WATERMARKING INFORMATION LAYERS IN MULTISPECTRAL IMAGES OF CULTURAL HERITAGE OBJECTS

Emanuele Salerno

National Research Council of Italy
Institute of Information Science and Technology, Pisa, Italy

ABSTRACT

We present a strategy for transparent, robust watermarking to protect intellectual property rights on cultural heritage images. This can be applied on any kind of vector image, from color to hyperspectral, and has a particular value when each individual channel has its own significance and can be used independently of the others, as often happens with quantitative diagnostic images. For color images, we can reduce the correlation between channels by relying on alternative color spaces. Dealing with images with any number of channels, we rather propose to work in the principal component space, whatever the watermarking algorithm chosen. We motivate why this is advantageous and show an example experiment using an embedding procedure proposed in the literature.

Index Terms— Robust watermarking, Multispectral images, Principal components, Cultural heritage

1. INTRODUCTION

Transparent, robust watermarking is currently one of the techniques of choice for the protection of intellectual property rights in digital objects. *Transparent* means that some extra information is included in the object so that it is not visible to the human eye. A watermark is *robust* if it remains detectable in a distorted, but still usable, version of the watermarked object. When this is a digital image, most watermarking strategies are conceived for single-channel data, and are extended to color images, mostly affecting the less perceptually sensitive channel. In this way, the information sources in the images are not equally protected, as removing a single channel is tantamount to remove the watermark. This is not a problem when protecting merely illustrative images, as removing one channel degrades significantly their perceptual quality.

In cultural heritage images, however, each channel is often important *per se*, as it carries essential information for diagnostics or other documentation purposes. In those cases, each individual channel should be protected by a watermark, but different channels in multichannel images are always strongly correlated and, as such, not suitable for watermarking. Multichannel data can be better protected by decreasing their mutual cross-correlations. When working with trichromatic images, color spaces less correlated than the classical RGB can be exploited but, in spaces with dimensionality higher than 3, a simple extension of an alternative color space does not exist. We suggest that working on principal components could be a good choice, as they are uncorrelated by construction. A further advantage of this approach, for images with dozens of channels, is in computational efficiency. Indeed, only the strongest principal components could be watermarked, and all the channels of the resulting image would be protected.

In this paper, with no reference to any particular watermarking technique, we elaborate on this idea, by both theoretical considerations and a simple example, and show why it offers advantages over the channel-by-channel approach.

2. TRANSPARENT, ROBUST WATERMARKING

Let us assume we have a scalar image A, of size $h \times w$, to be distributed without compromising its property rights. Watermarking consists in embedding additional information in the image, thus enabling us to claim our rights. With this purpose, the watermark must be *robust*, that is, it must be difficult for anyone to extract or forge it, or distort the image so that the watermark is deleted or made undetectable, with no significant loss in image perceptual quality. A further option, in the interest of possible users, is to make the watermark *transparent*, that is, invisible to the naked eye. General principles and specific procedures can be found in the literature (*e.g.* [1, 2] and references therein). Irrespective of any particular strategy, let us introduce some notation useful in the rest of the paper. We denote the operation of embedding a watermark W into A with

$$A_W = A \oplus W \tag{1}$$

This work has been partially supported by European funds, through the program POR Calabria FESR 2007-2013-PIA Regione Calabria Pacchetti Integrati di Agevolazione Industria Artigianato Servizi, project ITACA (Innovative Tools for cultural heritage ArChiving and restorAtion). Accepted for presentation at International Workshop on Computational Intelligence for Multimedia understanding (IWCIM), Reggio Calabria, Italy, 27-28 October, 2016

The image thus watermarked could then be distorted by an *attacker* to try and delete the watermark:

$$A_W^* = \mathcal{D}(A_W) \tag{2}$$

where $\mathcal{D}$ can be any distortion operator, such as compression, subsampling, low-pass filtering, geometric distortion, *etc.* Once the *extraction* procedure

$$W^* = A_W^* \ominus A \tag{3}$$

is performed, the *owner* can claim their rights if they prove that the information in W^* is the same contained in W.

