Temporal Anti-aliasing of a Stereoscopic 3D Video

Wook-Joong Kim, Seong-Dae Kim, Namho Hur, and Jinwoong Kim

Frequency domain analysis is a fundamental procedure for understanding the characteristics of visual data. Several studies have been conducted with 2D videos, but analysis of stereoscopic 3D videos is rarely carried out. In this paper, we derive the Fourier transform of a simplified 3D video signal and analyze how a 3D video is influenced by disparity and motion in terms of temporal aliasing. It is already known that object motion affects temporal frequency characteristics of a time-varying image sequence. In our analysis, we show that a 3D video is influenced not only by motion but also by disparity. Based on this conclusion, we present a temporal anti-aliasing filter for a 3D video. Since the human process of depth perception mainly determines the quality of a reproduced 3D image, 2D image processing techniques are not directly applicable to 3D images. The analysis presented in this paper will be useful for reducing undesirable visual artifacts in 3D video as well as for assisting the development of relevant technologies.

Keywords: Stereoscopy, three-dimensional video, temporal aliasing, anti-aliasing filter, Fourier transform.

I. Introduction

Temporal aliasing (also called a wagon-wheel effect or a strobing effect) is one of the fundamental artifacts in a time-varying image sequence caused by insufficient temporal sampling [1]. When objects in a video move quickly, viewers can notice unrealistic motions such as jerkiness, or typically, the reverse rotation of wheels. Especially in computer animations, temporal aliasing requires special attention. Each frame of a computer animation is created with pre-defined models which have clear edges, and this causes sharpness of motion which is easily noticed by a viewer. Hence, to remove such visual discomfort, a motion blurring technique, which stretches a moving object along the path of motion, is generally used in rendering computer animations [2]. Fortunately, in real scene videos, temporal aliasing is less serious than in computer-generated images because cameras naturally provide a blurring effect by accumulating light over the duration of the exposure time. Though we can partially rely on the interaction between light and cameras, temporal aliasing is still an issue to be studied because a camera can never guarantee freedom from temporal aliasing.

To identify the cause of aliasing, frequency domain analysis is necessary. The Fourier transform of conventional two-dimensional (2D) images can be easily obtained by a well-known method such as 2D fast Fourier transform (FFT), and a variety of image processing techniques have been developed by analyzing the transform results. However, the frequency analysis of stereoscopic three-dimensional (3D) images requires a different approach than that of 2D images because the depth of a 3D image is illusory data reproduced by the human brain. Many 3D displays provide a 3D scene using human binocular vision [3], [4]. A stereo image pair (left-eye and right-eye images) is multiplexed and projected onto the screen of a 3D display, and the 3D display delivers each image

Manuscript received July 9, 2008; revised Nov. 21, 2008; accepted Dec. 10, 2008.

This work was supported by Brain Korea 21 Project, the School of Information Technology, KAIST, in 2008 and by Top Brand Project of MEST, ETRI, in 2008.

Wook-Joong Kim (phone: + 82 42 350 5430, email: wjk@sdvision.kaist.ac.kr) and Seong-Dae Kim (email: sdkim@sdvision.kaist.ac.kr) are with the Department of Electrical Engineering, KAIST, Daejeon, Rep. of Korea.

Namho Hur (email: namho@etri.re.kr) and Jinwoong Kim (email: jwkim@etri.re.kr) are with Broadcasting and Telecommunication Convergence Research Laboratory, ETRI, Daejeon, Rep. of Korea.

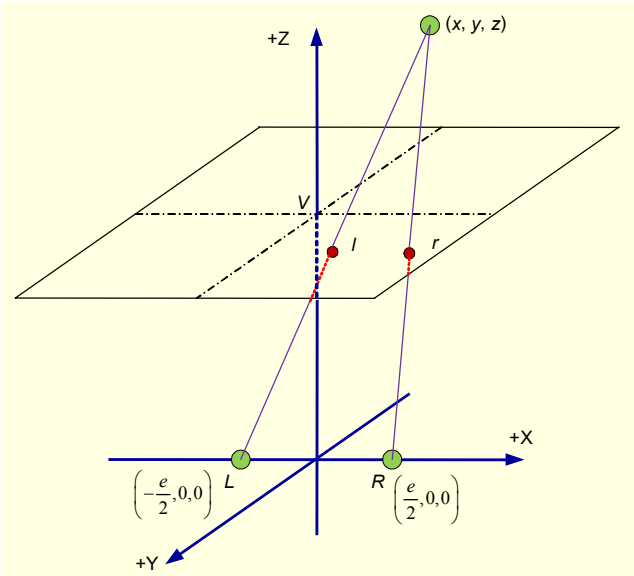

Fig. 1. Geometry model for a planar stereoscopic display. The origin of the coordinate system (X, Y, Z) is located at the center of the left eye (L) and right eye (R). The interocular distance is denoted by e and the display screen is assumed to be located with the viewing distance V.

to its corresponding eye through special optical mechanisms, such as polarized filters, lenticular sheets, or parallax barriers [3]. Hence, the frequency characteristics of a 3D image cannot be interpreted only with the projected images on a 3D display screen; rather, consideration of the process of human depth perception is also required.

