

Protection of Real and Artwork Human Objects based on a Chaotic Moments Modulation Method

Klimis Ntalianis¹, Paraskevi Tzouveli¹, Stefanos Kollias¹ and Athanasios Drigkas²

¹National Technical University of Athens, Electrical and Computer Engineering Department
9, Heroon Polytechniou str., Zografou 15773, Athens, Greece

²Net Media Lab, NCSR Demokritos, Athens, Greece
kntal@image.ntua.gr

Abstract: Content analysis technologies give more and more emphasis on multimedia semantics. However most watermarking systems are frame-oriented and do not focus on the protection of semantic regions. As a result, they fail to protect semantic content, especially in case of the copy-paste attack. In this framework, a novel unsupervised semantic region watermark encoding scheme is proposed. The proposed scheme is applied to real and artwork human objects, localized by two different face and body detection methods. Next, an invariant method is designed, based on Hu moments, for properly encoding the watermark information into each semantic region, using a chaotic pseudo-random number generator. Finally, experiments are carried out, to illustrate the advantages of the proposed scheme, such as: (a) robustness to RST, copy-paste and other attacks, and (b) low overhead transmission.

Keywords: Semantic region protection, artwork objects, Hu moments, Chaotic pseudo-random number generator.

1. Introduction

Copyright protection of digital images, video and artworks is still an urgent issue of ownership identification. Several watermarking techniques have been proposed in literature [1]-[11]. Among them some milestone methods include: (a) the technique of Cox et. al [9], who state that a watermark should be constructed as an i.i.d. Gaussian random vector and be imperceptibly inserted in a spread-spectrum-like fashion, into the perceptually most significant spectral components of the data. They report that the use of Gaussian noise ensures strong resilience to multiple-document, or collusion attacks. (b) In [10], a hybrid method is proposed, which combines two different watermark-embedding strategies for inserting information in the DCT coefficients of 8×8 blocks of the host video. (c) Additionally, in [11] quantization is used for watermark embedding in the “low frequencies” and spread spectrum watermarking is applied to the “high frequencies”, providing a way to maximize robustness to different types of attacks. Most of the abovementioned approaches rely on the insertion of pseudorandom noise into the original data. Generally, the resulting alterations do not change the essential properties of the data and cannot be perceived by the HVS.

However the majority of them are not resistant enough to geometric attacks, such as rotation, scaling, translation and shearing. Several researchers [12]-[18] have tried to overcome this inefficiency by designing watermarking techniques resistant to geometric attacks. Some of them are based on the invariant property of the Fourier transform.

Others use moment-based image normalization [19], with a standard size and orientation or other normalization techniques [20]. In most of the aforementioned techniques the watermark is a random sequence of bits [10] and it is retrieved by subtracting the original from the candidate image and choosing an experimental threshold value to determine when the cross-correlation coefficient denotes a watermarked image or not.

On the other hand, another deficiency of the majority of the existing techniques is that they are frame-based and thus semantic regions such as humans, buildings, cars etc., are not considered. These regions may need better protection or can be the only regions that need protection, depending on the specific application. Furthermore, typical watermark detection modules fail to extract watermark information in case of copying and pasting a semantic region (copy-paste attack), due to complete loss of synchronization. Even though a limited number of region watermarking schemes has also been proposed [21]-[24], the literature still lacks efficient algorithms for content authentication especially in case of the copy-paste attack.

At the same time, multimedia analysis technologies give more and more importance to semantic content and in several applications semantic regions, and especially humans, are addressed as independent video objects [25] and thus should be independently protected.

Towards this direction the proposed innovative system is specifically designed to provide geometrically resistant copyright protection of semantic content in two cases: in case of generic real world human objects and in case of artwork human objects, existing in Byzantine iconography. To achieve this goal, in these cases a human object detection module is required both in watermark encoding and during authentication. After object detection the watermark encoding phase is activated, where chaotic noise is properly generated and added to the detected human objects, producing the watermarked human objects. For authentication reasons, the watermark encoding procedure is guided by a feedback mechanism in order to satisfy a specific equality, formed as a weighted difference between Hu moments of the original and watermarked human objects. During authentication, initially every received image/artwork passes through the human object detection module. Then Hu moments are calculated for each detected human object, and a specific inequality is examined. A received human object is copyrighted only if the inequality is satisfied. Experimental

results on real sequences indicate the advantages of the proposed scheme in cases of mixed attacks, affine distortions and the highly innovative copy-paste attack.

The rest of this paper is organized as: in Section 2 we briefly describe Hu moment invariant functions. In Section 3 the human object extraction submodules are described. Next, in Section 4 the proposed watermark encoding module is discussed while Section 5 focuses on the watermark decoding module. Experimental results are presented in Section 6 to indicate the promising performance of the proposed system. Finally the paper is concluded in Section 7.

2. Moment Invariant Functions

Geometric moments and moment invariants are briefly presented in this section. Moments and functions of moments have been utilized as pattern features in a variety of applications [10], [19], [26], [27]. Geometric transformations have been based on a moments' constructive way for extracting features, which can provide global information about 2-D images. Moment invariants include: a) moments that are invariant under change of size, translation, and rotation only and b) moments that are invariant under all previous changes as well as reflection.

In this paper, Hu moments are used during the watermark encoding phase of semantic objects. Traditionally, moment invariants are computed based both on the shape boundary of the area and on its interior. Hu first introduced [20] the mathematical foundation of 2-D moment invariants, based on methods of algebraic invariants and demonstrated their application to shape recognition. Hu's method is based on nonlinear combinations of 2nd and 3rd order normalized central moments, providing a set of RST invariant functions.

