Robust Watermarking of Facial Images Based on Salient Geometric Pattern Matching

Athanasios Nikolaidis and Ioannis Pitas, *Senior Member, IEEE*

***Abstract*—We introduce a novel method for embedding and detecting a chaotic watermark in the digital spatial domain of color facial images, based on localizing salient facial features. These features define a certain area on which the watermark is embedded and detected. An assessment of the watermarking robustness is done experimentally, by testing resistance to several attacks, such as compression, filtering, noise addition, scaling, cropping and rotation.**

***Index Terms*—Chaos, copyright protection, correlators, feature extraction, image segmentation.**

I. Introduction

A FIELD of rapidly increasing interest during the last few years has been multimedia data protection. It emerges from the fact that many innovative techniques for digital data transfer, storage, and processing have been recently developed and used. Malicious users (attackers/pirates) of the stored/transferred data intend to present copyrighted material, such as digital images, audio, and video, as their own property. Watermarking provides efficient tools for ensuring that product ownership of these multimedia is preserved even after multimedia data processing by such attackers [1].

Many different watermarking methods have been proposed for still images, that embed a pseudo-random sequence in either the spatial or the transform image domain [2]–[7]. Most of these methods treat the image in a global sense, without taking into account any local properties that could be useful when the watermark is to be detected. The results presented in the respective papers usually display robustness against only some of the usual attacks, notably either filtering/compression or geometric distortions. Thus, these methods are customized in such a way that no care is taken about their resistance against most commonly considered attacks. There are also a few methods that attempt to ensure robustness against a wider range of attacks. One of the techniques that tries to cope with geometric transforms, while retaining robustness to other kinds of attack, is presented in [8]. The main concept is to find points in the image that could be warped according to their distance from specific line segments that form the watermark. However, it is not quite clear whether the construction of a theoretically infinite set of different line patterns in the detection stage is computationally feasible. Also the prominence of the selected points to be warped is a topic under question. In general, when a watermark is embedded on the entire image, scaling, rotation, or cropping may result in the destruction of the watermark, because no salient reference points exist that would lead in detecting the watermark after some degree of scaling, rotation, or cropping. The use of an image transform, with the exception of a Fourier transform, will suffer the same problems. The use of the Fourier transform provides rotation, translation, and scale invariance [6]. However, the robustness to filtering or compression depends on the range of frequencies that are used for watermark embedding. A significant step in robust watermarking would be to extract salient features based on their robustness against most of the frequent geometric attacks that do not degrade the visual image quality.

This paper presents a technique for watermarking a certain image class of special interest, namely color frontal face images with uniform background, based on extracting a finite set of salient facial features and on applying proper geometric pattern matching in the facial region. This image class has two important characteristics: 1) it contains several salient features, notably the eyes and mouth that can be used for matching and 2) it has particular applications related to copyright protection of digital portrait galleries, as well as digital frontal "head and shoulder" images or video that can be found in several applications, notably in broadcasting and in security. In this class of images, the facial region is the most important one. We shall therefore concentrate our study in the use of watermarks for copyright protection of facial images. The proposed technique provides sufficient immunity to the most commonly referenced attacks. Section II gives the outline of the proposed watermarking system. Section III presents the color face segmentation and region approximation technique, as well as the pattern matching process for localizing the salient features. In Section IV, the general class of chaotic watermarks is presented, together with adaptations for digital images. Section V explains the connection between the extracted features and the watermark to be embedded on the image. Section VI presents the watermark detection procedure. Simulation results for several kinds of manipulations on the watermarked image are presented in Section VII and, finally, conclusions are drawn and future work is addressed in Section VIII.

II. Technique Outline

This paper aims at providing a watermarking technique for the copyright protection problem of color frontal facial images with uniform background based on the selection of certain robust facial features for watermark embedding. Region-based image watermarking is introduced for the following reasons.

Manuscript received June 16, 1999; revised July 6, 2000. This work was carried out within the framework of the EU LTR project INSPECT. The associate editor coordinating the review of this paper and approving it for publication was Dr. Chung-Sheng Li.

The authors are with the Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece (e-mail: nikola@zeus.csd.auth.gr; pitas@zeus.csd.auth.gr).

Publisher Item Identifier S 1520-9210(00)07838-X.

1) This watermarking technique is related to the object-based coding/description approach followed in MPEG4 and MPEG7 (although the proposed method is described for still images only).
2) In some cases, only certain image regions have to be protected (e.g., facial regions in portraits).
3) Feature detection can be proven robust to certain geometrical transforms and other image processing operations.

The outline of the developed technique is as follows.

- *Feature selection*: This stage is concerned with the preprocessing that is necessary to extract the spatial image characteristics needed for the watermark embedding/detection stage.
  - *Image segmentation*: In this step, a skin-tone color segmentation technique is used that operates on the HSV color space, by selecting certain value ranges for the chrominance and luminance components.
  - *Feature detection*: The resulting facial region is approximated by an ellipse, by means of a properly chosen neural network. Afterwards, the eyes and the center of the mouth are being searched for inside this ellipse, by trying to match them with appropriate simple geometrical templates. These three reference points define a rectangular area of certain dimensions, center and orientation. These parameters are finally used as input parameters for watermark embedding and detection.
- *Watermark embedding/detection*:
  - *Chaotic watermark embedding*: A chaotic watermark that is constructed by Peano scanning of an one-dimensional (1-D) chaotic trajectory is embedded [9] according to the geometric parameters produced in the previous stage. The watermark is embedded on a rectangle corresponding to the facial image region of the previous stage.
  - *Chaotic watermark detection*: A watermark detector based on the correlation of a watermark template with the possibly watermarked and processed image is proposed. This detector acts on a rectangle defined on the watermarked image in the same way as in the original one. Consequently, the robustness of the localization of the spatial features after several attacks on the watermarked image ensures the robustness of the detection process. In this way, only small local searches in the geometric parameter space are required to find the correct position of the embedded watermark.

III. Face Segmentation and Salient Feature Localization

The first stage in the watermark embedding on a selected image region is to segment the facial region so that the search for salient features is limited in this area. Thus, this step is very significant, since the facial features that will be localized (eyes, mouth) via geometric templates are unique for the facial region.

In the following, we propose a technique for eye and mouth localization that is rotation, translation, and scale invariant. Any other technique proposed for this purpose in the literature can be used as well [10]–[16]. The method followed is based on exploiting color information in a similar way as in [10]. Our aim is to discriminate the skin-colored image region, which corresponds to the facial region. The original RGB image is converted to the HSV domain, because it is easier to perform skin-tone color segmentation in this color space. More specifically, the parameter ranges for hue (H), saturation (S), and value (V) that fulfil our requirements have experimentally been found to be [11]:

$$\begin{gathered} 0 \le H \le 25 \text{ or } 335 \le H \le 360 \\ 0.2 \le S \le 0.6 \\ 0.4 \le V \end{gathered} \tag{1}$$

according to tests on the M2VTS database of color frontal facial images [17]. More sophisticated methods to discriminate skin-tone color have later been implemented [18]. In our implementation of the segmentation method, the initial image is subsampled by a factor of two in both dimensions before thresholding. This helps eliminating isolated pixels that do not belong to the facial region but their color is considered as skin-color like because they satisfy (2) and, vice versa, pixels that were not considered as skin-colored ones, though they belong to the facial region. The skin-colored regions, and only them, are entirely labeled using (2). The choice of the aforementioned ranges in the HSV domain ensures that the segmented region of interest will approximately be the same even after some manipulation, as only an insignificant number of pixels will exceed these thresholds.

A connected component algorithm follows after binary thresholding, in order to isolate all the compact skin-colored regions. In order to get a good approximation of the facial region that does not contain useless areas, e.g., the neck, as well as to prevent the facial region areas from getting connected to the background, we employ an α-trimmed mean radial basis function network to get an elliptical approximation of the facial region [19]. This network is based on the approach that a percentage of the data samples pertaining to an object need to be taken into account in order to approximate the object. An α percentage of each end of the distribution of these samples can be trimmed away before approximating the object by an ellipsoid. Each hidden unit of the network corresponds to an object. If we consider the facial region as an object, this network can provide an estimate of the center of the RBF (radial basis function) by considering the marginal data samples [19]:

$$\hat{\mu}_k = \frac{\sum_{i=\alpha_k N_k}^{N_k - \alpha_k N_k} X_{(i)}}{N_k - 2\alpha_k N_k} \tag{2}$$

where

- $X_{(i)}$ marginal data samples sorted according to their values;
- N_k total number of data samples assigned to the kth hidden unit (equivalently, an object);
- α_k percentage of data samples to be trimmed away.

