STEREOSCOPIC VIDEO ADAPTATION IN MPEG-21 DIA

Manbae Kim, Haksoo Kim and Donghee Hong

Kangwon National University
Department of Computer, Information, and Telecommunication
192-1 Hoja2-dong, Chunchon 200-701, Republic of Korea
manbae@kangwon.ac.kr

ABSTRACT

This paper considers the method of adapting 2D video to stereoscopic video according to the descriptors of MPEG-21 Digital Item Adaptation (DIA) that provide a functionality of user preferences. For the adaptation, we consider the stereoscopic adaptation of 2D video (stereoscopic video adaptation) so that users can enjoy 3D stereoscopic video being converted from 2D video. The adaptation method deals with a variety of image motion types such as horizontal motion, non-horizontal motion, static, zoom and fast motion. For each type, we investigate how descriptors such as *parallax type*, *depth range*, and *maximum interval of a delayed frame* are applied to the stereoscopic video adaptation. The experiments have been performed to verify our proposed adaptation method.

1. INTRODUCTION

The Digital Item Adaptation (DIA) is one of main MPEG-21 parts [1, 2]. 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. DIA is composed of resource adaptation and descriptor adaptation that produce a newly adapted (modified) Digital Item (DI). The descriptor adaptation transforms the input descriptor to the output one. As well, the input resource is adapted into the output resource according to the descriptor. Those outputs are then delivered to a user that has requested the adapted Digital Item.

There could be a variety of Digital Items that need to be managed in the DIA framework. Among them, we are concerned primarily with the stereoscopic adaptation of 2D video. Unlike the stereoscopic video acquired from a stereoscopic camera with two sensors, the stereoscopic video adaptation (SVA) makes it possible for users to feel 3D depth perception from ordinary 2D video data by adapting or converting a single 2D image into a stereoscopic image. The adaptation algorithms are not normative part of the MPEG-21 DIA standardization. The main purpose of this paper is to provide a framework of implementing stereoscopic video adaptation according to the descriptors.

Fig. 1 describes a general structure of SVA in the server/client/network environment. User requests a DI to DIA server with a descriptor that is generally represented in XML. Upon receiving it, DIA server adapts the current DI by modifying the descriptor as well as the resource. In the descriptor adaptation, *ParallaxType*, *DepthRange*, and *MaxDelayedFrame* control the SVA. The resource (e.g., 2D video) is adapted to a new resource (e.g., stereoscopic video) according to the descriptors. The two outputs form a new digital item (e.g., stereoscopic DI), which is delivered to the user through the network. Our study will be limited to the functionalities of DIA server, not covering the network part.

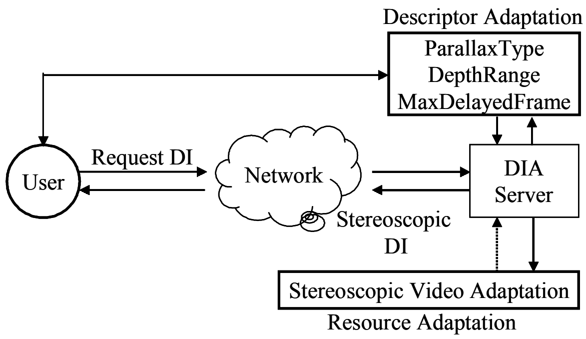

Fig. 1 The structure of SVA

This paper is organized as follows. Section 2 introduces DIA descriptors related to stereoscopic video adaptation. We introduce the basic methodology for the SVA in Section 3. Experimental results will be presented in Section 4. The feasibility of our proposed method will be examined based upon the experimental results. Section 5 summarizes this paper and discusses our future plan.

2. ADAPTATION DESCRIPTOR

This section introduces a description scheme (DS) for stereoscopic video adaptation, which is currently defined by DIA CD [2]. Table 1 shows the semantics of the stereoscopic video adaptation DS. Note that *Stereoscopic*

Video Adaptation supports the video adaptation from 2D video to stereoscopic video. A user can describe her/his own display presentation preferences by specifying the following descriptors: *ParallaxType*, *DepthRange*,and *MaxDelayedFrame* [3].