Actually, Hu described two different methods for producing rotation invariant moments. The first, based on principal axes, might present problems when images do not have unique principal axes (rotationally symmetric). In the second method, Hu described the concept of absolute moment invariants, which are derived through algebraic invariants applied to the moment generating function, under a rotation transformation. The result is a set of absolute orthogonal moment invariants, which can be used for RST invariant pattern identification. From the 2nd and 3rd order central moments, a set of six absolute orthogonal invariant moments can be computed as:

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} \\
\phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2 \\
\phi_4 &= (\eta_{30} - \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \\
\phi_5 &= (3\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \cdot \left[\eta_{30} + \eta_{12} \right]^2 - 3(\eta_{21} + \eta_{03})^2 \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \cdot \left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \\
\phi_6 &= (\eta_{20} - \eta_{02}) \cdot \left[\eta_{30} + \eta_{12} \right]^2 - (\eta_{21} + \eta_{03})^2 \\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})
\end{aligned} \tag{1}$$

The 7th moment (skew orthogonal invariant) is useful for distinguishing mirror images:

$$\begin{aligned}
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \cdot \left[\eta_{30} + \eta_{12} \right]^2 - 3(\eta_{21} + \eta_{03})^2 \\
&\quad + (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03}) \cdot \left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right]
\end{aligned} \tag{2}$$

The first six of these moments are also invariant under reflection, while ϕ_7 changes sign. These seven moments (ϕ_1 - ϕ_7) are used by the proposed method for watermark encoding.

3. Semantic Region Extraction

A very important subsystem of the proposed system, which supports content authentication in case of the highly innovative copy-paste attack, is the human objects' extraction module. Since the overall system is designed to provide geometrically resistant copyright protection of both real world and Byzantine art objects, the human object extraction module consists of two submodules: (a) the real world human object extraction submodule and (b) the artwork object extraction submodule. The first submodule depends on chrominance and topology modeling of face and body through Gaussian p.d.fs, while the second is based on fundamental knowledge and essential rules for analyzing and interpreting Byzantine artworks. These rules are described in detail in the theoretical approach of Dionysios from Fourni [28], an expert in Byzantine art. In the following subsections both submodules are analytically presented.

A. The real world human object extraction submodule

In the proposed scheme we focus on semantic content authentication; in this framework, a region is defined through segmentation. In this paper, human objects are selected as target regions, since they constitute independent entities in applications and may provide semantic information about a shot or an image. In such applications, it is often important to carefully handle and effectively protect them. Other semantic objects can also be selected as target regions, such as buildings, vehicles, animals, etc.

Having selected the type of target region, a semantic segmentation algorithm should be incorporated for human object extraction. In this paper, initially the human face is localized and then the human body is detected using topological information based on the human face (Figure 1). Both modules are analytically described in the next paragraphs.

Human face detection is a topic of extensive research for several decades. Face detection methods can be classified as either feature or image based. Among feature-based methods, those using skin color have gained strong popularity. The advantages of skin color based methods are the fast processing and the significant robustness to geometric variations of face patterns. Due to these advantages, detection of human faces is accomplished in this paper, by combining key ideas of the feature invariant method proposed in [29], based on a Gaussian p.d.f. According to [29], the distribution of chrominance values of each block, belonging to a human face, occupies a very small region of the colorspace. Based on this idea, the blocks of an image that are located inside this small region can be considered as face blocks.

Let Ω_f denote the face class. Then the histogram of

chrominance values corresponding to the face class can be initially modelled by a Gaussian p.d.f. as:

$$P(\mathbf{x} | \Omega_f) = \frac{\exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_f)^T \cdot \frac{-1}{f} \sum (\mathbf{x} - \boldsymbol{\mu}_f))}{2\pi \cdot |\boldsymbol{\Sigma}|^{1/2}} \quad (3a)$$

where $\mathbf{x} = [u \ v]^T$ is a 2×1 vector containing the mean chrominance components u and v of an examined block, $\boldsymbol{\mu}_f$ is the 2×1 mean vector of a face class and $\boldsymbol{\Sigma}$ is the 2×2 variance matrix of the p.d.f.:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_u^2 & \sigma_{u,v} \\ \sigma_{u,v} & \sigma_v^2 \end{bmatrix} \quad (3b)$$

where σ_u^2 is the variance of the chrominance component u , σ_v^2 is the variance of the chrominance component v and $\sigma_{u,v}$ corresponds to the covariance between u and v . Parameters $\boldsymbol{\mu}_f$ and $\boldsymbol{\Sigma}$ are estimated, based on a set of several face images, through a maximum likelihood approach [30]. Next, each block B_i of an image is considered to belong to the face class, if the respective probability of its chrominance values, $P(\mathbf{x}(B_i) | \Omega_f)$ is high ($P(\mathbf{x}(B_i) | \Omega_f) > 0.9$). Then, by fusing those blocks belonging to face class Ω_f , a binary mask M is produced, containing candidate face regions.

However, mask M may also contain non-face blocks that present similar chrominance characteristics (like hands, legs or other parts of the human body). To confront this problem, shape information of human faces is also considered, by using rectangles with certain aspect ratios [31]. In particular, the aspect ratio of face areas can be defined as $R = H_f / W_f$ where H_f is the height of the head, while W_f corresponds to the face width. According to this approach, R was experimentally found to lie within the interval [1.4 1.6]; consequently regions with aspect ratios within this interval are considered as face regions, while the rest are discarded. After checking all candidate face areas and discarding those that do not satisfy the aspect ratio rule, a final binary mask, say M_f , is built that contains only face areas.

Detection of the body area can be achieved using topological attributes that relate the locations of face and body. Initially the centre, width and height of the estimated face region, denoted as $c_f = [c_x \ c_y]^T$, w_f and h_f respectively, are computed. Human body is then localized by means of a probabilistic model, the parameters of which are estimated according to c_f , w_f and h_f .

In particular, if $r(B_i) = [r_x(B_i) \ r_y(B_i)]^T$ is the distance between the i -th block, B_i , and the origin, with $r_x(B_i)$ and $r_y(B_i)$ the respective x and y coordinates, the product of two independent 1-dimensional Gaussian p.d.f.s is used to model the location of human body. Thus, for each block B_i of an image, a probability $P(r(B_i) | \Omega_b)$ is assigned, expressing the degree of block B_i belonging to the human body class, say Ω_b

$$P(\mathbf{r}(B_i) | \Omega_b) = \frac{\exp(-\frac{1}{2\sigma_x^2}(r_x(B_i) - \mu_x)^2) \exp(-\frac{1}{2\sigma_y^2}(r_y(B_i) - \mu_y)^2)}{(2\pi)\sigma_x\sigma_y} \quad (4)$$

where μ_x , μ_y , σ_x and σ_y are the parameters of the human body localization model; these parameters are calculated based on the information derived from the face detection task, taking

into account the relationship between human face and body. In our simulations, the parameters of the human body localization model are estimated with respect to the face region as follows [32]:

$$\mu_x = c_x, \quad \mu_y = c_y + h_f, \quad \sigma_x = w_f, \quad \sigma_y = h_f/2 \quad (5)$$

Similarly to human face detection, a block B_i belongs to the body class Ω_b , if the respective probability, $P(r(B_i) | \Omega_b)$, is high, using a similar threshold as in the face detection case. The computed face and body masks can be properly used to extract human objects [32].