Let us now suppose that the object to watermark is a multichannel image, that is, an n-vector-valued matrix $\mathbf{A}$. It can range from a simple color image ($n = 3$) or a hyperspectral data cube, to any kind of multimodal object. An immediate choice to use procedure (1) with such an image is to embed one or more watermarks separately in all the channels:

$$\mathbf{A_W} = [A_1 \oplus W_1, A_2 \oplus W_2, ..., A_n \oplus W_n]^T \tag{4}$$

where superscript T means transposition, A_i are the image channels, and W_i are the components of a vector watermark $\mathbf{W}$. Having many channels often means needing too much computing. Moreover, especially when the channel watermarks are all equal, having many strongly correlated channels may compromise watermarking security. Also different watermarks derived from strongly correlated image channels produce very similar marks in all the image components. This is why it is suggested to lower the mutual channel correlations prior to embedding. Conversely, uncorrelated channels produce different patterns, thus making watermarking more robust to cancelation attacks. For color images, this result can be obtained by working in $L^*a^*b^*$, or in other color spaces less correlated than the classical RGB [3]. When $n > 3$, some extension to this approach must be devised.

3. WATERMARKING INFORMATION LAYERS

To decorrelate multispectral channels, we propose to rely on principal component analysis. This is a sufficiently fast approach, producing transformed channels representing the effective "information layers" in the image. Let us recall the essentials: let $\mathbf{a}$ be an $n \times hw$ matrix whose rows are the lexicographically ordered versions of the original channels A_i. Hereafter, we use indifferently the matrix or the lexicographic notations. We first compute the covariance matrix of the original image. If $\bar{\mathbf{a}}$ is the vector average of $\mathbf{a}$, this is:

$$C_{\mathbf{a}} = \frac{(\mathbf{a} - \bar{\mathbf{a}})(\mathbf{a} - \bar{\mathbf{a}})^T}{hw} \tag{5}$$

The principal components of $\mathbf{a}$ are given by

$$\mathbf{y} = V^T \cdot \mathbf{a} \tag{6}$$

where the columns of the $n \times n$ matrix V are the eigenvectors of C. It is easy to see that the covariance matrix $C_{\mathbf{y}}$ is diagonal, so all the principal components are uncorrelated. We can then watermark the principal components

$$\mathbf{Y_W} = [Y_1 \oplus W_1, Y_2 \oplus W_2, ..., Y_n \oplus W_n]^T \tag{7}$$

and come back to the original image space by

$$\mathbf{a_W} = V \cdot \mathbf{y_W} \tag{8}$$

Then, we first transform a possibly distorted version $\mathbf{a}_\mathbf{W}^*$ of $\mathbf{a_W}$ through

$$\mathbf{y}_\mathbf{W}^* = V^T \cdot \mathbf{a}_\mathbf{W}^* \tag{9}$$

and estimate the watermark as

$$\mathbf{W}^* = \mathbf{y}_\mathbf{W}^* \ominus \mathbf{y} \tag{10}$$

where $\ominus$ is the same operator in Eq. 3 applied to all the image components. The estimated vector watermark $\mathbf{W}^*$ must then be compared to $\mathbf{W}$ to see whether the "suspect" image $\mathbf{a}_\mathbf{W}^*$ is actually derived from our original $\mathbf{A}$.

This procedure ensures that not the strongly correlated original channels, but all the information layers contained in the multichannel image are protected by the watermark. Note that even though the embedding procedure (7) is limited to $q < n$ components, all the channels in $\mathbf{a_W}$ are protected through Eq. (8). When n is large, this can save computation time, especially if the $\oplus$ operation is complex. Moreover, as mentioned above, this watermarking is more difficult to break, as the property information can be made more diffuse in the image support. Another positive effect is in security: even knowing that the principal components have been marked, it is difficult for an attacker to extract or counterfeit the original patterns, as matrix V does not need to be made available.

To confirm experimentally these advantages, we should compare this approach to the channel-by-channel one, assuming the same embedding procedure and the same distortions in both cases. We do not make the complete case here, as exploring all the types of attack and watermarking strengths is out of the scope of this contribution, but present a single (and simple) example in Section 4 suggesting the way for a deeper evaluation.

