Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

journal homepage: www.elsevier.com/locate/aeue

Robust semi-blind video watermarking based on frame-patch matching

Q1 Ta Minh Thanh[a,b,*], Pham Thanh Hiep[c], Ta Minh Tam[d], Keisuke Tanaka[a]

[a] Tokyo Institute of Technology, 2-12-2, Ookayama, Meguro-ku, Tokyo 152-8552, Japan
[b] Le Quy Don Technical University, 100 Hoang Quoc Viet, Cau Giay, Hanoi, Viet Nam
[c] Department of Electrical and Computer Engineering, School of Engineering, Yokohama National University, Japan
[d] Satellite digital television Co. Ltd., Viet Nam

ARTICLE INFO

Article history:
Received 11 February 2014
Accepted 17 May 2014

Keywords:
KAZE feature
Video watermarking
Geometrically invariant
Feature matching
RST parameters estimation

ABSTRACT

In this paper, we present a frame-patch matching based robust semi-blind video watermarking using KAZE feature. The KAZE feature is employed for matching the feature points of frame-patch with those of all frames in video for detecting the embedding and extracting regions. In our method, the watermark information is embedded in Discrete Cosine Transform (DCT) domain of randomly generated blocks in the matched region. In the extraction process, we synchronize the embedded region from the distorted video by using KAZE feature matching. Based on the matched KAZE feature points, RST (rotation, scaling, translation) parameters are estimated and the watermark information can be successfully extracted. Experimental results show that our proposed method is robust against geometrical attacks, video processing attacks, temporal attacks, and so on.

1. Introduction

1.1. Background

In recent years, developments in PC and network technology have made digital content techniques widely available. It makes numerous advantages over the analog multimedia content, *i.e.* easy duplication, re-distribution and economic transmission both via network, wifi or by physical media (CD, DVD, *etc.*). However, these advantages have outstretched the security concerns of digital multimedia content such as copyright protection, owner's right problem, legal user verification and so on.

In order to protect and preserve the copyright of digital content, many copyright protection techniques have been proposed. Digital watermarking is a promising technique used for copy control, identification and traitor tracing. In digital watermarking, a watermark information is embedded into an image or video without affecting the quality but that can be detected using the dedicated algorithm. Ownership of the contents can be established by retrieving the embedded watermark from the copyright contents.

However, several attacks are effective against watermarking methods and watermark information can be destroyed under those attacks [1]. To the best of our knowledge, geometrical distortion, which derives the different error between the original content locations and those of the suspected content, is the most difficult attacks to resist. Since it resynchronizes the embedded location of the embedding method and then causes incorrect watermark extraction. Therefore, the watermark synchronization process is needed to detect the robust watermark locations before watermark embedding and extraction.

1.2. Related works

A number of research related to watermark synchronization has been proposed. In our best knowledge, the watermark synchronization methods using periodic sequence [2], templates [3], an invariant transform [4,5], and media contents [6,7] have been reported.

Nikolaidis et al. [8] proposed a watermark synchronization method based on image segmentation using an adaptive *k*-mean clustering technique. Their method segmented image based on the clustering region of *k*-mean and selected several largest regions. The bounding rectangles of the fitted ellipsoids from those regions are used as the patches for watermark embedding and extracting.

Q2 * Corresponding author at: Tokyo Institute of Technology, 2-12-2, Ookayama, Meguro-ku, Tokyo 152-8552, Japan. Tel.: +81 80 4168 7979.
E-mail addresses: thanhtm@ks.cs.titech.ac.jp, taminhjp@gmail.com (T.M. Thanh), hiep@kohnolab.dnj.ynu.ac.jp (P.T. Hiep), tam.ta@vstv.vn (T.M. Tam), keisuke@is.titech.ac.jp (K. Tanaka).

http://dx.doi.org/10.1016/j.aeue.2014.05.004
1434-8411/© 2014 Elsevier GmbH. All rights reserved.

The drawback of this method is that the image segmentation in *k*-mean clustering depends on the image contents, which can be changed under several image processing attacks. Therefore, image distortions severely affect the segmentation results for synchronization process.

With another view point, Dajun et al. [9], Ho et al. [10], and Lee et al. [11] proposed the watermark synchronization methods using object-based watermarking in which a set of angular radial transformation coefficients was selected as the feature to represent the object region and the background. In their extraction schemes, the scaling ratio was supposed to be known so that the received video object could be scaled back to its original resolution. But, such an assumption may not hold in the real situation.

In the synchronization method introduced by Tang et al. [12], first, the feature points by Mexican hat wavelet scale interaction are extracted. Second, the disks of fixed radius centered at each feature point are normalized and these disks are used as the patches for watermark embedding and extraction. However, when the images are distorted, the normalized disks cannot be correctly extracted from the image because the normalization is sensitive to the image contents. Therefore, the robustness of these patches will decrease when the image is attacked.

With a similar motivation, Viet and Aizawa [13] used SIFT (Scale-Invariant Feature Transform) feature to detect and match the object region. However, in their method, the object region is needed to select manually in advance. These conventional methods can be used only in “proof of ownership” applications and are not suitable for video broadcasting.

In order to eliminate the synchronization process, Wang et al. [14] employed the histogram shape of the low frequency subband in DWT domain that is insensitive to various geometric distortions. Wang et al. used two successive bins of histogram to embed and extract the watermark information. However, the number of the embedded watermark bits in one frame is clearly limited. If the number of watermark bits is increased, the method of Wang et al. [14] may not be robust against some attacks.

1.3. *Challenge issues*

Based on the related works, we summarize the following challenge issues:

(1) *Synchronization of extraction watermark process*. The previous methods tried to find out efficient ways to synchronize the extraction watermark process. However, some methods are not robust when the suspected image is degraded by several image processing attacks [8,12]. Other methods are supposed to know several parameters for synchronization process [9–11,13]. Those methods are not suitable for real applications. Therefore, the first challenge issue is how to improve the synchronization in order to correctly extract the watermark information.

(2) *Considering of the robust feature for synchronization*. To synchronize the embedded region in extraction process, we should employ the robust feature for restoration. In general, a set of robust feature points is used as the secret feature to restore the suspected image. Therefore, the second challenge issue is how to choose the feature points for restoration.

(3) *Problem of digital property dispute and illegal user tracing*. Based on the related works, only one watermark information is embedded into the digital content before delivering to a user. When a legal user redistributes the content via network without permission, the producer cannot detect the traitor even he/she can detect the illegal redistribution. Moreover, if the illegal user and the legal user have a problem of digital property dispute, the producer cannot judge based on the watermark information. Hence, the last challenge issue is how to provide the method to easily trace the traitor and to judge the illegal user and the legal user if the problem of digital property dispute happens.

