SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 2015; **8**:2102–2113
Published online 24 November 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.1158

RESEARCH ARTICLE

Detection of seam carving-based video retargeting using forensics hash

Wei Fei[1], Yang Gaobo[1]*, Li Leida[2], Xia Ming[1] and Zhang Dengyong[1]

[1] School of Information Science and Engineering, Hunan University, Changsha, 410082, China
[2] School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, 221116, China

ABSTRACT

Seam carving is a content-aware multimedia retargeting technique to adaptively resize multimedia data for different display sizes. However, it can also be used to remove objects from digital object or video for malicious purposes. In this paper, a forensics hash-based tampering detection and localization approach is proposed for seam carving-based video retargeting. It extracts the invariant Speeded-up Robust Feature points from every spatiotemporal image to represent the matching surface, and the relative position change of the neighboring matching surface is used to build the forensic hash in a compact and scalable way. Experimental results show that the proposed forensics approach can effectively estimate the exact amount and rough locations of deleted seam carving surfaces. It achieves desirable detection performance even when there are frames deleted. If the hash length is reasonably increased, it can estimate the rough location and exact amount of deleted frames. Moreover, the built forensics hash is of good robustness, scalability, and compactness. Copyright © 2014 John Wiley & Sons, Ltd.

KEYWORDS
video forensics; seam carving; video re-targeting; forensics hash; SURF

***Correspondence**
Yang Gaobo, School of Information Science and Engineering, Hunan University, Changsha, 410082, China.
E-mail: yanggaobo@hnu.edu.cn

1. INTRODUCTION

With the popularity of low-cost digital cameras, digital image and video are proliferating in our daily life. However, the large availability of image/video editing software tools makes their tampering extremely simple. Large amounts of doctored images and videos are spread over the Internet without obvious traces. Our traditional opinion that seeing is believing is no longer true. This leads to an increasing demand for automatic forgery detection to determine the trustworthiness of candidate image and video [1,2]. This is the so-called digital media forensics, which is generally classified into two categories: active and passive. Active forensic techniques use auxiliary data by embedding digital watermark or extracting digital signature in advance, whereas passive or blind forgery detection techniques simply uses the candidate image or video. The idea behind passive forensics is that although digital forgeries may leave no visual clues, it is high likely that they may disturb the underlying statistical properties or cause some artifacts in various forms of inconsistencies [3].

In the 34th International Conference and Exhibition on Computer Graphics and Interactive Techniques, seam carving was firstly presented for content-aware image resizing [4]. Because of its excellent performance, it has gained researchers' continuous attention [5,6]. It is incorporated into the most popular image-editing software Photoshop CS4 as an individual function, which is referred as content-aware scaling. In recent years, seam carving has also been extended to content-aware video retargeting by considering both spatial and temporal coherence [7–9]. However, seam carving can also be used for malicious tampering such as object removal. Especially, the content-aware mechanism of seam carving makes it preserve well the perceptually important contents without leaving any visually noticeable artifacts. This brings great difficulty for forensic analyst to determine whether an image or video has undergone seam carving. Figure 1 shows two examples of malicious object removal from digital image. When an object is deleted from a digital image, it might have a direct influence on the content of the digital image that it conveys. Therefore, the forensics of seam carving is a challenging but interesting topic in the field of information security.

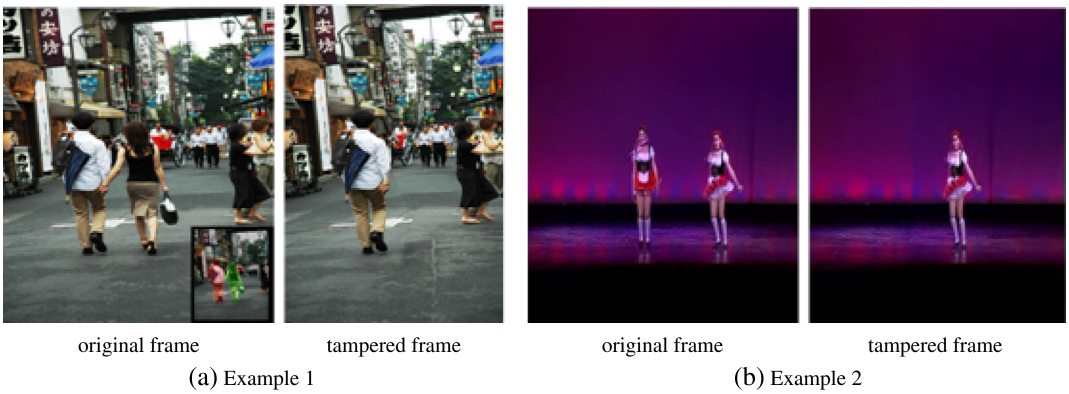


Figure 1. Two examples of malicious object removal by content-aware video retargeting. (a) Example 1. (b) Example 2.

In the literature, there are a few forensic approaches to detect the presence of seam carving. Sarkar *et al.* proposed a machine learning-based framework to distinguish between seam-carved (or seam-inserted) and normal images [10]. 324D Markov features, consisting of 2D difference histograms in the block-based discrete cosine transform domain, are used to train a classifier. It yields a detection accuracy of 80% and 85% for seam carving and seam insertion, respectively. Ryu *et al.* propose a detection method for content-aware image resizing by exploiting the energy bias of seam-carved images [11]. The correlation between adjacent pixels is also analyzed to estimate the inserted seams. These two forensic approaches can effectively make binary decision about whether an image has been resized by seam carving. However, it does not provide further information such as the amount or location of seams deleted or added. This is in fact the inherent limitation of passive forensics because it does not use any *a priori* information or auxiliary data. To answer a broader scope of forensic questions, a new concept of forensic hash is introduced [12]. In essence, a forensic hash is the robust features extracted from the original image but properly designed and compressed for forensic purposes. In Lu's work [12], a very compact forensic hash of around 50 bytes can reliably estimate both the amount and the location of seam carving, and further enable accurate alignment and tampering localization on a modified image.

Image seam carving is extended to video retargeting by considering the temporal coherence constrains. The existing seam carving-based video retargeting (SCVR) approaches [7–9] can also be exploited for malicious purposes such as object removal. Although there are many active and passive approaches for the detection of image seam carving, no work is reported for the forensic analysis of SCVR. In this paper, we are motivated by the works in [12,13] to present an active detection approach for SCVR using the forensic hash. By analyzing the possible traces left by typical SCVR algorithms, a compact and scalable forensic hash component is built to estimate the amount and location of seams deleted or inserted. The contributions of this work are twofold. First, because the proposed forensic hash is an active approach, it can provide not only the binary decision about whether a video has suffered from SCVR but also more information about the SCVR tampering such as the amount and location of seams involved in SCVR. Second, if the hash length is appropriately increased, it can also estimate the rough location and exact amount of frame-based manipulation. To the best of our knowledge, there are no similar works reported in the literature.

