

A SIFT-based Image Fingerprinting Approach Robust to Geometric Transformations

Xinghua Yu

School of Information
Graduate University of Chinese Academy of Science
Beijing, China
xhyu@jdl.ac.cn

Tiejun Huang

Institute of Digital Media
Peking University
Beijing, China
tjhuang@jdl.ac.cn

Abstract—Among approaches in implementing Digital Rights Management, image fingerprinting technique is considered to be one of the most attractive solutions, especially in detecting illegal use of image works. The deficiency of the existing image fingerprinting methods is that they can not deal with geometric transformations, such as aspect ratio changes, rotations, cropping, combining, *etc.* Aiming at this shortcoming, we propose a SIFT-based image fingerprinting algorithm which is robust to geometric transformations. Firstly, we introduce SIFT-based algorithm to extract features as a unique fingerprint. Secondly, a method based on area ratio invariance of affine transformation is utilized to verify valid matched keypoint pairs between the queried image and the pre-registered image. Finally, by counting the valid matched pairs, we estimate whether the two images are homologous or not. Experimental results demonstrate that the proposed method exhibits an excellent performance when geometric transformation occurs.

I. INTRODUCTION

As the digital media technique is developing rapidly, DRM (Digital Rights Management) for images has become an extraordinary urgent issue, and draws increasing attentions. Among the existing DRM approaches for images, the image fingerprinting technique [1] is considered as an efficient and feasible one. Generally, an image fingerprint [2] is a unique identifier of one image. The prerequisite qualification of an image fingerprint is that it should be highly distinctive from heterogeneous images and robust to extensive kinds of modified homologous images [3]. A practical image fingerprinting algorithm should generate fingerprints satisfying the above requirements.

To establish practical image fingerprinting algorithms, researchers have made considerable efforts. Jin et.al. [2] presented a Radon transform based method to extract the fingerprints of images. They obtained fingerprint bit sequences through nonlinear operation and random permutation of image's Radon feature. Paul et.al. [4] proposed a technique based on the Trace transform, which has been accepted by MPEG. After applying several kinds of functionals over all possible lines in an image, they generated

image representations from the acquired different Trace transforms of the image.

Although these methods can handle many cases, there are some complicated modifications listed in [3] that they do not deal with well, such as image aspect ratio change, cropping, and so on. Obviously, these geometric transformations damage global features of an image. Since the existing methods all make use of global features, they are not robust to such modifications.

To tackle this problem, we propose a SIFT-based image fingerprinting approach which is robust to geometric transformations. SIFT feature is a local feature that is robust to a substantial range of affine distortions, and are highly distinctive. Fortunately, most geometric transformations listed in [3] are affine transformations, thus the characteristics of SIFT can help us to establish an effective image fingerprinting algorithm. In general, the algorithm is designed in four steps. Firstly, the SIFT feature in an image is extracted and treated as the fingerprint. Secondly, the keypoints of the queried image and pre-registered image are compared to obtain the matched keypoint pairs, and the outliers are eliminated. Thirdly, we examine whether the triangles constructed by matched keypoints in the two images conform to the isomorphic relationship of affine transformations. Finally, the number of the conformed ones and the total number of triangles are counted and the ratio between them is used to decide whether the two images are homologous or not. The SIFT-based approach is firstly proposed in our previous work [5]. In this paper, it is improved so as to be not only robust to image local modifications but also robust to image aspect ratio modification and other geometric transformations.

The remainder of this paper is organized as follows. Section 2 reviews the SIFT algorithm briefly. Section 3 describes the proposed image fingerprinting approach in detail. In section 4, the experimental results are presented and analyzed. And we conclude the paper and discuss some future work in section 5.

II. REVIEW THE SIFT ALGORITHM

The SIFT algorithm [6] is a famous algorithm in the field of object recognition. Theoretically, SIFT algorithm is established basing on the solid theory of scale-space. In general, the SIFT algorithm is implemented in five steps.

