AFFINE TRANSFORM RESILIENT IMAGE FINGERPRINTING

Jin S. Seo[1], *Jaap Haitsma*[2], *Ton Kalker*[2] *and Chang D. Yoo*[1]

[1]Dept. of EECS, KAIST, 373-1 Guseong Dong, Yuseong Gu, Daejeon 305-701, Korea
jsseo@mail.kaist.ac.kr, cdyoo@ee.kaist.ac.kr
[2]Philips Research Eindhoven, Prof. Holstlaan 4, 5656AA, Eindhoven, The Netherlands
{jaap.haitsma,ton.kalker}@philips.com

ABSTRACT

Affine transformations are a well-known robustness issue in many multimedia fingerprinting systems. Since it is quite easy with modern computers to apply affine transformations to audio, image and video content, there is an obvious necessity for affine transformation resilient fingerprinting. In this paper we present a new method for affine transformation resilient fingerprints that is based upon the auto-correlation of the Radon transform, the log mapping and the Fourier transform. Besides robustness, we also address other issues such as security, database search efficiency and independence with perceptually different inputs. Experimental results show that the proposed fingerprints are highly robust to affine transformations.

1. INTRODUCTION

Multimedia fingerprinting (also known as robust hashing) is an emerging research area that is receiving increased attention. Fingerprints are perceptual features or short summaries of a multimedia object. This concept is an analogy with cryptographic hash function that maps arbitrary length data to a small and fixed number of bits [6]. Although cryptographic hashing is a proven method in message encryption and authentication, it is not possible to directly apply it to multimedia fingerprinting. Cryptographic hash functions are bit sensitive: an alteration of a single bit in the content will result in a completely different hash value. This renders cryptographic hash functions not applicable to multimedia objects that often undergo various manipulations including compression, enhancement, geometrical distortions and analog-to-digital conversion during distribution. By noting these deficiencies of cryptographic hash functions we arrive at the notion of multimedia fingerprinting, sometimes referred to as robust hash functions [1] [3]. Promising applications of multimedia fingerprinting are in authentication [8], filtering for file-sharing services [1], supporting digital watermarking [9], automated monitoring for broadcasting stations [2] and automated indexing of large multimedia archives.

Resilience to affine transformations has been one of the main issues in many image processing research areas, such as pattern recognition and watermarking. This paper deals with this important topic in the context of image fingerprinting. To improve robustness to affine transformations, we propose an image fingerprint extraction method that is based on the Radon transform. An image is first projected onto radial directions using Radon transform, and for each radial direction the affine invariant features are extracted based on the auto-correlation, the log mapping and the Fourier transform. The fingerprint bits are determined from the obtained features. The affine invariant features used in this paper have been successfully utilized in extracting speed-change resilient audio fingerprints [2]. This work is an extension of our audio fingerprinting methods in [1] [2]. A different fingerprinting method based on Radon transform was proposed in [5], where the medium point of each projection is used as a fingerprint. However, it needs search and energy modification for rotation and scaling respectively and does not provide keyed hash function [5]. The proposed method does not need any search or modification of fingerprint bits, provides keyed hashing scheme by random permutation and achieves collision-free property with relatively small amount of fingerprint bits (400 bits per image).

This paper is organized as follows. Section 2 describes the requirements of image fingerprints. Section 3 describes the extraction of the affine invariant features used in the proposed method. Section 4 evaluates the performance of the proposed fingerprint.

2. REQUIREMENTS ON IMAGE FINGERPRINTS

A cryptographic hash function $H(X)$ maps an (usually large) object X to a (usually small) hash value. It allows comparing two large objects X and Y, by only comparing their respective hash values $H(X)$ and $H(Y)$. *Mathematical equality* of $H(X)$ and $H(Y)$ implies the equality of X and Y with only a very low probability of error. For a properly designed cryptographic hash function this should be 2^{-L}, where L equals the number of bits in the hash value. However, in case of multimedia fingerprinting the *perceptual similarity* is more important rather than mathematical similarity. We should construct a fingerprint function in such a way that perceptually similar image objects result in similar fingerprints [1]. The modified version of the image should have the same or similar fingerprints with the original image. Requirements on image fingerprints are summarized in [4]. The main requirements for image fingerprints are as follows.

1) Robustness (Invariance under perceptual similarity): the fingerprinting system should give same or similar fingerprints to the severely degraded images originated from the same image.

2) Pairwise independence: if two images are different perceptually, the fingerprints from two images should be different considerably.

3) Randomization (Security): the fingerprint bits should have uniform distribution.