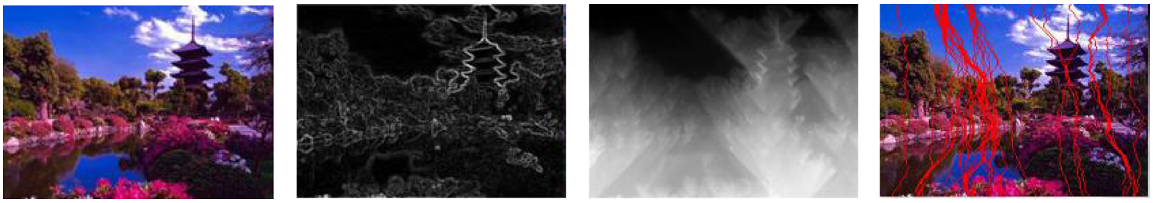
The rest of the paper is organized as follows. Section 2 briefly introduces related work. In Section 3, the proposed forensic hash approach is discussed in detail. Experimental results and analysis are given in Section 4, and Section 5 concludes this paper.

2. RELATED WORK

2.1. Typical SCVR approaches

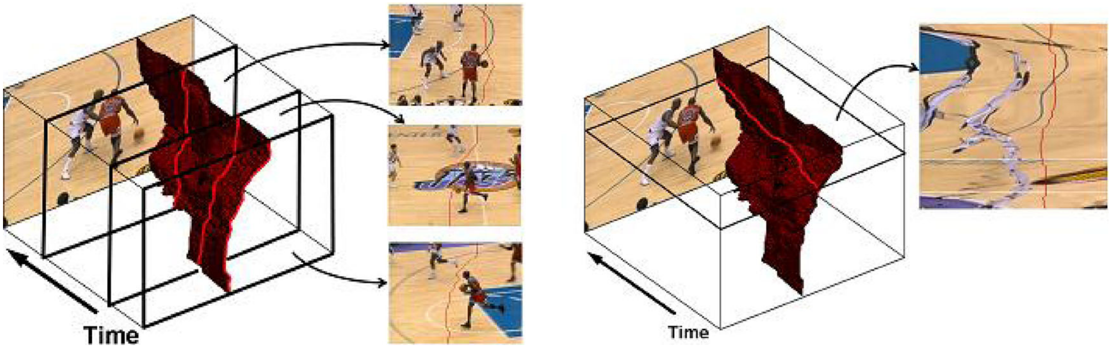
Seam carving is originally proposed for image resizing. A seam is defined as an eight-connected path of low-energy pixels crossing the image from top to bottom or from left to right. A dynamic programming technique is used to select the optimal seams in each direction, which are defined as the seams with the lowest accumulated energy. As shown in Figure 2, the energy map reflects the combined importance of all the pixels along a seam. SCVR is an extension of image seam carving by considering another temporal dimension. Its basic principle is similar to image seam carving, but SCVR is performed on a 3D video data, with temporal dimension as the z-axis. The energy map used is actually a combination of both the spatial and temporal energies. To avoid the visual distortion or discontinuousness of video content such as jittery artifacts, SCVR usually enforces both spatial and temporal coherence constrains for the energy function. In the following, three typical SCVR techniques are summarized.

The earliest SCVR technique is a straightforward extension of image seam carving to surface carving for digital video [7]. As shown in Figure 3, this extension defines 2D surfaces to be removed from a 3D video cube. And a



(a) The original image (b) energy map (c) accumulated energy (d) vertical seam

Figure 2. An example of seam carving for image resizing. (a) The original image. (b) Energy map. (c) Accumulated energy. (d) Vertical seam.



(a) the seam in every frame (b) the spatio-temporal seam surface

Figure 3. An example of carving-based video retargeting. (a) The seam in every frame. (b) The spatiotemporal seam surface.

temporal coherence constrain is introduced into the definition of energy function. Let video sequence be $\{I_t\}_{t=1}^{N}$, where N is the number of video frames and t is the frame index; the spatial and temporal energy function are defined as follows:

$$E_{\text{spatial}}(i,j) = \max_{t=1}^{N} \left\{ \left| \frac{\partial}{\partial_i}(i,j) \right| + \left| \frac{\partial}{\partial_j}(i,j) \right| \right\} \tag{1}$$

$$E_{\text{temporal}}(i,j) = \max_{t=1}^{N} \left\{ \left| \frac{\partial}{\partial_t}(i,j) \right| \right\} \tag{2}$$

The global energy function is defined by a linear combination of spatial and temporal energies. Apparently, $\alpha \in [0,1]$ is a parameter that balances spatial and temporal contributions.

$$E_{\text{global}}(i,j) = \alpha E_{\text{spatial}}(i,j) + (1-\alpha) E_{\text{temporal}}(i,j) \tag{3}$$

Matthias *et al.* proposed a discontinuous seam-carving scheme for video retargeting [8]. Different from geometrically smooth and continuous seams, an appearance-based temporal coherence is formulated to define temporally discontinuous seams. Moreover, a piecewise spatial seam is introduced based on the variation in the gradient of the intensity. For the retargeted video, more details are preserved, and a better visual quality is achieved. Yang Bo *et al.* proposed a matching-area-based SCVR technique [9]. The novel matching-area-based temporal energy adjustment allows the seam to track the object it previously carved and avoids carving the seam on different objects in two consecutive frames to achieve better spatial and temporal coherence.

The three aforementioned SCVR techniques have their own advantages and disadvantages. However, they share some common similarities. First, because video sequence is considered as a 3D array, they all achieve video retargeting by extending the definition of energy function. Second, to avoid the visual distortions such as jittering, the removal of optimal surfaces with the least energies in the array is enforced with constrains of both temporal and spatial coherence. Finally, because of the content-aware mechanism in SCVR, the retargeted video can preserve the content well. This implies that there must be some invariant features before and after video retargeting. As a result, we do not emphasize the specific tools and parameters used in these SCVR techniques. Instead, we are motivated to make use of the local invariant features, especially their relative position change to construct the forensic hash.