The depth perception process is commonly formulated by a geometric model (see Fig. 1) and analysis based on this geometric model is widely used to recognize 3D information, such as the amount of perceivable depth or structural distortions of a scene (such as the cardboard effect and puppet theater effect) [5]-[8]. Recently, a frequency domain analysis of a 3D image was presented with the geometry model [9], and it was shown that a 3D image has an intrinsic cause of spatial frequency aliasing due to frequency scaling by disparity. In this paper, as an extension of [9], we present the Fourier transform of a 3D video (time-varying 3D image sequence) with motion and disparity simplification. We then interpret the transformed result from a temporal aliasing perspective.

For a 2D video signal, temporal aliasing is a thoroughly studied problem. In [10]-[12], it was explained that object motion causes tilting of the spectrum plane toward the temporal frequency domain, and consequently, aliasing is more likely to occur as the motion becomes bigger. However, frequency analysis of 3D video has been rarely reported, and accordingly, designing a filter for 3D video lacks theoretical grounds. From our analysis, we found that the tilting of the spectrum plane for 3D video is affected not only by motion but also by disparity; hence, anti-aliasing of 3D video needs to be devised considering both motion and disparity. Based on this conclusion, we propose a temporal anti-aliasing filter for stereoscopic 3D video.

The remainder of this paper is organized as follows. In section II, we briefly describe background information for our analysis. In section III, we derive frequency domain characteristics of a 3D video and propose a temporal anti-aliasing filter. In section IV, simulation and subjective quality evaluation results are presented to support validity of the filter, and finally, we conclude in section V.

II. Background

As background information, we briefly describe the conclusion of [9] including a geometry model for binocular vision. Currently available 3D displays provide a depth impression primarily based on binocular vision (also called stereopsis) [3]. The degree of perceived depth, though it varies according to viewers' abilities, is generally formulated by a geometry model. Figure 1 is a simplified model of a planar stereoscopic display assuming the ideal display conditions (no optical diffraction, no crosstalk, and so on). It specifies the amount of perceived depth based on the disparity between stereo matching pairs in left and right images. A coordinate system (X, Y, Z) is defined to have its origin located at the center point between the two eyes. The left eye L and right eye R are separated by the interocular distance e and their locations are set to $(-e/2, 0, 0)$ and $(e/2, 0, 0)$, respectively. In the model, the display screen is located with the viewing distance V; that is, the screen is located on the plane of $Z=V$. Given a stereo pair $l=(x_l, y_p, V)$ and $r=(x_r, y_p, V)$ having horizontal disparity only (that is, the Y coordinate values are the same[1]), the coordinate of the perceived 3D pixel (x, y, z) from the stereo pair is defined as

$$\begin{aligned}(x,y,z) &= \left(\frac{e(x_l+x_r)}{2(e-d)}, \frac{ey_p}{e-d}, \frac{eV}{e-d}\right) \\ &= \left(\frac{(x_l+x_r)}{2}\cdot\frac{z}{V}, y_p\frac{z}{V}, z\right),\end{aligned} \quad (1)$$

where $d=x_r-x_l$ is the disparity of the stereo pair l and r.

In [9], the frequency spectrum of a 3D image was derived with the simplification of disparity distribution. When a left image $I_L(u, v)=f(u, v)$ and a horizontal disparity map $\delta_h(u,v)$

1) Left-eye and right-eye images must be rectified prior to their being projected onto a display screen for 3D reproduction. Without a rectification process, a stereo pair will not be aligned on the same epipolar line, and consequently, the quality of reproduced 3D scenes will be disturbing. Hence, for a 3D display model, vertical disparity is commonly ignored without loss of generality.

are given, where u and v are horizontal and vertical coordinates on a display screen (that is, $(u, v)=(x, y, V)$) , the corresponding right image $I_R(u, v)$ can be formulated as $I_R(u,v)=f(u+\delta_h(u,v),v)$. Given a stereo pair $c'_L=(u',v')$ and $c'_R=(u'+\delta',v')$, the position of a 3D pixel (x_i, y_i, z_i) to be perceived is determined using (1) as

$$(x_i,y_i,z_i)=\left(\frac{e(2u'+\delta')}{2(e-\delta')},\frac{ev'}{e-\delta'},\frac{eV}{e-\delta'}\right), \tag{2}$$

and modifying (2), we have

$$\begin{aligned} u' &= \frac{e-\delta'}{e}x_i-\frac{\delta'}{2}=\frac{V}{z_i}x_i-\frac{\delta'}{2}, \\ v' &= \frac{e-\delta'}{e}y_i=\frac{V}{z_i}y_i. \end{aligned} \tag{3}$$

Note that (3) guides us to find the position of a stereo pair on a display screen when a 3D pixel position and corresponding disparity are given.

Substituting the relation of (3) for (u, v), a 3D image $I_{3D}(s)$, where $z=(x, y, z)$ is a perceivable 3D pixel point in (X, Y, Z) coordinates, is defined as

$$I_{3D}(s)=\begin{cases} f\left(\frac{V}{z}x-\frac{\Delta(s)}{2},\frac{V}{z}y\right) & \text{if } z=\frac{eV}{e-\Delta(s)}, \\ \phi & \text{otherwise,} \end{cases} \tag{4}$$

where $\Delta(s)$ is the corresponding disparity for a 3D position s. Note that a 3D image is represented as a function of a 2D image. That is, colors of 3D positions on a perceivable 3D image are obtained from a left-eye image. The significance of (4) is that it allows us to find a property of the 3D image by analyzing the 2D function with disparity.

