Robust watermarking of facial images based on salient geometric pattern matching

Athanasios Nikolaidis Ioannis Pitas *

Department of Informatics

Aristotle University of Thessaloniki

Box 451, Thessaloniki 540 06, GREECE Tel,Fax: +3031-996304

E-mail: pitas@zeus.csd.auth.gr

Abstract

We introduce a novel method for embedding and detecting a chaotic watermark in the digital spatial domain of color face images, based on localizing salient facial features. A theoretic analysis of the robustness of the method is presented. Finally, resistance to several kinds of attack, such as compression, filtering, scaling, cropping and rotation is displayed based on simulation results.

EDICS number: 5-SECR

*Corresponding Author

1 Introduction

A field of rapidly increasing interest during the last few years has been multimedia protection. This has been addressed by the fact that many innovative techniques for digital data transfer, storage and processing have emerged. This has facilitated malicious users with the opportunity of manipulating images, audio or video and, thus, claiming product ownership.

Many different watermarking methods for still images have been proposed, that embed a pseudo-random sequence in either the spatial or the transform image domain [1]-[6]. All of these methods consider the image as a single object in the spatial domain and embed the watermark across the entire image content without exploiting possible local salient properties. These techniques usually consider immunity against only one category of attacks, either filtering and compression or geometric distortions. Usually, the methods are customized in such a way that no care is taken whether the watermark is resistant against the other category of attacks.

A technique that tries to cope with geometric changes while retaining robustness to other kinds of attack, is presented in [7]. 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, the cost of producing a theoretically infinite set of different line patterns in the detection stage, together with the questionable prominence of the selected points to be warped, render this technique unsuitable. A significant step would be to extract salient features based on their robustness against every possible attack.

The current paper presents a technique for watermarking a certain image class that has special interest, namely color frontal face images with uniform background, based on extracting a finite set of salient facial features and applying proper geometric pattern matching in the face region. This image class has two important characteristics: a) it contains several salient features, notably the eyes and the mouth, that can be used for matching, and b) it has particular applications related to copyright protection of portrait galleries. For this class of images, the facial region is the most important. We shall concentrate our study in the use of watermarks for copyright protection. The proposed technique provides sufficient immunity to any intended attack. Section 2 presents the color segmentation and region approximation technique, as well as the pattern matching process for localizing the salient features. In Section 3, the general class of chaotic watermarks is presented together with adaptations for digital images. Section 4 explains the connection between the extracted features and the watermark to be embedded on the image, and Section 5 presents the watermark detection procedure. Simulation results for several kinds of manipulations on the watermarked image are presented and explained in Section 6 and, finally, conclusions are drawn in Section 7.

2 Face Segmentation and Salient Feature Localization

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

In the following we propose one technique for eyes and mouth localization that is rotation and scale invariant. Any other technique proposed for this purpose in the literature can be used as well [8]-[9]. The method followed is based on exploiting color information in a similar way as in [8]. Our aim is to discriminate the skin-colored image region. The original RGB image is converted to its HSV representation, because it is easier to perform 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:

$$\begin{aligned} 0 \leq H \leq 25 \quad &or \quad 335 \leq H \leq 360 \\ 0.2 \leq \quad &S \quad \leq 0.6 \\ 0.4 \leq \quad &V \end{aligned} \tag{1}$$

The choice of the above mentioned ranges ensures that the segmented region of interest will approximately be the same even after some manipulation and only an insignificant number of pixels will exceed these thresholds. A brief theoretic analysis that proves the robustness of the method follows.

A connected components algorithm follows thresholding in order to isolate all the compact skin-colored regions. The downsampled map is scaled to the original image size. In order to get a good approximation of the face region that does not contain useless areas, e.g. the neck, as well as to prevent the face region areas from getting connected to the background, we employ an $\alpha-$trimmed Mean Radial Basis Function network to get an ellipsoidal approximation of the region [10]. This technique provides the marginal median estimation for the center of each object:

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

where $X_{(i)}$ are the marginal data samples sorted according to their values, N_k is the total number of data samples assigned to the k-th hidden unit, and α_k is the percentage of data samples to be trimmed away. The estimate of the covariance matrix is:

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

where $\mathbf{X}_{(i),\mathcal{M}}$ is the i-th ordered data sample according to the Mahalanobis distance, and $\alpha_{k,\mathcal{M}}$ is the trimming percentage in this case.

Once the face region is segmented, the trimmed ellipsoidal approximation is known and thus its orientation [11] can be computed by:

$$\theta = \frac{1}{2} \arctan \left[\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \right] \tag{4}$$

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

$$\mu_{p,q} = \frac{1}{N-1} \sum_{(m,n)} \sum_{\in \mathcal{R}} (m - \bar{m})^p (n - \bar{n})^q \tag{5}$$

and $(\bar{m}, \bar{n})$ is the mass center of the region $\mathcal{R}$. The input image should be rotated according to the angle given by (4) before pattern matching.

The most prominent features contained in the ellipsoidal area are the eyes and the mouth and can be approximated sufficiently well by proper geometric functions. These features are unique in such images and act as robust reference points, whatever the geometric distortion imposed to the image is. Other similar approaches use 2-D sinc functions for eye modeling [9]. The eye can be regarded as a circle of low, almost constant intensity centered inside an ellipse of very bright intensity. Ideally, the eye detector is described by:

$$\begin{aligned} \int\int_{(x,y)\in\mathcal{C}_{per}} w_c(x,y) I(x,y) dxdy &= \int\int_{(x,y)\in\mathcal{E}_{per}} w_e(x,y) I(x,y) dxdy \\ &- \int\int_{(x,y)\in\mathcal{C}_{per}} w_e(x,y) I(x,y) dxdy \end{aligned} \tag{6}$$

where $I(x,y)$ is the image intensity, $\mathcal{C}_{per}$, $\mathcal{E}_{per}$ are the sets of pixels lying on the perimeter of the circle and the ellipse, respectively, and $w_c(x,y)$ and $w_e(x,y)$ are weighting functions that compensate for the luminance differences between the two areas. In other words, this means that the approximate volumes of two cylindroids should be the same. The weighting functions cannot be easily estimated. However, constant values can be incorporated without significant loss of accuracy in the estimation of the feature position. These can be evaluated empirically by observing the average luminance of the corresponding areas for a few sample images of the database. In order to define a pattern matching criterion for eye detection, we discretize (6), use constant weighting functions and search for the absolute minimal response of the difference within the facial region:

$$R_{eye}(x,y) = \left| \sum_{(i,j)} \sum_{\mathcal{C}} w_c I(x+i, y+j) - \sum_{(i,j)} \sum_{\in \mathcal{E}-\mathcal{C}} w_e I(x+i, y+j) \right| \tag{7}$$

where $\mathcal{C}$ and $\mathcal{E}$ are the sets of points that belong to the circle and the ellipse, respectively.

