AUTHENTICATION OF DIGITAL DOCUMENTS USING SECRET KEY BIOMETRIC WATERMARKING

V.ANITHA[1] & R.LEELA VELUSAMY[2]

[1,2]Department of Computer Science and engineering, National Institute of Technology, Tiruchirappalli, India
E-mail : anitha.v2000@gmail.com

Abstract - Digital documents play a major role in modern era. They are easy to generate, modify and manage. The easy modifiable property of digital document makes it more vulnerable to forgery. It can be easily tampered or forged. So the challenge is to produce digital documents that are highly resistant to forgery and reliably confirms the real owner of the document. This can be resolved by biometric watermarking which make a direct relation between the document and its owner. A new biometric watermarking technique with secret key is proposed to digitize the authoritative documents issued by government / other organizations as a part of UID / Aadhar card project of India using biometric watermarking. Biometric code is generated from the biometric data collected from the owner of the document. The biometric code is watermarked in the document with a secret key to generate a biometric watermarked document that authenticates the real owner. De-watermarking the document with the same key yields the biometric code that can be used for authentication of the document. If the document is tampered in any way it will be indicated in the extracted watermark. Experimental results show that 100% accuracy is obtained in authenticating the genuine documents.

Keywords- *secret key watermarking, biometric watermarking, tamper detection, digital document authentication, ownership verification, biometric authentication.*

I. INTRODUCTION

In the modern era, digital document concept is becoming increasingly important as more of the world's information is stored as readily transferable bits.

Digital documents have so much of advantages over print documents. Digital documents are less expensive and easy to store, transport and search compared to traditional print documents. But it has its own limitations too. A simple image editor can be used to modify and make a forged document. Digital documents can be tampered easily. In order to utilize the whole benefits of digital document, these limitations have to be overcomed.

Digital watermarking technique has been used in order to overcome these limitations by embedding some text/logo/pseudorandom sequence that identifies the owner of the document. Some applications of watermark are:

- Establishing ownership by embedding identifying data
- Tracking the movement of authorized copies by embedding a unique serial number in each copy
- Attaching meta-data that pertains to the image such as a time, date, and location stamp

These digital watermarks suffer from certain limitations. Watermarks are least correlative to the owner of the digital document. Watermark doesn’t directly link the digital document to the owner of it. And the related characteristics of the watermark with the digital document may change over time. So the traditional watermarking method does not convincingly validate the claimed identification of the person. Using such kind of watermark may lead to easy forgery or tampering.

Recently biometrics is merged into watermarking technology to enhance the credibility of the conventional watermarking technique. Biometric watermarking is a special case of digital watermarking in which the watermark content is biometric data. Access control or authenticity verification has been addressed by both digital watermarking as well as by biometric authentication. By embedding biometrics in the host, it formulates a reliable individual identification as biometrics possesses exclusive characteristics that can be hardly counterfeited. Hence, the conflicts related to the intellectual property rights protection can be potentially discouraged [1]. Consequently, it has been decided by governmental institutions in Europe and the U.S. to include digital biometric data in future ID documents. In India, biometric based UID scheme, AADHAR is started with the goal of issuing a unique identification number to all the Indian citizens.

Adhaar card project of India is a major motivation beyond this project. It involves the collection of biometric data from all Indian citizens. Aadhaar is a 12-digit unique identity number which is to be issued by Unique Identification Authority of India (UIDAI) to all residents in India. The number will be stored in a centralized database and linked to the basic demographics and biometric information – photograph, ten fingerprints and iris – of each individual. UID is easily verifiable in an online, cost-effective way. And also, it is unique and robust enough to eliminate the large number of duplicate and fake identities in government and private databases. The random number generated will be devoid of any classification based on caste, creed, religion and geography [2].

The goal of this project is to propose a technique for the creation of robust, forgery resistant digital documents which replace/used along with the paper documents possessed by an individual. It makes use of the issued UID as well as the biometric data collected from the Indian citizens for the generation of biometric watermarked document which uniquely identify the owner of the document. This invokes an additional layer of authentication to the underlying system.

