The important requirements for stereo video retargeting are threefold: keeping temporal coherence, preventing depth distortion, and minimizing shape distortions of the retargeted video. To meet these requirements, the left and right video sequences are divided into groups of frames (GoFs), where the GoF is a basic unit for the seam carving and we assign a set of fixed seams for all frames within the GoF. To determine the fixed seams for each GoF, we need to find the GoF boundary in the video first. Then, the representative frame for each GoF is generated by considering the spatial saliency and temporal coherence. Also, the confidence of the stereoscopic correspondence between the left and right frames is considered to prevent depth distortion.

**Keywords:** Stereo video, stereo video retargeting, stereo video resizing, seam carving.

I. Introduction

With the growing popularity of 3D contents, the demand for displaying the stereoscopic 3D contents on various display devices is also increasing. As such, resizing of the original stereo videos to fit the target display device is necessary. Unlike the retargeting problems for still and 2D video data, however, one must consider the depth information as well as the spatial and temporal coherence for the stereo video retargeting, which makes the problem more challenging.

The first approach for the stereo image retargeting is to directly apply the seam carving approach [1]-[3] of 2D monoscopic images. For example, the 2D seams are determined for one view of the stereo pairs and then their corresponding stereo seams are found in the opposite view [4], [5]. Detecting corresponding seams in the opposite view is based on constraint energies of depth information (that is, the disparity) between the left and right images as well as the spatial coherence. Also, due to the occlusion that exists in the stereo images, discontinuous seams may be allowed [5]. In general, these methods are quite effective for stereoscopic still images. However, when these methods are applied directly for each stereoscopic pair of stereo videos, visually annoying artifacts, such as temporal jittering or flickering, are observed. Therefore, for the stereo video retargeting, we need to consider the temporal coherence for each stereoscopic pair of video sequences as well as the depth constraint. The importance of the temporal coherence was stressed in [6], the authors saying, “Even a mild inconsistency among adjacent frames would lead to the annoying flickering or jittering.” Similar artifacts can be observed in [7]. That is, the cropping-based retargeting method of [7] may create a temporal jittering problem because the cropped regions may vary from frame to frame.
When the seam carving method for 2D still images was directly applied for the retargeting of 2D videos in [2], visually annoying temporal jittering or flickering artifacts were observed. To alleviate this problem in [3], seams for videos were allowed to be discontinuous and a seam in the current frame was determined based on the reference seams in the previous frame. Also, the optimal balance between the spatial and temporal saliencies is required for video retargeting [8], [9]. In [8], to ensure sufficient temporal coherence, the motion information was adopted. In [9], the global scaling map derived from the panoramic mosaic was formed and was used to calculate a local scaling map for each frame. These local maps were further refined according to spatial coherence constraints, so the coherence between consecutive frames was ensured and spatial distortions were avoided. In [10], a seam carving based on matching area was released. This method has greater flexibility than that of [3] because the pixels in the current frame are divided into reward/punish regions according to their similarities to the pixels on the seam of the previous frame. Note that the aforementioned video retargeting methods are basically for monoscopic videos. So, if they are applied to stereo videos by treating left and right sequence separately as independent 2D videos, then depth distortions may occur in the retargeted stereo videos.

Our approach to alleviate spatial, temporal, and stereoscopic depth distortion is to assign the same seams for a set of consecutive stereo frames. That is, the left and right stereo video sequences are divided into non-overlapping groups of frames (GoFs) and a pair of representative frames is constructed for each stereoscopic pair of left and right GoFs. Then, the seams for the representative left and right frames are determined by considering the spatial, temporal, and disparity coherence of the representative frames. Since we fix the seams for a finite duration of the video, the resized video is temporally coherent without the flickering artifact. Also, since the coherent stereoscopic disparities between the left and right seams are maintained, the depth distortion is alleviated. In summary, our GoF-based stereo retargeting method consists of three steps: (i) detection of GoF boundaries in the stereo video, (ii) determination of representative frames for a stereoscopic pair of GoFs, and (iii) seam removal for all frames in the GoF.

The rest of this paper is organized as follows. In section II, we explain how to detect GoF boundaries for the stereo video sequence. Then, in section III, a representative frame is formed, and seams for the representative frame are determined. The energy criterion for the representative seams are addressed for vertical and horizontal resizing. We provide experiment results in section IV and conclude the paper in section V.

