

3D Conversion of 2D Video Encoded by H.264

Ho-Ki Hong*, Min Soo Ko**, Young-Ho Seo***, Dong-Wook Kim§ and Jisang Yoo†

Abstract – In this paper, we propose an algorithm that creates three-dimensional (3D) stereoscopic video from two-dimensional (2D) video encoded by H.264 instead of using two cameras conventionally. Very accurate motion vectors are available in H.264 bit streams because of the availability of a variety of block sizes. 2D/3D conversion algorithm proposed in this paper can create left and right images by using extracted motion information. Image type of a given image is first determined from the extracted motion information and each image type gives a different conversion algorithm. The cut detection has also been performed in order to prevent overlapping of two totally different scenes for left and right images. We show an improved performance of the proposed algorithm through experimental results.

Keywords: 2D/3D conversion, Stereoscopic video, H.264, MTD, Delay factor, Motion vector

1. Introduction

One of the desired features for realizing high-quality information and telecommunication services in the future is the sensation of reality. This can be achieved by three dimensional visual communications. The 3D video system has many potential applications, e.g., in education, training, movies, medical surgery and videoconferencing.

Because of the current interest in 3D video, various types of apparatuses and devices for obtaining 3D video have been developed by many research institutes and companies. In particular, several 3DTV projects combine both signal processing and human factor. In the early 1990s in Europe, the RACE DISTIMA project was initiated by EU for 3D-TV research in order to develop a system for acquisition, coding, transmission and presenting stereoscopic image sequences [1-3]. These projects led to another project, PANORAMA; its goal was to enhance visual information in telecommunications with 3D tele-presence [4, 5]. One of the IST projects, VIRTUE, is developing a videoconferencing system aimed at providing immersive effects for tele-presence [6]. It is being designed to achieve a three-way video conference supporting life-size, upper-body video images in a shared virtual environment. In Korea, ETRI (the Electronics and Telecommunications Research Institute) conducted a system test of broadcasting stereo images for the 2002 FIFA Korea-Japan World-cup Football Event [7]. In addition, in Japan, various researches related to the 3-

dimensional field including HDTV project by NHK are in progress [8, 9].

Alongside development in the field of obtaining, processing and displaying such 3D video, the stereoscopic conversion of 2D video is a new field that is growing gradually. An essential difference from general stereoscopic images acquired from two cameras is that the stereoscopic conversion must generate stereoscopic images from a 2D image sequence. The results from the conversion can also be applied directly to movie production or TV broadcasting. However, 3D content by means of 2D/3D conversion have not been popular until now. The main reasons are that 3D images themselves may cause dizziness, and the 2D/3D conversion method is not sufficient enough to give satisfactory stereoscopic effect to viewers. Conversion scheme from 2D video to 3D video has been studied since 1990 and a variety of research has been carried out.

Y. Matsumoto et al. presented a depth-based method. Deriving a depth from a sequence of successive image, they then generated two perspective-projected images that are displayed to the left and right eyes [10]. T. Okino developed a dedicated digital video processor to perform 2D/3D conversion, employing a technique called “Modified Time Difference (MTD)” algorithm. The technology automatically converts ordinary 2D images into binocular parallax 3D images according to the detected motion of objects in images in real-time [11]. In Korea, researchers in Kangwon National University developed a 3D video conversion method using delayed images [12], and Soft4D Co. in Korea, developed a 3D conversion program for MPEG-2 video, called 3D Plus [13]. Digital Dynamic Depth Co. also developed a DVD player in software and hardware for converting DVD content to 3D content [14]. [15] is a 2D to 3D video conversion scheme based on H.264. It utilizes the motion information between consecutive frames to approximate the depth map of the

† Corresponding Author: Dept. of Electronic Engineering, Kwangwoon University, Korea. (jsyoo@kw.ac.kr)

* R&D Team, Health & Medical Equipment Business Team, SAMSUNG Electronics Co., Ltd., Korea. (hoki79.hong@samsung.com)

** Dept. of Electronic Engineering, Kwangwoon University, Korea. (kmsqwet@kw.ac.kr)

*** College of Liberal Arts, Kwangwoon University, Korea. (yhseo@kw.ac.kr)

§ Dept. of Electronic Materials Engineering, Kwangwoon University, Korea. (dwkim@kw.ac.kr)

Received: September 2, 2011; Accepted: September 11, 2012

scene. It uses only the horizontal motion to estimate the depth map. [16] also proposed a depth map generation algorithm for the application in 2D to 3D video conversion with H.264 bit-stream. After analyzing and extracting the motion information in the compressed domain, the block-level depth map is estimated from the filtered motion vectors. Then the final depth map is generated by up-sampling.

2D/3D conversion is a research area including various techniques. The conversion techniques can be classified into off-line conversion and on-line conversion and the one in this paper is an on-line conversion technique. That is, it converts a 2D video sequence to a corresponding stereoscopic 3D video sequence in the same real time as for the received 2D sequence to be decoded and displayed in a real time. Therefore this technique is more applicable to a real time process such as DMB (Digital Multimedia Broadcasting). There have been various converting techniques. Among them, this paper uses the motion parallax, in which the direction and type of motion are determined by using the H.264 bit-stream itself. Also, this method directly generates a stereoscopic image pair from the motion parallax without extracting corresponding depth information. The method proposed in this paper consists of creating left or right images after determining a motion direction and motion type by using motion vector from H.264-encoded data. H.264 has more accurate motion information than MPEG because of the availability of a variety of block sizes. In order to decide motion direction, camera movement and object movement are analyzed separately by using a difference image. Then, image types can be categorized as zooming, static, vertical and horizontal by using these motion types and directions. Most of the motion information, however, is obtained from block matching for each image block, resulting in low accuracy. Due to such inaccurate motion information, the different motion directions and image types obtained for each frame may cause shaking in the displayed images. In order to solve this problem, the reliability of motion information should be improved by analyzing motion directions and motion types in every frame of the scene.