1.4. *Our contributions*

To address the issue (1), we propose the novel synchronization method using frame-patch matching. The frame-patch is randomly created based on the secret key for an individual legal user. Since the embedded region is defined by frame-patch matching, the extracted watermarks of the legal user are different. Therefore, although one digital content is delivered to many users, there are many frame-patchs that are created to distinguish the legal users. This proposed point is different from the previous methods.

In our understanding, for robust watermarking of watermark synchronization, the selection of features (global feature or local feature) is very important. It is considered that the local features are more useful than global ones [15]. The KAZE feature is considered as a local image property and it is invariant under rotation, scaling, translation, and partial illumination changes [16]. In this paper, we propose a robust video watermarking method using KAZE feature for synchronization of embedding and extracting watermark. Moreover, we extract the KAZE feature points from a video frame and match them with those of frame-patch to detect the matched region in all frames. Using the matched region, we generate the embedding and extracting region for watermark information. In our method, the watermark is embedded into the matched region based on DCT domain. After embedding the watermark, the producer sends the *license number* to the user for proving legal usage of content. To detect the embedded information in the video, we first detect the matched region by using KAZE feature matching. And by calculating the affine parameters, we can geometrically recover the video frame, and can easily extract the watermark information. By employing the KAZE feature for synchronization extraction, we can solve the issue (2).

In order to solve the issue (3), we consider to ask the legal user to register her/his information as watermark information. Based on the user information, the producer makes the secret key to create the frame-patch for each user and embeds the user information into the embedded region that is determined by frame-patch matching process. Since the user information is embedded into the digital content, the producer can extract the user information and detect exactly the traitor even if the legal user redistributes the content. If the problem of digital property dispute happens, the producer can request users to provide the *license number* to obtain the secret key, and restore the suspected content to extract watermark.

1.5. *Roadmap*

This paper is organized as follows. Section 2 gives the advantage of our proposed method in the real situation. Section 3 introduces our proposed frame-patch matching based video watermarking using KAZE feature. Section 4 presents the results of the experiments and Section 5 concludes our paper.

2. Advantage of our method in the real situation

2.1. *The flow of the system*

Our proposed method depends on the frame-patch that is used for synchronizing the embedded regions and the extracted regions. Therefore, the preparation of frame-patch is very important process in our method. The overview of our system is shown in Fig. 1.

We suppose our proposed method can be applied to the copyright content distribution system. Before delivering the content,

Please cite this article in press as: Thanh TM, et al. Robust semi-blind video watermarking based on frame-patch matching. Int J Electron Commun (AEÜ) (2014), http://dx.doi.org/10.1016/j.aeue.2014.05.004

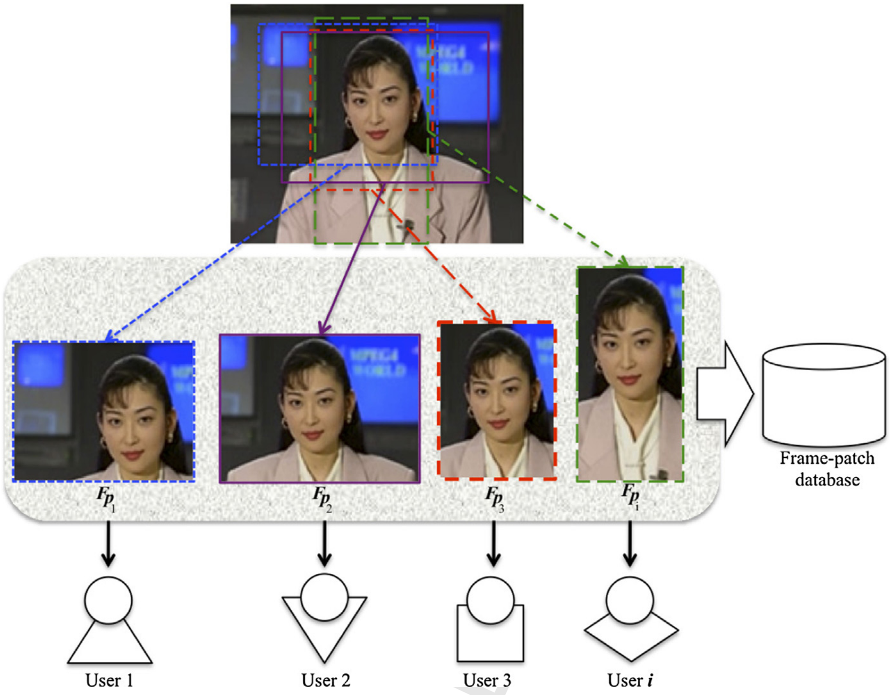


Fig. 1. Overview of frame-patch preparation. Producer can prepare many frame-patchs ($e.g.$ $F_{p_1}, F_{p_2}, \cdots, F_{p_i}$) for individual users ($e.g.$ $U_1, U_2, \cdots, U_i$). Best viewed in color.

the user is requested to register his/her information, which is used as watermark information W. Based on the user information, the producer makes the secret key K_i for creating the frame-patch for individual users. In the server side, the producer prepares the frame-patch F_{p_i} corresponding to the legal user U_i as shown in Fig. 1. The producer randomly generates many frame-patchs based on the secret key K_i and saves them into the frame-patch database. After preparing the frame-patch, the producer assigns each frame-patch to a legal user and implements the watermarking method based on Section 3.1. After embedding the user information into the content, the producer sends to the legal user the *license number* L_i. By using one frame-patch, the producer synchronizes the embedded regions and the extracted regions from all frames by using the KAZE feature point matching. That is why our proposed method is appropriate for video sequences. The synchronization process of watermark extraction is described in Section 3.2.

2.2. Detection of traitor

Since only the producer has the frame-patch database, so he/she can exactly detect the embedded region for extraction. Based on this idea, the security for watermark information is considered by the proposed method. Therefore, attackers have to require high computation cost to specify the embedded regions in the video sequences.

When the producer detected the redistribution via network, he/she can extract the user information by using the frame-patch database and compare it with user database. If the extracted watermark matches the user database, the user is legal. Conversely, if the extracted watermark is different from the user database, the user is illegal.

Moreover, by using user's information as watermark to embed into the content, the purpose of the proposed method is to inform the user about the existence of watermarking which can be used to exactly identify users. Therefore, it can limit the illegal redistribution in advance.