The estimate of the covariance matrix of the RBF is given by

$$\hat{\boldsymbol{\Sigma}}_k = \frac{\sum_{i=0}^{N_k - \alpha_{k,\mathcal{M}} N_k} (\boldsymbol{X}_{(i),\mathcal{M}} - \hat{\mu}_k)(\boldsymbol{X}_{(i),\mathcal{M}} - \hat{\mu}_k)^T}{N_k - \alpha_{k,\mathcal{M}} N_k} \tag{3}$$

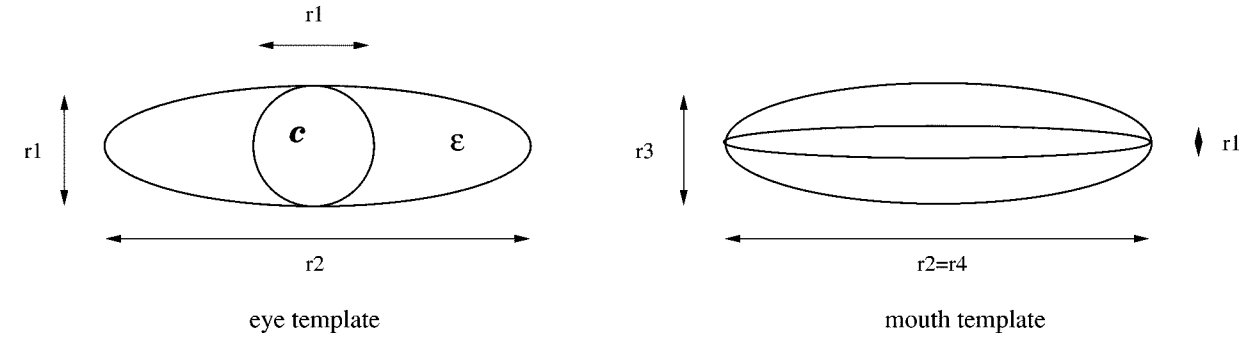


Fig. 1. Sample templates for the eye and the mouth.

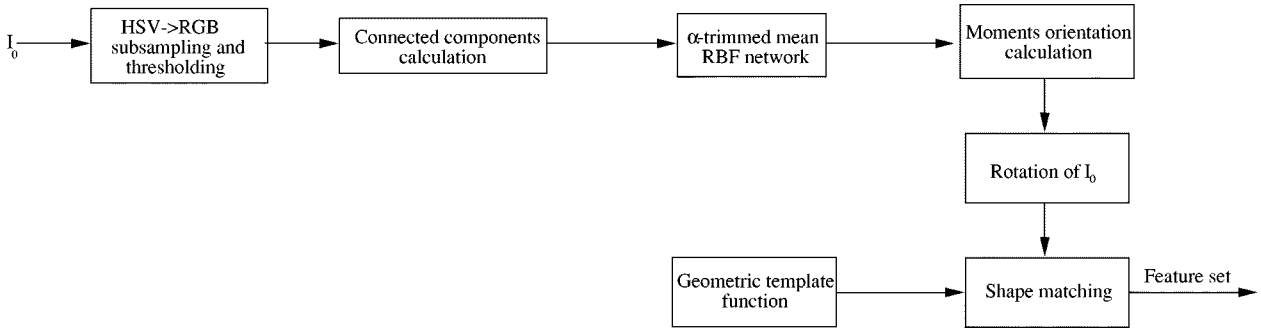


Fig. 2. Face segmentation and salient feature localization.

where $\boldsymbol{X}_{(i),\mathcal{M}}$ is the ith ordered data sample according to the Mahalanobis distance and $\alpha_{k,\mathcal{M}}$ is the trimming percentage. The Mahalanobis distance is defined as

$$r^2 = (\boldsymbol{X} - \hat{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{X} - \hat{\mu}_k) \tag{4}$$

where

- $\boldsymbol{X}$ marginal data sample;
- $\hat{\mu}_k$ center of the kth hidden unit (or, equivalently, the kth object);
- $\boldsymbol{\Sigma}_k$ its covariance matrix.

At this point all the original samples are used to compute the covariance matrix $\boldsymbol{\Sigma}_k$. The interested reader is referred to [19] for a detailed description of trimmed RBF's.

Once the trimmed elliptical approximation is known, its orientation can be computed by [20]

$$\theta = \frac{1}{2} \arctan\left[\frac{2m_{1,1}}{m_{2,0} - m_{0,2}}\right] \tag{5}$$

where $m_{p,q}$ are the central moments:

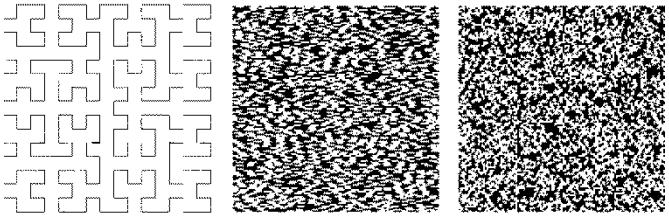
$$m_{p,q} = \frac{1}{N-1} \sum_{(x,y)\in\mathcal{R}} \sum (x - \overline{x})^p (y - \overline{y})^q \tag{6}$$

and $(\overline{x}, \overline{y})$ is the mass center of the elliptical region $\mathcal{R}$. The input image should be rotated according to the angle given by (5) before matching.

The most prominent features contained in the elliptical area are the eyes and the mouth. They can be approximated sufficiently well by proper geometric template functions. These features are unique in facial images and can act as robust reference points under any common geometric distortion. Other similar approaches use 2-D sinc functions for eye modeling [11]. The eye can be regarded as a circle (iris) having low, almost constant, intensity centered inside an ellipse of very high intensity, as can be seen in Fig. 1. According to this approach, we define the ideal eye detector by finding the image positions where the two sides of (7) have minimal difference:

$$\iint_{\mathcal{C}} w_c(x, y) I(x, y)\, dx\, dy = \iint_{\mathcal{E}} w_e(x, y) I(x, y)\, dx\, dy - \iint_{\mathcal{C}} w_e(x, y) I(x, y)\, dx\, dy \tag{7}$$

where $I(x, y)$ is the image intensity, $\mathcal{C}$, $\mathcal{E}$ are the sets of points lying on the circular (iris) and the elliptical disk, respectively, and $w_c(x, y)$ and $w_e(x, y)$ are the weighting functions that compensate for the luminance differences between the two areas. Equation (7) expresses the fact that a volume corresponding to the circular area of the eye, should be equal to a volume corresponding to the elliptical area of the eye, excluding the circular area. In practice, however, one has to



(a) (b) (c)

Fig. 3. (a) The 16×16 Peano curve. (b) Watermark produced by raster scan. (c) Watermark produced by Peano scan.

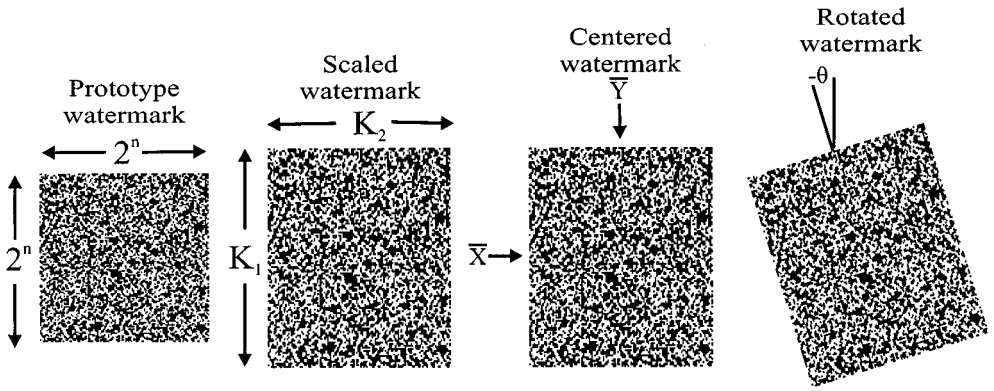


Fig. 4. Geometric watermark adaptation.

search for the minimum difference between the two sides of equation (7) in order to find the best candidate eye position. The weighting functions cannot be easily estimated. However, constant values can be used in (7) without significant loss of accuracy in the estimation of the eye position. These can be evaluated empirically by observing the average luminance of the corresponding areas for a few sample images of the facial image database. In order to define a pattern matching criterion for eye detection, we discretize (7), use constant weighting functions and search for the absolute minimal response of the following difference within the facial region:

$$R_{eye}(x, y) = \left| \sum_{(i,j) \in \mathcal{C}} \sum w_c I(x+i, y+j) - \sum_{(i,j) \in \mathcal{E}-\mathcal{C}} \sum w_e I(x+i, y+j) \right| \quad (8)$$

where $\mathcal{C}$ and $\mathcal{E}$ are the sets of points that belong to the circle and the ellipse that are centered at point (x, y), respectively.

To obtain a reasonable estimate of the relation between the magnitude of the circle and ellipse axes and the weighting constants, we have to compute the sums in (8). Since no knowledge of the precise intensity variation over these areas is available, we simplify (8) by assuming that the intensity is represented by its mean value I_c in the circular iris area, and its mean value I_e outside the iris area and inside the ellipse area. If we let r_1 be the magnitude of the iris radius as well as the minor axis of the ellipse, and r_2 be the magnitude of the major axis of the ellipse, when taking (8) equal to zero, we obtain

$$\frac{w_c \cdot I_c}{w_e \cdot I_e} = \frac{r_2}{r_1} - 1. \quad (9)$$

Considering constant values for I_c, I_e, w_c and w_e, a certain ratio between r_2 and r_1 can be established. The search for potential left and right eye positions is performed over the upper left and upper right quarter, respectively, of the rotated facial region that is covered by the elliptical area produced by the α-trimmed MRBF network. The correct eye position is the one for which the matching response $R_{eye}(x, y)$ is minimal.

