Watermarking Based on Principal Component Analysis

Wang Shuozhong

School of Communication and Information Engineering, Shanghai University

Abstract: A new watermarking scheme using principal component analysis (PCA) is described. The proposed method inserts highly robust watermarks into still images without degrading their visual quality. Experimental results are presented, showing that the PCA-based watermarks can resist malicious attacks including lowpass filtering, re-scaling, and compression coding.

Key words: watermarking, principal component analysis (PCA), Karhunen-Loeve transform (KLT)

1. Introduction

With the rapid development of computer network and multimedia technology, dissemination of information in the forms of audio, video and still image has become widespread. The problem of data piracy and copyright breach is a major concern when information is transmitted over networks, especially in the World Wide Web environment. As a means of intellectual property rights (IPR) protection, watermarking has received much research attention since mid-1990s[1,2,3], and is becoming a hot topic within the signal/image processing community. Watermarks are imperceptible[1] data embedded in the multimedia signals. The hidden data can be extracted and identified by the IPR owner, and used as an evidence of copyright.

This letter considers watermarking for still images. It is generally accepted that watermarks must possess the following properties to be useful in the IPR protection:

1. Invisibility. Watermarks should be invisible so that they do not affect the perceptive value of the image to be protected.
2. Identifiability. Watermarks can be extracted and recognized by means of special software, with or without the original image.
3. Robustness. Watermarks must not be destroyed, removed, altered, or forged by intentional attacks. They must also survive conventional signal processing operations such as filtering, cropping, scaling, compression coding, AD/DA conversion, etc.

Some early watermarking techniques are implemented in the space domain. For example, flipping the LSB of some pixel gray values does not significantly change the visual appearance of an image, and the embedded data can easily be extracted. But this simple technique is susceptible to most image processing operations, not to mention hostile attacks. Other space domain techniques with improved performance have been proposed[4,5]. In general, transform domain techniques[6,7,8] are considered more promising as in the transform domain the information in the original data spreads over the entire spectrum and different parts of the domain can be used to obtain different effects.

The main difficulty in watermarking lies in the conflicting requirements of invisibility and robustness. For a watermark to be invisible, the best region for data hiding is the part of the spectrum corresponding to the least perceptible information contents. But watermarks hidden in this region is likely to be destroyed by image manipulations such as lowpass filtering and compression coding. Therefore, a robust watermark must be embedded in the visually significant region so that it is not affected by any visual-effect preserving processing[9].

The most frequently used transforms in watermarking include DCT[6,7] and wavelet[6,8]. In this letter, we propose a new technique using Karhunen-Loeve transform, or principal component analysis (PCA). PCA is closely related to the statistical properties of the image, and has the advantage of high energy concentration and complete decorrelation. The new method is versatile in choosing a suitable region for data hiding. The method also allows a multi-layered key system, providing a high degree of data security.

2. Methodology

Consider vectors $\mathbf{f}_k$ sized $R\times1$, $k = 1, 2, \ldots, K$, as samples drawn from a stochastic process. The principal component transform, or Karhunen-Loeve transform, is defined as

[1] Although there are perceptible and imperceptible watermaks, the former is of less importance. Therefore we will concentrate on imperceptible watermarks.

$$\mathbf{g}_k = \mathbf{A}\,\mathbf{f}_k, \qquad k = 1,2,...,K \tag{1}$$

where **A** is an $R\times R$ transform matrix with columns being eigenvectors of the covariance matrix $\mathbf{C_F}$ of the image process **F**. The columns in **A** is arranged such that the corresponding eiganvalues are in a descending order. PCA decorrelates the elements in the vectors so that the covariance matrix $\mathbf{C_G}$ of the transform-domain process **G** is diagonal, with the diagonal entries equal to the variances of the vector elements in the transform domain.

An image **S**, sized $M\times N$, may be segmented into $I\times J = K$ small data groups, each of which arranged as a 2D array sized $P\times Q$ where $P=M/I$, and $Q=N/J$. These data groups can also be organized into $R\times 1$ vectors, where $R=P\times Q$. The way in which the image data are divided and reorganized can be varied considerably. In this study, a simple interleaving sub-sampling scheme is used:

$$f_{i,j}(p,q) = S\big[P(i-1)+p,\, Q(j-1)+q\big], \qquad \begin{matrix} i = 1,2,\ldots,I \\ j = 1,2,\ldots,J \\ p = 1,2,\ldots,P \\ q = 1,2,\ldots,Q \end{matrix} \tag{2}$$

or alternatively, using a single subscript and the vector representation:

$$\mathbf{f}_k = \mathbf{f}_{i,j}, \qquad k = (i-1)J + j = 1,2,\ldots,K \tag{3}$$

