AN IMAGE FINGERPRINTING SYSTEM BASED ON COLOR HISTOGRAM OF AFFINE COVARIANT REGION

Jinho Choi, Dalwon Jang and Chang D. Yoo

Div. of EE, School of EECS, KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon, Korea
cjh3836@kaist.ac.kr, dal1@kaist.ac.kr, cdyoo@ee.kaist.ac.kr

ABSTRACT

In this paper, we propose a content-based image fingerprinting system which uses color histograms of affine covariant regions as a local fingerprint of an image. Given a query, the system first extracts local fingerprints from affine covariant regions. The local fingerprints are based on the color histogram, and the affine covariant regions are detected using the maximally stable extremal region (MSER) detector. After extracting local fingerprints, the system searches local fingerprints in fingerprint database for the nearest neighbor of the query local fingerprints. Then, the query image is identified. Experimental results show that that our system is robust against cropping, resizing, rotation, and small color changes with small computational cost.

***Index Terms*—** Image fingerprinting system, local fingerprint, MSERs, color histogram

1. INTRODUCTION

An image fingerprinting system identifies a query image using feature vectors called fingerprints. Various image fingerprinting systems having have been proposed[1, 2]. The main application of image fingerprinting system is file filtering for copyright protection[1]. We propose an image fingerprinting system that is robust against geometric distortions such as cropping, rotation, resizing, and small color changes. For this, instead of using global fingerprint that is vulnerable to geometric transformations such as horizontal or vertical shift (translation), rotation, cropping, etc., we use the local fingerprint based on color histogram of affine region which is a connected region covariant with a class of affine transformations [3, 4]. In our system, given a query image, the system extracts local fingerprints of the query image and searches the fingerprint database (DB) for its nearest neighbor based on a distance threshold. Finally, the query image is identified based on a region rate threshold.

For various affine region detectors, we use the maximally stable extremal regions (MSER) detector [5] since it has less computational complexity. The MSERs are connected components of an image where local binarization is stable over a large range of thresholds. The MSERs are invariant to common photometric changes, and are covariant to adjacency preserving transformation, etc. In [4], it is evaluated that the MSERs detector performs better than other affine covariant region detectors such as the Harris or Hessian affine region detector while requiring low computational complexity.

Considering various image descriptors which can be used as a local fingerprint based on MSER, we select a color histogram since it is easily computed and has robustness against rotation without any normalization of rotation. Content-based image retrieval system uses various descriptors such as the color, texture, shape, structure of image, and so on [6, 7]. Among them, the color histogram is robust against the distortions except drastic color distortion [8] and is not influenced by rotation.

The rest of this paper is organized as follows-Section 2 describes our image fingerprinting system of four modules. Section 3 presents the system evaluation for various distorted images. Section 4 concludes the paper.

2. OUR IMAGE FINGERPRINTING SYSTEM

Fig. 1 illustrates our image fingerprinting system for four modules; pre-processing, resizing, MSERs detection, feature extraction and search modules.

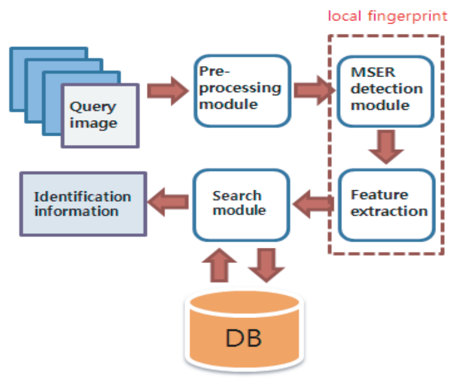

Fig. 1. Our image fingerprinting system

2.1. Pre-processing module

The pre-processing module decodes compressed query images and resizes the decoded images. This module can decode images compressed by various image compression format such as JPEG, GIF, PNG, and so on, and we can get raw RGB image data. After decoding, the images with various sizes are resized to $R \times R$ sized image for computational efficiency. The module uses CxImage library for decoding and resizing. To resize the image, we apply re-sampling with Lanczos filter, which is known as approximate optimal re-sampling filter. We have set R as 400 in our system.