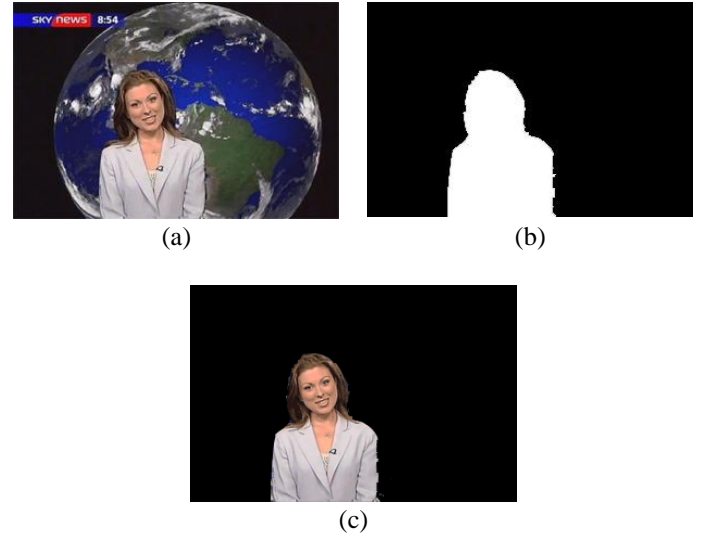


Fig. 1: Human Video Object Extraction Method: a) Initial Image b) Object Mask c) Object Extraction

This algorithm is an efficient method for finding face locations in complex backgrounds, when the size of faces is unknown. It can be used for a wide range of face sizes. The performance of the algorithm is based on the distribution of chrominance values corresponding to human faces, providing 92% segmentation success.

B. The artwork figure extraction submodule

Painters of Byzantine artworks follow the specific instructions that Dionysios from Fournia had recorded for painting holy figures. According to these instructions, initially a painter separates the painting area into seven semantic segments of equal size, each of which has specific characteristics that can make the figure distinguishable. An example is presented in Fig. 2, where the standing holy figure of Jesus Christ is presented. Starting from top to bottom, the first segment contains the head of the holy figure which is further separated into 4 equal smaller semantic parts

$$P = 4 \cdot H \quad (6)$$

The second segment, also with same height P , contains the part from neck to thorax while the third segment contains the part from thorax to elbow and waist, which always lay at the same height. Next, in the fourth segment the abdominal area is usually depicted, while the fifth segment contains the area from legs until the knees. The sixth segment contains from

ankle to foot and finally the feet of the figure are located at the seventh segment.

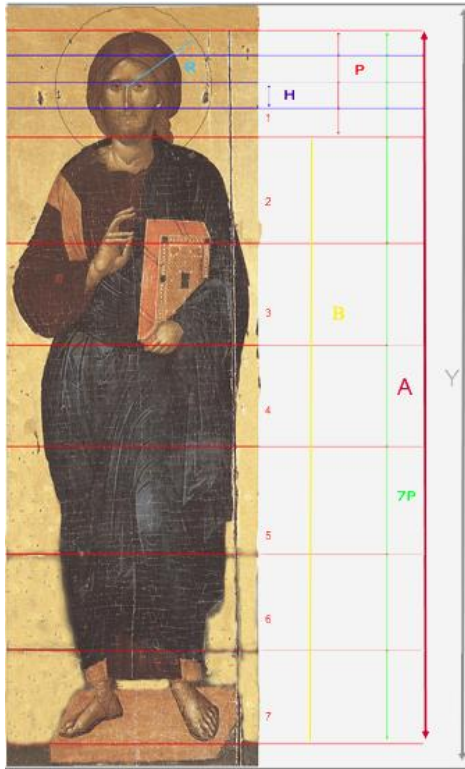


Fig. 2: Metric rules of Byzantine iconography

According to Dionysios from Fournia, the head of a saint should be surrounded by a halo, a circle that signifies the Holy Spirit (see Fig. 2). In order to paint the halo, the painter draws a circle, centered at the middle point of the nose with a radius $R = 2.5 \cdot H$ or according to (6):

$$P = 1.6 \cdot R \quad (7)$$

For the halo identification, we use the Hough transform, [33] by following the next steps:

1. Quantization of the parameter space with regard to the parameters a and b of the Hough transform.
2. Assign an accumulator to each cell in the parameter space and initialize all accumulators $M(a, b)$ to zero.
3. Compute the gradient direction $\theta(x, y)$ and magnitude $G(x, y)$ for all the edge points in the image.
4. For each edge point $G(x, y)$ increment all points in the accumulator array $M(a, b)$ along the line:
 $b = a \cdot \tan\theta - x \cdot \tan\theta + y$
5. Find the local maxima in the accumulator array and determine the center of the circles.

After the halo circle has been identified, we estimate two thresholds by drawing two concentric circles inscribed into the halo. From the small circle we estimate the average intensity value for the actual head and from the ring between the greater circle and the halo we estimate the intensity value for the halo. By choosing a threshold between these two values, we segment the area in the ring between the two circles in two regions, head and halo.

Finally, we apply a median filter with appropriate size to the segmented image in order to produce masks that better isolate the extracted head area. Application of these masks to

the original images produces the final images, which contain the extracted heads.

The pictures in Fig. 3(a) show, from left to right first, the stages for the head extraction for one image (extraction from halo location, threshold calculation, median filtering, and final extraction) and in Fig. 3(b) the extracted heads for nine such images are shown.

