Towards Second Generation Watermarking Schemes

M. Kutter S. K. Bhattacharjee T. Ebrahimi
Signal Processing Lab
Swiss Federal Institute of Technology
CH-1015, Lausanne, Switzerland

Abstract

The digital watermarking schemes of today use pixels (samples in the case of audio), frequency or other transform coefficients to embed the information. The drawback of such schemes is that the watermark is not embedded in the perceptually significant portions of the data. We refer to such techniques as first generation watermarking schemes. *In this paper, we introduce the concept of* second generation watermarking schemes *which, unlike first generation watermarking schemes, employ the notion of data features. We propose a scheme based on point features in images using a scale interaction technique based on 2D continuous wavelets. The features are used to compute a Voronoi partition of the image. The watermark is embedded in each segment using spread spectrum watermarking. In the recovery process the same features are detected, and again used to partition the image. Then the watermark is extracted from each segment separately.*

1 Introduction

In general, digital watermarking methods can be considered as digital communication schemes. Most methods use some kind of modulation and demodulation. Usually the modulated watermark is weighted, to decrease the perceptivity of the watermark, and then added to the host data. *First generation watermarking schemes* do not explicitly make use of perceptually significant features in the data. For a review of such watermarking methods for different types of media the interested reader is referred to [4, 10].

In this paper we introduce the concept of *second generation watermarking schemes*. The idea is as follows. Digital watermarking, by definition, requires the watermark to be robust. In order to design robust watermarking schemes, the data should be embedded in salient parts of the data [3]. The concept of second generation watermarking extends this idea by employing significant data features in the watermarking process. Maes and Overveld [7] have proposed a watermarking scheme that goes along the lines of second generation watermarking. They propose to modify geometric features of the image. The method is based on a dense line pattern, generated pseudo randomly, that represents the watermark. A set of salient points in the image is then computed, for example, based on an edge detection filter. The detected points are then warped such that a significant large number of points are within the vicinity of lines. In the detection process, the method verifies that a significantly large number of points are in the vicinity of the lines.

We start this paper by first introducing the concept of second generation watermarking and presenting the necessary properties of features to be used in such schemes. Then we give an example of second generation image watermarking.

2 Second generation watermarking schemes

Second generation watermarking involves the notion of perceptually significant features in the data. Features may be abstract, or may be semantically meaningful characteristics of the data. Taking the example of images, features can be edges, corners, textured areas or parts in the image with specific characteristics. For audio data the relationship between harmonics may be taken as a useful features. Of course, not all features are suitable for watermarking. For example the area of a uniform region in an image is not a good feature as it will change if the image is scaled. Features suitable for watermarking should have the following properties:

1. Invariance to noise (lossy compression, additive, multiplicative noise, etc.)

2. Covariance to geometrical transformations (rotation, translation, sub-sampling, change of aspect ratio, etc.)

3. Localization (cropping the data should not alter remaining feature points)

Property (1) ensures that only significant features are chosen. Attacks are likely not to alter significant features because otherwise the commercial value of the data would be lost. Therefore, selecting salient features implies that these features are resistant to noise. Property (2) describes the behavior of the feature if the host data is geometrically distorted. Moderate amounts of geometrical modification should not destroy or alter the feature. Property (3) implies that the features should have a well-localized support. The use of such features makes the watermarking scheme resilient to data modifications such as cropping.

Extracted features can be used in two ways. In the first way the features serve as some kind of helper scheme for standard watermarking techniques. For example, the features may be used as reference locations and orientation for a standard watermark scheme. The goal of such a scheme is basically to increase the robustness towards geometrical alterations, although other goals may also be envisaged. Another way is to use the features directly in the embedding process. That is, the extracted features are directly modified, to embed the watermark information.

3 Second generation watermarking: an example

As an example, we now outline a scheme which follows the second generation watermarking paradigm introduced above. Here, we use image features to determine the reference location to embed a spread spectrum modulated watermark. This scheme consists of two steps. First, feature-points are located in the image. The image is partitioned into regions based on the feature-points detected, and each segment is then watermarked independently of the other regions. The two steps are are now explained in detail.

3.1 Point Features

In this work, we detect feature-points that are reasonably stable under common geometrical transformations, such as rotation, scaling, cropping, and so on. In the recent literature, several methods have been proposed for identifying feature-points in images. Most methods are designed to suit particular applications, such as image registration [8] or tracking [9]. Our feature-detection scheme is similar to the *scale interaction model* proposed by Chellappa and colleagues [8].