II. GoF Boundary Detection

The basic idea of our approach is to fix seams for a subset of consecutive frames, namely, a GoF. So, the entire video sequence is divided into sets of non-overlapping GoFs, and we seek an optimal fixed seam for each GoF. Since we are dealing with stereo videos, we have a stereoscopic pair of left and right video sequences. Both of the stereoscopic sequences are divided into the same number of GoFs, and the GoFs of the stereo pair are synchronized with each other. Therefore, each GoF has a corresponding GoF at the opposite view. In this paper, we first divide one view (left or right) of the stereo sequence into GoFs, then the GoFs of the opposite view follow suit. The cost functions to detect the GoF boundaries are based on [11] but are modified for stereo data.

The determination of the GoF boundary can be done by measuring the change in energy when including the frame following the current GoF. Here, the energy cost includes not only the saliency of the left frame, as in the single video of [11], but also the stereoscopic confidence between the left and right frames by measuring the maximum amount of accumulated change in energy in the current GoF. After that, the spatial complexity of the maximum amount of accumulated change in energy is used to determine the boundary of the GoF.

The input of our method is an original stereo video $I$ that includes two sequences: left sequence $I_L$ and right sequence $I_R$. Each sequence consists of $N$ frames with $H \times W$ pixels for each frame, where $I_L = \{I_L^n(i,j) | n = 1,...,N, 1 \leq i \leq H, 1 \leq j \leq W\}$ and $I_R = \{I_R^n(i,j) | n = 1,...,N, 1 \leq i \leq H, 1 \leq j \leq W\}$. The basic unit of our video retargeting is the GoF and the entire stereo video sequence is divided into non-overlapping GoFs, as shown in Fig. 1.

To detect the GoF boundary, we need to calculate a disparity map, $D'$, where $D'$ can be calculated by any stereo matching algorithm (for example, [12]). Then, the confidence of correspondence between these two frames, which is called stereoscopic confidence map $S^{\gamma}_{i,j}$, can be determined based on the sum of absolute differences within a window, $\psi$, of a pixel $(i,j)$:

$$S^{\gamma}_{i,j} = \sum_{(x,y) \in \psi} |D'(i,x) - D'(i,y)|$$

Fig. 1. Structure of left and right sequences with GoFs.
where \( [\psi] \) denotes the number of pixels in window \( \psi \) and \( D^g(i, j) \) is the disparity value at pixel \((i, j)\). In stereo confi- dence map \( S^c \) of (1), \( S^c \) having a small value means the level of confidence matching between the left and right images is high. On the other hand, a large \( S^c \) value implies poor matching.

To measure temporal coherence for each sequence, we need a gradient map. Gradient map \( S^g_{/L} \) of frame \( I^k_{/L} \) is defined as follows:

\[
S^g_{/L}(i, j) = \frac{1}{[\psi]} \sum_{(i, j) \in \psi} \left[ | I^k_{/L}(i+u, j+v) - I^k_{/L}(i+u, j+D^g(i, j) + v) | \right],
\]

(2)

where the gradient magnitude is calculated by using the convolution masks \([-1, 1]\) and \([1, -1]^T\) for horizontal and vertical directions, respectively. Then, temporal salience cost, \( S^s_{/L} \), of frame \( I^k_{/L} \) can be determined by temporal gradient changes, as follows:

\[
S^s_{/L}(i, j) = \left| S^g_{/L}(i, j) - S^g_{/L}(i, j+1) \right| + \left| S^g_{/L}(i, j) - S^g_{/L}(i, j-1) \right|,
\]

(3)

According to (3), the higher temporal salience cost, \( S^s_{/L} \), implies the higher motion activities of the neighboring frames at that pixel.