2.2. MSERs detection module

To extract the fingerprint robust against cropping, resizing, rotation and small color change, local fingerprints are extracted based on affine covariant regions. The MSER detector [5] finds the affine covariant regions which is a connected region covariant with a class of affine transformations [4]. In our system, both the MSERs+ (containing regions with the inside brighter than outside) and MSERs- (the opposite contrast) are detected for an image. Two sets of MSERs leads to more robustness for various distortions[9]. Fig. 2 shows MSERs+, MSERs- for an example image. The ellipses in Fig. 2 are the boundaries of MSERs+, MSERs-. Even though the image is distorted with affine transformation, the MSER detector finds the covariant regions in both the original and distorted images. The covariant regions have similar color histograms. Fig.3 describe that the color histograms of covariant regions are very similar.

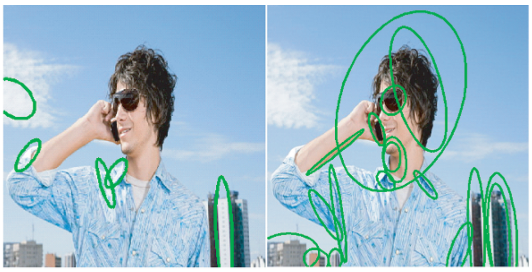

Fig. 2. Left : MSERs+, Right: MSERs-

2.3. Feature extraction module

Local fingerprint is extracted for each MSER using the color histogram. It is robust against the distortions except drastic color distortion [8]. Thus, this local fingerprint is proper to our system that has the restricted distortions; cropping, resizing, rotation and small color change. The color histogram of a MSER is obtained by counting the number of pixels satisfying two constraints- that the color values of pixels are in given set of ranges and that the pixels are in the MSER. For the ease of explanation, set of pixels in ith MSER is denoted as M_i.

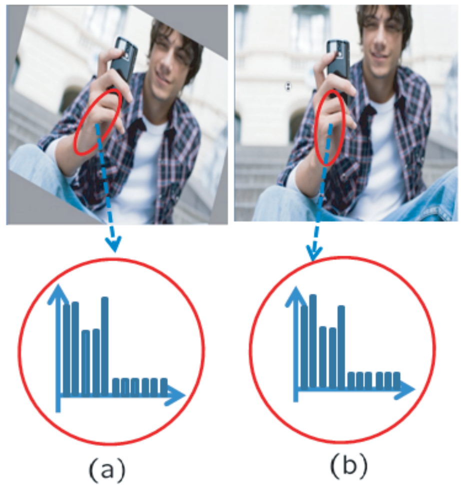

Fig. 3. The color histograms of covariant regions that are detected by MSER detector for a rotated image (a) and the original image (b)

The coordinate is denoted as (x, y), and R, G, and B values of pixel (x, y) is denoted as $R(x, y)$, $G(x, y)$, and $B(x, y)$, respectively . The 3 dimensional color space is divided into L^3 bins, and each bin is defined the range of R, G, and B values. Each R, G, and B space is divided by L levels by quantization function $Q(\cdot)$ which is defined as

$$Q(X) = \left\lfloor \frac{X}{(256/L)} \right\rfloor \quad (1)$$

where $\lfloor r \rfloor$ outputs the largest integer smaller than r. The quantized RGB color values $Q(R(x, y))$, $Q(G(x, y))$, and $Q(B(x, y))$ are in range of $\{0, 1, \cdot, L-1\}$. From these $Q(R(x, y))$, $Q(G(x, y))$, $Q(B(x, y))$, we construct the color histogram. Color histogram of M_i, $\mathbf{H}_i$ is L^3-length vector and jth element $H_i(j)$ is represented as

$$H_i(j) = \frac{1}{Z(M_i)} \sum_{(x,y) \in M_i \cap S_j} 1 \quad (2)$$

where $j = Q(R(x, y)) * L^2 + Q(G(x, y)) * L + Q(B(x, y))$ that is the index of bin. The set S_j is the set of pixels that satisfy $j = Q(R(x, y)) * L^2 + Q(G(x, y)) * L + Q(B(x, y))$. The normalization factor $Z(M_i)$ is the number of pixels in M_i. The local fingerprint as the color histogram of a MSER, is extracted for each MSER. In our system, quantization level L is 4.