The rest of the paper is organized as follows. In the next section, a brief discussion on the related work is given. Section III explains the proposed system in detail. The next section discusses the experiments and results and finally the concluding remarks and future work are mentioned in the last section.

II. RELATED WORK

Digital watermarking is a method that has received a lot of attention in the past few years. A digital watermark can be described as a visible or preferably invisible identification code that is permanently embedded in the data. Most of the watermarking model is used to embed a logo or private information into the document/other multimedia data for the purpose of copyright protection. By adding watermark, we add a certain degree of protection to the document/ image (or to the information that it contains). The goal is to embed some information in the image without affecting its visual content. In the copyright protection context, watermarking is used to add a key in the multimedia data that authenticates the legal copyright holder and that cannot be manipulated or removed without impairing the data.

F.Liu et.al [3] proposed a watermarking scheme for multiple cover images and multiple owners. It makes use of the visual cryptography technique, transform domain technique, chaos technique, noise reduction technique and error correcting code technique where the visual cryptography technique provides the capability to protect the copyright of multiple cover images for multiple owners, and the rest of the techniques are applied to enhance the robustness of the scheme. A new technique of color image digital watermarking based on visual cryptography is proposed by S.Kandar et,al [4]. In a digital image watermarking model proposed by H.Nyeem et.al. [5], a secret key is employed in both watermark encoding and decoding by which the model gains a resistant against many attacks. A very simple and effective watermarking technique is proposed by P.W.Wong et.al.[6]. Block by block the image is being watermarked with the help of secret key/ public key which provides additional protection to the watermarked image. S.Kandar et al. [7] proposed Visual Cryptographic scheme for color images where the divided shares are enveloped in other images using invisible digital watermarking. The shares are generated using Random Number.

Merging biometrics with the digital watermarking technique is called as biometric watermarking. A few proposals are made in biometric watermarking too. V.S.Inamdar et,al.[1] proposed a biometric watermarking scheme to watermark handwritten signature for signature authentication. Another one biometric watermarking technique is proposed by A.E. Hassanien [8] to hide iris data in digital images for protecting those iris data from tampering and stealing. A multiple watermarking technique based on wavelets and visual cryptography is proposed by S. Radharani et.al. [9]. Among all the reviewed watermarking techniques, the secret key based block by block watermarking technique proposed by P.W.Wong et.al.[6] is chosen to be the best and effective technique and is followed as the watermarking technique in this paper.

And for the generation of biometric code, a literature review is done on various biometric recognition methods. Among various existing biometrics, iris is considered to have more uniqueness and hence yield a high accuracy. Extensive research is been done in the field of iris recognition. Avila et al. [10] proposed iris recognition for biometric identification using dyadic wavelet transform zero-crossing and achieved a recognition rate of 98%. Li Ma et al. [11] proposed an iris recognition which uses circular symmetric filters for feature extraction and nearest feature line approach is used for iris matching. Tisse et al. [12] proposed iris based personal authentication technique based on gradient decomposed hough transform or integro-differential operators combination for iris localization and analytic image concept to extract information from iris texture. . Libor Masek [13] developed an iris identification system which employs hough transform for iris segmentation and 1D Log-Gabor filters for feature extraction. Simple and effective techniques such as majority voting and haar transform are applied to the template / iriscode generated by iris algorithms to reduce the storage space requirement and improve the accuracy in a iris recognition system proposed in [14]. It employs Libor Masek process to generate the iris template and it achieves a high accuracy rate of 99.82%. Naveen singh et al. [15] designed a iris recognition system using a Canny Edge Detection scheme and a Circular Hough transform, to detect the iris boundaries in the eye's digital image Among all reviewed iris biometric recognition techniques, iris recognition system based on haar transform and majority voting technique [14] is chosen to generate the biometric code for biometric watermarking, since it yield a good accuracy with less requirement of storage space.

