Electronic Imaging

SPIEDigitalLibrary.org/jei

Improved real-time video resizing technique using temporal forward energy

Daehyun Park Kanghee Lee Yoon Kim





Improved real-time video resizing technique using temporal forward energy

Daehyun Park Kanghee Lee Yoon Kim

Kangwon National University
Department of Computer and Communications Engineering
Hyoja 2-dong, Chuncheon Si, Gangwon-Do, Republic of Korea
E-mail: yooni@kangwon.ac.kr

Abstract. A novel video resizing algorithm that preserves the dominant contents of video frames is proposed. Because a visual correlation (a similarity) exists on consecutive frames within an identical shot, the energy distribution of the neighboring frames is also correlated and similar, and then the seams in a frame are analogous to those of neighboring frames. Thus, the seams of the current frame are derived from a specified range considering the coordinates of the seams of the previous frame. The proposed method determines the two-dimensional connected paths for each frame by considering both the spatial and temporal correlations between frames to prevent jittering and jerkiness. For this purpose, a new temporal forward energy is proposed as the energy cost of a pixel. The proposed algorithm has a fast processing speed similar to the bilinear method, while preserving the main content of an image to the greatest extent possible. In addition, because the memory usage is remarkably small compared with the existing seam carving method, the proposed algorithm is usable in mobile devices, which have many memory restrictions. Computer simulation results indicate that the proposed technique provides a better objective performance, subjective image quality, and content conservation than conventional algorithms. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JEI.22.1.013007]

1 Introduction

With the rapid development of wireless communication and mobile display devices, mobile multimedia have been available in many commercial areas, and among them, video contents have been considered an important form of information. In particular, since a variety of display devices such as tablet computers, cellular phones, smartphones, and handheld personal computers are released in the market, and such devices have different display resolutions, video needs to be changed according to different resolution sizes or aspect ratio of display panels. That is to say, the spatial resolution of video is downsized or upsized by resizing algorithms in order to use a video contents more effectively. However, because the simple resizing techniques such as scaling and cropping do not take into account the dominant contents (i.e., a person at the picture) in video, the primary transformation or distortion of such salient objects is inevitable. Therefore, it is necessary to

Paper 12047 received Mar. 20, 2012; revised manuscript received Nov. 28, 2012; accepted for publication Dec. 7, 2012; published online Jan. 7, 2013.

develop a new content-based video resizing method that can preserve the dominant contents in an image while changing its size.

Figure 1(b)–1(e) are 30% horizontally reduced images of Fig. 1(a). Because the scaling method adjusts the sampling rate uniformly over a whole image, if the scaling ratio is different from the aspect ratio of the source image, the contents of the source image are distorted [Fig. 1(b)]. Therefore, contentbased image resizing methods have been studied in order to prevent this visual distortion. Cropping is effective for displaying a region of interest (ROI) where dominant objects are located. Santella et al. proposed a semi-automatic cropping technique, which finds the important content and crops an image. However cropping-based methods discard the exterior region of ROI when the resolution of the target display is much smaller than that of original video and cannot correctly preserve the sparse multiple objects [Fig. 1(c)]. To solve this problem, Liu et al.² proposed the fisheye-view warping technique, that preserves the dominant region while other region is warped [Fig. 1(d)]. Fisheye-view warping preserves the main content of an image as much as possible but has the disadvantage of severely distorting the rest of the video information. Recently, Avidan et al.³ introduced the seam carving technique, which is known to have high image scaling performance with low quality loss of a retargeted image [Fig. 1(e)]. In order to resize images, this method removes or inserts pixels of a seam which is defined as a vertical (or horizontal) connected path of pixels with the minimum gradient energy. In addition, studies on additional methods to preserve the contents and change the size of an image are in progress.^{4–15}

In a video, the geometric transformations ^{16,17} of a content-based image are different from those in static images. For static images, each image is processed only in the spatial dimensions; in contrast, in a video, a consideration of the relationship between adjacent frames is needed because the concept of the time dimension is added. Without the consideration of the relationship, the contents of each frame's image can be preserved. However, the irregular movement of the contents' location in a video generates a shaking phenomenon (jitter) for the contents, because the connectivity of the time axis is lost. Therefore, it is essential to protect the time continuity of the contents to prevent this shaking phenomenon, which implies that a new content-based geometric conversion algorithm should be applied to videos.

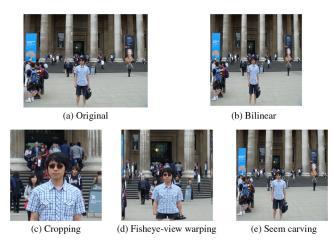


Fig. 1 Comparison of various methods to resize images.