2.3. Resolving of digital property dispute

Our proposed method is designed to address the problem of digital property dispute. Suppose there are many users who disputes digital contents. In order to judge the legal user, the producer asks those users to provide their *license number* L_i and obtains the secret key K_i. According to K_i, the producer can obtain the frame-patch and extract the watermark information. The extracted watermark according to K_i will judge whether the user is right.

3. Proposed semi-blind video watermarking

In order to synchronize the embedded region and the extracted region, we decide to choose the KAZE feature that is used to match the feature points of all frames and those of frame-patch. According to the matched KAZE feature points, the distorted video can be restored and the watermark information can be successfully extracted.

3.1. Embedding based on frame-patch matching

As shown in Fig. 2, we describe how to embed the watermark information in our method. There are six steps in the scheme:

- **Step 1:** Extract one frame F from the original video V and extract the KAZE feature points of frame F and frame-patch F_p. After that, those feature points are matched each other to detect the region for embedding. The KAZE matching method is explained in Section 3.3.
- **Step 2:** Convert the RGB frame of the matched region to YCbCr color space.
- **Step 3:** Transform Y-component to a frequency domain using Discrete Cosine Transform (DCT).
- **Step 4:** Embed $W(i, j) \in \{0, 1\}$, $1 \le i, j \le L$ to Y-component in the frequency domain, where $L \times L$ is the size of watermark. $W(i, j)$ is converted into a linear array $W_l(k) = W(i, j)$, $k = i + jL$, $1 \le i, j \le L$.

Please cite this article in press as: Thanh TM, et al. Robust semi-blind video watermarking based on frame-patch matching. Int J Electron Commun (AEÜ) (2014), http://dx.doi.org/10.1016/j.aeue.2014.05.004

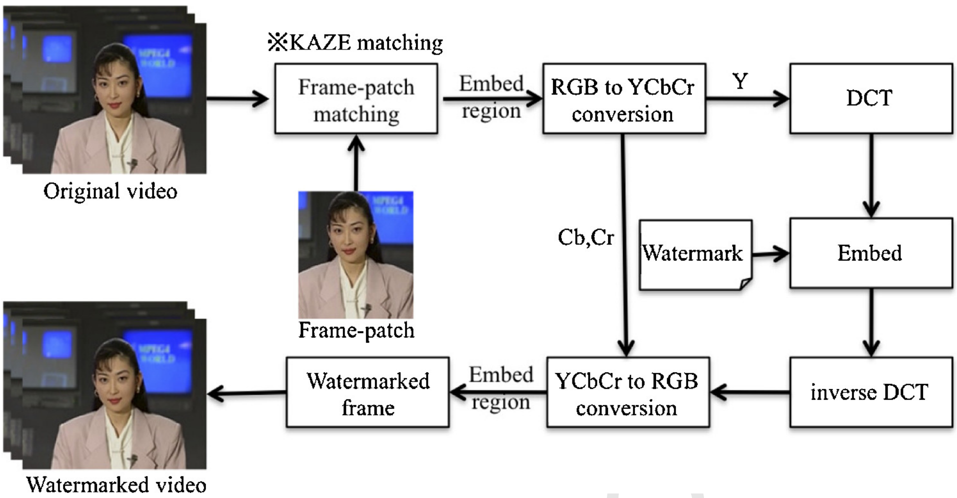


Fig. 2. Embedding based on frame-patch matching.

One bit $W_l(k)$ is embedded into the DCT coefficient in the frequency domain. The detailed explanation of this step is described in Section 3.4.

- **Step 5:** Compute the inverse DCT to obtain the modified Y-component and compose it with the Cb and Cr components.
- **Step 6:** Convert the modified YCbCr frame to obtain the modified RGB frame.

Repeat Step 1 to Step 6 for all frames in video, we can get the watermarked video.

3.2. *Extraction based on frame-patch matching*

Fig. 3 describes how to extract the embedded information from the watermarked video by using KAZE feature points matching in the frame-patch. This procedure consists of the following steps.

- **Step 1:** Extract one frame F' from the watermarked video V' and extract the KAZE feature points of it. Next, the feature points of frame-patch F_p is used to match with those of F' and detect the embedded region.
- **Step 2:** Based on matched feature points, the rotation, the scaling, and the translation parameters of the distorted video are calculated (see Section 3.5). Then, the distorted video is restored.
- **Step 3:** Convert the RGB frame of the matched region to YCbCr color space.
- **Step 4:** Transform Y-component to a frequency domain using DCT.
- **Step 5:** Here, the embedded information $W_l(k)$ can be extracted from the matched region. The detail explanation of this step will be described in Section 3.4.

Repeat Step 1 to Step 5 for all frames in video, we can get all watermark information from the watermarked video.

3.3. *Frame-patch matching method*

In order to synchronize the embedding region and extracting region, we have to find the common local features between the distorted video and the original video. By using the KAZE feature matching, we find that we can detect the matched feature for recovering the distorted video.

First, the KAZE feature points extracted from a target frame F are matched with those of a frame-patch F_p. Here, the P.F. Alcantarilla et al. method [16] is used for matching.

Suppose that the KAZE feature points p_l, q_k are extracted from frame F and frame-patch F_p, respectively:

$$p_l = (x_l, y_l, \lambda_l, o_l, \boldsymbol{f}_l), \quad \text{for } l \in 1, \ldots, L, \tag{1}$$

$$q_k = (x\prime_k, y\prime_k, \lambda\prime_k, o\prime_k, \boldsymbol{f}\prime_k), \quad \text{for } k \in 1, \ldots, K, \tag{2}$$

where x_l, y_l, λ_l, and o_l are, respectively, the x- and y-position, the scale and the orientation of the lth detected feature point of frame F. The element $\boldsymbol{f}_l$ is a 64-dimensional or 128-dimensional local edge orientation histogram of the lth point. The symbol "$\prime$" denotes the parameters of KAZE feature points in F_p. In our matching experiment, we use 128-dimensional local edge orientation histograms.

To find the matched points, for each point p_l in F, we compute its distances d_{1l} and d_{2l} to its two nearest neighbors in F_p:

$$d_{1l} = \underset{k}{argmin}||\boldsymbol{f}\prime_k - \boldsymbol{f}_l||, \tag{3}$$

$$d_{2l} = \underset{k \neq k_{\min}}{argmin}||\boldsymbol{f}\prime_k - \boldsymbol{f}_l||, \tag{4}$$

where $k_{\min}$ is the index of a feature point which had the minimum distance d_{1l}. Next, a ratio r_l is defined as $r_l = \frac{d_{1l}}{d_{2l}}$. Given a threshold τ, we can obtain a set of the matched points is $\mathbf{M} = \{(p_l, q_k)|r_l < \tau\}$. Using $\mathbf{M}$, we can detect the matched region from all frames for embedding and extraction as shown in Fig. 4. Note that, the matched region is found by the rectangle which covers all matched points in F.