To obtain a reasonable estimate of the relation between the magnitudes of the circle and ellipse axes and the weighting constants, we have to compute the integrals in (6). For simplicity, we assume that the intensity is represented by its mean value I_c in the circle area, and its mean value I_e outside the circle area and inside the ellipse area. If we let r_1 be the magnitude of the circle radial as well as the minor axis of the ellipse, and r_2 be the magnitude of the major axis, the desired relation is:

$$\frac{w_c \cdot I_c}{w_e \cdot I_e} = \frac{r_2}{r_1} - 1 \tag{8}$$

The search for potential left and right eye positions is performed over the upper left and upper right quarter of the rotated facial region that is covered by the ellipsoidal area, respectively. 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 magnitudes. This pattern is again unique for the mouth, and the search is performed in the lower half of the ellipsoidal region. A relation similar to (8) can be drawn in order to decide on the ellipse axes length values.

A schematic diagram showing the processing steps of this stage is shown in Figure 1.

The robustness issues of the proposed technique is presented in Section 6.

3 Watermark Construction

In the previous section we developed a method of 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. We choose to construct a watermark based on a chaotic trajectory [12] because of its controlled lowpass properties compared to a usual pseudo-random sequence. The first step to construct such a watermark, is to produce a sequence of real numbers by using a mapping function $\mathbf{F} : U \to U, U \subset \mathbb{R}$ of the form:

$$z(n+1) = \mathbf{F}(z(n), \lambda), \qquad z(n) \in U, \lambda \in \mathbb{R} \tag{9}$$

where $n = 0, 1, 2, ...$ denotes the current iteration and λ is a parameter that controls the chaotic behaviour of the system. The number of iterations is arbitrary and can be adapted to our needs. The system theoreticaly produces trajectories of an infinite period. By changing the value of the parameter λ, the set of real numbers is divided in two subsets. The decision on whether the trajectory presents regular or chaotic behaviour 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 $z_{th} \in [z_{min}, z_{max}]$ 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, whereas when $\lambda \simeq 2$ 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 one-dimensional. To embed it 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. 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 point along the scan is topologically closer to the previous and next pixels than in the case of raster scan [13]. In addition, it is possible to use cellular smoothing to eliminate spontaneous transitions that emerged after the Peano scan [12]. Example of a watermark of size 128×128 constructed using raster scan order and cellular smoothing for $\lambda = 1.3$ is shown in Figure 2a. A watermark of the same size and for the same value of λ for a Peano scan followed by cellular smoothing is shown in Figure 2b. We can notice the lowpass characteristics of the watermark produced by Peano scan. By using this technique, the output watermark has local neighbourhoods 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 embedding and detection sections we shall show that this does not seriously affect the performance of our method.

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 value provide sufficiently uncorrelated trajectories, because the set $\mathbf{K}$ 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 2D watermark.

4 Watermark Embedding

In this stage we use the extracted salient feature set and ellipsoidal 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 is going to be embedded. 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:

$$K_1 = k(x_{(mouth)} - \bar{x}_{(eyes)}), \qquad K_2 = l(y_{(right\_eye)} - y_{(left\_eye)}) \tag{10}$$

where $x_{(\cdot)}$ and $y_{(\cdot)}$ are feature coordinates, $\bar{x}_{(\cdot)}$ 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 face region. Afterwards, we rotate the prototype watermark by the angle θ as defined in (4).

The scaled watermark is centered in the mass center of the feature points set $\mathcal{F} = \{F_i, i = 1, ..., M\}$:

$$(\bar{x}, \bar{y}) = \left(\frac{1}{M} \sum_{(x,y)} \sum_{\in \mathcal{F}} x, \frac{1}{M} \sum_{(x,y)} \sum_{\in \mathcal{F}} y \right) \tag{11}$$

In our case $M = 3$. The mass center is also the center of the A_{em} region. After centering the watermark in the proper image area, it is rotated by $-\theta$ with respect to the center of the image. It then covers a new area A_{rot}. Before superimposing it on the original image, a visual masking stage is introduced. In this stage the variance is computed for every point of the original image $f(x, y)$ over a proper neighbourhood $N \times N$:

$$Var(x,y) = \frac{1}{N^2-1} \cdot \sum_{i=-(N-1)/2}^{(N-1)/2} \sum_{j=-(N-1)/2}^{(N-1)/2} (f(x+i, y+j) - \mu(x,y))^2 \tag{12}$$

where $\mu(x,y)$ is the mean value over the same neighbourhood. The local variance is then normalized according to its maximum value, and is compared against a threshold T_{var}, which is a function of the watermark power h. If the variance exceeds this threshold, then the local neighbourhood contains a big amount of texture or edge information, and the watermark can be embedded without being visually perceptible. Otherwise, the region is considered to be close to uniform, like background, and is not suitable for watermark casting. The dependency of the variance threshold on the watermark power is such that, if the watermark power is increased, the threshold is also increased non-linearly, so that the watermark remains imperceptible. This means that a trade off should be made so that the watermark is still both perceptually insignificant and recoverable to a reasonable degree.

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

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

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

Alternatively h can become a function of the local variance:

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

where $s(\cdot)$ takes values in the range $[0,1]$. s is chosen to increase monotonically with the variance. In our case, the watermark is casted in the spatial domain and, thus, the watermark power is quantized to two integer values, 0 and h_{max}, depending on the value of the variance. The main steps of the embedding process are depicted in Figure 3.