The image is now viewed as K samples drawn from an R-dimensioned stochastic process, **F**, on which PCA as described in (1) may be performed. The resulting transform-domain image **G** can be considered as K vectors sized $R\times 1$, arranged in the same way as Eq.(2) for F, namely,

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1K} \\ g_{21} & g_{22} & \cdots & g_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ g_{R1} & g_{R2} & \cdots & g_{RK} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_K \end{bmatrix} \tag{4}$$

Reorganize the rows in **G** into R blocks of size $I\times J$ ($=K$):

$$\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_r, \ldots, \mathbf{q}_R \tag{5}$$

with respective eigenvalues of the covariance matrix $\mathbf{C_F}$ equal to λ_r:

$$\lambda_1 > \lambda_2 > \cdots > \lambda_r > \cdots > \lambda_R \tag{6}$$

The component $\mathbf{q}_r$ in (5) can be written in a matrix form:

$$\mathbf{q}_r = \begin{bmatrix} g_{r,11} & g_{r,12} & \cdots & g_{r,1J} \\ g_{r,21} & g_{r,22} & \cdots & g_{r,2J} \\ \vdots & \vdots & \ddots & \vdots \\ g_{r,I1} & g_{r,I2} & \cdots & g_{r,IJ} \end{bmatrix} \tag{7}$$

The eigenvalue λ_r is the variance (energy) of block $\mathbf{q}_r$. The first block $\mathbf{q}_1$ contains most of the energy, while the energy contents decreases rapidly as the subscript rising due to the high degree of energy concentration of PCA.

It is clear that, in order to make a watermark robust against various attacks, the embedded data should be put into $\mathbf{q}_1$, resulting in a modified block $\mathbf{q}_1'$. In fact, by using the particular grouping and ordering scheme described in the above, $\mathbf{q}_1$ bears the major spatial features of the sub-sampled original image. Therefore, adding a watermark into $\mathbf{q}_1$ shares some merits of spatial approaches, although it is essentially a transform domain method.

After the secret code is embedded into the transform domain, an inverse transform is performed to obtain a watermarked image. The watermark can be extracted by PCA with the same matrix **A**.

A large number of different data arrangement schemes are possible. Every scheme will lead to a different transform matrix **A** and different watermarking performance. It should be noted that the data grouping scheme can be used as an encryption key. Another level of key is the way in which the invisible data are coded and inserted. It is also possible to use, in addition to $\mathbf{q}_1$, some of the less significant blocks in the PCA domain to increase the complexity of the multi-level key, presumably at the expense of slightly reduced robustness. A compromise is needed in deciding a specific method of data organization and embedding.

3. Computer Experiment and Performance Study

In the experiment, a 224×200 image shown in Fig.1 (a) was segmented into small pixel groups

following Eqs. (2) and (3) with P=4 and Q=4. Therefore a 16×16 covariance matrix $\mathbf{C_F}$ was obtained based upon 2800 samples of the image process, leading to a 16×16 principal component transform matrix **A**. Arranged in 4×4 squares, these 2800 transform-domain samples may be organized as a PCA domain "image", illustrated in Fig.1(b). In each 4×4 data group, the top-left element contains most of the energy. Thus, all the 2800 most significant elements form a rectangular data block sized 56×50. After linear re-scaling, this block appears to be a low-resolution version of the original image.

A pseudo-random binary sequence, consisting of $+s$ and $-s$, was inserted into the 56×50 data block, evenly distributed across the entire block, and an inverse pricipal component transform was then taken to obtain a watermarked image. The length of the sequence and the strength of the embedded watermark may be varied for different watermarking performance. Fig.1(c) presents the watermarked image with a 256-bit PN sequence embedded. The watermark strength was Q = 2.5%, defined as:

$$Q = \frac{s}{\max(\mathbf{q}_1) - \min(\mathbf{q}_1)} \tag{8}$$

It is observed that the inserted watermark is invisible although it causes an RMS error of 0.59% in the image.

The watermark was extracted with PCA, and cross-correlated with the embedded PN sequence for authentication. Ideally, the extracted and embedded sequences should have a correlation of 1 when in perfect alignment. The correlation would reduce under attack, as shown in Fig.2.

In the experiment, three types of attack with a wide range of strength were used to test the robustness of the proposed method: lowpass filtering, re-scaling, and JPEG compression. The lowpass filtering was performed using a 5×5 Gaussian-shaped convolution mask, characterized by the standard deviation σ. A large σ corresponds to a narrow-band filter. Image re-scaling was done by first sub-sampling with an integer factor K and then enlarging to the original size by repeating the pixel values. JPEG compression was characterized by a quality factor Q ranging from 0 to 100%. A small Q value represents a high compression ratio and low reconstruction quality. Effects of these attacks are illustrated in Fig.1(d), (e) and (f).