There have been several classes of video retargeting approaches. Setlur et al. ¹⁶ generates a motion illustration by using a principal motion direction in video to detect and accentuate a moving object's motion in a single static frame. Liu et al. ¹⁷ performs a video retargeting using an automatic pan-and-scan method by moving the cropping window in each frame. Furthermore, additional studies on methodologies that can maintain the dominant contents and change the size of an image or a video are in progress. ^{4–15}

Video carving, 18,19 which is the application of seam carving to a video, uses a three-dimensional (3-D) cube to connect the frames to the time axis. Rubinstein et al. 18 introduces an improved seam carving algorithm for image and video retargeting, which applies forward energy instead of gradient value to evaluate the energy of a pixels. Chen et al. 19 proposed a video carving to handle two-dimensioanl (2-D) connected surface of pixels in 3-D space-time volume by constructed consecutive frames in video. However, since the location and geometric shape of contents are changed in the video frames, the 2-D connected surface considering spatial and temporal connectivity in whole video is not obtained simply. Therefore, in order to attain effective video retargeting, the entire 3-D space-time volume has to be analyzed while considering the energy in spatial and temporal connectivity of 2-D surface (Fig. 2). At this time, because 2-D connected surface is obtained by applying the graph cut technique^{20,21} that is required to a large amount of memory and high-complexity operations within the both Rubinstein's and Chen's methods, novel real-time image retargeting technique is required for a systems with limited resources, such as a mobile devices.

In this paper, a novel video resizing algorithm that preserves the dominant contents of video frames is proposed. The proposed method determines the 2-D connected paths for each frame by considering both the spatial and the temporal correlation between frames to prevent jitter and jerkiness with a reduced computational cost. Therefore, this method is performed in real-time and with low memory consumption.

The proposed technique operates by shot unit, which means that the consecutive images are taken by a single camera, and all of the frames within a shot have similar features. First, in order to separate each shot effectively in a video, a shot change is detected by monitoring the brightness

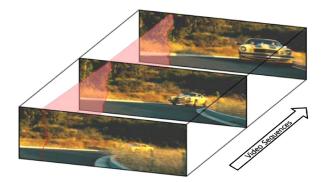


Fig. 2 Video carving.

differences and the histogram differences, which are susceptible to movement and color change, ^{22,23} respectively. If a shot change is generated and a new shot begins, the first frame of the shot is resized using the conventional seam carving technique on the static image. At this time, the seams extracted by the seam carving technique and the coordinates of the seams are stored. The proposed image resizing technique can calculate the new seams of the next frame in real-time by the newly proposed forward energy instead of creating a 3D cube which requires information on all of the video frames. And then the image resizing is carried out by the seams.

This paper is organized as follows. In the next section, the conventional seam carving algorithm is briefly introduced. The proposed algorithm is presented in Sec. 3. Section 4 presents and discusses the experimental results. Finally, our conclusions are given in Sec. 5.

2 Review on Conventional Seam Carving

The seam carving method extracts the seam of which the change of the energy is the lowest in the image, and controls the image size by adding or removing the pixel to the each coordinate of the seam. Seam is a line which is connected widthwise or lengthwise and composed of one pixel per a row and/or a column. In $W \times H$ image, the seam is defined as Eq. (1).

$$\mathbf{s}^{v} = \{ [\mathbf{X}(j), j] \}_{j=1}^{H}, \quad \text{s.t. } \forall j, \quad |\mathbf{X}(j) - \mathbf{X}(j-1)| \leq 1, \\
\mathbf{s}^{h} = \{ [i, \mathbf{Y}(i)] \}_{i=1}^{W}, \quad \text{s.t. } \forall i, \quad |\mathbf{Y}(i) - \mathbf{Y}(i-1)| \leq 1, \\
\mathbf{X} \colon \{ 1, \dots, H \} \to \{ 1, \dots, W \}, \\
\mathbf{Y} \colon \{ 1, \dots, W \} \to \{ 1, \dots, H \},$$
(1)

where s^v is the vertical seam, s^h is the horizontal seam, X and Y are the mapping functions for the row and column coordinates of the image, respectively. The column seam is the lengthwise connected coordinates set, and similarly the width seam is the widthwise connected coordinates set. In one image, several seams exist, and among them the optimum seam S^* required in the seam carving process is defined as Eq. (2).

$$e_{\min} = \min_{a \in S} [E(a)], \quad S = E^{-1}(e_{\min}),$$
 (2)

where S is a set of all seams obtained from one image, and $E(\bullet)$ is the cumulative energy function about one seam. That is, the optimum seam has the minimum energy value among

the whole seams in one image. The many operation quantities are required to calculate all seams in the image in order to find the optimum seam. The optimum seam is obtained by applying the dynamic programming technique^{24,25} in order to reduce these calculation quantities. The method finding the cumulative minimum energy map M, that is the first stage of the dynamic programming, by using the condition of the vertical seam of Eq. (1) and the matrix structure of image in $W \times H$ image shows up in Eq. (3).

$$M(i,j) = e(i,j) + \min[M(i-1,j-1),M(i,j-1),M(i+1,j-1)],$$

$$0 \le i < W, \quad 0 \le j < H,$$
(3)

where $e(\bullet)$ is the function finding the energy of the corresponding coordinates. The vertical cumulative minimum energy values are stored to the last row of M obtained by Eq. (3), and the vertical seam is found from each cumulative minimum energy values through the reverse search. The number of the vertical seams is identical with the horizontal size of the image since the number of the cumulative minimum energy values are like the horizontal size of the image. The optimum seam among the vertical seams is found through the reverse search from the pixel of which the cumulative minimum energy value is the smallest. The horizontal optimum seam can be found in the same way.