The GoF boundary is determined by the accumulated cost of the sequence of consecutive frames. That is, for the \( k \)-th GoF, \( G^k_{/L} \), starting at frame number \( m_0 \), we sequentially check the frames that follow to determine the GoF boundary. So, if we already examine \( M-2 \) frames from \( m_0 \) without finding the GoF boundary, then \( G^k_{/L} \) currently has \( M-1 \) frames, including the first frame at \( m_0 \) and the next frame to check is located at \( m_0+M-1 \). The decision to include the frame at \( m_0+M-1 \) in \( G^k_{/L} \) or not is determined by the spatial smoothness of the pixel values with the maximum temporal and disparity complexity among the pixels in the same location of all frames from \( m_0 \) to the current frame, \( m_0+M-1 \). Therefore, the spatial smoothness of the pixel values with the maximum temporal and depth activities serves the criterion for the GoF boundary. To measure the smoothness of the temporal and depth activities, we first define the current representative frame, \( I^{m_0}_{/L} \), from \( m_0 \) to \( m_0+M-1 \), as follows:

\[
I^{m_0}_{/L}(i, j) = \begin{cases} I^{m_0}_{/L}(i, j) & \text{if } L \leq i \leq n \text{ and } 0 \leq j \leq W, \hfill \\
I^{m_0}_{/L}(i, j) + \alpha S^s_{/L}(i, j) & \text{if } L \leq i \leq n \text{ and } 0 < j < W, \hfill \\
I^{m_0}_{/L}(i, j) - \alpha S^s_{/L}(i, j) & \text{if } L \leq i \leq n \text{ and } W < j \leq W. \hfill \\
I^{m_0}_{/L}(i, j) & \text{otherwise}. \hfill 
\end{cases}
\]

(4)

where weighting factors \( \alpha_1 \) and \( \alpha_2 \) satisfy \( \alpha_1+\alpha_2=1 \). According to (4), for each pixel location, we select the pixel in the frame of \( n^* \) with the maximum value of the linear combination of temporal gradient \( S^s_{/L} \) and amount of stereoscopic disparity \( S^d_{/L} \) for all frames in the GoF (see Fig. 2). Then, our representative frame is formed through a pixel-by-pixel selection of maximum temporal and depth activities (pixel-wise mosaicking) from any of the frames in the GoF, and the spatial smoothness of it serves as a criterion for the GoF boundary. Now, the decision of including the \((m_0+M-1)\)th frame in \( G^k_{/L} \) or not is made by the spatial continuity of \( I^{m_0}_{/L} \) and \( C^{m_0}_{/L} \), defined as follows:

\[
C^{m_0}_{/L} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \left| I^{m_0}_{/L}(i,j) - I^{m_0}_{/L}(i,j+1) \right| + \left| I^{m_0}_{/L}(i,j) - I^{m_0}_{/L}(i+1,j) \right|,
\]

(5)

If \( C^{m_0}_{/L} < \tau \), then the \((m_0+M-1)\)th frame is included in \( G^k_{/L} \) and we proceed to the next frame to check \( C^{m_0}_{/L} \) for the inclusion of the next \((m_0+M)\)th frame. Otherwise, we complete the \( k \)-th GoF, \( G^k_{/L} \), with \( M \) consecutive frames, starting from \( m_0 \). As a result, the left sequence is divided into \( K \) GoFs, \( \{ G^k_{/L} \mid k = 1, \ldots, K \} \), where \( \sum_{k=1}^{K} | G^k_{/L} | = N \). After determining the GoFs for the left sequence of the stereoscopic video, the same GoF boundaries are applied to the right sequence to have \( K \) corresponding right GoFs, \( \{ G^k_{/R} \mid k = 1, \ldots, K \} \).

### III. Representative Seams for GoF

In this section, we present how to determine a pair of left and right representative seams for each stereoscopic pair of GoFs.
We first detect the left representative seam for the left GoF. Then, based on the disparity maps of all frames in the left GoF, we can calculate the horizontal disparity between the left and right representative seams. Finally, the right representative seam of the corresponding right GoF is determined by using the calculated disparities. Specifically, with each pixel in the representative frame, \( I_{\text{left}}^{k/M-1} \), of the \( k \)-th left GoF, \( G_{L}^{k} \), we can calculate the cost of removing that pixel for retargeting. The cost includes four components: saliency cost \( S_{\text{max}}^{k} \), confidence cost \( S_{\text{p}/\text{GoF}}^{k} \), spatial coherence cost \( S_{\text{v}/\text{GoF}}^{k} \), and temporal coherence cost \( T_{\text{GoF}/L}^{k} \). The saliency cost, spatial cost, and temporal cost are based on those in [3]. However, the basic difference between the costs in [3] and the costs in our study lies in the coverage for the computation of the costs, that is, the frame-based costs in [3] and the GoF-based costs in our study.

