in DIA engine according to user preference.

Fig. 7 shows real view, intermediate view and its stereoscopic conversion. (a), (b),and (c) are real views (first column). (a2), (b2) and (c2) are intermediate views third column). Finally, (a1), (b1) and (c1) are stereoscopic views which are interlaced using real views and intermediate views. Each row has sequential relations. That is, (a) is captured at 1st frame by camera number 0, (b) is captured at 2nd frame by cam #1 and (c) is captured at 3rd frame by cam #2. The second column's images have less camera baseline than the stereoscopic images which made from only real cameras. So their three dimensional effects are efficiently reduced.

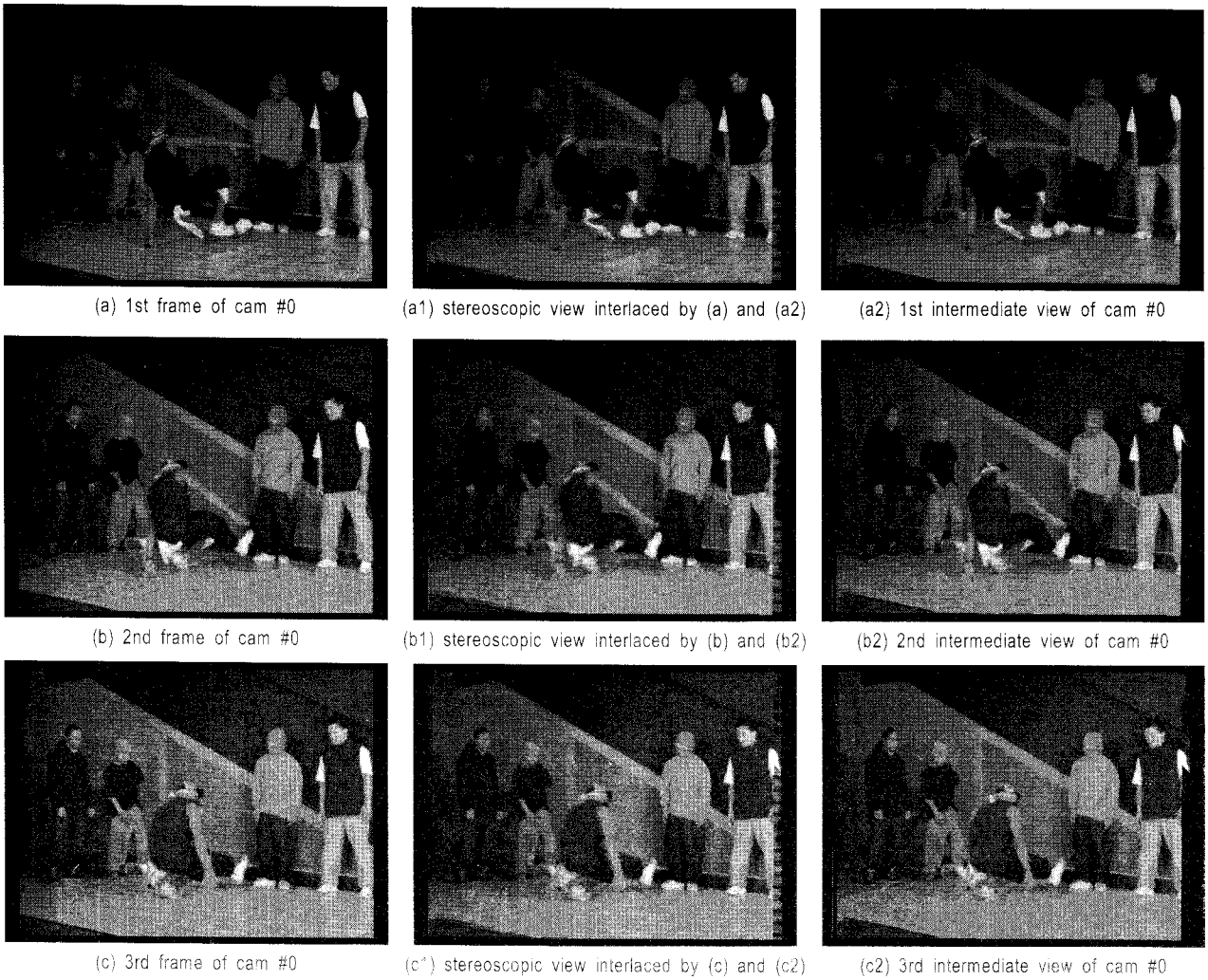


Fig 7. left image, intermediate image and stereoscopic image

2. View Change Smoothing

One of the desirable advantages of using multi-view camera is the "continuous look-around." It can be applied to various applications.

In these applications, natural view with conferees at the other end is essential; as moving his head, a user should be able to "see around" the conferees.

Virtual Travel can be another example. We can see a distant world in real-time by watching television. It may show the most important and popular scene. However, TV provides us only a single view of a real 3D world. If we don't want to see the scene with such position and wish to see different position, there is no way to changing viewpoint. Because of the view is determined not by users but by a camera. That is very different from realistic experience.

Such a problem could be overcome by using multiple cameras. We can see an object or scene at different viewpoint by jumping between shots of cameras. If we compose cameras like a circle, we can see front and side as well

as back of an object. And if we set cameras a hemisphere, we can see a scene like we are flying. In this way, multi-camera can give us the feeling which we are there.