3. PROPOSED FINGERPRINT EXTRACTION METHOD

An overview of the proposed algorithm is shown in Figure 1. First, an image is projected onto N (typically, $N = 512$) radial directions using the Radon transform, and the auto-correlation of each projection is calculated. Through the log mapping and the Fourier transform of the auto-correlation, the affine invariant features are extracted. From the affine-invariant features, a sub-fingerprint (typically 20 bits) is obtained. A sub-fingerprint does not contain enough information to identify an image, but a sequence of sub-fingerprints, which we refer to as a fingerprint block, does. An image fingerprint (fingerprint block) typically contains M (typically, $M = 20$) sub-fingerprints and consequently 400 bits. Details of the proposed method are in the next subsections.

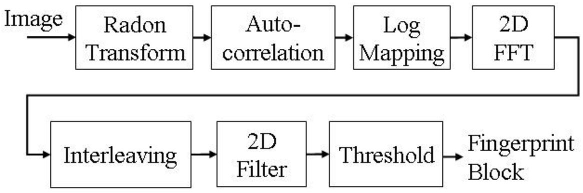

Fig. 1. Overview of Affine transformation resilient fingerprint extraction

3.1. Radon transform and its properties

The Radon transform of an image $f(x, y)$, denoted as $g(s, \theta)$, is defined as its line integral along a line inclined at an angle θ from the y-axis and at a distance s from the origin [7]. Mathematically, it is written as

$$g(s,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - s)dxdy \qquad (1)$$

where $-\infty < s < \infty,\ 0 \le \theta < \pi$. The Radon transform $g(s, \theta)$ is the one-dimensional projection of $f(x, y)$ at an angle θ. The Radon transform has the following useful properties for the affine transformations of an image.

P1) The translation of an image by (x_0, y_0) causes the Radon transform to be translated in the direction of s, i. e.,

$$f(x - x_0, y - y_0) \longleftrightarrow g(s - x_0\cos\theta - y_0\sin\theta, \theta).$$

P2) The scaling (retaining aspect ratio) of an image by a factor ρ ($\rho > 0$) causes the Radon transform to be scaled through the same factor, i. e.,

$$f(\rho x, \rho y) \longleftrightarrow \frac{1}{|\rho|} g(\rho s, \theta).$$

P3) The rotation of an image by an angle θ_r causes the Radon transform to be shifted through the same amount, i. e.,

$$f(x\cos\theta_r - y\sin\theta_r, x\sin\theta_r + y\cos\theta_r) \longleftrightarrow g(s, \theta - \theta_r).$$

3.2. Affine invariant feature extraction

Affine transformations, we consider here, are translation, scaling (retaining aspect-ratio) and rotation. By using the above properties of Radon transform, affine invariant features are obtained.

For translation invariance, the normalized auto-correlation of each radial projection is calculated that is given as follows:

$$c(l,\theta) = \frac{\int_{-\infty}^{\infty} g(s,\theta)g(s-l,\theta)ds}{\int_{-\infty}^{\infty} g(s,\theta)g(s,\theta)ds}. \qquad (2)$$

From P1 the translation of an image causes translation in the Radon domain, but the amount of translation in each projection is different. By taking auto-correlation, we get translation-invariant signal $c(l, \theta)$. Among the affine transformations, scaling and rotation are remained in $c(l, \theta)$. Consider the auto-correlation $c(l, \theta)$ of an original image. From P2 and P3, the auto-correlation of a scaled and rotated image is given as $c'(l, \theta) = c(\rho l, \theta - \theta_r)$ where ρ ($\rho > 0$) and θ_r are the amount of scaling and rotation respectively. To achieve invariance on the scaling and rotation, the log mapping and the 2D Fourier transform are used. The log mapping translates the scaling of the signal to a shift. The subsequent Fourier transform translates this shift into a phase change. By the log mapping $l = e^{\mu}$, the signal $c'(l, \theta)$ can be written as

$$\begin{aligned} c'(l,\theta) &= c(\rho l, \theta - \theta_r) \\ &= c(\exp[\mu + \log\rho], \theta - \theta_r). \end{aligned} \qquad (3)$$

Then the log-mapped signal $\tilde{c}'(\mu, \theta)$ is given by

$$\tilde{c}'(\mu,\theta) = \tilde{c}(\mu + \log\rho, \theta - \theta_r). \qquad (4)$$

