IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 13, 2019, accepted August 22, 2019, date of publication September 2, 2019,
date of current version September 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938865

SmartGrid: Video Retargeting With Spatiotemporal Grid Optimization

HO SUB LEE[1], (Student Member, IEEE), GYUJIN BAE[2], (Student Member, IEEE), SUNG IN CHO[3], (Member, IEEE), YOUNG HWAN KIM[1], (Senior Member, IEEE), AND SEOKHYEONG KANG[1], (Member, IEEE)

[1]Department of Electrical Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea
[2]LCD Division, LG Display, Seoul 07796, South Korea
[3]Department of Multimedia Engineering, Dongguk University, Seoul 04620, South Korea

Corresponding author: Seokhyeong Kang (shkang@postech.ac.kr)

This work was supported in part by the LG Display company Ltd., in part by the IDEC, and in part by the Ministry of Science and ICT (MSIT), South Korea, through the ICT Consilience Creative program under Grant ITP-2019-2011-1-00783 supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP).

ABSTRACT We propose SmartGrid, a new video-retargeting method that preserves a shape of the salient object and maintains a temporal coherence for the static background regions in the video. Previous methods mainly focus on preserving the shape of the salient object or maintaining the temporal coherence for each region. However, these methods do not explicitly adjust the sizes of the grids corresponding to the salient objects and the static background regions. Thus, they have difficulty in maintaining the consistency of the salient object shape and the temporal coherence for the static background regions. The basic idea of SmartGrid is to maintain the consistency of the contents in consecutive frames by analyzing the degree of the spatiotemporal consistency. Compared to the best results obtained using eight previous methods, SmartGrid achieved improvements of 1.19× in Bidirectional Similarity Measure, 7.59× in Jittery Metric 1, and 13.16× in Jittery Metric 2, and reduced the average computation time per pixel by 6.14×.

INDEX TERMS Temporal coherence, video retargeting, deformation, computational complexity.

I. INTRODUCTION

Recently, many mobile devices such as smartphones, tablets, and portable game consoles with different resolutions and aspect ratios are widely used. With the advancement of mobile devices and display technologies, there is a growing need to reproduce video contents on a display platform with various aspect ratios. In addition, sharing of video contents through mobile devices becomes popular. Therefore, it is necessary to adjust the video contents to various display platforms. Video retargeting is a technique that adjusts the size of an image to the aspect ratio and resolution of an output device. Video retargeting can be used for devices with various display platforms such as televisions, tablets, personal computers, and smartphones. Movies generally have different aspect ratio that of device displays, so video retargeting is an essential process to adapt movies to various display platforms.

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqi Wang.

Standard video retargeting methods (Fig. 1), such as uniform scaling, cropping, and inserting letter boxes have obvious drawbacks. The uniform scaling method resizes the image and keeps the same scaling factor for all the pixels to fit the desired aspect ratio; compared to the original image (Fig. 1a), this method can stretch or shrink a salient object in the image (Fig. 1b), and this artifact can be noticeable to the viewer. The cropping method removes unnecessary surrounding regions from the given image to fit the desired aspect ratio; this method can remove salient objects or region (Fig. 1c). The letter box insertion method inserts black areas into the upper and lower regions of the image to fit the desired aspect ratio; this method wastes display area (Fig 1d). Various content-aware video retargeting methods have been developed [1]–[29] to overcome the limitations of standard methods. Content-aware video retargeting methods deform regions that the human visual system regards as relatively unimportant [1], so these methods minimize distortion of important regions. A key insight of the methods is to adjust the size of the image by analyzing the visual importance of

(a) (b) (c) (d)

FIGURE 1. Examples of the image retargeting from 16:9 to 4:3: (a) original image, (b) linear scaling method, (c) cropping method, and (d) letter box insertion method. ©Love Dis KBS.

the pixels. The goal of content-aware video retargeting is change the aspect ratio of the image while preserving the salient object, and maintaining the temporal coherence of video contents. The preservation of the salient object is to keep the important video contents and regions which are most attractive to the viewers. Maintenance of the temporal coherence is to keep the smooth and consistent motion of the objects or background regions in the temporal domain. The content-aware video retargeting methods can be grouped into three categories: pixel-based, region-based, and object-based retargeting methods. We will introduce the details of these methods in Section II.

Generally, content-aware video retargeting consists of two steps. In the first step, the retargeting method generates a saliency map that represents the region of salient object roughly in the image by analyzing the contents to detect the visually important regions of the given image. Most previous methods use the edge magnitude of pixels, face/object detection, and motion information in consecutive frames to quantify the visual importance of the video contents. In the second step, the retargeting method uses the saliency map to deform relatively unimportant regions or pixels in the image.

The simplest content-aware video retargeting method performs retargeting independently for each frame [1]. However, this independent retargeting can cause a temporally unstable result for objects or background regions in the video frames. The result can yield visually-annoying artifacts such as stretching or shrinking of a moving object when it exists in consecutive frames. This artifact occurs because this method does not consider the temporal correlation between consecutive frames.

To ensure the maintenance of temporal coherence, several video retargeting methods perform retargeting by considering the relationship between the current frame and neighboring frames [2]–[6]. However, these methods do not use temporal information for the entire region of the image to maintain the temporal coherence. Results for a moving object can be temporally unstable when a moving object occurs beyond the range of the temporal information. To maintain the temporal coherence, other video retargeting methods use information extracted from the entire video frame [7]–[11]. Methods in [9]–[11] divide each frame into grids, representing the structure of the squares made from horizontal and vertical lines crossing each other and automatically align the grids by using an image registration process between consecutive frames. Image registration aligns each video frame by estimating the object motion or a camera motion between consecutive frames. Based on the alignments, these methods formulate an optimization problem over all frames. Then the aligned grid is coherently deformed by finding an optimal solution; this process can maintain the temporal coherence of the entire video frame.

However, previous methods have two problems. First, the results of the previous methods show that the shape of the salient object is not consistently maintained in consecutive frames. Second, they have difficulty maintaining the temporal coherence of the static background regions. The reason for these problems is that these methods do not explicitly adjust the sizes of the grids that correspond to the salient objects and the static background regions, depending on the degree of the spatiotemporal consistency. These problems can significantly degrade the retargeted video (Fig. 2). Therefore, maintaining temporal coherence without deformation of the salient object is a key factor to improve the quality of video retargeting. Here we demonstrate SmartGrid, a method that solves both of these problems. The goal of SmartGrid is to maintain the temporal coherence and to preserve the shape of the salient object by analyzing the degree of the spatiotemporal consistency. This paper makes three contributions:

- SmartGrid analyzes the degree of the spatial consistency to maintain the consistency of salient object shapes between consecutive frames. Use of the resulting saliency map prevents deformation of salient object in regions where the spatial consistency is high.
- SmartGrid maintains temporal coherence for the static background regions by analyzing the consistency of video contents. This method uses the grid information of the previous frame to prevent the occurrence of temporal incoherence artifacts in the retargeted images.
- We provide an efficient and simple method of grid interpolation on retargeted grids. By using this interpolation, our proposed method can generate high-quality retargeted images with much less computation time than the previous methods.

Section II surveys recent video retargeting methods. Section III explains the overview of SmartGrid. Section IV describes SmartGrid. Section V compares the video quality

FIGURE 2. Comparison of video retargeting results: (a) two consecutive original images, (b) temporal incoherence is noticeable (result images generated by [23]), (c) temporal incoherence is noticeable (result images generated by [27]), and (d) temporal incoherence-free (result images generated by SmartGrid). Blue and red lines are grid positons. ©Love Dis KBS.

produced by the video retargeting methods. Section VI concludes the paper.

II. RELATED WORK

Video retargeting methods resize the aspect ratio of the video frame by considering saliency values. The methods can be broadly categorized into pixel-based, region-based, and object-based retargeting methods.

A. PIXEL-BASED AND HYBRID METHODS

Pixel-based retargeting methods resize the image to a desired resolution by removing or inserting relatively less-important pixels. Seam carving methods [7], [12], [13] are the most representative of this class. They calculate a gradient-based saliency map and find a seam that consists of low-saliency pixels. These methods continuously remove or insert the seams until the size of given image reaches the desired resolution. However, these methods may not preserve the contents when the video shows a new type of content or a complex motion.

Hybrid methods have been proposed to solve this problem. A multi-operator approach [14] combines seam carving with scaling and cropping methods. One method [15] combines seam carving and scaling to preserve both object structure and the background region in the image. A retargeting method [16] uses a saliency map that combines the saliency values and spatio-temporal coherence values. Sparse seam carving [17] adjusts the degree of seam separations to preserve the object structure; this method introduced an additional parameter to allow seams to be sparsely assigned to each other. The sparse seam carving method has higher retargeting quality than conventional seam carving-based retargeting methods [7], [12], [13] for a single image, but may lose temporal coherence in video sequences.

A retargeting method [2] uses saliency, face detection, and motion analysis to solve the retargeting problem by applying the least-squares method. To improve the temporal coherence, retargeting method [18] based on the saliency map [30] uses 2D Fourier Transform and motion information to generate the retargeted image; to attain a high-quality video display, the method uses a 2D version of EWA rendering [31]. The central component of this retargeting method [18] is a non-uniform and the pixel-wise warping to adjust the aspect ratio of the image. However, this method fails when the video sequences include visible elements such as buildings, or show complex motions.

