社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2006－ＡＶＭ－55（７）
2006／12／14

SIFT 特徴量を用いた 幾何学的不変のオブジェクトベース電子透かし

ファンヴェトクォク† 味八木 崇‡ 山崎 俊彦* 相澤 清晴*

東京大学 †工学部 ‡大学院新領域創成科学研究科 ＊大学院情報理工学系研究科
113-8656 東京都文京区本郷 7-3-1 工学部２号館
E-mail: {pqvietvn, miyaki, yamasaki, aizawa}@hal.k.u-tokyo.ac.jp

あらまし 現在、パソコンとインターネットの普及と高性能化によってディジタルコピーが容易になった。コンテンツの不正コピーに対処する方法の一つとして、電子透かし技術が注目されている。電子透かしとはマルチメディアデータに、画質や音質にはほとんど影響を与えずに情報を埋め込む技術である。その中でも、画像中のオブジェクト領域に情報を埋め込む技術はオブジェクトベース電子透かしと呼ばれる。一般に、オブジェクトベース電子透かしは幾何学的変形に弱い。本稿では幾何学的攻撃に強いオブジェクトベース電子透かしを実現するために、SIFT 特徴量を用いたオブジェクトマッチングとロバストな埋め込み方式を組み合わせた電子透かしを提案する。

キーワード 電子透かし、オブジェクトベース、幾何学的不変、SIFT 特徴量、オブジェクトマッチング

Geometrically Invariant Object-based Watermarking Using SIFT Features

Pham Viet Quoc† Takashi MIYAKI‡ Toshihiko YAMASAKI* and Kiyoharu AIZAWA*

† Faculty of Engineering ‡ Graduate School of Frontier Sciences ＊Graduate School of Information Science and Technology , The University of Tokyo
Eng. Building #2, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
E-mail: {pqvietvn, miyaki, yamasaki, aizawa }@hal.k.u-tokyo.ac.jp

Abstract Owing to the development of the Internet, digital content has become easier to access. In order to stop illegal content copy, some methods have been proposed, and digital watermarking is one of them. Digital watermarking is a technology that embeds a short message (a watermark) in the image without affecting the usability but that can be detected using dedicated analysis software. To apply to video protection, digital watermarking needs to become an object-based method, because the objects, not the whole image are usually attacked. One of the biggest weaknesses of such methods is that they are sensitive to geometrical distortions. In this paper, by using the combination of a hiding method and object matching, we propose an object-based watermarking that is invariant to geometrical distortions.

Keyword Digital Watermarking, Object-based, Geometrically Invariant, SIFT Feature, Object Matching

1. Introduction

Owing to the development of the Internet, digital imaging has seen tremendous growth over the last decade. We now can easily find and download a large number of images within few seconds. Despite the fact that the digital media has become easier to copy, manipulate or convert without any control, the authors can not stop showing their work in public. In order to protect and preserve the owner's right, a number of copyright protection methods have been proposed.

Digital watermarking is a technology used for copy control and media identification and tracing. In digital watermarking, they embed a short message (a watermark) in an image without affecting the usability but that can be detected using dedicated analysis software. But like other security problems, new protection methods always come along with new attacks. Geometrical distortion is one of the most difficult attacks to solve.

In video editing and transmission, it is the object of interest, not the whole video, which needs to be processed [5]. To protect videos against object-based attacks, the watermarking needs to become an object-based method. And how to make the object-based watermarking able to resist geometrical distortions is one of the most important jobs for watermarking researchers these days. In this paper, we propose an object-based watermarking that is

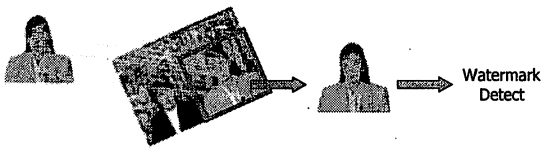


Figure 1: Method demonstration.

