Security of Human Video Objects by Incorporating a Chaos-Based Feedback Cryptographic Scheme

Tzouveli Paraskevi
National Technical University of Athens
9, Iroon Polytechniou str.
Zografou 157-73, Athens, Greece
+30 210 772 4352
tpar@image.ntua.gr

Ntalianis Klimis
National Technical University of Athens
9, Iroon Polytechniou str.
Zografou 157-73, Athens, Greece
+30 210 772 4352
kntal@image.ntua.gr

Kollias Stefanos
National Technical University of Athens
9, Iroon Polytechniou str.
Zografou 157-73, Athens, Greece
+30 210 772 2488
stefanos@cs.ntua.gr

ABSTRACT

Security of multimedia files attracts more and more attention and many encryption methods have been proposed in literature. However most cryptographic systems deal with multimedia files as binary large objects, without taking into consideration regions of semantic information. These regions may need better protection or can be the only regions that need protection, depending on the specific application. Towards this direction, in this paper we propose a human video object encryption system based on the chaotic logistic map. Initially face regions are efficiently detected and afterwards body regions are extracted, using geometric information of the location of face regions. Next the pixels of extracted human video objects are encrypted using an iterative cipher module, which is based on a feedback mechanism responsible for mixing the current encryption parameters with encrypted information of the previous step. The system presents robustness against known cryptanalytic attacks, and can save us a great amount of computational resources and time devoted for encrypting the whole contents of a multimedia file.

Categories and Subject Descriptors

I.4 [**Image Processing and Computer Vision**]: I.4.m Miscellaneous – *multimedia cryptography*

General Terms

Algorithms, Design, Security

Keywords

Cryptographic systems, chaos, logistic map, face and body detection, human video objects.

1. Introduction

During the last decade the increasing need for multimedia communications has induced a growing need for security of visual content. There are several applications such as medical imaging systems, military communications, confidential video conferencing and pay-TV where content of crucial importance is stored, transmitted and processed.

The most common way to protect large multimedia files is by using conventional encryption techniques. Implementations of popular public key encryption methods, such as RSA or El Gamal cannot provide suitable encryption rates, while security of these algorithms relies on the difficulty of quickly factorizing large numbers or solving the discrete logarithm problem, topics that are seriously challenged by recent advances in number theory and distributed computing.

On the other hand, private key bulk encryption algorithms, such as Triple DES or Blowfish, are more suitable for transmission of large amounts of data. However, due to the complexity of their internal structure, they are not particularly fast in terms of execution speed and cannot be concisely and clearly explained, so that to enable detection of cryptanalytic vulnerabilities.

In order to confront these problems systems based on chaotic maps where proposed. Chaotic maps present many desired cryptographic qualities such as simplicity of implementation that leads to high encryption rates, and excellent security. In literature, some chaos-based cryptographic systems have been proposed: In [1], a chaotic key-based algorithm (CKBA) for image encryption is proposed, functioning as a value substitution cipher. An encryption algorithm that uses the iterations of the chaotic tent map is proposed in [2]. In [3], each character of the messages is encrypted as the integer number of iterations performed by the logistic equation. Another encryption algorithm based on synchronized chaotic systems is proposed in [4], where it is suggested that each byte of a message should be encrypted using a different chaotic map. However the security of most chaos-based block encryption algorithms is not analyzed in terms of known cryptanalytic techniques. Furthermore none of the algorithms considers human video objects or any other type of semantic information.