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 feature set and orientation of the approximated face region are derived. The prototype watermark is again scaled, centered, and rotated according to the information obtained from the segmentation and feature extraction stage. The response of the correlation between the geometrically adapted watermark at the detection region A_{det} is given by:

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

where $\mathbf{A} = \{(x,y) \in A_{det} | \hat{w}_n(x,y) = 1\}$ and $\mathbf{B} = \{(x,y) \in A_{det} | \hat{w}_n(x,y) = -1\}$. $N_{\mathbf{A}}$ and $N_{\mathbf{B}}$ are the number of pixels of the sets $\mathbf{A}$ and $\mathbf{B}$ respectively. This detector expresses the difference of two sample means [16]:

$$\bar{\mathbf{w}} = \bar{\mathbf{x}} - \bar{\mathbf{y}} \tag{17}$$

where:

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

The mean value and variance of the detector output are:

$$\eta_{\bar{w}} = \eta_{\bar{x}} - \eta_{\bar{y}} \quad \sigma_{\bar{w}} = \left(\frac{1}{N_{\mathbf{A}}} + \frac{1}{N_{\mathbf{B}}}\right) \sigma_{\hat{f}_w} \tag{19}$$

In the case that the watermark is casted all over the embedding region, the detector output is assumed to follow a normal distribution. If the correct watermark is casted on the image, then the mean value is $\eta_{\bar{r}} = 2h$ and the variance is $\sigma_{\bar{r}}^2 = \left(\frac{1}{N_{\mathbf{A}}} + \frac{1}{N_{\mathbf{B}}}\right)(\sigma_f^2 + \sigma_w^2)$ where σ_f^2 is the variance of the initial image and σ_w^2 is the variance of the watermark. Otherwise, if there is no watermark present, the mean value of the detector is $\eta_{\bar{r}} = 0$ and the variance is $\sigma_{\bar{r}}^2 = \left(\frac{1}{N_{\mathbf{A}}} + \frac{1}{N_{\mathbf{B}}}\right)\sigma_f^2$ 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.

We choose not to use masking in the detection 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 presense or the absense of the watermark. According to the above mentioned analysis, this threshold should ideally be set to $R_{thr} = h$. A schematic of the detection stage is presented in Figure 4.

6 Experimental Results

In order to test the robustness of the face segmentation method to any kind of attack, we chose to simulate the possible distortions by Gaussian noise. This was done for all three channels of the initial RGB image. The output is of the form:

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

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 a face database and computed the moment-based features, that is, the mass center and orientation of the resulting ellipse before and after noise addition. Figures 5a and 5b show an original sample image and the same image after noise addition with $\sigma = 0.3$. In Figures 5c and 5(d) the corresponding face regions are shown. The derived ellipsoidal approximations are depicted in Figures 5e and 5(f). Finally, Figures 5g and 5h show the relative change in center position (in pixels) and in orientation (in rads) of the approximated ellipse, respectively. The mean values of center distortion and orientation distortion are 1.2974 pixels and 0.0205 rads, respectively. The results show that the method is quite robust even when significant distortion is present.

In our implementation of the segmentation method, the initial image is subsampled by a factor of 2 in both dimensions before thresholding. This helps eliminating isolated pixels that do not belong to the face region but their color is considered as skin-color because they satisfy the hue, saturation and value inequalities mentioned in Section 2 and, vice versa, pixels that were not considered as skin-colored ones, though they belong to the face region. 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 face image database [14]. The feature extraction success rate was 84% for combined eye and mouth detection. The algorithm can perform correctly only if all three features (the eyes and the mouth) are present. Figure 6 shows some detection examples. Figure 7 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 k and l parameters for defining the watermark spread over the face region are both fixed at value 1.25. The normalized threshold variance for masking was chosen $T_{var} = 0.002$. Figure 8 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 because their geometric models are robust enough.

For optimized detection performance we choose to cast the watermark in the entire image, by keeping the center of the embedding rectangle and its orientation, and afterwards scaling the watermark according to the largest distance of the rectangle center to each image boundary. However, only the aspect ratio of the embedding region is useful in this case, and the possible cropping should be coped with a search for the correct cropping factor. In Figure 9a a sample image after watermarking and feature extraction is shown. The prototype watermark size in this case was chosen to be 256×256 for reasons of increased performance.

For the same reasons, no visual masking was incorporated in the following experiments. Figure 9b shows the normal distribution of the detector output for 100 different watermarks searched in the unmarked original image (on the left) and in the watermarked image. The detector output is normalized for both cases by dividing the right part of (16) by $2h$, which is the maximum theoretical value of the detector. The left side distribution in Figure 9b has mean value approaching 0. The mean value for the right side distribution is approximately 1.02, which is close to 1.0, as expected according to the previous analysis. We can find a threshold that separates completely the two distributions. The threshold for this case is $R_{thr} = 0.52$. Figure 10 shows the result after JPEG compression of approximate ratio 1:30. We can see that, though the degradation is significant, the detection of the features is quite efficient and the normalized response of the correlator for the watermarked images is kept at a level of almost 0.49. A detection threshold e.g. $R_{thr} = 0.25$ separates the two distributions resulting to zero error detection. In Figure 11 the watermarked image has been modified by applying a median filter of size 3×3. The image is smoothened but the distortion affects the image in a uniform way. The detector output is again decreased, with the mean value of the distribution for the watermarked images falling to 0.53. We can already notice that the theoretically derived threshold of h (0.5 in the normalized case) cannot be used in practice due to several heavy distortions that may degrade the watermark power. A sufficient threshold for the above mentioned attack is approximately $R_{thr} = 0.28$. Figure 12 shows the result for a rotation of 12^0 around the center of the image. The watermark is recovered to a big extent after some local search, because only some points of the embedding region are interpolated, whereas the others remain intact. Thus, the mean value of the detector output pdf for the watermarked images is approximately 0.69. A detection threshold that completely separates the two distributions is $R_{thr} = 0.35$. Figure 13 displays a cropped version of the watermarked image, resulting in an image of size

188×248. The detector output mean value is approximately 1.03 for the signed images. This is explained by the fact that the correct position of the watermark is found after search and without any degradation present, but the variance has risen to a value that is almost triple to the original one because the content of the watermarked image has severely changed in total. A proper threshold that completely separates the two distributions in this case is $R_{thr} = 0.53$. Finally, in Figure 14 a scaled version of the watermarked image by a factor of 1.16 at each dimension is depicted. The produced image has points that are more densely interpolated than in the case of rotation, causing the empirical mean value for the watermarked images to fall to 0.54. The detection threshold should be set to $R_{thr} = 0.53$ to completely separate the two distributions.

For the class of images under consideration, the features are not expected to be localized at exactly the same positions in the watermaked and possible 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 ellipsoidal approximation. The detection is expected to give a peak for the correct height, width orientation and center of the originally embedded watermark, because of the high degree of sensitivity of the watermark to geometric operations. Response to a false watermark is not expected to give a significantly different peak for some false geometric parameters selection, because different watermarks are highly uncorrelated. Therefore, the detection response in the diagrams is always shown for the correct dimensions, orientation and center of the watermark. The method is expected to perform better for larger images than the ones used in our experiments.