Suppose that there are multiple frames of the original video, only one F_p is used to detect the matched regions in every frames. Therefore, it is considered that the matched regions of the distorted video can be synchronized with those of the original video as shown in Fig. 4.

3.4. *Embedding and extraction algorithm*

The embedding and the extraction method are performed on the DCT frequency domain of the matched region. First, we segment the DCT coefficients into an (8×8)-block. In each block, two coefficients at (x_i, y_i) and at (y_i, x_i) are selected randomly from the low-frequency region by using the secret key. Their DCT coefficients

Please cite this article in press as: Thanh TM, et al. Robust semi-blind video watermarking based on frame-patch matching. Int J Electron Commun (AEÜ) (2014), http://dx.doi.org/10.1016/j.aeue.2014.05.004

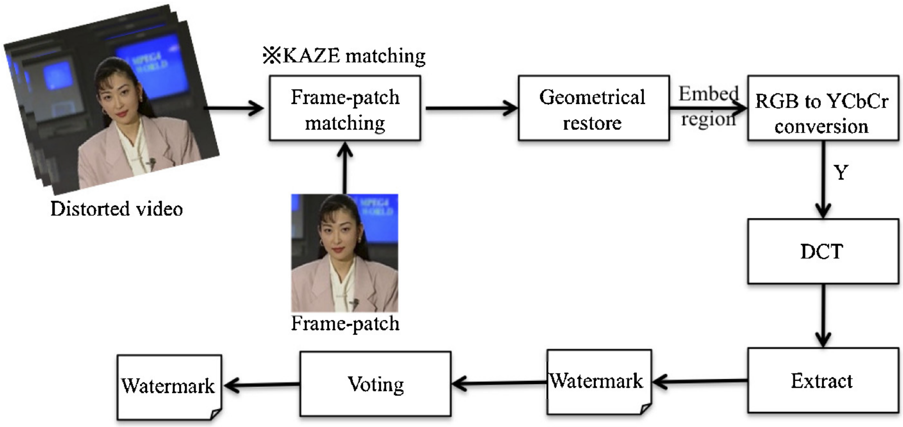


Fig. 3. Extraction based on frame-patch matching.

$f(x_i, y_i)$ and $f(y_i, x_i)$ are modified with the watermarking strength a $(a > 0)$:

When $W_l(k) = 0$,

$$\begin{cases} f'(x_i, y_i) = \dfrac{f(x_i, y_i) + f(y_i, x_i)}{2} - \dfrac{a}{2}, \\ f'(y_i, x_i) = \dfrac{f(x_i, y_i) + f(y_i, x_i)}{2} + \dfrac{a}{2}. \end{cases} \tag{5}$$

When $W_l(k) = 1$,

$$\begin{cases} f'(x_i, y_i) = \dfrac{f(x_i, y_i) + f(y_i, x_i)}{2} + \dfrac{a}{2}, \\ f'(y_i, x_i) = \dfrac{f(x_i, y_i) + f(y_i, x_i)}{2} - \dfrac{a}{2}. \end{cases} \tag{6}$$

Note that, in our method, $x_i \neq y_i$. The secret key is saved for information extraction.

In the information extraction, the embedded bit can be extracted by comparing $f'(x_i, y_i)$ and $f'(y_i, x_i)$, where f' indicates that corresponding distorted DCT coefficients: If $f'(x_i, y_i) > f'(y_i, x_i)$ then $W'_l(k) = 1$, otherwise $W'_l(k) = 0$.

The watermark strength a effects to quality of watermarked video. In our experiments, set the basis value of a to 0.15. This value will be increased or decreased based on the feature of the local region.

After extracting the watermark $W'_l(k)$, two dimensional watermark $W'(i, j)$ is formed from $W'_l(k)$ as,

$$W'(i, j) = W'_l(k), \quad k = i + jL,\ 1 \leq i, j \leq L. \tag{7}$$

The watermark is extracted from all the frames. Since there are F_N frames, the number of watermarks extracted at the receiver side will also be F_N. The final watermark W_v needs to be constructed from these F_N watermarks based on some decision. In our method, the voting method is used to find the maximum occurrence of bit value (either bit “0″ or bit “1″) corresponding to the same pixel location in all extracted watermarks. This voting method can be represented as the following equation,

$$W_v(i, j) = voting(W'_f(i, j), \quad 1 \leq f \leq F_N),\ 1 \leq i, j \leq L. \tag{8}$$

$W_v(i, j)$ is the final watermark information that is extracted from F_N frames of the distorted video. Our proposed method is robust against several attacks as shown in Section 4.2.2 even if $W_v(i, j)$ is similar to $W(i, j)$.

3.5. Estimation of RST parameters

According to the KAZE feature point matching, the rotation, the scaling, and the translation parameter can be estimated based on some set **M** of the matched points.

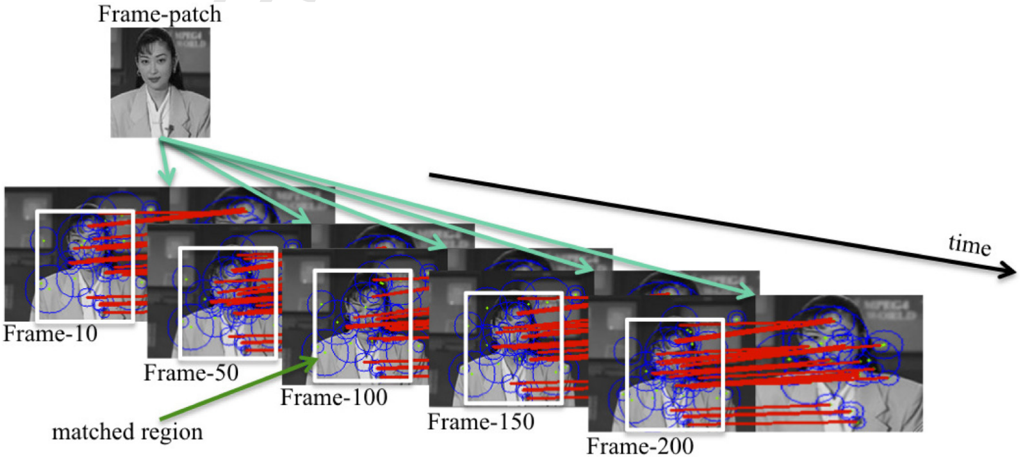


Fig. 4. Matched region using KAZE feature matching.

