FINGERPRINTING SCHEME FOR FILE SHARING IN TRANSFORM DOMAIN

S.Manikandaprabu[1]

[1]Department of CSE,

Tamilnadu College of Engineering,

Coimbatore, India

smaniprabume@gmail.com

P.Kalaiyarasi[2]

[2]Department of EIE

Sri Ramakrishna Engineering College,

Coimbatore, India

kalaimepsgtech@gmail.com

Abstract

Fingerprinting scheme plays an important role for file sharing. In this paper, a novel method is introduced for fingerprinting an image. The proposed method uses wavelet and Principal Component Analysis (PCA) techniques for fingerprint generation. Fingerprinting is done two phases. During, first phase wavelet techniques are used to obtain low frequency representation of the test image. In second phase, PCA finds the features of the representation; set of fingerprint matrices can be created based on a proposed algorithm. Finally, each matrix combined with the low frequency representation to become a unique fingerprinted matrix. Size of the fingerprinted image is very small but it has most important information. Without this information, quality of the reconstructed image is very poor. We use different wavelet for fingerprint generation. Based on the performance, it is found that the DB6 is the best choice for fingerprint generation. Fingerprinted file is more suitable for distribution of file in peer-2-peer (p2p) networks.

Keywords:-Fingerprinting, Mean square error (MSE), Discrete Wavelet Transform and principal component analysis (PCA)

I. INTRODUCTION

Digital Fingerprinting is an emerging technology to protect multimedia from unauthorized redistribution. It embeds a unique ID into each user's copy, which can be extracted to help identify culprits when an unauthorized leak to found. The goal of digital fingerprinting is to deter people from illegally redistributing digital data. It strongly suggests that the fingerprinting technique is much needed, especially for the multimedia producers who like to share their valuable multimedia with the subscribed customers privately within the public P2P networks.

A good fingerprint technique should be robust against attacks and have negligible impact on the quality of the multimedia file.

In this paper, a digital fingerprinting technique for an image file based on wavelet and principal component analysis (PCA) is proposed.

II. LITERATURE REVIEW

Literature survey shows that very few researchers have worked on fingerprinting for P2P applications so far. Tsolis *et al.* proposed their watermarking scheme recently for P2P application. There are multiple keys as the watermark is cast into the image by the pseudorandom sequence of a Gaussian distribution generator. However, the paper did not mention the robustness of the scheme against common attacks [1]. Many researchers have proposed algorithms mainly for watermarking, and among those watermarking schemes, there are two main streams: the one which embed a watermark directly in the spatial domain and the others which implement it in a frequency domain. It is found that the transform domain watermarking schemes are typically much more robust to image manipulation as compared to the spatial domain schemes [2].

Among the schemes applying wavelet techniques, Kaarna *et al.* proposes an algorithm in the PCA and wavelet-transform domain. They first apply PCA to produce eigen images and then decompose them into multi resolution images. Correspondingly, the watermark image is also decomposed into a multi resolution image in same scale. Finally, the human visual system (HVS) as the strength parameter is adopted for watermark embedding. The scheme is applicable for embedding one mark because of the uniqueness of the strength parameter [3]. Liu *et al.* proposed their algorithm based on singular value decomposition (SVD). The host image is originally presented as USV^{-1}, where the matrix S contains the singular values and UV are singular vectors. The algorithm adds the watermark to the singular values S. Thus, the modified singular value is presented by

$U_wS_wV_w^{-1}$. Then the newly generated singular value S_w will replace the original S to generate the watermarked image. The singular vectors U_w and V_w^{-1} are kept by the owner just for watermark detection. Since S_w approximately equal to S, the visual quality of the image is preserved. To extract the watermark, the watermarked image will be decomposed again using SVD. The main issue of this method is that the attacker can also claim his/her watermark easily by providing another set of singular vectors .It proves that embedding a fingerprint only on singular values is unreliable[4].Hien *et al.* also proposed a PCA method. The difference is they embed the watermark into the eigenvectors. First, the PCA process decomposes the image into eigenvectors and eigen values. Then the image is projected onto each eigenvector and becomes a coefficient matrix. The watermark is embedded into the coefficient matrix based on the selected components. Finally, the watermarked image is obtained by applying the inverse PCA process. The robustness becomes the issue of this method. Because the eigenvectors are normalized, the numerical value of each component of the eigenvector is very small and can be easily corrupted by distortion methods [5]. Some research group uses wavelet and PCA techniques as features were utilized to detect the image information [6] [7].