Seam carving method which uses a discontinuous seam has been proposed [19]. The method obtains the discontinuous seams by allowing seams to move freely in homogeneous regions of the consecutive frames. Discontinuous seams are low-saliency pixels which jump over the salient object in the adjacent frames to preserve a fast moving salient object. The method calculates spatial and temporal coherence costs for all pixels. Then it adopts a genetic algorithm to find optimal seams. This method considers both visual consistency and computational complexity for preserving the spatial and temporal coherence effectively. However, calculation of the spatial coherence costs using the edge magnitude of pixels leads to inaccurate coherence costs in the images with high texture components. Very recently, a multiple seam searching method [20] establishes connections between consecutive frames by creating a dynamic spatiotemporal buffer. The buffer is a concept to maintain temporal coherence by uniformly applying the seam carving method between consecutive frames. However, uniformly applying the seam carving between consecutive frames can produce temporally unstable results because identical seams are found in consecutive frames with different motion characteristics.

B. REGION-BASED METHODS

Region-based retargeting methods divide a video frame into several regions, then use an optimization technique to assign scaling factors to the separate regions. An algorithm suitable for efficient hardware architecture [21] uses axis-aligned grid deformation based on the 2D saliency estimation and 1D saliency projection for retargeting; this algorithm first transforms the RGB image to a complex number with four components (quaternion), and then transforms them to frequency

space by using the quaternion Fourier Transform. Then phase information is extracted to estimate the saliency values in the image. Based on the saliency map [30], the scaling factors for each column of the image are computed. However, this method maintains the temporal coherence by applying a finite impulse response (FIR) and an infinite impulse response (IIR) filters, so the results can be temporally unstable.

A graph model [22] retargets the image by estimating grid-cell-wise motion. The graph is constructed to obtain the context relationship and to estimate motion. Then the graph-cut algorithm is executed to partition the layers iteratively. The objective function is formulated as coherence preservation and context awareness. Minimization of this function yields new grid vertices and the final retargeted image is obtained using a rendering process. However, when this method is used for retargeting, the important object can be vertically stretched because the grid cell is rectangular.

Stretchability-aware block scaling [23] is an efficient method to retarget images. It first uses gradient, saliency, and color features to measure image stretchability. The optimal size of the stretched image is determined under the constraint of stretchable space. The image is divided into stretchable blocks and non-stretchable blocks, then scaling factors are assigned using the result of the stretchability measure. Finally, the stretched image is uniformly scaled to generate the retargeted image. However, inaccurate partitioning of the image can cause distortion in the retargeting results.

A Bayesian network can be used to construct the grid flows of a video for retargeting [24]. The method first constructs a grid flow and extracts a set of key frames that summarize a video clip. Then a retargeted grid is generated using convex quadratic programming for the set of key frames. Finally, non-key frames are resized using grid interpolation guided by contents of the nearest resized key frames. This method showed better shape preservation of salient object and temporal coherence than the conventional retargeting methods.

One retargeting method [25] divides a video frame into several strips and resizes it using scaling factors obtained by Fourier analysis. Another [26] combines scaling and cropping methods; it uses an optimization technique to balance the loss of detail due to scaling with the loss of content due to cropping, and showed good retargeting results in various video sequences for small displays.

The method in [10] combines cropping and warping operations; it uses the pixel motions that imply inter-frame pixel correspondence to consistently warp corresponding objects, therefore, video quality is highly dependent on the accuracy with which pixel motions are estimated. However, this method can discard important regions in the result image.

A matching-area-based temporal saliency adjustment method for video retargeting [27] allows the seam to track an object that had been previously carved, and tries to avoid carving the seam on different objects in consecutive frames. This method provides better spatial and temporal coherence than the conventional seam carving method. However, this method does not detect regions in which grid position should be maintained, so the temporal stability is not guaranteed for all regions where temporal coherence should be maintained.

The most recent method [28] is to perform video retargeting in the saliency histogram domain. The method adjusts the grid sizes by dividing the video frames into several grids using saliency histogram values. It applies seam carving method to adjust the grid sizes. This method has advantage of low memory and fast run time, so it is especially suitable for mobile applications. However, since it divides frame into several grids, the results of the retargeting can cause a distortion of the saliency object shape when the saliency object is included in adjacent grids.

C. OBJECT-BASED METHODS

Object-based retargeting methods separate the image into foreground and background regions, then perform retargeting for each region [9], [29]. Reference [29] detects the region of interest (ROI) by analyzing color and motion features. Then direct retargeting is applied to background regions excluding the ROI to fit the desired aspect ratio. Finally, the frame is recomposed using the extracted ROI and the resized background regions. However, object-based retargeting methods usually fail to retarget a video when the segmentation result for the ROI is inaccurate.

III. OVERVIEW OF PROPOSED METHOD

The review of related work has demonstrated that the main challenge in video retargeting to maintain the spatiotemporal coherence of the entire video frame. In this paper, we propose SmartGrid, a new video retargeting method that generates optimal retargeted grids by minimizing the position differences of the corresponding grids in neighboring frames. SmartGrid generates a saliency map that represents the rough area of the salient object, then calculates the spatial grid sizes (which consider spatial coherence) by using the extracted saliency values. Then SmartGrid calculates the temporal grid sizes (which consider temporal coherence) by analyzing the grid position of the previous image. SmartGrid formulates the objective function by using the spatial and temporal grid sizes for each column of the image. By minimizing the objective function with various spatiotemporal constraints, SmartGrid finds the optimal grid sizes, then generates the final retargeted image by performing image interpolation on the grids.

IV. PROPOSED METHOD

To perform video retargeting, SmartGrid analyzes the degree of spatiotemporal consistency and adjusts the sizes of the grids. This approach can maintain the consistency of video contents better than the previous methods.

The proposed video retargeting method consists of three steps (Fig. 3): Saliency value extraction, grid optimization, and retargeted image generation steps. Specific operations of the proposed method are described in detail as follows.

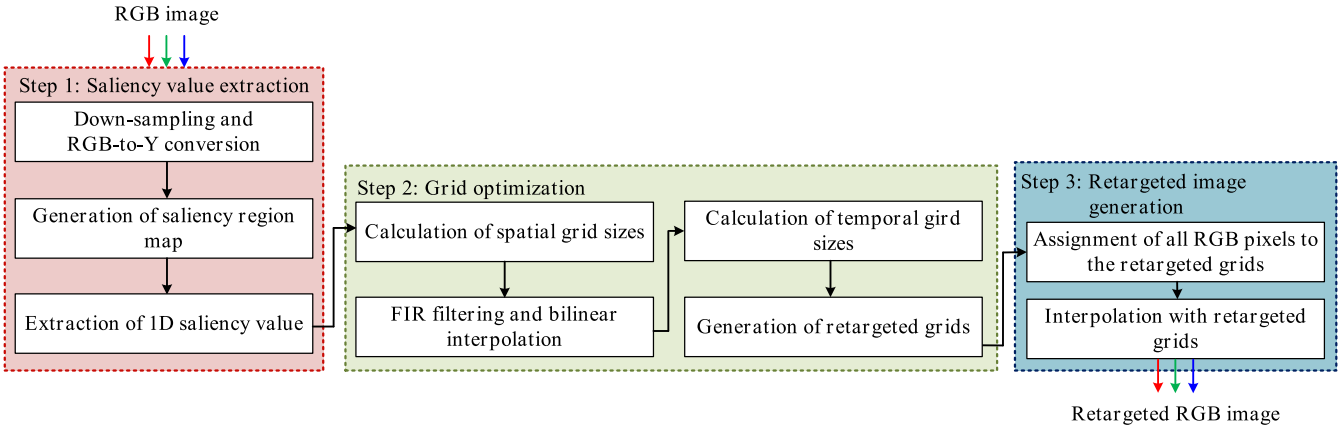


FIGURE 3. Overall flow of the proposed video retargeting method.

A. SALIENCY VALUE EXTRACTION

The first purpose of the video retargeting is to preserve the important regions in the image. Therefore, analyzing the visual importance of the pixels in the image is a prerequisite for our proposed method. A saliency map is a representation of pixel information that is topographically encoded for stimulus conspicuity over the visual scene [32]. It is used in various applications to identify important regions of an image. Similar to previous methods [1]–[29], the proposed method uses this technique to represent important regions of an image.

SmartGrid extracts the saliency value by three processes: down-sampling and RGB-to-Y conversion of the RGB images, generation of the saliency region map, then extraction of the 1D saliency value. These processes are described in detail as follows.

The image (Fig. 4a) is down-sampled to reduce the computational complexity before the saliency map is computed. For the down-sampling, the horizontal and vertical pixel resolutions are reduced by 1/2 [21], [33]–[35]. Next, we generate the luminance image by converting RGB color values to YCbCr color values and extracting only Y, which is the luminance image. The proposed method is performed on the luminance image to extract the saliency values of the image in the next process.

The saliency regions (Fig. 4b) are extracted by analysis of image characteristic. For this process, an image signature algorithm [36] is used because it operates fast by using a discrete cosine transform (DCT). Image signature is defined as the sign function of the DCT of an image; the signature preferentially contains the foreground information, so this algorithm uses image signature to separate the foreground and background regions. Using this concept, the saliency region can be calculated as

$$IS(i,j) = sign(DCT(i,j)),\ (\overline{i,j}) = IDCT(IS(i,j)), \quad (1)$$

where IS denotes an image signature, $IDCT$ denotes the inverse DCT, $sign\ (\cdot)$ is the signum operator, and $\overline{i,j}$ denotes the reconstructed image at pixel (i,j). Using (1), the saliency region is defined as

$$H(i,j) = g * (\overline{i,j}) \circ (\overline{i,j}), \quad (2)$$

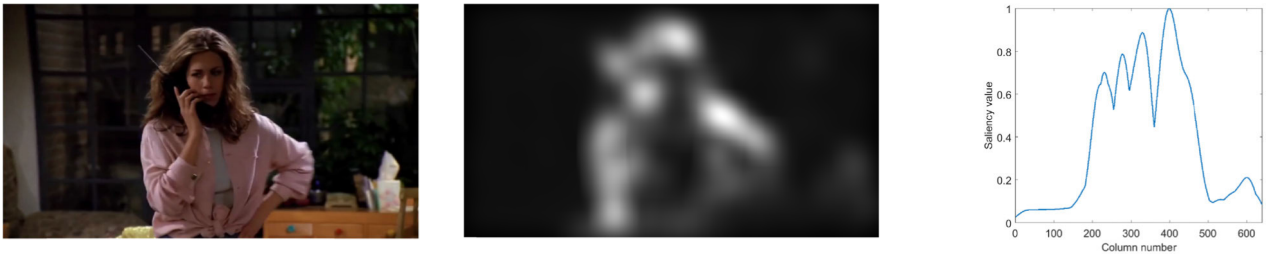
where g denotes a Gaussian kernel, $\circ$ denotes the Hadamard product operation [37], $*$ denotes the convolution operation, and $H(i,j)$ is the saliency region obtained by using the image signature algorithm [36] (Fig. 4b).