The saliency cost of the left GoF, \( S_{\text{max}}^{k} \), represents the most significant temporal gradient changes for each pixel of the consecutive frames in \( G_{L}^{k} \), as follows:

\[
S_{\text{max}}^{k}(i, j) = \arg \max_{n=0, \ldots, n_{\text{max}}=n_{\text{min}}+1} \left\{ \alpha_i S_{\text{h}}^{k}(i, j) + \alpha_j S_{\text{v}}^{k}(i, j), \right. \quad 1 \leq i \leq H, \quad 1 \leq j \leq W \tag{6}
\]

Note that (4) and (6) are based on the same criterion, but the outputs are different: the representative image with gray levels in (4) and the representative gradient map in (6). On the other hand, the output of the confidence cost, \( S_{\text{p}/\text{GoF}}^{k} \), is formed by the confidence of the stereoscopic correspondence, as follows:

\[
S_{\text{p}/\text{GoF}}^{k} = \left\{ S_{\text{h}}^{k}(i, j) \mid n = \arg \max_{n=0, \ldots, n_{\text{max}}=n_{\text{min}}+1} \left\{ \alpha_i S_{\text{h}}^{k}(i, j) + \alpha_j S_{\text{v}}^{k}(i, j), \right. \quad 1 \leq i \leq H, \quad 1 \leq j \leq W \tag{7}
\]

The temporal coherence cost, \( T_{\text{GoF}/L}^{k} \), and the spatial coherence cost, \( S_{\text{v}/\text{L}}^{k} \), take different formulas for vertical and horizontal resizing. So, we will present these two costs separately in the following subsections.

1. Vertical Resizing

In the case of vertical resizing, the spatial coherence cost, \( S_{\text{v}/L}^{k} \), represents the loss of the spatial continuity by removing a pixel in the representative frame, \( I_{\text{left}}^{k/M-1} \), where \( S_{\text{v}/L}^{k} = S_{\text{h}/L}^{k} + S_{\text{v}/L}^{k} \), combining the gradient changes in horizontal \( S_{\text{h}/L}^{k} \) and vertical \( S_{\text{v}/L}^{k} \) directions. The horizontal gradient change for removing a pixel \((i, j)\) in \( I_{\text{left}}^{k/M-1} \) is determined as follows:

\[
S_{\text{h}/L}^{k}(i, j) = \begin{cases}
I_{\text{left}}^{k/M-1}(i, j - 1) - I_{\text{left}}^{k/M-1}(i, j), & (i, j) \in \Omega_a, \\
I_{\text{left}}^{k/M-1}(i, j + 1) - I_{\text{left}}^{k/M-1}(i, j), & (i, j) \in \Omega_b,
\end{cases}
\]

\[
S_{\text{v}/L}^{k}(i, j) = \begin{cases}
I_{\text{left}}^{k/M-1}(i - 1, j) - I_{\text{left}}^{k/M-1}(i, j - 1), & (i, j) \in \Omega_a, \\
I_{\text{left}}^{k/M-1}(i - 1, j + 1) - I_{\text{left}}^{k/M-1}(i, j - 1), & (i, j) \in \Omega_b,
\end{cases}
\]

where \( \Omega_a \) and \( \Omega_b \) denote the interior and border pixels, respectively. For the vertical gradient changes, we consider three cases that depend on the relative pixel locations in the previous row and the current one. That is, the vertical gradient \( S_{\text{v}/L}^{k} \) must be differentiated according to the three cases (that is, \( S_{\text{v}/L}^{k}, S_{\text{v}/R}^{k}, \) and \( S_{\text{v}/L}^{k} \), where \( S_{\text{v}/L}^{k}, S_{\text{v}/R}^{k}, \) and \( S_{\text{v}/L}^{k} \) represent the gradient changes for upper pixel removal, left pixel removal, and right pixel removal, respectively. For the upper pixel removal, we have \( S_{\text{v}/L}^{k} = 0 \). For \( S_{\text{v}/L}^{k} \) and \( S_{\text{v}/R}^{k} \), we have