The image size can be controlled by adding or deleting video data on the coordinates of the optimum seam. Several seams are required in order to control the image size variously. After excluding the pixels corresponding to the seam which firstly is extracted in order to extract several seams, the next seam is extracted by the renewal of M. The reason for excluding the pixel corresponding to the previous seam coordinates in order to find the new seam is to satisfy the definition of the seams. The energy of the pixels comprising the optimum seam is low. Therefore, if the pixels of the already selected seam are not removed, the possibility that these pixels are again selected is high, and the overlapped pixels between the seams are generated, so the definition of the seam cannot be satisfied. If the definition of the seam is not satisfied, when converting the image size, the same pixel is repeatedly referred and the distortion of the result image is generated. Because the renewal of M is needed in order to prevent this distortion, the total processing time delay is inevitable. If the resolution of the image to adjust is big, the delay time increases exponentially.

3 Proposed Image Resizing Algorithm in Video

As shown in Fig. 3, the proposed real-time content-aware video resizing system is composed of three parts: shot change detection (SCD), generating seam, and image resizing (Appendix). If shot change is detected and a new shot is initiated, seam information stored of the previous frame is ignored and the new seams are searched using the seam carving technique on the static images. And then, after storing the information about the searched seam, the frame is resized to the target size. On the other hand, if a shot has been continued, the seams of the current frame are calculated by using the seam information stored of the previous frame, and then the frame is resized by generated seam.3.1.Detecting Shot Change.

Because the frame rate of a video is more than 10 fps, the shot change detection is performed every 10 frames. First, the feature values are extracted between two consecutive frames.

$$f_{i}(n) = \sum_{j=0}^{\text{height-1}} \sum_{i=0}^{\text{width-1}} |i_{n}(i,j) - i_{n-1}(i,j)|,$$

$$f_{h}(n) = \sum_{k=0}^{255} |h_{n}(k) - h_{n-1}(k)|,$$
(4)

where $i_n(i, j)$ is the (i, j)'th pixel value in the n'th frame, and f_i represents the brightness change susceptible to movement. In addition, $h_n(k)$ indicates the histogram of gray level k in the n'th frame, and the difference between h(k)s of consecutive frames is defined as f_h of the histogram change susceptible to color change.

For the stability of the algorithm, the shot change detection is not performed until 10 feature values are gathered. After 10 feature values are gathered, the largest and the second largest feature values are extracted and the difference between the two values is calculated. The shot change between two consecutive frames is detected through the following equations.

$$SCD = \begin{cases} 1, & \text{if } m_i > 3s_i \text{ and } m_h > 3s_h \\ 0 & \text{otherwise} \end{cases},$$

$$m_i = \max_{\alpha \in F_i}(\alpha), \quad s_i = \max_{\alpha \in F_i, \alpha \neq m_i}(\alpha),$$

$$m_h = \max_{\alpha \in F_h}(\alpha), \quad s_h = \max_{\alpha \in F_h, \alpha \neq m_h}(\alpha),$$

$$F_i = \{f_i(n-9), f_i(n-8), \cdots, f_i(n-1), f_i(n)\},$$

$$F_h = \{f_h(n-9), f_h(n-8), \cdots, f_h(n-1), f_h(n)\}, \quad (5)$$

where F_i and F_h are the sets of the feature values calculated on the previous 10 frames, and m_i and s_i is the largest and the second largest value within the set F_i , respectively Also, m_h and s_h is the largest and the second largest value within the set F_h , respectively. In the case where m_i and m_h are three times greater than s_i and s_h , respectively, it is determined that the shot change has happened. If a shot change is detected, as mentioned above, the shot change detection process is not performed until 10 new feature values are obtained.

Since the conventional seam carving for a static image is applied to the first frame after a shot change, the frequency of shot change has an effect on the speed of the algorithm. However, in the case of the general video, the scene change does not occur frequently as much as the real time processing is obstructed.

3.1 Deriving Seam in the First Frame

After a shot change is generated, the conventional seam carving for a static image is applied to the first frame. All the coordinate and energy values of the seams of the first frame are stored in order to use this information when finding seams in the next frame. The following equations show the stored information of the seams in frame of $W \times H$ size.

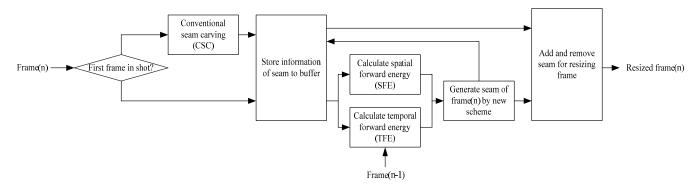


Fig. 3 Overall system block diagram.

$$S = [S_1, S_2, S_3, \cdots], \quad S_n = \{C, E\},$$

$$C = [x_0, x_1, \cdots, x_{H-1}] \quad \text{or} \quad [y_0, y_1, \cdots, y_{W-1}],$$

$$E = [ev_0, ev_1, \cdots, ev_{H-1}] \quad \text{or} \quad [eh_0, eh_1, \cdots, eh_{W-1}], \quad (6)$$

where the set S_n includes the information for one seam, and S is the array of S_n found in frame. The number of seams is determined by the target image size. S_n is comprised of the array C of the seam's coordinates and the array E of the energy in each coordinate of seam. The set C stores only x coordinates in case of the vertical seam or only y coordinates in case of the horizontal seam. W and H define the width and height of image, respectively. Seams are numbered in the ascending order of their energy values. The corresponding coordinate sets and energy values for each seam are stored systematically in the buffer.