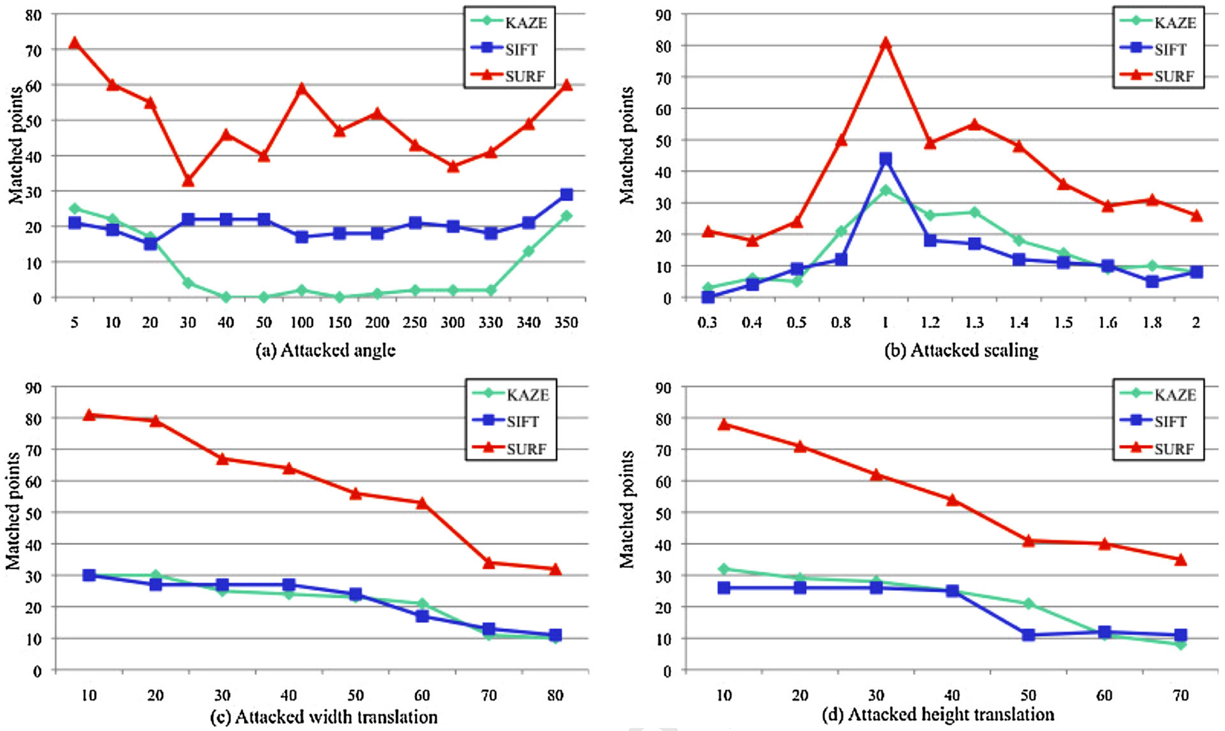


Fig. 5. Comparison of matched points of akiyo: (a) clockwise rotation attacks with angle from 5° to 350°; (b) scaling attacks with scale factor from 0.3 to 2.0; (c) Width translation attacks with δx from 0 to *Width*/2; (d) Height translation attacks with δy from 0 to *Height*/2. Best viewed in color.

First, we can estimate the scaling parameter Λ between the frame and the frame-patch as follows:

$$\Lambda = \frac{\sum_{i=1}^{M} \lambda\prime_i}{\sum_{i=1}^{M} \lambda_i}, \tag{9}$$

where λ_i and λ'_i are the scales of matched feature points of frame F and the frame-patch F_p, respectively, and $M = |\mathbf{M}|$.

Next, we estimate the angle α of rotation by using the feature points matched to each other. The rotation angle is estimated as follows:

$$\alpha = \frac{\sum_{i=1}^{M} (\alpha\prime_i - \alpha_i)}{M}, \tag{10}$$

where α_i and α'_i denote the centre angle of the feature point i of the frame-patch and that of the corresponding feature point i of the rotated distorted frame, respectively.

After adjusting the differences of scale and rotation, we calculate the translation parameters δx and δy, which correspond to the differences in width and height, respectively. Let the coordinates of frame-patch feature points i is (x_i, y_i) and that of the corresponding distorted frame feature points (x'_i, y'_i). Then the translation parameters are estimated as follows:

$$\delta x = \frac{\sum_{i=1}^{M} (x\prime_i - x_i)}{M}, \delta y = \frac{\sum_{i=1}^{M} (y\prime_i - y_i)}{M}. \tag{11}$$