(a) Generate the scale space. In order to detect effective stable keypoint locations in scale space, we build the difference-of-Gaussian scale-space using the difference-of-Gaussian function to convolve with the original image.

(b) Detect scale-space extremes. The difference-of-Gaussian value of every pixel is compared with that of its 8 neighbors in the same scale and 9×2 neighbors in the two adjacent scales to find the extremes of scale-space. In this way, we can detect all the actual extreme points as candidates for the keypoints.

(c) Localize accurate keypoints. At each candidate location detected in the previous step, the fitted 3D quadratic function is used to localize it to sub-pixel accuracy. Meanwhile the candidates with low contrasts or unstable edge responses will be eliminated. Accordingly, the others will be reserved as the accurate keypoints.

(d) Assign orientations of the keypoints. An orientation histogram is formed from the gradient orientations of sample points within a region around one keypoint. And the highest peak corresponds to its dominant orientation. After this step, every keypoint will have a triple of attributes: location, scale and orientation.

(e) Build the local image descriptor for each keypoint. This is based on a patch of neighboring pixels in the local neighborhood of each keypoint. Their gradient orientation histogram information is used to generate the descriptor of the keypoint. The descriptor and the triple of attributes together constitute the feature of a keypoint.

SIFT algorithm possesses prominent merits in several aspects: (1) the feature is highly distinctive, and is based on local features which can be robust to geometric modifications; (2) the descriptor is carefully designed to be invariant to image changing in scale, location and orientation, and be resilient to affine transformations; (3) matching speed of SIFT feature can be optimized to satisfy practical use. Thus it is attractive to introduce the SIFT approach into the image fingerprinting field.

III. SIFT-BASED IMAGE FINGERPRINTING APPROACH

Intuitively, since the SIFT algorithm is able to localize objects in an image, it can also help us to determine whether two images contain the identical content. Basing on such consideration, we introduce the SIFT algorithm into the image fingerprinting system. Generally, the proposed approach consists of three major steps: extract fingerprint, match keypoint pairs and detect homology. Figure 1 depicts the detail flow chart of the proposed method.

A. Extract Fingerprint

Initially, a pre-process is applied to every original image. While maintaining the aspect ratio, we resize the original image to a size of $256 \text{ pixels} \times W \text{ pixels}$ or $W \text{ pixels} \times 256 \text{ pixels}$,

where $W \leq 256$. The resized image will replace the original one in the following algorithm.

After the pre-process, the SIFT keypoints and their corresponding descriptors of the image will be extracted, and then constitute the fingerprint of this image. Also, the fingerprints of the pre-registered images have been extracted analogically and stored in the database beforehand. The fingerprint of queried image will be compared with each fingerprint in the database. In the following descriptions, we will take the queried image A and one pre-registered image B as an example.

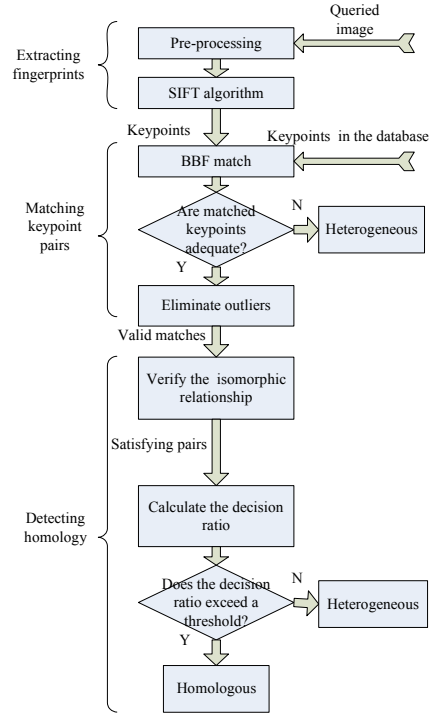


FIGURE 1. THE FLOW CHART OF THE PROPOSED METHOD.