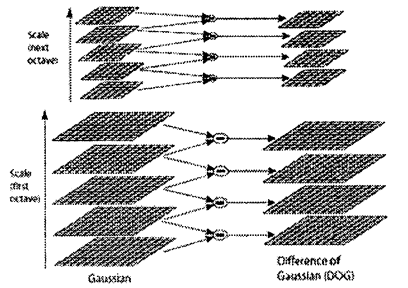


Figure 2: Scale space and DoG images [1].

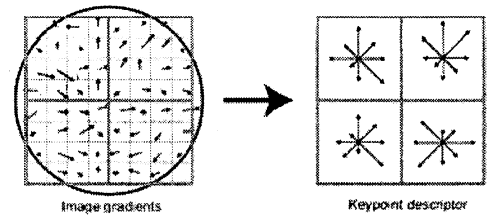


Figure 3: Keypoint descriptor [1].

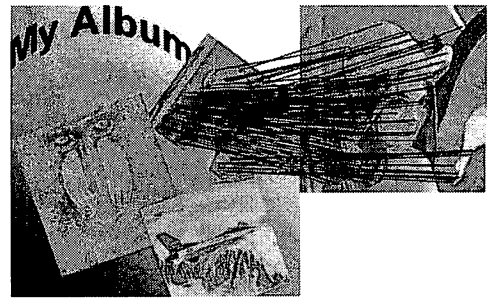


Figure 4: Object matching from keypoint pair matching.

invariant to geometrical distortions by using the combination of a robust hiding method and object matching. The general idea of this method is shown in Fig. 1. The watermarked object "akiyo" in Fig. 1 is attacked by being mixed with another object and then geometrical transformed. To detect the hidden information in the object, we firstly detect the object region by using the feature matching. And by calculating the affine parameters, we can geometrically recover the object, and can easily read the hidden message. In our method, we employed the SIFT algorithm [1] for the object matching job.

The experiment results demonstrate that our proposed method can resist to very strong attacks such as 0.4x scaling, 180° rotation, 30° shearing, JPEG 20% or the combination of both of them.

2. Scale-Invariant Feature Transform

Scale-Invariant Feature Transform (SIFT) [1] is an algorithm for extracting distinctive features from images. The algorithm has been used for matching different views of an object or scene and object recognition. The features (called "SIFT features") are invariant to the image scale, rotation, and partially invariant to changing viewpoints, and changing in illumination.

The first stage of the computation searches for extrema over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian (DoG) function (Fig. 2) to identify potential interest points that are invariant to scale and orientation.

Next, keypoints are selected from the candidates based on measures of their stability. Finally, a keypoint descriptor is created by computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location (Fig. 3).

As shown in the Fig. 4, two objects are matched by searching the nearest keypoint pairs from the two objects. The nearest keypoint is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector.

3. Watermarking Scheme

3.1. Embedding Scheme

This scheme (Fig. 5) describes how to embed hidden messages and register information for the detecting scheme. There are three steps in the scheme:

-Step 1: In this step, we select the object region from the original work (common name for digital contents, such as image, video) to embed a hidden message. The object shape can be arbitral – the bounding rectangle or the segmentation results of the object. In our experiment, we choose the bounding rectangle as the region to be the message (Fig. 6).

-Step 2: A hidden message is embedded in the selected

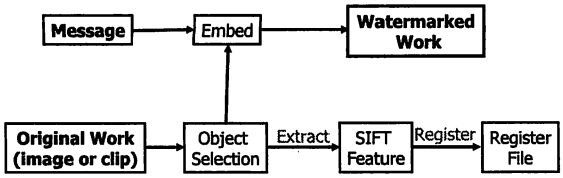


Figure 5: Embedding scheme.

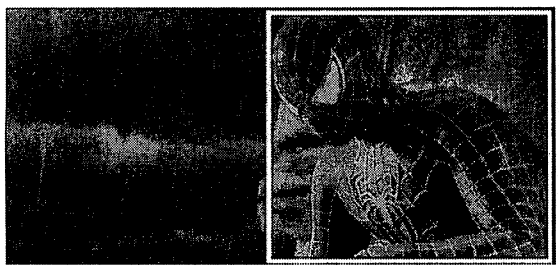


Figure 6: Bounding rectangle of the object.