III. PROPOSED SYSTEM

The proposed solution has the following four modules listed below:

- Biometric code generation
- Biometric watermark Generation
- Secret key biometric watermarking of the digital document
- Authentication of biometric watermarked digital document

A. *Biometric code generation*

The first module collects the biometric data from the user and generates the biometric code. Any biometric data such as face, fingerprint, palmprint, etc., or a combination of them can be used as the source to generate the biometric code. In this paper, the biometric code is generated using iris since iris is more popular among all biometrics because of its greater uniqueness and accuracy.

It is wise to choose a biometric model that need less storage space and high accuracy. Storage space of the biometric templates in the database is of major concern because applications that employ biometric recognition system deal with a huge set of data. So the template which has to be created from the biometric data for storage should occupy less space. So iris recognition system based on haar transform and majority voting technique [14] is used for the iriscode generation. The technique applies multilevel haar transform on biometric templates in order to reduce the storage space and to increase the accuracy by combining the important features of the templates after the haar transform. The generated biometric code is efficient in terms of accuracy and storage space.

Iris image is collected from the Aadhaar database. .Libor Masek iris recognition model [13] is used to generate the iris template from the iris image. Fig.1 explains the steps involved in generating the iris template.

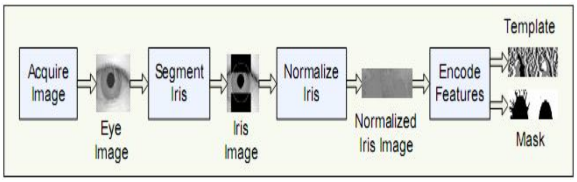


Figure 1: Iris template generation using Libor Masek process

A method based on multilevel haar transform is applied to the generated iris template and the important features of it are combined to generate a feature vector. Then the feature vector is encoded to give biometric code in binary bits. MVIC45 (Majority voting Iris Code 45) method described in [11], is used to generate a 225 iriscode. MVIC45 method is chosen because of it has a good tradeoff between low storage space and high accuracy. This 225 bit iriscode is considered as the biometric code that is to be used in the watermark generation. Fig.2 shows the generated biometric code.

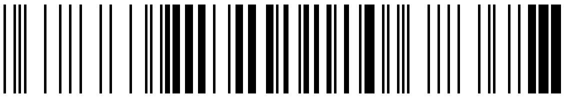


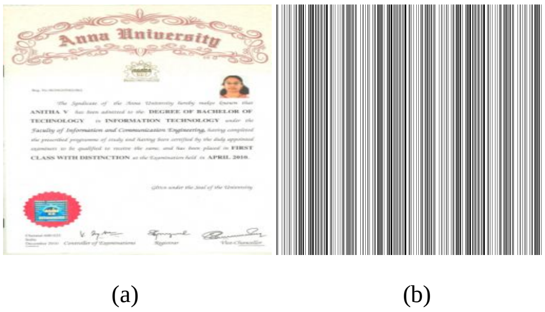
Figure 2: Generated biometric code

The generated biometric code is stored in the database for the authentication purpose. The UID (bio_uid) allotted by the government to each person can be used as an index to store the biometric code in the database. This enables a faster retrieval of the biometric code from the database when many documents of a particular user have to be authenticated.

B. *Biometric watermark generation*

A watermark has to be generated such that it covers the entire size of the digital document. An authoritative document ‘X’ of size $M_X \times N_X$ is considered as the cover image for the watermarking procedure. Let ‘A’ is considered as the binary biometric code image which is of dimension $M_A \times N_A$. ‘A’ is the source for the generation of the invisible watermark which is to be inserted ‘X’. Note that always the size of ‘A’ should be less than or equal to the size of ‘X’.

If ‘A’ is less than ‘X’, from ‘A’ of size $M_A \times N_A$, another binary image ‘B’ of size $M_X \times N_X$ (same as X) is formed by tiling process i.e., periodically replicate the image A to generate an image B which is of the same size as ‘X’. Else if ‘A’ is of the same size as that of ‘X’, ‘A’ is considered to be ‘B’. Fig.3 (a)(b) shows the document ‘X’ (degree certificate) and the generated biometric watermark ‘B’ for that document.