A similar pattern matching technique is used for the localization of the mouth, except that the model now consists of two concentric ellipses having major semiaxes of the same magnitude and minor semiaxes of considerably different magnitude. This pattern is again unique for the mouth, and the search is performed in the lower half of the elliptical region. The form of eye and mouth templates that are used are shown in Fig. 1. A relation similar to (9) can be derived in order to decide on the length values of the ellipse axes. A schematic diagram showing the processing steps of this stage is shown in Fig. 2. The robustness issues of the proposed technique are presented in Section VII.

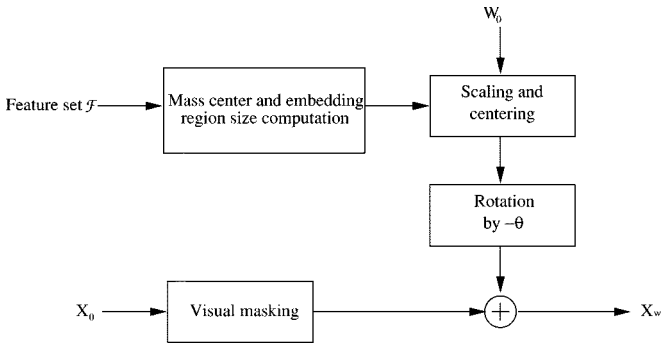


Fig. 5. Watermark embedding stage.

IV. Watermark Construction

In the previous section we developed a method for locating salient features, so that they can be used as reference points to embed our watermark. We should now define the class of watermarks that will be embedded in the spatial image domain. In our experiments, we chose to construct a watermark based on a chaotic trajectory [21], that enables, to a point, control of the spectral properties compared to a pseudo-random sequence. In particular, we employed the Renyi map, which is a recursive function that contains a parameter λ which controls how finely structured the binary watermark is, after the thresholding step that is introduced subsequently. It is a strongly chaotic map, which means that, for slight changes of the initial value, the resulting trajectories diverge very quickly from each other. These properties are desirable in order to have both an acceptable performance against filtering and compression attacks, as well as cryptographic security against attempts of reconstructing the generator function based on a finite set of sample values. Alternatively, the low-passed pseudorandom number generator

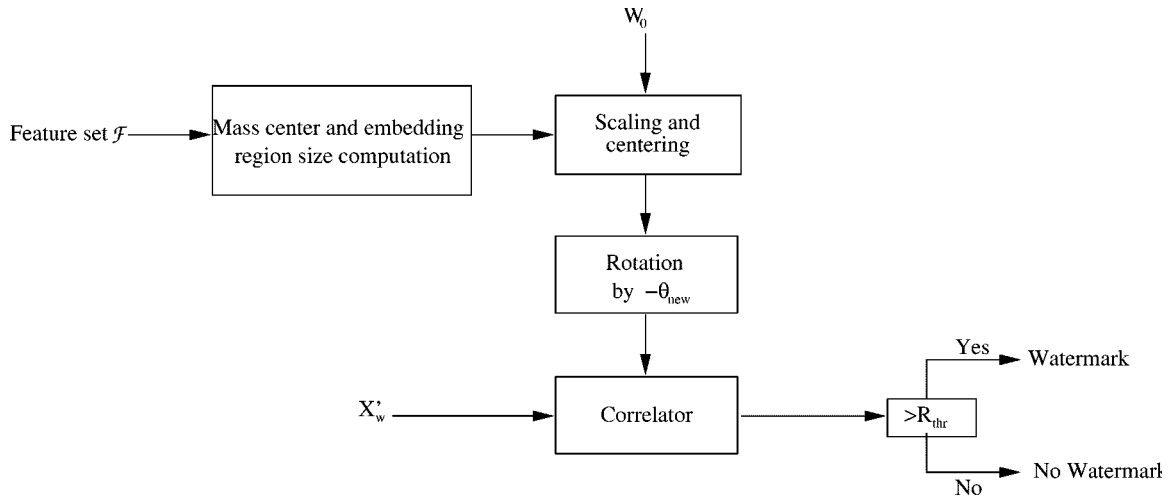


Fig. 6. Watermark detection stage.

output can be used for watermark generation [22], as well as some other chaotic map, like the piecewise affine Markov maps, which have tunable spectral shape [23].

The first step to construct such a watermark is to produce a sequence of real numbers by using a mapping function $\boldsymbol{F}$: $U \rightarrow U$, $U \subset \mathbb{R}$ of the form

$$z(n+1) = \lambda z(n) \bmod 2\pi, \qquad z(n) \in U, \lambda \in \mathbb{R} \tag{10}$$

called the Renyi map [9], $n = 0, 1, 2, \ldots$ denotes the current iteration and λ is a parameter that controls the chaotic behavior of the system. The number of iterations is arbitrary and can be adapted to our needs. The system theoretically produces trajectories of an infinite period. For any value of the parameter λ, the set of real numbers is divided in two subsets: U_{reg} and U_{ch}. The decision on whether the trajectory presents regular or chaotic behavior depends on the seed value $z(0)$. If $z(0) \in U_{reg}$, the produced sequence proves to be periodic, whereas if $z(0) \in U_{ch}$, it is chaotic. The values of the produced trajectory oscillate inside an interval $[z_{\min}, z_{\max}] \subset U$ that is related to λ. Thus, we can define a threshold level $z_{\text{th}} \in [z_{\min}, z_{\max}]$ 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 $-1 \rightarrow 1$ and $1 \rightarrow -1$. For $\lambda > 1$ and values close to 1, we get a chaotic watermark with low number of transitions and, thus, lowpass properties, whereas, when $\lambda \simeq 2$ is chosen, the transitions are very frequent, the lowpass properties degrade and the sequence degrades to a pseudorandom one.

However, the sequence we produced so far is 1-D. To embed it on a two-dimensional (2-D) signal, such as a digital image, we need to scan across the sequence in such a way that the lowpass properties are preserved. The classic raster scan is not proper for this task because the number of transitions is not any more under control in the vertical dimension. To avoid this, we use Peano scan order, which has the property that every pixel along the scan is topologically closer to the previous and subsequent pixels than in the case of raster scan. A Peano curve is a space-filling curve that represents a linear traversal of a multidimensional grid [24]. Fig. 3(a) shows a Peano curve of size 16×16. An analysis of its locality property can be found in [25]. In addition, it is possible to use cellular smoothing to eliminate spontaneous transitions that emerge after the Peano scan [21]. An example of a watermark of size 128×128 constructed using raster scan order and cellular smoothing for $\lambda = 1.3$ is shown in Fig. 3(b). A watermark of the same size and for the same value of λ for a Peano scan followed by cellular smoothing is shown in Fig. 3(c). We can notice the lowpass nature of the watermark produced by the Peano scan. By using this technique, the output watermark has local neighborhoods of 1s (or -1s) that are more compact. The main disadvantage of the Peano scan is that it produces square $2^n \times 2^n$ watermarks only. In the watermark embedding and detection sections, we shall see that this does not seriously affect the performance of our method.

In order to construct different watermarks, we use a key K that corresponds to the seed value $z(0)$ of the chaotic trajectory. Keys of slightly different value provide sufficiently uncorrelated trajectories, because the set $\boldsymbol{K}$ of possible keys is quite large. This reduces the possibility of the watermark being tampered. This also ensures non-invertibility of the watermark. Thus, the corresponding key cannot be extracted from the 2-D watermark.

V. Watermark Embedding

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

The prototype watermark of size $2^n \times 2^n$ is first scaled to the size of the facial area where it will be embedded, using nearest-neighbor interpolation. We choose to construct a prototype watermark whose dimensions will be smaller than these of the embedding area, so that we have both a large set of different watermarks and an insignificant loss of watermark energy. If A_{em} is the embedding area, its size $K_1 \times K_2$ is defined by

$$\begin{aligned} K_1 &= k(x_{(mouth)} - \overline{x}_{(eyes)}), \\ K_2 &= l(y_{(right\_eye)} - y_{(left\_eye)}) \end{aligned} \tag{11}$$

where $x_{(\cdot)}$ and $y_{(\cdot)}$ are the feature coordinates, $\overline{x}_{(\cdot)}$ are mean feature coordinates, and k, l are normalizing factors that control

the size of A_{em} so that it covers at least the entire facial region. The scaled watermark is centered in the mass center of the feature point set $\mathcal{F} = \{F_i, i = 1, \ldots, M\}$:

$$(\overline{x}, \overline{y}) = \left(\frac{1}{M} \sum\sum_{(x,y) \in \mathcal{F}} x, \frac{1}{M} \sum\sum_{(x,y) \in \mathcal{F}} y \right). \quad (12)$$

In our case $M = 3$. The mass center is also the center of the region A_{em}. After centering the watermark in the proper image region, it is rotated by $-\theta$ according to (5) with respect to the image center. It then covers a new area A_{rot}. The various steps of the watermark adaptation procedure are depicted in Fig. 4. Before superimposing it on the original image, a visual masking stage can be introduced.