We adopt a rectilinear deformation approach to simplify the proposed algorithm. Therefore, SmartGrid calculates the 1D saliency values $S_V(j)$ for column j of the saliency map by extracting the maximum value among all saliency values in column j (Fig. 4c). This process preserves a high saliency value even if the salient object includes only one structure.

B. GRID OPTIMIZATION

The second purpose of the video retargeting is to maintain consistency of the contents for each region in the image. To achieve this purpose, the sizes of the grids that correspond to each region in consecutive frames must be consistently adjusted. We formulate this task as an optimization problem. The solutions of the optimization problem are the sizes of the optimal grids corresponding to the salient objects and the static background regions. We use both spatial and temporal grids to solve this optimization problem. The spatial grids are the rectilinear resampling patterns obtained from the 1D saliency values to preserve the shape of the salient object. The temporal grids are the rectilinear resampling patterns obtained from the retargeted grids of the previous image to maintain the temporal coherence of the static background regions. The optimal grids represent the deformation information of the original image to retarget the image. Therefore, the optimal grids can be used to generate the final retargeted image in the next step.

SmartGrid optimizes the sizes of the grids by following four processes: (i) calculation of spatial grid sizes for each column of the image, (ii) FIR filtering on the spatial grid sizes to smooth the grid and bilinear interpolation, (iii) calculation of temporal grid sizes for each column of the image using the retargeted grids of the previous image, and (iv) generation



(a) (b) (c)

FIGURE 4. Extraction of the saliency value: (a) original RGB image, (b) result of the saliency region map, and (c) result of the 1D saliency value. ©prison break fOX.

of the retargeted grids for each column of the image. These processes are described in detail as follows.

To retain the aspect ratio of the important regions, SmartGrid calculates a spatial grid size that equals 1 in such regions. In addition, the sum of all elements of the spatial grid size should be the desired resolution. Therefore, we use the 1D extracted saliency values to calculate the spatial grid sizes (Fig. 5b) for column k of the image:

$$\begin{aligned} &SG^{Down}(k) = R_{\max} - \min(\beta S_V(k), R_{\min}), \\ &s.t. \sum_{k=1}^{W} SG^{Down}(k) = W'_{down}. \end{aligned} \tag{3}$$

In (3), $SG^{Down}(k)$ and $S_V(k)$ denote the down-sampled spatial grid size and the 1D extracted saliency value at the k-th column of the image. W'_{down} denotes the half width of the desired resolution. Also, R_{max} and R_{min} denote the maximum and the minimum values of the spatial grid size. R_{max}, R_{min}, and constant value β were set to 1.4, 1.0, and 100, respectively, on the basis of preliminary experiments. A description of the preliminary experiments method is described in the next section. In this process, we use a binary search method that can always converge to the exact solution, as did a previous paper [21]. By collecting these down-sampled spatial grid sizes for each column of the image, the down-sampled spatial grids (*SGs*) are generated (Fig. 5 b).

The *SGs* obtained from the previous process are directly related to the 1D saliency values. Therefore, high-frequency and low-frequency fluctuations in the 1D saliency will lead to spatiotemporal fluctuations in the retargeted video. To compensate for these fluctuations, we apply the FIR filter to *SGs* that are likely to include moving objects. To identify the moving objects in the candidate area, SmartGrid detects *distinguished pixels* by thresholding the absolute luminance differences of the current and previous images; for this purpose, a Just Noticeable Distortion (JND) model [38] is used. JND refers to the minimum visibility threshold that the pixel can be perceived by the human visual system. JND is an efficient perceptual model obtained from the extensive experiments. Therefore, appropriate JND model is widely used for the perceptual threshold to guide an image/video processing task. After calculating these pixels for each column, we use the rate of the *distinguished pixels* among all pixels in each column of the image. The equation related to the FIR filter is

$$\begin{aligned} SG^{Down}_{filtered}(k) &= \sum_{j=k-WS}^{k+WS} w_S(j) \times SG^{Down}(j), \\ w_S(k) &= \exp(\frac{Rate_{DP}(k)}{h}), \quad Rate_{DP}(k) = \frac{N_{dist}(k)}{N(k)}. \end{aligned} \tag{4}$$

In (4), $SG_{filtered^{Down}}(k)$ denotes the FIR filter result of the down-sampled spatial grid size in the k-th column of the image. $Rate_{DP}(k)$ and $w_s(k)$ denote the rate of the *distinguished pixels* among all pixels and the weight value in the k-th column of the image. $N_{dist}(k)$ and $N(k)$ denote the number of *distinguished pixels* and the total number of the pixels in the k-th column of the image. Also, WS and h denote the half window size and smoothing strength. In this paper, WS was set to 4 and h was set to 1. FIR filtering yields down-sampled *SG*s (Fig. 5c) for each column of the image. Then we perform bilinear interpolation on $SG_{filtered^{Down}}$ to generate the original resolution of the spatial grid sizes, *SGs* that were lost due to down-sampling for the computation of the saliency value (Fig. 5d).

To maintain the temporal coherence of the static background regions, the grid position of the current image should be equal to the grid position of the previous image in the static background regions. We calculate the temporal grid sizes at column j of the image using the previous *RGs* and the current *SGs* to achieve this purpose:

$$TG_{cur}(j) = \sum_{k=1}^{j} RG_{prev}(k) - \sum_{k=1}^{j-1} SG_{cur}(k), \tag{5}$$

where $RG_{prev}(k)$ and $SG_{cur}(k)$ respectively denote the retargeted grid size of the previous and spatial grid size of the current images at the k-th column. The temporal grids (*TGs*) are generated by collecting these computed $TG_{cur}(j)$. The proposed method applies the computed *TGs* to the static background regions for maintaining the temporal coherence in the next process (Fig. 5e).

Our goal is to maintain temporal coherence without degrading shape preservation of salient objects. Therefore, SmartGrid should perform retargeting to preserve the shape of the salient object in the region where the spatial consistency is high. SmartGrid also should perform retargeting

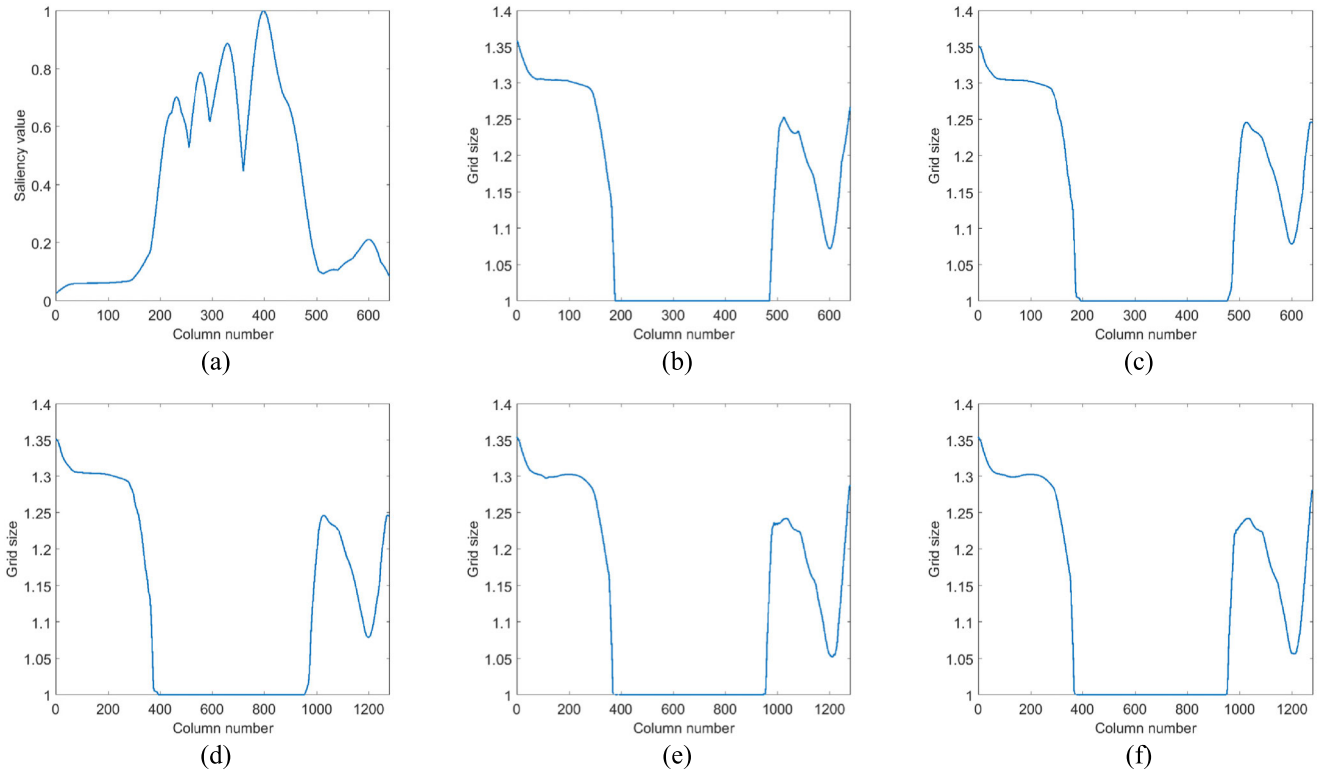


FIGURE 5. Generation of the optimal grids: (a) result of the 1D saliency value, (b) result of spatial grid sizes based on saliency value, (c) FIR filter result of the spatial grid sizes, (d) bilinear interpolation result of (c), (e) result of the temporal grid sizes using the previous retargeted grids, and (f) result of retargeted grids using (d) and (e).