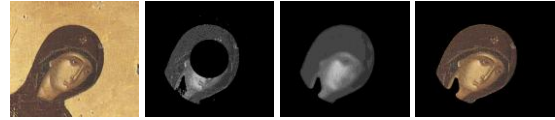


Fig. 3(a): Head extraction for one Holy figure

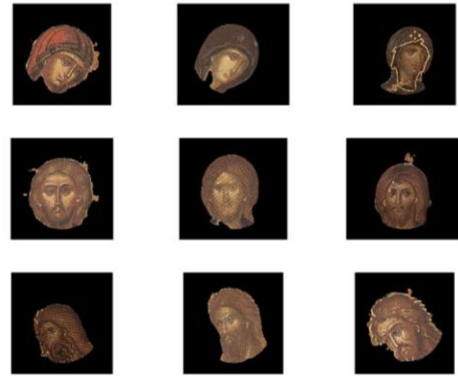


Fig. 3(b): Extracted heads of several Holy figures

Then, the height of the image Y is given according to Dionysios' manual as:

$$Y = A + \frac{P}{2} \rightarrow A = Y - \frac{P}{2} \quad (8)$$

where A is the area in which the holy figure is illustrated and P the height of each semantic part (see Fig.2).

Then the body area of the figure is given by

$$B = A - P \Rightarrow B = Y - 2.4 \cdot R \quad (9)$$

Having the values of the homocentric circles inscribed into the main halo circle, the intensity values of the background of the image is known. So, the extraction of the body area is achieved by cropping the proper area with height B and extract the background part as it is depicted in Fig. 4.



Fig. 4: Extracted Holy figures

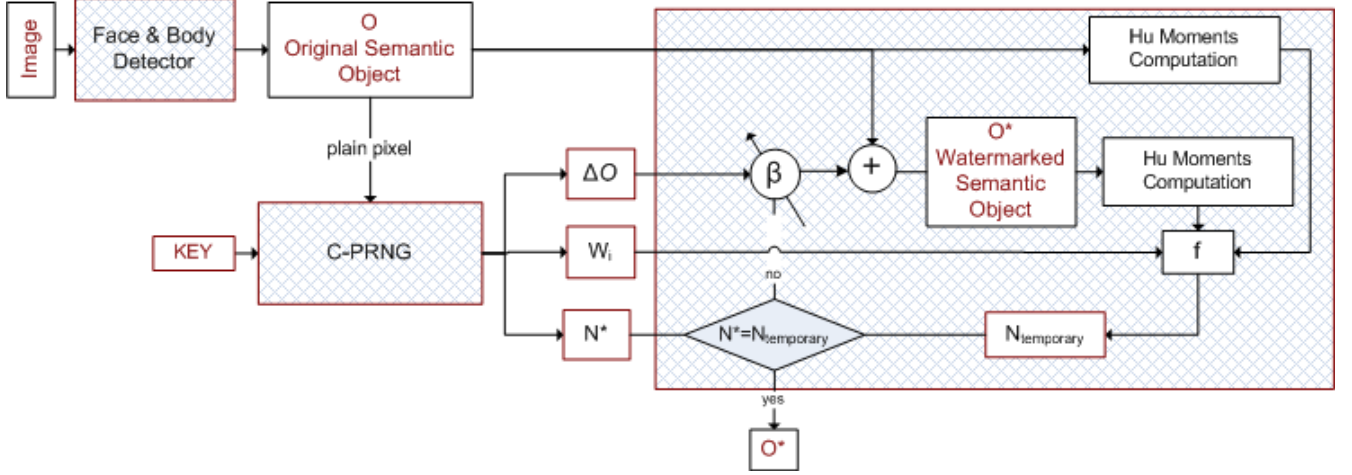


Fig 5: Block diagram of the encoding module

4. The Watermark Encoding Module

Let us assume that human object O has been extracted from an image or frame, using the object extraction modules described in Section 3. Initially, Hu moments of human object O are computed [20], providing an invariant feature of an object. Traditionally, moment invariants are computed based both on the shape boundary of the area and its interior object. Hu first introduced the mathematical foundation of 2-D moment invariants, based on methods of algebraic invariants and demonstrated their application to shape recognition. Hu's method is based on nonlinear combinations of 2nd and 3rd order normalized central moments, providing a set of absolute orthogonal moment invariants, which can be used for RST invariant pattern identification. Hu derived seven functions from regular moments, which are rotation, scaling and translation invariant. In [32], Hu's moment invariant functions are incorporated and the watermark is embedded by modifying the moment values of the image. In this implementation, exhaustive search should be performed

in order to determine the embedding strength. The method that is proposed in [32] provides an invariant watermark in both geometric and signal processing attacks based on invariant of moments.

Hu moments are seven invariant values computed from central moments through order three, and are independent of object translation, scale and orientation. Let $\Phi = [\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7]^T$ be a vector containing the Hu moments of O . In this paper, the watermark information is encoded into the invariant moments of the original human object. To accomplish this, let us define the following function:

$$f(X, \Phi) = \sum_{i=1}^7 w_i \left(\frac{x_i - \phi_i}{\phi_i} \right) \quad (10)$$

where X is a vector containing the ϕ values of an object, Φ contains the ϕ invariants of object O and w_i are weights that put different emphasis to different invariants.

Each of the weights w_i receives a value within a specific interval, based on the output of a chaotic random number generator. In particular chaotic functions, first studied in the 1960's, present numerous interesting properties that can be used by modern cryptographic and watermarking schemes. For example the iterative values generated from such

functions are completely random in nature, although they are limited between some bounds. The iterative values are never seen to converge after any number of iterations. However the most fascinating aspect of these functions is their extreme sensitivity to initial conditions that make chaotic functions very important for applications in cryptography. One of the simplest chaotic functions that are incorporated in our work is the logistic map. In particular, the logistic function is incorporated, as core component, in a chaotic pseudo-random number generator (C-PRNG) [34].