Anna University

The Syndicate of the Anna University hereby makes known that ANITHA V has been admitted to the DEGREE OF BACHELOR OF TECHNOLOGY in INFORMATION TECHNOLOGY under the Faculty of Information and Communication Engineering, having completed the prescribed programme of study and having been certified by the duly appointed examiners to be qualified to receive the same, and has been placed in FIRST CLASS WITH DISTINCTION at the Examination held in APRIL 2010.

Given under the Seal of the University

Chennai
India
December 2010

Controller of Examinations
Registrar
Vice-Chancellor

(a) (b)

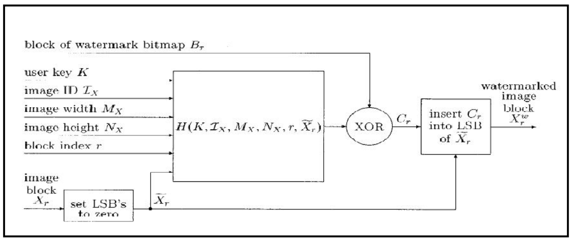
Figure 3: (a) Authoritative document X (b) Generated biometric watermark of the document B

C. *Secret key biometric watermarking of the digital document*

A secret key watermarking scheme based on cryptographic hash function [6] is used. The document can be a grayscale image or a RGB image. Watermarking procedure for a single plane of image is explained below. If it is a grayscale image, procedure can be applied directly in the single plane. If the document is a RGB image, the watermarking has to be applied independently to each of the 3 planes to get the final watermarked RGB image. Fig 4 explains how the generated watermark is inserted into the cover image/document.

The document 'X' and the generated watermark 'B' is divided into blocks of certain size say i×j for the watermarking procedure. Each block in the image is indexed with an integer number 'r'. X_r, B_r represents the r^{th} block of X and B respectively. For each block of X, say X_r, a block X $_r$ is generated by setting the LSB of all the pixels of block X_r to 0. Fig.4 explains how the r^{th} block of generated watermark is inserted into the r^{th} block of cover image / document.

A cryptographic hash function H is used in the watermarking insertion and extraction process of each image block. A cryptographic hash function is a hash function that can be defined as a deterministic procedure that takes an arbitrary block of data and returns a fixed-size bit string, the (cryptographic) hash value, such that an accidental or intentional change to the data will change the hash value.



block of watermark bitmap B_r
user key K
image ID $\mathcal{I}_X$
image width M_X
image height N_X
block index r
image block X_r
set LSB's to zero
$\widetilde{X}_r$
$H(K, \mathcal{I}_X, M_X, N_X, r, \widetilde{X}_r)$
XOR
C_r
insert C_r into LSB of $\widetilde{X}_r$
watermarked image block X_r^w

Figure 4: Biometric watermark insertion into the document

The ideal cryptographic hash function has four main or significant properties: 1) it is easy to compute the hash value for any given message 2) it is infeasible to generate a message that has a given hash 3) it is infeasible to modify a message without changing the hash 4) it is infeasible to find two different messages with the same hash.

A cryptographic hash function H(S) is used to generate a message digest of a fixed size p from the string S. String S is a concatenated string which is composed from the hash inputs secret key k, Unique id of the image I_x, width of the document M_x, height of the document N_x, block index r. String S is used as the input to the hash function to generate the digest D of p bits as the hash output. Hash function is applied to each and every block of the image. H (k, I_x, M_x, N_x r, X $_r$) = H(S) = (d_1, d_2, d_3,…, d_p)where d_i is the i^{th} bit of the message digest and p is the number of bits in the message digest.

The secret key k is a character array of arbitrary length. Secret key is known to the issuer as well as the verifier. Issuer and the verifier of the document is same in most of the cases. For image id $I_{x,}$ a table containing a list of documents type and its corresponding unique id say doc_uid is maintained. Each person in the database will be assigned with a unique id say bio_uid. It is a 12 digit number of Aadhar card that uniquely identifies each person with his/her biometrics. The unique image id $I_{x,}$ is a combination of the UID of the person and the UID of the document i.e. the concatenation of bio_uid and doc_uid. Table 1 lists some of the common authoritative documents and its assigned unique id.