The 2D Fourier transform of the log-mapped signal is written as

$$\begin{aligned} C'(\zeta_l,\zeta_\theta) &= \int_0^{\pi}\int_{-\infty}^{\infty} \tilde{c}'(\mu,\theta)\exp[-j\mu\zeta_l - j\theta\zeta_\theta]d\mu d\theta \\ &= \exp[j\zeta_l\log\rho - j\zeta_\theta\theta_r]C(\zeta_l,\zeta_\theta). \end{aligned} \qquad (5)$$

Then the magnitude $|C'(\zeta_l, \zeta_\theta)|$ and phase $\phi'(\zeta_l, \zeta_\theta)$ of the complex signal $C'(\zeta_l, \zeta_\theta)$ are given by:

$$|C'(\zeta_l,\zeta_\theta)| = |C(\zeta_l,\zeta_\theta)| \qquad (6)$$

$$\begin{aligned} \phi'(\zeta_l,\zeta_\theta) &= \zeta_l\log\rho - \zeta_\theta\theta_r + \angle C(\zeta_l,\zeta_\theta) \\ &= \zeta_l\log\rho - \zeta_\theta\theta_r + \phi(\zeta_l,\zeta_\theta) \end{aligned} \qquad (7)$$

where $\angle C(\zeta_l, \zeta_\theta)$ is the phase of the complex signal $C(\zeta_l, \zeta_\theta)$.

As shown above, the log mapping translates scaling into a shift, and the subsequent Fourier transform translates the shift into a phase change. By using the properties, we find features that are invariant to scaling. From Eqn. (6), $|C'(\zeta_l, \zeta_\theta)|$ is affine invariant. Since $\zeta_l \log\rho - \zeta_\theta\theta_r$ in Eqn. (7) is a linear function of ζ_l and ζ_θ, the double differentiation of $\phi'(\zeta_l, \zeta_\theta)$ on ζ_l or ζ_θ is also affine invariant.

3.3. Fingerprint bit extraction

In some applications (for example image authentication), the security of the fingerprint extraction algorithm is an issue. More precisely, it is sometimes required that the fingerprint function depends on a key K. For two different keys K_1 and K_2, the fingerprinting function H should have the property that $H_{K_1}(X) \neq H_{K_2}(X)$ for any image X. To satisfy this requirement, we use interleaving. The coefficients of $C'(\zeta_l, \zeta_\theta)$ are randomly interleaved in either ζ_l or ζ_θ direction by the permutation table (this is key information). After interleaving, the coefficients of $\log |C'(\zeta_l, \zeta_\theta)|$ and $\phi'(\zeta_l, \zeta_\theta)$ are filtered by a simple 2D filter F_{ζ_l,ζ_θ} (along both ζ_l and ζ_θ axes), of which the kernel F_{ζ_l,ζ_θ} equals

$$F_{\zeta_l,\zeta_\theta} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{8}$$

Finally the filter output is converted to bits by taking the sign of the resulting value (thresholding). The output of the filter F_{ζ_l,ζ_θ} is invariant to scaling and translation as we have seen in 3.2. Interleaving does not have any effect on the affine invariance of the filter output.

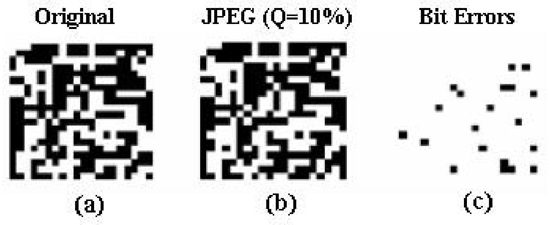

Fig. 2. (a) Fingerprint of original Lena image, (b) Fingerprint of compressed Lena image, (c) the difference between a and b showing bit errors in black (BER=0.05)

Figure 2 shows an example of 20 subsequent 20-bit sub-fingerprints (fingerprint block) extracted with the proposed method from the Lena image. A ‘1’ bit corresponds to a white pixel and a ‘0’ bit to a black pixel. Figure 2a and Figure 2b show a fingerprint block from an original image and JPEG compressed (quality factor 10%) version of it respectively. Ideally these two fingerprints should be identical, but due to the compression some of the bits are erroneous. These bit errors, which are used as *the similarity measure*, are shown in black in Figure 2c.