In order to decide for an efficient threshold indicating watermark existense, we can follow an experimental approach. A big number of watermarks is casted and detected after an attack on both the original and the respective watermarked image. The mean value of the means of each pair of experimental distributions, corresponding to each attack as presented in the above referenced figures, represents an acceptable threshold for the current attack, considering that the distribution variance is almost the same for both normal approximations. The median value of the acceptable thresholds is considered as the common threshold for watermark detection. This was the approach that was followed in our experiments. Table 1 shows the false rejection rate (FRR) and the false acceptance rate (FAR) for each of the attack, considering a threshold of 0.32. From the above discussion we can notice that, in order to obtain both an acceptable false acceptance and false rejection ratio, we should define a decision threshold that is in general $R_{thr} < h$ (or, $R_{norm} < 0.5$ in the case of normalization).

7 Conclusions

In the present paper we developed a method for embedding and detecting watermarks in color frontal face images. Color information was exploited in order to obtain a good approximation of the skin-colored facial region, in which to search for salient features like the eyes and the mouth, using a geometric model matching method. The prototype watermark used for casting 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 face region orientation. A correlation detector is employed in order to decide about the presense of a possible watermark. The color segmentation and feature localization technique precedes both embedding and detection stages. A visual masking technique is added in order to avoid visible artifacts produced by the casted watermark. Experimental results display the robustness of the method and address the fact that the method is especially suitable for databases of large images.

Acknowledgements

The present work has been carried out within the framework of the EU LTR project INSPECT.

References

[1] N. Nikolaidis and I. Pitas, "Copyright Protection of Images using Robust Digital Signatures", in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP '96*, May 1996, Atlanta, Georgia, vol. 4, pp. 2168-2171.

[2] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding", *IBM Systems Journal*, vol. 25, pp. 313-335, 1996.

[3] I.J. Cox, J. Killian, T. Leighton and T. Shamoon, " Secure Spread Spectrum Watermarking for Multimedia", in *IEEE Trans. on Image Processing*, **6**(12), pp. 1673-1687.

[4] A. Piva, M. Barni, F. Bartolini and V. Capellini, "DCT-based watermark recovering without resorting to the uncorrupted original image", in *Proc. IEEE Int. Conf. on Image Processing (ICIP'97)*, October 1997, Santa Barbara, California, vol. 1, pp. 520-523, 1997.

[5] J. O'Ruanaidh, T. Pun, "Rotation, Scale and Translation Invariant Digital Image Watermarking", *Proceedings of ICIP '97*, October 1997, Santa Barbara, California, vol. 1, pp. 536-539.

[6] X.-G. Xia, C. G. Boncelet and G. R. Arce, "A Multiresolution Watermark for Digital Images", *Proceedings of ICIP '97*, October 1997, Santa Barbara, California, vol. 1, pp. 548-551.

[7] M.J.J.J.B. Maes and C.W.A.M. van Overveld, "Digital Watermarking by Geometric Warping", *Proceedings of ICIP '98*, October 1998, Chicago, Illinois, vol. 2, pp. 424-426.

[8] K. Sobottka and I. Pitas, “A novel method for automatic face segmentation, facial feature extraction and tracking”, *Signal Processing: Image Communication*, **12**(3), pp. 263-281.

[9] S. Tsekeridou and I. Pitas, “Facial feature extraction in frontal views using biometric analogies”, in *Proc. of EUSIPCO ’98*, September 1998, Rhodes, Greece, vol. 1, pp. 315-318.

[10] A.G. Bors and I. Pitas, “Object segmentation in 3-D images based on alpha-trimmed mean radial basis function network”, in *Proc. of EUSIPCO ’98*, September 1998, Rhodes, Greece, vol. 2, pp. 1093-1096.

[11] A.K. Jain, “Fundamentals of Digital Image Processing”, Prentice-Hall, New Jersey, 1989.

[12] G. Voyatzis and I. Pitas, “Chaotic Watermarks for Embedding in the Spatial Digital Image Domain”, in *Proc. of ICIP ’98*, October 1998, Chicago, Illinois, vol. 2, pp. 432-436.

[13] D. Coltuc and I. Pitas, “Memory Mappings for Fast Peano Scan”, in *Proc. of ECCT 1995 European Conference on Circuits Theory and Design*, August 1995, Istanbul, Turkey, pp. 817-821.

[14] 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.), **1206**, pp. 403-409, 1997.

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

[16] A. Papoulis, “Probability and Statistics”, Prentice Hall, 1990.

List of Tables

List of Figures

attack	FRR	FAR
no attack	$1.3094 \cdot 10^{-16}$	$2.1585 \cdot 10^{-4}$
1:30 JPEG compression	0.0273	$2.1015 \cdot 10^{-4}$
rotation by 12^0	$7.7982 \cdot 10^{-5}$	$9.6364 \cdot 10^{-4}$
scaling by 1.16	0.0085	$4.7922 \cdot 10^{-4}$
median (3×3)	0.0077	$3.1819 \cdot 10^{-4}$
cropping (188×248)	$2.9899 \cdot 10^{-5}$	0.049

Table 1: Detector rates for several attacks.

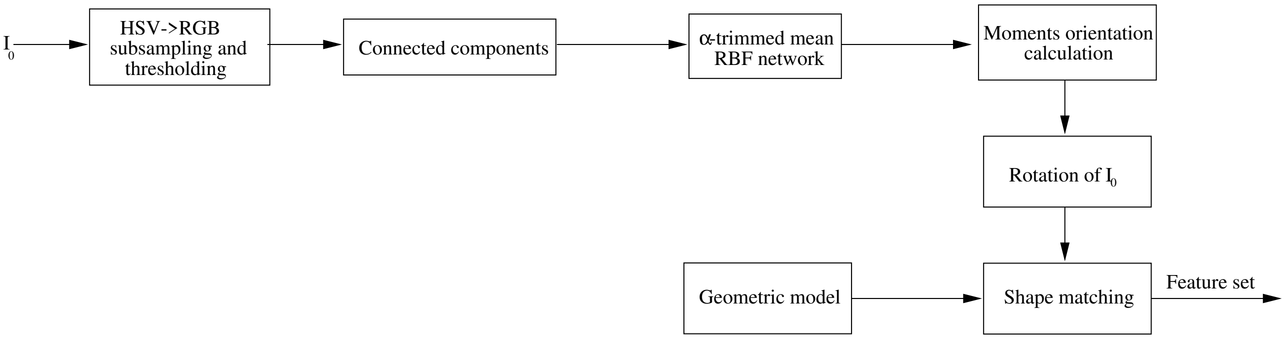


Figure 1: Face segmentation and salient feature localization.