The proposed fingerprinting technique is illustrated in Section II, and its results are compared in Section III. Section IV discusses the conclusion and future scope.

II. PROPOSED SYSTEM

The proposed technique, Decompose the source file is into two parts: base file and supplementary file. The base file then will carry the embedded unique fingerprint for each peer and be distributed using the traditional server–client mode, while the supplementary file will be freely distributed in P2P networks. Fig. 1 shows the structure of the fingerprint distribution.

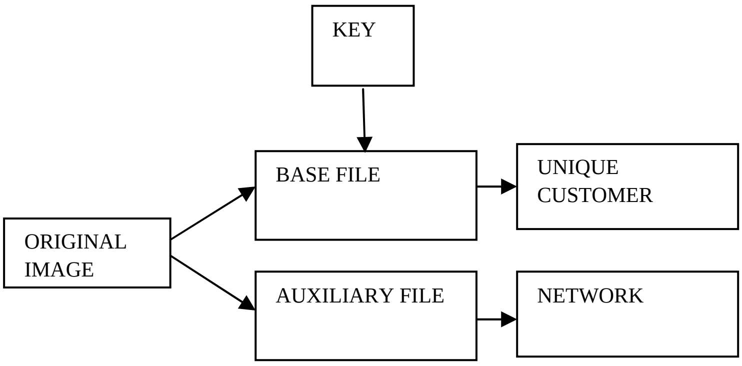

Fig. 1 Fingerprint distribution.

A. FINGERPRINT GENERATION

In our approach, the base file will be distributed from the central server to all the clients. So it should be designed to have small size but contain the most important information. Thus, the load of the server can be alleviated to some extent, while the supplementary file can be larger but contains less important information. By doing this, the peer who has the supplementary file has no commercial motivation to leak the supplementary file alone because of its low quality without the base file. One possible approach to derive a small size base file is to decompose the image into two parts, the base pixel matrix and the detail pixel matrix. The base pixel matrix can give us a rough outline of the image. Since the base pixel matrix has higher correlation information, its entropy value is small so that it can be compressed into a very small size with no quality loss. This is the first time that wavelet and PCA techniques are employed for fingerprint generation and embedding .Fig. 2 shows the flowchart of the fingerprint generation and embedding.

The original image is decomposed into two coefficients which are approximation and detailed coefficients. PCA is carried out on the approximation coefficients. PCA is added with the secret key to generate the fingerprinted image. Size of fingerprinted image is very small but contains important information. To recognize the embedded fingerprint, the receiver needs to decompose the fingerprinted image some level using the wavelet technique.