4. AN ILLUSTRATIVE EXAMPLE

For convenience, our example is based on an RGB image (Fig. 1), but the procedure followed can be applied with any number of channels. The RGB channels are shown in Fig. 2.

The embedding procedure we use here is directly derived from [4]. We just give a few details about how the procedure works. The watermarks W_i are images with the same support as $\mathbf{A}$, all derived from a single *basis* watermark W_o, which is modified for every A_i, so as to be stronger in the less perceivable regions. Each resulting pattern is then suitably

Fig. 1. 1074×722 example image: Soft color restoration on the basis of Mona Lisa copy by Francesco Melzi (downloaded from https://upload.wikimedia.org /wikipedia/commons/7/7d/Mona_Lisa_color_restoration.jpg). Attribution: CC BY-SA 3.0 (http://creativecommons.org/ licenses/by-sa/3.0).

superimposed to the related image component. In our example, W_o is obtained by periodic repetition of the $signature$ pattern shown in Fig. 3.

As the RGB channels are strongly correlated, the derived watermarks will be all similar. This is immediately apparent from Fig. 4, where the darkest zones are clustered around the color edges, which are nearly the same in the three channels.

Conversely, the principal components (Fig. 5) are totally uncorrelated, so the watermarks (Fig. 6) are significantly different from each other, and the watermark patterns projected back onto the RGB space as in Eq. (8) are more diffuse in all the image domain, and then more robust. The color versions of both watermarked images (not shown here) are visually indistinguishable from the original in Fig. 1.

To test the efficiency of the two approaches, we first estimate the watermarks from the undistorted watermarked images. Rather than showing the entire watermarks, in Fig. 7, we show the signatures extracted from them by averaging over all the sub-images with the same size as the signature in Fig. 3. Apparently, the principal component watermarks are more legible than the others, because of the more uniform coverage of the image. To have a quantitative index, we compute the Pearson correlation coefficients ρ between the basis watermark W_o and the sums of the component watermarks estimated in the two cases. For the color case, $\rho = 0.384$; for the principal component case, $\rho = 0.480$. The p-values of the

Fig. 2. From left to right: the R, G and B channels of the example image.

Fig. 3. 41×127 $signature$ pattern used to build the basis watermark W_o.

associated tests with null hypothesis "no correlation" are both zero.

To demonstrate the better robustness of the principal component approach, we only show a single case with the distortion consisting in jpeg compression with 50% output quality. We do not display here the distorted image because, at this scale, it is indistinguishable from the original. The extracted signatures are shown in Fig. 8. It is difficult to say which of them is more similar to W_o or, better, which set of extracted signatures (color or PC) is more correlated with the original. The quantitative evaluation is now necessary: in the color case, $\rho = 0.020$; in the PC case, $\rho = 0.040$. To assess the significance of this result, we first note that the correlation coefficients between the basis watermark and a number of randomly generated images are always at least one order of magnitude smaller; secondly, the p-values of the Pearson tests

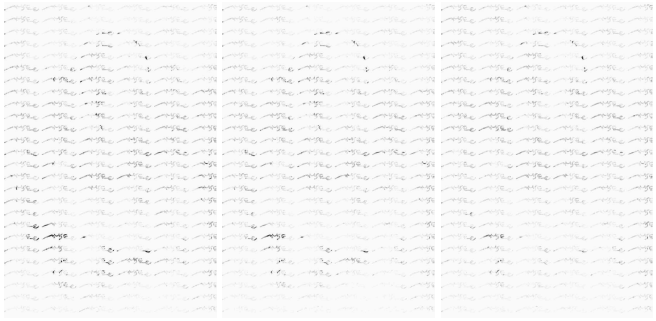

Fig. 4. Watermark patterns for the RGB channels of the example image (graylevels nonlinearly stretched and inverted for visibility).

Fig. 5. Principal components of the example image (graylevels translated and rescaled for display).