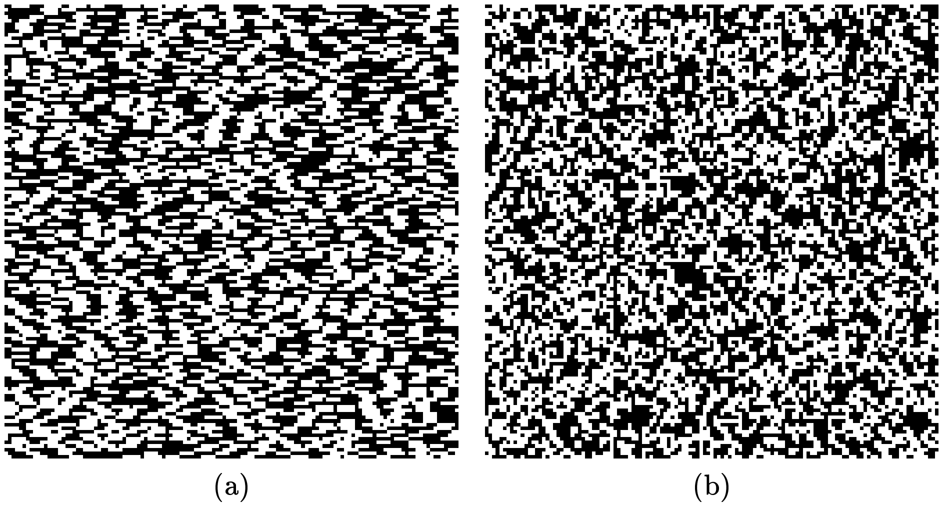


Figure 2: Watermarks produced by different scan order: (a) Raster scan. (b) Peano scan.

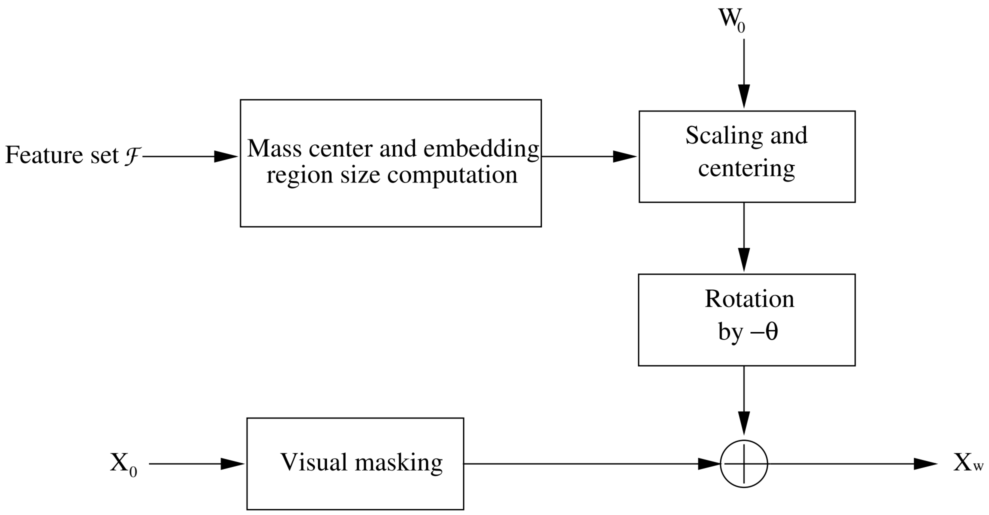


Figure 3: Watermark embedding stage.

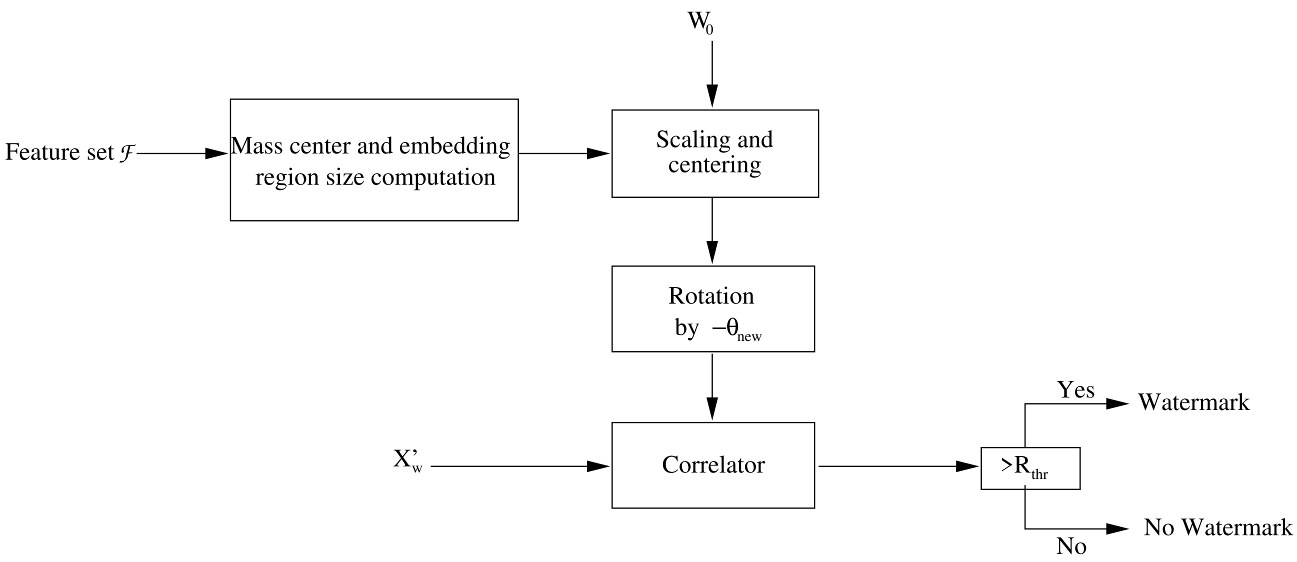


Figure 4: Watermark detection stage.