3.2 Generating Seam of Current Frame by New Scheme

The seams of the current frame are extracted with reference to the seams information stored in the buffer when a shot change not occurs, that is the current frame belongs to the same shot as the previous frame. Because a visual correlation (a similarity) exist on consecutive frames within an identical shot, the energy distribution of the neighboring frames are also correlated and similar, and then the seams in a frame are analogous to those of neighboring frame. Thus the seams of the current frame are derived from specified range considering the coordinates of the seams of the previous frame because of correlation. At this time, the seams of temporal connection have to be considered. If the seams for each frame in video are generated independently without correlation, the jitter and jerkiness are occur. The visual artifact of jitter mainly occurs, in particular, because of a difference in the numbers of the seams around the dominant contents each frame. For example, assume that in the first frame, three seams and five seams were extracted from the left and right of some contents, respectively. And in the consecutive second frame, five seams and three seams were extracted from the left and right of the same contents of first frame, respectively. If the image size is changed identically for the two frames, the relative locations of the contents between the two frames have a difference of two pixels. This problem is jitter, which occurs on the contents of frame by repeating process of extracting seams independently for each frame. Figure 4 shows the results of independently expanding the size of the consecutive frames by the seam carving.

If we give attention to the picture in the red circle each frame in Fig. 4, we can observe that seven seams and one seam exist to the left and right of the red circle in the first frame, respectively, whereas six seams and two seams exist to the left and right of the red circle in the second frame, respectively. In the original video, the picture in the red circle exists in a fixed location. However, in the images expanded independently by the seam carving, the picture in the second frame moves one pixel to the left compared to the first frame. If these processes are repeated, the contents in the red circle shake tremendously.

Therefore, in a video, preventing the shaking phenomenon is more important than finding the optimum seam. This section presents a new process to extract seams that prevents the shaking phenomenon and preserves the form of the dominant content.

3.2.1 Seam-ordering of current frame

Since seams of frame can be overlapped, the conventional seam carving extracts the next seam after removing the previous seam. Figure 5 shows the overlapped coordinate between the first seam and the second seam.

In Fig. 5, overlapped coordinates are generated at the location where the first seam and the second seam meet. If the coordinate of the overlapped part is used when the

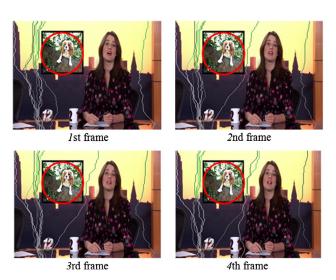


Fig. 4 Independent seam carving result for each frame.

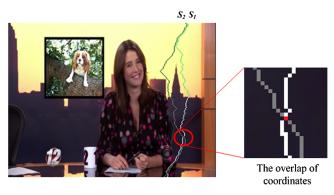


Fig. 5 Order of seams and example of coordinate overlap.

image size is modified by the seams, it will be incorrect by one pixel at the location of the overlap. In conclusion, a distortion of the image occurs. Therefore, a specific order is used for the seams. That is, the seam order of the current frame is identical to that of the previous frame. For example, the information of the 4th seam of the previous frame is stored in order to get the 4th seam of the current frame. Equation (7) indicates that the seam information of the previous frame is referred in order to produce the seam of the current frame.

$$S_{\text{ref}} = S_i(n-1), \tag{7}$$

where S_{ref} has the same structure as S_n in Eq. (4), and is the reference to produce the new seam. Also, n is the number of the current frame, and i is the number of the current seam.

3.2.2 Energy cost of pixel

The conventional seam carving method considers the energy of each pixel to determine a seam, and there exist the various energy functions. The amount of change of the pixel value, the spatial forward energy, the standard deviation, the edge information, ²⁶ gradient vector flow, ²⁷ the energy of high tasks (e.g., face detector), etc., can be used as the energy, and the other result image is generated according to each energy function. Among them, the spatial forward energy

having the good performance uses the difference between adjacent pixels of a pixel. If the pixel is selected as a seam and removed, the adjacent pixels are smoothly connected. The spatial forward energy is defined as Eq. (8).

$$SFE_{left-up}(i, j) = |p(i+1, j) - p(i-1, j)| + |p(i, j-1)| - p(i-1, j)|,$$

$$SFE_{up}(i, j) = |p(i+1, j) - p(i-1, j)|,$$

$$SFE_{right-up}(i, j) = |p(i+1, j) - p(i-1, j)| + |p(i, j-1)| - p(i+1, j)|,$$
(8)

where SFE(\bullet) is the spatial forward energy according to the position of the pixel to be removed, and p(i,j) is the (i,j)'th pixel value. Equation (8) is used to find the vertical seam, and the horizontal seam is obtained by the same method. In calculating SFE(\bullet), one among the left-up, up, and right-up is selected only for the pixels of which the spatial connectivity is maintained.

The spatial forward energy shows the good performance about the static images, but not about the videos because the correlation between frames is not considered. In this paper, the temporal forward energy is proposed as the energy considering the correlation between frames. The temporal forward energy can guarantee the continuity of the seam in the time domain.