Fig.3(a), (b), and (c) show the calculated correlation between the embedded and extracted sequences as functions of σ, K, and Q, respectively. It is observed that the proposed watermarking scheme can resist all the three types of attack. Even if the attack was so strong that the image actually lost its value, the extracted watermark was still unambiguously recognizable.

Compared with some methods introduced in the literature, the PCA technique demonstrates better robustness against attacks. In [9], for example, a DCT-based approach was implemented, and the effects of various attacks were described. In their experiments, the detected watermark signal was reduced by over 50% when the watermarked image was scaled down to half of its original size, while ours only reduced by less than 25% as shown in Fig.3(b). When JPEG-encoded with parameters of 10% quality, the detected signal reduction in [10], which also used a DCT technique, was 78%, while ours was less than 45%. When the JPEG quality factor was 40%, the correlation was 100% in our case, but dropped to less than 40% in [10].

4. Conclusions

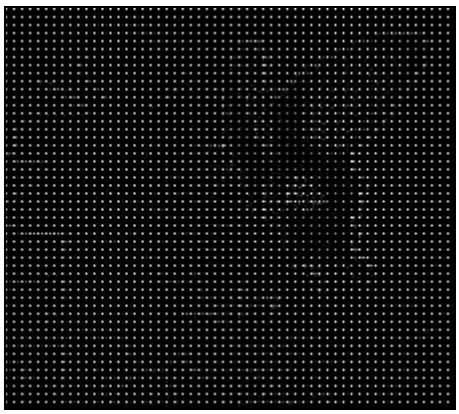
With a high degree of energy concentration, PCA is suitable for data hiding. By slightly altering the most significant data block in the transform domain, an invisible watermark can be inserted into the image. It has also been shown that the proposed method is very robust against severe attacks such as lowpass filtering, re-scaling, and compression coding, with the performance considerably better than some other methods. More aspects of the attack-resistant capability are currently under investigation.

In this letter, a specific blocking and data organization scheme is used in performing the PCA, which makes the most significant data block in the transform domain resembling a reduced-resolution version of the original image. This may allow the introduction of certain space domain techniques.

Different data grouping schemes may be devised to provide a wide variety of encryption keys. The number of ways in which a PN sequence is embedded into the target block is also enormous. This implies that a PCA watermark is very difficult to be extracted and counterfeited by any unauthorized attempts.

Current studies are concentrated in using m-sequence or other types of author-id data, combining PCA with other transform techniques, applying the HVS, etc., for the enhancement of reliability and robustness.

(a) Original image



(b) PCA

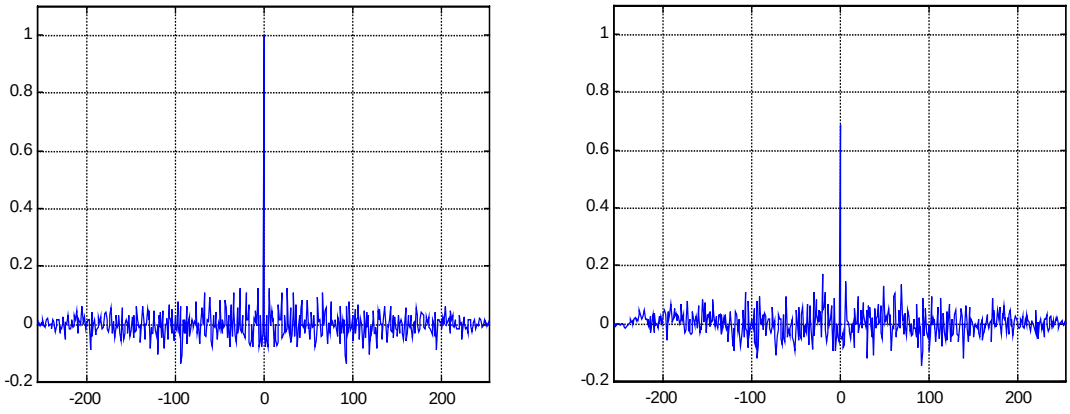
(c) Watermarked without attack

(d) Watermarked and blurred with a 5×5 Gaussian mask ($\sigma = 1.2$)

(e) Watermarked, and scaled down and up (K=3)