Since $I_{3D}(s)$ is defined as a variation of f, as in (4), the Fourier transform of $I_{3D}(s)$ can be also represented as a function of the Fourier transform of f. Assuming constant disparity for all positions, that is, $\Delta(s)=p$ where p is constant, the Fourier transform of $I_{3D}(s)$, $FT\{I_{3D}(s)\}$, is obtained as

$$\begin{aligned} FT\{I_{3D}(s)\} &= \iiint_{x,y,z} I_{3D}(s)e^{-j[\omega_x x+\omega_y y+\omega_z z]}dxdydz \\ &= e^{-j\omega_z Z_0}\cdot FT\left\{f\left(\frac{V}{Z_0}x-\frac{p}{2},\frac{V}{Z_0}y\right)\right\} \\ &= \left(\frac{Z_0}{V}\right)^2 F\left(\frac{Z_0}{V}\omega_x,\frac{Z_0}{V}\omega_y\right)e^{-j\left(\frac{pZ_0}{2V}\omega_x+Z_0\omega_z\right)}, \end{aligned} \tag{5}$$

where $Z_0=Ve/(e-p)$, and $F(\omega_x,\omega_y)$ is the Fourier transform of $f(x, y)$. Note that ω_x and ω_y are normalized frequencies ranging over an interval of length 2π .

The key finding from (5), which was described in [9], is frequency scaling by the factor of Z_0/V . This frequency scaling causes stretching or shrinking of F according to the magnitude and direction of p. Especially, when p is negative (crossed), $I_{3D}(s)$ is likely to contain frequency aliasing due to stretching of the spectrum on spatial domains. When this aliasing occurs, viewers are likely to notice jagged and/or twisted edges especially in high-frequency regions. In the next section, using the notations and definitions mentioned above, we derive the frequency spectrum of a 3D video.

III. Temporal Frequency Analysis of a 3D Video

We now extend our analysis to a 3D video. Using the model of Fig. 1 and (4), we define a 3D video as

$$I_{3D}(s;t)=\begin{cases} g\left(\frac{V}{z}x-\frac{\Delta(s;t)}{2},\frac{V}{z}y;t\right) & \text{if } z=\frac{eV}{e-\Delta(s;t)}, \\ \phi & \text{otherwise,} \end{cases} \tag{6}$$

where $I_{3D}(s;t)$ and $\Delta(s;t)$ denote a perceivable color of a 3D video and disparity at position s and time t, respectively; and $I_L(u, v; t)=g(u, v; t)$ denotes a 2D video signal for a left eye. Note that (6) is simply defined by adding time variation t into (4). Assuming constant disparity again for all 3D positions and time (that is, $\Delta(s;t)=p$), we obtain the Fourier transform of $I_{3D}(s;t)$ similarly to (5) as

$$FT\{I_{3D}(s;t)\}=e^{-j\omega_z Z_0}\cdot FT\left\{g\left(\frac{V}{Z_0}x-\frac{p}{2},\frac{V}{Z_0}y;t\right)\right\}. \tag{7}$$

In the remainder of this section tilting of the spectrum plane by motion is described. This property is extended to include disparity, and we analyze temporal aliasing of a 3D video. Finally, the ideal transfer function of a temporal anti-aliasing filter is presented.

1. Tilting of a Spectrum Plane by Motion

In [11] and [12], the relation between temporal aliasing and motion for a 2D video was described. They analyzed the frequency spectrum of a 2D video and interpreted the cause of temporal aliasing as tilting of the plane where the spectrum of the video exists. Suppose that a video signal $g(u, v; t)$ has rigid-body transformation with constant motion velocities as

$$g(u,v;t)=g_0(u+m_u t,v+m_v t), \tag{8}$$

where m_u and m_v denote the constant motion velocity in u and v directions, respectively, and $g_0(u, v)$ is the frame at $t=0$, that is, $g_0(u, v)=g(u, v; 0)$. Using (8), the Fourier transform of $g(u, v; t)$

