A REGION-BASED TECHNIQUE FOR CHAOTIC IMAGE WATERMARKING

Athanasios Nikolaidis *Ioannis Pitas*
Department of Informatics, Aristotle University of Thessaloniki,
Box 451, Thessaloniki 540 06, GREECE
Tel,Fax: +3031-996304
e-mail: [nikola, pitas]@zeus.csd.auth.gr

ABSTRACT

A novel method for embedding and detecting a chaotic watermark in the digital spatial image domain, based on segmenting the image and locating regions that are robust to several image manipulations, is presented in this paper. Each selected region is approximated by an ellipse. The watermark is embedded on its bounding rectangle. This representation proves robust under geometric attacks. The controlled lowpass nature of the chaotic watermark ensures its immunity to lowpass filtering and JPEG compression. Experimental results display the robustness of the method under several kinds of attacks, such as JPEG compression, mean and median filtering, scaling, cropping and rotation.

1 INTRODUCTION

Protection of multimedia information has attracted a lot of attention during the last few years. The aim of such methods is to protect the copyright of broadcast or publicly exposed multimedia data. Attackers have the freedom to obtain copies of copyrighted electronic material via the Internet and manipulate them at will. The most popular method to protect such kind of information is watermarking [1]. Most of the proposed watermarking techniques do not consider simultaneous robustness to several kinds of attacks. Many of them focus on robustness against JPEG compression only, others consider also noise addition as well as lowpass filtering, while others only attempt to face geometric distortions efficiently [2]-[6]. These techniques are either applied in the spatial digital image domain or in some image transform domain (e.g., DCT, DFT, DWT, etc.). None of them has covered the entire range of different processing attacks at the same time, without resorting to the original image.

In the case of image watermarking, employing spatial characteristics is essential for ensuring immunity to geometric transformations. When a watermark is embedded on the entire image, scaling, rotation or cropping will result in the destruction of the watermark because no reference points exist that would lead in finding the amount of scaling, rotation or cropping. The current paper proposes a novel technique that succeeds to embed a watermark that is robust to several kinds of manipulation, based on locating robust region-based spatial features on the image, so that they will be used as reference for compensating geometric attacks, while preserving robustness to other types of attack such as filtering and compression.

Section 2 presents the preprocessing step for the determination of the spatial constraints to be used in the embedding and detection stages. In section 3, the general class of chaotic watermarks is presented together with adaptations for digital images. Section 4 provides an explanation of the connection between the spatial features and the watermark that is embedded on the image. Section 5 presents the watermark detection procedure. Simulation results before and after manipulation of the watermarked image are presented in section 6. Finally, conclusions are drawn in section 7.

2 REGION SEGMENTATION AND FEATURE DETERMINATION

In order to embed the watermark on some selected image regions, first a segmentation or clustering technique should be found, that will provide us with a robust representation under image processing. The technique which was chosen as the most robust one is a multilevel implementation of the adaptive clustering method proposed in [7]. It is a variation of the ICM (iterated conditional modes) algorithm. This technique works well especially on images containing objects with smooth surfaces. The algorithm may not be optimal in the case of some textured images. However, the merging step that completes the algorithm provides a set of regions that may not correspond to real objects but are approximately as large as required.

First the classical K-means algorithm is performed on a subsampled version of the original image to get an initial coarse segmentation estimate. However, we wish to obtain a smoothed segmentation output containing a rather small number of large regions that would be suitable for spatial watermark embedding. According to the approach in [7], we present an adaptive method

that takes under consideration both similarity potentials between current and neighboring pixel cluster assignments, as well as greylevel relation between current pixel and possible centers.

By applying Bayes theorem, we can obtain a model for the *a posteriori* probability density function that describes the desired segmentation:

$$p(x_s|y_s, x_q, q \in N_8(s)) \quad \propto$$
$$exp\left\{-\frac{1}{2\sigma^2}(y_s - \mu_{x_s})^2 - \sum_{x_s \in C} V_C(x)\right\} \qquad (1)$$

where x_s is the cluster assignment of pixel s, y_s is the luminance of pixel s, μ_{x_s} and σ^2 are the mean value and variance of cluster x_s, C is the clique of s, $V_C(x)$ is the potential function of this clique and $N_8(s)$ is the 8-neighbourhood of s, over which the potentials are summed. By maximizing this probability with respect to the cluster center, each pixel is assigned to a certain cluster. Finally, a region merging process according to the mean value similarity between adjacent regions is employed in order to eliminate useless small regions.

The resulting regions are then ordered according to their size, excluding the ones along the image boundaries to avoid problems resulting from image cropping. The largest regions are preferred for watermarking, so that a largest data set will be present in the detection stage and a bigger percentage of watermark power will be preserved.