The procedure is triggered and guided by a secret 256-bit key that is split into 32 8-bit session keys (k_0, k_1, \dots, k_{31}). Two successive session keys k_n and k_{n+1} are used to regulate the initial conditions of the chaotic map in each iteration. The robustness of the system is further reinforced by a feedback mechanism, which leads to acyclic behavior, so that the next value to be produced depends on the key and the current value. In particular the first 7 output values of C-PRNG are linearly mapped to the following intervals: [1.5 1.75] for w_1 , [1.25 1.5] for w_2 , [1 1.25] for w_3 , [0.75 1] for w_4 and w_5 , and

[0.5 0.75] for w_6 and w_7 . These intervals have been experimentally estimated based on the importance and robustness of each of the ϕ invariants. Then watermark encoding is achieved by enforcing the following condition:

$$f(\Phi^*, \Phi) = \sum_{i=1}^7 w_i \left(\frac{\phi_i^* - \phi_i}{\phi_i} \right) = N^* \quad (11)$$

where Φ^* is the moments vector of the watermarked human object O^* and N^* is a target value also properly determined by the C-PRNG, taking into consideration a tolerable content distortion. N^* value expresses the weighted difference among the ϕ invariants of the original and the watermarked human objects. The greater the value is, the larger perturbation should be added to the original video object and the higher visual distortion would be introduced. This is achieved by generating a perturbation region ΔO of the same size as O such that, when ΔO is added to the original human object O , it produces a region

$$O^* = O + \beta \cdot \Delta O \quad (12)$$

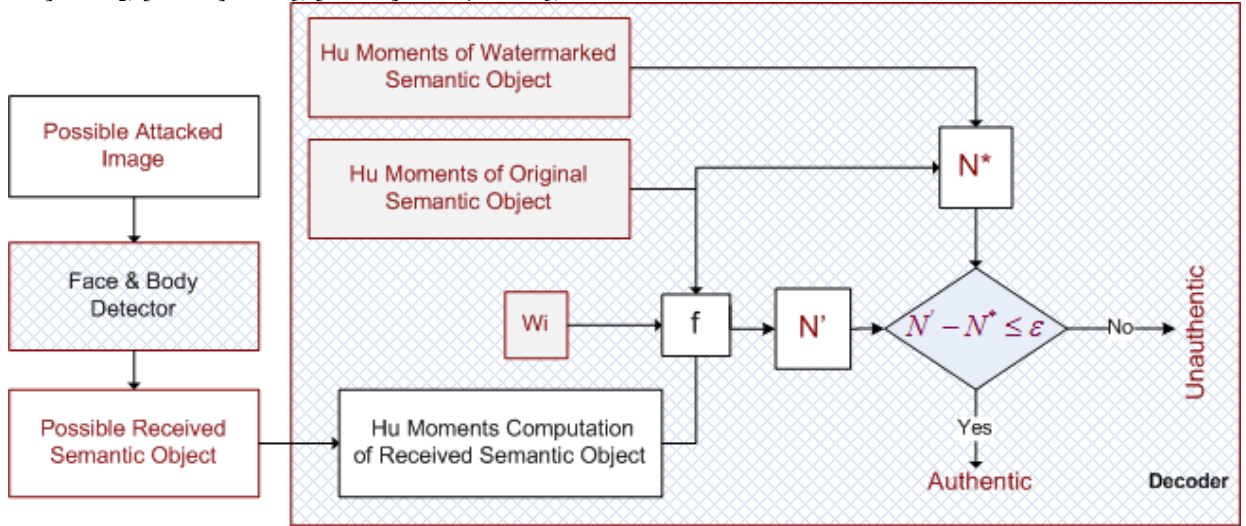


Fig 6: Block Diagram of the decoding module

that satisfies Eq. (11). Here, β is a parameter that controls the distortion introduced to O by ΔO . C-PRNG generates values until mask ΔO is fully filled. After generating all sensitive parameters of the watermark encoding module, a proper O^* is iteratively produced using Eqs. (11) and (12). In this way, the watermark information is encoded into the ϕ values of O producing O^* . An overview of the proposed watermark encoding module is presented in Fig. 5.

5. The Decoding Module

The decoding module is responsible for detecting copyrighted human objects. The decoding procedure is split into two phases (Fig. 6). During the first phase, the received image passes through the human object extraction module described in Section 3. During the second phase each human object undergoes an authentication test to check whether it is copyrighted or not.

In particular let us consider the following sets of objects and respective ϕ invariants: (a) (O, Φ) for the original human object, (b) (O^*, Φ^*) for the watermarked human object and

(c) (O', Φ') for a candidate copyrighted human object. Then O' is declared authentic if:

$$\left| f(\Phi^*, \Phi) - f(\Phi', \Phi) \right| \leq \varepsilon \quad (13)$$

where $f(\Phi^*, \Phi)$ is given by Eq.(15), while $f(\Phi', \Phi)$ is given by:

$$f(\Phi', \Phi) = \sum_{i=1}^7 w_i \left(\frac{\phi'_i - \phi_i}{\phi_i} \right) = N' \quad (14)$$

Then Eq. (13) becomes

$$N_d = |N^* - N'| \leq \varepsilon \Rightarrow \left| \sum_{i=1}^7 w_i \left(\frac{\phi_i^* - \phi'_i}{\phi_i} \right) \right| \leq \varepsilon \quad (15)$$

where ε is an experimentally determined, case-specific margin of error and w_i are the weights.

Two observations need to be stressed at this point. It is advantageous that the decoder does not need the original image. It only needs w_i , Φ , Φ^* and the margin of error ε . Secondly, since the decoder only checks the validity of Eq. (15) for the received human object, the resulting watermarking scheme answers a yes/no, (i.e. copyrighted or not) question. As a consequence, this watermarking scheme belongs to the family of algorithms of 1-bit capacity.

Now in order to determine ε , we should first observe that Eq. (15), delimits a normalized margin of error between Φ and Φ^* . This margin depends on the severity of the attack, i.e., the more severe the attack, the larger the value of N_d will be. Thus, its value should be properly selected so as to keep false reject and false accept rates as low as possible (ideally zero). More specifically, the value of ε is not heuristically set, but depends on the content of each distinct human object. In particular, each watermarked human object, O^* , undergoes a sequence of plain (e.g. compression, filtering etc.) and mixed attacks (e.g. cropping and filtering, noise addition and compression) of increasing strength. The strength of the attack increases until, either the SNR falls below a predetermined value, or a subjective criterion is satisfied.