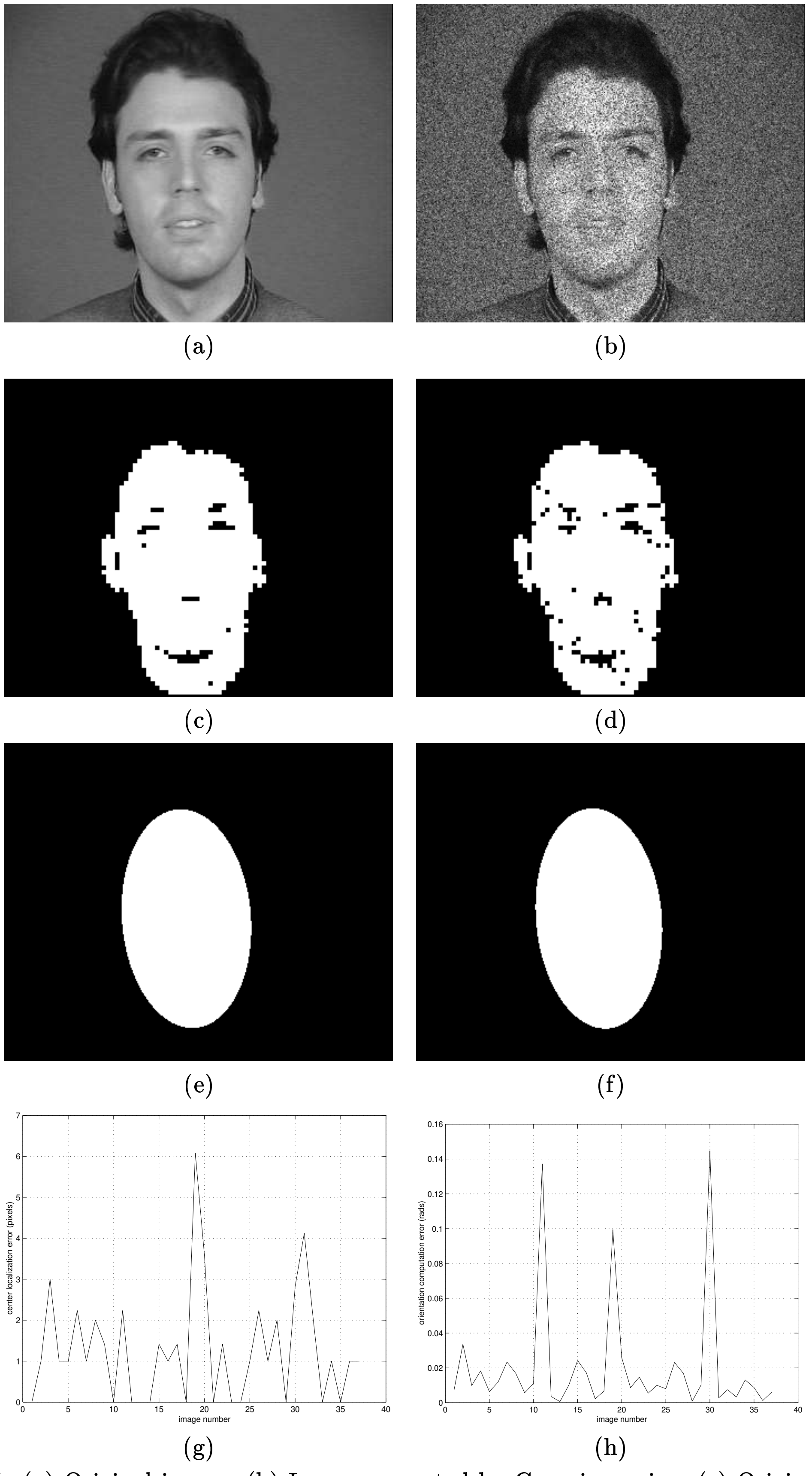
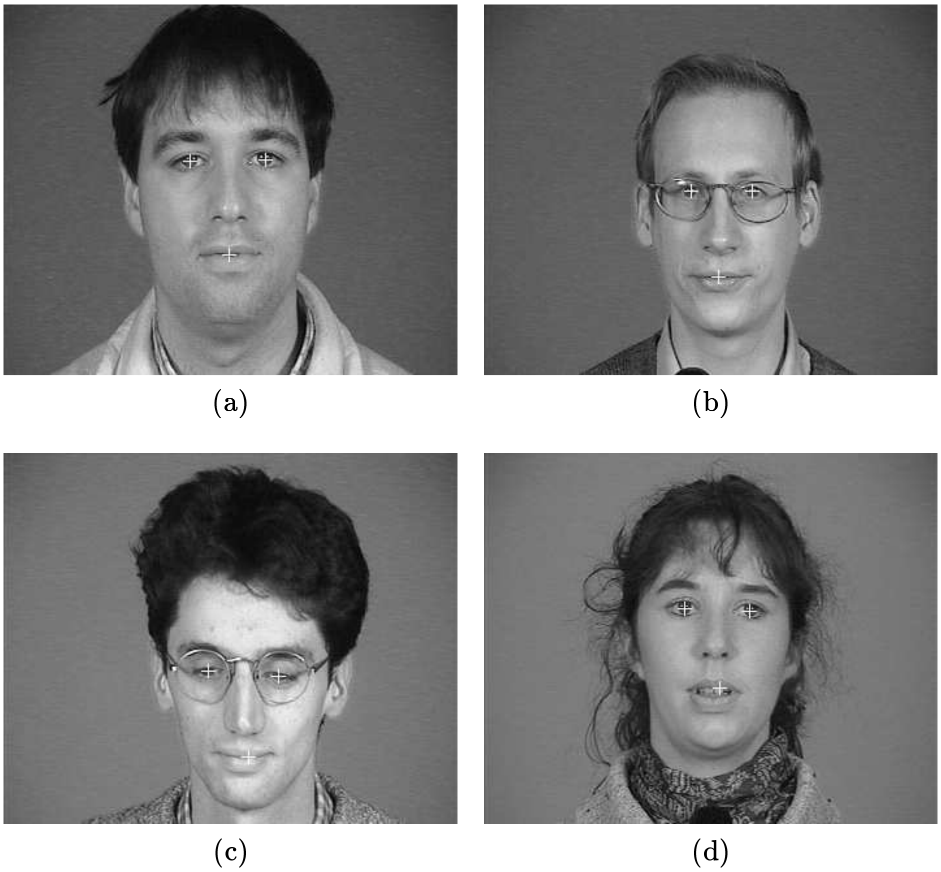
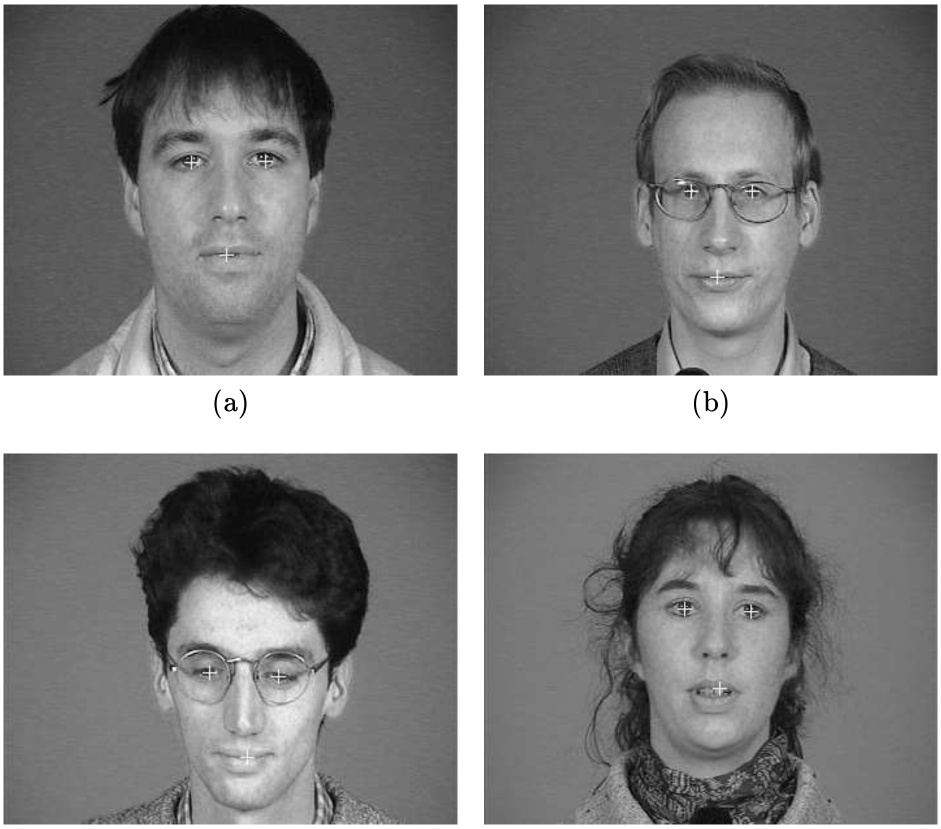