This paper is divided into five sections. In section II, conventional 2D/3D stereoscopic conversion will be described. In section III, the new 2D/3D video stereoscopic conversion algorithm is explained. In section IV, the experimental results will be shown. In section V, conclusions and future research directions will be described.

2. 2D/3D Conversion

2.1 Basic principles of 2D/3D conversion

Stereoscopic images allow a viewer to feel the depth with three types of binocular parallax; positive parallax,

negative parallax, and zero parallax. The type of parallax is decided based on the location of a convergent point of an image relatively from the screen as shown in Fig. 1. When the LAP (left accommodation point) and the RAP (right accommodation point) are laid on the same point on screen, it produces zero parallax. When the LAP and RAP have different location out of screen, positive or negative parallax can be obtained.

Fig. 2 illustrates the principle of Ross phenomenon, which provides a theoretical background for 2D/3D image conversion [15, 16]. The left image serves as a current image and the right image serves as a delayed image. The airplane in the image moves from left to right and the background mountain is stationary. While both left and right eyes see the airplane and the mountain, the location of the airplane changes with respect to the location of the mountain in between left and right images. This phenomenon creates the parallax effect.

The convergent point in this case for the airplane is generated in front of the screen (negative parallax), and the convergent point for the mountain is generated right on the screen (zero parallax).

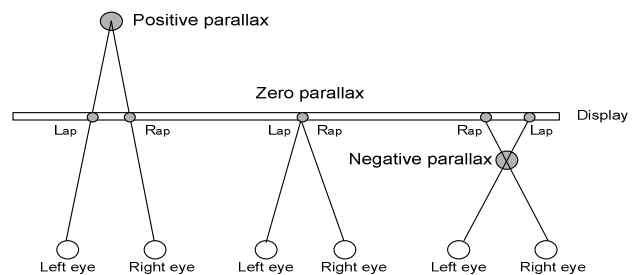


Fig. 1. Three types of binocular parallax

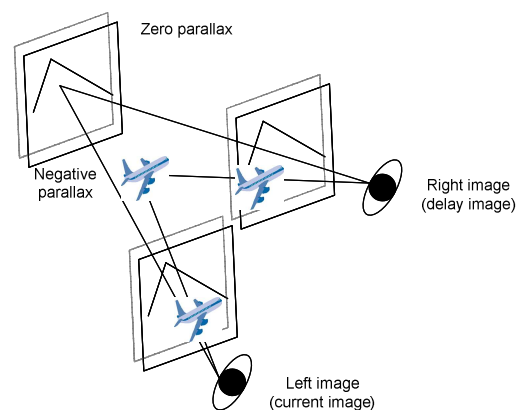


Fig. 2. Ross phenomenon

2.2 Structure of 2D/3D conversion

Fig. 3 shows a proposed conversion algorithm by using the standard decoding process of H.264. In this algorithm, special analysis of motion vectors while H.264-encoded

video is decoded is required to determine motion type of a current image. This additional information of motion type plays a critical role in the 3D image conversion process. Once motion type of each image is decided, 3D images are created by applying different 3D conversion algorithms depending on these image types.

As shown in Fig. 3, the motion vector analysis module first decodes H.264-encoded video and decoded motion vectors are analyzed to decide image type of each frame. In the next step, the 3D image synthesis and display module creates and displays 3D images, using an efficient 2D/3D conversion algorithm depending on the analyzed motion vector type. The algorithms utilized in this process will be presented in the next section.

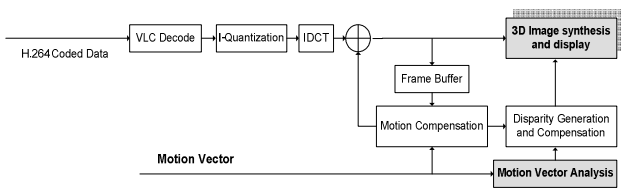


Fig. 3. H.264 decoder and structure of 2D/3D conversion

3. Proposed 2D/3D Conversion Algorithm

3.1 Overview

This section explains the method proposed in this paper. By comparing with the existing methods, the proposed has several distinguishing characteristics: It directly converts the 2D image sequence into 3D stereoscopic sequence during deciding the H.264 bit-stream without generating a depth map. Differently from other method using H.264 bit-stream as [15] and [16], it considers the stationary, vertical, and zooming movement as well as the horizontal movement to increase the reliability of the conversion. Accordingly a methodology to generate a stereoscopic image pair for each type of motion is provided. Also, the

information of intra blocks in the inter slices is used to increase the conversion reliability more. The 2D/3D conversion process proposed in this paper is illustrated in Fig. 4. Current image, delayed image, and motion vectors that are reconstructed by H.264 decoding stage in Fig. 3 serve as input to the conversion module.