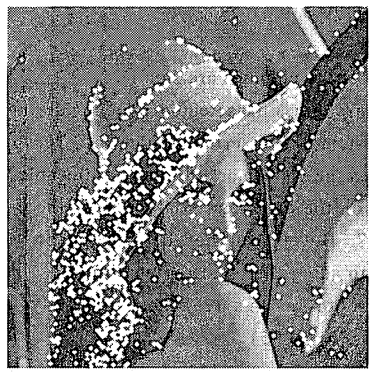


Figure 7: 1082 SIFT feature points are extracted.

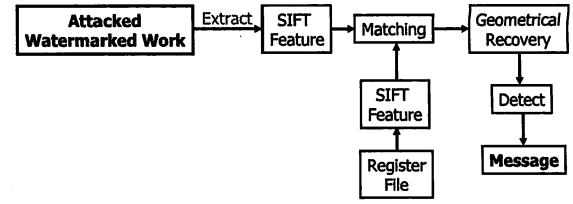


Figure 8: Detecting scheme.

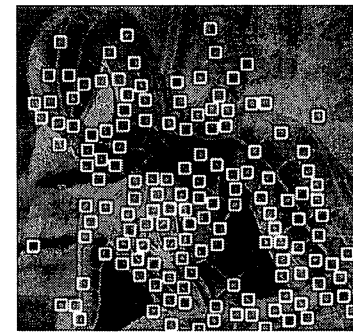
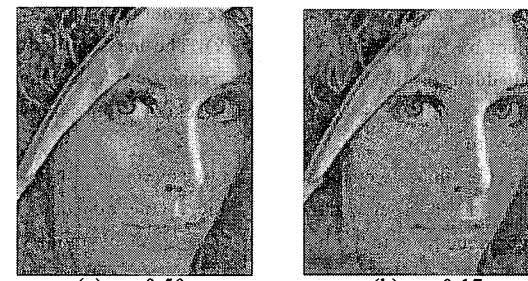


Figure 9: Generated random blocks.



(a) s = 0.50 (b) s = 0.17

Figure 10: Watermarking strength.

region. The detailed explanation of the embedding method is described in section 3.3.

-Step 3: In this final step, we extract SIFT features from the object region (Fig. 7) and register them for the object matching in the detection procedure. The position of the selected region is also registered.

3.2. Detecting Scheme

This scheme (Fig. 8) describes how to detect the embedded message from the object that was attacked. We divide the detecting scheme in three steps:

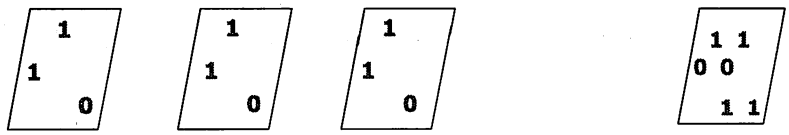
-Step 1: We extract the SIFT features from the attacked watermarked work. Then the extracted features will be matched with the registered features (step 3 in section 3.1).

-Step 2: Based on the matching results, we calculate the parameters of the geometrical transformation (six parameters for the affine transformation, or nine parameters for the perspective transformation). In our experiment, we only deal with the affine transformation. The perspective transformation will be solved in the future work.

-Step 3: We recover the attacked object region based on the transformation parameters found in step 2. Then the embedded message can be detected (section 3.3).