Table 1: The semantics of *Stereoscopic Video Adaptation* DS

Name	*Definition*
ParallaxType	Identifies the type of parallax being composed of negative and positive parallaxes.
DepthRange	Describes the range of 3D depth perceived by the user
MaxDelayedFrame	Indicates the maximum interval of a previous frame.

ParallaxType represents the type of the parallax being composed of positive parallax and negative parallax [9]. This description can be used by the resource adaptation of stereoscopic video in order to deliver the perception of 3D depth. In the negative parallax, the 3D depth is perceived between the monitor screen and human eyes. On the contrary, the 3D depth is perceived behind the monitor screen in the positive parallax. *DepthRange* indicates the range of 3D depth perceived by the user and is defined as the distance between the monitor screen and the object in 3D. It applies identically to the positive and negative parallaxes. The amount of DepthRange is varied at [Min, Max], which can be normalized to [0, 1]. For positive and negative parallaxes, shifting a right image to the right direction increases the range of depth. On the contrary, shifting it to the left direction decreases the range of depth.

One of the SVA schemes is to make use of a previous (delayed) image. Suppose the image sequence is $\{\ldots, I_{K-3}, I_{K-2}, I_{K-1}, I_{K}, \ldots\}$ and I_K is the current frame. One of the previous frames, I_{K-i} ($i \geq 1$) is chosen. Then, a stereoscopic image consists of I_K and I_{K-i}. If the current and previous images are appropriately presented to both human eyes according to camera and object motions as shown in Table 2, the user then feels the stereoscopic perception. *MaxDelayedFrame* determines the amount of i value. Thus, the larger it is, the more depth the user feels.

Table 2: Motion types and left/right image selection

Camera motion	Object Motion	Left Image	Right Image
Right	None	Previous	Current
Left	None	Current	Previous
None	Right	Current	Previous
None	Left	Previous	Current

3. STEROSCOPIC VIDEO ADAPTATION

This section describes the methodology underlying the stereoscopic video adaptation (SVA). In SVA, a stereoscopic image for each image frame needs to be generated based on the properties of the descriptors. Usually, the motion and image information is used, from which a stereoscopic image is generated. To effectively deal with various motions, many methods on stereoscopic adaptation/conversion of 2D video have been proposed. However, there are currently no unique solutions to deal with all kinds of motions. For instance, the approach of processing static images will be different from that of moving images. For the latter, motion information is a useful tool that can realize the motion parallax of the stereoscopic image. On the contrary, because there is no such information in the static images, other cues such as color and texture might be utilized. Furthermore, other complex types of image (e.g., zoom, fast motion) need special conversion algorithms. The main purpose of this section is to present one of possible adaptation methods for each image type and more significantly to show how parameters associated with the methods are related to the descriptors.

An image type analysis needs to be carried out for each image. Note that the image analysis is beyond the scope of this paper. For more detail, refer to [4, 8]. Suppose that the image type is classified into a *horizontal motion image* in which a camera and object(s) move in the horizontal direction, a *non-horizontal motion image* in which a camera and object(s) move in the non-horizontal direction, a *fast motion image* containing fast camera and object movement, a *static image* having no movement, and a *zoom image*. We will describe how each image type is managed according to the descriptors.

A) *Horizontal motion image*: For a horizontal motion image, a stereoscopic image is generally composed of a current image and a previous image, where each eye sees one of the two images [4, 5, 6, 7, 8] as shown in Table 2.

There are two ways of selecting an appropriate eye to receive a previous image as described. The first one is to determine the current image as the left image and the previous image as the right image. The second one is for determining the current image as the right image and the previous image as the left image. The factor that chooses a previous frame plays an important role in the stereoscopic perception. MaxDelayedFrame determines one of the previous images according to its value. The greater it is, the better depth perception is. ParallaxType is controlled by switching the left and right images. For varying values of DepthRange, a left image is horizontally shifted to the left direction.

B) *Non-horizontal motion image*: The stereoscopic image may be generated by two methods. First, either a left or a right image may be vertically shifted for removing the vertical parallax existing in two images. This requires the calculation of the amount of a vertical motion. Then, the adaptation by descriptors is similar to that of the horizontal motion image. Second, after estimating the depth of each block (e.g., 8x8 block) from motion information, a new image can be synthesized by shifting each block by its associated parallax value that could be computed from the depth. Then, the adaptation by descriptors follows the adaptation of a static image. The block shifting causes empty pixel regions in the new image and thus region-filling methods can be used to remove such artifacts.