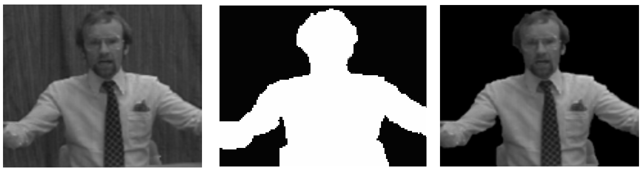
In this paper we propose an automatic chaos – based human video object encryption system. The proposed system is novel as no other object-based chaotic encryption systems have been proposed in literature, even though video objects are defined in the framework of MPEG-4/7 standards. The main subsystems of the proposed system include a face detection module [11], a body detection module and an iterative cipher mechanism based on the logistic function. After a human region is detected, the encryption module encrypts the region pixel-by-pixel, taking into consideration, in each iteration, the values of the previously encrypted pixels. This feedback property, combined with the 256-

ACM Multimedia'04, October 10–16, 2004, New York, NY USA.

bit key, makes our algorithm robust even when homogeneous segments of human video objects are encrypted. This is better illustrated after performing cryptanalytic attacks to the proposed system. Furthermore a simple implementation of the new cipher achieves high encryption rates on general-purpose computers. Finally, our system makes rational use of computational resources since only semantic information is encrypted.

2. Human Video Object Detection

In order to detect human video objects an effective method is used. According to the proposed method, initially, a human face detection module is incorporated based on chrominance values of face areas. Then, the extracted face areas are used for detecting the body areas of video objects, using geometric attributes that relate face and body areas. Finally face and body regions are combined and human video objects are extracted [10].



(a) Initial Image (b) Object Mask (c)Object Extraction

Figure 1. Human Face and Body Extraction Method

Figure 1 illustrates the phases of the proposed method. Firstly the human video object region is detected within the initial image (Figure 1a) and an object mask is produced (Figure 1b). Then, the human video object (Figure 1c) is extracted using the object mask. More details about the proposed method for face and body extraction are given in the next subsection.

2.1 Training Set Construction and Final Segmentation

The human face and body detection modules provide an initial estimation of the human video object forming the foreground training set, say Df. Similarly, a background set, say Db, should also be created. For this reason initially a region of uncertainty is created around the selected foreground masks (face (Mf) and body (Mb)). In particular for each connected component (representing face or body region), the confidence interval of the Gaussian pdf model increases further than 80%, leading to an expansion of the face and body areas. Under this consideration, the new blocks, which are classified to the face or body region, compose the region of uncertainty. Then, the background mask Db is comprised of the blocks that do not belong either to the face/body masks or to the uncertainty zone. As a result, the neural network training set consists of the blocks of sets Df and Db. Since there is a large number of a similar training blocks, Principal Component Analysis (PCA) is incorporated to reduce their number and the remaining blocks are used for training the network. Finally the trained neural network classifies the image-pixels to extract the human video object [11].

3. Chaos and logistic map

Chaotic functions, first studied in the 1960's, present numerous interesting properties that can be used by modern cryptographic schemes. In particular, 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 chaos functions that has been studied recently for cryptography applications is the logistic map. The logistic map function is expressed as $x_{n+1} = r \cdot x_n (1 - x_n)$ Eq.(1) where x takes values in the interval [0,1]. It is one of the simplest models that presents chaotic behavior [8].

If parameter r takes values between (0,3), the logistic map which is defined by Eq.(1) is seen to converge to a particular value after some iterations. As the parameter 'r' is further increased, the curves bifurcations become faster and faster and when r takes values greater than 3.57 (the r=3.57 known as the "point of accumulation") periodicity gives way to complete chaos.

Finally for r=3.9 to 4, the chaos values are generated in the complete range of 0 to 1. In our system we take this advantage of the logistic map and apply it in order to generate chaotic values that are used during the encryption of the human video objects.

4. Cryptographic Scheme

The conventional encryption techniques take the unprotected data called plaintext, apply a key-depended encryption algorithm producing the ciphertext. The process of turning ciphertext back into plaintext is called decryption. In this paper plaintext is denoted by P and in our case it is a human video object. Ciphertext is denoted by C.

In modern effective cryptographic schemes, the secrecy of encrypted information does not depend on the cryptographic algorithm but on the key that is used with the algorithm to produce the encrypted data. The largest the key is the better the secrecy of the encrypted information. In our system, the keyspace is 256-bit which is split into 32 subkeys of 8 bits each (session keys). Our algorithm is a symmetric cipher as the same key is used both by the encryption and decryption modules.