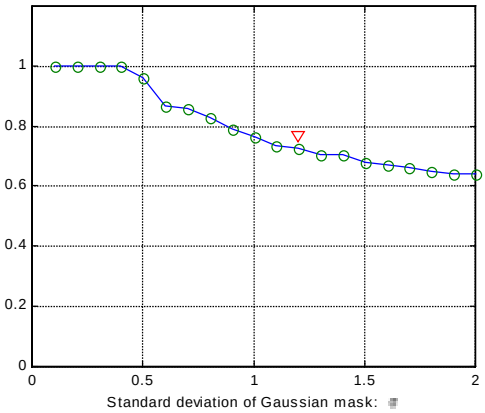
(f) Watermarked, JPEG-compressed and reconstructed (Q = 15%, compression ratio = 15.3)

Fig.1 The original image, its PCA, and the watermarked images, without attack and with severe attacks, respectively

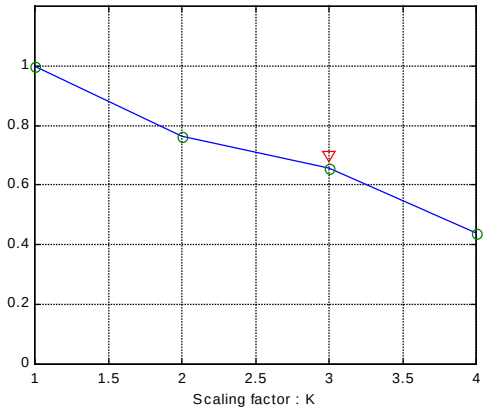


(a) Without attack (b) Lowpass filtered as in Fig.1(d)

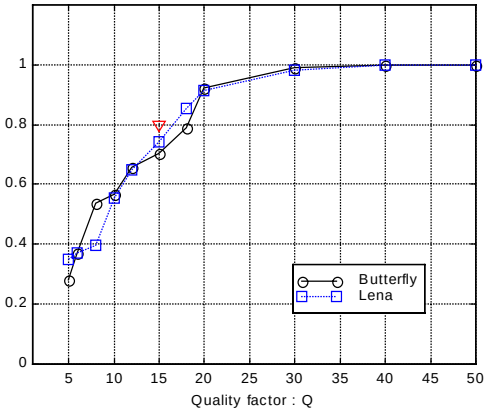
Fig.2 Correlation between the embedded PN sequence and the extracted sequence. (a) With a perfect watermarked image, correlation is 1 when the two sequences are in alignment. (b) When the watermarked image is lowpass-filtered with a 5×5 Gaussian mask ($\sigma = 1.2$), the correlation drops to 0.69.



(a) Lowpass filtering



(b) Re-scaling



(c) JPEG compression

Fig.3 Correlation between the embedded and extracted sequences, as functions of the strength of various attacks. The ∇ signs indicate the respective conditions under which Fig.1(d), (e), and (f) were obtained. In addition to Butterfly, the results of Lena in the JPEG-coding experiment is also shown for comparison with [10].

References:

[1] G. Caronni, “Assuring Ownership Rights for Digital Images,” in Proc. Reliable IT Systems, VIS’95, Vieweg Publishing Co., 1995

[2] B. M. Macq, et al., “Cryptology for Digital TV Broadcasting,” Proc. of IEEE, **83**(6), 1995: 944–957

[3] W. Bender, et al., “Techniques for Data Hiding,” *IBM System Journal*, **35**(3, 4), 1996: 313–336

[4] O. Bruyndoncky, et al., “Spatial Method for Copyright Labeling of Digital Images,” in *Nonlinear Signal Processing Workshop*, Thessaloniki, Greece, 1995: 456–459

[5] R. Wolfgang, et al., “A Watermark for Digital Images,” in Proc. Int. Conf. on Image Processing, vol.3, 1996: 219–222

[6] C. I. Podilchuk, et al., “Image-Adaptive Watermarking Using Visual Models,” *IEEE J. Selected Areas in Communications*, **16**(4), 1998: 525–539

[7] B. Tao, et al., “Adaptive Watermarking in the DCT Domain,” in *Proc. IEEE Int. Conf. ASSP*, 1997

[8] D. Kundur, et al., “A Robust Image Watermarking Scheme Using the Wavelet-based Fusion,” in *IEEE Signal Society Int. Conf. on Image Processing*, 1997

[9] I. J. Cox, ‘A Secure, Robust Watermark for Multimedia,” *Workshop on Information Hiding*, Newton Inst., Univ. of Cambridge, 1996

[10] M. D. Swanson, et al., “Transparent Robust Image Watermarking,” in *Proc. IEEE International Conference on Image Processing*, vol.3, 1996: 211–214