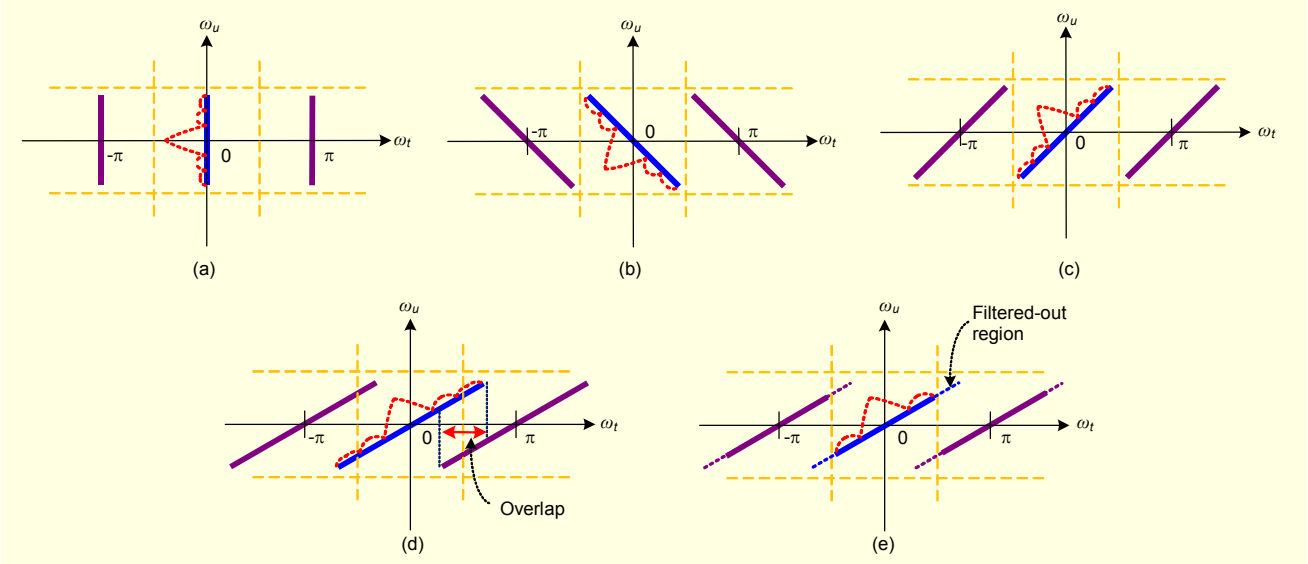

Fig. 2. Variation of spectrum planes in terms of motion on the $\omega_t - \omega_u$ domains. The dotted curve on a center line represents the frequency spectrum. Each diagram shows how the slope of a spectrum plane changes according to motion as (a) zero motion (m_u=0), (b) negative motion (m_u=-1), (c) positive motion (m_u=1), (d) positive motion bigger than 1 (m_u=2), and (e) desired spectrum shapes for temporal alias-free.

can be represented as

$$\begin{aligned} &FT\{g(u,v;t)\} \\ &= \iiint_{u,v,t} g(u,v;t)e^{-j[\omega_u u+\omega_v v+\omega_t t]}dudvdt \\ &= \iiint_{u,v,t} g_0(u+m_u t, v+m_v t)e^{-j[\omega_u u+\omega_v v+\omega_t t]}dudvdt \\ &= FT\{g_0(u,v)\}\delta(\omega_t - m_u\omega_u - m_v\omega_v) \\ &= G_0(\omega_u,\omega_v)\delta(\omega_t - m_u\omega_u - m_v\omega_v), \end{aligned} \quad (9)$$

where $G_0(\omega_u,\omega_v)$ is the Fourier transform of $g_0(u, v)$, and $\delta(\cdot)$ is the delta function. One important property that we can find from (9) is that the spectrum of $g(u, v; t)$ resides on the plane $\omega_t - m_u\omega_u - m_v\omega_v = 0$. We define this plane as a *spectrum plane*. Figure 2 illustrates various spectrum planes in the $\omega_t - \omega_u$ domains.

When there is no motion, as shown in Fig. 2(a), the baseband spectrum and its replicas exist on the planes of $\omega_t = n\pi$, where n is an integer. However, when there is non-zero constant motion, the motion causes tilting of the spectrum plane. Figures 2(b) and (c) respectively illustrate the spectrum planes of a 2D video signal which has negative (m_u=-1) and positive (m_u=1) motion velocities. Furthermore, if the velocity of motion becomes greater than 1, as shown in Fig. 2(d), the spectrum replicas may have an overlapping region in the ω_t domain, which consequently causes temporal aliasing. Therefore, to prevent temporal aliasing, the spectrum of $G_0(\omega_u,\omega_v)$ must be band-limited not to contain any components above $\pi/2$ in the temporal frequency domain as shown in Fig. 2(e), and its cutoff frequency needs to be determined according to motion velocities.

2. Temporal Aliasing of a 3D Video

We now describe the influence of disparity on the frequency spectrum of a 3D video. Combining (7) and (9), we obtain the $FT\{I_{3D}(s;t)\}$ as

$$\begin{aligned} &FT\{I_{3D}(s;t)\} \\ &= e^{-j\omega_z Z_0}\cdot FT\left\{g\left(\frac{V}{Z_0}x-\frac{p}{2},\frac{V}{Z_0}y;t\right)\right\} \\ &= e^{-j\omega_z Z_0}\cdot FT\left\{g_0\left(\frac{V}{Z_0}x-\frac{p}{2}+m_x t,\frac{V}{Z_0}y+m_y t\right)\right\} \\ &= \left(\frac{Z_0}{V}\right)^2 G_0\left(\frac{Z_0}{V}\omega_x,\frac{Z_0}{V}\omega_y\right) \\ &\quad\times\delta\left(\omega_t - m_x\frac{Z_0}{V}\omega_x - m_y\frac{Z_0}{V}\omega_y\right)e^{-j\left(\frac{pZ_0}{2V}\omega_x+Z_0\omega_z\right)}. \end{aligned} \quad (10)$$

Note that u and v are respectively substituted by x and y in the (X, Y, Z) coordinates. Surely, this result has clear limitation for representing all the frequency characteristics of a 3D video because it is obtained with the simplification of motion and disparity distributions. We assumed constant rigid body motion and constant disparity for all pixels. However, this brings up a significant point for consideration, especially regarding how the 3D video spectrum is influenced by disparity and motion. For convenience, we define a new value α_p as $\alpha_p = Z_0/V = e/(e-p)$.