2.2. Forensic hash approaches for image resizing

The idea behind passive image forensics is to analyze the detectable changes specific to image tampering [10,11] or the intrinsic traces left by devices. It can only provide a simple binary decision of authenticity, which is inadequate for the purpose of image forensics. To answer a broader range of questions regarding the processing history of image tampering, a novel conception of forensic hash is recently proposed [12]. Later, Wang *et al.* propose an image forensic signature for content authenticity analysis [13]. Discrete wavelet transform and adaptive Harris operator are comprehensively used to extract image feature points; then the statistics of feature point neighborhood are used to construct forensic signature. Based on this forensic signature, a search-based forensic analysis method is proposed for analyzing the processed history of the received image, including geometric transform estimation, tampering detection, and tampering localization. It is straightforward to understand that the concept of a forensic signature in [13] is very similar to the forensic hash in terms of functionality [12,14]. Moreover, it is also a compact representation of robust features that are properly selected from the original image.

In essence, the forensic hash is a robust feature representation generated by properly selecting the local statistics of robust feature points, which are extracted from the original image as side information. Apparently, a key issue for forensic hash construction is the selection of robust features from the original image that are robust to specific tampering. In addition, because the forensic hash is required to be transmitted to the receiver as a side information, its compactness and scalability are also preferable.

3. PROPOSED FORENSIC HASH APPROACH FOR SCVR

3.1. Construction of forensic hash

A scale-invariant feature transform (SIFT) is widely used in computer vision because it is a stable and distinctive feature. In the forensic hash scheme for image seam carving [12], SIFT feature points in an image are utilized to build the forensic hash. Because a SIFT feature is invariant to image translation, scaling, rotation, and illumination change, it is partially robust to local geometric transform. The scale and dominant orientation of a SIFT point can be used to estimate geometric transforms such as rotation and scaling. It is reported that a very compact forensic hash of around 50 bytes can reliably estimate both the amount and location of seams involved in seam carving and further enable accurate alignment and tampering localization on a tampered image.

Because digital video is actually a series of still images, it can be naturally treated as a 3D cube, where each frame is a surface in the cube. As shown in Figure 2, the seam carving on 2D images is directly extended to the seam surface in SCVR. Therefore, we are motivated to extend the matching of feature points in [12] to feature surfaces for the detection of SCVR. Specifically, the seam carving is extended from a 1D path in a 2D image to a 2D surface in 3D video data. The relative displacement of stable feature points between two matching surfaces is used to construct the forensic hash for the estimation of seams involved in SCVR. The construction of the forensic hash is detailed as follows.

First, the Speeded-up Robust Feature (SURF) is chosen as the invariant local feature, instead of SIFT in [12]. Luo *et al.* made a lot of experimental comparisons among SURF, SIFT, and principal component analysis SIFT [15]. It is claimed that SURF has better robustness and more efficient processing speed, although it is inferior to SIFT in terms of stability to rotation and illumination changes. For digital video, processing, speed is usually a major concern because of its huge data amount. Moreover, rotation and illumination change are not the main concerns for the detection of SCVR. Therefore, SURF is more suitable for the detection of SCVR than SIFT because it can improve detection efficiency.

As shown in Figure 4, all the pixels in the digital video are treated as a 3D array, where X and Y are the horizontal and vertical axes and *Time* is the temporal axis. All the SURF points in single frame are used for surface matching. Let $X1$ and $X2$ be two matching planes and N be the interval between $X1$ and $X2$. For every plane, m SURF points that are most stable are extracted for feature matching. That is, these feature points are used as a feature set for forensic hash construction. Apparently, the values of N and m are closely related with the length of the forensic hash. This will have a direct influence on forensic accuracy. In general, the bigger is N, the less planes are involved in forensic analysis, and the worse is the forensic accuracy, but the smaller is the length of the forensic hash, and vice versa. The bigger is m, the more SURF points are involved in forensic analysis and the better will be the forensic accuracy, but the bigger the length of the forensic hash. Instead of a 128D vector for a SIFT descriptor, a SURF descriptor is a 64D vector. To further decrease the forensic hash length, the SURF descriptor is compactly represented using visual word representation [12]. Only the visual word *ID* is needed to be stored, rather than the full descriptor.

Figure 5 shows the process of visual word representation. All the feature point descriptors are gathered and quantized into discrete feature vectors. Every feature vector is mapped into visual words to build a vocabulary tree. The vocabulary tree is shared by a data sender and a data receiver, which can be used every detection after its generation. Apparently, this process of hash generation not only keeps the local content of the seam surface but also realizes the dimension reduction. The leads to the significant decrease of forensic hash length. The feature set H_i for surface matching is expressed as Equation (8), which compactly represents the content in a single frame.

$$H_i = \{ID_1, ID_2, \ldots ID_i, \ldots, ID_k\} \tag{4}$$

where every *ID* is a visual word of corresponding feature point descriptors and k is the number of feature points in every matching surface.

In summary, the forensic hash generation involves the following steps: first, the top-k stable SURF feature points

DOI: 10.1002/sec

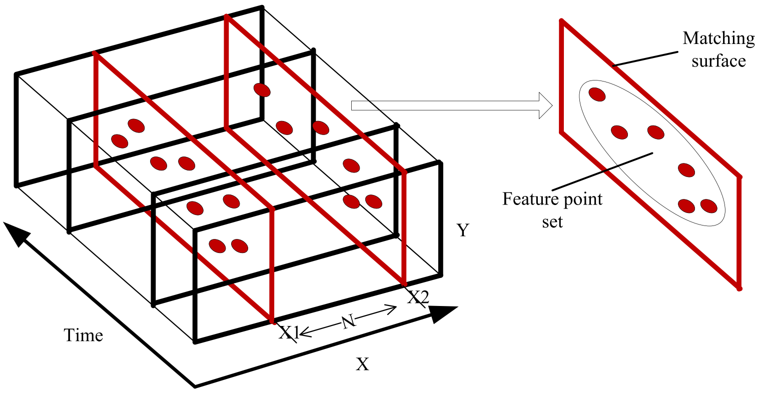


Figure 4. An illustration of the matching surface.

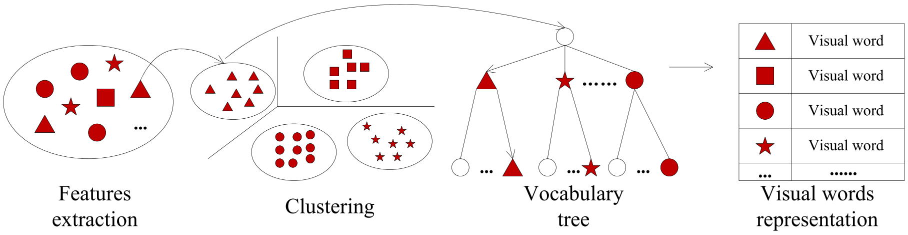


Figure 5. The visual word representation of the forensic hash.

are selected, and their descriptors are hierarchically quantized based on a pre-trained vocabulary tree to obtain the visual word ID_s; second, every point is represented by a vector of five parameters: visual word ID, x and y positions in a frame, its scale, and dominant orientation. Every visual word ID_i is denoted by $(id, x, y, \sigma, \theta)$. For a vocabulary tree with 1000 visual words and an image size of 1024 × 1024, each vector occupies about 50 bits.

