Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Digital Forensics: Electronic Evidence Collection, Examination and Analysis by Using Combine Moments in Spatial and Transform Domain

Hani Saleh
Intel Corporation
Austin, TX, 78749
hani.saleh@intel.com

Sos Agaian
University of Texas at San Antonio
San Antonio, TX 78249
sos.agaian@utsa.edu

Khader Mohamamd
Intel Corporation
Austin, TX, 78749
khader.mohammad@intel.com

***Abstract*—A novel digital forensics tool is developed by combining wavelet invariant with spatial moments. A forensic printed circuit board image matching system is presented that is capable of probing a large database of digital images of circuit boards and compare them for similarity to provide investigation leads for electronic crimes digital forensic science investigations. The developed system has been implemented, and proved to be very efficient in detection similarities between a target image and a large image database even when the target image is noisy, scaled or mirrored.**

Keywords— Invariant Moments, Wavelet Transform, FFT Trandfor, Object Recognition, Edge Detection, Image Matching, Forensic, Digital forensics data, network forensics and digital images.

I. INTRODUCTION

Digital forensics [1]-[2], also referred to as computer forensic analysis, electronic discovery, electronic evidence discovery, digital discovery, data recovery, data discovery, computer analysis, and computer examination, is the process of methodically examining computer media (hard disks, diskettes, tapes, etc.) for evidence. A thorough analysis by a skilled examiner can result in the reconstruction of the activities of a computer user. Digital forensics is now an integral part of many criminal and civil investigations [3]. Basic Forensic Science is concerned with both the law enforcement view of forensics as well as general lab policies. This is an exciting new area bringing together an interesting mix of talent – computer science, law enforcement, judicial, computer engineering, and others in an effort to address computer crime.

Digital forensics is an emerging science. It is an extension of traditional forensics and used to protect and analyze the digital evidence rather than the physical one. Digital forensics has become prevalent, since much more criminal events are recorded by digital devices and the law enforcement recognizes these digital records as legal evidence [4]. Digital forensic is used in the crime investigation for the national investigation organization like prosecution, police and the necessity of digital forensic technique is increasing even from civil field like general enterprise and banking company [5].

Unfortunately, crime methods evolve with technology. New forms of theft started to appear in the industrial evolution era where some companies steal the innovations and designs of their competitors and start manufacturing these products for their own interest without licensing or benefiting the original innovators or manufacturers. Nowadays the wide range of digital devices is often part of an investigation. However, while devices such as phones, digital media players, and game consoles may harbor relevant information, there are some significant challenges associated with forensic analysis of such devices. Cell phones are perhaps the most diverse, as they tend to have no standard interface, either at the hardware or software levels, essentially making the analysis process unique to each device model. Furthermore, forensic tools often cannot handle new or less commonly encountered devices, leaving an investigator to either develop custom tools, or lose the opportunity to examine the device. In addition to the number of incompatible devices of a particular type, such as cell phones, the number of device types, especially integrated devices, is also growing rapidly [7].

The current research efforts in computer forensics are improving digital forensics tools and techniques [8].

There have been many techniques developed which attempt to identify file types; file header, data and network forensics [9]. However, currently available techniques do not have useful for investigation of the acquired digital devices is done by identifying fraud by using their application.

In this article we focus on sub-phase digital forensics such as (1) finding the hardware, which may contain evidence by using the image processing methods , (2) collecting physical electronic evidence hardware which may become evidence, (3) estimating the reliability of evidence, (4) accountability in case of legal action.

The rest of the paper is organized as follows. System design is presented in Section II. Section III presents the system implementation results. Section VI Concludes the paper with conclusion and future research.

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

II. SYSTEM DESIGN

Many systems and/or algorithms have been presented to classify and identify similarity between images [10]-[11]. Most of the systems presented are applicable to certain class of images and sensitive to image mirroring or scaling.

Our goal is to allow for the more accurate identification of hardware circuit theft by creating a tool that determines similarity between PCB images of the investigated subject. The developed tool meets the following requirements:

1. The tool is capable to probe a large database of digital images (up to five gigabyte images) and compare them for similarities.
2. A comparison method suitable for circuit boards is developed.
3. The investigated circuit boards may have ancillary components attached that are included in the analysis.
4. Some of the evidence may be fragmentary and in general the boards will not be identical. Controlled image resolutions at a standard number of pixels per inch (300 dpi) and defined image layout may simplify image analysis.
5. The system is able to compare images and identify similarity even if the target image is scaled or mirrored.