The basic mechanism of this process is to utilize the current or the reconstructed current image for left and right images depending on image types and possible scene change situation. The cut detection module covers a problem with MTD method by detecting any possible rapid image change (cut) before the actual conversion process, and provides a better solution to image reconstruction. Different image reconstruction mechanisms are applied depending on image types while the conventional MTD method can be applied only to the horizontal image.

3.2 Cut detection

In a situation where a video cuts from one shot to the next, the use of current and delayed frames leads to overlapping of two totally different scenes and it would be impossible to obtain proper stereoscopic 3D images. The cut detection module has been developed in order to prevent this problem.

In a cut detection process, information of macro-blocks from P slice is used since P slice has more intra-blocks than inter-blocks when a rapid scene change occurs. Therefore, it is determined as a cut when the portion of intra-blocks is over a threshold value. When a cut is detected, conversion is done with the same method as for static images which will be described later.

3.3 Decision of image type

The image type defined in this paper is categorized as a static, zoom, vertical, or horizontal.

1) Static image

A static image is defined as an image without movement of objects or a camera. Since there is no motion, motion vector in the image has zero value. Also, the static image mainly consists of copy-blocks from the reference frame. Therefore, it is determined as a static image when the portion of copy blocks are greater than a threshold and both values of all horizontal (dx) and vertical motion vectors (dy) in a frame have values of zero

2) Zoom image

A zoom image is created by zoom-in and zoom-out functions of a camera, and shows scenes with gradually expanded or scaled-down images. Therefore, the zoom image has motion vectors with their directions either toward the center or toward the outside of the image. If a given image has this feature, it is categorized as a zoom

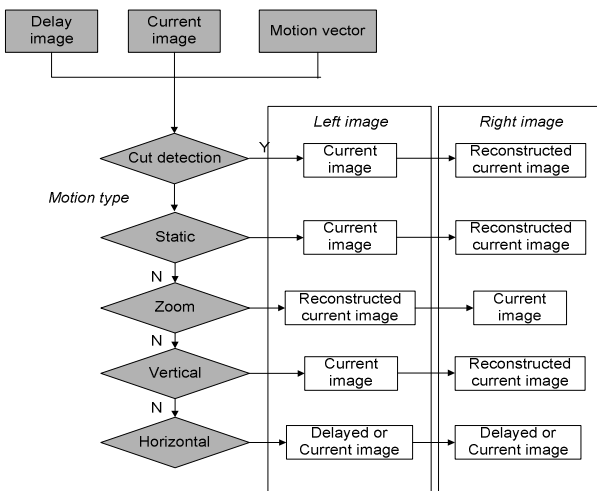


Fig. 4. The proposed 2D/3D conversion algorithm

image. To determine this, a given image is divided into four quadrants as shown in Fig. 5. If the direction of a motion vector defined in Eq. (1) in each quadrant satisfies Eq. (2), then the value of N_i is increased by 1. If the total number of N_i in an image is bigger than a threshold value, the image is defined as a zoom image.

$$\theta = \tan^{-1} \frac{dy}{dx} \quad (1)$$

$$\frac{\pi}{2}(i-1) \leq \theta < \frac{\pi}{2}i \quad i=1,2,3,4 \quad (2)$$

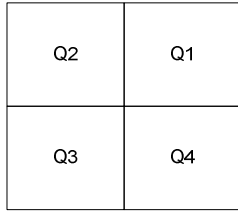


Fig. 5. Image divided into four quadrants

In order to decide between the cases of zoom-in and zoom-out, N_R and N_L which are the maximum of the N_i for the left and right side of quadrants are calculated as shown in Eq. (3).

$$N_L = \text{MAX}(N_2, N_3), \quad N_R = \text{MAX}(N_1, N_4) \quad (3)$$

Since motion vectors are decided by a block matching algorithm in H.264 that produces a low level of precision, the maximum number of vectors in between two quadrants is selected so that the conversion process uses more distinguishable characteristics of zoomed images.

Here, the positive (+) and the negative (-) directions of motion vectors are decided based on the assumption where the positive direction is toward the right side, and the negative direction is toward the left side. When the relation between N_L and N_R satisfies Eq. (4), it is zoom-in, and Eq. (5) shows the condition for the zoom-out

$$N_{L+} < N_{L-}, \quad N_{R+} < N_{R-} \quad (4)$$

$$N_{L+} > N_{L-}, \quad N_{R+} > N_{R-} \quad (5)$$

3) Vertical image

A vertical image is defined when motion vectors with vertical movement are dominant in an image. The angle of a motion vector in Eq. (1) is used to find how much vertical motion is present. If the value of the angle is larger than 45° , it is said that a vertical motion exists. When the portion of macro-blocks with vertical motion exceeds a threshold value, then the image is classified as a vertical image.

4) Horizontal image

While going through the motion type decision process, if an image is not determined to be one of the three previous

cases (static, zoom, or vertical), it is classified as a horizontal image.

3.4 Stereoscopic 3D conversion algorithm

The stereoscopic image conversion process utilizes an original 2-D image for generating left and right (delayed) images. In this section, we present an algorithm that generates left and right images based on the image type defined in the previous section.

1) Static Image

Since a static image does not have movement of any object, a current image and its delayed image are the same. In this case, a reconstructed current image is usually used in the place of the delayed image. In Fig. 6, the left image is a current image and the right image is the reconstructed current image by the assumption that the lower part (ground) of the image usually appears closer to a viewer than the upper part (sky). In order to increase depth of the image, the negative and the positive parallax are applied to the lower and the upper parts of the image, respectively.