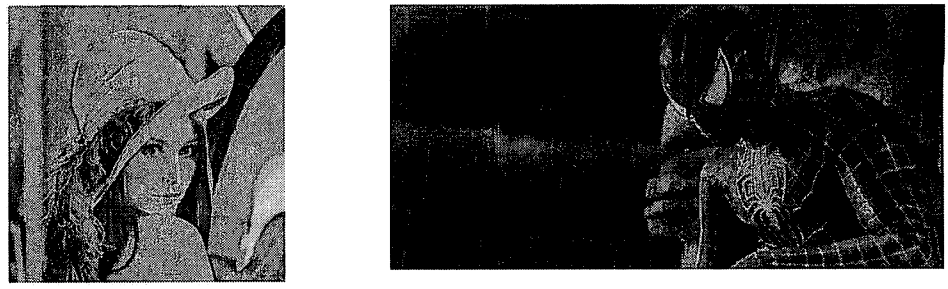
3.3. Embedding Method

We embed a bit array in the object by modifying the Discrete Cosine Transform (DCT) coefficients of its region. The geometrical distortion does not change the waveform of the region, so that the embedded message is preserved well even after the attacks.



(a) Within a single frame (b) In multiple frames

Figure 11: Repetitive Embedding.



(a) Still image "LENA" (b) Video clip "Spider Man" [7]

Figure 12: Experiment samples.

First, random 16x16 blocks are generated within the object by the secret key K (Fig. 9). The message will be embedded in these blocks. The reason for the random block sampling is to prevent attackers from knowing where to attack.

After that, two coefficient indices *(x0, y0), (x1, y1)* are selected from each 16x16 block. One bit (mark) is then embedded in both *(xi, yi)* (*i* = 0, 1) by modifying DCT coefficients *f(xi, yi), f(yi, xi)*

$$(f(xi, yi), f(yi, xi)) \rightarrow (f'(xi, yi), f'(yi, xi))$$

When mark = 0:

$$f'(x_i,y_i)=\frac{f(x_i,y_i)+f(y_i,x_i)}{2}-\frac{s}{2}$$

$$f'(y_i,x_i)=\frac{f(x_i,y_i)+f(y_i,x_i)}{2}+\frac{s}{2}$$

When mark = 1:

$$f'(x_i,y_i)=\frac{f(x_i,y_i)+f(y_i,x_i)}{2}+\frac{s}{2}$$

$$f'(y_i,x_i)=\frac{f(x_i,y_i)+f(y_i,x_i)}{2}-\frac{s}{2}$$

s is the watermarking strength.

In the detecting scheme, the embedded mark can be detected by comparing $\Sigma f(xi, yi)$ and $\Sigma f(yi, xi)$:

if $\Sigma f(xi, yi) > \Sigma f(yi, xi)$ → mark=1, else → mark=0.

(x0, y0), (x1, y1) can not be too small like (1, 1) or (1, 2), because such coefficients have a big affection on the image, changing their values will make big change on the image. As a result, the 2 indices *(x0, y0), (x1, y1)* are selected from the 7 candidates (1,4), (2,3), (1,5), (2,4), (1,6), (2,5), (3,4). It can be simply generated by the secret key K. In our algorithm, *(x0, y0), (x1, y1)* are selected as the 2 indices that have the smallest value of $|f(xi, yi)-f(yi, xi)|$ (so that the changes are small).

The strength *s* is selected by considering the human vision system. As described in the Fig. 10, the high value of *s* makes the watermarked image noticeable.

Based on the empirical study, we set the basis value of *s* to 0.17. But it can not be a constant; it should be changed to adapt the local region feature:

→in smooth regions (changes are easily noticeable) : decrease s (by -0.05)

→in busy regions (changes are hardly noticeable) : increase s (by +0.05)

The smoothness of the 16x16 block is defined as the smallest variance value of all the 12x12 sub-regions included in it:

$$Variance(R_{n\times n})=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}f_R^2(i,j)-f_R^2(0,0)$$

$$Smoothness(R_{16\times 16})=\min_{r\subset R}(Variance(r_{12\times 12}))$$

($f_R(i, j)$:DCT coefficient of R region)

3.4. Robustness- Repetitive Embedding

To attain higher detection rates, we apply the repetitive embedding method in our research (Fig. 11). In case of the video watermarking, a same bit (mark) is repetitively embedded in continuous frames. In case of the still image watermarking, a same bit is repetitively embedded in different 16x16 blocks within the same object.