Figure 6 shows the three possible vertical seam by temporal forward energy, and p(i, j, n) is the (i, j)'th pixel value in the n'th frame. As shown in Fig. 6, we search for the seam whose removal inserts the minimal amount of energy between two consecutive frames. These are seams that are not necessarily minimal in their energy, but will leave less artifacts in the resulting image, after removal. This coincides with the assumption that two neighboring images have piecewise smooth intensity at the same position of the pixel, which is a popular assumption in the literature. The temporal forward energy according to the position of the pixel to be removed is defined as Eq. (9).

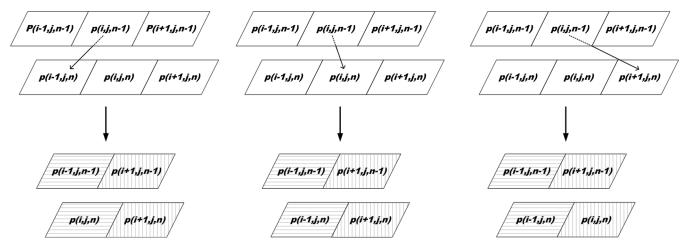


Fig. 6 The three possible vertical seam by temporal forward energy.

$$\begin{aligned} \text{TFE}_{\text{left-down}}(i,j) &= |p(i-1,j,n-1) - p(i,j,n)| \\ &+ |p(i+1,j,n-1) - p(i+1,j,n)|, \\ \text{TFE}_{\text{down}}(i,j) &= |p(i-1,j,n-1) - p(i-1,j,n)| \\ &+ |p(i+1,j,n-1) - p(i+1,j,n)|, \\ \text{TFE}_{\text{right-down}}(i,j) &= |p(i-1,j,n-1) - p(i-1,j,n)| \\ &+ |p(i+1,j,n-1) - p(i,j,n)|. \end{aligned} \tag{9}$$

In calculating $TFE(\bullet)$, one among the left-down, down, and right-down is selected only for the pixels of which the temporal connectivity are maintained.

3.2.3 Generating seam of continuous frames

The coordinates of the pixels which are temporally connected with the reference seam of the previous frame are selected as the starting coordinates of a seam. The p is set of coordinate of seam candidate. The next coordinate p_{n+1} of p_n is obtained with reference to p_n and the reference seam S_{ref} . The condition to find p_{n+1} is given by

- 1. p_n and p_{n+1} are spatially connected (spatial connection).
- 2. p_{n+1} and $C(\subset S_{ref})$ are connected to the time axis (temporal connection).

Equation (10) is the process of finding the candidate pixel (CanPix) satisfying the above condition.

$$\begin{aligned} & \operatorname{CanPix} = \operatorname{SPA} \cap \operatorname{TEM}, \\ & \operatorname{SPA} = \{x | p_n - 1 \le x \le p_n + 1\}, \\ & \operatorname{TEM} = \{x | \textbf{\textit{C}} \subset \textbf{\textit{S}}_{\mathrm{ref}}, \textbf{\textit{C}}(n+1) - 1 \le x \le \textbf{\textit{C}}(n+1) + 1\}, \end{aligned}$$

where n is x coordinate (horizontal seam) or y coordinate (vertical seam) of p_n . SPA and TEM is the pixel set satisfying the spatial connection and the temporal connection, respectively. That is, SPA includes the adjacent pixels to p_n , and SPA includes the adjacent pixels to $C(\subset S_{\text{ref}})$. Figure 7 shows an example of the spatial connection condition, temporal connection condition, and the set CanPix satisfying two conditions.

The set Canpix is composed of the pixels satisfying the spatial connection and temporal connection altogether and the pixels becomes the candidate for the seam guaranteeing the continuity in the time domain. The spatial forward energy and the temporal forward energy of the candidate pixels are

obtained, and the pixel with the smallest sum of the two energy values is included in the seam as Eq. (11).

$$p_{n+1} = \underset{\alpha \in \mathbf{CanPix}}{\arg\min}[\mathbf{SFE}(\alpha) + \mathbf{TFE}(\alpha)], \tag{11}$$

where $SFE(\bullet)$ is the function finding the spatial forward energy, $TFE(\bullet)$ is the function finding the temporal forward energy. The seam which guarantees the spatial connectivity and the temporal connectivity can be obtained by Eqs. (8), (9), and (11), and therefore, the proposed technique resizes the video without distortion of the primary contents and visual artifacts.

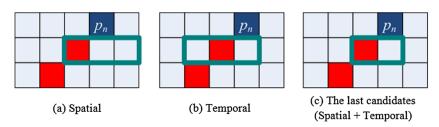
3.3 Image Resizing

The image size is modified by the coordinates of all the seams that are finally determined in the current frame. When reducing the image size, as many seams as the difference in size between the original video and target video are removed in the order of the seams, one at a time. On the other hand, when expanding the image size, pixel values are inserted to the coordinates of the seams in the order of the seams. Figure 8 shows examples of the process to control the image size. First, a seam map is generated by the coordinates of seams in seam information stored. The size of the seam map is identical to that of the original image, and the corresponding seam numbers are stored with the coordinates of the seams as shown in Fig. 8(a). The image size is controlled by the produced seam map. When reducing the image size, as shown in Fig. 8(b), the seam map is searched and the pixels with the coordinates of the first seam are removed. After the size of the image is reduced by one seam, in order to update the coordinates by removed seam, the referred seam is removed from the seam map. The image size is reduced by repeating this process for the number of seams.