For each of the chosen regions we employ an $\alpha-$trimmed Mean Radial Basis Function network to get an ellipsoidal region approximation [8]. This technique provides the marginal median estimation for the center and the covariance matrix describing each object.

The orientation of the trimmed ellipsoidal approximation can easily be computed using central moments [9]. The bounding rectangle of the ellipsoidal approximation can also be found. It defines the area where the watermark is to be embedded. The center coordinates of the bounding rectangle of each selected region, its dimensions and its orientation are the output of the segmentation stage. This information is used in both watermark embedding and detection stages.

3 WATERMARK CONSTRUCTION

After locating robust regions in the input image, so that they can be used as reference areas to embed our watermark, a watermark is constructed based on a chaotic trajectory [10], because of its controlled lowpass properties. This cannot be accomplished using a usual pseudo-random sequence, because this type of sequence produces noisy-like binary watermarks that would very easily be distorted by lowpass filtering or JPEG compression. The employed chaotic trajectory is of the form:

$$z(n+1) = \mathbf{F}(z(n), \lambda), \qquad z(n) \in U, \lambda \in I\!R \qquad (2)$$

where $\mathbf{F}$ is the Renyi map [11] with $\mathbf{F} : U \to U, U \subset I\!R$, $n = 0, 1, 2, ...$ denotes the current iteration and λ is a parameter that controls the chaotic behaviour of the system. The trajectory is recursively constructed and can be theoretically of an infinite period. The values of the produced trajectory oscillate inside an interval $[z_{min}, z_{max}]$ that is related to the parameter λ [11]. Thus, we can define a threshold level $z_{th} \in [z_{min}, z_{max}]$ in a way that, after thresholding the sequence numbers, a bipolar sequence $s(n) \in \{-1, 1\}$ is produced with approximately equal number of -1s and 1s. Parameter λ controls the frequency characteristics of the chaotic sequence, i.e. the frequency of the transitions $-1 \to 1$ and $1 \to -1$. For $\lambda > 1$ and values close to 1, we get a chaotic watermark with low number of transitions and, thus, lowpass properties. To embed the one-dimensional sequence in a two-dimensional signal, such as a digital image, we need to scan across the sequence in such a way that the lowpass properties are preserved. In order to do this, we employ the Peano scan order which has the property that every point along the scan is topologically closer to the previous and subsequent pixels than in the case of raster scan. In addition, it is possible to use cellular smoothing to eliminate spontaneous transitions that emerge after the Peano scan [10].

In order to construct different watermarks we use a key K that produces the seed value $z(0)$ for the generation of a chaotic trajectory. Keys of slightly different values provide sufficiently uncorrelated trajectories, reducing the possibility of the watermark being tampered and ensuring non-invertibility of the watermark. Thus, the corresponding key cannot be extracted from the 2D watermark.

4 WATERMARK EMBEDDING

In this stage we use the extracted salient feature set to embed the produced watermark in a specific image region that will be easy to detect even after intentional or unintentional attacks.

A prototype watermark serves as a reference pattern which can be adapted according to the dimensions, center and orientation of the bounding rectangle of each selected region before embedding. When the new region parameters are computed in the detection stage, each potential prototype watermark that is tested for presence in the watermarked and possibly manipulated image, is again adapted to these parameters before applying the detector.

The watermarked image $f_w(x, y)$ is defined as:

$$f_w(x, y) = f(x, y) \qquad (x, y) \notin A_{emb} \quad (3)$$
$$f_w(x, y) = f(x, y) + h \cdot w_n(x, y) \qquad (x, y) \in A_{emb} \quad (4)$$

where A_{emb} is the embedding image area, w_n is the watermarking sequence and h is the strength of the watermark. In our case, the watermark is embedded in the

spatial domain and, thus, the watermark strength has an integer value.

5 WATERMARK DETECTION

When a prototype watermark is to be detected inside a watermarked and possibly manipulated image, the image has to be first segmented, so that the salient feature set and orientation of the approximated regions are derived. These features include the center coordinates, dimensions and orientation of the bounding rectangle of each approximated region. A prototype watermark of standard dimensions is constructed. Afterwards, this watermark is adapted to each embedding region by scaling, centering, and rotating it according to the bounding rectangle features. For each detection region A_{det_i}, $i = 1, ..., M$, where M is the number of selected regions, the response of a hypothesis testing detector is computed:

$$R(\hat{f}_w, \hat{w}_i) = \bar{\mathbf{a}}_i - \bar{\mathbf{b}}_i \tag{5}$$

where:

$$\bar{\mathbf{a}}_i = \frac{1}{N_{\mathbf{A}_i}} \sum_{(x,y)\in\mathbf{A}_i} \hat{f}_w(x,y) \qquad \bar{\mathbf{b}}_i = \frac{1}{N_{\mathbf{B}_i}} \sum_{(x,y)\in\mathbf{B}_i} \hat{f}_w(x,y) \tag{6}$$