3.2. Forensic estimation using forensic hash

Let the size of the original video V be $rows \times cols$. The number of matching surfaces is $M = \lceil cols/N \rceil$, where "$\lceil \cdot \rceil$" is the up round operator. Thus, the forensic hash will be represented as follows:

$$H_i = \{N, \{H_1, H_2, \dots H_i, \dots, H_M\}\} \tag{5}$$

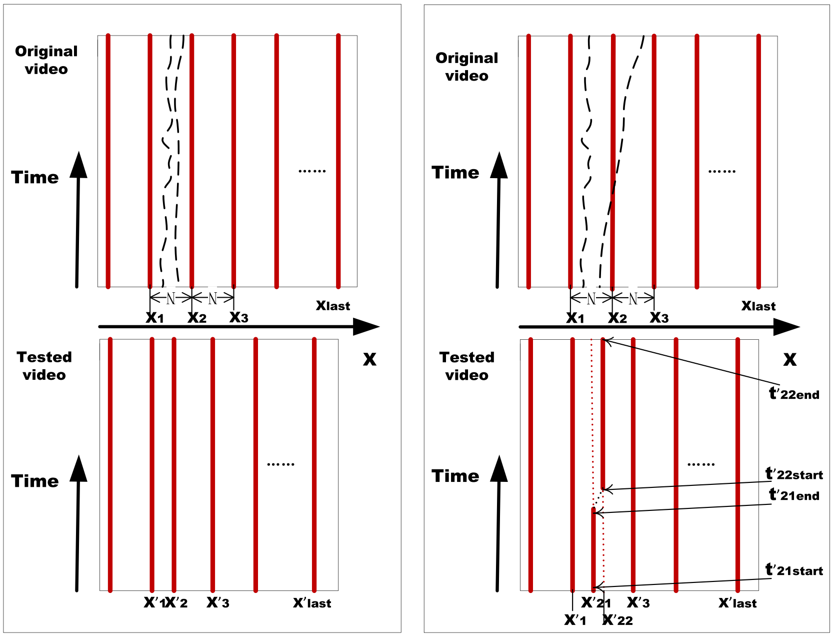
Let the size of the investigated video V' be $rows' \times cols'$. The feature point set H_i of the original video is matched with every Y-time plane along the X-axis of V'. That is, the matching surface is obtained by matching those feature point set in V' with the hash H_i $(i = 1, 2, \dots, M)$. As shown in Figure 6, a red line is a plane in the Y-time, and the red lines at the positions of x'_1, x'_2, and x'_3 in the investigated video are those planes matching the surfaces at x_1, x_2, and x_3 in the original video. That is, the extracted feature point sets are similar at the positions of x_1 and x'_1, x_2 and x'_2, and x_3 and x'_3. If the amount of matched feature points in the Y-time plane of investigated video exceeds a predefined threshold, it is considered that this surface is a matching surface in the Y-time with H_i. From Figure 6(a), it can be observed that there are seams deleted between x_1 and x_2 in the original video, which leads to the relative displacement of investigated video. For seam adding, the forensic process is similar except that the interval between matching surfaces becomes bigger.

In fact, there are two different conditions for seam adding and deleting. One is that the seam surface is completely located between two matching surfaces, and the other is that the seam surface passes through the matching surface, as shown in Figure 6(a and b), respectively. For the former condition as shown in Figure 6(a), the change of seams between x_1 and x_2 is defined as follows:

$$\Delta N = \left(x'_2 - x'_1\right) - N \tag{6}$$

Apparently, if it satisfies $\Delta N > 0$, then it implies that there are some seam surfaces added. Otherwise, if it satisfies $\Delta N < 0$, then it implies that there are some seam surfaces deleted. If $\Delta N = 0$, it means that there is no seam carving here. For the later condition, it is possible that there are two matching surfaces x'_{21} and x'_{22} in the investigated video with surface x_2 in the original video. The change of the last matching surface x_{last} is estimated as follows.

$$\Delta N = (col - X_{\text{last}}) - \left(col' - X'_{\text{last}}\right) \tag{7}$$



(a) seam surface locating between two matching surface

(b) seam surface passing through the matching surface

Figure 6. The detection of seam adding/deleting.

From Figure 6(b), it can be observed that there are some amounts of feature points in the Y-time surfaces x'_{21} and x'_{22}, which are matched with those feature points of matching surface at x_2. Of course, it is also possible that only one surface has some amounts of feature points. When there is only one matching surface between x'_{21} and x'_{22}, the estimation method is similar to the condition shown in Figure 6(a). When the surfaces in the positions of both x'_{21} and x'_{22} are matched with the surface x_2, record the starting frame index $(x'_{21\text{start}}, x'_{22\text{start}})$ and the ending frame index $(x'_{21\text{end}}, x'_{22\text{end}})$ of feature points in the surfaces x'_{21} and x'_{22}, respectively. That is, record the frame indexes of the first and last feature points in the feature point set. Then, the seam surface is estimated as follows.

$$\Delta N = \begin{cases} \left(x'_{21} - x'_1\right) - N & t \in \left[t'_{21start}, t'_{21end}\right] \\ \left(\left\lceil \frac{x'_{22}+x'_{21}}{2} \right\rceil - x'_1\right) & t'_{21end} \neq t'_{22\text{start}} \quad \text{and} \quad t \in \left(t'_{21end}, t'_{22start}\right) \\ \left(x'_{22} - x'_1\right) - N & t \in \left[t'_{22start}, t'_{22end}\right] \end{cases} \tag{8}$$

However, when a surface in the investigated video matching the surface x_2 in the original video cannot be found, it means that the position of x'_2 in Figure 6(a) or the positions of x'_{21} and x'_{22} in Figure 6(b) do not exist. Then, we can simply utilize the matching surface in position x'_3. The number of seams added or deleted between x_1 and x_3 is estimated as follows.

$$\Delta N = \left(x'_3 - x'_1\right) - 2N \tag{9}$$

Thus, it guarantees the estimation of seam carving to the greatest extent, even though the accuracy of estimation range is lowered.