4. Experimental result

4.1. Experimental environment

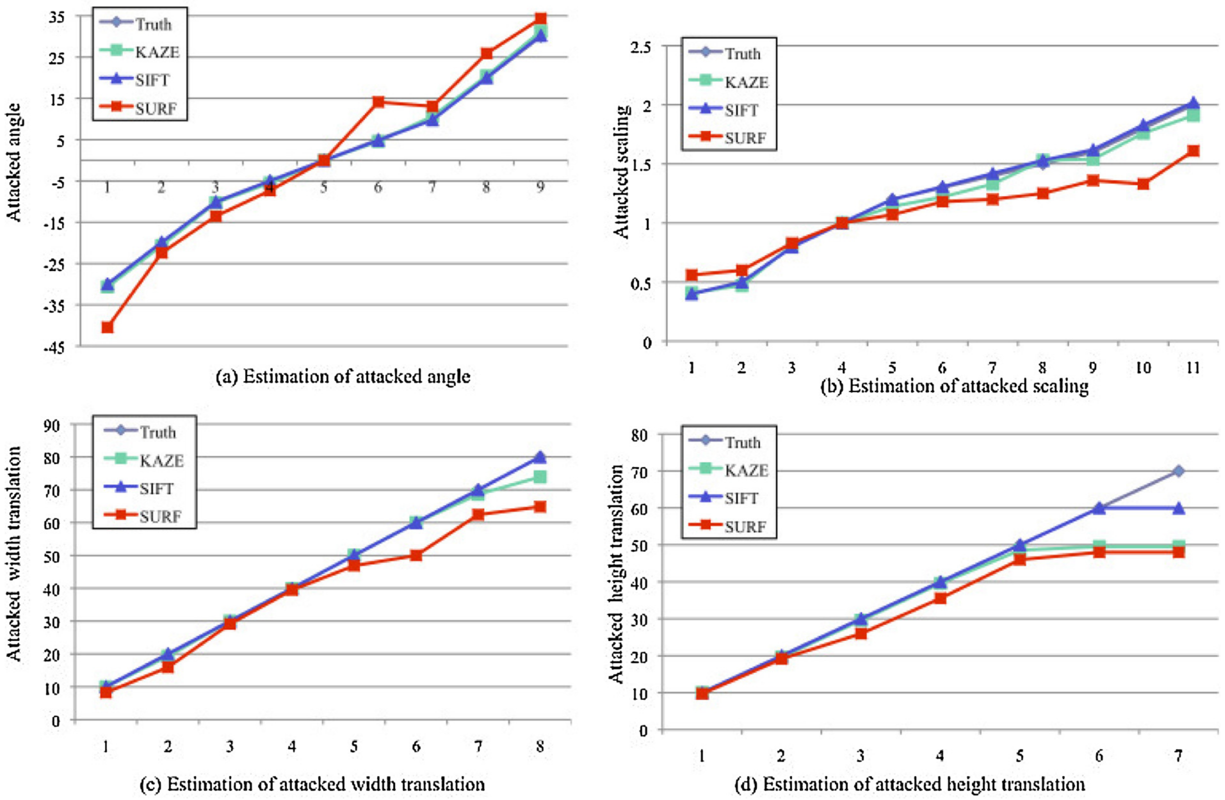
To evaluate the proposed method fairly, we use six standard video sequences[1] (Akiyo: 300 frames, 25 fps; Coastguard: 300 frames, 25 fps; Carphone: 382 frames, 12 fps; Miss: 150 frames, 25 fps; Mobile: 300 frames, 25 fps; Suzie: 150 frames, 25 fps) in QCIF format (*Width* × *Height* = 176 × 144). All the experiments are performed in the Mac OSX 10.6.8 system. The watermark is a binary image with the size $L \times L = 64 \times 64$. GCC version 4.0.1[2] and the MPlayer version 1.1-4.2.1[3] are used to convert and to view the experimental video data.

In order to evaluate the performance of the proposed method, the transparency and the robustness for watermarking are used to measure the system performance.

Transparency is important to evaluate the watermarking. The peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are used as criteria to estimate the invisibility [17]. In these experiments, the PSNR and SSIM are calculated for every video sequences. Since PSNR and SSIM do not take the temporal activity into account, to compare the perceptual quality, the video quality metric (VQM)[18] is also employed. This metric is between zero and one; zero means not having any distortion while one shows maximum impairment. The original sequences and the watermarked sequences are used as the original and the processed videos.

Robustness is also important factor in watermarking. The normalized cross correlation (NCC) [19] measures the difference

[1] http://media.xiph.org/video/derf/.
[2] http://gcc.gnu.org/.
[3] http://www.mplayerhq.hu/design7/dload.html.



(a) Estimation of attacked angle

(b) Estimation of attacked scaling

(c) Estimation of attacked width translation

(d) Estimation of attacked height translation

Fig. 6. Comparison of estimated parameters of akiyo: (a) estimated rotation attacks with angle from −30° to 30°; (b) estimated scaling attacks with scale factor from 0.4 to 2.0; (c) estimated width translation attacks with δx from 0 to $Width/2$; (d) estimated height translation attacks with δy from 0 to $Height/2$. Best viewed in color.

between the extracted watermark $W_v(i, j)$ and the original watermark $W(i, j)$.

4.2. Simulation results

4.2.1. Feature points matching

In order to evaluate our method using KAZE feature, we apply the rotation, the scaling, and the translation attacks to the video sequences. With those attacks, the resulting feature points are matched with those of the frame-patch. We compare the number of matched points of KAZE feature with those of SIFT feature [20] and SURF [21].

As shown in Fig. 5, we observe that the number of the matched points of SURF (red) is larger than those of KAZE feature (green) and SIFT feature (blue). The reason for this is that a SURF vector has 64 dimensions for matching, whereas a SIFT feature vector and a KAZE feature vector has 128 dimensions. Therefore, the mismatching points of SURF are more than others [16]. Thus, SURF is not appropriate since such mismatched points can derive incorrect result.

We note that at least two matched points are required for geometrical recovery. If we use KAZE feature, we can estimate the rotation parameters from −30°(=330°) to +30° (Fig. 5(a)), the scaling parameters from 0.3 to 2.0 (Fig. 5(b)), and the translation parameters from 0 to $Width/2$ for width translation and from 0 to $Height/2$ for height translation (Fig. 5(c) and (d)). In general, the cropping rotations are small for video and their angles are no more than 5 degrees. Therefore, the estimation of the rotation angles from −30° to +30° is enough for rotation attacks.

4.2.2. Estimation of attacked parameters

In order to estimate the geometrical parameters (RST) from the matched points, we implement the rotation, the scaling, and the translation attack. Fig. 6 shows the actual parameters called *Truth* and the estimated parameters of KAZE, SIFT, and SURF.

As described below, we can see that the KAZE feature gives the estimation results almost equal to SIFT feature and better than SURF.

First, for rotation (Fig. 6(a)), we see that we can estimate the rotation angle from −30° to 30° as for the SIFT feature. There is large error if we use SURF in the rotation attack. The average errors of KAZE, SIFT, and SURF are 0.54°, 0.11°, and 4.57°, respectively.

Second, for scaling (Fig. 6(b)), we also see that the estimated scaling factors are close to *Truth* factors by using the KAZE feature and SIFT feature.

However, if the scaling factor equals to 0.3, the scale parameter cannot be estimated by SIFT because the matched point equals to 0 (see Fig. 5(b)) and SURF because the all matched points are mismatching. By using KAZE, it can be estimated successfully. The average errors of KAZE, SIFT, and SURF are 0.05, 0.01, and 0.23, respectively.

Table 1
Average PSNR[dB], SSIM and VQM value.