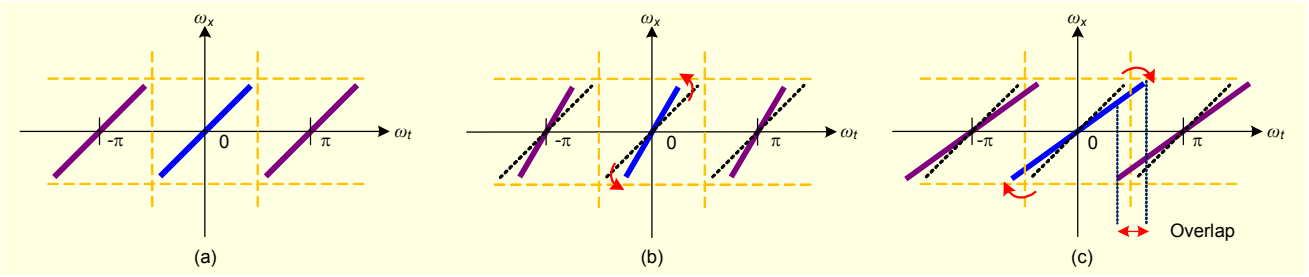

Fig. 3. Slope variation of spectrum planes with respect to disparity. The slope of a spectrum plane for a 3D video is influenced by disparity even with fixed motion. Assuming the motion velocities as m_x=1, each diagram shows the slope changes according to p: (a) zero disparity (p=0), (b) negative (crossed) disparity (p<0), and (c) positive (uncrossed) disparity (p>0).

When the magnitude of p is small enough compared to the interocular distance ($e >> |p|$), α_p is almost equal to 1 since $e/(e-p) \approx 1$. However, as the magnitude of disparity or the dimensions of a display screen increase, α_p becomes considerable.

As (10) shows, the spectrum $FT\{I_{3D}(s;t)\}$ resides on a plane $\omega_t - m_x\alpha_p\omega_x - m_y\alpha_p\omega_y = 0$, which is the spectrum plane for a 3D video. Figure 3 shows an example of how the slope of a spectrum plane varies according to disparity. Figure 3(a) shows the baseband spectrum of a 3D video and its replicas on the $\omega_t - \omega_x$ domains when m_x=1 and p=0. Since there is no disparity, the slope of the spectrum planes is determined only by m_x, and this is identical to Fig. 2(c). When p has a negative value (α_p<1), as shown in Fig. 3(b), the spectrum plane has a bigger slope than that seen in Fig. 3(a). Consequently, the planes are titled up toward the ω_x axis. For the opposite (positive or uncrossed) disparity, the spectrum planes are tilted down as shown in Fig. 3(c), and the spectrum replicas likely have overlapping regions on the ω_t domain, whereas there is no overlapping in Figs. 3(a) and (b). As previously described with Fig. 2, this overlapping implies that temporal aliasing is occurring.[2)]

3. Temporal Anti-alias Filter

Viewing comfort for stereoscopic images has been widely studied [13]-[15], and one of the key findings is that image smoothing by an appropriate low-pass filter contributes to reducing the visual discomfort of a 3D scene [13]. Based on the analysis presented previously, we propose a filter for temporal anti-aliasing of a 3D video.

To prevent spectrum overlapping in the temporal frequency domain, the spectrum has to be limited within the range of $(-\pi/2, \pi/2)$. This limitation can be achieved by a low-pass filter whose transfer function is defined in the temporal frequency domain as

$$H\left(e^{j\omega_t}\right) = \begin{cases} 1, & \text{if } |\omega_t| < \frac{\pi}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

Practically, however, an image filter defined in the temporal frequency domain is less desirable than a filter in the spatial frequency domains due to a computational burden caused by frame memories and filtering delay. Using the equation of the spectrum plane $\omega_t = m_x\alpha_p\omega_x + m_y\alpha_p\omega_y$, we can transform this filter to the $\omega_x - \omega_y$ domains as

$$H\left(e^{j\omega_x}, e^{j\omega_y}\right) = \begin{cases} 1, & \text{if } \left|m_x\alpha_p\omega_x + m_y\alpha_p\omega_y\right| < \frac{\pi}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Note that the frequency cutoff of this filter is determined adaptively by pixel motion and disparity regardless of viewing distance, object, or background. Furthermore, temporal anti-aliasing can be achieved by spatial-domain filtering while a desirable performance is maintained.

Meanwhile, the transfer function of the temporal anti-aliasing filter (12) is ideal only when a 3D video has uniform motion and disparity, which is unrealistic for most 3D videos. However, this transfer function is still valid in practically designing the filter. An FIR-type filter is most desirable for image filtering (5-tap or 7-tap). Though uniform disparity and motion for the whole image region is rarely possible, we can assume that the uniformity would be satisfied for a small region due to the similarity among neighboring pixels. That is, a small region in an image can be considered as a rigid body having constant motion and disparity. Therefore, assuming that the region-of-support of an FIR filter is a rigid body, we can obtain close results to the ideal by adaptively changing the transfer function of the filter according to motion and disparity of each pixel.

Additionally, we may need extra consideration from the

2) Though negative (crossed) disparity does not directly cause temporal aliasing, it can cause spatial aliasing for a 3D still image [9]. Therefore, to eliminate spatial and temporal frequency aliasing artifacts in a 3D video both disparity types have to be considered in an anti-aliasing process. In this paper, we limit our consideration to the temporal frequency aliasing of a 3D video.