5. The Proposed Chaos-Based Cryptographic System

The proposed chaos-based cryptographic system consists of two main modules the encryption and decryption module.

5.1 The Encryption Method

An overview of the proposed system's encryption module is depicted in Figure 3. Initially the human video object is extracted by the face and body detection modules, as described in section 2. Afterwards, the pixels of the human video object are scanned from top-left to the bottom-right providing the p_i pixels (plaintext pixels). The output of the chaotic map is been added to the plaintext producing the ciphertext. The two successive session keys k_i and k_{i+1} of the keyspace are used to regulate the initial conditions of the chaotic map. The robustness of the system is further reinforced by a feedback mechanism, which leads the cipher to acyclic behavior so that the encryption of each plain pixel depends on the key, the value of the previous cipher pixel and the output of the logistic map.

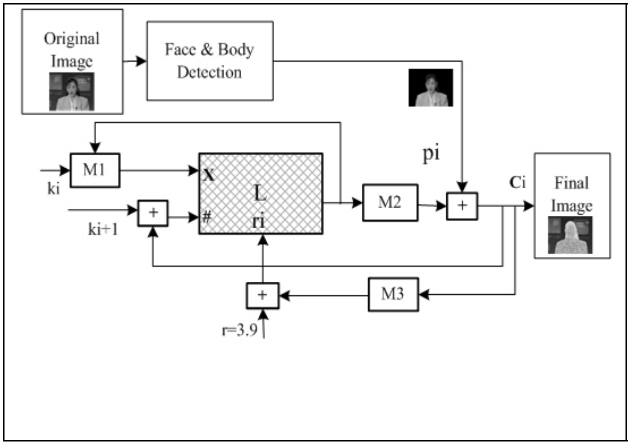


Figure 2. Diagram of Encryption Module

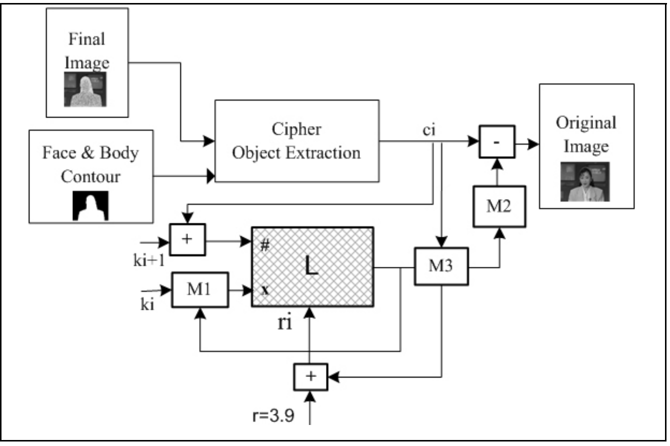


Figure 3. Diagram of Decryption Module

In particular the feedback mechanism includes three operations: With the first operation the output value of the logistic map is input to box M1. The session key k_i is also input to M1 box. The box M_1 represents a mapping function from the input interval to the domain of the logistic map (real numbers in the interval [0, 1]) and fixes the initial value of x.

The second operation interjects in the computation of the number of iterations the logistic map performs (Eq.1). Specifically, the cipher pixel is added to the k_{i+1} session key. The result of this addition provides the value of parameter n which controls the number of iterations the logistic map performs. Finally, the third operation is responsible for the computation of parameter r_i. The M_3 box is a mapping function from the interval [0, 255] to the interval [0, 0.1] so that r_i takes values in the interval [3.9, 4.0], where the logistic map presents chaotic behavior as we mentioned in section 3.

The box M2 is used with the purpose of normalizing the output of the logistic map. Normalization is performed by box M_2 which represents a fuzzy membership function that maps interval [0, 1] into the interval [0, 255].