C) *Static Image*: Unlike moving images, the static image requires a different approach. The human visual system uses many psychological depth cues to disambiguate the relative positions of objects in a 3-D scene. The instances are linear perspective, shading and shadowing, aerial perspective, interposition, texture gradient, and color [9]. Human can make perception of 3-D position with the aid of such depth cues. Due to the difficulty in the realization of all the depth cues, two depth cues such as color and texture gradient might be used. Bright-colored objects will appear to be closer than dark-colored objects. We can perceive detail more easily in objects that are closer to us. As objects become more distant, the texture becomes blurred. Texture is finer in the foreground and grows coarse as the distance increases. The approach is to represent and integrate two depth cues for each image block. Suppose that we have computed for depth data for all the image blocks. Then, the depth, D is transformed into a horizontal parallax value, P.

$$P = P_{max} \quad (D \ / D_{max})$$

Where D_{max} is the maximum depth value and P_{max} is usually chosen as a value less than inter-ocular distance in order to satisfy the image fusion.

For stereoscopic image generation, each block is moved by P in the horizontal direction, to generate a right image. The current image is fixed to be a left image for effects of positive parallax. For negative parallax, P is simply subtracted from P_{max}. DepthRange is controlled by shifting the left or right image to the appropriate direction.

D) *Fast Image*: The fast motion belongs to horizontal motion image, where fast moving region(s) exists in the image. This makes the stereoscopic fusion difficult even in the horizontal direction. To reduce the horizontal motion parallax, the image shift is needed. The amount of the shift can be computed by averaging the *x* components of motion vectors larger than a maximum horizontal-parallax threshold. Due to the horizontal characteristics, the adaptation by descriptors is similar to horizontal motion image.

E) *Zoom Image*: One of approaches of generating a stereoscopic image is to compute the parallax of each image pixel and shift the pixel by its associated parallax, which is computed as follows.

Step 1: Compute the distance, r between each pixel (x,y) and the image center.

Step 2: Compute the parallax of each pixel as follows:

$$P = P_{max} \quad (r \ / r_{max})$$

Step 3: Shift each pixel in the horizontal direction for the right image by P.

P_{max} controls the range of depth. The adaptation of a zoom image is similar to one of a static image.

4. EXPERIMENTAL RESULTS

In this section, we present the experimental results that demonstrate the 2D images being stereoscopically adapted according to descriptors. The feasibility of descriptors that are applied to a resource adaptation from 2D image to stereoscopic image is verified. The format of all the output images is the interlaced one so that the disparity between left and right images is easily observed.

Fig. 2 shows the interlaced stereoscopic images of 2D MPEG *Fun* sequence, where a merry-go-round is horizontally moving to the right. DepthRange is controlled by shifting the right image. A current image and a previous image are viewed to the left and right eyes, respectively. Since the image has negative parallax, 3D depth is perceived between a monitor screen and eyes. The parallax types being composed of positive and negative parallax are demonstrated in Fig. 3, where two images are switched for parallax exchange. As shown in Fig. 4, it is observed that as the *MaxDelayedFrame* increases from one to four, the disparity between the left and right images accordingly increases, thus resulting in the greater 3D depth.

Fig. 5 demonstrates adapted static images whose parallax is negative. The original image, the stereoscopic image with P_{max} =12 (on the top), and two images whose left image is shifted by 5 and 10 pixels (at the bottom) are shown. Two bottom images shows the output with varying DepthRange. Similarly, for the positive parallax, Fig. 6 shows the original image and the stereoscopic image with P_{max} =12. Two images whose right image is shifted by 5 and 10 pixels are shown at the bottom.