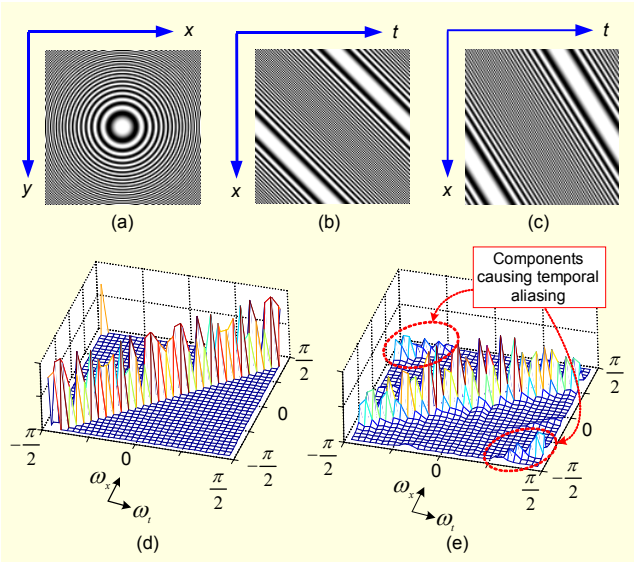

Fig. 4. Test image sequences and the spectrum magnitude in the $\omega_x - \omega_t$ domains: (a) a zoneplate image, (b) a slice image on the *x-t* plane from the 2D video, (c) a slice image on the *x-t* plane from the virtually created 3D video, (d) the spectrum magnitude of (b), and (e) the spectrum magnitude of (c).

aspect of computational complexity. Adaptive filtering for each pixel could be a burden for fast computing since each pixel would require different filter coefficients. Hence, the filtering process would need to be further simplified while maintaining the desired anti-aliasing performance. As a simple solution, we may determine the filter coefficients for a whole image frame using the maximum magnitude motion and disparity. Even though over-blurring for some pixels is unavoidable, removal of temporal anti-aliasing can be achieved.

IV. Experiments

In this section, we show the results of proposed temporal anti-aliasing filtering on a 3D video. First, we present the frequency limitation performance simulation under a virtual viewing environment with a zoneplate image sequence, and second, the results of subjective quality assessment are presented.

1. Simulations with a Zoneplate Image Sequence

We managed a simulation under a virtual stereo-viewing environment to show the effect of the proposed temporal anti-aliasing filter. First, we generated a 2D image sequence with a zoneplate image as shown in Fig. 4(a), assuming the motion of 1 pel/frame for all pixels in horizontal and vertical directions (that is, m_x=m_y=1). Figure 4(b) shows a slice on the *x-t* plane from the sequence. The diagonal lines illustrate the translational motion in the *x* direction, and their slopes represent its velocity. Figure 4(d) shows the spectrum magnitude of Fig. 4(b). As explained with (9) and Fig. 2(c), most frequency components exist on a line which is the spectrum plane of the sequence.

Based on (6), we created a 3D video using the zoneplate image sequence. The viewing distance V and the interocular distance e are set to 600 mm and 65 mm, respectively, and disparity for matching stereo pairs is set to be 22 mm (that is, p=22).[3] Figure 4(c) shows a slice image on the *x-t* plane from the created 3D video, and Fig. 4(e) is the spectrum magnitude of the slice.

As explained with (10) and Fig. 3(c), the spectrum of the 3D video contains additional frequency components above and below the spectrum plane, which are the cause of temporal aliasing. Even though the 2D and 3D videos have the same number of motions, the lines in Fig. 4(c) have a different slope from those of Fig. 4(b) due to disparity. Those of the spectrum plane of Fig. 4(e) also differ from those of Fig. 4(d).

With the synthesized zoneplate 3D video, we now show the performance of the temporal anti-aliasing filter. Since m_x=m_y=1 and $\alpha_p = e/(e-p) \approx 1.51$, the transfer function of the 3D temporal anti-aliasing filter for this 3D video becomes

$$H\left(e^{j\omega_x}, e^{j\omega_y}\right) = \begin{cases} 1, & \text{if } -\frac{\pi}{2} < 1.51\cdot\omega_x + 1.51\cdot\omega_y < \frac{\pi}{2}, \\ 0, & \text{otherwise}. \end{cases} \quad (13)$$

Figures 5(a) and (b) are the output zoneplate image and its Fourier transform after filtering, respectively. Note that the upper-right and lower-left parts of Fig. 5(a) are smoothed by the filter. Figure 5(b) is the spectrum magnitude of the filtered 2D video in the $\omega_t - \omega_x$ domains, and comparing this with Fig. 4(d), it can be easily noticed that the high frequency components are reduced. Finally, Fig. 5(c) illustrates the spectrum of the filtered 3D video in the $\omega_t - \omega_x$ domains. The frequency components causing temporal aliasing, which exist in Fig. 4(e), are successfully suppressed by the filter.

2. Subjective Quality Assessment

Since a 3D scene is created by the human brain, subjective viewing is the best measure for 3D image/video quality evaluation. We conducted a subjective test to evaluate the performance of temporal anti-aliasing filtering. It is known that the outcome of subjective viewing varies according to various factors such as the viewers' experiences, 3D perception abilities, lighting condition, crosstalk of a 3D display, and so on. To

3) For general LCD monitors, the size of a single pixel is about 0.4 mm. We obtained p=22 assuming 55 pixels for the disparity (that is, 55×0.4=22).