\[
S_{\text{v}/L}^{k}(i, j) = \begin{cases}
I_{\text{left}}^{k/M-1}(i - 1, j) - I_{\text{left}}^{k/M-1}(i, j - 1), & (i, j) \in \Omega_a, \\
I_{\text{left}}^{k/M-1}(i - 1, j - 1) - I_{\text{left}}^{k/M-1}(i, j - 1), & (i, j) \in \Omega_b,
\end{cases}
\]

and

\[
S_{\text{v}/R}^{k}(i, j) = \begin{cases}
I_{\text{left}}^{k/M-1}(i, j + 1) - I_{\text{left}}^{k/M-1}(i, j), & (i, j) \in \Omega_a, \\
I_{\text{left}}^{k/M-1}(i, j + 1) - I_{\text{left}}^{k/M-1}(i, j - 1), & (i, j) \in \Omega_b,
\end{cases}
\]

The temporal coherence cost \( T_{\text{GoF}/L}^{k} \) is similar to that of [3], where the seam of the previous frame is used to measure the temporal coherence between the consecutive frames. For our GoF-based approach, instead of the previous seam, we use the seam of the previous GoF, \( G_{L}^{k-1} \), to measure the coherence between the seams of the consecutive GoFs. For more details on the temporal costs as well as the spatial coherence, refer to [3].

Now, we are ready to calculate the cost of removing a pixel in the representative frame of a GoF. For the first GoF, \( G_{L}^{1} \), the initial combined cost, \( M_{\text{0}/L}^{k} \), is a linear combination of \( S_{\text{max}/L}^{k}, S_{\text{p}/L}^{k}, \) and \( S_{\text{p}/\text{GoF}}^{k} \):

\[
M_{\text{0}/L}^{k} = \alpha_i S_{\text{max}/L}^{k} + \alpha_j S_{\text{p}/L}^{k} + S_{\text{p}/\text{GoF}}^{k}.
\]

From the second GoF of \( G_{L}^{k} \) with \( k=1 \), since the previous
GoF is available, we can include the temporal coherence, $T_{\text{GoF}}^k$, as well, and the initial combined cost for the first frame of the GoF is defined as follows:

$$M_{0/L}^k = \beta_i S_{\text{L}}^k + \beta_2 (S_i^k + S_{\text{GoF}}^k) + \beta_3 T_{\text{GoF}}^k,$$

(12)

where $k>1$ and $\beta_i+\beta_2+\beta_3=1$.

From the initial combined costs of (11) and (12), we can calculate the seams for the representative frame by using dynamic programming similar to that in [1]. To do that, we need a cumulative minimum energy matrix, $M_{L/L}^k$, for seam tracking in each representative frame, $I_{m_i/M-1}^k$. Matrix $M_{L/L}^k$ starts with $M_{L/L}^k(1,j) = M_{0/L}^k(1,j)$ for all $j=1, \ldots, W$. For $M_{L/L}^k(i,j)$ with $i>1$, we have

$$M_{L/L}^k(i,j) = \min \left\{ M_{L/L}^k(i-1,j-1) + \gamma \times S_{\text{L}}^k(i,j), \right.$$  
$$M_{L/L}^k(i-1,j) + \min \left\{ M_{L/L}^k(i-1,j-1), \right.$$  
$$M_{L/L}^k(i-1,j+1) + \gamma \times S_{\text{L}}^k(i,j) \right\},$$

(13)

where

$$\gamma = \begin{cases} \alpha_2, & k = 1, \\ \beta_2, & k > 1. \end{cases}$$

(14)

Then, by using a simple backtracking process from the last to the first row of $M_{L/L}^k$, we can find an optimal seam for the representative frame of each GoF. The resulting optimal seam is then applied for all frames in $G_{L}^k$ for retargeting.