If w_0 is the prototype watermark, then the scaled, centered, and rotated watermark w_n of size $K_1 \times K_2$ is embedded to the region A_{rot}. The watermarked image $f_w(x, y)$ is defined as

$$f_w(x, y) = f(x, y) \quad (x, y) \notin A_{rot} \vee ((x, y) \in A_{rot} \wedge \mathrm{Var}(x, y) \leq T_{\mathrm{var}}) \quad (13)$$

$$f_w(x, y) = f(x, y) + h \cdot w_n(x, y) \quad (x, y) \in A_{rot} \wedge \mathrm{Var}(x, y) > T_{\mathrm{var}} \quad (14)$$

where h is the watermark power, $\mathrm{Var}(x, y)$ is the local image variance, and T_{var} an appropriate variance threshold that must be chosen in such a way that a sufficiently large image area is watermarked. Alternatively, h can become a function of the local variance

$$h(x, y) = h_{\max} \cdot s(\mathrm{Var}(x, y)) \quad (15)$$

where $s(\cdot)$ takes values in the range $[0, 1]$ and s is chosen to increase monotonically with the variance. In our case, the watermark is embedded in the spatial domain and, thus, the watermark power must be of integer value. In accordance to masking, it is quantized to either of two values, zero or $h_{\max}$, depending on the value of the variance. Visual masking produces interesting results provided that the local image variance is high in several image regions, so the threshold is exceeded and the watermark can be embedded. However, this is not the case for frontal facial images that contain mostly homogeneous image regions. Therefore, we used low watermark power h to ensure watermark invisibility and we did not employ visual masking in our experiments, to ensure increased performance. Larger images would also provide larger areas for embedding and, thus, additional watermark energy. The main steps of the embedding process are depicted in Fig. 5.

VI. Watermark Detection

When a prototype watermark is to be detected inside a watermarked and possibly manipulated image, the image has first to be segmented, so that the feature set and orientation of the approximated facial region are derived. The prototype watermark is again scaled, centered, and rotated according to the information obtained from the segmentation and feature extraction stage. For the detection region A_{det}, the response of a hypothesis testing detector is computed

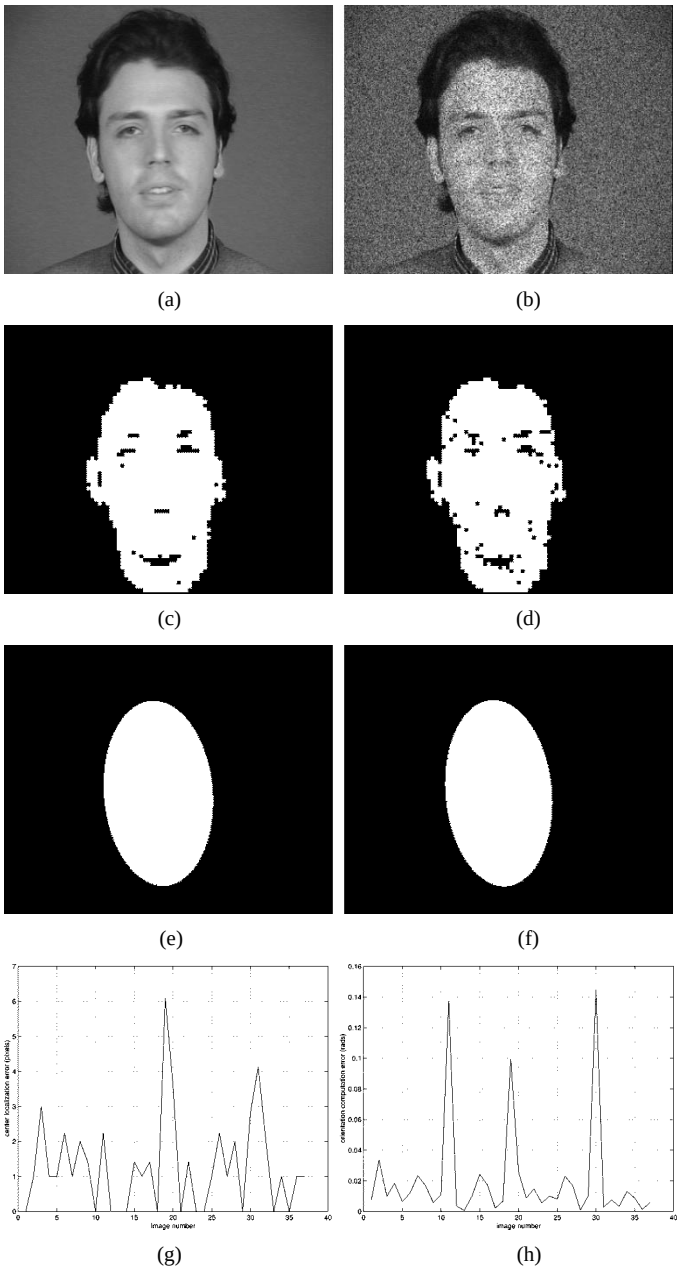


Fig. 7. (a) Original image. (b) Image corrupted by Gaussian noise. (c) Original image facial region. (d) Corrupted image facial region. (e) Elliptical approximation of original face region. (f) Elliptical approximation of corrupted facial region. (g) Center localization error. (h) Orientation computation error.

$$R(\hat{f}_w, \hat{w}_n) = \frac{1}{N_{\boldsymbol{A}}} \sum_{(x,y) \in \boldsymbol{A}} \hat{f}_w(x, y) - \frac{1}{N_{\boldsymbol{B}}} \sum_{(x,y) \in \boldsymbol{B}} \hat{f}_w(x, y) \quad (16)$$

where $\boldsymbol{A} = \{(x, y) \in A_{\mathrm{det}} | \hat{w}_n(x, y) = 1\}$ and $\boldsymbol{B} = \{(x, y) \in A_{\mathrm{det}} | \hat{w}_n(x, y) = -1\}$. $N_{\boldsymbol{A}}$ and $N_{\boldsymbol{B}}$ are the number of pixels of the sets $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. This detector expresses the difference of two sample means [26]:

$$\overline{\boldsymbol{w}} = \overline{\boldsymbol{x}} - \overline{\boldsymbol{y}} \quad (17)$$

Fig. 8. Salient feature extraction results on original images.

Fig. 9. Salient feature extraction results on watermarked images.

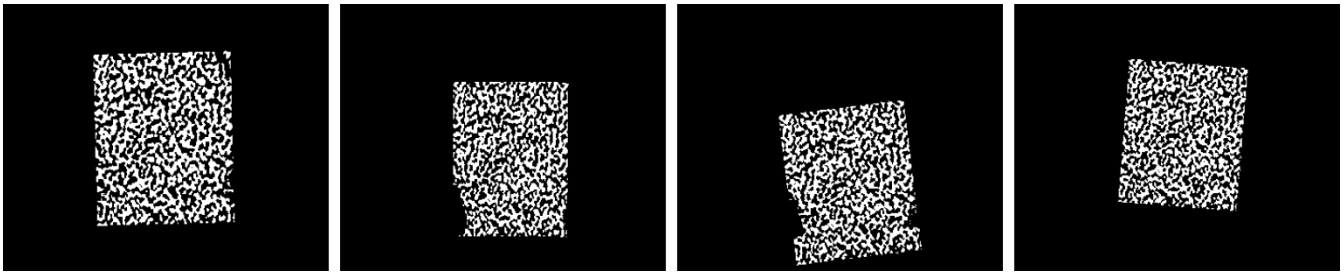


Fig. 10. Amplified difference of watermarked and original images.

where

$$\overline{\boldsymbol{x}} = \frac{1}{N_{\boldsymbol{A}}} \sum_{(x,y)\in \boldsymbol{A}} \hat{f}_w(x, y), \quad \overline{\boldsymbol{y}} = \frac{1}{N_{\boldsymbol{B}}} \sum_{(x,y)\in \boldsymbol{B}} \hat{f}_w(x, y). \tag{18}$$

Considering that the pixels of the original, as well as the watermarked image are i.i.d., the mean value and variance of the detector output are

$$\eta_{\overline{w}} = \eta_{\overline{x}} - \eta_{\overline{y}} \quad \sigma^2_{\overline{w}} = \left(\frac{1}{N_{\boldsymbol{A}}} + \frac{1}{N_{\boldsymbol{B}}}\right) \sigma^2_{\hat{f}_w}. \tag{19}$$

In general, both f and $\hat{f}_w$ are not i.i.d. Therefore, the watermark correlation detector is suboptimal and whitening should be performed to obtain optimal detection [27]. The experimental results are proven to be in accordance with this simplified model, though the model is not exact. In general, the detector output is assumed to follow a normal distribution. If the correct watermark is embedded on the image, the mean value of the detector output is $\eta_{\overline{w}} = 2h$ and its variance is $\sigma^2_{\overline{w}} = ((1/N_{\boldsymbol{A}}) + (1/N_{\boldsymbol{B}}))(\sigma^2_f + \sigma^2_w)$, where σ^2_f is the variance of the initial image and σ^2_w is the variance of the watermark. Otherwise, if no watermark is present, the mean value of the detector is $\eta_{\overline{w}} = 0$ and the variance is $\sigma^2_{\overline{w}} = ((1/N_{\boldsymbol{A}}) + (1/N_{\boldsymbol{B}}))\sigma^2_f$. This variance is not significantly different than in the case the watermark is present, because the factor $(1/N_{\boldsymbol{A}}) + (1/N_{\boldsymbol{B}})$ is very small and $\sigma^2_w \ll \sigma^2_f$. This is only a suboptimal type of detector and other techniques (e.g., pre-whitening) could be employed to improve performance. Still, this is a simple and widely used detector.