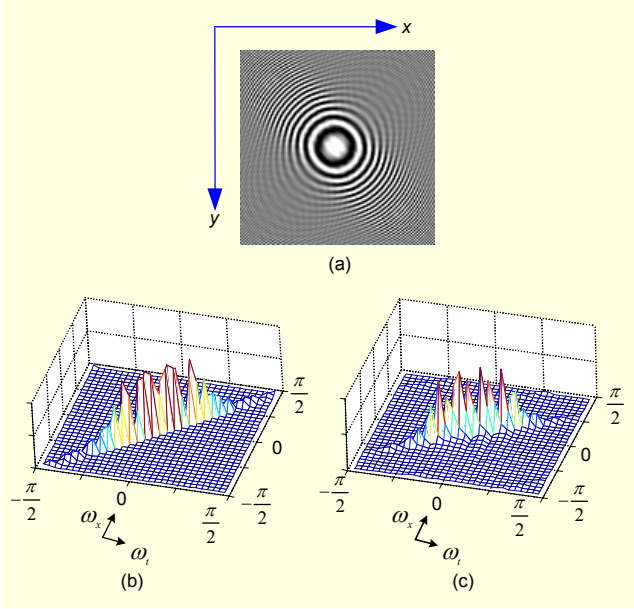

Fig. 5. Performance of the proposed temporal anti-aliasing filter: (a) an output image of the temporal anti-aliasing filter, (b) the spectrum magnitude of the filtered 2D video in the $\omega_x - \omega_t$ domains, and (c) spectrum magnitude of the 3D video after temporal anti-aliasing filtering.

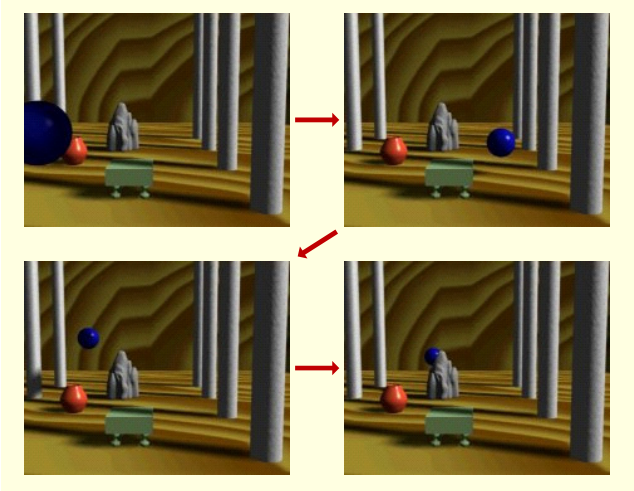

Fig. 6. Screenshots of the 3D video used for the subjective quality assessment. The zero-disparity plane is placed at the front side of the table in the middle. A ball appears from the left-side with negative (crossed) disparity (the left-top figure); passes the zero-disparity plane (the right-top figure); and it moves backward swinging from side to side with positive (uncrossed) disparity.

minimize other influences apart from temporal aliasing and filtering, we intentionally selected viewers who had a certain level of knowledge about temporal aliasing of digital videos, and we asked them to pay more attention to moving objects which are susceptible to temporal aliasing as well as smoothing by the proposed filter.

Ten viewers participated in this test, and we used a autostereoscopic 3D display from [16]. A synthesized 3D video was used for this test, and Fig. 6 shows four frames of the video. In synthesizing the test 3D video, we made the speed of the ball movement fast enough for us to perceive jerky motion when it was presented without filtering. With the original 3D video, we prepared two different 3D videos: one by a motion blurring method for left and right-eye images (that is, 3D video processed by a 2D-image temporal anti-aliasing method), and the other was processed by the proposed anti-aliasing filter. We presented the two 3D videos to the viewers and asked for their comments regarding two aspects: perceivable motion naturalness and edge reproduction of the moving ball.

Regarding the motion naturalness, five viewers commented that both 3D videos provided almost same motion naturalness and the others preferred the 3D video filtered by the proposed method. For edge reproduction, seven viewers judged that boundary blurring was less visible in the 3D video filtered by the proposed filter. This latter observation suggests that the proposed anti-aliasing filter can be effective in reducing artifacts in terms of 3D object motion that might lead to visual discomfort. However, more tests with a wider range of video content would be required to reinforce and confirm this initial finding.

V. Conclusion

In this paper, we presented the frequency analysis of a 3D video from the aspect of temporal aliasing. From the analysis, we found that both motion and disparity in a 3D video influence the temporal frequency characteristics (titling of a spectrum plane), and temporal aliasing of a 3D video is more likely to happen due to the presence of positive (uncrossed) disparity. Based on this conclusion, we proposed a temporal anti-aliasing filter whose frequency cutoff is adaptively determined by pixel motion and disparity. It is known that the amount of screen disparity needs to be limited for viewing comfort, and a certain level of low-pass filtering contributes to enhancing the perceptual quality of a 3D video [13]. Our analysis is consistent with this knowledge, which supports the reliability of this analysis, and also provides theoretical grounds for further understanding of 3D video. A number of 2D image/video processing methods were developed based on the frequency analysis of visual data. However, such 2D methods are not directly applicable for 3D images/videos since the process of depth perception has to be incorporated. We believe that the presented analysis in this paper will be useful for improving 3D scene reproduction performance, and this will contribute to the further development of 3D image/video services.

References

[1] J. Korein and N. Badler, "Temporal Anti-aliasing in Computer Generated Animation," *ACM SIGGRAPH Computer Graphics*, vol. 17, no. 3, July 1983, pp. 377-388.

[2] M. Potmesil and I. Chakravarty, "Modeling Motion Blur in Computer-Generated Images," *Proc. of Conf. on Computer Graphics and Interactive Techniques*, 1983, pp. 389-399.