On the other hand, when the image size is enlarged, as shown in Fig. 8(c), empty spaces are inserted at the same coordinates as the coordinates of a seam. In addition, the pixel values generated by an interpolation method are filled in the empty spaces, and the image size is expanded. After the size of the image is expanded by one seam, in order to update the coordinates by inserted seam, the referred seam is inserted in the seam map. The target image is obtained by repeating this process.

4 Experimental Results

In this section, the performance of three image resizing techniques are evaluated, namely, the bilinear method, the technique of applying Avidan's algorithm³ to a video, and the



(10)

Fig. 7 Example of coordinate candidates.

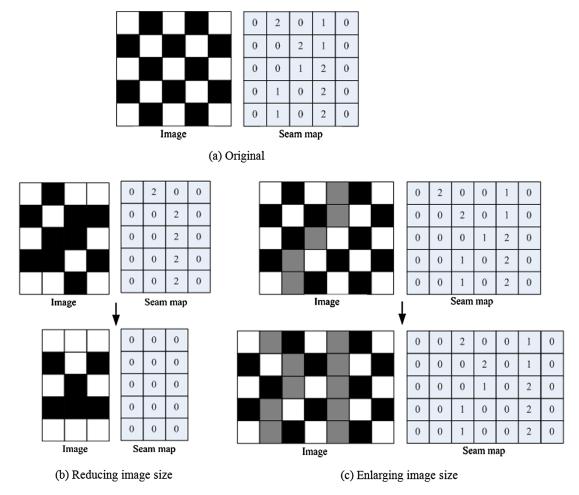


Fig. 8 Examples of image resizing process.

proposed technique. Extensive experimental testing and comparison were performed on several sequences with different characteristics: "SOCCER," "COASTGUARD," and "MOTHER & DAUGHTER" are in CIF format (352 × 288 pixels), and "IN TO TREE" are in 720p format (1280 × 720 pixels). All sequences have 300 frames, and were horizontally enlarged by 30%. First, each method was evaluated on the basis of its runtime and the average memory usage, which are the most important factors in real-time processing. The experiments were performed in the 1.86 GHz dual core with 2 GB memory. In order to enhance the reliability in the measured value, the same process was repeated 10 times, and the averages of the result values were compared.

Table 1 Run-times for different algorithms (s).

Algorithm	352×288 pixels	1280×720 pixels
Bilinear	16.266	49.353
Avidan's	582.541	2671.998
Proposed	23.761	110.562

Tables 1 and 2 show the runtime and the average memory usage of each algorithm, respectively.

As the Avidan's algorithm needs many operations and the large storage space in order to analyze the entire frames in video, it cannot be performed on a system with limited resource such as a mobile terminal. However, the proposed algorithm runs about 25 times faster than the Avidan's algorithm and achieves the comparable runtime as compared with the bilinear method as shown in Table 1. Since the proposed algorithm can process 12 frames per second in case of CIF, real-time processing is possible for systems with a frame rate of 12 frames per second.

Since the proposed algorithm is designed for mobile terminal, the memory usage is also important. As shown

Table 2 Memory usages for different algorithms (KB).

Algorithm	352×288 pixels	1280×720 pixels
Bilinear	637.8	6749.6
Avidan's	2455.9	32252.8
Proposed	831.3	13842.3

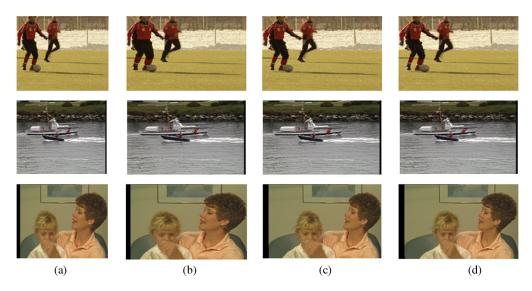


Fig. 9 Subjective quality comparison for the different algorithms: (a) original; (b) bilinear method; (c) Avidan's method; (d) proposed method.

Table 3 Error rates for different algorithms.

Algorithm	352×288 pixels	1280×720 pixels
Avidan's	67.3	82.1
Proposed	5.7	13.3

in Table 2, the proposed method requires lower memory about three times than the Avidan's algorithm. Because the new seam of the current frame is computed with reference to the seam information of the previous frame, the memory usage of the proposed method is similar to that of the bilinear method which is usually performed to resize image on mobile device.

Next, whether the main content was maintained and whether the shaking phenomenon exits or not were compared through each result frame. Figure 9 shows "SOCCER" (174th frame), "COASTGUARD" (62th

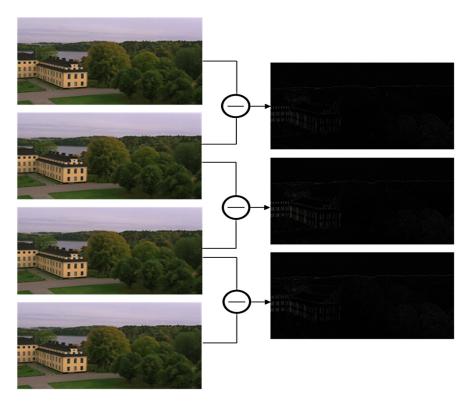


Fig. 10 Differences between adjacent frames in original video.