3.3. Performance analysis of the proposed forensic hash

The proposed forensics hash approach is built using SURF feature points. Because SURF features can be extracted in real time, which guarantees that the computational complexity of the proposed approach is well controlled. Moreover, SURF is robust to noise, compression, and image blurring. This implies that desirable results can be achieved even when there is some other changes to digital video such as illumination change. In the feature-matching process, k most stable SURF points are selected for surface matching, instead of all the feature points or single point. This guarantees the stability of forensic performance and the compactness of the forensic hash.

In this paper, the forensic hash is represented by $H = \{N, \{H_1, H_2, \cdots, H_i, \cdots, H_M\}\}$, where N is the interval between two neighboring matching surfaces. If N does not exceed 50, only 6 bits is required. For a vocabulary tree with visual words of no more than 1000, only 10 bits is required to represent every *ID* in H_i. Consequently, for a video sequence in Source Input Format (360 × 240, 500 frames), when *N* equals 30 and the number of matching points is 50 for every matching surface on average, the length of forensics will be 10,006 bits. This is an acceptable length of the forensic hash, which proves its compactness.

Another performance metrics of the forensic hash is its scalability. Apparently, the length of the forensic hash is controlled by two key parameters, that is, the interval *N* between two matching surfaces and the number of matching points *k* in every matching surface. Moreover, the frame index of feature points can be recorded to estimate the frame-based manipulation such as frame adding and deleting. That is, the forensic hash will be $H_i = \{(ID_1, t_1), (ID_2, t_2), \cdots, (ID_i, t_i), \cdots, (ID_k, t_k)\}$, where t_i represents the frame index of feature points in the whole video sequence. The forensic process is summarized as follows. Let t_1 and t_2 be the frame index of two neighboring frames without feature points in the original video and t'_1 and t'_2 be the frame indexes of the matching surface in the investigated video. The relative distances of feature points in the original and investigated videos are $d = t_2 - t_1$ and $d' = t'_2 - t'_1$, respectively. The relative displacement of video frames will be $\Delta D = d' - d$. Apparently, if ΔD is larger than zero (a positive number), it implies that there is frame adding. Otherwise, if ΔD is less than zero, it means that there is frame deleting. Furthermore, the spatial resolution might be changed during video transmission because of video transcoding, which will have influences on video forensics. Therefore, we can consider adding the scaling information into feature points. If the scaling information is kept in the construction of the forensic hash, it can be used to restore the original video using a scaling transform. This will benefit the forensics.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Experimental setup

In order to prove the performance of the proposed approach, experiments are performed on a PC with Intel Core2 2.7 GHz CPU, 2 GB RAM, Windows XP. The test video sequences including *Indoors* and *Car park* are chosen from the forensic analysis library of the University of Surrey, UK [16]. These sequences are in Source Input

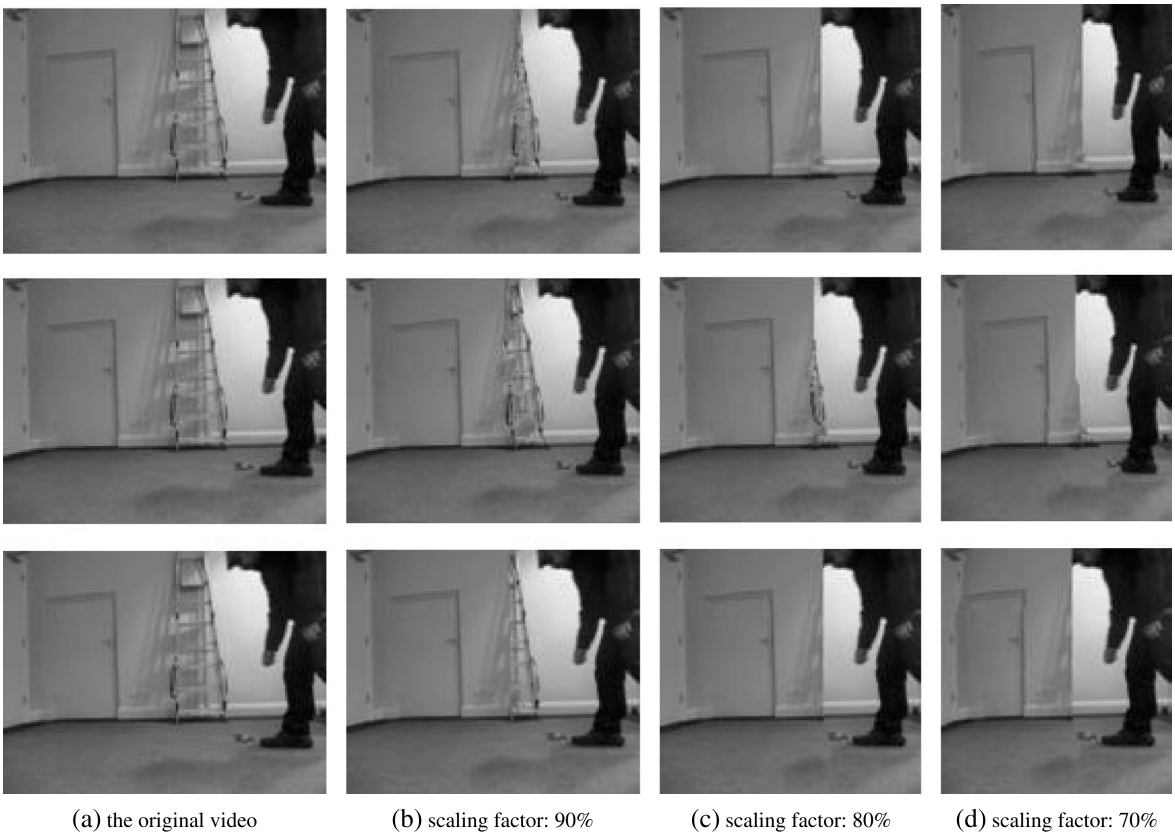


Figure 7. An example of resized videos by different seam carving-based video retargeting algorithms. (a) The original video. (b) Scaling factor: 90%. (c) Scaling factor: 80%. (d) Scaling factor: 70%.

DOI: 10.1002/sec

Format (360 × 240, 500 frames). They are manipulated by three typical SCVR schemes [7–9]. These SCVR schemes are briefly introduced in Section 2.1. The scaling factors for SCVR are 90%, 80%, and 70%, respectively. Note that object removal is also achieved during video resizing by SCVR. As shown in Figure 7, the object, that is, the ladder, is removed from the video. To keep the feature points stable, the response value of feature points is no less than 0.05 when extracting the forensic hash.