In the following the subjective criterion is selected, which is related to the content's visual quality. According to this criterion and for each attack, when the quality of the human object's content is considered unacceptable for the majority of evaluators, an upper level of attack, say A_h , is set. This upper level of attack can also be automatically determined based on SNR, since a minimum value of SNR can be defined before any attack is performed. Let us now define an operator $p(\cdot)$ that performs attack i to O^* (reaching upper level A_{h_i}) and producing an object O_i^* :

$$p(O^*, A_{h_i}) = O_i^*, i = 1, 2, \dots, M \quad (16)$$

Then for each O_i^* , N_{di} is calculated according to Eq. (15). By gathering N_{di} values, a vector is produced:

$$\vec{N}_d = [N_{d_1}, N_{d_2}, \dots, N_{d_m}] \quad (17)$$

Then the margin of error is determined as:

$$\varepsilon = \max \vec{N}_d \quad (18)$$

Since ε is the maximum value of \vec{N}_d , it is guaranteed that human objects should be visually unacceptable in order to deceive the watermark decoder.

6. Experimental Results

Several experiments were performed to examine the advantages and open issues of the proposed method. Firstly, face and body detection was performed on different images, both real world and artistic. The following experimental results concern the real world object of Figure 1(c) and the middle artwork holy figure of Figure 4. After objects' extraction, the watermark was encoded to each object and the decoding module was tested under a wide class of geometric distortions, copy-paste and mixed attacks. When an attack of specific type was performed to each one of the watermarked

objects (real world and artwork), it led to SNR reduction that was proportional to the severity of the attack.

Firstly we examined JPEG compression for different quality factors in the range of 10 to 90. Result sets (N^* , SNR, N_d) are provided in the first group of rows of Table I. It can be observed that N_d changes rapidly for SNR < 9.6 dB. Furthermore, the subjective visual quality is not acceptable for SNR < 10 dB for both categories of human objects. Similar behaviors can be observed in the cases of Gaussian noise for SNR < 11 dB (using different means and deviations) and median filtering for SNR < 10 dB (changing the filter size). By summarizing the results in Table I, it can be observed that in most cases the proposed system can successfully authenticate watermarked content.






Fig 7: Copy-paste attack. (a) Watermarked human object (b) Modified watermarked human object in new content

In the following, we also illustrate the ability of the method to protect watermarked content in case of the very innovative and widespread copy-paste attack. The encoding module receives an image which contains a weather forecaster and provides the watermarked human object (Fig 7a). In this case ε was automatically set equal to 0.65 according to Eq. (18), so as to confront even cropping inaccuracy of 6 %. It should be mentioned that, for larger ε , larger cropping inaccuracies can be addressed, however, the possibility of false alarms also increases.

Now let us assume that a malicious user initially receives Fig. (7a) and then copies, modifies (cropping inaccuracy of 2%, scaling 5%, rotation 10°) and pastes the watermarked human object in a new content (Fig. 7b). Let us also assume that the decoding module receives Fig. (7b). Initially the human object is extracted and then the decoder checks the validity of Eq. (18). In this case $N_d=0.096$, a value that is smaller than ε . As a result the watermark decoder certifies that the human object of Fig. 7b is copyrighted, even though it is inserted to a completely new background. Results for different percentages of cropping inaccuracy are presented in the last row of Table I, both for the real and the artwork objects.

As mentioned, a crucial issue for the determination of ε has to do with the accuracy of cropping. In particular, let us assume that a malicious user crops the watermarked holy figure object of Table II, in order to reuse it. We examine three different ways that the malicious user can incorporate in order to extract the holy figure from the rest of the artwork (Figure 4a). In the first case the malicious user applies rectangular cropping, while in the second and third he uses lasso cropping with different accuracy. The simulation of these attacks and results are presented in Table II. As it can be observed the cropping attack has a larger impact on the value of N_d , compared to the rest of the attacks of Table I, and for this reason selection of ε is mainly based on these outcomes.

Table II: Cropping and holy figure object authentication

| Cropped Object |  Rectangle area |  Lasso technique |  Object extraction |
|----------------|-----------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| SNR | 8.7 | 4.68 | 10,45 |
| N_d | 1.9862 | 1.2033 | 0.2877 |

7. Conclusions

The latest multimedia systems and technologies give more and more emphasis on semantic regions, their detection, analysis, recognition and protection. However most of the existing watermarking schemes are frame-based and do not independently protect semantic regions. These regions within a frame may need better protection, compared to the rest, semantically indifferent, content or can be the only regions

that need protection. Currently, typical watermark detection modules fail to authenticate semantic regions, due to complete loss of synchronization. They are only able to authenticate a frame as a whole. Thus the copy-paste attack is not addressed.

In this paper we have proposed an unsupervised, robust to geometric attacks and low complexity semantic objects watermarking scheme. Two cases have been studied: the case of generic real world human objects and the case of Byzantine art objects. For the first case initially human objects are extracted, using skin-tone color and shape and topology constraints that are built into Gaussian probabilistic models. For the second case, fundamental knowledge and essential rules from the handbook of Dionysios from Fournai are incorporated, for analyzing and interpreting Byzantine artworks. Next, a watermark is encoded to each human object by properly modifying its Hu moments. Finally during authentication, initially the human objects' extraction module is incorporated and then authentication is performed on the detected regions.

Table I: Experimental results for the real world human object of Figure 1(c) and the middle artwork holy figure of Figure 4.