Hash digest D contains a p bit array. If p<i×j, the p bits are replicated to form a bit array of dimension i×j. If p>i×j, then the exceeding bits of p can be truncated to form a bit array of dimension i×j. The resultant array of size i×j is said to be $d_{i\times j}$. The watermark block B_r of size i×j is combined with $d_{i\times j}$ by XOR operation to form C_r. Then C_r is inserted into the least significant bit of X $_r$ to generate the watermarked block of image . Repeat this procedure to generate the watermarked block for all the blocks. All the output watermarked blocks of the image are merged together to form the watermarked document Y. The watermarked document Y is issued to the user and also stored in the database.

The biometric watermarked document 'Y' is stored in the central database using the image id Ix as index i.e. 'Y' is stored in the bio_uid directory (a folder which belongs to a specific person with his/her biouid as the folder name) with doc_uid as its document name. This makes the database to be more efficient and responsive while searching a specific document of a specific person (or) all the documents of the same person. Storing of the issued document and the generated watermark in the database is optional as the biometric code required for verification is already been stored in the Aadhaar database.

TABLE I. DOCUMENT TYPES AND ITS CORRESPONDING UID

Document type	Document UID
Birth certificate	001
Death certificate	002
PAN card	003
Driving license	004
Degree certificate	005
10th certificate	006
12th certificate	007
Land registration	008
Passport	009

D. Authentication of biometric watermarked document

Authentication of biometric watermarked document includes watermark extraction from the biometric watermarked document. Watermark extraction is the exact reverse procedure of watermark insertion procedure which is shown in Fig.5. The watermarked document ‘Y’ is divided into blocks of fixed size i×j with block index r. The LSB bit of Y_r is extracted to generate G_r before gen

$_r$ by setting the LSB of Y_r

$_r$ $_r$, all the other inputs are same. The hash output is replicated using the same procedure mentioned in insertio

$_{i×j,}$ which is XOR ed with G_r to give which is the watermark block. All the blocks of the image Y are combined to give the extracted watermark W_e. The same secret key should be used in order to reconstruct the correct watermark.

In the authentication phase, the biometric code of the person is retrieved from the database using the bio_uid as index. ‘A’ watermark W_g is generated from the retrieved biometric code by the same tiling procedure mentioned above to generate the watermark for insertion. A comparison is made between the generated watermark ‘W_g’ and the extracted watermark ‘W_e’. The biometric watermarked document is authenticated if a hamming distance of 0 is achieved between them. Otherwise the document is not authenticated.

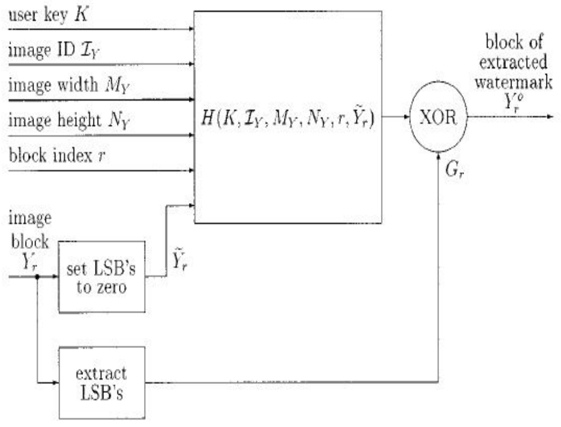


Figure 5: Watermark extraction from the watermarked document

IV. EXPERIMENTS AND RESULTS

A total of 25 documents belonging to 5 different persons (each person has 5 documents) are chosen for the experiment. Biometric watermark for those 5 different persons are generated from their iris. Each biometric watermark is embedded in all the 25 documents i.e., a person’s biometric watermark is embedded in all his/her documents as well as all the other person’s documents too. And this watermarking is done with 5 different secret keys. Totally there are 25 documents, 5 biometric code and 5 keys. Various combinations of it generate 625 different documents among which only 25 documents are genuine with the watermark of its owner and with its own secret key. The 25 biometric watermarked documents having the original owner’s biometric code embedded with the correct secret key alone got authenticated among the 625 different biometric watermarked documents. This shows 100% accuracy in authenticating the real owner of the biometric watermarked document.