The forensic hash generated from the ground truth is matched with that of the tampered video after SCVR, so as to estimate the positions and number of seams involved in SCVR. The performance metrics used for evaluation are the correct detection rate P_r and false detection rate P_f. They are defined as follows.

$$P_r = \frac{\sum_i \min\left(\Delta N - \Delta\tilde{N}\right)}{\Delta N_{\text{total}}} \tag{10}$$

$$P_f = \frac{\sum_i \max\left(\Delta N - \Delta\tilde{N}, 0\right)}{\Delta N_{\text{total}}} \tag{11}$$

where ΔN is the amount of actually deleted seams in every frame, $\Delta\tilde{N}$ is the number of detected seams that are deleted in every frame, i is the total number of video frames, and ΔN_{total} is the total number of deleted seams in all the frames.

4.2. Experimental results and comparison

The detection results are summarized in Table I. From the experimental results, it can be observed that the correct detection rate P_r decreases with the decrease of interval N between matching surfaces. This is mainly because the estimated position increases with the decrease of span, and a small span implies a more accurate estimation of seam positions. Thus, the possibility of false estimation is increased. Furthermore, the correct detection rate of the rescaled video by the method in [7] is better than those of the scaled video by the methods in [8,9]. The reason is that the seams between successive frames in [7] are smoother than those in [8,9]. For example, for the methods in [8,9], there might be seams deleted in the range [10, 40] of the X-axis in the 10th frame, and the seams might occur within the range [100, 140] for the 50th frame. This implies that there are discontinuities of seams among frames. However, the proposed approach achieved desirable detection performance for those video by three typical SCVR algorithms. On average, the correct detection rate is higher than 90%. Moreover, the false detection rate is quite slight especially when the scaling factor is small. Figure 8 illustrates the detection conditions to those videos suffering from different SCVR algorithms when the span N between matching surface is 40. A line in it is the position of matching surface, and the curve is the positions of all the seams in the current frame. From this figure, it is apparent that there is a possibility of false detection.

In the following, we detect the video with 5% of frames deleted. The results are summarized in Table II. It can be observed that there is only a slight decrease of correct detection rate when some frames are deleted from the investigated video. In other words, the frame deletion does not have an obvious influence on the detection of seams by different SCVR methods. Figure 9 is the detection performance via a receiver operating characteristic (ROC) curve when the spans between the matching surface are different. The left is the result when there is simply SCVR, whereas the right is the result when there are simultaneously SCVR and frame deleting. The ROC curve reflects the total actual positives (TPR = true positive rate) versus the total actual negatives (FPR = false positive rate) for every estimated

Table I. Detection accuracy of tampered video by different seam carving-based video retargeting algorithms.

		N							
		50		40		30		20	
Scaling factor (%)		P_r (%)	P_f (%)	P_r (%)	P_f (%)	P_r (%)	P_f (%)	P_r (%)	P_f (%)
90	Method [7]	98.3	2.17	97.5	3.53	96.8	4.22	95.2	5.97
	Method [8]	98.2	2.32	96.4	3.87	95.2	6.72	90.8	9.39
	Method [9]	98.2	2.24	97.2	3.61	95.3	5.31	91.2	9.34
	Average	98.2	2.24	97.0	3.67	95.8	5.42	92.4	8.23
80	Method [7]	97.9	3.54	95.4	5.72	95.1	6.74	90.3	12.78
	Method [8]	95.7	4.77	92.6	8.81	90.8	9.75	86.4	14.03
	Method [9]	96.5	4.39	93.5	8.17	91.9	9.10	87.6	13.46
	Average	96.7	4.23	93.8	7.57	92.6	8.53	88.1	13.42
70	Method [7]	95.5	6.82	92.0	9.61	89.1	11.48	86.4	15.02
	Method [8]	92.9	7.22	89.3	10.03	87.6	13.09	84.3	18.17
	Method [9]	93.8	7.10	90.1	10.76	88.2	12.69	84.7	17.39
	Average	94.1	7.05	90.5	10.13	88.3	12.42	85.1	16.86

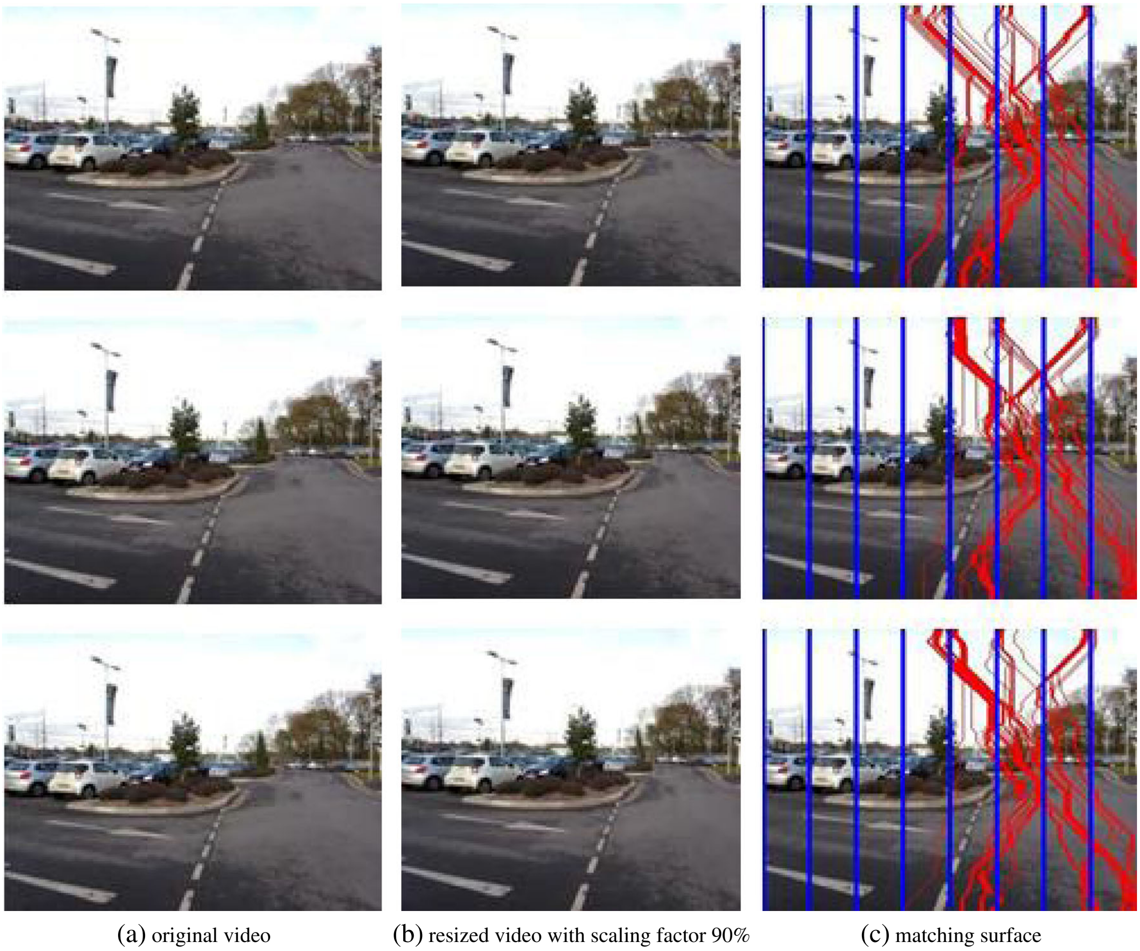
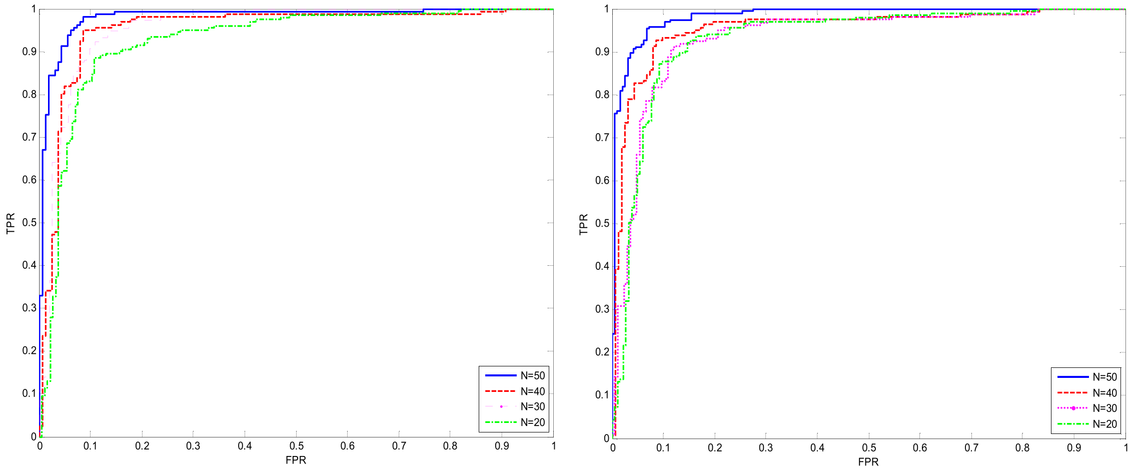