According to the diagram of Figure 2, for each pixel pi two successive session keys k_i and k_{i+1} are used to regulate the initial conditions of the chaotic map. Then, each pixel p_i is encrypted by adding the normalized output of the logistic map to the pixel value. As we encrypt each new pixel *i*, the counter that keeps track of the current session key, is incremented so that the next session key will be consider in the next iteration. The output of the proposed system is an image that contains encrypted human video objects as foreground video objects.

5.2 The Decryption Method

The diagram of the proposed decryption module is given in Figure 3. The decryption module receives an image with encrypted human video objects together with their contours and returns the original image.

In particular, there is an encrypted region extraction module that extracts the encrypted human video object from the received image by using its contour. The decryption module works in the same way as the encryption module but now the output of the logistic map is subtracted from the corresponding cipher pixel c_i providing the plain pixel. The output of the decryption module is the original image (with decrypted human video object).

6. Experimental Results

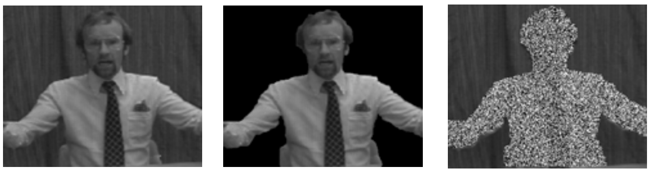
To evaluate the efficiency of the proposed cryptographic scheme one frame of the Trevor sequence is incorporated (Figure 4(a)) as the original image. Initially, human video object is extracted from the image by the face and body detection module. The output of this module is depicted in Figure 4(b).

In our system, the key size is 256-bit which is splited into 32 sub keys of 8 bits each (session keys). This key is used both by the encryption and decryption operations. Using this key as an input to the logistic map, the output of this module provides the values that will be added to the human video object. The encrypted human video object is depicted in Figure 4(c). As can be seen from this figure, the encrypted human video object region is totally invisible.

The decryption method takes as input the final image with the encrypted human video object (Figure 5(a)), together with the contour of the human video object to detect the encrypted object as is shown in Figure 5(b). Then, the output of the logistic map module is now subtracted from the cipher human video object region giving the original image as it is shown in Figure 5 (c).

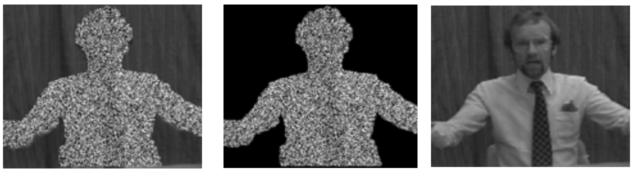
The encryption and decryption examples that are presented in Figures 4 and 5 use the same key in both methods. The proposed system is robust to different-key attacks. A desirable property that the proposed algorithm presents is that it is highly sensitive to the key. Statistical analysis shows that changing one bit in the decryption key causes heavy changes in the corresponding decrypted content of the image.

To illustrate the security of the algorithm, a hacker's approach to crack an encrypted image provided by this algorithm is presented here. Since the 256-bit sequences of key is used as an input to the logistic map, it is computationally infeasible for a hacker to guess each and every output of the logistic map.For the algorithm that has been presented, the keyspace contains 2^{256} different values. Additionally the number of iterations supported by the logistic map module is between 0 and 767, as cipher pixels take values in the interval [0,512] and the session keys take values in the interval [0, 255].



(a) Original Image (b) Original Object Extraction (c) Final Image

Figure 4: Phases of the Encryption Method



(a) Final Image (b) Cipher Object Extraction (c) Original Image

Figure 5: Phases of the Decryption Method

All these parameters lead to the fact that the only way to break the proposed system is by brute–force attack scanning the whole keyspace, that is to try 2^{256} different keys.