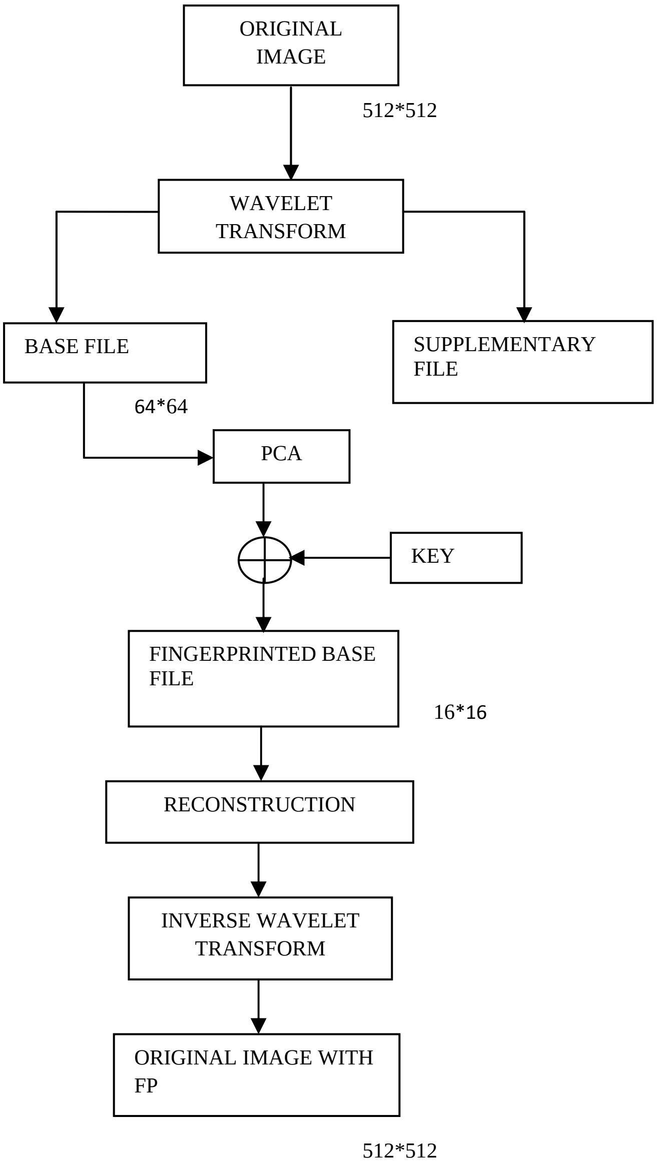

Fig. 2 Fingerprint generation

B. DECOMPOSITION

Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. In wavelet transform, source image is split into two one approximation coefficient and three details in horizontal, vertical, and diagonal direction coefficient. The approximation is then itself split into a second-level approximation and details, and the process is repeated.

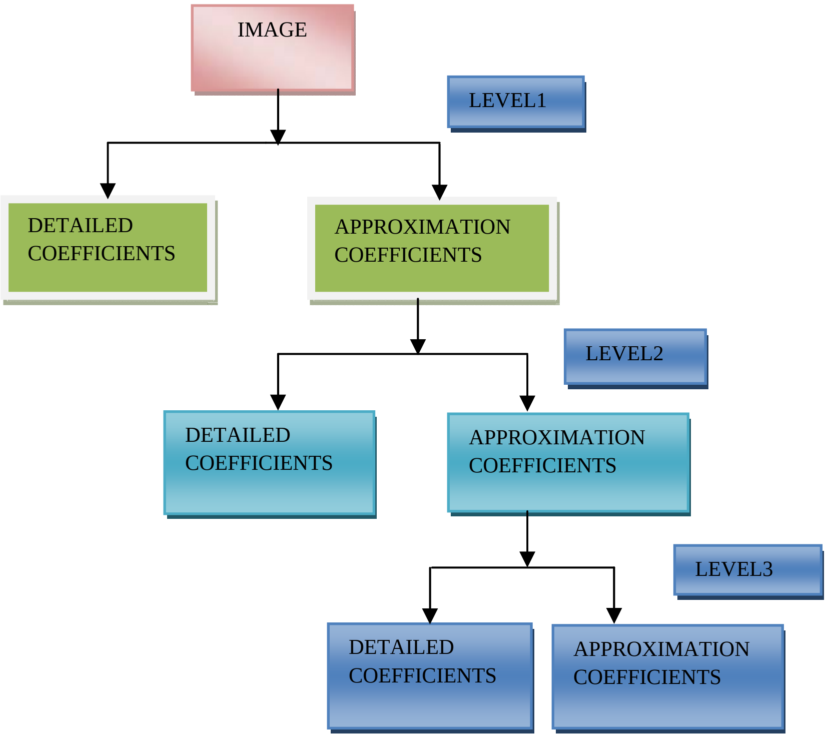

Fig.3.Wavelet Decomposition

For J-level decomposition, the approximation and the details are described in equation 1.

$$
\begin{aligned}
w_{aJ} &= [I.A_J] \\
w_{hj} &= [I.Hj] \\
w_{vj} &= [I.Vj] \\
w_{dj} &= [I.Dj] \qquad j=1,2,3\ldots J \qquad (1)
\end{aligned}
$$

Where I denotes the input image and A_J, Hj , Vj and Dj are approximation, horizontal, vertical and diagonal coefficients. For image decomposition, even the size of the coefficients in a different level is different, but the coefficients are still a 2-Dmatrix. Equation (2) indicates how the image is recovered:

$$
\begin{aligned}
I &= a_J + \sum_{k=j}^{J} d_j \\
&= w_{aJ}\, AJ + \sum_{k=j}^{J} wh_{j.Hj} \\
&\quad + \sum_{k=j}^{J} wvj.Vj + \sum_{k=j}^{J} wdj.Dj \qquad (2)
\end{aligned}
$$

Where I is the reconstructed image and d is the detailed coefficients.

In this method, the fingerprint is small but strong and robust compared to the multimedia file. In this paper, we implement the fingerprint method on both the gray scale and color images. First, the host image with size 512 * 512 was decomposed into the three levels by different discrete wavelet transform. For three-level decomposition, the coefficient set is W= [w_{a3}, w_{h3}, w_{v3}, w_{d3}, w_{h2}, w_{v2}, w_{d2}, w_{h1}, w_{v1}, w_{d1}].At the third level, the size of the approximate coefficient is significantly reduced from original to 64 * 64. This coefficient is called from now on as the base file .Correspondingly; the other coefficients are defined as the supplementary file.