Since our approach is to determine the left representative seams first for the stereoscopic GoFs, the issue raised is how to find the representative seam of the corresponding right GoF, given the representative seam of the left. To this end, for the left representative seam, $L_{L}^k = \{p_{L}^k, i = 1, \ldots, H\}$, $p_{L}^k = (i, j_0(i))$, of the $k$-th left GoF with consecutive frames starting from $m_0$, let us define the horizontal disparity between the left and right representative seams as follows:

$$d^k = \arg \max_{d} \left\{ \left\{ p_{L}^k \mid D^k(p_{L}^k) \right\} \right\},$$

(15)

where we first calculate the disparity population over $G_{L}^k$, and the disparity that appears most frequently (that is, the maximum count of the pixel with the disparity $d$) is then chosen for the final horizontal disparity. This disparity is then used to calculate the representative seam of the right GoF, and we have $L_{R}^k = \{p_{R}^k, i = 1, \ldots, H\}$, $p_{R}^k = (i, j_0(i) + d^k)$ (see Fig. 3(a)). This seam is then applied to all frames in $G_{R}^k$ for retargeting.

2. Horizontal Resizing

In the case of horizontal resizing, the spatial coherence cost, $S_{\text{H}}^k$, of removing a pixel in the representative frame, $I_{m_i/M-1}^k$, is also the combination of the gradient changes in horizontal ($S_{hL}^k$) and in vertical ($S_{vL}^k$, $S_{vR}^k$) directions as $S_{p}^k = S_{hL}^k + S_{vL}^k$. However, the vertical gradient change for removing the pixel $(i, j)$ in $I_{m_i/M-1}^k$ is different and is determined as follows:

$$S_{L}^k(i,j) = \begin{cases} I_{kL}^M(i-1,j) - I_{kL}^M(i,j), & (i, j) \in \Omega_v, \\ -I_{kL}^M(i,j) - I_{kL}^M(i+1,j), & (i, j) \in \Omega_v, \\ I_{kL}^M(i+1,j) - I_{kL}^M(i,j), & (i, j) \in \Omega_v. \end{cases}$$

(16)

The horizontal gradient change, $S_{hL}^k$, has three different cases (that is, $S_{hL}^k$, $S_{hU}^k$, and $S_{hD}^k$), where $S_{hL}^k$, $S_{hU}^k$, and $S_{hD}^k$ represent the gradient changes for left pixel removal, left-upper pixel removal, and left-lower pixel removal, respectively. For the left pixel removal, we have $S_{hL}^k = 0$. For $S_{hU}^k$ and $S_{hD}^k$, we have

$$S_{hU}^k(i,j) = \begin{cases} I_{kL}^M(i-1,j) - I_{kL}^M(i+1,j), & (i, j) \in \Omega_v, \\ -I_{kL}^M(i,j) - I_{kL}^M(i-1,j), & (i, j) \in \Omega_v, \\ I_{kL}^M(i,j) - I_{kL}^M(i-1,j), & (i, j) \in \Omega_v. \end{cases}$$

(17)

and

$$S_{hD}^k(i,j) = \begin{cases} I_{kL}^M(i+1,j) - I_{kL}^M(i,j), & (i, j) \in \Omega_v, \\ -I_{kL}^M(i,j) - I_{kL}^M(i+1,j), & (i, j) \in \Omega_v, \\ I_{kL}^M(i,j) - I_{kL}^M(i+1,j), & (i, j) \in \Omega_v. \end{cases}$$

(18)

Again, the temporal coherence cost, $T_{\text{GoF}}^k$, is similar to that of [3] except that the GoF is used instead of the frame.
Therefore, it is similar to the vertical resizing, but, this time, we use the involved columns instead of the involved rows to calculate the seam of the current GoF. For more details regarding involved columns and rows, refer to [3].

Now, we have the initial combined cost, \( M_{0/L}^k \), with cases \( k=1 (19) \) and \( k>1 (20) \), as follows:

\[
M_{0/L}^k = \alpha_1 S_{\text{max/L}}^k + \alpha_2 (S_{\text{r/L}}^k + S_{\text{G/L}}^k),
\]

\[
M_{0/L}^k = \beta_1 S_{\text{max/L}}^k + \beta_2 (S_{\text{r/L}}^k + S_{\text{G/L}}^k) + \beta_3 T_{\text{G/L}}^k.
\]