Consequently, the lower part of image becomes protruded and the upper part of image looks deeper so that the whole image shows an improved 3-D effect when both images are combined. Here, the parallax (p) can be calculated as shown in Eq. (6):

$$p = \text{MAX}_p \times \frac{\text{verticallength}/2}{|\text{verticallength}/2 - y|} \quad (6)$$

where MAX_p is a user-defined maximum parallax value and y represents the value of the y-coordinate in an image plane. Once the conversion of a static image is performed by applying parallax values, the resulting image will show more depth along the direction of the arrows in Fig. 6.

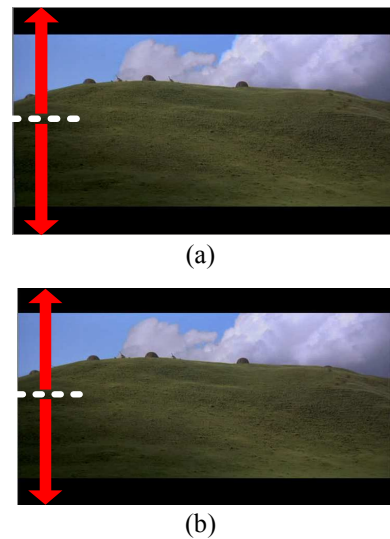


Fig. 6. Static image conversion: (a) left image (current image); (b) right image (reconstructed current image)

2) Zoom image

A zoom image has two different types of conversion processes based on camera function: either zoom-in or zoom-out. The zoom-in is a function that expands an image. Therefore, negative parallax can be applied as focus approaches the center of an image, which produces an image that stands out. In contrast, the zoom-out is a function that scales down objects in an image. Positive parallax is applied in this case so that the reconstructed image shows more depth.

The reconstructed image for zoom-in and zoom-out images are created by moving pixels in a current image on the horizontal axis by the value of the calculated in Eq. (7) and (8), respectively.

$$r = \sqrt{x^2 + y^2}, \quad p = MAX_p \times \frac{MAX(r)}{r} \quad (7)$$

$$r = \sqrt{x^2 + y^2}, \quad p = MAX_p \times \frac{r}{MAX(r)} \quad (8)$$

where r is a distance from the center of an image to the coordinate (x, y) of a given pixel, MAX_p is a user-defined maximum parallax value and $MAX(r)$ is a distance from the center to the corner of the image. The value from Eq. (7) is used for negative parallax, and Eq. (8) is used for positive parallax.

While a camera is fixed in most of zoomed images, a special case that may occur is when a camera is panning so that the central point of an image keeps changing. In order to find a new central point for zoom-in or zoom-out, a spiral path from the original central point is found in searching for a minimal motion vector. Since the motion vector around the object in a zoomed image is smaller than other area of an image, identifying the minimal value of motion vector is used to find the new central point in an image. The conversion process described beforehand is then applied at the new central point, completing a zoomed image conversion.

3) Vertical image

When current and delayed images are generated for a vertical image, it is impossible to have the stereoscopic effect because a convergent point cannot be created by vertical parallax. The conventional MTD method uses current and delayed images; in this case, the delayed image would be the one with vertical move. However, as is evident in Fig. 7, due to different motions of objects and the background in the image, canceling the vertical parallax for the background will generate another vertical parallax for the object and vice versa. This problem can be fixed by generating a reconstructed image from the current image. The key point for generating the reconstructed image is to use the fact that the vertical motion vector of an object is bigger than the one for the background.

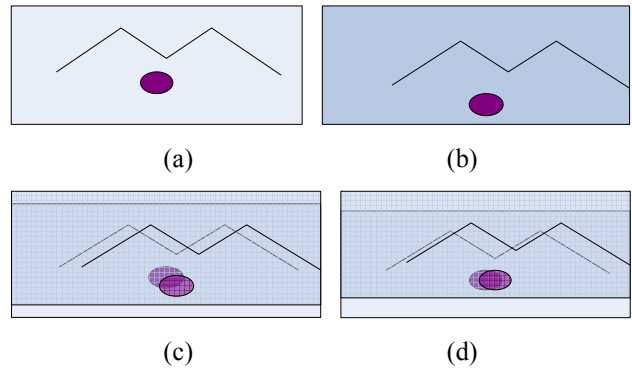


Fig. 7. Vertical image conversion using current and delayed images: (a) delay image; (b) current image; (c) vertical parallax of object; (d) vertical parallax of background

In the vertical image, motion vectors are categorized as global camera motion or local object motion. Details of how to decide the motion category for horizontal image conversion will be presented in the next section. Since an image affected by global camera motion has motion vectors in most of its macro-blocks, if the motion vector in the vertical direction for a macro-block is subtracted by the average value of the motion vectors in vertical direction of the image, then the result will include only proper object motion.

Eq. (9) describes how parallax for the vertical image is determined:

$$parallax = dy_{MB} - dx_{cur} \quad (9)$$

where dy_{MB} is the vertical motion vector of the macro-block, and dx_{cur} is the average value of vertical motion vectors of the image.

For the local object motion, motion vectors are non-zero only when an object moves. Therefore, the vertical motion vector can be used for creating negative parallax by shifting the object in horizontal direction. The last step for the conversion is motion vector stabilization with a median filter for removing noise components.