(a) original video (b) resized video with scaling factor 90% (c) matching surface

Figure 8. The detection results of a tampered video by different seam carving-based video retargeting algorithms. From top to bottom, they are the detection results of [7–9], respectively. (a) Original video. (b) Resized video with scaling factor 90%. (c) Matching surface.

Table II. The detection performance of seams in the tampered video by different seam carving-based video retargeting algorithms.

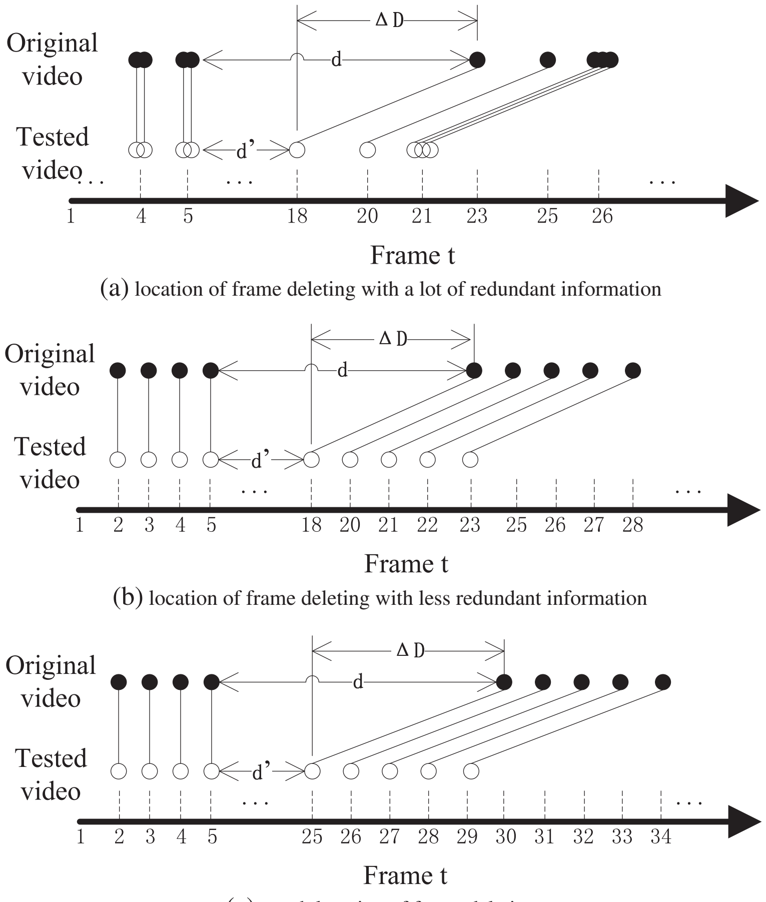
		N							
		50		40		30		20	
Scaling factor (%)		P_r (%)	P_f (%)	P_r (%)	P_f (%)	P_r (%)	P_f (%)	P_r (%)	P_f (%)
90	Method [7]	98.1	2.36	97.1	4.30	96.2	4.78	94.6	7.03
	Method [8]	97.6	2.39	96.8	4.45	93.8	6.31	90.7	10.25
	Method [9]	97.7	2.48	97.1	4.31	94.7	5.84	90.9	10.34
	Average	97.8	2.41	97.0	4.35	94.9	5.64	92.1	9.21
80	Method [7]	96.4	4.79	94.2	6.99	93.1	7.39	89.1	12.92
	Method [8]	94.6	7.05	91.8	10.00	89.7	10.87	87.2	13.93
	Method [9]	95.7	5.65	92.0	8.93	90.5	10.52	87.5	13.58
	Average	95.6	5.83	92.7	8.64	91.1	9.59	87.9	13.48
70	Method [7]	94.2	7.14	90.8	11.89	88.7	12.41	84.1	17.48
	Method [8]	92.8	7.92	89.7	11.97	86.4	14.77	85.6	18.91
	Method [9]	92.7	8.34	89.6	12.34	87.4	14.55	84.0	17.48
	Average	93.2	7.80	90.0	12.07	87.5	13.91	84.6	17.96



(a) Without frame-based operation

(b) With frame deleting

Figure 9. The receiver operating characteristic curve when the interval N is different. (a) Without frame-based operation. (b) With frame deleting.