This work created a scientific tool (system) that can aid the investigation of electronic design crimes (block diagram shown in Figure1).

A library with large number of images was created. The image enhancement block was implemented as function that could call any image enhancement method of choice.

The feature vector extraction method used image invariant moments combined with DWT feature vectors with invariant moments in the DWT domain. The simulation results showed that the selected feature vectors gave the best results that are insensitive to image mirroring or scaling.

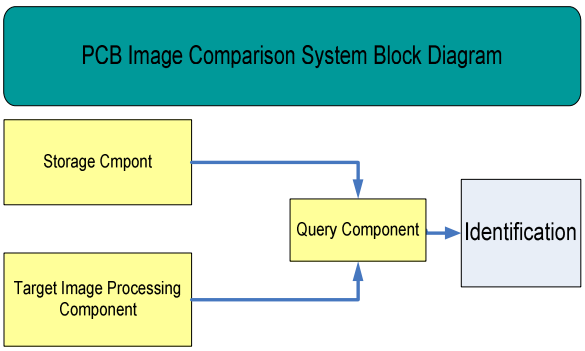

Figure 1. PCB Images Matching System Block Diagram.

The PCB image matching system is composed of three components:

- Storage component,
- Target image processing component
- Query component.

A. *Storage component*

The storage component (shown in Figure 2) is used to create the database of the library images. The input images go through necessary enhancement, proper transformation, feature vectors extraction and the database storage for the library images and their feature vector matrix.

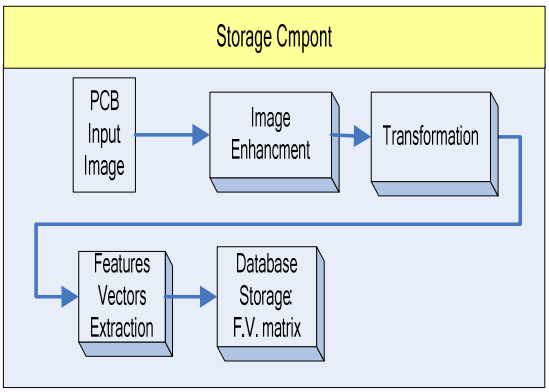

Figure 2. System Storage Components.

The Storage component is composed of the following functions:

1. Image Enhancement: The image passes through an automatic image enhancement algorithm (shown in Figure 3). The goals of this step is:

- To reduce the effect of noise.
- To stress the important features of the image.
- To ease the process of automatic image recognition

This algorithm is an extension to the image enhancement algorithms presented in [12]-[15].

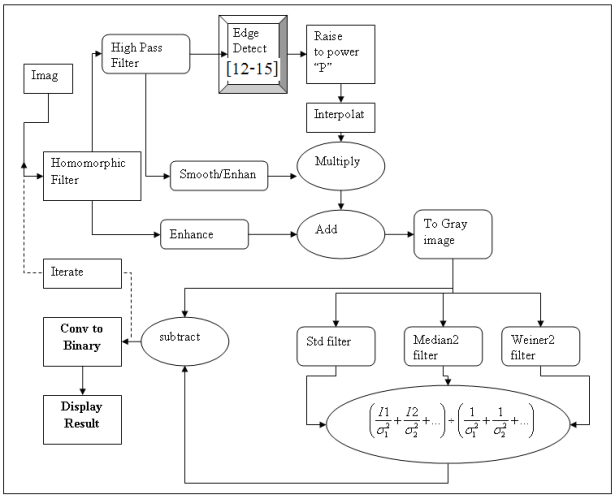

Figure 3. Image Enhancement algorithm.

2. Transformation: In this step a suitable transformation is performed to allow the process of extraction of the feature vectors of choice. The Discreet Wavelet transform "db4" (DWT) was used.
3. Feature vector extraction: The feature vectors of choice are calculated off-line and stored for each image so efficient computation is not a bottleneck for the matching process in the query stage.
4. Database storage: The library PCB images features vectors are stored for each image (shown in Figure 4).

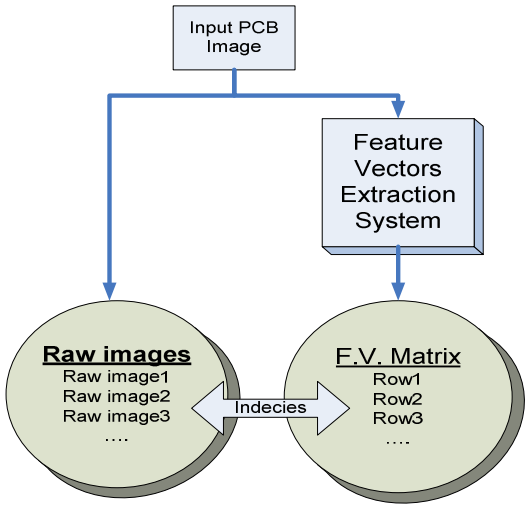

Figure 4. Database Storage.