The final value of motion vectors from the above process is used for horizontal image shifting and creating the reconstructed image. Consequently, conversion process using the current and the reconstructed images will generate the stereoscopic 3D effect on the image with a negative parallax.

4) Horizontal image

The stereoscopic conversion of the horizontal image with conventional MTD method utilizes the current image and its delayed image by one frame. The new model to be developed in this paper introduces a new concept, the delay factor f_d , for better quality stereoscopic images. The delay factor provides information critical to selecting the most

appropriate frame for the delay image by deciding how many frames need to be skipped from the current image. Eq. (10) shows how to calculate the delay factor;

$$f_d = \text{ROUND} \left[\frac{dx_{del} + dx_{cur}}{dx_{cur}} \right] \quad (10)$$

where dx_{cur} and dx_{del} are the sums of horizontal motion vectors in current and delayed images respectively. The delay factor varies in proportion to the velocity of objects since the sum of motion vectors is used in this equation. In case of rapid motion, for example, a motion vector gets larger due to high velocity. In contrast, slow motion has small value of the motion vector.

The reason why the sum of dx_{cur} and dx_{del} is divided by dx_{cur} is to make it possible to compare the motion vectors between current and delayed images, i.e. the value of a delay factor varies linearly due to the linear variation of motion vectors in two consecutive frames. Therefore, horizontal parallax can be maintained by using a delayed image near the current frame for fast motions (high velocity), and by using a delayed image far from the current frame for slow motions (low velocity).

Fig. 8 shows a stereoscopic conversion using a delay factor. Fig. 8(b) shows three consecutive frames of a car moving to the right. The third image is the current image, and the first and second delayed images are images one frame and two frames before the current image frame, respectively.

As shown in Fig. 8(a), a negative parallax can have different value based on the values of the delay factor. The left figure shows what happens when the first delayed image is used ($f_d=1$) and the right figure shows when the second delayed image is used ($f_d=2$). The negative parallax in the right pair of images has a larger value and has a stronger depth effect than the left pair due to a better choice of a delayed image based on a delay factor. A viewer can have flexibility of selecting the depth level of the 3-D stereoscopic image by choosing the value of a delay factor. However, one should note that using too large value for a delay factor can cause problems creating improper eye matching point on screen because the delayed image for the conversion process might be too far from the current frame. Determining an optimal delay factor for a given image is necessary, but it requires input from viewers at this moment.

After a delay factor is determined, motion directions are used to choose which of the current and delayed frames is presented to the right eye, and which is presented to the left eye.

First, in order to determine a motion direction, a motion is determined to be due to either a global camera motion or a local object motion. The decision can be made by obtaining the difference image between the delayed image (I_{t-1}) and the current image (I_t) as seen in Eq. (11). If the resulting value is larger than a threshold T , it is classified

as a global camera motion. Otherwise, it is a local object motion.

$$\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |I_t(i, j) - I_{t-1}(i, j)| > T \quad (11)$$

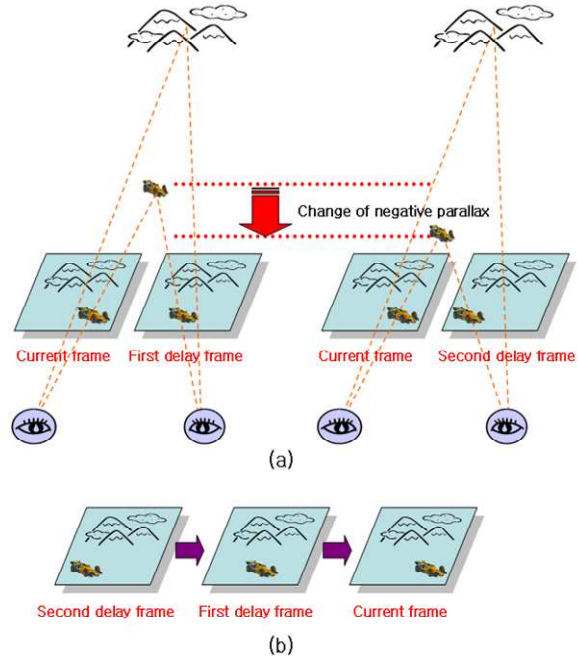


Fig. 8. Stereoscopic conversion using a delay factor

Fig. 9 shows an example of deciding between a global camera motion and an object motion. Since Fig. 9(a) shows an object motion, most of the pixel values in the difference image are zero, and for the global camera motion in (b), most of the pixels in the difference image are non-zero.

When either the global camera or the object motions for the image have been decided, a motion direction is determined from the motion vector. Table 1 shows how to determine left and right images depending on motion directions of the camera and the object. For example, if the camera moves to the right, the motion vector will have a positive value. The delayed image is chosen to be the left, and the current image is chosen to be the right image [12].