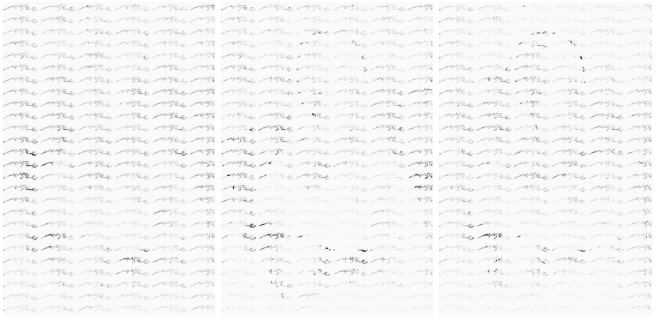

Fig. 6. Watermark patterns for the principal components in Fig. 5 (graylevels nonlinearly stretched and inverted for visibility).

in the two cases are of the order of 10^{-12} or less. Also, the correlation of W_o with the sum of the extracted watermarks is better than the correlation with any single component. This demonstrates the advantage of multichannel watermarking in terms of robustness, and the advantage of working with uncorrelated over strongly correlated channels.

5. CONCLUSION

We propose a strategy to watermark multichannel/multimodal digital images through their principal components rather than their channels. By our considerations and experimental demonstration, and independently of any particular embedding algorithm, this presents advantages in terms of security and robustness. Indeed, apart from possible cryptographic strategies adopted in embedding, a potential attacker would also need the eigenvector matrix to attempt extracting the watermarks. Normally, the eigenvector matrix is not made available, and can only be evaluated precisely from the original, not watermarked, image. Both robustness and security are also increased by the fact that projecting the watermarked principal components back onto the original channel space produces marks that are more uniformly distributed over the image support. This projection operation can also offer an immediate advantage in terms of computational efficiency: in the limit, even though a single principal component, that is, a single row in matrix $\mathbf{y_W}$, is watermarked, procedure (8) spreads it all over matrix $\mathbf{a_W}$. This option should be investigated carefully. Intuitively, one should mark the strongest components to maintain robustness, but some advantage could also be drawn by exploiting the eigenvalues of the correlation matrix. This aspect will be considered in the future, along with a deeper experimental analysis including more images, more types of attack, and different embedding and extraction strategies and parameters.

Today, finding an efficient, transparent and robust watermarking technique for high-quality and/or specialised images is particularly important in cultural heritage management, since many digital collections are being made available to the public at large, and a strategy to protect the rights on these objects without decreasing their value for general or expert users is strongly needed.

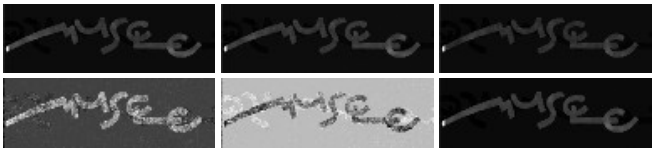

Fig. 7. Top: extracted signatures from the R, G and B watermarked channels of the example image. Bottom: extracted signatures from the principal components. Graylevels stretched for visibility.

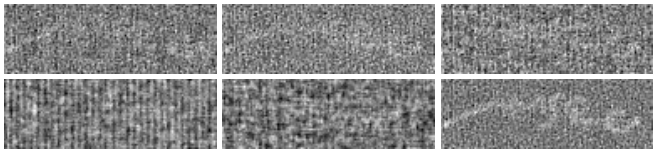

Fig. 8. Top: extracted signatures from the RGB channels of the distorted color-watermarked image. Bottom: extracted signatures from the principal components of the distorted pc-watermarked image. Graylevels stretched for visibility.

6. REFERENCES

[1] S. Stankovic, I. Orovic, and N. Zaric, "An application of multidimensional time-frequency analysis as a base for the unified watermarking approach," *IEEE Trans. Im. Proc.*, vol. 19, pp. 736–745, 2010.

[2] A. Kumar and V. Santhi, "A review on geometric invariant digital image watermarking techniques," *Int. J. Comp. Appl.*, vol. 12, pp. 31–36, 2011.

[3] S.A.M. Gilani, I. Kostopoulos, and A.N. Skodras, "Multipurpose color image watermarking," *Int. J. Comp. Appl.*, vol. 15, pp. 1–9, 2008.

[4] T.V. Nguyen and J.C. Patra, "A simple ica-based digital image watermarking scheme," *Digital Signal Processing*, vol. 18, pp. 762–776, 2008.