The cumulative minimum energy matrix, \( M_L^k \), for seam tracking in each representative frame, \( T_{0/L}^k = \frac{M_{0/L}^k}{M-1} \), is formed with \( M_L^k(i,j) = M_{0/L}^k(i,1) \) for all \( i=1, \ldots, H \) and for \( M_L^k(i,j), \ j>1 \), and we have the following dynamic programming:

\[\begin{align*}
M_L^k(i,j) &= M_L^k(i,j-1) + \gamma \times S_{\text{M/L}}^k(i,j), \\
M_L^k(i,j) &= \min \left\{ M_L^k(i,j-1), \\
M_L^k(i+1,j-1) + \gamma \times S_{\text{D/L}}^k(i,j) \right\}.
\end{align*}\]

So, the representative seam of each left GoF can be detected by using a simple backtracking process from the last to the first column of \( M_L^k \). Now, the corresponding right representative seam can be determined from the given left representative seam, \( L_L^k = \{p_L^k, j=1, \ldots, W\} \), \( p_L^k = (t_L^k(j), j) \), with \( M \) consecutive frames, starting from \( m_0 \). To this end, we first determine the actual disparity, \( d^* \), as follows:

\[
d^* = \arg\max_d \left\{ \left\{ |p_L^k| D'(p_L^k) = d, \right\} \right\},
\]

Finally, we have a pair of left and right representative seams for the corresponding left and right GoFs: \( L_L^k = \{p_L^k, j=1, \ldots, W\} \) with \( p_L^k = (t_L^k(j), j) \) and \( L_R^k = \{p_R^k, j=1, \ldots, W\} \) with \( p_R^k = (t_R^k(j), j+d^*) \) (see Fig. 3(b)).

IV. Experiment Results

For all of our experiments, we use a 5\times5 window, \( \psi \), for (1). According to the suggested ratio of saliency to spatial coherence to temporal coherence, \( \beta_1: \beta_2: \beta_3 \) equals 2.5:0.2, as in [3], we fix the weighting factors to \( \alpha_1 = 0.3, \alpha_2 = 0.7 \) for (11) and (19), and \( \beta_1 = 0.28, \beta_2 = 0.69, \beta_3 = 0.03 \) for (12) and (20). On the normalized representative frame, threshold value \( \tau \) is determined by experiments as \( \tau = 0.12 \). All test stereo videos are from the FhG-HHI database (http://sp.cs.tut.fi/mobile3dtv/stereo-video/).

First, we show the representative frames of the GoF formed by (4). In Fig. 4(a), the “Pole” test video with 44 frames is divided into four GoFs with \( \tau = 0.06 \), and their corresponding representative frames are shown. If we increase the threshold to \( \tau = 0.08 \), then all 44 frames will be combined into a single GoF, as in Fig. 4(b). As one can see in the representative frame of

![Fig. 4. GoFs and representative frames: (a) four GoFs and their representative frames for \( \tau = 0.06 \), (b) one GoF and its representative frame for \( \tau = 0.08 \), (c) representative seam for left GoF of (b), and (d) representative seam for right GoF of (b).](image-url)
Fig. 5. 1st column: original anaglyph, 2nd column: retargeted anaglyph by [5], 3rd column: seams for left frame by [5], 4th column: seams for right frame by [5], 5th column: retargeted anaglyph by our method, 6th column: seams for left frame by our method, and 7th column: seams for right frame by our method.

Fig. 6. Comparison of vertical alignment after retargeting for (a) “Pole” and (b) “Stone”: 1st column: original anaglyph; 2nd column: retargeted anaglyph by [5]; and 3rd column: retargeted anaglyph by our method.

Fig. 4(b), the salient region of interest (ROI) of the representative frame around the tip of the sharpening wood has a high spatial contrast. Based on this representative frame, as well as the confidence map, the seams for the representative frame of the left GoF and the right GoF avoid the ROIs shown in Figs. 4(c) and 4(d). Once we have the seams for the left GoF, their corresponding right seams are determined by (15). These representative seams are applied to all frames in the GoF to yield temporally coherent retargeted video without any visual artifacts, such as flickering or jittering. Similarly, in Fig. 5, by considering the motion activities of the man, the man on the right side is considered salient and remains intact after the retargeting in our method.