Video	PSNR	SSIM	VQM
Akiyo	37.14	0.95	0.02
Coastguard	36.46	0.98	0.04
Carphone	36.93	0.96	0.06
Mobile	35.48	0.99	0.04
Miss	37.59	0.93	0.02
Suzie	37.52	0.95	0.03

Table 2
Q3 Experimental results of robustness. The NCC values of ours/those of [13]

Attack	Akiyo	Coastguard	Carphone	Mobile	Miss	Suzie
No attack	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Rotation w. crop (−20°)	0.98/1.0	1.0/1.0	1.0/1.0	0.99/1.0	1.0/1.0	1.0/1.0
Rotation w. crop (−10°)	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Rotation w. crop (−5°)	0.99/1.0	1.0/1.0	1.0/1.0	0.99/1.0	1.0/1.0	1.0/1.0
Rotation w. crop (+5°)	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Rotation w. crop (+10°)	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Rotation w. crop (+20°)	1.0/1.0	1.0/1.0	1.0/1.0	0.99/1.0	1.0/1.0	1.0/1.0
Scaling 0.3	**0.69**/–	**0.70**/–	**0.69**/–	**0.53**/–	**0.61**/–	**0.56**/–
Scaling 0.5	0.91/0.92	0.90/0.90	**0.91**/0.90	0.88/0.94	0.80/0.90	0.79/0.99
Scaling 0.8	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Scaling 1.2	1.0/1.0	1.0/1.0	1.0/1.0	1.0/0.99	1.0/1.0	1.0/1.0
Frame dropping 10%	0.70/0.70	0.68/0.68	0.66/0.69	0.70/0.76	**0.66**/0.53	0.66/0.66
Frame dropping 20%	0.69/0.69	**0.70**/0.69	**0.69**/0.66	**0.70**/0.69	**0.73**/0.69	0.69/0.70
Frame insertion 10%	**0.70**/0.69	**0.71**/0.68	**0.68**/0.63	0.70/0.72	**0.72**/0.61	0.65/0.71
Frame insertion 20%	**0.70**/0.69	**0.70**/0.69	**0.70**/0.69	**0.70**/0.68	**0.66**/0.65	**0.71**/0.70
Frame transposition 10%	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Frame transposition 20%	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Frame averaging 10%	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Frame averaging 20%	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Blur	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Gaussian 3 × 3	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
MPEG-4	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
Xvid	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0

Finally, for translation (Fig. 6(c) and (d)), we can estimate the translation factors from 0 to *Width*/3 and from 0 to *Height*/3 by using KAZE; from 0 to *Width*/2 and from 0 to *Height*/2 by using SIFT. In these parameters, the average errors of KAZE, SIFT, and SURF are 0.92, 0.01, and 4.79 for width translation; 3.74, 1.12, and 5.29 for height translation, respectively.

Therefore, according to our results, the KAZE feature is almost the same as the SIFT [13] and much better than SURF. Thus, KAZE feature can be employed for video watermarking.

4.2.3. Quality evaluation

As shown in Table 1, the average PSNR values of six watermarked video sequences are 37.14, 36.46, 36.93, 35.48, 37.59, and 37.52 decibels (dB). All the values are higher than 36 dB. In addition, the average SSIM values of those are 0.95, 0.98, 0.96, 0.99, 0.93, and 0.95. All the values are also very close to 1.0. The original video and the watermarked video are visually indistinguishable. This implies that the watermarking scheme can achieve visual transparency. By observing the VQM values of six watermarked videos, we confirm that the videos with more motion (Coastguard, Mobile, and Carphone), are degraded less than those with less motion (Akiyo, Miss, and Suzie).

4.2.4. Evaluation of robustness

We use a tool for video watermarking attack named VirtualDub[4] to modify the watermarked videos. We also apply the Vidmark Benchmark [22] to simulate the temporal desynchronization attacks. We regard the watermarking scheme as robust if the NCC values are over than 0.9.

In general, the cropping rotations are small for video and their angles are no more than 5 degrees. In our experiments, we use more than 5°. In particular, from −20° to +20°. We confirm that the NCC values remain over than 0.9.

We also apply several scale factors (from 0.3 to 1.2) to the watermarked videos.

Table 2 shows the experimental results of robustness: rotations, scaling, frame dropping, frame insertion, frame transposition, frame averaging, blur, gaussian, and compression. We compare our proposed method with [13]. As it can be observed, when the scaling factor equals to 0.3, the method of [13] cannot detect the matched region, while our method can.

For example, the watermarks of the experimental results for Akiyo video sequences are given in Fig. 7. As shown in Fig. 7, we can see clear images of watermark.

4.2.5. Discussion

In this paper, we propose a robust frame-patch matching based semi-blind watermarking method using KAZE feature which can be applied to the copyright content distribution system. The point of our proposed method is that the producer can use many frame-patchs for distinguishing legal users while only one object layer can be used in [13]. Therefore, our technique is more flexible than [13]. From this idea, our method is applied not only to copyright protection, but also distinguishing legal user and detecting illegal redistribution. For this purpose, the user ID can be used as watermark to be embedded into the frame-patch region.

The method of Wang [14] was shown to be robust against some attacks, however, a small amount of watermark can be embedded because the number of histogram bins on frequency is limited. In [14], only 60 bits of watermark information are repeatedly embedded within a frame. If the number of bits for watermark is increased, the robustness of Wang's method may decrease. However, our proposed method does not dependent on the embedding range and bin width that affects to the robustness of method. In this paper, we use 4096 bits of binary logo as watermark information and our method can be more robust if the number of bits for watermark is reduced.

In addition, as mentioned in [16], compare to others, KAZE feature has the disadvantage of time consuming. Therefore, the proposed method is suitable for offline system for checking copyrights and detecting the illegal users. To adapt for real-time applications, we have to reduce the matching process, *e.g.*, once for every 10 frames. We leave it to the future works.

[4] http://www.virtualdub.org/.

Please cite this article in press as: Thanh TM, et al. Robust semi-blind video watermarking based on frame-patch matching. Int J Electron Commun (AEÜ) (2014), http://dx.doi.org/10.1016/j.aeue.2014.05.004