to maintain the temporal coherence in the region where the temporal consistency is high. To simultaneously meet the two requirements, we use *SGs* and *TGs* obtained from the previous processes. The underlying idea of this step is that the retargeted grids (*RGs*) are determined by the combination of *SGs* and *TGs*, depending on the degree of spatiotemporal consistency for each region. To analyze the degree of the spatiotemporal consistency for each region, SmartGrid uses the *distinguished pixel* concept in each column of an image. The description of the *distinguished pixel* concept was already explained in the previous process. To achieve this goal, we formulate the objective function by considering the spatial and temporal energy functions in the j-th column as

$$\begin{aligned} &\min E(j) = E_S(j) + E_T(j) = w_S(j) \times (RG(j) - SG(j))^2 \\ &\qquad\qquad + w_T(j) \times (RG(j) - TG(j))^2, \\ &s.t.\ SG(j) - \lambda \le RG(j) \le TG(j) + \lambda \\ &\quad w_S(j) = \frac{N_{dist}(j)}{N(j)}, \quad w_T(j) = 1 - w_S(j). \end{aligned} \tag{6}$$

In (6), $E_S(j)$ and $E_T(j)$ are respectively the spatial and temporal distortion functions at the j-th column of the image. $SG(j)$ and $TG(j)$ respectively denote the spatial and temporal grid sizes of the j-th column of the image. $RG(j)$ represents the retargeted grid size at the j-th column of the image. Also, $w_S(j)$ and $w_T(j)$ respectively denote the weight values for the spatial and temporal energy functions at the j-th column of the image. In addition, λ was set to 0.1 based on results of preliminary experiments. To determine the RG at the j-th column of the image, we minimize the objective function in (6) by using the second derivative

$$\frac{dE(j)}{dRG(j)} = 0,\ RG(j) = \frac{w_S(j) \times SG(j) + w_T(j) \times TG(j)}{w_S(j) + w_T(j)}. \tag{7}$$

We recalculate the *RGs* to minimize the objective function value to satisfy the following constraints:

$$RG(j) = \begin{cases} RG(j), & if\ TG(j) - \lambda \le RG(j) \le TG(j) + \lambda \\ TG(j) - \lambda, & if\ RG(j) \le TG(j) - \lambda \\ TG(j) + \lambda, & if\ RG(j) \ge TG(j) + \lambda. \end{cases} \tag{8}$$

By collecting these computed the *RGs* for each column of the image, the optimal grids are generated (Fig. 5f). The optimal grids will be used to generate the final retargeted image in the next step.

C. RETARGETED IMAGE GENERATION

The optimal grids obtained from the previous step represent the deformation information of the original image to retarget the image. Using the deformation information of the original image, we can generate a final retargeted image. To retarget the input image, we must predict the unknown pixels in the corresponding grid regions. An image interpolation is a technique to generate the unknown pixels from the provided pixels. Therefore, we use this technique on the

(a) (b) (c)

FIGURE 6. Generation of the retargeted image: (a) result of the retargeted image with the grid hole, (b) gradient result on (a) using the Sobel operation, and (c) result of the retargeted RGB image. ©prison break fox.

grids to predict the unknown pixels of the optimal grids. The specific operations of this step are described in detail as follows.

SmartGrid generates the final retargeted RGB image by applying two processes.

In the first process, SmartGrid rounds off the position values of the *RGs* obtained in the previous step. Then, in the current image, SmartGrid finds all RGB pixel values that correspond to each index of the *RGs*, then assigns all RGB pixel values to the grid positions that correspond to each index of the *RGs*. This process generates a retargeted image that has an empty region in which the RGB pixel values are not defined (a grid hole) (Fig. 6a). In the next process, SmartGrid predicts the pixel information for this region.

In the second process, SmartGrid generates the retargeted RGB image by performing image interpolation on the *RG*s in which the RGB pixel values of the current image are not assigned. To compensate for edge blurring in the retargeted image, the image interpolation technique should consider edge pixels as outliers. Therefore, in this process SmartGrid uses Sobel operation [39] to calculate the gradient map for the retargeted image that has a grid hole. Then we perform the image interpolation on the retargeted image with the grid hole at pixel (i, j), by using the neighboring pixels of the grid hole as

$$C_{retargeted}(i,\ j) = \frac{\sum\limits_{(x,\ y)\in D_{(i,\ j)}} e^{SE(x,\ y)} C_{grid\_hole}(x,\ y)}{\sum\limits_{(x,\ y)\in D_{(i,\ j)}} e^{SE(x,\ y)}}, \quad (9)$$

where $\mathbf{C}_{retargeted}(i,\ j)$ and $\mathbf{C}_{grid\_hole}(x,\ y)$ represent respectively the RGB values of the retargeted image at pixel (i, j) and the RGB values of the retargeted image with the grid hole at pixel $(x,\ y)$, $D_{(i,j)}$ represents the neighboring pixels centered at pixel (i, j), and $SE\ (x,\ y)$ represents the gradient result of the Sobel edge mask for the retargeted image with the grid hole at pixel $(x,\ y)$. In this paper, the block size $D_{(i,j)}$ was set to 3×3 pixels. The gradient result (Fig. 6b) of the retargeted image is used to generate the final retargeted RGB image (Fig. 6c). The method proposed in Step 3 has fewer artifacts and lower computational complexity than the existing EWA rendering method [31]. The reason for this advantage is that SmartGrid generates the retargeted RGB image by assigning all of the original pixel values to the *RG*s, and performs image interpolation only for *RG* regions in which pixel values are not assigned.

V. EXPERIMENTAL RESULTS

We conducted experiments to compare the quality of retargeting by SmartGrid and by previous methods. First, we quantitatively compared the video-retargeting quality and the temporal coherence obtained using SmartGrid and the previous methods by using the Bidirectional Similarity Measure (*BSM*) [40] and two Jittery Metrics (*JM1*, *JM2*) [41], [42]. Second, we visually assessed the quality of SmartGrid and the previous methods by using a pairwise comparison. Third, we compared the computation complexity of the previous methods and SmartGrid. Finally, we evaluated the overall video-retargeting quality of SmartGrid and the previous methods by combining these three criteria. The pairwise comparison is the most popular subjective evaluation methodology in video retargeting. For the subjective evaluation, these pairwise comparisons were performed as in previous work [8], [18], [19], [22], [24], [27], and [28].

A. EXPERIMENTAL ENVIRONMENT

We used 17 video sequences [43] (Table 1, 2). We collected datasets from four Korean broadcasting companies (KBS, MBC, SBS, TVN, SPOTV), CNN, MSNBC, NBC, TBS, BBC, FOX, and also collected a movie sequence (*Big Buck Bunny*). The experiments were conducted on 9,782 frames from HD resolution (1280 × 720 pixels) to Full HD resolution (1920 × 1080 pixels). Various parameters used in SmartGrid were optimized to values obtained in preliminary experiments. The sequences of *Video 15*, *16*, and *17* were used to train and select the parameters because these sequences contained diverse types of motion such as camera motion, object motion, and static motion and diverse types of the saliency objects. The rest of the sequences were used as test sequences.

We used eight previous methods [7], [17], [19], [20], [21], [23], [27], and [28]. Of these, [7], [17], [19], and [20] are pixel-based methods. Reference [21] is a typical warping-based method that calculates the grid size. Reference [27] is region-based method to estimate the matching area for video retargeting and [23] is the most representative warping-based method to calculate the stretchability for video retargeting. Reference [28] is the state-of-the-art.

TABLE 1. Test video sequences used in experiments.

Test video	Number of frames	Resolution	Category	Title
1	500	1280 × 720	Entertainment	My Little Television (MBC)
2	352	1280 × 720	Drama	Oh My Ghost (TVN)
3	810	1280 × 720	Sitcom	High Kick (MBC)
4	820	1280 × 720	Drama	Discovery of Love (KBS)
5	800	1280 × 720	Drama	Friends (NBC)
6	700	1280 × 720	Talk show	Conan Show (TBS)
7	500	1280 × 720	Documentary	The Human Body Secrets of Your Life Revealed (BBC)
8	600	1280 × 720	Drama	Discovery of Love (KBS)
9	500	1280 × 720	Drama	Prison Break (FOX)
10	700	1920 × 1080	Animation movie	Big Buck Bunny
11	600	1920 × 1080	News	CNN News (CNN)
12	500	1920 × 1080	News	MSNBC News (MSNBC)
13	500	1920 × 1080	Music program	Popular Songs (SBS)
14	700	1920 × 1080	Sport	Major League Baseball (FOX)
Total	8,582			

TABLE 2. Training video sequences used in experiments.

Training Video	Number of frames	Resolution	Category	Title
15	300	1280 × 720	Entertainment	Radio Star (MBC)
16	300	1280 × 720	Drama	MALE (KBS)
17	600	1280 × 720	Sports	Volleyball (SPOTV)
Total	1,200			

TABLE 3. *BSM* results of the previous methods and SmartGrid.