Table 1. Assignment of left and right images according to motion direction

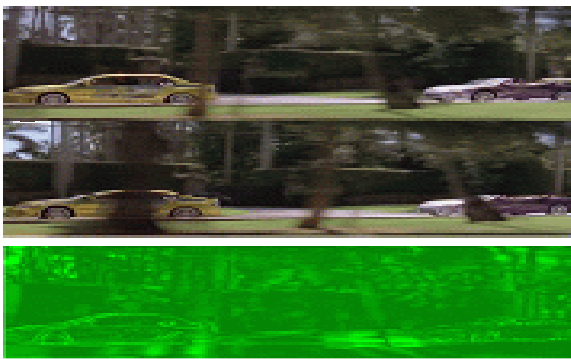
Type	Direction (MV)	Left image	Right image
Object	left (+)	delay frame	current frame
Object	right (-)	current frame	delay frame
Camera	left (-)	current frame	delay frame
Camera	right (+)	delay frame	current frame

4. Experimental Results

The 2D/3D conversion algorithm that has been developed in this paper is implemented in a new H.264 player so that



(a)



(b)

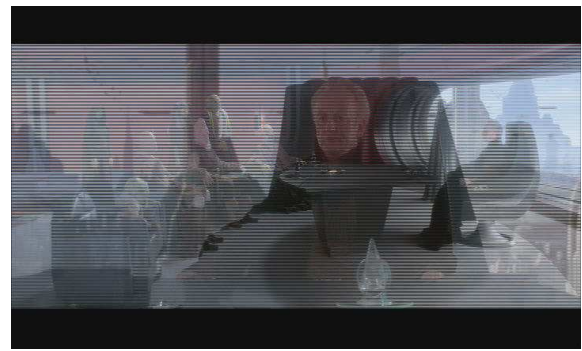
Fig. 9. Difference image (a) local object motion and (b) global camera motion

converted 3D stereoscopic images are displayed properly. To verify the quality of converted 3D images, a 3D stereoscopic monitor with a pair of polarized glasses was used.

Fig. 10 shows an improved converted image with cut detection process when a rapid scene change occurs. During a scene change, if cut detection is not available, both the current and totally different delayed images are overlapped as shown in Fig. 10(a), resulting in no 3D stereoscopic effect. In contrast, Fig. 10(b) shows much better result when cut detection is enabled.

Fig. 11 shows the stereoscopic conversion of a static image. Fig. 11(a) shows motion vectors of a given image. As evident in this figure, motion vectors are negligible; therefore the image is categorized as a static image. Fig. 11(b) shows the result after the stereoscopic conversion. The red box in the figure indicates a reconstructed image shape after conversion process has been applied. The stereoscopic image shows more disparity in the corners than at the center of the image.

In Fig. 12, the stereoscopic conversion of a zoom image is illustrated, which is a zoom-in case. As shown in Fig. 12(a), direction of motion vectors is towards outside and magnitudes are also getting larger towards the outside of the image. In the case of a zoom-out image, motion vectors would have the opposite direction. The converted image



(a)



(b)

Fig. 10. Improved stereoscopic conversion result by cut detection (a) no stereoscopic 3D effect and (b) an improved version

will have positive parallax at the center, which creates a stereoscopic zoom-out effect.

The stereoscopic conversion of a vertical image is shown in Fig. 13. As evident in Fig. 13(a), the directions of most of motion vectors are down; consequently, it is classified as a vertical image. Fig. 13(b) shows the result after the conversion process. Since a horizontal shift is applied for the conversion rather than vertical shift, negative parallax is obtained and an object in the image protrudes out from background.

The result of the stereoscopic conversion for a horizontal image is shown in Fig. 14. As depicted in Fig. 14(a), most of motion vectors point to the right or left, which indicates that the horizontal image conversion process needs to be applied. The delay factor is calculated according to the velocity of an object in the image based on the magnitude of motion vectors; therefore a consistent 3D effect can be obtained regardless speed of objects in the image.

To estimate the quality of the converted results, we performed a subjective test for the two test sequences, which are the same ones as [15] used. The number of subjects was 15 and it was also the same as the number in [15]. The results are shown in Table 2 with the result from [15] and the ones using the actual depth that [15] provided. Here, the actual depth cases were based on the ITU Recommendation BT.500-11 [17]. The items to be tested



Fig. 11. Motion vectors and converted result for a static image: (a) direction of motion vectors; (b) interlaced stereoscopic converted result

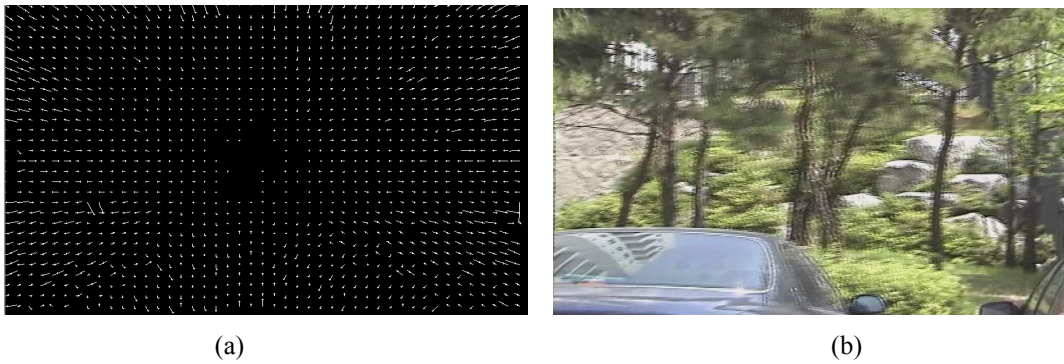


Fig. 12. Motion vectors and converted result for a zoom image: (a) direction of motion vectors; (b) interlaced stereoscopic converted result