with $\mathbf{A}_i = \{(x,y) \in A_{det_i} | \hat{w}_i(x,y) = 1\}$ and $\mathbf{B}_i = \{(x,y) \in A_{det_i} | \hat{w}_i(x,y) = -1\}$. $N_{\mathbf{A}_i}$ and $N_{\mathbf{B}_i}$ are the number of pixels of the sets $\mathbf{A}_i$ and $\mathbf{B}_i$ respectively. Thus, the detector expresses the difference $\bar{\mathbf{r}}_i$ of two sample means. The mean value and variance of the detector output are:

$$\eta_{\bar{r}_i} = \eta_{\bar{a}_i} - \eta_{\bar{b}_i} \qquad \sigma^2_{\bar{r}_i} = \left(\frac{1}{N_{\mathbf{A}_i}} + \frac{1}{N_{\mathbf{B}_i}}\right) \sigma^2_{\hat{f}_w} \tag{7}$$

In the case that the watermark is embedded on the entire embedding region, the detector output is assumed to follow a normal distribution. If the correct watermark is embedded on the image, then the mean value is $\eta_{\bar{r}_i} = 2h$ and the variance is $\sigma^2_{\bar{r}_i} = \left(\frac{1}{N_{\mathbf{A}_i}} + \frac{1}{N_{\mathbf{B}_i}}\right)(\sigma^2_f + \sigma^2_{w_i})$, where σ^2_f is the variance of the initial image and $\sigma^2_{w_i}$ is the variance of the watermark, as is adapted for the certain region. Otherwise, if there is no watermark present, the mean value of the detector is $\eta_{\bar{r}_i} = 0$ and the variance is $\sigma^2_{\bar{r}_i} = \left(\frac{1}{N_{\mathbf{A}_i}} + \frac{1}{N_{\mathbf{B}_i}}\right)\sigma^2_f$, which is not significantly different than in the case the watermark is present, because the factor $\frac{1}{N_{\mathbf{A}}} + \frac{1}{N_{\mathbf{B}}}$ is very small and $\sigma^2_{w_i} \ll \sigma^2_f$. The detection is done over all regions where the watermark was embedded, and the overall detector output is defined as the maximal detector output for all watermarked image regions. This is expressed by:

$$R = \max_{1\le i\le M} R(\hat{f}_w, \hat{w}_i) \tag{8}$$

The detector output (8) must be compared against a proper threshold R_{thr} that will inform us with a satisfying certainty about the presence or the absence of the watermark. The distribution of the resulting output is not anymore normal, both in the case that no watermark is detected and in the case the correct watermark is detected. The expected mean values are now greater than 0 and $2h$, respectively. However, when searching for an efficient detection threshold, we will consider the approximating distributions as normal, for simplicity reasons.

6 EXPERIMENTAL RESULTS

The robustness of our technique was tested against several processing attacks on several images like the one of size 800×800 shown in Figure 1a. Figure 1b shows the final segmentation result for $K = 4$, after the small regions elimination result. The several regions (which are 7 in this case) are represented by different greylevels. In Figure 1c the two largest regions of the above referenced image are shown, after excluding the regions lying at the image borders, and finally Figure 1d shows the result of the ellipse approximation stage. Figures 1e, 1f, 1g and 1h show respectively a watermarked image, the two largest regions of its segmentation, their ellipse approximations and two experimental distributions for the normalized detector output. These are obtained after detecting 100 different watermarks on the original image and on the correctly watermarked image. The vertical axis shows the number of watermarks that give a certain detector output, and the horizontal axis shows the detector output values. The distributions are approximated by normal ones. The corresponding results for a watermarked image that is afterwards rotated by 12 degrees are shown in Figures 1i, 1j, 1k and 1l. A threshold can still be found for separating the distributions.

We can see that the region features remain almost intact after watermarking, and even after significant rotation. This is also the case after JPEG compression, lowpass filtering, scaling and cropping, or even after a combination of the above attacks, though a local search for the exact center, rotation angle and aspect ratio of the watermarked region may be necessary. The immunity of the watermark under filtering and compression is explained by its lowpass nature.