Video	Rubinstein [7]	Greisen [21]	Yan [27]	Du [23]	Choi [17]	Li [19]	Zhu [20]	Hsin [28]	SmartGrid
1	2.3988	3.4894	2.4743	6.5987	3.4766	3.7550	2.5368	1.7459	**1.4886**
2	1.0761	1.2977	0.8233	2.5962	1.1794	1.4767	1.1268	0.5442	**0.4272**
3	0.8957	1.2274	1.1215	2.8510	1.5912	2.2091	0.6574	0.9781	**0.6506**
4	1.1783	1.5790	0.9356	3.5614	1.2458	1.6858	1.4214	0.7868	**0.7230**
5	1.3198	1.7098	**0.5644**	3.8035	1.5882	1.9590	1.4385	0.6472	0.6294
6	1.3703	1.6641	1.1427	3.1911	1.8813	2.1709	1.4788	0.8344	**0.7963**
7	1.3937	1.5697	0.9922	3.2311	1.8632	1.6834	2.6079	0.6663	**0.4182**
8	1.2201	1.2384	1.1714	2.8998	1.3098	1.7864	1.2306	0.9098	**0.7311**
9	0.9016	0.7848	0.5160	1.7105	1.1133	1.2414	0.6352	**0.4080**	0.4556
10	1.1369	1.5230	1.0069	2.6949	1.6519	1.9475	1.5875	0.8417	**0.5976**
11	1.5456	2.1318	1.9861	7.6121	2.5468	3.9928	1.6699	1.1328	**1.1284**
12	3.3524	3.9478	3.5109	7.7106	3.8040	4.0510	4.5309	**1.2271**	1.2733
13	3.7934	4.2947	3.2485	7.9213	5.0271	3.5122	3.6163	1.3079	**1.0095**
14	1.7151	2.0185	1.2987	3.4151	1.5692	1.9592	1.8227	0.8884	**0.5680**
Avg.	1.5919	1.9602	1.4095	4.1471	2.0419	2.3424	1.7911	0.9141	**0.7673**

For all previous methods, various parameters were optimized and set to the values guided by their corresponding papers. The 16:9 aspect ratio is widely accepted as a standard for televisions, projectors, monitors, and other devices [44]. However, many smartphones, tablets, personal computers, projectors, and televisions adopt a wide-screen format from 18:9 to 21:9 [44]. Among them, the 18:9 aspect ratio is widely used for smartphones and tablets that have been recently produced. Therefore, in our experiments, we converted the aspect ratio of each video from 16:9 to 18:9 for the applications of recently produced devices.

B. OBJECTIVE EVALUATION

In the first objective evaluation, we compared the temporal coherence maintenance and shape completeness achieved by SmartGrid and the previous methods.

TABLE 4. *JM1* (top) and *JM2* (bottom) results of the previous methods and SmartGrid.

Video	Rubinstein [7]	Greisen [21]	Yan [27]	Du [23]	Choi [17]	Li [19]	Zhu [20]	Hsin [28]	SmartGrid
1	2.92e-04	7.70e-04	2.92e-04	2.70e-03	2.98e-04	3.70e-04	2.59e-04	9.63e-04	**4.35e-05**
2	3.36e-04	1.20e-03	3.17e-04	3.40e-03	3.50e-04	4.26e-04	2.86e-04	2.40e-03	**3.32e-05**
3	2.52e-04	3.79e-04	2.69e-04	1.70e-03	2.34e-04	3.41e-04	2.26e-04	1.80e-03	**3.13e-05**
4	3.09e-04	6.28e-04	2.97e-04	3.20e-03	3.11e-04	3.88e-04	2.78e-04	1.90e-03	**2.87e-05**
5	2.78e-04	8.20e-04	2.45e-04	3.00e-03	2.85e-04	3.69e-04	2.45e-04	1.90e-03	**4.27e-05**
6	2.20e-04	5.11e-04	2.11e-04	1.50e-03	2.68e-04	3.28e-04	2.13e-04	5.71e-04	**1.00e-05**
7	3.52e-04	1.40e-03	3.52e-04	2.90e-03	3.52e-04	4.31e-04	3.42e-04	1.36e-02	**1.03e-04**
8	2.80e-04	3.46e-04	2.26e-04	2.00e-03	2.66e-04	3.68e-04	2.53e-04	1.50e-03	**9.88e-06**
9	2.68e-04	7.33e-04	2.68e-04	2.50e-03	2.80e-04	3.79e-04	2.19e-04	2.60e-03	**3.42e-05**
10	1.70e-04	6.15e-04	1.72e-04	2.20e-03	2.03e-04	2.03e-04	1.47e-04	2.22e-04	**1.28e-05**
11	1.23e-04	3.10e-04	1.24e-04	3.20e-03	1.51e-04	2.11e-04	1.22e-04	2.42e-04	**3.95e-06**
12	2.09e-04	7.46e-04	1.83e-04	2.70e-03	2.22e-04	2.57e-04	1.62e-04	1.81e-04	**4.48e-06**
13	2.52e-04	1.10e-03	2.42e-04	2.20e-03	2.57e-04	2.98e-04	2.31e-04	1.90e-03	**5.32e-05**
14	2.43e-04	8.93e-04	2.35e-04	2.30e-03	2.50e-04	2.81e-04	2.22e-04	3.20e-03	**2.59e-05**
Avg.	2.53e-04	7.09e-04	2.42e-04	2.50e-03	2.63e-04	3.29e-04	2.27e-04	2.20e-03	**2.99e-05**

Video	Rubinstein [7]	Greisen [21]	Yan [27]	Du [23]	Choi [17]	Li [19]	Zhu [20]	Hsin [28]	SmartGrid
1	2.20e-03	6.91e-02	3.90e-03	3.30e-01	3.20e-03	3.10e-03	2.60e-03	2.10e-03	**2.12e-04**
2	4.20e-03	1.05e-01	5.90e-03	4.76e-01	5.70e-03	6.10e-03	6.10e-03	4.90e-03	**9.41e-05**
3	1.80e-03	2.77e-02	3.10e-03	2.36e-01	2.90e-03	3.80e-03	1.80e-03	4.00e-03	**1.37e-04**
4	1.80e-03	7.38e-02	5.20e-03	4.61e-01	2.70e-03	2.90e-03	3.20e-03	4.00e-03	**9.82e-05**
5	2.60e-03	7.42e-02	3.20e-03	3.67e-01	3.20e-03	4.00e-03	3.00e-03	4.10e-03	**2.54e-04**
6	1.60e-03	4.81e-02	2.40e-03	2.47e-01	2.90e-03	3.80e-03	1.90e-03	1.20e-03	**2.14e-05**
7	4.70e-03	1.07e-01	5.90e-03	4.30e-01	6.50e-03	5.00e-03	1.04e-02	7.40e-02	**7.10e-04**
8	1.40e-03	3.19e-02	2.00e-03	2.62e-01	1.70e-03	2.40e-03	1.30e-03	3.60e-03	**7.63e-06**
9	4.10e-03	7.41e-02	4.90e-03	3.14e-01	4.70e-03	6.30e-03	2.90e-03	5.60e-03	**1.26e-04**
10	1.80e-03	7.32e-02	3.00e-03	4.07e-01	3.30e-03	2.90e-03	2.20e-03	5.34e-04	**1.63e-04**
11	8.49e-04	2.79e-02	1.20e-03	5.65e-01	1.60e-03	3.00e-03	1.10e-03	6.01e-04	**1.62e-05**
12	2.50e-03	8.42e-02	2.40e-03	5.87e-01	3.20e-03	3.10e-03	3.40e-03	4.74e-04	**3.12e-05**
13	4.20e-03	1.39e-01	5.10e-03	4.96e-01	6.00e-03	3.80e-03	5.60e-03	4.90e-03	**7.37e-04**
14	3.70e-03	1.15e-01	5.90e-03	5.27e-01	4.30e-03	5.20e-03	4.70e-03	3.54e-02	**2.12e-04**
Avg.	2.50e-03	7.19e-02	3.80e-03	4.00e-01	3.50e-03	3.80e-03	3.30e-03	9.70e-03	**1.90e-04**

Bidirectional Similarity Measure (*BSM*) [40] was used as the evaluation metric:

$$d(S,\ T) = \frac{1}{N_S}\sum_{P\subset S}\min_{Q\subset T} D(P,\ Q) + \frac{1}{N_T}\sum_{Q\subset T}\min_{P\subset S} D(Q,\ P), \tag{10}$$

where S is a source image and T is a retargeted image, P and Q denote patches in S and T, respectively, and N_S and N_T denote the number of patches in S and T, respectively. $D(\cdot)$ denotes the *SSD* (Sum of Squared Distances) between two patches in CIE $L*a*b$ color space.

BSM calculates patch-wise bidirectional comparisons between the original images and the retargeted images. It measures whether all patches of the original image have been preserved in the retargeted image and whether any new patches occur in the retargeted image but not in the original image. If the content of the original image is well preserved and the temporal coherence is well maintained in the retargeted image, *BSM* is low.

In the second objective evaluation, we compared the temporal coherence of SmartGrid and the previous methods. The evaluation criteria were the Jittery Metrics (*JM1*, *JM2*) [41], [42]. The original input video sequences are absolutely jittery-free. In general, temporal incoherence can appear as jittery artifacts. To maintain temporal coherence, the grid sizes and positions of the corresponding pixels in consecutive frames should be constant. The first jittery artifact between the k-th frame and the k-*1*-th frame at grid $(i,\ j)$ can be defined as

$$DS^k_{i,j} = S^k_{i,j} - S^{k-1}_{i,j}, \tag{11}$$

where $S^k_{i,j}$ denotes the grid size of the k-th retargeted image at grid $(i,\ j)$. If the magnitude of $DS^k_{i,j}$ is small, the horizontal jittery artifact at grid $(i,\ j)$ between the k-th retargeted image and the k-*1*-th retargeted image is low. The first jittery metric (*JM1*) between the k-th and k-*1*-th images is defined as

$$JM1_k = \frac{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(DS^k_{i,j})^2}}{M\cdot N}, \tag{12}$$

TABLE 5. Average BSM, *JM1*, and *JM2* values for case 1 and SmartGrid.