Experiments with original document and forged document show how the changes made to the original document reflected in the extracted watermark. Fig.6 (a)(b) shows the original document and the extracted watermark pattern. The extracted watermark shown in Fig.6 (b) has no distortion in the extracted pattern which means the document is not tampered and it can be authenticated. Fig.6 (c) shows the forged document in which the photograph of the original document has been modified. Fig.6 (d) shows the extracted watermark from the forged document. It has a heavy noise in the modified portion and mild noise in the neighborhood portion around it. From the unmodified portion/pattern, the original owner of the forged document can be found. Original owner can be identified by performing an identification search in the database with the extracted biometric code.

If the biometric watermarked document is extracted with the wrong key, then the extracted watermark will be an image full of noise without any pattern as shown in Fig7. Full noise image is also produced when the watermarked document is cropped/rescaled/when an extraction is made from an un-watermarked document.

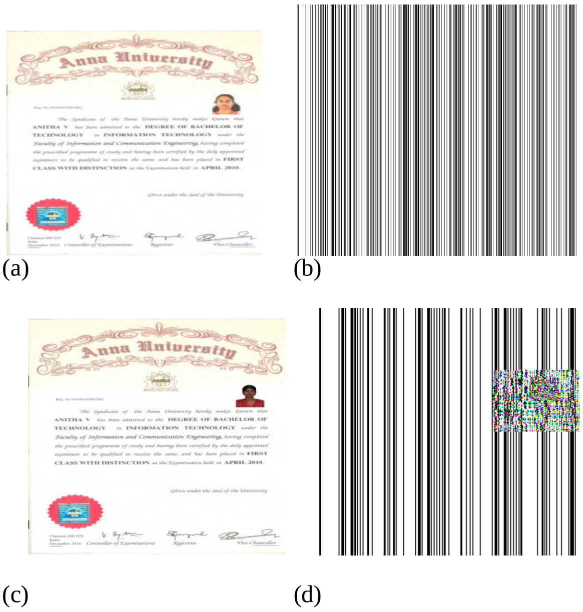


Figure 6: (a)Original document (b) Forged document (c) Extracted watermark from original document (d) Extracted watermark from the forged document

Any change in the hash inputs during the watermark extraction process will make a huge distortion in the hash output. According to the avalanche effect property of hash algorithm, a change of a single bit in the hash input will lead to a major change in the hash output. In addition to the input document image block X□□□$_r$, the other inputs to the hash algorithm are (k, I_x, M_x, N_x r). Since the dimension parameters (M_x, N_x) are included in the hash input, any change in the dimension i.e. rescaling/cropping of the watermarked document will result a noisy image during the extraction pattern. The same result will occur in case of wrong secret key (k) /wrong image id (I_x). The block index r is also added to the hash input such that it adds more resistant against attacks. Resistance to Holliman-Memon attack [16] is increased by adding the block index r. If a attacker attempts to reveal the biometric code from a block of image, all the blocks of the other similarly biometric watermarked documents can't be used for the attack. Only the same block from other similar watermarked documents can be used since we are using the block index r. Hence, the size of the code book is small, i.e., one entry in the codebook per watermarked image in the collection. So the chance of the attacker to find the watermark from the document is less.

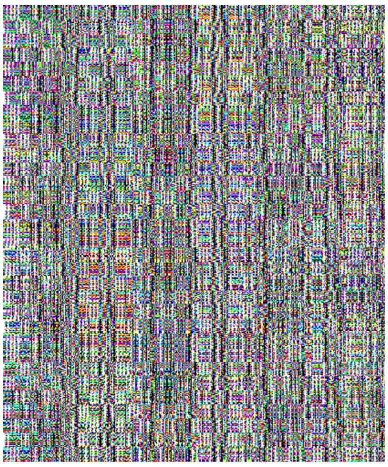


Figure 7: Noisy extracted watermark

V. CONCLUSION AND FUTURE WORK

A simple and efficient technique for digitizing the authoritative documents of an individual based on secret key biometric watermarking key is proposed. It reliably authenticates the real owners of the authoritative documents. Only the authorized person (government/ any organization issuing the document) who owns the secret key, can extract the biometric watermark from the biometric watermarked document for authenticating the document. Hence a third person can't de-watermark and forge the document, since he doesn't possess a valid secret key. If a document is rescaled/ cropped/modified, the changes are directly reflected in the extracted watermark showing where it modified. The original document which is used for forgery can also be figured out. Experiments show that the proposed technique show100% accuracy in authenticating the genuine documents and is resistant against attacks. The future work is to make a study on the performance analysis of the proposed model under attacks.