4. EXPERIMENTAL RESULTS

To evaluate the proposed method, we tested our method on over a hundred images. In the experiments we used a version of the proposed algorithm where we took the luminance of the test image, resized it to 512×512 and filtered it by median filter to make it robust against small geometric attacks. Then we took the 21 by 21 lowest coefficients of $C'(\zeta_l, \zeta_\theta)$ and interleaved the coefficients in ζ_l direction according to the key information. The magnitude and the phase of the 21 by 21 coefficients were filtered by F_{ζ_l,ζ_θ} and thresholded to obtain 20 by 20 bits. Finally, fingerprint bits are determined by the exclusive OR of the 20 by 20 bits from the magnitude and the 20 by 20 bits from the phase. Thus we obtained fingerprint block, consisting of 20 subsequent 20-bit sub-fingerprints, extracted from each image. Using the extracted fingerprints, we tested the proposed method in terms of three requirements in Section 2.

4.1. Robustness of the proposed method

To test robustness of the proposed method, the original images were subjected to various kinds of image processing steps (see [10] for a detailed description of the processing steps) and their respective fingerprint blocks were extracted. The bit error rate (BER) between the original and the processed image fingerprints is shown in Table 1 for four images (Airplane, Boat, Lena and Peppers). The table clearly shows that the proposed method is highly robust to affine transformations, which preserve aspect ratio, (all possible angles of the rotation and the scaling factor ρ larger than 0.15) and other image processing steps including compression and various filtering. However, it was not robust against cropping and shearing (more than 5%). A region based approach is promising to improve robustness to them. This will be covered in future work.

Table 1. BER for different kinds of signal degradations

Processing	AIR	BOAT	LENA	PEP
JPEG (Q=10%)	0.0450	0.0625	0.0500	0.0400
Gaussina filtering	0.0325	0.0300	0.0200	0.0150
Sharpening filtering	0.0825	0.0525	0.0750	0.0325
Median filtering (4×4)	0.0525	0.0525	0.0955	0.0350
Rotation (worst case $45.176°$)	0.1550	0.1625	0.1600	0.1750
Rotation $(90°)$	0.0925	0.0925	0.1025	0.0725
Scaling $(\rho = 0.5)$	0.0350	0.0325	0.0175	0.0175
Scaling $(\rho = 0.15)$	0.1625	0.2075	0.1575	0.0825
Cropping (2%)	0.1350	0.1625	0.1800	0.2025
Cropping (5%)	0.3275	0.3175	0.3450	0.3300
17 column 5 row removed	0.0300	0.0275	0.0450	0.0250
Shearing (1%)	0.1150	0.1100	0.1050	0.1275
Shearing (5%)	0.3400	0.3225	0.2700	0.3450
Random bending attack	0.2150	0.2000	0.3150	0.2800

It is important for any fingerprinting method that it not only results in a low BER but also allows efficient searching. In [1] a search algorithm is presented that exploits the fact that each of the 20 sub-fingerprints in a fingerprint block has a list of most probable candidates for being an *original* sub-fingerprint. This list is generated from soft decoding information during fingerprint extraction of the processed image. In our experiments, a list of 1024 most probable candidates was created for each sub-fingerprint. From these lists the fingerprint database can be searched very efficiently, and with high probability at least one of the 20 lists of the fingerprint block contains a corresponding *original* sub-fingerprint. Table 2 shows the number of lists that contain a corresponding original sub-fingerprint. This number is referred to as the number of database hits. Table 2 shows that the number of database hits are sufficient to database search for most of the image processing steps. We recall from [1] that only a single database hit is needed for a successful search.

4.2. Pairwise independence and security

To test pairwise independence, we extract a fingerprint database from 100 images. Thereafter the BER between all possible pairs of the fingerprints were calculated. Figure 3a shows the histogram of the measured BER. All the measured BER were in the range between 0.39 and 0.6. This shows that the proposed method is approximately pairwise independent. The mean of the measured BER was 0.4917 that is close to 0.5, and the standard deviation of it was 0.0282 that is similar to 0.025 one would expect from random i.i.d. bits. This result is similar to the case of the audio fingerprints in [1]. Through the same false positive analysis in

Table 2. Hits in the database for different kinds of signal degradations

Processing	AIR	BOAT	LENA	PEP
JPEG (Q=10%)	17	17	20	18
Gaussina filtering	20	20	19	20
Sharpening filtering	17	18	16	19
Median filtering (4×4)	19	15	17	18
Rotation (worst case $45.176°$)	11	12	11	13
Rotation ($90°$)	14	15	14	16
Scaling ($\rho = 0.5$)	20	20	19	20
Scaling ($\rho = 0.15$)	8	9	11	13
Cropping (2%)	9	6	9	7
Cropping (5%)	4	2	6	2
17 column 5 row removed	20	18	20	20
Shearing (1%)	10	15	16	11
Shearing (5%)	3	3	7	2
Random bending attack	4	6	5	3

[1], the threshold for the BER was determined to be 0.3 with very low false positive rate. It means that out of 400 bits there must be less than 120 bits in error in order to decide that the fingerprint blocks originate from the same image. Details of the false positive analysis are in [1].