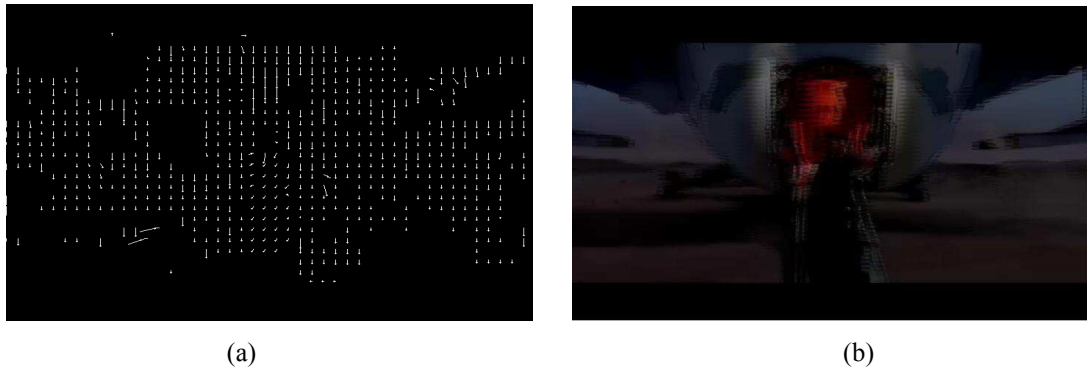


Fig. 13. Motion vectors and converted result for a vertical image: (a) direction of motion vectors; (b) interlaced stereoscopic converted result.

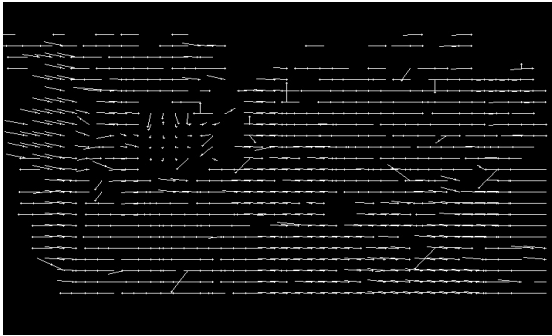
Table 2. Subjective test scores

Subjective Score (out of 10)		Interview		Orbi	
		Active	FPR	Active	FPR
3D Perception	Actual depth	5.50		7.08	
	[15]	6.51		7.28	
	Ours	6.82	6.65	7.49	7.25
Visual Quality	Actual depth	7.91		6.93	
	[15]	6.65		6.15	
	Ours	8.56	7.12	6.55	6.29

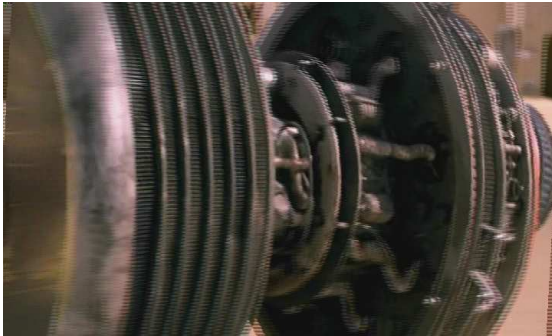
were 3D perception and visual quality as [15], each of which was scored from 1 to 10. Each stream was displayed in the active and the film-type patterned retarder (FPR) 3D displays. In all the tests, the test scores for the active 3D display were a little higher than ones for the FPR 3D display. For visual quality, the cases using the actual depth were better than ours. But for 3D perception ours showed the best. In all the tests, ours showed better scores than [15].

Fig. 15 and Fig. 16 show two examples of anaglyph

images made from the 2D/3D converted results by the method from [16] and ours for the sequence Hall and Coastguard, respectively. It is easily sensible that ours show higher 3D effects without losing the high visual qualities.



(a)



(b)

Fig. 14. Motion vectors and converted result for a horizontal image (a) direction of motion vectors and (b) interlaced stereoscopic converted result



(a)



(b)

Fig. 15. The rendering results for the Hall sequence using the algorithm of (a) [15] and (b) ours.



(a)



(b)

Fig. 16. The rendering results for the Coastguard sequence using the algorithm of (a) [15] and (b) ours.

5. Conclusion

An improved 3D stereoscopic image conversion algorithm for H.264 encoded video has been proposed in this paper. H.264 encoded video has more accurate information on motion vectors than MPEG video so that it produces better stereoscopic image conversion results. While the conventional MTD method is only applicable to horizontal images, the newly developed conversion algorithm can be applied to more general types of images and generates enhanced results.

The 3D conversion algorithm in this paper categorizes given 2D images into one of four different image types; static, zoom, vertical or horizontal images. The global motion of objects and camera motion of an image provide information for the image type decision. Based on its image type, the motion vector information is utilized for the proper image conversion process. In addition to the actual conversion process, cut-detection by using information of macro-blocks from P slice has been added in order to avoid image overlap in the case of a rapid scene change.

For improvement of MTD method for horizontal images, a delay factor has been defined, and it calculates the optimal number of frames between the current image and the delayed (reconstructed) image in order to have maximized stereoscopic 3D effect.

The result of this improved conversion process is that an image flickering has been fixed, and that various image types such as static, zoom and vertical can also have the 3D effect as well as the horizontal image. There may be cases where an image has multiple motion types in a single image frame, and it will require more advanced research work on interpolation of image holes produced by applying parallax for different motion types in a given image.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST). (2010-0026245).