C.PRINCIPAL COMPONENT ANALYSIS (PCA)

The approximation coefficient at the third level was then used to calculate its principals. It goes through three steps according to the following equations:

$$X' = X - \text{mean}(X) \qquad (3)$$

$$X'' = \text{cov}(X', X'^{T}) \qquad (4)$$

$$X''P = PA \qquad (5)$$

Where P and A are the set of eigenvectors and the set of Eigen values of X''. There are a total of 16 eigenvectors which make up the columns of P. It is represented as

$$P = \{P_{16}, P_{15} \ldots.. P_1\} \qquad (6)$$

Where the eigenvectors are arranged in descending order according to their principal components and each eigenvector is a 16 * 1 vector. Equation (7) illustrates how the fingerprint matrix is derived

$$FP = Y * S * P^{T} \qquad (7)$$

Where S is a scale vector, which is a 16 * 1 vector, is multiplied with and one of the eigenvectors is a visually meaningful full matrixes all positive elements, and only known by the source owner. Thus, it can be a company's logo, another low resolution of a portion of the original host image, or simply a portion of the host image. It is utilized to prove the right ownership fingerprint.

D.FINGERPRINT DETECTION

Since only the owner, e.g., the media producer keeps the mapping between the fingerprint and the customer, as long as the producer successfully tracks back the fingerprint for a suspect video, for example, the pirate customer can be revealed. The suspected video is defined as a video which is freely distributed out of the scope of owners' authorized P2P networks.

In our case, to identify the embedded fingerprint, the multimedia producer needs to decompose the fingerprinted image into three level using the wavelet technique so that a 64*64 approximate512*512 matrix is obtained. To prove that the ownership of the fingerprint is issued by the right source owner, the matrix, approximately equal to , is first extracted. Then the correlation coefficient is used to decide if the matrix exists.

III.SIMULATION RESULTS

The images are decomposed up to 3 levels using different wavelets. Principal components are calculated on the approximation coefficients of last level. The decomposed images were reconstructed using different wavelets. We have tested nearly 150 images. In this section two results are presented.

Fig.4.a) shows the original Lena image and Fig.4.b) shows the corresponding fingerprinted image. The image is fingerprinted using 'db6' wavelet.

Fig.5.a) and b) shows the original and fingerprinted image of pepper. The image is fingerprinted using 'db6' wavelet.

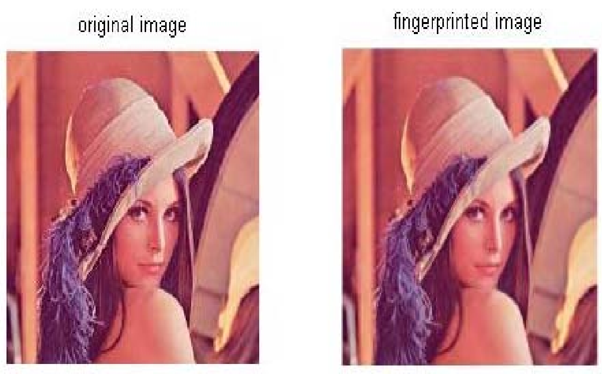

Fig.4. a) Original image b) Fingerprinted image

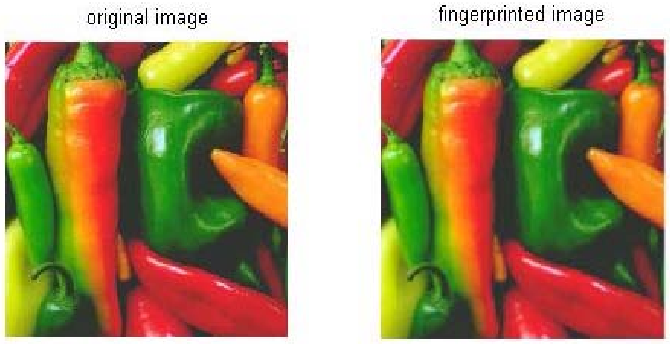

Fig.5. a) Original image b) Fingerprinted image

Decomposed images are reconstructed using inverse wavelet. For best reconstruction, Mean Square Error (MSE) value should be low and Peak Signal-to-Noise Ratio (PSNR) is high. Table 1 show the MSE and PSNR values for lena and pepper image. These values are obtained using different wavelets. By comparing PSNR value MSE for all wavelets it is found that ‘DB6’ gives high value. So DB6 is the best choice for fingerprint generation.