To test security of the proposed method, we generated 100 fingerprints from the Lena image using different interleaving. Similar with the above analysis, the BER between all possible pairs of the fingerprints were calculated. The histogram of the measured BER is shown in Figure 3b. The mean and the standard deviation of the measured BER were 0.5002 and 0.0264 respectively. This result clearly shows the fingerprint is significantly dependent on the key information (interleaving). In terms of security, such a strong dependency on the key is significant. Once a key is broken, the user can simply change it, like a password [3] without modifying overall system.

5. CONCLUSION

For the multimedia fingerprinting, extracting features that allow direct access to the relevant distinguishing information is crucial. The features used in fingerprint extraction are directly related to the performance of the fingerprinting system. In this paper, we presented a new fingerprint extraction method that is resilient to affine transformations. The robustness against affine transformations are essential because it is quite easy to impose affine transformations to images with modern computers. The experimental results show that the proposed method is highly robust against affine transformations and most of the other image processing steps. It was experimentally verified that the proposed image fingerprints satisfy the main requirements of fingerprints; robustness under quality preserving signal processing steps, pairwise independence with different inputs and sufficient randomization (uniform distribution) of fingerprint bits. Future work includes more robust image fingerprinting and extension of the proposed method to video.

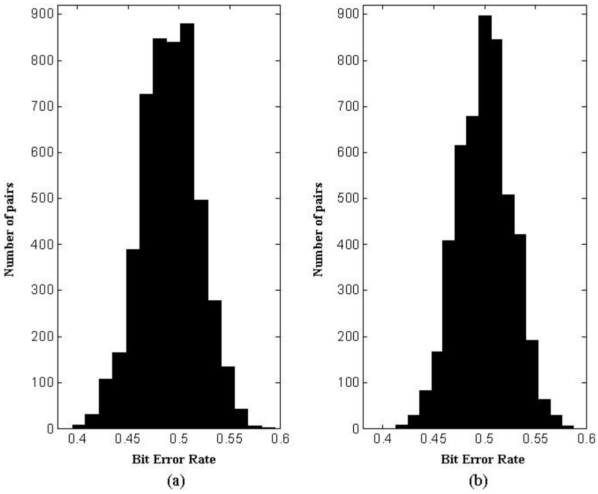

Fig. 3. (a) Histogram of measured BER between the fingerprints from different images, (b) Histogram of measured BER between fingerprints of Lena generated with different keys

6. REFERENCES

[1] J.A. Haitsma and T. Kalker, “A Highly Robust Audio Fingerprinting System,” *Proc. International Conf. on Music Information Retrieval (ISMIR) 2002*, Paris, Oct. 2002.

[2] J.S. Seo, J.A. Haitsma and T. Kalker, “Linear Speed-Change Resilient Audio Fingerprinting,” *Proc. IEEE Benelux Workshop on Model Based Processing and Coding of Audio (MPCA) 2002*, Leuven, Belgium, Nov. 2002.

[3] R. Venkatesan, S.-M. Koon, M. H. Jakubowski, and P. Moulin, “Robust Image Hashing,” *Proc. IEEE ICIP 2000*, Vancouver, CA, Sept. 2000.

[4] M. K. Mihcak and R. Venkatesan, “New Iterative Geometric Methods for Robust Perceptual Image Hashing,” *Proc. ACM Workshop on Security and Privacy in Digital Rights Management*, Philadelphia, PA, Nov. 2001.

[5] F. Lefebvre, B. Macq and J.-D. Legat, “RASH:Radon Soft Hash Algorithm,” *Proc. European Signal Processing Conf. 2002*, Toulouse, France, Sept. 2002.

[6] A. Menezes, P. Oorshot, and S. Vanstone, *Handbook of Applied Cryptography*, CRC press, 1997.

[7] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989.

[8] M. Schneider and S.-F. Chang, “A Robust Content Based Digital Signature for Image Authentication,” *Proc. IEEE ICIP 96*, Laussane, Switzerland, Oct. 1996.

[9] J. Fridrich, “Robust Bit Extraction from Images,” *Proc. IEEE International Conf. on Multimedia Computing and Systems (ICMCS) 99*, Florence, Italy, June 1999.

[10] F.A.P. Fetitcolas, “Watermarking Schemes Evaluation,” *IEEE Signal procssing Mag.*, vol. 17, no. 5, Sept. 2000.