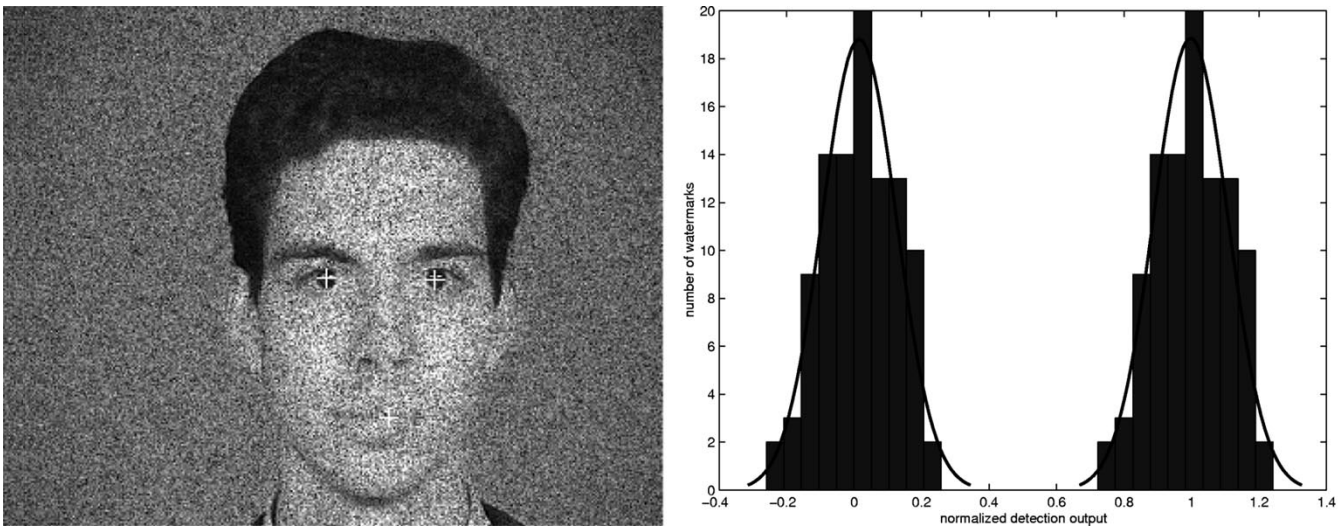
For the class of facial images under consideration, the features are not expected to be localized at exactly the same positions in the watermarked and possibly processed image as in the original image. This can be faced efficiently by testing the correlation output for small changes in the height, width, orientation, and center of the prototype watermark, compared to those computed after extracting the new features and the new orientation of the elliptical approximation. The detection is expected to give a strong peak for the correct height, width, orientation, and center coordinates of the originally embedded watermark, because of the high degree of watermark sensitivity to geometric operations. The detection output is weak if wrong geometrical parameters are used. These geometrical parameters can be found by a local search. This need not be more than 2 pixels in the average for height, width, and center coordinates and no more than 0.02 radians in the average for the orientation. Therefore, the detection response in the diagrams is always shown for the correct size, orientation, and center coordinates of the watermark. The method is expected to perform better for larger images than the ones used in our experiments.

(a) (b) (c)

Fig. 11. (a) Original image. (b) Watermarked image using face region. (c) Normalized detector output.

(a) (b) (c)

Fig. 12. (a) Original image. (b) Watermarked image using extended region. (c) Normalized detector output.

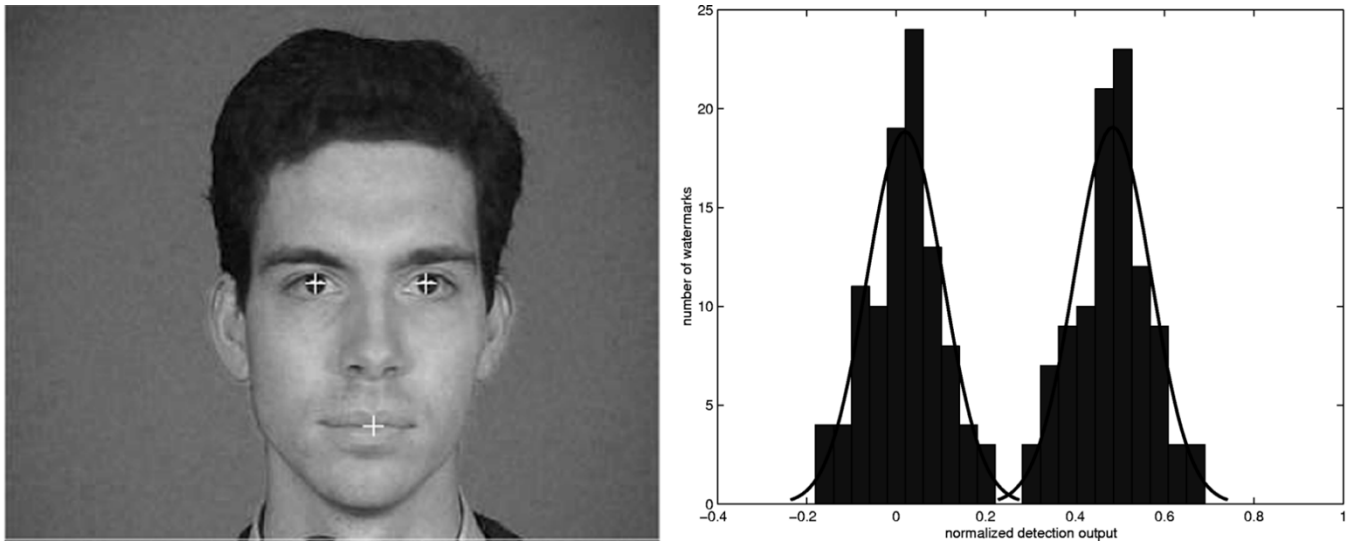


(a) (b)

Fig. 13. (a) Watermarked image after distortion by multiplicative Gaussian noise with $\sigma = 0.3$. (b) Normalized detector output.

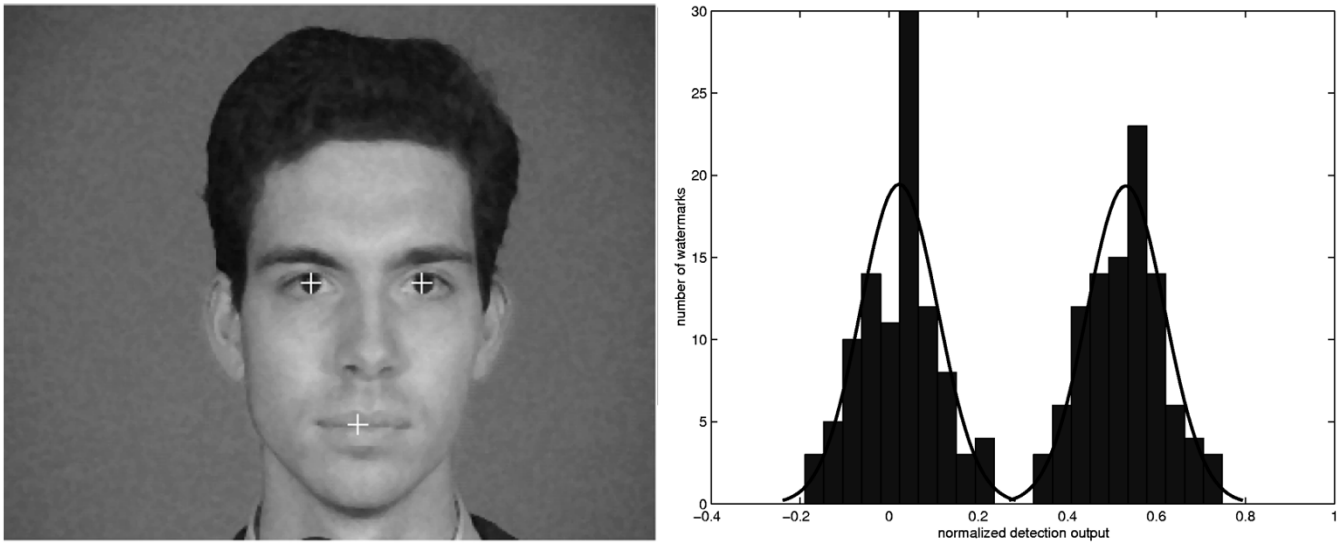
We choose not to use masking in the detection stage, even if masking was employed in the embedding stage, because the local variance may have changed significantly due to manipulations. The response is thus computed over the entire expected area of the embedded watermark. The detector output (16) must 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. A schematic diagram of the detection stage is presented in Fig. 6.