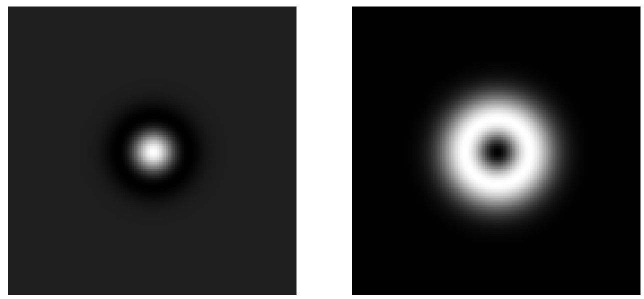
Our feature extraction scheme is based on a decomposition of the image using Mexican-Hat wavelets [1]. In two dimensions the Mexican-Hat mother wavelet (also called Marr wavelet), $\psi(\vec{x})$, at location $(\vec{x})$, is defined as:

$$\psi(\vec{x}) = (2 - |\vec{x}|^2)exp(-\frac{\vec{x}^2}{2}). \qquad (1)$$

where $\vec{x}$ gives the two-dimensional coordinate of a pixel. In the spatial-frequency domain, the wavelet is written as:

$$\hat{\psi}_H(\vec{k}) = (\vec{k} \cdot \vec{k})e^{-1/2(\vec{k}\cdot\vec{k})} \quad , \qquad (2)$$

where $\vec{k}$ represents the 2D spatial-frequency variable. The shape of the wavelet, in spatial domain and in frequency domain, is shown in Fig. 1.



(a) (b)

Figure 1: The 2D Mexican Hat wavelet: (a) in spatial domain; and (b) in frequency domain. Note that the filter is well localized in both domains.

We note that, in the frequency domain, the Mexican Hat is always centered at the origin. This implies that the response of a Mexican Hat wavelet is invariant to rotation. Therefore, it is well suited for detecting point-features in a rotation-invariant fashion. Different features in a given image have different intrinsic scales. The wavelet properties of the Mexican Hat can be exploited to detect features at different scales. For our purposes, feature-points at very fine scales are not very stable, and features at extremely low scales may be artifacts of lighting or other such phenomena. Therefore, we use a feature-detector that locates feature-points within a certain range of scales.

The approach for detecting feature-points is as follows.

1. Define the feature detector function, Q as:

$$Q_{ij}(\vec{x}) = |M_i(\vec{x}) - \gamma \cdot M_j(\vec{x})|, \qquad (3)$$

where $|\cdot|$ returns the absolute value of its argument, and $M_i(\vec{x})$ and $M_j(\vec{x})$ represent the

Figure 2: Image feature extraction: the extracted features are marked with white crosses on the input image.

responses of Mexican-Hat wavelets at the image-location ($\vec{x}$) for scales i, and j respectively. We only consider wavelets on a dyadic scale. Thus, the normalizing constant, γ, is given by $\gamma = 2^{-2(i-j)}$. The feature detector function, Q, is the absolute difference of the responses of the Mexican Hat wavelet applied to the image at two different scales.

2. Determine points of local maxima of Q. These maxima correspond to the set of potential feature-points.

3. Accept a point of local maximum in Q as a feature-point based on a threshold.

In previous work we have shown that such feature-points are quite stable, even in the face of moderately lossy compression [2]. Figure 2 shows an example for this feature extraction process. The locations of the extracted features are indicated with white crosses superimposed on the input image.

3.2 Watermark Embedding and Extraction

The watermarking process employed is similar to the one presented by Kutter [5, 6]. The method is based on spread spectrum modulation, where the binary watermark of length N is represented by N orthogonal, two-dimensional spread spectrum functions. Orthogonality among functions is achieved by forcing the functions to be non-overlapping, that is the intersection of the sets of non-zero locations is the empty set. The resulting watermark is weighted based on the image content and added to the blue component of the image. The weighting procedure is based on a weighting matrix which is proportional to the luminance. This approach exploits the contrast sensitivity model of the human visual system (HVS) which says that the detection threshold is approximately proportional to the local luminance. The watermark is embedded only in the blue image component because the HVS is least sensitive to modifications in the blue band.

The set of extracted feature points are used to segment the image using *Voronoi diagrams*, a partitioning of the space into segments such that all points in one segment are closer to a particular object (in our case the location of the extracted features) than to any other. Figure 3 illustrates the segmentation of the input image based on the feature points shown in Figure 2. The spread spectrum watermark is independently embedded in each segment, using the feature location as origin.