The temporal variations due to the misalignment of objects in the retargeted video sequence cause flickering or jittering artifacts. This is demonstrated in Fig. 6. The figure clearly shows that the objects are misaligned in the previous method [5], resulting in temporal jittering artifacts. However, our
method has steady seam removals in the temporal domain, yielding few temporal jittering artifacts. In fact, the temporal jittering can be quantitatively measured by calculating the average seam displacement between the seams of two consecutive frames [11]. The average seam displacements for our method and those of [3] and [5] are quantitatively compared in Table 1. As shown in the table, our method has the least seam deviations. Note that the average seam displacement depends on threshold \( \tau \) for the proposed method. Fixing the value of the threshold, the length of the GoF is inversely proportional to the complexities of the video content. If the transition between two consecutive GoFs is visually noticeable and causes flickering, then the threshold should be lowered, shortening the lengths of the GoFs.

Another important requirement for a stereo video retargeting problem is to prevent depth distortion. In our approach, by using the stereoscopic confidence in left-right correspondence and the majority voting for disparity throughout the GoF in (15), the depth distortion is prevented in our video retargeting results (compare the 3D effects at the arched windows in Fig. 7). As one can see in Figs. 7 and 8, the spatial distortions

Table 1. Average seam displacements of consecutive frames.

<table>
<thead>
<tr>
<th></th>
<th>Pole</th>
<th>Stone</th>
<th>Rope jump</th>
<th>Alt moabit</th>
<th>Book arrival</th>
<th>Butterfly</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau = 0.06 )</td>
<td>0.96</td>
<td>0.11</td>
<td>2.83</td>
<td>3.91</td>
<td>0.58</td>
<td>0.89</td>
</tr>
<tr>
<td>( \tau = 0.08 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.64</td>
<td>0.88</td>
<td>0.00</td>
<td>0.52</td>
</tr>
<tr>
<td>( \tau = 0.10 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
<td>0.49</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \tau = 0.12 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>[3]</td>
<td>4.12</td>
<td>1.03</td>
<td>6.41</td>
<td>8.95</td>
<td>6.45</td>
<td>3.34</td>
</tr>
<tr>
<td>[5]</td>
<td>100.62</td>
<td>30.09</td>
<td>69.24</td>
<td>98.12</td>
<td>101.52</td>
<td>97.47</td>
</tr>
</tbody>
</table>

Fig. 7. Depth distortions in stereo video retargeting: 1st column: original anaglyph; 2nd column: retargeted anaglyph by [5]; and 3rd column: retargeted anaglyph by our method.
and the depth distortions affect each other.

V. Conclusion

We proposed a retargeting method for stereo video. Since the seam carving for still images was successfully developed, we decomposed a stereo video sequence into groups of frames (GoFs) and formed a representative frame for each GoF to exploit the existing 2D seam carving of still images. By combining the GoF into a representative 2D image we found seams by treating the representative frame as a still image. Then, by applying the seams of the representative frame to all frames within the GoF, we avoided the visually annoying flickering and jittering artifacts. For a stereo video, the seams of the representative frame for one view (say, the left video) can be used for the GoF of the other view (say, the right video) by determining the horizontal disparities between the representative frames of the left and right GoFs. Experiment results confirmed that our GoF-based retargeting method for a stereo video can effectively avoid the spatial, temporal, and stereoscopic depth artifacts.

References


Hai Thanh Nguyen received her B.S. degree in 2010 from the Department of Telecommunication Systems, Faculty of Electronics and Telecommunications, Hanoi University of Science and Technology (HUST), Hanoi, Vietnam. She is currently working toward her Ph.D degree at Dongguk University Division of Electronics and Electrical Engineering, Seoul, Korea. Her research interests include image and video retargeting, stereoscopic 3D images, and video signal processing.

Chee Sun Won received his B.S. degree in electronics engineering from Korea University, Seoul, Rep. of Korea, in 1982, and his M.S. and Ph.D. degrees in electrical and computer engineering from the University of Massachusetts, Amherst, MA, USA, in 1986 and 1990, respectively. From 1989 to 1992, he was a senior engineer with GoldStar Co., Ltd. (LG Electronics), Seoul, Rep. of Korea. In 1992, he joined Dongguk University, Seoul, Rep. of Korea, where he is currently a professor in the Division of Electronics and Electrical Engineering. He was a visiting professor at Stanford University, Stanford, CA, USA, and at McMaster University, Hamilton, ON, Canada. His research interests include MRF image modeling, image segmentation, robot vision, image resizing, stereoscopic 3D video signal processing, and image watermarking.