7 CONCLUSIONS

In the present paper we developed a method for embedding and detecting chaotic watermarks in large images. An adaptive clustering technique is employed in order to approximate selected regions by ellipsoids, whose bounding rectangles are chosen as the embedding areas for the watermark. The chaotic prototype watermark used for embedding is modified in such a way as to retain certain lowpass properties. The watermark is geometrically adapted before embedding, using the orientation, center coordinates and dimensions of each bounding rectangle. A hypothesis testing detec-

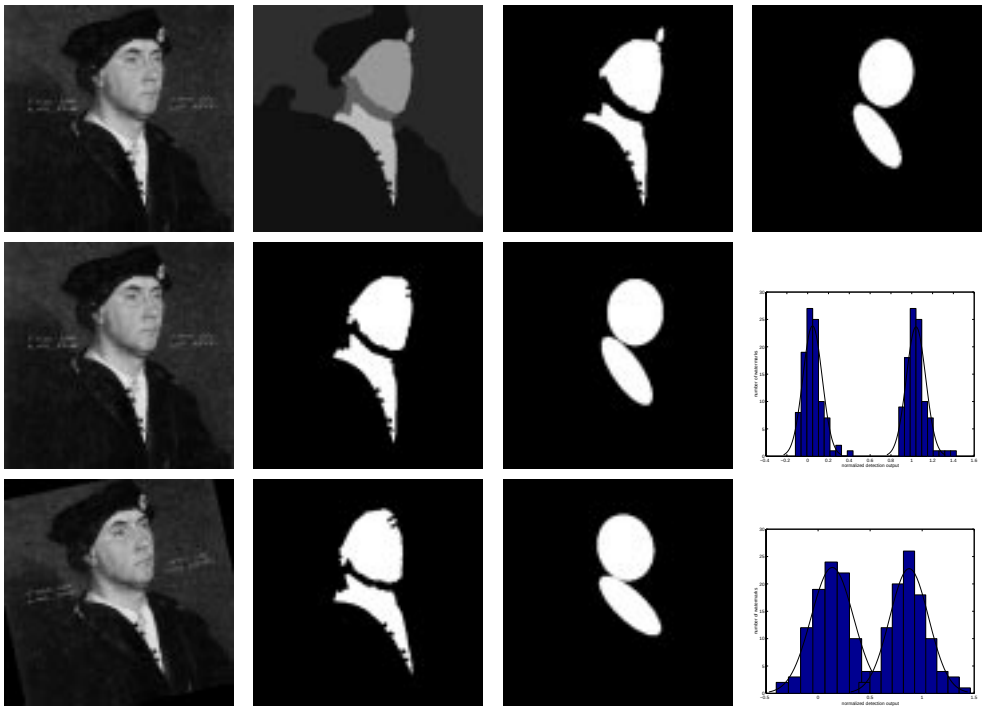

Figure 1: (a) Original image. (b) Segmentation result. (c) Two largest regions of the segmented image. (d) Ellipsoid approximation of the regions in (c). (e) Watermarked image. (f) Largest regions of the segmented image. (g) Ellipsoid approximation of the regions in (f). (h) Experimental distributions of the normalized detector output. (i) Rotated watermarked image. (j) Largest regions of the segmented image. (k) Ellipsoid approximation of the regions in (j). (l) Experimental distributions of the normalized detector output.

tor is employed in order to decide about the presence of a potential watermark. Experimental results display the robustness of the method for a variety of attacks on different images.

References

[1] G. Voyatzis and I. Pitas, "Protecting Digital-Image Copyrights: A Framework," IEEE Computer Graphics and Applications. vol. 19, pp. 18-24, January/February 1999.

[2] N. Nikolaidis and I. Pitas, "Copyright Protection of Images using Robust Digital Signatures," Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP '96, Atlanta, Georgia, USA, May 1996, pp. 2168-2171.

[3] I.J. Cox, J. Killian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Trans. on Image Processing, vol. 6, pp. 1673-1687, December 1997.

[4] A. Piva, M. Barni, F. Bartolini and V. Capellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," Proc. IEEE Int. Conf. on Image Processing (ICIP'97), Santa Barbara, California, USA, October 1997, pp. 520-523.

[5] J. O'Ruanaidh, T. Pun, "Rotation, Scale and Translation Invariant Digital Image Watermarking," Proceedings of ICIP '97, Santa Barbara, California, USA, October 1997, pp. 536-539.

[6] X.-G. Xia, C. G. Boncelet and G. R. Arce, "A Multiresolution Watermark for Digital Images," Proceedings of ICIP '97, Santa Barbara, California, USA, October 1997, pp. 548-551.

[7] T.N. Pappas, "An Adaptive Clustering Algorithm for Image Segmentation," IEEE Trans. on Signal Processing, vol. 40, pp. 901-914, April 1992.

[8] A.G. Bors and I. Pitas, "Object segmentation in 3-D images based on alpha-trimmed mean radial basis function network," Proc. of EUSIPCO '98, Rhodes, Greece, September 1998, pp. 1093-1096.

[9] A.K. Jain, Fundamentals of Digital Image Processing. New Jersey: Prentice-Hall, 1989.

[10] G. Voyatzis and I. Pitas, "Chaotic Watermarks for Embedding in the Spatial Digital Image Domain," Proc. of ICIP '98, Chicago, Illinois, USA, October 1998, pp. 432-436.

[11] R.L. Devaney, An introduction to dynamical systems. Penjamine/Cummings, 1986.