5. CONCLUSION AND FUTUREWORK

In this paper, we presented the stereoscopic video adaptation method based upon MPEG-21 DIA. The feasibility of descriptors such as *ParallaxType*, *DepthRange*, and *MaxDelayedFrame* in adaptation methods has been verified by the experiments. In the experiments, horizontal, non-horizontal, and static images have been used to demonstrate that the stereoscopic adaptation of them is possible for the varying properties of the descriptors. Our proposed adaptation methods are one of many possible ones. Our paper does not consider presenting an optimum solution, but show how any stereoscopic conversion methods could be adapted according to the descriptors.

Currently, we are working on the DIA testbed being composed of client/DIA-server/network that implements our works under the network environment.

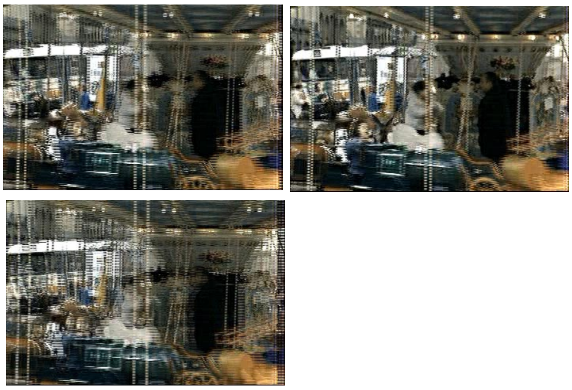

Fig. 2 Horizontal motion image. DepthRange is varied.

(a) (b)

Fig. 3 Horizontal motion image. (a) Positive parallax and (b) Negative Parallax.

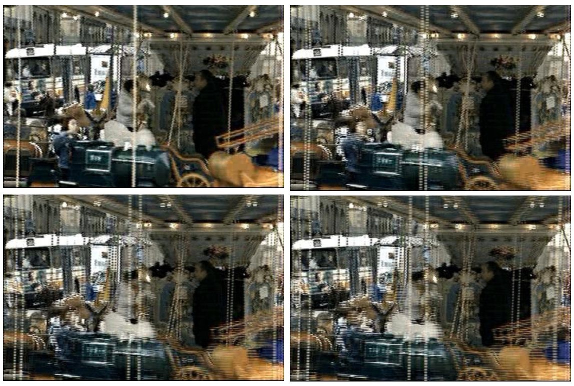

Fig. 4 Horizontal motion image. MaxDelayedFrame is varied in [1, 4].

ACKNOWLEDGMENT

This research was supported by University IT Research Center (ITRC) Project of Korean Ministry of Information and Telecommunication. We would like to thank the anonymous reviewers for their helpful suggestions.

Fig. 5 Static Image. DepthRange is varied in [0.0,1.0].

Fig. 6 Static Image. DepthRange is varied in [0.0,1.0].

References

[1] "Information Technology – Multimedia Framework- Part 1: Vision, Technology, and Strategy", ISO/IEC JTC1/SC29/WG11, March 2001.

[2] "MPEG-21 Digital Item Adaptation CD (v6.0)", ISO/IEC JTC1/SC29/WG11 N5178, Trondheim, May 2003.

[3] "MPEG-21 DIA Description Tools for 3D Stereoscopic Video Adaptation", ISO/IEC JTC1/SC29/WG11, M8613, Klagenfurt, Austria, July 2002.

[4] M. B. Kim and S. H. Lee, "A new method for the conversion of MPEG encoded Data into Stereoscopic Video", *Journal of the Society for 3D Broadcasting and Imaging*, January 2002.

[5] Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa, "Conversion system of monocular image sequence to stereo using motion parallax," *SPIE. Vol 3012, Photonic West*, pp. 108-115, 1997.

[6] T. Okino, M. Murata, K. Taima, T. Inimura, and K. Oketani, "New Television with 2D/3D image conversion technologies", *SPIE Photonic West*, vol 2653, pp. 96-103, 1995.

[7] B. J. Garcia, "Approaches to stereoscopic video based on spatial-temporal interpolation," *SPIE Photonic West*, vol. 2635, pp. 85-95, San Jose, 1990.

[8] M. B. Kim, "Video processing of MPEG compressed data for 3-D stereoscopic conversion," *Korean Broadcasting Engineering Conference*, pp. 3-8, Dec. 1998.

[9] D. F. McAllister, Stereo computer graphics and other true 3D Technologies, Princeton University Press, Princeton, NJ, 1993.