2.4. Search module

Fig.4 describes MSERs and local fingerprints of query and original images. The ellipses are detected MSERs. H_n^q and H_m^o represent the local fingerprints of nth , mth MSER of the query image (superscript q) and the original image (superscript o), respectively. In DB, local fingerprints of original

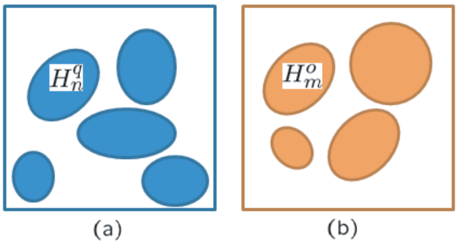

Fig. 4. (a) MSERs (ellipses) and those local fingerprints of a query image, (b) MSERs (ellipses) and those local fingerprints of the original image

images are stored. The local fingerprint of the query image are compared with those. Let N, M as the number of local fingerprints for a query image and an original image, respectively. We use two similarity measures for the similarity of two images. One is the similarity of local fingerprints. This is simply defined as distance between local fingerprints like,

$$D(H_n^q, H_m^o) = |H_n^q - H_m^o|, \tag{3}$$

where $n = 1, ..., N,\ \ m = 1, ..., M$. This distance is calculated for all possible (n, m) combinations. The number of (n, m) combinations that those distances are below a fixed distance threshold, T_1 (determined by training) is defined as $N_{D(H_n^q, H_m^o)<T_1}$. $N_{D(H_n^q, H_m^o)<T_1}$ is used for the second similarity measure as the region rate. The region rate is defined as,

$$R = \frac{N_{D(H_n^q, H_m^o)<T_1}}{N} \tag{4}$$

R is compared with a fixed region rate threshold, T_2 (also determined by training). If the R of a query is more than T_2, the query image is identified as a distorted version of original image.

3. SYSTEM EVALUATION

The system evaluation was performed on Intel(R) Core(TM)2 2.13GHz CPU with 1.0GB of RAM. Image processing and visualizing of the system is created on Visual C++ 6.0 based MFC. Three distortions considered in our system are subject to the following specifications:

- Image cropping: 50 - 90 % of the central portion of the image are retained while the boundaries are removed.
- Image resizing: Isotropic/non-isotropic resizing with scaling factors from 0.5 to 1.5 for the size (width, height) of an original image.
- Image rotation: at angle from 1 to 359 degrees.

3.1. Training thresholds T_1, T_2

The distance threshold T_1 and region rate threshold T_2 are determined by training with various distorted images. The number of original images is 3000. For each original image, Three classes of the distorted images with cropping, resizing and rotation is made. The original and distorted images are paired to train the thresholds. There are two kinds of pairs; **corresponding image pair**, which is a pair of an original image and its distorted version, and **non-corresponding image pair**, which is a pair of an original image and a distorted version of a different original image. The thresholds are determined by minimizing two error rates; false negative rate, F_n and false positive rate,F_p. F_n is defined as the rate that corresponding image pairs are determined as non-corresponding image pairs, and F_p is defined as the rate that non-corresponding image pairs are determined as corresponding image pairs.

The table 1 shows two error rates according to T_1, T_2. We find the proper T_1 and T_2 to minimize two error rates simultaneously for three distortions. In table 1, especially, the trade off of two error rates for cropping is notable. Fig.5 describes that the trade off between F_p, F_n for cropping according to T_1, T_2. Therefore, we should have the balance of the performances to be robust for various distortions including composite distortion of cropping, resizing and rotating. We determine the optimal $T_1 = 0.17, T_2 = 0.22$ as the values of cross points in Fig.5.

T_1	T_2	F_n	F_p
0.16	0.21	cr=0.033, re=0, ro=0	cr=0.024, re=0.030, ro=0.033
	0.23	cr=0.058, re=0, ro=0	cr=0.021, re=0.025, ro=0.027
	0.25	cr=0.070, re=0, ro=0	cr=0.020, re=0.022, ro=0.023
0.18	0.21	cr=0.025, re=0, ro=0	cr=0.033, re=0.036, ro=0.041
	0.23	cr=0.033, re=0, ro=0	cr=0.027, re=0.031, ro=0.041
	0.25	cr=0.041, re=0, ro=0	cr=0.033, re=0.027, ro=0.031
0.20	0.21	cr=0.025, re=0, ro=0	cr=0.043, re=0.049 ,ro=0.054
	0.23	cr=0.025, re=0, ro=0	cr=0.038, re=0.042 ,ro=0.045
	0.25	cr=0.025, re=0, ro=0	cr=0.033, re=0.036 ,ro=0.040