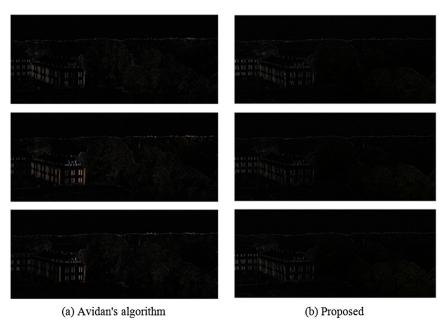


Fig. 11 Differences between adjacent frames after applying Avidan's and proposed algorithm.

frame), and "MOTHER & DAUGHTER" (60th frame) from the results of each algorithm.

Compared to the source image in Fig. 9(a), the result of the bilinear technique in Fig. 9(b) indicates that the shapes of the primary contents have been broadened. However, in the images results from Avidan's algorithm and the proposed algorithm, the shapes of the contents are similar to those in the original image. Thus, it is seen that the proposed algorithm maintains the main content of the image.

Finally, the differences between the experimental results and source image are shown as the Error Rate given by

$$D_n = \sum_{j=0}^{\text{height}-1} \sum_{i=0}^{\text{width}-1} |f_n(i,j) - f_{n+1}(i,j)|,$$

$$\zeta = \frac{1}{\text{height} \times \text{width}} \sum_{k=1}^{K-1} D_k, \text{Error Rate} = \left|1 - \frac{\zeta_0}{\zeta}\right| \times 100,$$
(12)

where f_n indicates R, G, and B values of the n'th frame, and D_n shows the error per pixel between n'th frame and (n+1)'th frame. K is the number of total frames, and ζ_0 is the error between frames in the original video. Error Rate represents the difference between original video and the result video. The Table 3 shows numerically how many differences the result images by the proposed method and the Avidan's method shows with the original video by Error Rate.

As shown in the Table 3, the result images by the proposed method have the smaller error rate and are more similar to the original video than those of the Avidan's method.

Figure 10 shows the differences between adjacent frames in "IN TO TREE" (frames 33–36). Because these frames belong to a single shot, any differences between adjacent frames are small.

As shown in Fig. 11(a), because the technique applying Avidan's algorithm to video does not consider the relation between adjacent frames, the shaking phenomenon occurs and many differences between neighboring frames are generated. On the other hand, because the proposed algorithm considers the correlation between adjacent frames, there is no shaking phenomenon and the differences between neighboring frames are similar to those in original video as shown in Fig. 11(b).

The results have been presented only for horizontal direction. In order to control the image size in both directions, the proposed algorithm is just applied twice: once in the horizontal direction and once in the vertical direction.

5 Conclusion

A novel video resizing algorithm that preserves the dominant contents of video frames was proposed. Because a visual correlation (a similarity) exist on consecutive frames within an identical shot, the energy distribution of the neighboring frames are also correlated and similar, and then the seams in a frame are analogous to those of neighboring frame. Thus, the seams of the current frame are derived from specified range considering the coordinates of the seams of the previous frame because of correlation. The proposed method determines the 2-D connected paths for each frame by considering both the spatial and temporal correlations between frames to prevent jitter and jerkiness. The conventional seam carving requires too much complexity and a large amount of memory because the entire frames in video have to be analyzed. Therefore, the conventional seam carving cannot be performed on a system with mobile terminal. The proposed algorithm has a fast processing speed similar to that of the bilinear method, while preserving the main content of an image to the greatest extent possible. In addition, because the memory usage is remarkably small compared with the existing seam carving method, the proposed algorithm is usable in mobile terminals which have limited memory resources. Computer simulation results indicate that the proposed technique provide better objective performance, subjective image quality, shaking phenomenon removal, and content conservation than conventional algorithms.

Appendix

Pseudocode of the framework of the proposed algorithm:

```
F = \text{number of frame}
N = number of seam
for (f = 1; f <= F; f + +)
  perform shot change detection
  for (n = 1; n \le N; n + +)
  if f is first frame or shot change occurred
    calculate SFE to pixel of frame without including seam extracted
    extract one seam using dynamic programming on SFE
    update and accumulate seam information
  else
    calculate SFE and TFE to pixel satisfying spatial and temporal
    connectivity
    for the nth seam of previous frame
    generate one seam considering SFE, TFE value and the location
    of seam of
    previous frame
    update and accumulate seam information
  }
create new resizing frame to use seam information
```

Acknowledgments

This study was supported by the Research Grant from Kangwon National University.

References

- 1. A. Santella et al., "Gaze-based interaction for semi-automatic photo
- Tr. Salica Cettal, State State Intertaction for State and Action Coropping," in Proc. SIGCHI Conf. Human Factors Comput. Syst., pp. 771–780, ACM, New York (2006).
 F. Liu and M. Gleicher, "Automatic image retargeting with fisheye-view warping," in Proc. ACM Symposium on User Interface Software and Technology, pp. 153–162, ACM, New York (2005).
- 3. S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," ACM Trans. Graph. 26(3), 1-9 (2007).
- S. Battiato et al., "Content-based image resizing on mobile devices," in Int. Conf. Comput. Vis. Theory App. (VISAPP), Rome, Italy, pp. 87-90 (2012).