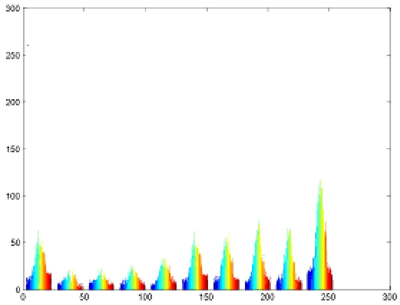


Figure 6: Histogram of Cipher Image

Another advantage of the proposed algorithm is in the amount of computational time spent during the encryption module. In particular, our algorithm spends less time in encryption as only human video objects are considered and not the whole image. The human video objects extraction cost depends on the size of the image. Furthermore as the proposed system is based on feedback mechanisms, periodicities in the encrypted data do not appear. Another desirable effect of feedbacks is that even small changes in the plain image will lead to a completely different cipher image. This sensitivity is also a plus to the security of the proposed algorithm.

Finally, the histogram of a plain image contains large spikes. These spikes correspond to color values that appear more often in the plain image. The histogram of the cipher image, Figure 6, is more uniform and bears no statistical resemblance to the plain image.

7. Conclusion

The recent growth of networked multimedia systems has increased the need for protection of digital media. Most cryptography systems are not taking into consideration regions of semantic information which actually need to be encrypted.

Our proposed system takes as input images containing human video object regions. The human video objects are extracted by a face and body region detection module. Afterwards, the extracted video objects are encrypted by an encryption module that is based on the logistic map and three independent feedback mechanisms. The final image consists of the encrypted human video object and the original background. Experimental results illustrate the security of the proposed scheme showing that the only method to break the system is by brute-force attack.

8. References

[1]. Jui-Cheng Yen and Jiun-In Guo, "A new chaotic key-based design for image encryption and decryption", *in Proc. IEEE Int. Conf. Circuits and Systems*, 2000, vol. 4, pp. 49–52.

[2]. T. Habutsu, Y. Nishio, I. Sasase, and S. Mori, "A secret key cryptosystem by iterating a chaotic map", *in Proc. Advances in Cryptology,* Berlin, Germany: Springer-Verlag, 1991, pp. 127–140.

[3]. M. S. Baptista, "Cryptography with chaos", *Phys. Lett*. A, vol. 240, pp. 50–54, 1998.

[4]. Bruce Schneier, "Applied Cryptography", Second Edition

[5]. G. Jakimoski and L. Kocarev, "Chaos and Cryptography: Block Encryption Ciphers Based on Chaotic Maps", *IEEE Transactions on Circuits and Systems—I*: Fundamental Theory and Applications, vol. 48, no. 2, February 2001.

[6]. E. Biham, "Cryptanalysis of the chaotic-map cryptosystem suggested at EUROCRYPT'91", *in Proc. Advances in Cryptology,* Berlin, Germany: Springer-Verlag, 1991, pp. 532–534.

[7]. Shujun Li, Xuan Zheng, "Cryptanalysis of a Chaotic Image Encryption Method", *in Proceedings of 2002 IEEE International Symposium on Circuits and Systems*, vol. 2, pp 708-711, 2002.

[8]. R. Devaney, "An Introduction to Chaotic Dynamical Systems", 2nd ed. Redwood City, *CA: Addison-Wesley*, 1989.

[9]. Wang H., Chang, Shih-Fu, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video Sequences", *IEEE Trans. CSVT*, Vol. 7, No. 4 (1997) 615-628

[10]. Papoulis A., "Probability, Random Variables, and Stochastic Processes", *McGraw Hill*, New York (1984)

[11]. P. Tzouveli, K. Ntalianis and S. Kollias "Automatic Videoconference Objects Watermarking Using Object Adapted Qualified Significant Wavelet Trees," *in Proc. of International Workshop* VLBV, Madrid, Spain, September 2003.

[12]. H. Wang and Shih-Fu Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video Sequences," IEEE Trans. CSVT, vol. 7, No. 4, pp. 615-628, August 1997