Table1.Comparison of MSE and PSNR value

S.NO	**WAVE NAME**	**LENA IMAGE**		**PEPPER IMAGE**	
		MSE	**PSNR**	**MSE**	**PSNR**
1	db1	283.344	23.6077	244.7524	24.2435
2	db2	203.6546	25.0419	100.3815	28.1143
3	db3	182.3728	25.5212	107.4404	27.8191
4	db4	177.9596	25.6276	87.0419	28.7335
5	db5	175.0453	25.6993	104.1507	27.9542
6	**db6**	**168.2295**	**25.8718**	**78.5383**	**29.18**
7	db7	170.4434	25.815	98.4649	28.198
8	db8	170.5407	25.8125	80.8809	29.0523
9	sym2	203.6546	25.0419	100.3815	28.1143
10	sym3	182.3728	25.5212	107.4404	27.8191
11	sym4	169.0473	25.8507	97.3872	28.2458
12	sym5	177.2524	25.6449	96.4967	28.2857
13	coif1	186.9919	25.4126	122.1511	27.2618
14	coif2	176.6437	25.6598	84.9282	28.8403
15	coif3	183.7821	25.4878	105.9478	27.8799
16	coif4	174.3761	25.7159	81.4537	29.0217
17	coif5	174.6715	25.7086	95.1848	28.3451
18	bior1.1	283.344	23.6077	244.7524	24.2435
19	bior1.3	274.4122	23.7468	188.9695	25.3669
20	bior1.5	297.4032	23.3973	228.1274	24.549
21	haar	283.344	23.6077	244.7524	24.2435

IV .CONCLUSION

The newly proposed P2P fingerprinting scheme is specific for P2P networks and will benefit those multimedia producers who want to share their big files, such as video files, utilizing the convenience of P2P networks. The fingerprint embedding scheme involves the wavelet and PCA techniques. The reason for using the wavelet technique is because it can provide a scalable approximation matrix, and the approximation matrix contains the most important low-frequency information, which gives it strong robustness. The PCA technique, on the other hand, determines the orthogonal eigenvectors, which makes it possible to maximally distinguish the different fingerprints. Unlike other conventional fingerprinting techniques which suffer from poor scalability this scheme is scalable, not only because the scheme reduces the burden of the media owner's server by only sending the small-size base file and making use of the P2P network infrastructure to support the majority of the file transfer process because it provides a large number of unique fingerprints. Moreover, the wavelet technique makes the base file into necessary information for the customer. In addition, the PCA technique is calculated on a small size matrix, which causes low computation complexity Although the scheme proposed in the paper has shown some promising results, much work remains to be done. We are also investigating a new feature to generate the fingerprint that can counter the collusion attack.

REFERENCES

[1] D. Tsolis, S. Sioutas, and T. Papatheodorou, "Digital watermarking in Peer to Peer networks," in *16th Int. Conf. Digital Signal Processing*, Greece, Jul. 2009, pp. 1–5.

[2] R. Dugad, K. Ratakonda, and N. Ahuja, "A new wavelet-based scheme for watermarking images," in *Proc. Int. Conf. Image Processing, 1998 (ICIP 98)*, Oct. 4–7, 1998, vol. 2, pp. 419–423.

[3] A. Kaarna and P. Toivanen, "Digital watermarking of spectral images in PCA/wavelet-transform domain," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp., 2003 (IGARSS '03)*, Jul. 21–25, 2003, vol. 6,pp. 3564–3567.

[4] R. Liu and T. Tan, "An SVD-based watermarking scheme for protecting rightful ownership," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp.121–128, Mar. 2002.

[5] T. D. Hien, Z. Nakao, K. Miyara, Y. Nagata, and Y. W. Chen, "A new chromatic color image watermarking and its PCA-based implementation," in *ICAISC 2006, LNAI 4029*, 2006, pp. 787–795.

[6] E. Chang, J. Wang, C. Li, and G. Wiederhold, "RIME: A replicated image detector for the world wide web," in *SPIE Multimedia Storage and Archiving Systems III*, Bellingham, VA, Nov. 1998, pp. 58–67.

[7] M. Sanchez, X. Binefa, J.Vitria, and P. Radeva, "Local color analysisfor scene break detection applied to TV commercials recognition," in*Proc. Int. Conf. Vis. Inf. Syst.*, 1999, pp. 237–244