In order to decide for an efficient threshold indicating watermark existence, we can follow an experimental approach. One hundred different watermarks are embedded and detected after an attack on both the original and the respective watermarked image. The experimental output of the detector for both cases is approximated by a normal distribution. The pdf of the detector output has zero mean when evaluated on an unwatermarked image (left-hand distribution) and $2h$ when evaluated on a watermarked image (right-hand distribution). We wish to



(a) (b)

Fig. 14. (a) Watermarked image after 1 : 30 JPEG compression. (b) Normalized detector output.



(a) (b)

Fig. 15. (a) Watermarked image after 3×3 median filtering. (b) Normalized detector output.

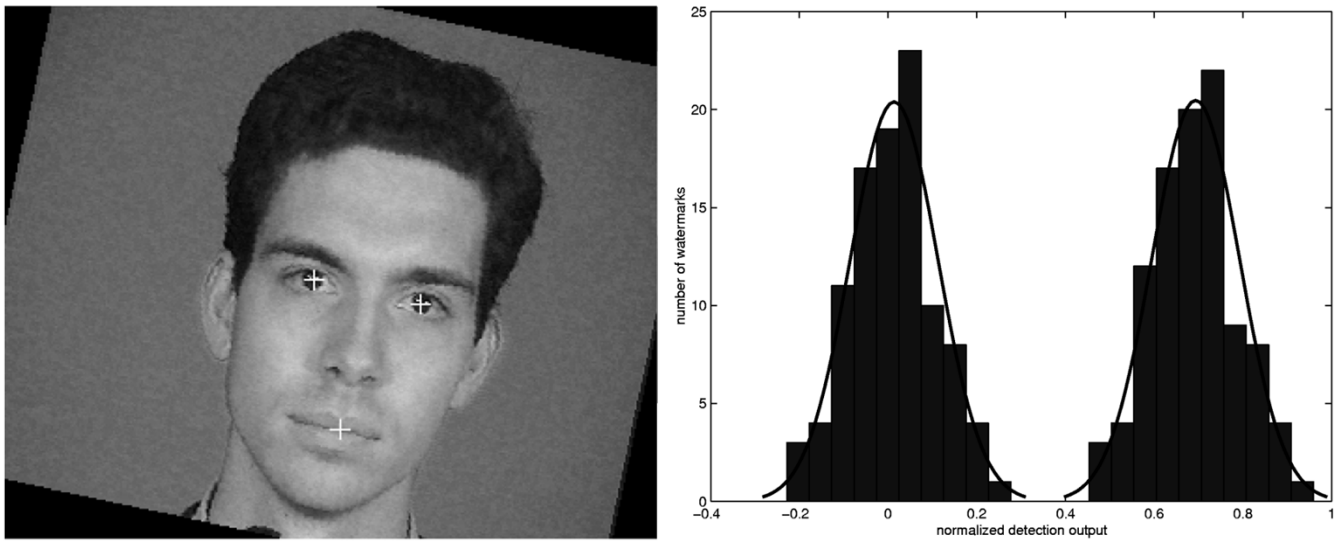
obtain an acceptable compromise between the false acceptance and the false detection ratio. After each attack, we can evaluate the two distributions and choose the optimal threshold. If they have equal variances, the optimal threshold is given by $R_{thr} = (mean_l + mean_r)/2$, where $mean_l$ and $mean_r$ are the two distribution means. The median value of the acceptable thresholds R_{thr} chosen for each attack is considered as the common threshold for watermark detection under any attack. This was the approach that was followed in our experiments.

VII. Experimental Results

In order to test the robustness of the facial feature extraction method to some usual attacks, we first choose to simulate the possible distortions by using Gaussian noise. This is done for all three channels of the initial RGB image. The output is of the form:

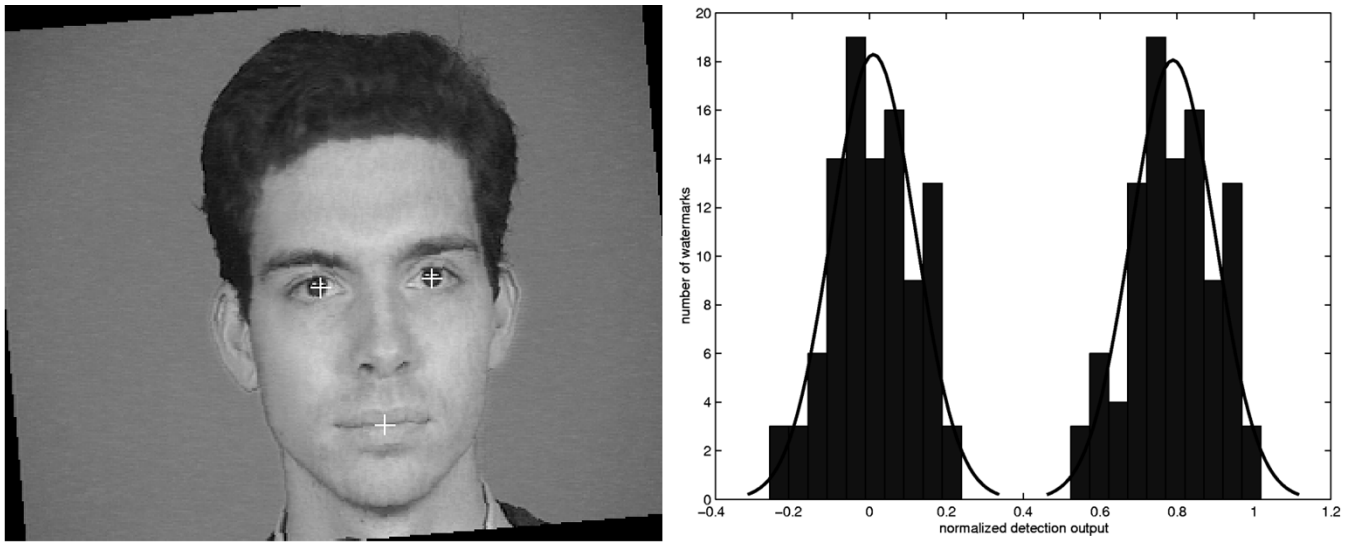
$$g_i(x, y) = f_i(x, y)(1 + \sigma \cdot n(x, y)) \qquad i = 1, 2, 3 \quad (20)$$

where $n(x, y)$ follows a Gaussian distribution of zero mean and unit variance and σ defines the standard deviation of the multiplicative Gaussian noise. We added noise to 37 images of the M2VTS face database and computed the moment-based features, i.e., the mass center and orientation of the resulting ellipse before and after noise addition. Fig. 7(a) and (b) show an original sample image and the same image after noise addition with $\sigma = 0.3$. The corresponding facial regions are shown in Fig. 7(c) and (d). The derived elliptical approximations are depicted in Fig. 7(e) and (f). Finally, Fig. 7(g) and (h) show the relative change of the center position (in pixels) and of the orientation (in rads) of the approximating ellipse, respectively. The



(a) (b)

Fig. 16. (a) Watermarked image after rotation by 12°. (b) Normalized detector output.

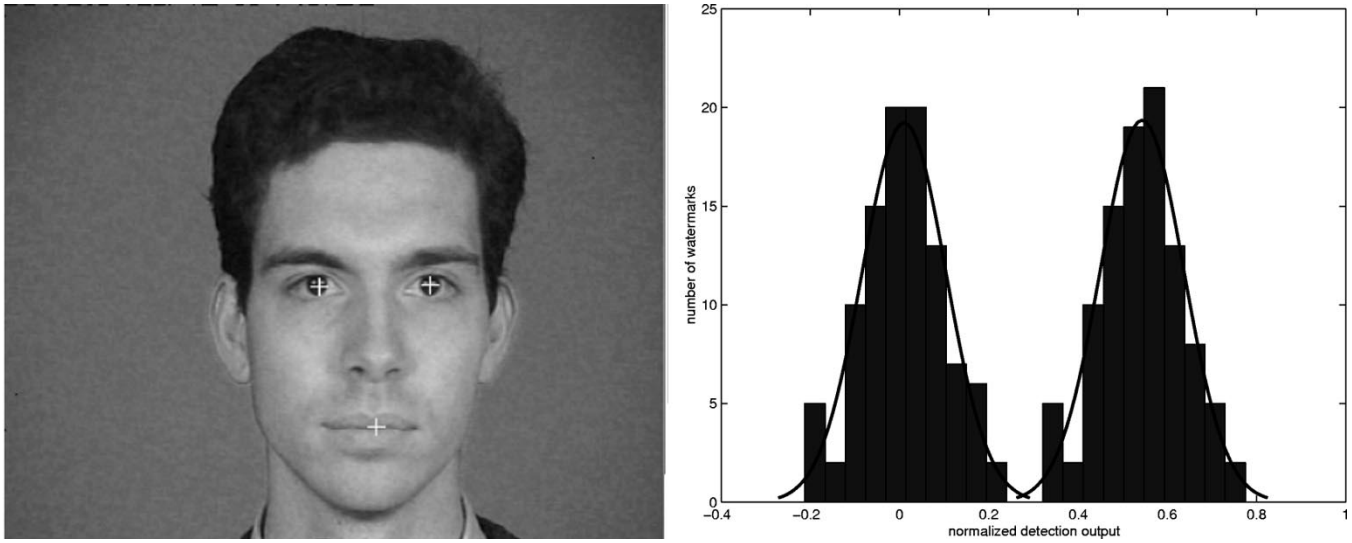


(a) (b)

Fig. 17. (a) Watermarked image after rotation by -5°. (b) Normalized detector output.