Measure	Case 1*	SmartGrid
Avg. *BSM*	1.3792	**0.7673**
Avg. *JM*	2.79e-04	**2.99e-05**
Avg. *JM*	3.84e-02	**1.90e-04**

*Case 1: SmartGrid considers the spatial grids only in the video retargeting process.

TABLE 6. Pairwise comparison of videos generated by the previous methods and SmartGrid (The value of entry (i, j) represents how many times method i was preferred over method j).

	j									
i	**[7]**	**[21]**	**[27]**	**[23]**	**[17]**	**[19]**	**[20]**	**[28]**	**Smart Grid**	**Total**
[7]	–	468	579	671	573	534	418	232	91	3,566
[21]	232	–	532	667	453	508	259	149	36	2,836
[27]	121	168	–	663	262	298	210	105	50	1,877
[23]	29	33	37	–	52	23	18	15	6	213
[17]	127	247	438	648	–	204	207	76	26	1,973
[19]	166	192	402	677	496	–	265	83	31	2,312
[20]	282	441	490	682	493	435	–	224	39	3,086
[28]	468	551	595	685	624	617	476	–	127	4,143
Smart Grid	609	664	650	694	674	669	661	573	–	**5,194**

TABLE 7. Comparison of the average computation times C_T (μs/pixel) for the previous methods and SmartGrid.

Rubinstein [7]	Greisen [21]	Yan [27]	Du [23]	Choi [17]	Li [19]	Zhu [20]	Hsin [28]	SmartGrid
192.590	21.665	410.136	41.898	205.720	119.834	104.326	60.270	**3.529**

where M and N respectively denote the number of rows and columns in the input image. The second jittery artifact between the k-th and k-*1*-th images at grid (i, j) is defined as

$$DP_{i,j}^{k} = P_{i,j}^{k} - P_{i,j}^{k-1}, \tag{13}$$

where $P_{i,j}^{k}$ denotes the grid position of the k-th retargeted image at grid (i, j). If the magnitude of $DP_{i,j}^{k}$ is small, the horizontal jittery artifact is small at grid (i, j) between the k-th retargeted image and the k-*1*-th retargeted image. The second jittery metric (*JM2*) between the k-th image and the k-*1*-th image is defined as

$$JM2_k = \frac{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(DP_{i,j}^{k})^2}}{M \cdot N}. \tag{14}$$

Small values of *JM1* and *JM2* are good. SmartGrid achieved lower average values of *BSM* (Table 3, Fig. 7), *JM1* and *JM2* (Table 4) than the eight previous methods on all 14 test video sets. Therefore, SmartGrid consistently produced better video-retargeting quality than the previous methods. This improvement by SmartGrid can be attributed to the use of 1D extraction of the saliency map information, because this extraction can preserve the shape of the salient object in the video sequences. In addition, to determine the optimal grid sizes, we minimized the objective function in (6) that aims to reduce deformation for the static background regions of the current image. This process ensures that the positions of the retargeted grids for the static background regions in the current image are similar to the positions of the retargeted grids for the static background regions in the previous image.

Specifically, SmartGrid achieved *BSM* that was 0.8246 lower than [7], 1.1929 lower than [21], 0.6422 lower than [27], 3.3798 lower than [23], 1.2746 lower than [17], 1.5751 lower than [19] 1.0238 lower than [20], and 0.1468 lower than [28]. SmartGrid also achieved the lowest *JM1* and *JM2* among the eight previous methods. We also evaluated the effect of temporal grid sizes in SmartGrid (Table 5). The experiment was conducted using all test sequences for two cases. In Case 1, SmartGrid considered only spatial grid sizes and ignores temporal grid sizes; the SmartGrid case considered both sequences. SmartGrid greatly improved the video-retargeting quality and temporal coherence because it accurately predicts the regions where the grid position should be maintained in the corresponding regions between neighboring frames by applying the spatiotemporal grid optimization. In comparisons (Fig. 8), SmartGrid was superior to the previous methods for the salient object region without visual artifacts or shape distortion. The reason for this improvement is that SmartGrid can detect the saliency regions of the video contents and smoothen the grid size of the region in which moving objects are likely to occur; i.e., SmartGrid can minimize the visual artifacts during the video retargeting process, when compared to the previous methods. The methods of [7], [17], [19], [20], [21], [23], [27], and [28] found inaccurate regions where temporal coherence should be maintained for the given images. As a result, human faces or bodies, and background regions are stretched or distorted. In contrast, SmartGrid can preserve the human face and body without stretching, because this method analyzes the temporal consistency of video contents and maintains the grid positions of the regions where temporal coherence should be maintained. Furthermore, SmartGrid analyzes the spatial consistency of video contents and deforms the grids in regions where motion occurs.

C. SUBJEC TIVE EVALUATION

We conducted pairwise comparisons to evaluate the video-retargeting quality of SmartGrid and the previous methods.

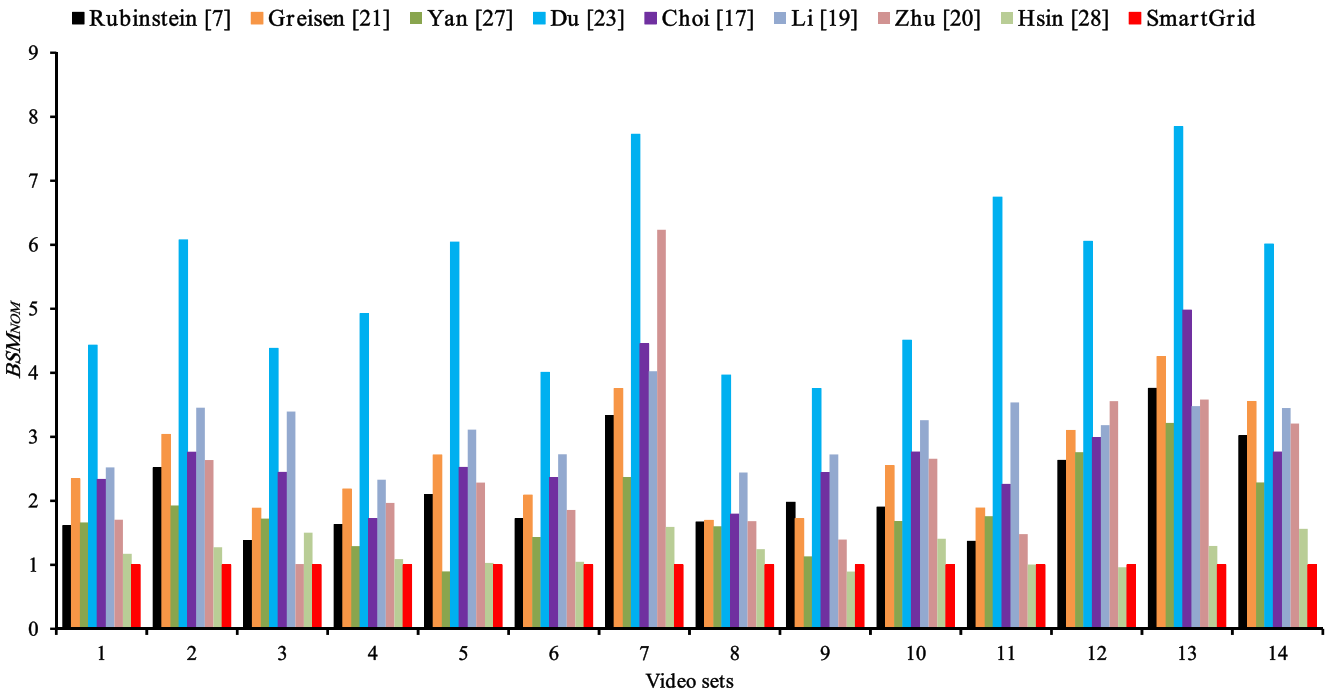


FIGURE 7. Comparison of *BSM* values among different video retargeting methods (normalized values to SmartGrid).

For subjective evaluation, the pairwise comparison has been widely used as an evaluation method in previous work [8], [18], [19], [22], [24], [27], and [28]. We invited 50 participants of different ages and professions, then showed them the original video and a pair of retargeted videos generated by two different retargeting methods. The retargeted videos were displayed in random order to prevent participants from inferring the retargeting methods. The participants were not provided with any technical information; we simply asked the participants to state which retargeted video was better. We generated 14 retargeted videos using SmartGrid and the eight previous methods for the subjective evaluation. As there were nine retargeting methods, we performed ${}_9C_2$ (=36) pairwise comparisons for each retargeted video. Thus, we received answers for $36 \times 14 = 504$ comparisons from each participant, and we received $504 \times 50 = 25{,}200$ answers from all participants. Also, the results of each method were compared $8 \times 14 \times 50 = 5{,}600$ times with the results of the other methods. We quantified the preference rate of each method among 5,600 pairwise comparisons. SmartGrid was preferred over all previous methods used in the experiment (Table 6). In Table 6, the value of entry (i, j) represents how many times method i was preferred over method j. Specifically, the preference rate of the proposed method was 5,194/5,600 (92.75%). In comparison, the preference rate of [7] was 3,566/5,600 (63.68%), of [21] was 2,836/5,600 (50.64%), of [27] was 1,877/5,600 (33.52%), of [23] was 213/5,600 (3.80%), of [17] was 1,973/5,600 (35.23%), of [19] was 2,312/5,600 (41.29%), of [20] was 3,086/5,600 (55.11%), and of [28] was 4,143/5,600 (73.98%).