[3] P. Benzie et al., "A Survey of 3DTV Displays: Techniques and Technologies," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, Nov. 2007, pp.1647-1658.

[4] H. Ozaktas and L. Onural, *Three-Dimensional Television: Capture, Transmission, Display*, Springer, Dec. 2007.

[5] A. Woods, T. Docherty, and R. Kock, "Image Distortions in Stereoscopic Video Systems," *Proc. SPIE:* Stereoscopic Displays and Applications IV, vol. 1915, Feb. 1993, pp. 36-48.

[6] Z.Y. Alpaslan and A.A. Sawchuk, "Multiple Camera Image Acquisition Models for Multi-view 3D Display Interaction," *Proc. IEEE 6th Workshop Multimedia Signal Processing*, Sept. 30, 2004, pp. 256-262.

[7] B. Froner and N.S. Holliman, "Implementating an Improved Stereoscopic Camera Model," *Proc. Eurographics Theory and Practice of Computer Graphics*, June 2005.

[8] H. Yamanoue, M. Okui, and F. Okano, "Geometrical Analysis of Puppet-Theater and Cardboard Effects in Stereoscopic HDTV Images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, June 2006, pp. 744-752.

[9] W. Kim and J. Kim, "Disparity Adaptive Filter for Anti-aliasing of Stereoscopic 3D Images," *Proc. of IEEE 3DTV-Conference*, May 2008, pp. 41-44.

[10] A. Watson, A. Ahumada, and J. Farrell, "Window of Visibility: A Psychophysical Theory of Fidelity in Time-Sampled Visual Motion Displays," *Journal of the Optical Society of America A*, vol. 3, no. 3, Mar. 1986, pp. 300-307.

[11] H. Kang at al., "3-D Band Limitation by Motion Adaptive Spatial Filtering," *Proc. ICASSP*, Apr. 1994, pp. 513-516.

[12] S. Kim and S.-D. Kim, "Analysis of the Effect of Motion Estimation Error on the Motion Adaptive Spatial Filter and Its Application," *Optical Engineering*, vol. 35, no. 7, July 1996, pp. 2045-2050.

[13] G. Um et al, "Investigation on the Effect of Disparity-Based Asymmetrical Filtering on Stereoscopic Video," *Proc. of SPIE*, vol. 5150 (Visual Communications and Image Processing), 2003, pp. 110-118.

[14] C. Moller and A. Travis, "Correcting Interperspective Aliasing in Autostereoscopic Displays," *IEEE Trans. on Vis. Comput. Graphics*, vol. 11, no. 2, Mar./Apr. 2005, pp. 228-236.

[15] J. Konrad and P. Agniel, "Subsampling Models and Anti-alias Filters for 3-D Automultiscopic Displays," *IEEE Trans. Image Process.*, vol. 15, no. 1, Jan. 2006, pp.128-140.

[16] http://www.vrlogic.com/html/stereographics/synthagram.html

Wook-Joong Kim received his BS, MS, and PhD degrees from the Department of EE, KAIST, Daejeon, Korea, in 1993, 1995, and 1999, respectively. From December 1999 to September 2007, he was with Electronics and Telecommunications Research Institute (ETRI) as Senior Research Staff and he was a visiting scholar with Michigan State University from December 2005 to November 2006. He has been currently a research professor with the Department of EE, KAIST, since October 2007. His research interests include stereoscopic 3D image processing, video compression, and image-based modeling.

Seong-Dae Kim received his BS in electronics engineering from Seoul National University, Seoul, Korea, in 1977, and his MS in electrical engineering from KAIS, Seoul, Korea, in 1979. He obtained his Dr.Ing in electrical engineering from ENSEEIHT, INPT, Toulouse, France, in 1983. Since 1984, he has been a professor with the School of Electrical Engineering and Computer Science at KAIST. His research interests are image processing, computer vision, pattern recognition, and image coding.

Namho Hur received the BS, MS, and PhD degrees in electrical and electronic engineering from Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 1992, 1994, and 2000, respectively. He is currently with the Broadcasting and Telecommunications Convergence Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. He is also an associate professor of Mobile Communication and Digital Broadcasting Engineering with the University of Science and Technology (UST). He was with the Communications Research Centre Canada (CRC) from 2003 to 2004, as a research scientist. His main research interests are 3D-TV broadcasting systems for realistic 3D audio-visual services.

Jinwoong Kim received the BS and MS degrees in electronics engineering both from Seoul National University, Seoul, Korea, in 1981 and 1983. He received the PhD degree in electrical engineering from Texas A&M University, Texas, USA, in 1993. He has been with ETRI since 1983 and is now a principal member of research staff with the Broadcasting and Telecommunications Media Research Division. He has been involved

in many big projects in the telecommunications and broadcasting area, such as development of TDX-1 and TDX-10 digital telephone switching systems, an HDTV video encoding chipset, and a real-time encoder system. As the director of the Broadcasting Media Research Department, he carried out several projects on new digital broadcasting technologies, including data broadcasting, customized broadcasting, and content protection. He also led projects on MPEG-7 and MPEG-21 technology development, resulting in several important contributions to the MPEG standards. He is currently a 3DTV project leader, focusing on 3D DMB and multiview 3DTV system. He was a Far-East Liaison of the 3DTV Conference 2007 and 2008. He is the vice president of the Association of Realistic Media Industry (ARMI) of Korea.