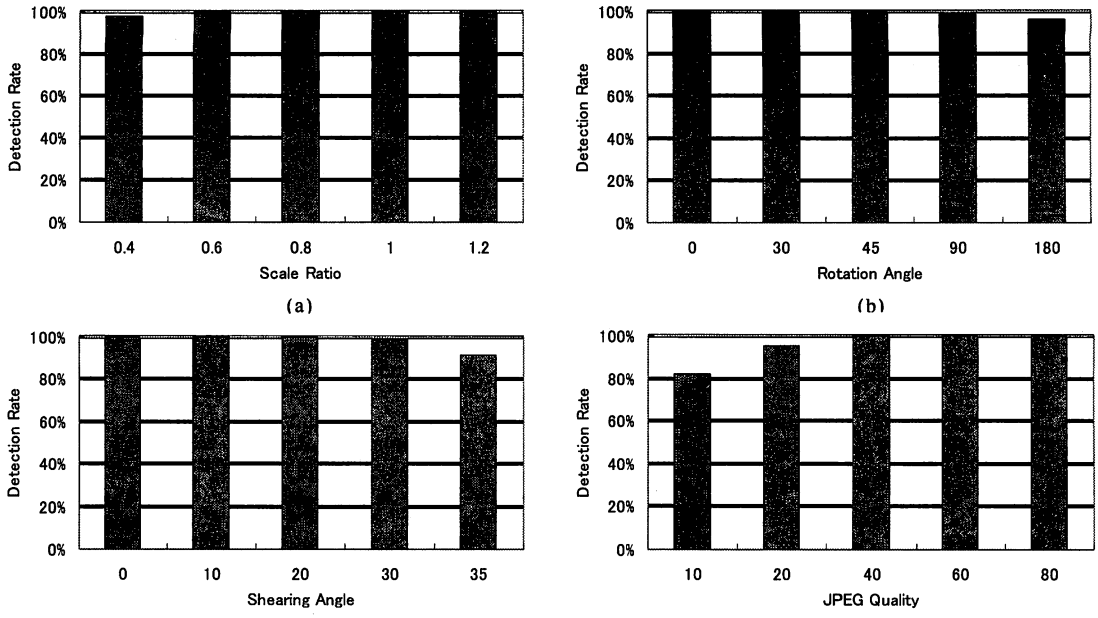


Figure 13: Attack resistance evaluation.

(a) Scaling attack (fixing rotation angle = 30°, shearing angle = 10°)

(b) Rotation attack (fixing scale ratio = 0.6, shearing angle = 10°)

(c) Shearing attack (fixing scale radio = 0.8, rotation angle = 30°)

(d) JPEG compression (fixing scale ratio = 0.6, rotation angle = 30°, shearing angle = 10°).

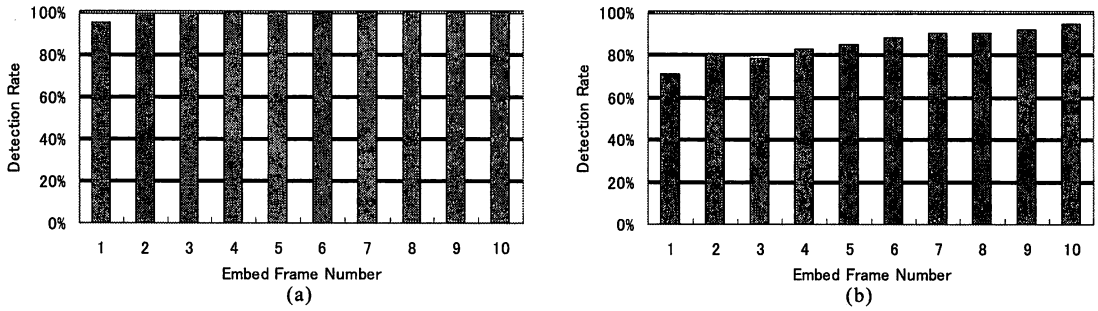


Figure 14: Repetitively embedded frame number evaluation.