Table 1. two error rates according to T_1, T_2 (cr=cropping, re=resizing, ro=rotation)

3.2. Testing for our image fingerprinting system

Testing for our image fingerprinting system is performed for 6000 **composite distorted** images. 3000 images distorted not from original images and 3000 images distorted from original images are made by hand. In fingerprint DB, there are the fingerprints of 3000 original image. All original images are different from each other. The distortions are also subject to the above specifications. The original image of the most similar local fingerprints (the original image of the highest region rate R) are chosen by T_1. After it, the system determines the identification of the image by T_2. The performance is evaluated

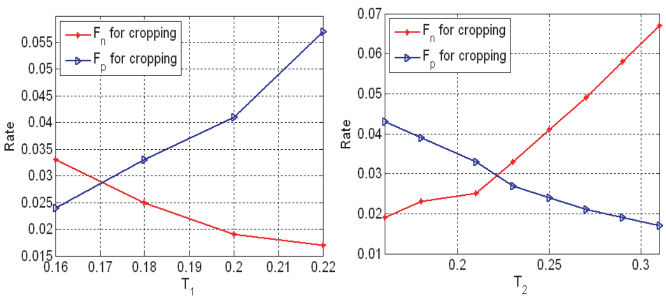

Fig. 5. The trade off between F_p, F_n for cropping according to T_1, T_2

as the identification rate $\mathbf{I_r}$ for detecting the corresponding original image and make the correct choice whether the query image is or not in DB. For the above optimal $T_1 = 0.17$, $T_2 = 0.22$, the identification rate, $\mathbf{I_r}$, is 0.986. respectively. Fig.6 shows that an example of visualization for our system.

Fig. 6. An example of visualization for our system

3.3. Query response time

Query response time (QRT) is defined as follows; the time the system takes to output the determination of the identification for a query image. The average QRT of the system is computed as taking the mean of QRTs of 6000 images. It depends on the number of MSERs which can be adjusted in MSER detection module. Fig.7 shows The average QRT(sec) of our system according to the number of MSERs. We determine the number of MSERs considering the QRT (less than 1 sec) and the above identification rates, $\mathbf{I_r}$.

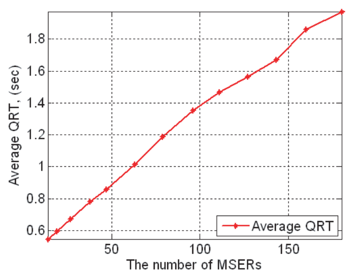

Fig. 7. The average QRT(sec) of our system according to the number of MSERs

4. CONCLUSION

In this paper, an image fingerprinting system based color histogram of affine covariant regions is developed. As local fingerprints, the color histograms of MSERs for an image are used. The experimental results show that our image fingerprinting system is successful for identifying images with high rate for various distortions. Also our system will be enough to work for on-line processing with low complexity in terms of query response time. The further work is to detect the real distorted images linked with high level semantic information.

5. REFERENCES

[1] J. S. Seo, J. Haitsma, T. Kalker, and C.D. Yoo, "A robust image fingerprinting system using the Radon transform," *Signal Processing: Image Communication*, vol. 19, no. 4, pp. 325–339, 2004.

[2] C. Kim, "Content-based image copy detection," *Signal Processing: Image Communication*, vol. 18, no. 3, pp. 169–184, 2003.

[3] C. Schmid and R. Mohr, "Local Grayvalue Invariants for Image Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 530–535, 1997.

[4] C. Schmid A. Zisserman J. Matas F. Schaffalitzky T. Kadir K. Mikolajczyk, T. Tuytelaars and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1/2, pp. 43–72, 2005.

[5] M. Urban J. Matas, O. Schum and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *British Machine Vision Conference*, pp. 384–393, 2002.

[6] A. Natsev, R. Rastogi, and K. Shim, "WALRUS: A Similarity Retrieval Algorithm for Image Databases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 301–316, 2004.

[7] R. Datta, J. Li, and J.Z. Wang, "Content-based image retrieval: approaches and trends of the new age," *The 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 253–262, 2005.

[8] A. M. Ferman, A. M. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 497–508, 2002.

[9] O. Chum and J. Matas, "Geometric hashing with local affine frames," *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.