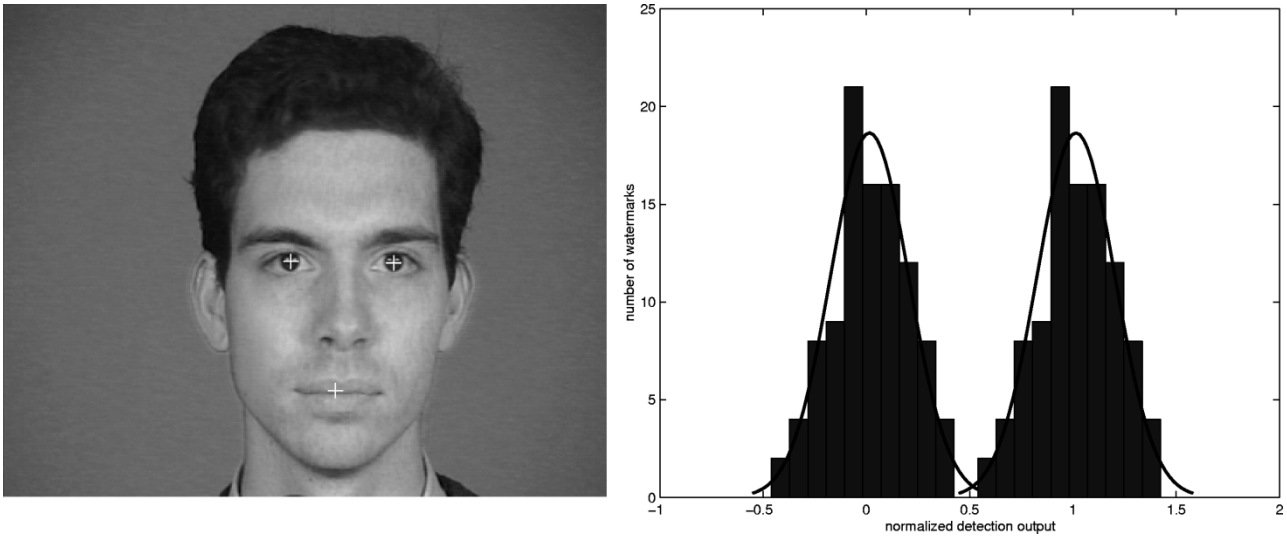
mean values of center estimation error and orientation estimation error are 1.2974 pixels and 0.0205 rads, respectively. The results show that the region approximation is quite robust even under significant distortion.

In order to demonstrate the robustness of the watermarking method to various attacks, we tested it on 37 color images (of size 350×286) of the M2VTS frontal facial image database [17]. The feature extraction success rate was 84% for combined eye and mouth detection, providing an equal success rate for watermark detection over the images. However, the watermarking scheme can perform correctly even if some features are not detected at their correct position, provided that all three features (the eyes and the mouth) are present. This raises the performance to about 95% of the images during watermark detection. It is, anyway, essential to obtain the three feature points (eyes, mouth locations) in order to define the dimensions, center, and orientation of the watermarking region. Fig. 8 shows some detection examples. Fig. 9 shows the corresponding results after watermark embedding. The size of the prototype watermark is 128×128, the watermark strength is $h = 3$, and the chaotic map parameter is $\lambda = 1.8$. The value for watermark strength is a good compromise between watermark robustness and invisibility. Larger values would result in visually significant watermarks, whereas smaller ones would result in hardly separated experimental distributions. Since no masking is employed, three could be considered an upper bound value. The k and l parameters for defining the watermark spread over the facial region are both fixed at value 1.25. The normalized threshold variance for masking is $T_{\text{var}} = 0.002$. Fig. 10 shows the amplified difference of the watermarked and original images. We can clearly see the parts of the images that have not been watermarked at all due to masking. It is also obvious that, although the features have been distorted by the watermark, the positions of the salient features in the watermarked image are very close to the original ones



(a) (b)

Fig. 18. (a) Watermarked image after scaling by a factor of 1.16. (b) Normalized detector output.

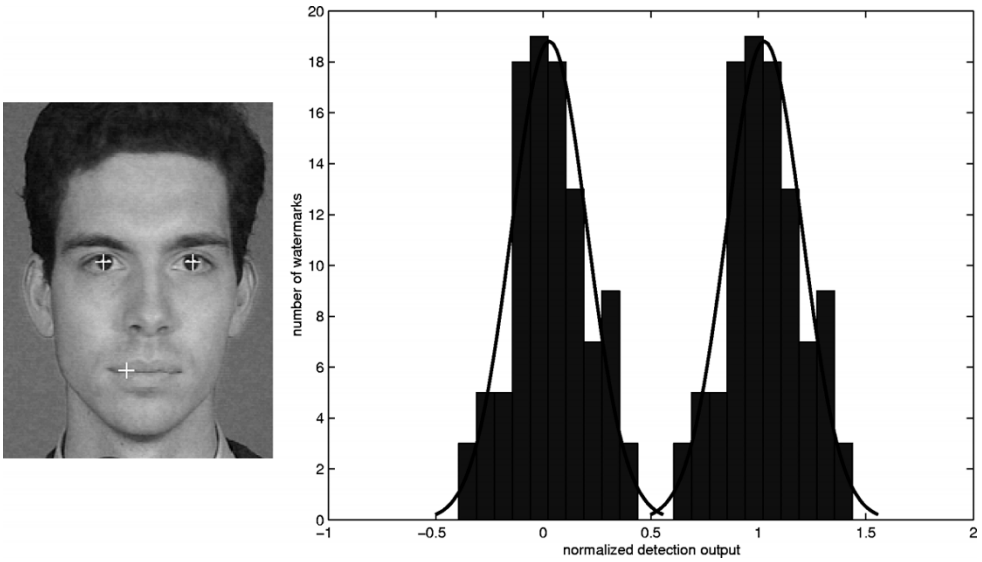


(a) (b)

Fig. 19. (a) Watermarked image after scaling by a factor of 1.21. (b) Normalized detector output.

because their geometric models are robust enough. However, introduced visual masking may drastically reduce the area over which the watermark is embedded, due to the lowpass spectral characteristics of the frontal facial images. Thus, in the subsequent experiments we did not employ masking.

A sample image after watermarking using a 128×128 prototype watermark and feature extraction using the previously mentioned parameter values is shown in Fig. 11(b). The corresponding original is shown in Fig. 11(a). Fig. 11(c) shows the normal distribution of the detector output for 100 different watermarks searched in the original image (on the left) and in the watermarked image (on the right). The detector output is normalized in both cases. We can notice the rather large variance of both distributions that may cause problems when having to decide on a sufficient detection threshold. Fig. 12(b) shows the result of feature extraction when embedding a 256×256 prototype watermark on an image region that is extended to the image boundaries, while keeping the center and aspect ratio of the embedding area unmodified. As Fig. 12(c) shows, the variance of both distributions has decreased, thus providing a wider range of possible threshold values that would result in satisfactory detection ratios. However, this causes problems in the case of cropping attacks, because the embedding area is extended to the false boundaries of the cropped image, instead of the watermarked one. This can only be faced by searching for the correct cropping factor. We can see that a threshold that separates completely the two distributions can be found. The common detection threshold is decided after performing a set of attacks, considered as a training set, on the watermarked image corresponding to Fig. 12(a). The common threshold was decided to be 0.32, after computing the median of the thresholds obtained for the following attacks: $1:30$ ratio JPEG compression, 3×3 median filtering, rotation by $12°$, scaling by a factor of 1.16, asymmetric cropping to a size of 188×248, and no attack. All these attacks, as well as any other, are evaluated using this common threshold.



(a) (b)

Fig. 20. (a) Watermarked image after cropping to size 188×248. (b) Normalized detector output.