TABLE 8. Comparison of the average computation times C_T (μs/pixel) for each step of SmartGrid.

Step 1	Step 2	Step 3	Total
0.941	0.889	1.699	**3.529**
27%	25%	48%	**100%**

These results show that SmartGrid gives retargeting results that are more visually pleasing than those of the previous methods.

D. COMPUTATIONAL COMPLEXITY

We compared the computation times of each method by using MATLAB on a PC with an Intel E5-2697 processor at 2.60 GHz. The comparison metric was computation time per pixel C_T [μs]. The proposed method significantly reduced the average of C_T by 98.17% compared to [7], by 83.71% compared to [21], by 99.14% compared to [27], by 91.58% compared to [23], by 98.28% compared to [17], by 97.06% compared to [19], by 96.62% compared to [20], and by 94.14% compared to [28] (Table 7). SmartGrid achieves this high speed because it uses simple spatiotemporal optimization without any iterative process, and performs simple image interpolation on the retargeted grids; the most time-consuming step is generation of the final retargeted image (Table 8).

E. OVERALL VIDEO-RETARGETING QUALITY

We compared the overall utility of the proposed and previous methods by combining video-retargeting quality and

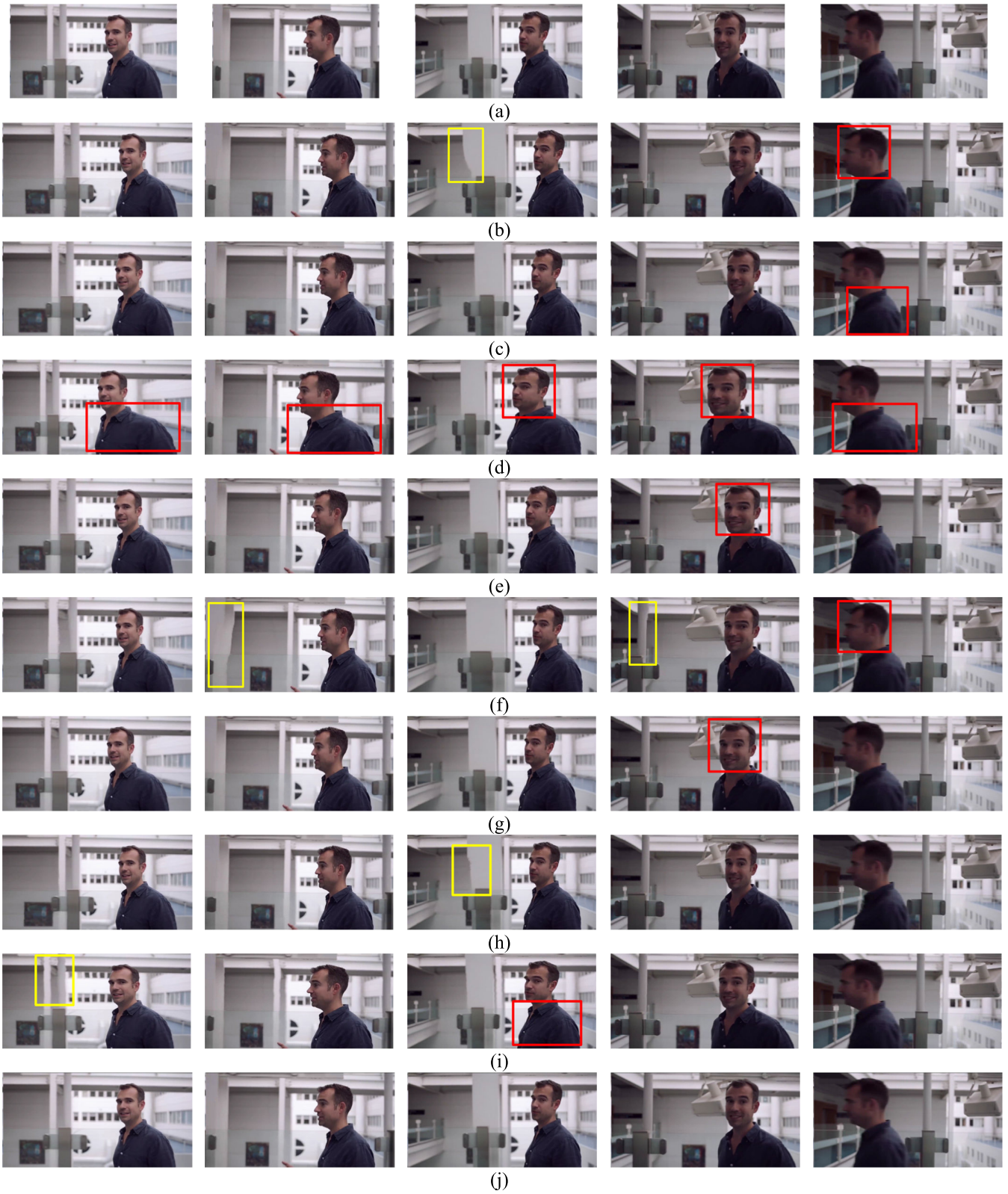


FIGURE 8. Comparison with previous methods for video 7 sequences. Rows from top to bottom: (a) consecutive original images (1st column: 180th frame, 2nd column: 221st frame, 3rd column: 244th frame, 4th column: 308th frame, and 5th column: 338th frame), (b) result images generated by [7], (c) result images generated by [21], (d) result images generated by [27], (e) result images generated by [23], (f) result images generated by [17], (g) result images generated by [19], (h) result images generated by [20], (i) result images generated by [28], and (j) result images generated by SmartGrid. Red box represents an unreasonably stretched region in the salient object. Yellow box represents a distortion of the vertical structure in the background region. ©The human body secrets of your life revealed BBC.

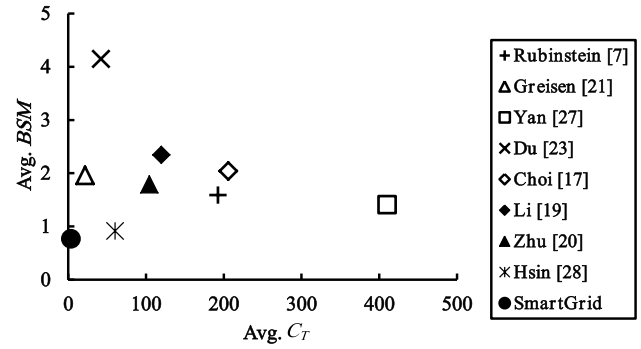


FIGURE 9. Comparison of the average *BSM* values and average computation times (C_T) obtained using SmartGrid and the previous methods.

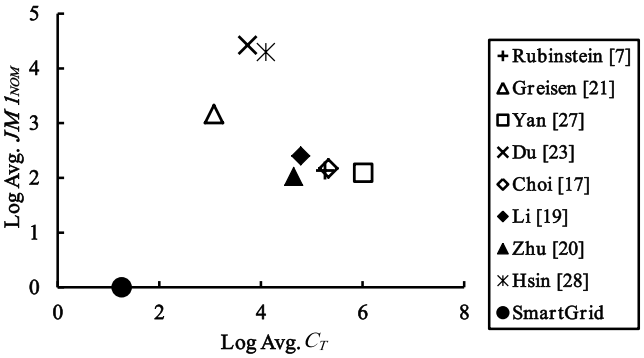


FIGURE 10. Comparison of the average *JM1* and average computation times (C_T) obtained using SmartGrid and the previous methods (normalized values to SmartGrid).

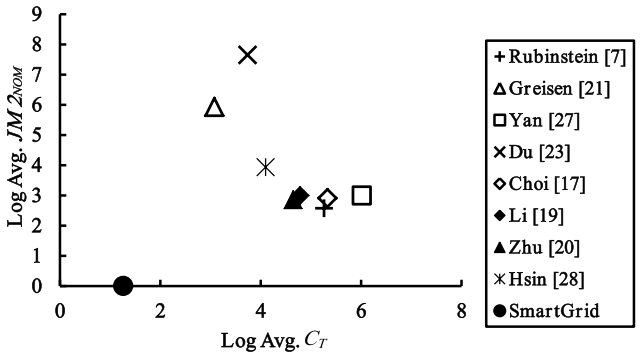


FIGURE 11. Comparison of the average *JM2* and average computation times (C_T) obtained using SmartGrid and the previous methods (normalized values to SmartGrid).

computation time. The *BSM* values were averaged over all test video sequences. SmartGrid produced was faster and produced lower *BSM* (Fig. 9), *JM1* (Fig. 10), and *JM2* (Fig. 11) than all of the previous methods. The experimental results demonstrate that SmartGrid provides the best video-retargeting quality, and does it fast.

VI. CONCLUSION

We proposed a new video-retargeting method that uses an optimization method by minimizing the position differences of corresponding grids of the current and previous images. SmartGrid uses the saliency values and the positions of *RGs* in the previous image to calculate the spatial and temporal grid sizes. The goal of our proposed method is to maintain the consistency of the contents for each region in consecutive frames. To achieve this goal, we adjust the sizes of the grids that correspond to the salient objects and the static background regions. SmartGrid uses both spatial and temporal grid sizes to formulate the retargeting problem as an optimization problem. Based on the spatiotemporal constraints, we minimize an objective function to calculate the optimal grid sizes. SmartGrid collects these optimal grid sizes to generate *RGs*. Finally, it generates the retargeted image by performing an image interpolation on the generated *RGs*. The benefits of our proposed method were verified in extensive experiments on 14 video datasets. In experiments, SmartGrid improved the *BSM*, *JM1*, and *JM2* by 1.19×, 7.59× and 13.16×, respectively, and reduced computational complexity by 6.14×, compared to the best results of previous methods [7], [17], [19], [20], [21], [23], [27], and [28]. Furthermore, subjective evaluation proved that the preference rate of SmartGrid was 92.75%. From the experimental results, we conclude that SmartGrid provides superior video-retargeting quality with much lower computational complexity than the previous methods.