| | N* | 0,0193 | | | | | | | | | |
|------------------|-------------------------|---------------|---------------|---------------|---------------|---------------|----------------------|---------------|---------------|---------------|---------------|
| | Real world human object | | | | | | Artwork human object | | | | |
| jpeg Compression | Quality | 10 | 30 | 50 | 70 | 90 | 10 | 30 | 50 | 70 | 90 |
| | SNR | 2,80 | 9,63 | 12,99 | 14,37 | 15,36 | 5,30 | 10,24 | 13,56 | 15,01 | 17,34 |
| | N_d | 0,1620 | 0,0141 | 0,0110 | 0,0057 | 0,0076 | 0,1840 | 0,0162 | 0,0105 | 0,0090 | 0,0079 |
| Gaussian Noise | v=0, σ | 1,80 | 1,40 | 1,00 | 0,06 | 0,02 | 1,80 | 1,40 | 1,00 | 0,06 | 0,02 |
| | SNR | 9,49 | 11,44 | 13,90 | 17,41 | 23,17 | 10,20 | 12,64 | 15,14 | 19,44 | 28,98 |
| | N_d | 0,2340 | 0,0109 | 0,0065 | 0,0030 | 0,0025 | 0,0174 | 0,0099 | 0,0035 | 0,0027 | 0,0004 |
| Median Filtering | [n×n] | 11 | 9 | 7 | 5 | 3 | 11 | 9 | 7 | 5 | 3 |
| | SNR | 9,49 | 10,92 | 12,27 | 12,93 | 20,68 | 10,85 | 11,19 | 12,81 | 15,34 | 24,63 |
| | N_d | 0,1166 | 0,1026 | 0,0970 | 0,0117 | 0,0032 | 0,1299 | 0,1124 | 0,0991 | 0,0043 | 0,0043 |
| Rotation | Degrees | 180 | 90 | 60 | 40 | 20 | 180 | 90 | 60 | 40 | 20 |
| | SNR | 3,13 | 3,29 | 3,96 | 4,78 | 6,30 | 3,49 | 3,63 | 4,78 | 5,56 | 8,14 |
| | N_d | 0,1490 | 0,1489 | 0,1487 | 0,1483 | 0,1480 | 0,1492 | 0,1487 | 0,1481 | 0,1480 | 0,1477 |
| Scaling | % | 0,20 | 0,60 | 1,00 | 1,40 | 1,80 | 0,20 | 0,60 | 1,00 | 1,40 | 1,80 |
| | SNR | 3,16 | 3,48 | 4,00 | 4,68 | 6,32 | 3,22 | 3,67 | 4,35 | 5,56 | 8,43 |
| | N_d | 0,0690 | 0,0700 | 0,0700 | 0,0710 | 0,0690 | 0,0580 | 0,0582 | 0,0583 | 0,0582 | 0,0586 |
| Free cropping | % | 1 | 3 | 5 | 7 | 9 | 1 | 3 | 5 | 7 | 9 |
| | N_d | 0,3840 | 0,4560 | 0,5090 | 0,6020 | 0,6602 | 0,3503 | 0,4956 | 0,5869 | 0,6923 | 0,6989 |

Here it should be mentioned that the authentication module only uses the moment values of the original and watermarked human object and the moment weights. For these reasons, both the encoding and decoding modules have low complexity.

Experimental results on both real sequences and Byzantine artworks indicate the robustness of the proposed watermarking method under various signal distortions, mixed processing and especially the highly innovative copy-paste

attack. Finally we should note that the proposed scheme reaches a very high performance, which however depends on the accuracy of the human object extraction module. In a future work this problem will be further addressed and sub-object authentication methods will be proposed, so that next generation watermarking schemes can protect the semantic content of images more effectively than the existing methods.

Acknowledgment

This research is performed in the framework of the Greek Secretariat of Research and Technology Project “MORFES: Spatiotemporal Modeling of Standing, Full-Length and Dimidiate Figures in Works of Art, by Retrieval of Proportions, based on the Monuments of the Foundation of the Holy Monastery of Mount Sinai”, which is co-funded by the European Social Fund (75%) and National Resources (25%).

References

- [1] J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, San Mateo, Morgan Kaufmann, 2001.
- [2] S. Katzenbeisser and F. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, 2001.
- [3] F. Petitcolas, R. Anderson, and M. Kuhn, “Attacks on Copyright Marking Systems”, in *Proceedings of the 2nd International Workshop of Information Hiding*, pp. 218-238, 1998.
- [4] J. Wang, S. Lian, Y. Dai, G. Liu and Z. Ren, “Secure Semi-Fragile Multi-Feature Watermarking Authentication Scheme,” *Journal of Information Assurance and Security*, Vol. 1, p.p. 265-274, 2006.
- [5] H.O. Altun, A. Orsdemir, G. Sharma and M.F. Bocko, “Optimal Spread Spectrum Watermark Embedding via a Multistep Feasibility Formulation,” *IEEE Transactions on Image Processing*, Vol. 18, No.2, February 2009.
- [6] N. Bi, Q. Sun, D. Huang, Z. Yang and J. Huang, “Robust Image Watermarking Based on Multiband Wavelets and Empirical Mode Decomposition,” *IEEE Transactions on Image Processing*, Vol. 16, No. 8, August 2007.
- [7] S.P. Maity and S. Maity, “Multistage Spread Spectrum Watermark Detection Technique Using Fuzzy Logic,” *Signal Processing Letters*, Vol. 16, No. 4, pp. 245-248, April 2009.
- [8] S.K. Kapotas and A.N. Skodras, “Real time data hiding by exploiting the IPCM macroblocks in H.264/AVC streams,” *Journal of Real-Time Image Processing*, Vol. 4, No. 1, March 2009.
- [9] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoan “Secure Spread Spectrum Watermarking for Multimedia” in *IEEE Transactions on Image Processing*, Vol. 6, No. 12, 1997.
- [10] M. Alghoniemy and A. H. Tewfik, “Geometric Invariance in Image Watermarking” in *IEEE Transactions on Image Processing*, Vol. 13, No.2, February 2004.
- [11] B. Chen and G. W. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” in *IEEE Transactions on Information Theory*, vol. 47, pp. 1423–1433, 2001.
- [12] C. Y. Lin, M. Wu, J. Bloom, I. Cox, M. Miller, and Y. Lui, “Rotation, scale, and translation resilient watermarking for images,” *IEEE Trans. on Image Processing*, vol. 10, pp. 767–782, 2001.
- [13] M.Wu and H. Yu, “Video access control via multi-level data hiding,” in *Proc. of the IEEE ICME*, N.Y. York, 2000.
- [14] S. Pereira and T. Pun, “Robust template matching for affine resistant image watermarks,” *IEEE Transactions on Image Processing*, vol. 9, no. 6, 2000.
- [15] L. Coria¹, P. Nasiopoulos, R. Ward and M. Pickering, “An Access Control Video Watermarking Method that is Robust to Geometric Distortions,” *Journal of Information Assurance and Security*, Vol. 2, p.p. 266-274, 2007.
- [16] D. Zheng; S. Wang and J. Zhao, “RST Invariant Image Watermarking Algorithm With Mathematical Modeling and Analysis of the Watermarking Processes,” *IEEE Trans. on Image Processing*, vol. 18, pp. 1055–1068, May 2009.
- [17] Y. Wang and A. Pearmain “Blind MPEG-2 video watermarking robust against geometric attacks: a set of approaches in DCT domain,” *IEEE Trans. on Image Processing*, vol. 15, pp. 1536–1543, June 2006.
- [18] X.Y. Wang and C.Y. Cui, “A novel image watermarking scheme against desynchronization attacks by SVR revision,” *Journal of Visual Communication and Image Representation*, Elsevier, No. 5, pp. 334-342, July 2008.
- [19] Y. Abu-Mostafa and D. Psaltis, “Image normalization by complex moments,” in *IEEE Trans. on Pattern Analysis and Machine Intelligent*, vol.7, 1985.
- [20] M. K. Hu, Visual pattern recognition by moment invariants, in *IEEE Trans. on Information Theory*, vol. 8, pp. 179–187, 1962.
- [21] Yiping Chu, Yin Zhang, Sanyuan Zhang, Xiuzi Ye, "Region of Interest Fragile Watermarking for Image Authentication," *1st International Multi-Symposiums on Computer and Computational Sciences*, vol. 1, pp. 726-731, 2006.
- [22] H. Liu and M. Steinebach, “Non-Ubiquitous Watermarking for Image Authentication by Region of Interest Masking,” in *Proceedings of the Picture Coding Symposium*, Portugal, 2007.
- [23] K. Zebbiche and F. Khelifi, “Region-Based Watermarking of Biometric Images: Case Study in Fingerprint Images,” *International Journal of Digital Multimedia Broadcasting*, vol. 2008, 2008.
- [24] X. Guo and T.-G Zhuang, “A Region-Based Lossless Watermarking Scheme for Enhancing Security of Medical Data,” *Journal of Digital Imaging*, Springer, Vol. 22, No. 1, p.p. 53-64, February 2009.
- [25] MPEG-4 Part 10, *Advanced Video Coding*, ISO/IEC, May 2003.
- [26] J. Wood, “Invariant pattern recognition: a review,” *Pattern Recognition*, vol. 29, no. 1, pp. 1–17, 1996.
- [27] S. O. Belkasim, M. Shridhar, and M. Ahmadi, “Pattern Recognition with Moment Invariants: A Comparative Study and New Results,” *Pattern Recognition Society*, Vol. 24, No. 12, pp. 1117-1138, 1991.
- [28] P. Hetherington, *The “Painter’s Manual” of Dionysius of Fournas*. London: Sagittarius Press, 1974.
- [29] G. Yang and T. S. Huang, “Human Face Detection in Complex Background,” in *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, 1994.
- [30] Papoulis A., *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, New York, 1984.