(a) Fixing scale ratio = 0.4, rotation angle = 30°

(b) Fixing scale ratio = 1/3, rotation angle = 45°, shearing angle = 20°.

4. Experimental Results

In this experiment, we considered some attacks including affine transformations and the JPEG compression. The still image watermarking is generally the same as the video watermarking. The only difference is that the repetitive embedding frequency for the video watermarking is much higher than that of the still image watermarking. Therefore, we used still images as samples to evaluate the resistance of the method to the above attacks. Then we evaluated the affection of the high repetitive embedding frequency in the video watermarking.

The sample used in our experiments is "LENA" for the still image and "Spider Man" for the video (Fig. 12). For the image "LENA", we generated 166 blocks within the object, and embedded 83 bit into them (each bit was repetitively embedded into 2 blocks). For the video "Spider Man", we selected 10 continuous frames, and then generated 146 blocks within the object of each frame. 146 bits were embedded into these frames (each bit was

repetitively embedded into different frames).

4.1. Resistance Evaluation

The resistance to scaling, rotation, shearing and JPEG compression attacks are shown in Fig. 13. For example, for evaluating the resistance to the scaling attack, we changed the scale ratio while fixing the other parameters and then calculated the detection rates. The other attacks were considered in the same way.

In scaling attacks (Fig. 13(a)), the detection rate can reach the value of 97% when the scale ratio = 0.4, and 100% when the ratio $\geqq$ 0.6.

In rotation attacks (Fig. 13(b)), the detection rate can reach the value of 96% when the rotation angle = 180°, and 100% when the angle $\leqq$ 45°.

In shearing attacks (Fig. 13(c)), the detection rate can reach the value of 91% when the shearing angle = 35°, and 100% when the angle $\leqq$ 20°.

In JPEG compression attacks (Fig. 13(d)), the detection rate can reach the value of 81% when the JPEG quality = 10%, and 100% when the quality $\geqq$ 40%.

4.2. Repetitively Embedding Frequency

By increasing the repetitively embedding frequency, we can attain higher detection rates. That can be proved by the results shown in Fig. 14.

In the normal level of the geometrical attack (Fig. 14(a) - scale ratio = 0.4, rotation angle = 30°), we need only 2 embedded frames to make the detection rate reach the value of 99%.

In the high level (Fig.14 (b) - scale ratio = 1/3, rotation angle = 45°, shearing angle = 20°), 10 frames are needed for the acceptable detection rate (94%).

5. Conclusions

By using the combination of the object matching and the robust watermarking scheme, we have developed a robust geometrically invariant object-based watermarking. The experimental results show that our proposed method can resist to very strong attacks such as 0.4x scaling, 180° rotation, 30° shearing, JPEG 20% or the combination of both of them. By applying the repetitively embedding robustness, we can get much better results for the video watermarking.

Despite the fact that this method is a non-blind watermarking, its robustness to a wide variety of attacks is suitable for many applications requiring high watermarking reliability and capacity.

Reference

[1] D.G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, in press (2004).

[2] P. Bas, J.-M. Chassery, and B. Macq, "Geometrically invariant watermarking using feature points", IEEE Trans. Image Processing, vol. 11, pp. 1014--1028, Sept. 2002.

[3] CW Tang and HM Hang, "A feature-based robust digital image watermarking scheme", IEEE Trans. Signal Process., vol.51, no.4, pp.950–959, April 2003.

[4] Yu-Kuen Ho, Mei-Yi Wu, "Robust object-based watermarking scheme via shape self-similarity segmentation", Pattern Recognition Letters, v.25 n.15, p.1673-1680, November 2004.

[5] Dajun He, Qibin Sun and Qi Tian, "An Object Based Watermarking Solution for MPEG4 Video authentication", submitted to *ICASSP 2003*.

[6] Fabien A.P. Petitcolas, Ross J. Anderson, and Markus G. Kuhn, "Attacks on Copyright Marking Systems", Information Hiding 2nd International Workshop, April 1998.

[7] "Spider Man 3" trailer from www.themovieinsider.com.