B. Match Keypoint Pairs

With the fingerprints of image A and B, we then detect the matched keypoint pairs between them. Two steps are involved to find the accurate and effective pairs.

Firstly, we obtain the preliminary matched keypoints by constructing a k-d tree [7]. For the sake of stability, the threshold of distance ratio in [7] is reset to be 0.4. In addition, the Best-Bin-First (BBF) algorithm [8] is used to accelerate the feature matching speed. If the number of matched keypoint pairs is larger than a threshold T , which is suggested to be 6 by experiments, it is possible that image A and B are homologous. Thus we will continue to the next step. Otherwise, we will arrive at a conclusion that the two images are heterogeneous.

Secondly, we find the valid matched keypoints by decimating the outliers among the preliminary matched keypoints. Inevitably, there are always some incorrectly matched keypoints named outliers. Since outliers can deteriorate the result seriously, they must be eliminated

previously. Intuitively, the matched keypoints in one image should be viewed as a cluster, if image A and B are homologous and all matches are correct. Hence we can determine whether one matched keypoint is an outlier or not by its distance from the center of the cluster. In practice, we design an iterative process as below to eliminate the outliers in an image.

(a) Suppose all the preliminary matched keypoints as valid ones initially.

(b) Calculate the centroid $P_c(x_c, y_c)$ of all valid matched keypoints as

$$x_c = \frac{1}{M} \sum_{i=1}^M x_i \text{ and } y_c = \frac{1}{M} \sum_{i=1}^M y_i, \quad (1)$$

where M is the number of the valid matched keypoints in the image, $P_i(x_i, y_i)$ is the i th matched keypoint. In addition, the distance from keypoint i to the centroid is defined to be its central distance, denoted as d_i .

(c) Compute the mean and the standard deviation of central distances for all matched keypoints, signified as μ_d and σ_d .

(d) Remove outliers. A matched keypoint i is considered as an outlier if

$$d_i \geq \mu_d + \alpha \cdot \sigma_d, \quad (2)$$

where α is the tolerating factor, and is set to be 3 empirically.

(e) If no keypoint satisfies the outlier condition (2), the process will be terminated. Otherwise, return to step (b) and execute iteratively.

C. Detect Homology

After selecting the valid matched pairs, we will detect homology of image A and B by verifying whether the valid pairs satisfy the isomorphic relationship of affine transformation. An important property of affine transformations is that they can preserve the ratio of areas [9]. Accordingly, the original plane and its transformed plane are isomorphic [10].

A typical affine transformation is given by

$$x' = c_{11}x + c_{12}y + c_{13}, \quad y' = c_{21}x + c_{22}y + c_{23}, \quad (3)$$

where the (x, y) , (x', y') are the Cartesian coordinates of the point P and of its transformed point P' . Moreover, for a nonsingular transformation, we have

$$C = c_{11}c_{22} - c_{12}c_{21} \neq 0. \quad (4)$$

When the origin is fixed, we can discard the translation expressed by c_{13} , c_{23} and obtain the *centro-affine* transformation expressions as

$$x' = c_{11}x + c_{12}y, \quad y' = c_{21}x + c_{22}y. \quad (5)$$

Taking the *centro-affine* transformation as example, we proof the invariance of the ratio of areas. The area S of a directed triangle OPQ , with $P(x_1, y_1)$, $Q(x_2, y_2)$, is

$$S = k \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}, \quad (6)$$

where k is a constant which is equal to $(\sin\theta)/2$ and $\theta = \angle XOY$ in the Euclidean interpretation of affinities. After the transformation, the area S' of the transformed triangle $OP'Q'$, with $P'(x_1', y_1')$, $Q'(x_2', y_2')$ can be calculated as

$$S' = k \begin{vmatrix} x_1' & x_2' \\ y_1' & y_2' \end{vmatrix} = k \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} = kC_1S. \quad (7)$$

By a passage to the limit, this result can be extended to the area of any figure in the plane. The ratio of areas is therefore an affine invariant.