- C. Tao, J. Jia, and H. Sun, "Active window oriented dynamic video retargeting," in *ICCV Proc. Workshop Dynamic. Vis.*, pp. 1–12 (2007).
 S. Cho et al., "Image retargeting using importance diffusion," in *Proc.*
- IEEE Int. Conf. on Image Process., pp. 977-980, IEEE, Cairo (2009).
- L. Chen et al., "A visual attention model for adapting images on small displays," *Multimedia Syst.* 9(4), 353–364 (2003).
- 8. H. Liu et al., "Automatic browsing of large pictures on mobile devices," in Proc. Eleventh ACM Int. Conf. Multimedia, pp. 148-155, ACM, New York (2003).
- 9. B. Suh et al., "Automatic thumbnail cropping and its effectiveness," Proc. 16th Ann. ACM Symp. User Interface Software Technology, pp. 95–104, ACM Press, New York (2003).
- X. Fan et al., "Looking into video frames on small displays," in Proc. Eleventh ACM Int. Conf. Multimedia, pp. 247-250, ACM, New York
- 11. J. Xiao et al., "A novel adaptive interpolation algorithm for image resizing," Int. J. Innov. Comput. Inform. Cont. 3(6(A)), 1335-1345 (2007).
- Y. Zhang et al., "Application of a bivariate rational interpolation in image zooming," *Int. J. Innov. Comput. Inform. Cont.* 5(11(B)), 4299–4307 (2009).
- 13. P. Lin, H. Chu, and T. Lee, "Smooth shape interpolation for 2D polygons," Int. J. Innov. Comput. Inform. Cont. 4(9), 2405-2417 (2008).
- 14. Y. Tian et al., "An iterative hybrid method for image interpolation," in Proc. Int. Conf. on Intelligent Computing (ICIC), Vol. 1, pp. 10-19, Springer, Berlin, Heidelberg (2005).
- 15. J. Xiao et al., "Adaptive interpolation algorithm for real-time image resizing," in *Proc. Innov. Comput. Inform. and Control (ICICIC)*, Vol. 2, pp. 221–224 (2006).
- V. Setlur et al., "Retargeting images and video for preserving informa-tion saliency," *IEEE Comput. Graphics Appl.* 27(5), 80–88 (2007).
- 17. F. Liu and M. Gleichar, "Video retargeting: automating pan and scan," in *Proc. ACM Int. Conf. Multimedia*, pp. 241–250, ACM, New York (2006).
- M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.* 27(3), 1–9 (2008).
 B. Chen and P. Sem, "Video carving," in *Short Papers Proc. Eurographics* (2008).
- P. Kohli and P. H. S. Torr, "Dynamic graph cuts for efficient inference in Markov random fields," *IEEE Trans. Pattern Anal. Mach. Intel.* 29(12), 2079-2088 (2007)
- V. Kwatra et al., "Graphcut textures: image and video synthesis using graph cuts," ACM Trans. Graph. 22(3), 277–286 (2003).
 U. Gargi, R. Kasturi, and S. H. Strayer, "Performance characterization of video-shot-change detection methods," IEEE Trans. Circuits Sys. Video Technol. 10(1), 1–13 (2000).
- 23. Y. Gong, "An accurate and robust method for detecting video shot boundaries," in *Proc. IEEE Int. Conf. on Multimedia Comput. and Sys.*, Vol. 1, pp. 850–854, IEEE, Florence (1999).
 24. R. Bellman, "Some problems in the theory of dynamic programming,"
- *Econometrica* **22**(1), 37–48 (1954).
- 25. D. P. Bertsekas, Dynamic Programming and Optimal Control, Vol. II, 3rd ed., Athena Scientific (2007).
- 26. K. S. Choi and S. J. Ko, "Fast content-aware image resizing scheme in the compressed domain," ACM Trans. Consum. Electron. 55(3), 1514–1521 (2009).
- 27. S. Battiato et al., "Content-aware image resizing with seam selection based on gradient vector flow," in *Proc. Int. Conf. on Image* Processing (ICIP) (2012).



Daehyun Park received BS and MS degrees in computer engineering with the Department Computer and Communications Engineering from Kangwon National University in 2007 and 2009, respectively. He is now a PhD candidate in computer engineering with the Department of Computer and Communications Engineering at Kangwon National University. His research interests are in the areas of video signal processing and multimedia communications.



Kanghee Lee received BS and MS degrees in computer engineering with the Department Computer Communications and Engineering from Kangwon National University in 2009 and 2011, respectively. His research interests are in the areas of video signal processing and multimedia communications.



Yoon Kim received a BS degree in 1993, an MS degree in 1995, and a PhD degree in 2003, in electronic engineering with the Department of Electronic Engineering from Korea University. In 2004, he joined the Department of Computer and Communications Engineering, Kangwon National University, where he is currently an associate professor. From 1995 to 1999, he was with the LG-Philips LCD Co., where he was involved in research and

development on digital image equipment. His research interests are in the areas of video signal processing, multimedia communications, and wireless sensor networks.