However if the base line between cameras is not sufficiently small, the viewer must feel uncomfortable when view change occurs, because of the scene will not be changed continuously. The more cameras, the more the viewer's realistic feeling will be got. But the more cameras, the more expansive and bulky system will be. In practice, a 'virtual camera' is considered and the reconstructed outputs (intermediate views) based on the images from the true cameras. This is shown in Fig. 6. Intermediate views are reconstructed based on images acquired by cam 5 and cam 6. Similarly, images from cam 6 and 7 generate virtual images. These generated virtual images will improve view changes smoother by fill up the blank between the true cameras.

V. Conclusion

In this paper, we presented the multi-view stereoscopic adaptation and transmission system based on MPEG-21 DIA. Our proposed system is focused on the reflection of user preferences and the transmission of adapted digital items in the network environments. The goal of MPEG-21 DIA is to achieve interoperable transparent access to multimedia contents by shielding users from network and terminal installation, management and implementation issues. This will enable the provision of network and terminal resources on demand to form user communities where multimedia content can be created and shared. On comparing the purpose of MPEG-21 DIA framework, we can see our work clearly consistent with it.

For the implementation, we developed a testbed program which includes all of them. For the real-time streaming broadcasting, RTP/RTSP and TCP/IP are used. And RTP/RTSP also brings transmissive efficiency because there are various types of digital items and RTP/RTSP is suitable protocols for treat the DIs.

Our work is meaningful as a suggestion of the description that depicts stereoscopic view selection from multi-view and its adaptation. However, in the intermediate view generation process, depth estimation takes long time. So it is not suitable for live broadcasting. Improvements of the depth estimation speed and make more accurate intermediate view will be our future work direction.

REFERENCES

[1] Meesters, Lydia M.J., IJsselsteijn, Wijnan A., Seuntiens, Pieter J.H., "A survey of perceptual evaluations and requirements of three-dimensional TV", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 3, pp. 381-391, 2004.

[2] "MPEG-21 Overview V.4", ISO/IEC JTC1/SC29/ WG11 N4801, May 2002

[3] HS Sohn, HS Kim and MB Kim, "MPEG-21 DIA Testbed for Stereoscopic Adaptation of Digital Items," LNCS (vol. 3333) PCM2004, 449-456, 2004.

[4] IETF RFC 1889, RTP: A Transport Protocol for Real-Time Applications, January 1996

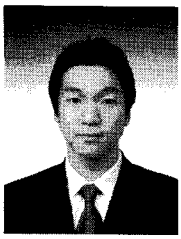
[5] IETF RFC 2326, RTSP: Real Time Streaming Protocol (RTSP), January 1996

[6] Serdar Ince, "Correspondence Estimation and Intermediate View Reconstruction", Boston University, Department of Electrical and Computer Engineering Technical Report No. ECE-2004-01

[7] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. A. J. Winder, and R. Szeliski, "High-quality Video View Interpolation Using a Layered Representation,"ACM SIGGRAPH and ACM Trans. on Graphics Aug. 2004.

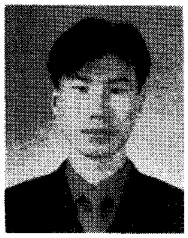
[8] D. Scharstein, and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47, 1, 7-42. 2002

저 자 소 개



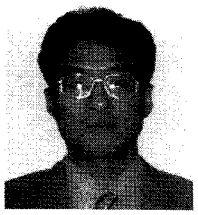
Seungwon Lee

- 2005년 : 홍익대학교 컴퓨터공학과 졸업(공학사)
- 2007년 8월 : 연세대학교 컴퓨터과학과 졸업예정(공학석사)
- 주관심분야 : 패턴인식, 영상처리, 다시점영상처리, MPEG-21 DIA



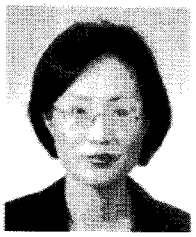
Ilkwon Park

- 2002년 : 군산대학교 컴퓨터과학과 졸업(이학사)
- 2005년 : 연세대학교 컴퓨터과학과 졸업(공학석사)
- 2005년~현재 : 연세대학교 컴퓨터과학과 박사과정 재학 중
- 주관심분야 : 컴퓨터 비전, 패턴인식, 3D 비디오처리, 다시점영상처리



Manbae Kim

- 1983년 : 한양대학교 전자공학과 학사
- 1986년 : University of Washington 전기공학과 공학석사
- 1992년 : University of Washington 전기공학과 공학박사
- 1992년~1998년 : 삼성종합기술원 수석연구원
- 1993년 : Georgetown University 의과대학 객원연구원
- 1996년 : University of Rochester 전기공학과 객원연구원
- 1998년~현재 : 강원대학교 컴퓨터정보통신공학과 교수
- 주관심분야 : 3D 비디오처리, 다시점영상처리, 3DTV 시스템



Hyeran Byun

- 1980년 : 연세대학교 수학과 졸업(이학사)
- 1983년 : 연세대학교 대학원 수학과 졸업(이학석사)
- 1987년 : University of Illinois, Computer Science(M.S.)
- 1993년 : Purdue University, Computer Science(Ph.D.)
- 1994년~1995년 : 한림대학교 정보공학과 조교수
- 1995년~1998년 : 연세대학교 컴퓨터과학과 조교수
- 1998년~2003년 : 연세대학교 컴퓨터과학과 부교수
- 2003년~현재 : 연세대학교 컴퓨터과학과 교수
- 주관심분야 : 영상인식, 영상처리, 패턴인식