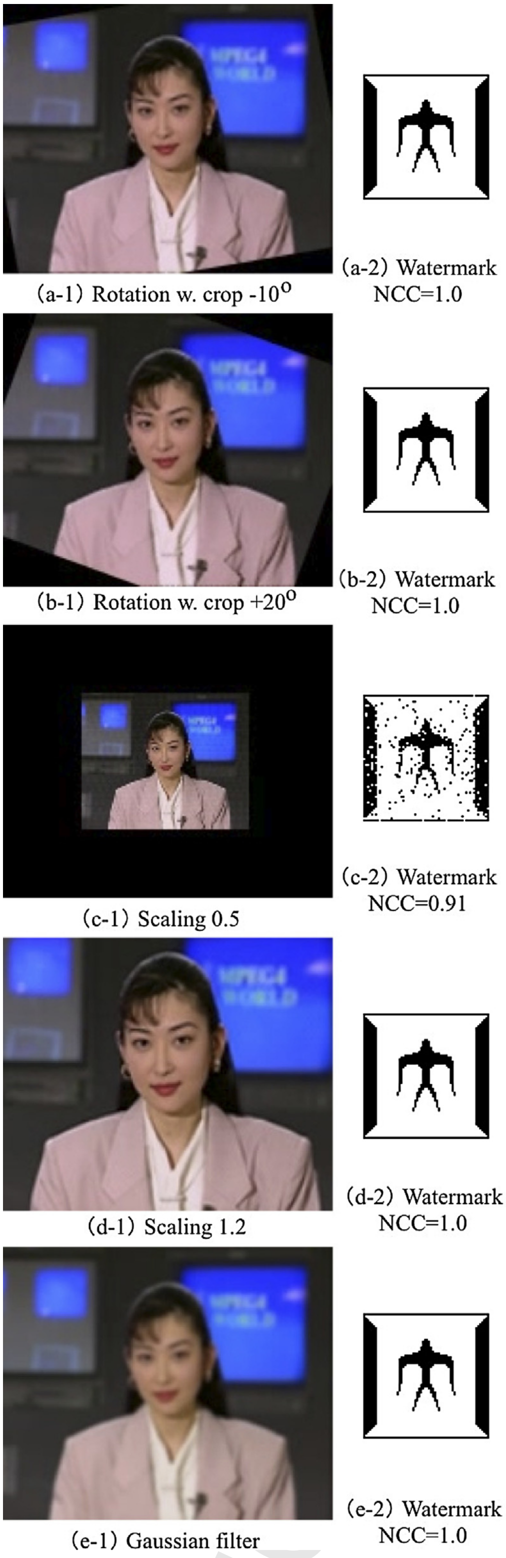

(a-1) Rotation w. crop -10°
(a-2) Watermark NCC=1.0
(b-1) Rotation w. crop +20°
(b-2) Watermark NCC=1.0
(c-1) Scaling 0.5
(c-2) Watermark NCC=0.91
(d-1) Scaling 1.2
(d-2) Watermark NCC=1.0
(e-1) Gaussian filter
(e-2) Watermark NCC=1.0

Fig. 7. Example of Akiyo videos (#frame 2).

5. Conclusion

In this paper, we have employed the KAZE feature to develop a robust semi-blind video watermarking. By using the proposed method, we can solve three challenges mentioned before. First, we have proposed the frame-patch matching technique using the KAZE feature for synchronizing the embedded and extraction regions in the watermarking scheme. Second, by employing the KAZE feature, we can say that the local feature is more robust and it can help our method to restore the distorted the video. Finally, based on the advantage of our method, we can provide the ways to trace the illegal redistribution and to judge the legal users when the problem of digital property dispute happens. There are still some issues left for future work. For instance, we want to reduce the computation cost to apply our method for realtime video watermarking. For this issue, we have to reduce the cost of matching process, *e.g.*, performing the matching once for every 10 frames. In this paper, we have focused on the robustness against geometrical attacks, video processing attacks, and temporal attacks. As future work, it is interesting to combine different techniques with our proposed method to achieve the robustness also against ratio change, perspective transform, and so on.

References

[1] Petitcolas FAP, Anderson RJ, Kuhn MG. Attacks on copyright marking systems. In: Proc. int. workshop on information hiding. Springer-Verlag; 1998. p. 218–38.

[2] Kutter M. Watermarking resisting to translation, rotation and scaling. In: Proc. SPIE 3528. 1998. p. 423–31.

[3] Pereira S, Pun T. Robust template matching for affine resistant image watermark. IEEE Trans Image Process 2000;9(6):1123–9.

[4] Lin C-Y, Cox IJ. Rotation, scale and translation resilient watermarking for images. IEEE Trans Image Process 2001;10(5):767–82.

[5] Ruanaidh O, Pun T. Rotation, scale and translation invariant spread spectrum digital image watermarking. Signal Process 1998;66(3):303–17.

[6] Bas P, Chassery J-M, Macq B. Geometrically invariant watermarking using feature points. IEEE Trans Image Process 2002;11(9):1014–28.

[7] Lee HY, Kim H, Lee HK. Robust image watermarking using local invariant features. Opt Eng 2006;45(3):037002.

[8] Nikolaidis A, Pitas I. Region-based image watermarking. IEEE Trans Image Process 2001;10(11):1726–40.

[9] Dajun H, Sun Q, Tian Q. A secure and robust object-based authentication system. EURASIP J Appl Signal Process 2004;14:1–14.

[10] Ho YK, Wu MY. Robust object-based watermarking scheme via shape self-similarity segmentation. Pattern Recogn Lett 2004;25(15):1673–80.

[11] Lee JS, Kim WY. A new object-based image watermarking robust to geometrical attacks. In: Proc. the 5th Pacific Rim conf. On advances in multimedia information processing. 2004. p. 58–64.

[12] Tang C-W, Hang H-M. A feature-based robust digital image watermarking scheme. IEEE Trans Signal Process 2003;51(4):950–9.

[13] Viet PQ, Miyaki T, Yamasaki T, Aizawa K. Robust object-based watermarking using feature matching. IEICE Trans Inform Syst 2008;E91-D(7):2027–34.

[14] Wang L, Ling H, Zou F, Lu Z. Real-time compressed domain video watermarking resistance to geometric distortions. IEEE MultiMedia 2012;19(1):70–9.

[15] Lee HY, Kim H, Lee HK. Robust image watermarking using local invariant features. J SPIE Opt Eng 2006;45(3), 037002-1–11. Q4

[16] Alcantarilla PF, Bartoli A, Davison AJ. KAZE features. In: European conf. on computer vision (ECCV). 2012. http://www.robesafe.com/personal/pablo.alcantarilla/kaze.html

[17] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12.

[18] Pinson M, Wolf S. A new standardized method for objectively measuring video quality. IEEE Trans Broadcast 2004.

[19] Huang HY, Yang CH, Hsu WH. A video watermarking technique based on pseudo-3-D DCT and quantization index modulation. IEEE Trans Inform Forensics Secur 2010;5(4):625–37.

[20] Lowe D. Distinctive image features from scale-invariant keypoints. Int J Comput Vision 2004;60:91–110.

[21] Bay H, Ess A, Tuytelaars T, Gool LV. SURF: speeded up robust features. Comput Vision Image Understand 2008;110:346–59.

[22] Pedro AH, Claudia F, Cumplido R, Garcia-Hernandez JJ. Towards the construction of a benchmark for video watermarking systems: temporal desynchronization attacks. In: Proc. of the 53rd IEEE international midwest symposium on circuits and systems (MWSCAS). 2010. p. 628–31.

Please cite this article in press as: Thanh TM, et al. Robust semi-blind video watermarking based on frame-patch matching. Int J Electron Commun (AEÜ) (2014), http://dx.doi.org/10.1016/j.aeue.2014.05.004