Profiting from the robustness of SIFT, most of the valid matched keypoints can keep their relative locations invariant to affine transformations. Abstractly, valid matched keypoints in image A and B can be treated as two spaces, while the keypoints pairs matching can be seen as a mapping. In a formulation way, we say

$$L(x_1, x_2, \dots, x_n) = L(f(x_1), f(x_2), \dots, f(x_n)), \quad (8)$$

where L is a functional operator and f is the isomorphic mapping. x_i and $f(x_i)$ belongs to the two spaces respectively. Basing on this isomorphism, we design the following method to detect the homology between image A and B. In practice, we make use of the invariance of ratio of areas and utilize it as the functional operator.

In the first step, we use S_i to denote the area of the triangle with $P_i(x_i, y_i)$, $P_{i+1}(x_{i+1}, y_{i+1})$ and $P_c(x_c, y_c)$ as vertices in image A, use S_i' to denote the corresponding area in image B. $P_i(x_i, y_i)$ and $P_{i+1}(x_{i+1}, y_{i+1})$ here are two valid matched keypoints and $P_c(x_c, y_c)$ is the centroid of all valid matched keypoints. If the isomorphism exists, in term of (8), we have

$$\frac{S_i}{S_{i+1}} = \frac{S_i'}{S_{i+1}'} \Rightarrow \left(c_i \triangleq \frac{S_i}{S_{i+1}} \right) = \left(c_{i+1} \triangleq \frac{S_{i+1}}{S_{i+1}'} \right) = C. \quad (9)$$

If S_i' is zero which means three points are collinear, c_i is excluded from the following manipulations. We estimate C by the mean value of all valid c_i . In order to handle perturbations, the standard deviation σ_c is calculated and condition (9) is relaxed to

$$C - \varepsilon \cdot \sigma_c \leq c_i \leq C + \varepsilon \cdot \sigma_c, \quad (10)$$

where ε is the fluctuating factor and is set to 1.2 experimentally. Suppose the number of valid c_i is L , and n of them satisfies condition (10), then we will have the decision ratio ρ as

$$\rho = \frac{n}{L}. \quad (11)$$

When ρ is larger than a threshold T_0 , it is determined that a homology is detected, otherwise, heterogeneity. Experimentally, T_0 is set to 0.7.

IV. EXPERIMENTAL RESULTS

In this section, the experimental results are presented and discussed. We construct a dataset of images on which the proposed method is run. The dataset is composed by 3000 independent images selected from the CD-ROMs "Art Explosion 800000" by Nova [11]. Fingerprints are extracted

from all images in the dataset previously to build the pre-registered image database. By applying six types of modifications to images in dataset, we obtain the corresponding modified versions of them. The modifications include aspect ratio changes, rotation, scaling, cropping, embedding and combining.

For a queried image, we perform the proposed matching process with each fingerprint of the pre-registered images in the database. To evaluate the performance of our method, we calculate the *false-alarm rate* and the *recall* as

$$\text{false-alarm rate} = \frac{\text{number of incorrect matches}}{\text{total number of processes}} \text{ and} \quad (12)$$

$$\text{recall} = \frac{\text{number of correct matches}}{\text{total number of homologous image pairs}}. \quad (13)$$

A *match* is a pair of images which are claimed to be homologous. An *incorrect match* is a *match* that is claimed falsely and a *correct match* is one that is claimed correctly. For each kind of modification, there are 3000 pre-registered images and 3000 corresponding transformed images as queried images. Therefore, the total number of processes is $3000 \times 3000 = 9000000$ and the total number of homologous image pairs is 3000. For comparisons, the MPEG proposal [4] is also tested.

The results are depicted in Table 1. As we can see, the proposed method can handle complicated geometric transformations fairly well. Contrarily, the performance of the MPEG proposal is not so excellent. This indicates that our method can deal with complicated modifications much better than MPEG proposal. The reason is that the MPEG method uses the global features while our method uses local ones, which are robust to geometric transformations.