- [31] H. Wang, S-F. Chang “A highly efficient system for automatic object detection in MPEG video,” in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 615-628, 1997.
- [32] P. Tzouveli, K. Ntalianis, S. Kollias, “Human Semantic Object Watermarking Based on HU Moments” in *Proceedings of the IEEE Workshop on Signal Processing Systems*, Athens, Greece, November 2005.
- [33] Ballard D. H., “Generalizing the Hough transform to detect arbitrary shapes,” *Pattern Recognition*, Vol. 13, No. 2, pp. 111–122, 1981.
- [34] R. Devaney, *An Introduction to Chaotic Dynamical Systems*, Redwood City, CA: Addison-Wesley, 1989.

Authors' Biographies

Dr. Klimis Ntalianis was born in Athens, Greece, in 1975. He received the Diploma degree and the PhD degree in electrical and computer engineering, both from the National Technical University of Athens (NTUA), Athens, Greece, in 1998 and 2002 respectively. He is the author of more than 50 scientific articles and a reviewer of several international journals and conferences. His research interests include 3-D image processing, video organization, multimedia cryptography and data hiding. During the last decade, Dr. Ntalianis has received prizes for his academic achievements. His PhD studies were supported from the National Scholarships Foundation and the Institute of Communications and Computers Systems of the NTUA. Dr. Klimis Ntalianis has participated in 12 Greek and European projects as researcher and in 4 Greek and European projects as senior researcher. Dr. Ntalianis is a member of the Technical Chamber of Greece.

Dr. Paraskevi Tzouveli was born in Athens, Greece. She obtained her Diploma from School of Electrical and Computer Engineering of National Technical University of Athens in 2001 and she is currently pursuing her Ph.D. degree at the Image, Video, and Multimedia Systems Laboratory at the same University. She is the author of more than 30 papers and a reviewer of several international journals and conferences. Her current research interests lie in the areas of image and video analysis, information retrieval, knowledge manipulation, cryptography and e-learning systems. She has been a member of the Technical or Advisory Committee

Prof. Stefanos Kollias was born in Athens, Greece. He obtained his Diploma from NTUA in 1979, his M.Sc. in Communication Engineering in 1980 from UMIST in England and his Ph.D in Signal Processing from the Computer Science Division of NTUA. He is with the Electrical Engineering Department of NTUA since 1986 where he serves now as a Professor. Since 1990 he is Director of the Image, Video, and Multimedia Systems Laboratory of NTUA. He has published more than 120 papers in the above fields, 50 of which in international journals. He has been a member of the Technical or Advisory Committee or invited speaker in 40 International Conferences. He is a reviewer of 10 IEEE Transactions and of 10 other journals.

Dr. A.S.Drigas (Eng & psych) is Senior Researcher at IIT-NCSR Demokritos. He is Coordinator of Telecoms & founder of Net Media Lab since 1996. 1985 to 1999 was Operational manager of Greek Academic network. Coordinator of Several International & National Projects, in the fields of ICTs – Telecoms, e-services (e-learning, e-psychology, e-government, e-inclusion, e-culture, e-business etc), He has published more than 200 international & national articles in ICTs, 7 books, 25 educational CD-Roms, & several patents. He has been member in several International & National committees for design and coordination Network & ICT services & activities, and also in several committees of international conferences & journals. He has also received several distinctions for his scientific work (articles, projects, patents).