REFERENCES

[1] V. Inamdar, P. Rege, M. Arya,"Offline handwritten signature based blind biometric watermarking and authetication technique using biorthogonal wavelet transform," International Journal of Computer Applications (0975 – 8887), vol. 11, Dec. 2010, pp. 19-27

[2] Wikipedia, the free encyclopedia, "Unique identification authority of India", http://en.wikipedia.org/wiki/Unique_Identification_Authority of_India. Retrieved on : 10 th April, 2012.

[3] F. Liu and K. Wu, "Robust visual cryptography-based watermarking scheme for multiple cover images and multiple owners," IET Information Security, June 2010, doi: 10.1049/iet-ifs.2009.0183, ISSN 1751-8709

[4] S. Kandar, A. Maiti, C. Dhara, "Visual cryptography scheme for color image using random number with enveloping by digital watermarking," International Journal of Computer Science Issues, May 2011 ,Vol. 8, pp.543-549, ISSN : 1694-0814.

[5] H.Nyeem, W. Boles, C. Boyd, " Developing a Digital Image Watermarking Model," Proc. IEEE Symp. International Conference on Digital Image Computing: Techniques and Applications, 2011, pp.468-473,doi: 10.1109/DICTA.2011.85.

[6] W. Wong and N. Memon, "Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification," IEEE transactions on image processing, vol. 10, no. 10, Oct. 2001.

[7] S. Kandar, A. Maiti, C. Dhara,"Visual Cryptography Scheme for Color Image Using Random Number with Enveloping by Digital Watermarking," International Journal of Computer science Issues, Vol. 8, Issue 3, No. 1, May 2011,ISSN :1694-0814.

[8] E.Hassanien,"Hiding iris data for authentication of digital images using wavelet theory," GVIP 05 Conference, Egypt, Dec 2005.

[9] S. Radharani and L. Valarmathi,"Multiple watermarking scheme for image authentication and copyright protection using wavelet based texture properties and visual cryptography," International Journal of Computer Applications (0975 – 8887), Vol. 23, June 2011,pp.29-36

[10] S. C. Avila, S. R. Reillo R, I. D. Martin, "Iris recognition for biometric identification using dyadic wavelet transform zero-crossing," Proc of the IEEE 35th International. Camahan Conference on Security Technology, 2001, pp. 272 -277

[11] L. Ma, Y. Wang, T. Tan, " Iris Recognition Using Circular Symmetric Filters ," Proc. of the 16th International Conference on Pattern Recognition, IEEE press, Vol. 2, 2002, pp. 414 -417.

[12] C. L. Tisse, L. Torres and M. Robert, "Person Identification Technique Using Human Iris," Proc. of the 15th International Recognition conference on Vision Interface, 2002

[13] L. Masek, "Recognition of human iris patterns for bio-metric identification," Tech. Rep., The School of Computer Science

and Software Engineering, The university of Western Australia, 2003.

[14] V. Anitha, L. Velusamy, "Iris recognition systems with reduced storage and high accuracy using Majority Voting and Haar Transform," Springer Proc. Advances in Intelligent and Soft Computing book Series, International Conference on communications security and information assurance, May 2012.,in press.

[15] N. Singh, D. Gandhi and K. P. Singh, "Iris recognition system using a canny edge detection and a circular hough transform," International Journal of Advances in Engineering & Technology, vol 1, May 2011, pp.221-228.

[16] M. Holliman and N. Memon, "Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes," IEEE Trans.Image Processing, vol. 9, pp. 432–441, Mar. 2000.

◈◈◈