Figure 5: (a) Original image. (b) Image corrupted by Gaussian noise. (c) Original image face region. (d) Corrupted image face region. (e) Ellipsoidal approximation of original face region. (f) Ellipsoidal approximation of corrupted face region. (g) Center localization error. (h) Orientation computation error.



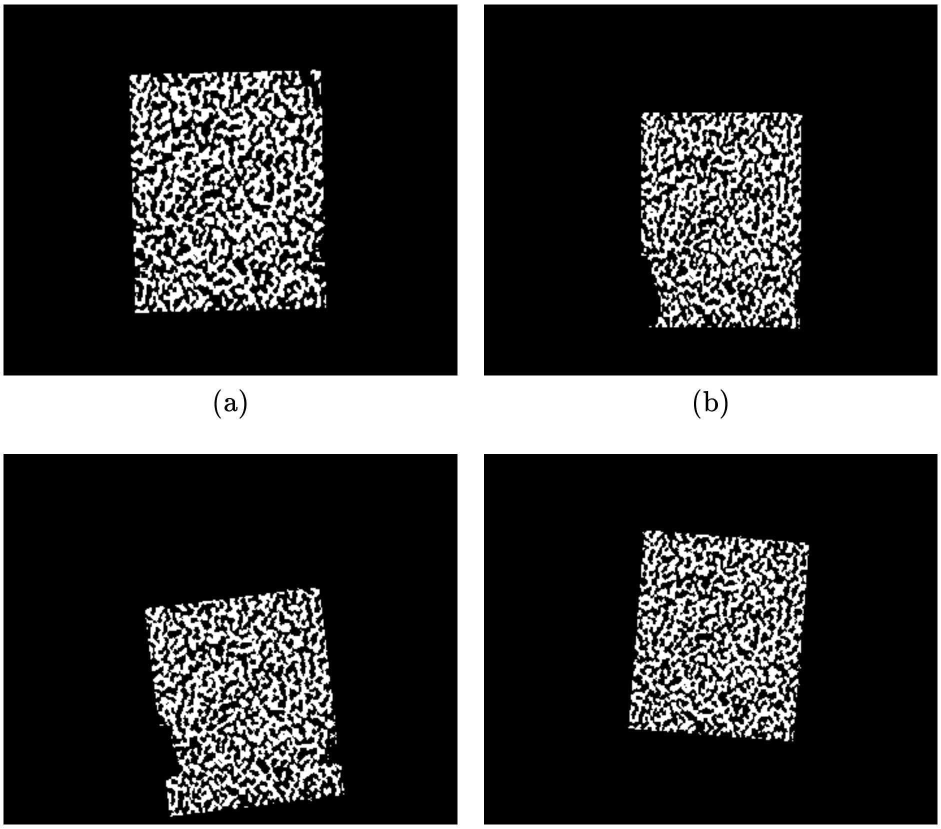
(a) (b) (c) (d)

Figure 6: Salient feature extraction results on original images.



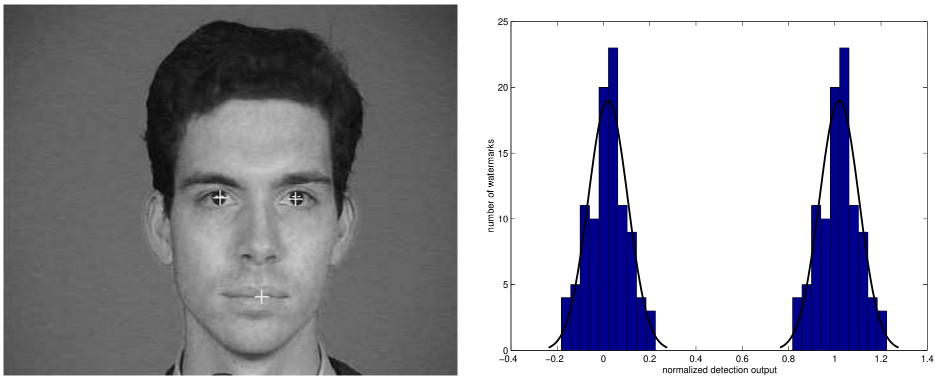
(a) (b) (c) (d)

Figure 7: Salient feature extraction results on watermarked images.



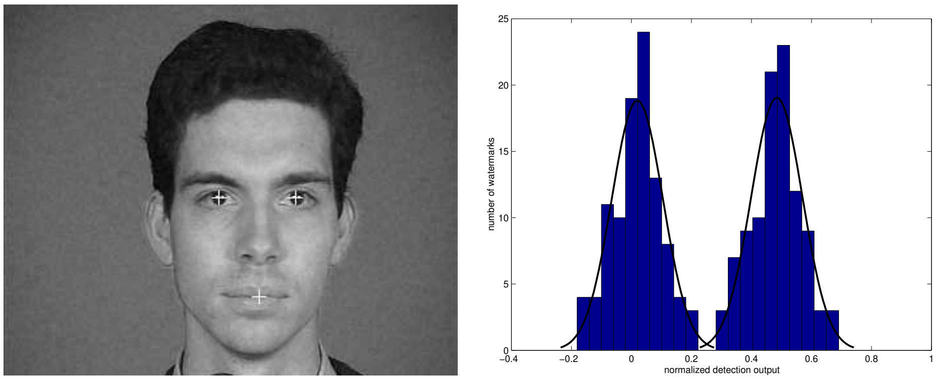
(a) (b) (c) (d)

Figure 8: Amplified difference of watermarked and original images.