The computational complexity of the proposed method is acceptable. On a PC with Pentium 3.4GHz processor, the average time to extract a fingerprint is about 0.500 second, and the matching process is about 0.001 second. Although the original SIFT features demand a considerably large storage size, they can be compressed without decreasing the performance of the SIFT features obviously. To be more concrete, we can decrease the dimensions of descriptors using PCA [12] or change parameters of SIFT algorithm to extract only the most stable keypoints.

V. CONCLUSION

Aiming at the geometric transformations of image, we propose a novel SIFT-based image fingerprinting solution and improve it to be robust to aspect ratio change and other affine transformations. With the fingerprint comprised by the SIFT feature, we design a keypoints matching strategy according to area ratio invariance of affine transformations. Experiments prove that the proposed method performs quite superiorly in detecting geometric transformed homology and in distinguishing heterogeneity. In the future, we will explore ways to compress the fingerprint more effectively, so as to make the algorithm more applicative.

ACKNOWLEDGMENT

This work is supported by the National Key Technology R&D Program [2006BAH02A10 and 2006BAH02A13] of China.

REFERENCES

- [1] P. Cano, E. Batlle, T. Kalker, J. Haitsma: Review of Algorithms for Audio Fingerprinting. Multimedia Signal Processing IEEE workshop (2002)169- 173.
- [2] J. S. Seo, J. Haitsma, T. Kalker: A robust image fingerprinting system using the Radon transform. Signal Processing: Image Communication, Vol. 19. (2004)325-339.
- [3] MPEG Video Sub-Group: Call for proposal on Image & Video Signature Tools. MPEG Doc No.N9216 (July, 2007).
- [4] P. Brasnett, M. Bober: Proposed Improvements to Image Signature XM 31.0. MPEG Doc No.M14983 (Oct. 2007).
- [5] Xinghua Yu, Tiejun Huang: An image fingerprinting method robust to complicated modifications. International Conference of Intelligent Information Hiding and Multimedia Signal Processing (2008).
- [6] D.G. Lowe: Distinctive image features from scale invariant keypoints. International Journal of Computer Vision, Vol. 60(2). (2004)91-110.
- [7] J.H. Friedman, J.L. Bentley, and R.A. Finkel: An algorithm for finding best matches in logarithmic expected time. ACM Transactions on Mathematical Software, Vol. 3 (1977)209-226.
- [8] J. Beis, and D.G. Lowe: Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. Conference on Computer Vision and Pattern Recognition, Puerto Rico (1997)1000-1006.
- [9] D.J. Struik: Analytic and Projective Geometric. Addison-Wesley Publishing Company, Inc. Cambridge 42, Massachusetts (1953)91-95.
- [10] S. Buchin: Affine Differential Geometry. Gordon and Breach, New York (1983)25-70.
- [11] Avanquest Software: Art Explosion 800000 (Oct. 2004).
- [12] Y. Ke, R. Sukthankar: PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In Proc. the IEEE Conference on Computer Vision and Pattern Recognition, Vol.2 (2004)506- 513.

TABLE I. EXPERIMENTAL RESULTS

	MPEG		Proposed method	
	false-alarm rate	recall	false-alarm rate	recall
Aspect ratio change (width*1.33)	0.0004%	7.7%	0.093%	88.3%
Aspect ratio change (width*1.5)	0.0005%	4.5%	0.097%	85.6%
Rotation (25°, 45°, 90°, 180°)	0.0006%	99.7%	0.0003%	99.8%
Cropping (50%, 60%, 90%)	0.0004%	11.2%	0.0005%	95.6%
Scaling (50%, 70%, 90%)	0.0006%	99.3%	0.0003%	99.7%
Embedding and combining	0.0006%	4%	0.0002%	93.4%
Average	0.00052%	37.73%	0.032%	93.73%