(a) location of frame deleting with a lot of redundant information

(b) location of frame deleting with less redundant information

(c) rough locating of frame deleting

Figure 10. The estimation of frame index for frame deleting.

seam. The results reported are the average results for all the seams in every frame with different scaling factors. In Figure 9(a), the TPR is bigger than 90% for different spans when the FPR is 10%. This implies that the proposed forensic hash achieves satisfactory detection results. Figure 9(b) reports the detection results of the tampered video by both SCVR algorithms and frame deleting. It also shows that the detection performance decreases slightly when some frames are deleted.

If the frame index is added into the forensic hash, the index and position of the frame deleted or added can be estimated. To verify this, three groups of experiments are conducted. The first is to store the frame index information for all the feature points. The second is to partially store the frame index of feature points. That is, if the frame number is the same for different feature points in the same matching surface, just keep the frame number of one feature point. The third is to just store non-repetitive information. That is, only one different frame number information for all the matching surfaces is stored, and the feature points that keep their frame number are randomly selected. Figure 10 shows the experimental results of frame number estimation when some frames are added or deleted. The solid points represent the feature points in the original video, whereas the hollow points are the feature points in the investigated video. The lines between the solid points and hollow points imply that these points are matched with each other. The horizontal axis is the corresponding frame number of feature points. From the experimental results shown in Figure 10(a, b), it can be observed that there are five frames deleted between the 5th and 23rd frames of the original video. In Figure 10(c), there are also five frames deleted between the 5th and 30th frames. Although the detection results are the same in Figure 10(a, b), there are much more redundant information in Figure 10(a) than in Figure 10(b). For Figure 10(c), although the exact number and rough position of the deleted frames are correct, its accuracy is decreased because the range of the deleted frames is wider than that reported in Figure 10(a, b). From the preceding analysis, we can conclude that the proposed forensic approach can accurately estimate the number of deleted frames if the frame index is kept in the forensic hash. Moreover, the accuracy of estimation depends on the number and distributions of feature points in the Y-time surface. If there are more non-repetitive feature points, it will achieve higher accuracy of the estimated range for those deleted frames. However, the length of the forensic hash will also be bigger. For frame adding, similar results are also obtained, which are omitted here owing to space restrictions.

5. CONCLUSIONS

In this paper, a detection approach is proposed for SCVR using a forensic hash. It extracts the SURF points every several frames to build the forensic hash. The change of the relative position between two matching surfaces is used to detect the possible seam carving in SCVR. We comprehensively consider the robustness and compactness of the forensic hash and the forensic accuracy. Experimental results show that the number and positions of seams deleted or added by SCVR can be accurately estimated. Moreover, the proposed forensic hash approach is scalable. If the frame index is recorded into the forensic hash, it is also possible to estimate frame-based manipulation. For future research, because all the existing forensic hash approaches including the proposed approach are tampering specific, it will be much better if some theoretical support can be developed for the construction of the forensic hash. Furthermore, considering that digital videos usually suffer from several kinds of tampering, we will attempt to model the tampering process by a complex processing chain and estimate the full processing history.

ACKNOWLEDGEMENTS

This work is supported in part by the National Natural Science Foundation of China (61072122 and 61379143), the program for New Century Excellent Talents in University (NCET-11-0134), the key project of Hunan Provincial Natural Science Foundation (11JJ2053), S&T Program of Xuzhou City (XM13B119), and the Specialized Research Fund for the Doctoral Programme of Higher Education (20120161110014). We also appreciate the kind help from Dr. Jeff Tan for proofreading.

REFERENCES

1. Rocha A, Scheirer W. Vision of the unseen: current trends and challenges in digital image and video forensics. *ACM Computing Surveys* 2011; **43** (4): 26–40.
2. Redi J, Taktak W. Digital image forensics: a booklet for beginners. *Multimedia Tools and Applications* 2011; **51**(2): 133–162.
3. Milani S, Fontani M, Bestagini P, Barni M, Piva A, Tagliasacchi M, Tubaro S. An overview on video forensics. *APSIPA Transactions on Signal and Information Processing* 2012; **1**(e1): 1–18.
4. Avidan S, Shamir A. Seam carving for content-aware image resizing. *ACM Transactions on Graphics* 2007; **26**(3): 1–9.
5. Frankovich M, Wong A. Enhanced seam carving via integration of energy gradient functionals. *IEEE Signal Processing Letters* 2011; **18**(6): 375–378.
6. Luo S, Zhang J, Zhang Q, Yuan X. Multi-operator image retargeting with automatic integration of direct and indirect seam carving. *Image and Vision Computing* 2012; **30**(9): 655–667.
7. Rubinstein M, Shamir A, Avidan S. Improved seam carving for video retargeting. *ACM Transactions on Graphics* 2008; **7**(3): 1–9.

8. Grundmann M, Kwatra V, Han M, Essa I. Discontinuous seam-carving for video retargeting, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, 2010; 569–576.
9. Yan B, Kairan S, Liu L. Matching-area-based seam carving for video retargeting. *IEEE Transactions on Circuits and Systems for Video Technology* 2013; **23**(2): 302–310.
10. Sarkar A, Nataraj L, Manjunath B. Detection of seam carving and localization of seam insertions in digital images, *Proceedings of the 11th ACM Multimedia Security Workshop (MM&Sec09)*, Princeton, New Jersey, USA, 2009; 107–116.
11. Ryu SJ, Lee HY, Lee HK. Detection of content-aware image resizing using seam properties. *Applied Mechanics and Materials* 2013; **284**: 3074–3078.
12. Lu W, Wu M. Seam carving estimation using forensic hash[C], *Proceedings of the 13th ACM Workshop on Multimedia and Security (MM&Sec09)*, Buffalo, NY, USA, 2011; 9–14.
13. Wang X, Xue J, Zheng Z, Liu Z, Li N. Image forensic signature for content authenticity analysis. *Journal of Visual Communication and Image Representation* 2012; **23**(5): 782–797.
14. Battiato S, Farinella GM, Messina E, Puglisi G. A robust forensic hash component for image alignment. In *Image Analysis and Processing-ICIAP*. Springer: Berlin, Heidelberg, 2011; 473–483.
15. Luo J, Gwun O. A comparison of SIFT, PCA-SIFT and SURF[J]. *International Journal of Image Processing* 2009; **3**(4): 143–152.
16. Surrey University Library for Forensic Analysis [EB/OL]. http://sulfa.cs.surrey.ac.uk/videos.php [Accessed date: 2013-08-10].