To extract an embedded watermark, we start by extracting feature point from the watermarked image. If the feature selection method is robust, the set of original feature points and the set of feature points extracted from the watermarked image should be equivalent. The extracted feature points are again used to segment the image. The watermark in each segment is then extracted using the feature location as reference origin for the modulation functions. In theory, this approach is inherently resistant to translation and cropping since the used features are covariant to these transformations. However, tests indicate that, due to the added watermark, the feature locations may move by $1-2$ pixels. This is not a problem – a limited search can be used to compensate for this misalignment. Robustness to rotation and scaling can be achieved by applying a log-polar mapping to the modulation function and the segments, and then correlating in the transformed domain to detect the watermark and to determine the scaling factor and angle of rotation. Preliminary results show that this approach is very promising. Moreover, the proposed scheme is resistant to high quality JPEG compression.

4 Conclusion

In this paper, we have introduced the new concept of second generation watermarking. We argue that these schemes, when properly designed, are more robust to attacks. This is due to the fact that perceptually important features are watermarked, or used as a helper scheme for standard watermarking schemes. Watermarks are supposed to be imperceptible and should resist attacks, as long as the attacks do not

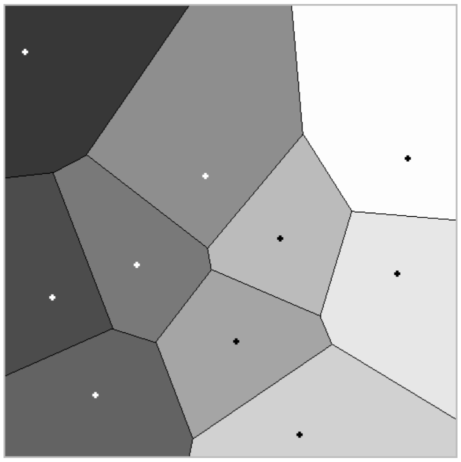


Figure 3: Image segments, computed from the feature locations, used for the watermarking process.

destroy the commercial value of the data. This implies that attacks should not visibly distort (or destroy) perceptually important features, such as edges, corners, texture segments.

As an example we have presented an image watermarking scheme based on the concept of second generation watermarking. The features are extracted using a scale interaction method. The location of the features is then used to perform a segmentation using *Voronoi diagrams*. The resulting segments are watermarked individually using a spread spectrum watermarking approach. The watermark in each segment is centered at the location of the corresponding feature. In the watermark extraction process the feature are again computed and used to segment the image. The watermark in each segment is extracted independently of the other segments. The proposed approach is inherently resistant to translation and cropping. Of course, the stability of the method depends on the features. The features used in the technique proposed in this paper have the drawback that their location may change by some pixels, due to the attack or even the watermarking process. Nevertheless, the proposed approach is quite efficient since the search space is quite small. We are exploring other ways of identifying more robust feature-points.

The concept of second generation is very promising because it links the watermark to the image and hence makes the watermarks less vulnerable to attacks. Of course, the strength of such a system lies in the feature selection method and the techniques used for embedding the watermark.

References

[1] J.-P. Antoine, P. Vandergheynst, and R. Murenzi. Two-dimensional directional wavelets in image processing. *Int. J. Imaging Sys. and Tech.*, 7:152–165, 1996.

[2] Sushil Bhattacharjee and Martin Kutter. Compression tolerant image authentication. In *Proceedings of the International Conference on Image Processing (ICIP)*, 1998.

[3] I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for images, audio and video. In *Proceedings of the International Conference on Image Processing (ICIP)*, September 1996.

[4] Frank Hartung and Martin Kutter. Multimedia watermarking techniques. *to appear in Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, 87, 1999.

[5] Martin Kutter. Watermarking resisting to translation, rotation, and scaling. In *Proceedings of SPIE International Symposium on Voice, Video, and Data Communications*, November 1998.

[6] Martin Kutter, F. Jordan, and Frank Bossen. Digital watermarking of color images using amplitude modulation. *Journal of Electronic Imaging*, 7(2):326–332, April 1998.

[7] M.J.J.J.B Maes and C.W.A.M Overveld. Digital watermarking by geometric warping. In *Proceedings of the International Conference on Image Processing (ICIP)*, volume 1, 1998.

[8] B. S. Manjunath, C. Shekhar, and R. Chellappa. A new approach to image feature detection with applications. *Pattern Recognition*, 29(4):627–640, 1996.

[9] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, June 1994.

[10] M.D. Swanson, M. Kobayashi, and A.H. Tewfik. Multimedia data embedding and watermarking techniques. *Proceedings of the IEEE*, 86(6):1064–1087, June 1998.