TABLE I
DETECTOR FRR, FAR FOR SEVERAL ATTACKS (COMMON DETECTION THRESHOLD = 0.32)

attack	FRR	FAR
no attack	$1.3094 \cdot 10^{-16}$	$2.1585 \cdot 10^{-4}$
multiplicative Gaussian noise ($\sigma = 0.3$)	$3.3269 \cdot 10^{-10}$	0.0028
1:30 JPEG compression	0.0273	$2.1015 \cdot 10^{-4}$
median (3×3)	0.0077	$3.1819 \cdot 10^{-4}$
rotation by 12^0	$7.7982 \cdot 10^{-5}$	$9.6364 \cdot 10^{-4}$
rotation by -5^0	$8.4085 \cdot 10^{-6}$	0.0023
scaling by 1.16	0.0085	$4.7922 \cdot 10^{-4}$
scaling by 1.21	$1.1913 \cdot 10^{-4}$	0.0536
rotation by -5^0 and scaling by 1.16	0.0016	0.0712
cropping (188×248)	$2.9899 \cdot 10^{-5}$	0.049

Fig. 13 shows the result after corruption of the watermarked image by multiplicative Gaussian noise of $\sigma = 0.3$. In Fig. 14 a result is shown for the same watermarked image after JPEG compression of ratio 1 : 30. Fig. 15 shows a result after performing 3×3 median filtering. Figs. 16 and 17 show results for rotation by $12°$ and $-5°$, respectively. Figs. 18 and 19 show results for scaling by a factor of 1.16 and 1.21, respectively. Finally, Fig. 20 shows a result for asymmetric cropping to a size of 188×248.

Table I shows the false acceptance ratio (FAR) and the false rejection ratio (FRR) for each of the attacks on the considered image using the common detection threshold. We can see that, in some cases, the FAR is increased in favor of the FRR, and vice versa. This depends on how much the common threshold differs from the optimal threshold for the particular attack. The performance of the method will become better for larger images. This is because the facial region would also be larger, thus allowing the watermark to be embedded on a larger set of points. Another point is that the prototype watermark could also be greater, allowing for smaller cross-correlation between different watermarks and, thus, providing detector response pdfs of smaller variance.

VIII. CONLUSIONS AND FUTURE WORK

In this paper, we developed a method for embedding and detecting watermarks in color frontal facial images. Color information was exploited in order to obtain a good approximation of the skin-colored facial region, in which we search for salient features like the eyes and the mouth, using a geometric model matching method. The prototype watermark used for embedding was chosen to be a chaotic one, modified in such a way as to retain certain lowpass properties. The watermark is geometrically adapted before embedding, using the extracted feature positions and facial region orientation. A correlation detector is employed in order to decide on the possible presence of a watermark. The color segmentation and feature localization technique precedes both embedding and detection stages. Experimental results display the robustness of the method. Future directions of the current work include development of more robust techniques for salient feature extraction, improvement of the watermark detection stage performance, as well as examination of alternative chaotic generators that may perform better than the one employed in this work.

REFERENCES

[1] G. Voyatzis and I. Pitas, "The use of watermarks in the protection of digital multimedia products," *Proc. IEEE*, vol. 87, pp. 1197–1207, July 1999.
[2] N. Nikolaidis and I. Pitas, "Copyright protection of images using robust digital signatures," in *Proc. IEEE ICASSP'96* Atlanta, GA, May 1996, vol. 4, pp. 2168–2171.
[3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Syst. J.*, vol. 25, pp. 313–335, 1996.

[4] I. J. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.
[5] A. Piva, M. Barni, F. Bartolini, and V. Capellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," in *Proc. IEEE ICIP'97* Santa Barbara, CA, Oct. 1997, vol. 1, pp. 520–523.
[6] J. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant digital image watermarking," in *Proc. ICIP '97* Santa Barbara, CA, Oct. 1997, vol. 1, pp. 536–539.
[7] X.-G. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," in *Proc. ICIP '97* Santa Barbara, CA, Oct. 1997, vol. 1, pp. 548–551.
[8] M. J. J. J. B. Maes and C. W. A. M. van Overveld, "Digital watermarking by geometric warping," in *Proc. ICIP '98* Chicago, IL, Oct. 1998, vol. 2, pp. 424–426.
[9] R. L. Devaney, *An Introduction to Dynamical Systems*: Penjamine/Cummings, 1986.
[10] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal Processing: Image Communication*, vol. 12, no. 3, pp. 263–281, 1998.
[11] S. Tsekeridou and I. Pitas, "Facial feature extraction in frontal views using biometric analogies," in *Proc. of EUSIPCO '98* Rhodes, Greece, September 1998, vol. 1, pp. 315–318.
[12] G. Yang and T. S. Huang, "Human face detection in a complex background," *Pattern Recognition*, vol. 27, no. 1, pp. 53–63, 1994.
[13] R. Chellapa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.
[14] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, 1993.
[15] X. Li and N. Roeder, "Face contour extraction from front-view images," *Pattern Recognition*, vol. 28, no. 8, pp. 1167–1179, 1995.
[16] M. J. T. Reinders, P. J. L. van Beek, B. Sankur, and J. C. A. van der Lubbe, "Facial feature localization and adaptation of a generic face model for model-based coding," *Signal Processing: Image Communication*, vol. 7, no. 1, pp. 57–74, 1995.
[17] S. Pigeon and L. Vandendorpe, "The M2VTS multimodal face database," in *Lecture Notes in Computer Science: Audio- and Video-Based Biometric Person Authentication*, J. Bigun, C. Chollet, and G. Borgefors, Eds., 1997, vol. 1206, pp. 403–409.
[18] J. C. Terrillon and S. Akamatsu, "Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images," in *Vision Interface*, Canada, 1999.
[19] A. G. Bors and I. Pitas, "Object classification in 3-D images using alpha-trimmed mean radial basis function network," *IEEE Trans. on Image Processing*, vol. 8, no. 12, pp. 1744–1756, 1999.
[20] A. K. Jain, *Fundamentals of Digital Image Processing*, NJ: Prentice-Hall, 1989.
[21] G. Voyatzis and I. Pitas, "Chaotic watermarks for embedding in the spatial digital image domain," in *Proc. ICIP '98* Chicago, IL, Oct. 1998, vol. 2, pp. 432–436.
[22] T. Kalker, J. P. Linnartz, G. Depovere, and M. Maes, "On the reliability of detecting electronic watermarks in digital images," in *Proc. EUSIPCO '98* Rhodes, Greece, September 1998, vol. 1, pp. 13–16.
[23] S. H. Isabelle and G. W. Wornell, "Statistical analysis and spectral estimation techniques for one-dimensional chaotic signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 1495–1506, June 1997.
[24] D. Coltuc and I. Pitas, "Memory mappings for fast Peano scan," in *Proc. ECCT 1995 Eur. Conf. Circuits Theory and Design*, Istanbul, Turkey, Aug. 1995, pp. 817–821.
[25] C. Gotsman and M. Lindenbaum, "On the metric properties of discrete space filling curves," in *Proc. Int. Conf. Pattern Recognition* Jerusalem, Israel, Oct. 1994, vol. 3, pp. 98–102.
[26] A. Papoulis, *Probability and Statistics*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
[27] G. Depovere, T. Kalker, and J.-P. Linnartz, "Improved watermark detection reliability using filtering before correlation," in *ICIP '98* Chicago, IL, Oct. 1998, vol. 1, pp. 430–434.

Athanasios Nikolaidis was born in Serres, Greece, in 1973. He received the Diploma in computer engineering from the University of Patras, Greece, in 1996. He is currently a Research and Teaching Assistant pursuing the Ph.D. degree in the Department of Informatics, Aristotle University of Thessaloniki, Greece. His research interests include nonlinear image and signal processing and analysis, face detection and recognition, and copyright protection of multimedia.

Mr. Nikolaidis is a member of the Technical Chamber of Greece.

Ioannis Pitas (S'83–M'84–SM'94) received the Diploma of electrical engineering in 1980 and the Ph.D. degree in electrical engineering in 1985, both from the University of Thessaloniki, Greece.

Since 1994, he has been a Professor at the Department of Informatics, University of Thessaloniki. From 1980 to 1993, he served as Scientific Assistant, Lecturer, Assistant Professor, and Associate Professor in the Department of Electrical and Computer Engineering at the same University. He served as a Visiting Research Associate at the University of Toronto, Toronto, ON, Canada, University of Erlangen–Nuernberg, Germany, Tampere University of Technology, Finland, and as Visiting Assistant Professor at the University of Toronto. He was Lecturer in short courses for continuing education. His current interests are in the areas of digital image processing, multidimensional signal processing and computer vision. He has published over 300 papers and contributed to eight books in his area of interest. He is the co-author of the book *Nonlinear Digital Filters: Principles and Applications* (Norwell, MA: Kluwer, 1990) and author of *Digital Image Processing Algorithms* (Englewood Cliffs, NJ: Prentice-Hall, 1993). He is the editor of the book *Parallel Algorithms and Architectures for Digital Image Processing, Computer Vision and Neural Networks* (New York: Wiley, 1993). He was Co-Editor of *Multidimensional Systems and Signal Processing*.

Dr. Pitas has been member of the European Community ESPRIT Parallel Action Committee. He has also been an Invited Speaker and/or member of the program committee of several scientific conferences and workshops. He was Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS and is currently an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS. He was Chair of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing (NSIP'95). He was Technical Chair of the 1998 European Signal Processing Conference. He is General Chair of IEEE ICIP 2001.