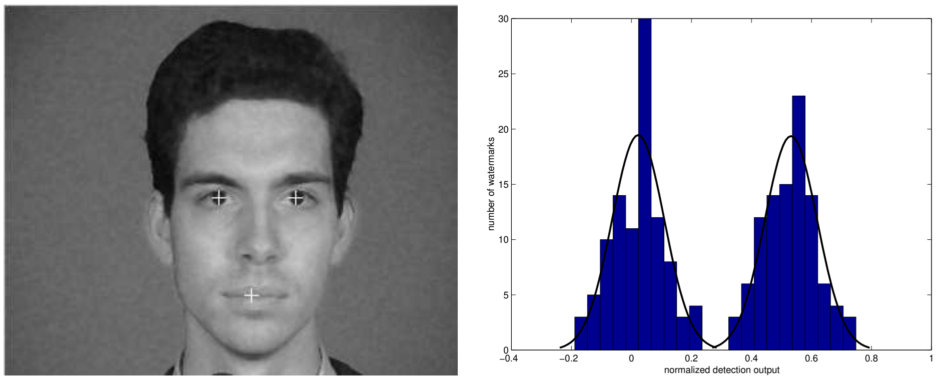
(a) (b)

Figure 9: (a) Watermarked image. (b) Normalized detection output.



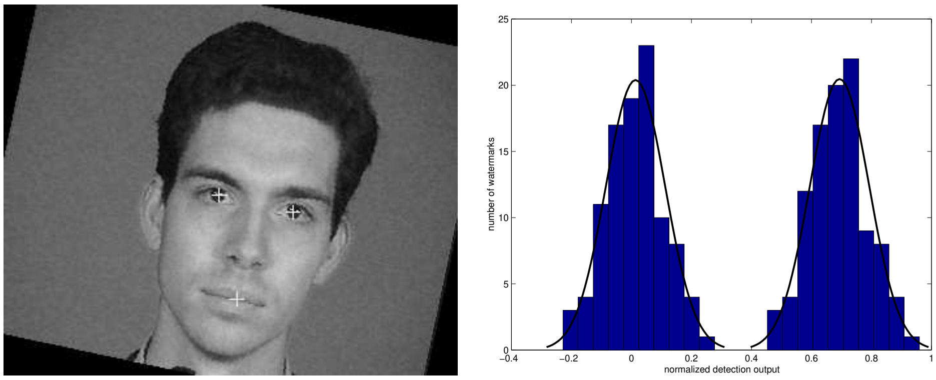
(a) (b)

Figure 10: (a) Watermarked image after 1:30 JPEG. (b) Normalized detection output.



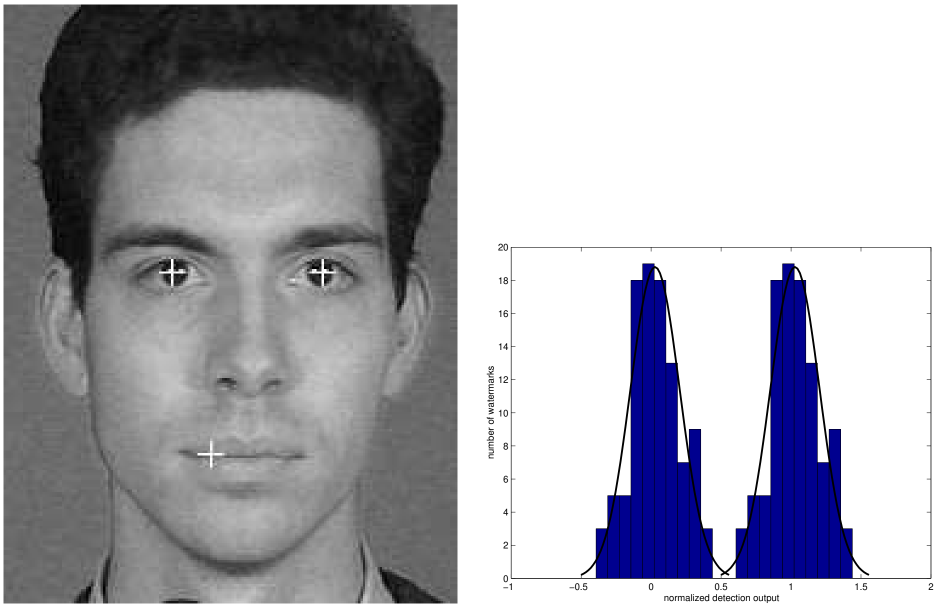
(a) (b)

Figure 11: (a) Watermarked image after 3×3 median. (b) Normalized detection output.



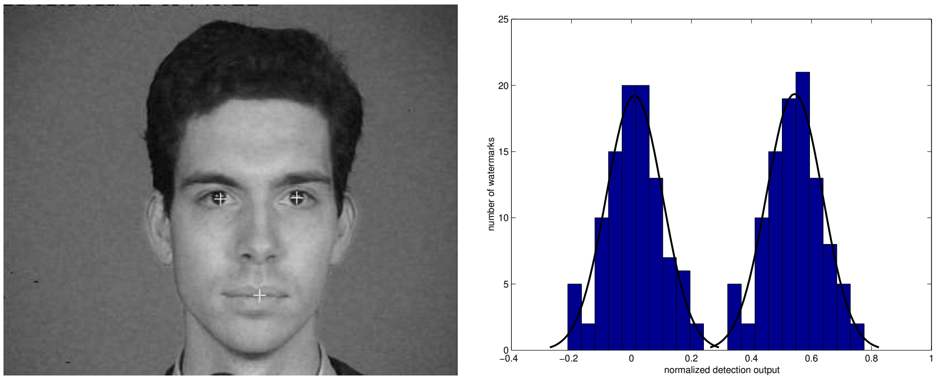
(a) (b)

Figure 12: (a) Watermarked image after rotation by 12^0. (b) Normalized detection output.



(a) (b)

Figure 13: (a) Watermarked image after cropping to size 188×248. (b) Normalized detection output.



(a) (b)

Figure 14: (a) Watermarked image after scaling to size 406 × 332. (b) Normalized detection output.