REFERENCES

[1] A. Shamir and O. Sorkine, "Visual media retargeting," in *Proc. ACM SIGGRAPH ASIA Courses*, 2009, pp. 1–13.
[2] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. IEEE 11th ICCV*, Oct. 2007, pp. 1–6.
[3] Y. Niu, F. Liu, X. Li, and M. Gleicher, "Warp propagation for video resizing," in *Proc. IEEE Int. Conf. CVPR*, Jun. 2010, pp. 537–544.
[4] Y.-F. Zhang, S.-M. Hu, and R. R. Martin, "Shrinkability maps for content-aware video resizing," *Comput. Graph. Forum*, vol. 27, no. 7, pp. 1797–1804, Oct. 2008.
[5] L. Shi, J. Wang, L. Duan, and H. Lu, "Consumer video retargeting: Context assisted spatial-temporal grid optimization," in *Proc. 17th ACM Int. Conf. Multimedia*, Oct. 2009, pp. 301–310.
[6] W.-L. Chao, H.-H. Su, S.-Y. Chien, W. Hsu, and J.-J. Ding, "Coarse-to-fine temporal optimization for video retargeting based on seam carving," in *Proc. IEEE ICME*, Jul. 2011, pp. 1–6.
[7] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 16:1–16:9, Aug. 2008.
[8] T.-C. Yen, C.-M. Tsai, and C.-W. Lin, "Maintaining temporal coherence in video retargeting using mosaic-guided scaling," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2339–2351, Aug. 2011.
[9] Y.-S. Wang, H. Fu, O. Sorkine, T.-Y. Lee, and H.-P. Seidel, "Motion-aware temporal coherence for video resizing," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 127:1–127:10, Dec. 2009.
[10] Y.-S. Wang, H.-C. Lin, O. Sorkine, and T.-Y. Lee, "Motion-based video retargeting with optimized crop-and-warp," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 90:1–90:9, Jul. 2010.
[11] Y.-S. Wang, J.-H. Hsiao, O. Sorkine, and T.-Y. Lee, "Scalable and coherent video resizing with per-frame optimization," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 88:1–88:8, Aug. 2011.
[12] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 10:1–10:8, Jul. 2007.
[13] A. Mansfield, P. Gehler, L. Van Gool, and C. Rother, "Scene carving: Scene consistent image retargeting," in *Proc. ECCV*, Sep. 2010, pp. 143–156.

[14] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, Aug. 2009, Art. no. 23.
[15] W. Dong, N. Zhou, J.-C. Paul, and X. Zhang, "Optimized image resizing using seam carving and scaling," *ACM Trans. Graph.*, vol. 29, no. 5, Dec. 2009, Art. no. 125.
[16] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 569–576.
[17] J. Choi and C. Kim, "Sparse seam-carving for structure preserving image retargeting," *J. Signal Process. Syst.*, vol. 85, no. 2, pp. 275–283, Nov. 2016.
[18] P. Krähenbhl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, p. 126, Dec. 2009.
[19] C. Li, R. Hu, C. Liang, C. Xiao, and W. Ruan, "Faster seam carving for video retargeting," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 823–827.
[20] Z. Chuning, "Fast video retargeting based on seam carving with parental labeling," 2019, *arXiv:1903.03180*. [Online]. Available: https://arxiv.org/abs/1903.03180
[21] P. Greisen, M. Lang, S. Heinzle, and A. Smolic, "Algorithm and VLSI architecture for real-time 1080p60 video retargeting," in *Proc. Eurograph. Conf. High-Perform. Graph.*, 2012, pp. 57–66.
[22] Z. Qu, J. Wang, M. Xu, and H. Lu, "Context-aware video retargeting via graph model," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1677–1687, Nov. 2013.
[23] H. Du, Z. Liu, J. Jiang, and L. Shen, "Stretchability-aware block scaling for image retargeting," *J. Vis. Commun. Image Represent.*, vol. 24, no. 4, pp. 499–508, May 2013.
[24] B. Li, L. Duan, J. Wang, R. Ji, C. Lin, and W. Gao, "Spatiotemporal grid flow for video retargeting," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1615–1628, Apr. 2014.
[25] J.-S. Kim, J.-H. Kim, and C.-S. Kim, "Adaptive image and video retargeting technique based on Fourier analysis," in *Proc. IEEE Conf. CVPR*, Jun. 2009, pp. 1730–1737.
[26] F. Liu and M. Gleicher, "Video retargeting: Automating pan and scan," in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 241–250.
[27] B. Yan, K. Sun, and L. Liu, "Matching-area-based seam carving for video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 302–310, Feb. 2013.
[28] H.-C. Hsin, "Video retargeting based on SH equalisation and seam carving," *IET Image Process.*, vol. 13, no. 8, pp. 1333–1340, Jun. 2019.
[29] W.-H. Cheng, C.-W. Wang, and J.-L. Wu, "Video adaptation for small display based on content recomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 43–58, Jan. 2007.
[30] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.
[31] P. Greisen, M. Schaffner, S. Heinzle, M. Runo, A. Smolic, A. Burg, H. Kaeslin, and M. Gross, "Analysis and VLSI implementation of EWA rendering for real-time HD video applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 11, pp. 1577–1589, Nov. 2012.
[32] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, 2001.
[33] Y. D. Ahn and S.-J. Kang, "Backlight dimming based on saliency map acquired by visual attention analysis," *Displays*, vol. 50, pp. 70–77, Dec. 2017.
[34] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480–1494, Sep. 2015.
[35] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[36] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, Jan. 2012.
[37] R. A. Horn, "The Hadamard product," in *Proc. Symp. Appl. Math.*, vol. 40, pp. 87–169, May 1980.
[38] C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.
[39] I. E. Sobel, *Camera Models and Machine Perception*. Stanford, CA, USA: Stanford Univ., 1970, pp. 64–68.
[40] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[41] B. Yan, B. Yuan, and B. Yang, "Effective video retargeting with jittery assessment," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 272–277, Jan. 2014.
[42] K. Li, B. Yan, and B. Yuan, "A new metric to assess temporal coherence for video retargeting," *Proc. SPIE*, vol. 9273, Oct. 2014, Art. no. 92732Z.
[43] (2019). Video Retargeting Project. [Online]. Available: http://ee-cad.postech.ac.kr/Retargeting.egg.
[44] C. M. Maloney, M. A. Goheer, and J. P. Harvey, "Dimensional conversion in presentations," U.S. Patent 9 563 630, Feb. 7, 2017.

HO SUB LEE (S'14) received the B.S. degree in electrical and electronic engineering from Kyungpook National University, South Korea, in 2014, and the M.S. degree in electrical and electronic engineering from the Pohang University of Science and Technology, South Korea, in 2016, where he is currently pursuing the Ph.D. degree. His current research interests include image analysis, computer vision, and circuit design for display and multimedia systems.

GYUJIN BAE (S'13) received the B.S. degree in electronics engineering from Kyungpook National University, South Korea, in 2013, and the M.S. and Ph.D. degrees in electrical and electronic engineering from the Pohang University of Science and Technology, South Korea, in 2015. He is currently a Senior Researcher with LG Display, South Korea. His current research interests include image processing and computer vision.

SUNG IN CHO (S'10–M'17) received the B.S. degree in electronic engineering from Sogang University, South Korea, in 2010, and the Ph.D. degree in electrical and computer engineering from the Pohang University of Science and Technology, in 2015. From 2015 to 2017, he was a Senior Researcher with LG Display. From 2017 to 2019, he was an Assistant Professor of electronic engineering with Daegu University. He is currently an Associate Professor of multimedia engineering with Dongguk University, Seoul. His current research interests include image analysis and enhancement, video processing, multimedia signal processing, and circuit design for LCD and OLED systems.

YOUNG HWAN KIM (S'86–M'89–SM'14) received the B.E. degree in electronics from Kyungpook National University, South Korea, in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley, Berkeley, CA, USA, in 1985 and 1988, respectively. From 1977 to 1982, he was with the Agency for Defense Development, South Korea, where he was involved in various military research projects, including the development of autopilot guidance and control systems. From 1983 to 1988, he was a Postgraduate Researcher with the Electronic Research Laboratory, University of California at Berkeley, where he was involved in developing VLSI CAD programs. He is currently a Professor with the Division of Electronic and Computer Engineering, Pohang University of Science and Technology, South Korea. His current research interests include plasma and liquid crystal display systems, multimedia circuit design, MPSoC and GPGPU system design for display and computer vision applications, statistical analysis and design technology for deep-submicron semiconductor devices, and power noise analysis. He served as a Committee Member and a General Chair of various Korean domestic and international technical conferences, including International SoC Design Conference, IEEE ISCAS 2012, and IEEE APCCAS 2016. He has served as an Editor for the *Journal of the Institute of Electronics Engineers of Korea*.

SEOKHYEONG KANG received the B.S. and M.S. degrees in electrical engineering from the Pohang University of Science and Technology, Pohang, South Korea, in 1999 and 2001, respectively, and the Ph.D. degree from the VLSI CAD Laboratory, University of California at San Diego, San Diego, in 2013. He was with the System-on-Chip (SoC) Development Team, Samsung Electronics, Suwon, South Korea, from 2001 to 2008, where he was involved in development and commercialization of optical disk drive SoC. He has been an Assistant Professor with the Department of Electrical Engineering, Pohang University of Science and Technology, since 2018. His current research interest includes low-power design optimization and cost-driven methodology for chip implementation.