References

- [1] R. Franich, R. Lagendijk and R. Horst, *Reference model for hardware demonstrator implementation*, RACE DISTIMA deliverable 45/TUD/IT/DS/B/003/bl, October, 1992.
- [2] <http://www.tnt.uni-hannover.de/plain/project/eu/distima/>.
- [3] A. Raulo, *State of the art of autostereoscopic displays*, RACE DISTIMA deliverable 45/THO/WP4.2/DS/R/57/01, December 1995.
- [4] Berlin, *PANORAMA Final Demonstrations*, AC092/SIE/FinalDemo/DS/P/032/b1, October 1998.
- [5] S. Malassiotis, and M.G. Strintzis, "Coding of video-conference stereo image sequences using 3D models", *Signal Processing: Image Communications* 9 (1), pp. 125-135, January 1997.
- [6] <http://www.virtue.eu.com/>.
- [7] N. Hur and C. Ahn, "Experimental service of 3DTV broadcasting relay in Korea", *Proc. SPIE* 4864, pp.1-13, 2002.
- [8] C.V. Berkel, and D.W. Parker, "Multi-view 3D-LCD", *Proc. SPIE* 2653, pp. 32-39, 1996.
- [9] R. Borner, "Autostereoscopic direct-view displays and rear-projection for short viewing distances by lenticular method", *Proceedings of the First International Symposium on Three Dimensional Image Communication Technologies*, pp.1-14, Tokyo, December 1993.
- [10] Y. Matsumoto and et al., "Conversion system of monocular image sequence to stereo using motion parallax", *SPIE Vol. 3012*, Photonic West, 1997.
- [11] T. Okino and et. al, "New television with 2D/3D image conversion techniques" *SPIE Vol. 2653*, Photonic West, 1995.
- [12] Man Bae Kim, Sang Hoon Park "2D/3D video conversion based on key frame", *HCI conference*, Feb. 2002.
- [13] <http://www.soft4d.com/>.
- [14] <http://www.ddd.com>.
- [15] M.T. Pourazad, P. Nasiopoulos, and R.K. Ward, "An H.264-based scheme for 2D to 3D video conversion," *IEEE Trans. on Consumer Electronics*, Vol. 55, No. 2, pp. 742-748, May, 2009.
- [16] Feng Jie, Huang Hai, Zheng Yayu, and Zhu Wei, "Stereoscopic 3D conversion from 2D video in H.264/AVC compressed domain," 4th International

Congress on Image and Signal Processing (CISP), Vol. 1, No. 15-17, pp. 565-568, Oct. 2011.

- [17] Ross, J., "Stereopsis by binocular delay", In *Nature*, Vol.248, pp.354-364, 1974.
- [18] Ross, J. and Hogben, J. H., "Short-term memory in stereopsis", In *Vision Research*, Vol. 14, pp. 1195-1201, 1974.
- [19] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU-R Recommendation BT.500-11.



Ho-ki Hong He received B.S. and M.S. degrees in electronics engineering from Kwangwoon University in 2005 and 2007, respectively. He is currently a research engineer at SAMSUNG Electronics. His research interests include 3D image processing and video codec.



Min Soo Ko He received B.S. and M.S. degrees in electronics engineering from Kwangwoon University in 2010 and 2012, respectively. He is currently a PhD student at Kwangwoon University. His research interests include stereo matching and 3D image processing and image compression.



Young-Ho Seo He has received his M.S and Ph.D degree in 2000 and 2004 from Dept. of Electronic Materials Engineering of Kwangwoon University in Seoul, Korea. He was a researcher at Korea Electrotechnology Research Institute (KERI) in 2003 to 2004. He was a research professor of Dept. of Electronic and Information Engineering at Yuhan College in Buchon, Korea. He was an assistant professor of Dept. of Information and Communication Engineering at Hansung University in Seoul, Korea. He is now an associated professor of College of Liberal Arts at Kwangwoon University in Seoul, Korea and a director of research institute in TYJ Inc. His research interests include realistic media, digital holography, SoC design and contents security.



Dong-Wook Kim He has received his M.S degree in 1985 from Dept. of Electronic Engineering of Hangyang University in Seoul, Korea and his Ph.D degree in 1991 from Dept. of Electrical Engineering of Georgia institute of Technology in GA, U.S.A. He is a Professor of Dept of Semiconductor and New Materials Engineering

at Kwnagwoon University in Seoul, Korea. He is now the chief of professional activities executive group of SSCS/EDS Joint chapter of IEEE Seoul section. His research interests include digital image processing VLSI design, and design of digital signal processor.



Jisang Yoo Jisang Yoo was born in Seoul, Korea in 1962. He received the B.S. and M.S. degrees from Seoul national university, Seoul, Korea in 1985 and 1987, all in electronics engineering, and Ph.D. degree from Purdue University, West Lafayette, IN, USA, in electrical engineering in 1993,

respectively. From September 1993 to august 1994, he worked as a senior research engineer in industrial electronics R&D center at Hyundai Electronics Industries Co., Ltd, Inchon, Korea, in the area of image compression and HDTV. He is currently a professor with the department of electronics engineering, Kwangwoon University, Seoul, Korea. His research interests are in signal and image processing, nonlinear digital filtering, and computer vision. He is now leading 3DTV broadcast trial project in Korea.