B. *Background and Feature Vector Elements*

Template matching is the process of determining the position of a template inside an image. Invariant based image recognition is an important problem in pattern recognition. The key is to find out the invariant features which are robust to geometrical transformations such as translation, rotation, size changes and reflection. Invariant moments have become a classic tool for recognition in recent years [16]-[19]. Since the conception of invariant moment emerged in 1962, several new moments have brought forward to improved efficiency.

Since these moments are designed to capture global information about the image, they are not suitable for classifying similar object when corrupted by a significant amount of random noise. In contrast wavelet moment invariants [18]-[19] presented for capturing both global and local information from the objects of interest and a method of selecting discriminative features based on a set of discriminative measures defined for the features. The Invariant moments used were a set of seven 2-D moment invariants that are insensitive to translation, scale, mirroring and rotation [20]-[25].

The system was implemented with invariant moments extracted from the image spatial domain and from the DWT "db4" transform with two level of decomposition.

In special case Hu defined seven values computed from central moments through order three, that are invariant to object scale, position, and orientation. The seven moment invariants are given by:

$$\phi_1 = \eta_{20} + \eta_{02}, \qquad \phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\times\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

$$\phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}\left[(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})\right]$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})^2\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\times\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

It is also useful to normalize the magnitude of invariants by using a statistical approach:

$$\Psi_i = \frac{\phi_i - m_i}{\sigma_i}$$

Where m_i and σ_i are mean and the standard deviation of the i-th moment invariant, respectively

C. *Target image processing component*

The target images go through necessary enhancement, proper transformation, and feature vectors extraction. This is similar to the process that library PCB images went through.

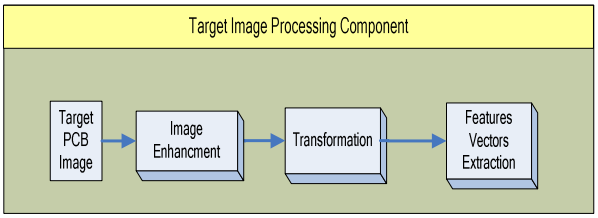

Figure 5. Target Image Processing Component.

The functions of the target image processing system are as follows:

1. Image enhancement: The image passes

through an image enhancement algorithm to stress the important features of the image to ease the process of automatic image recognition; the used method has to be similar to the enhancement method used in the feature vector extraction system.

2. Transformation: The same transformation technique that has been used to create the PCB images library feature vector matrix is applied on the target image.
3. Feature vector extraction: The features of choice are computed for the target image to be used for the comparison with the library feature vectors matrix rows.

D. Query component.

The inputs of the query component (shown in Figure 6) are composed by the PCB library images feature vector matrix, the target image feature vector and the matching threshold. It computes the distance between the target image feature vector and the rows of the matrix and reports the images with distance below input matching threshold.

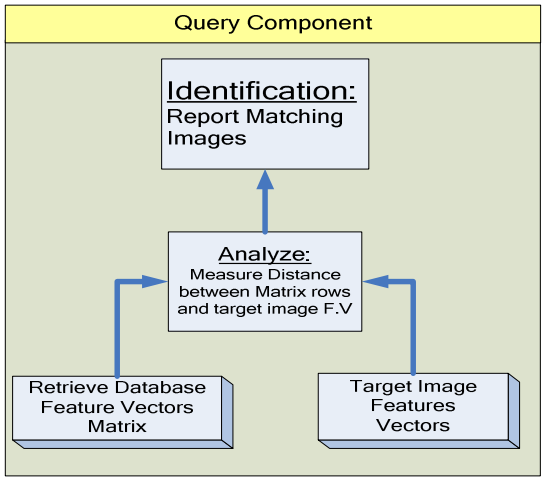

Figure 6. Query Component.

For a match to be reported the distance between the target image feature vector and the library image feature vector should be less or equal to the selected threshold.

III. System Implementation Results and Evaluation

A library of large set of images was constructed and processed by the database storage component to extract the feature vector matrix. Figure 7 shows a snapshot of 25 images from this library. A block was extracted from one of the images; the following target images were constructed based on the selected block:

- The original extracted block as is.
- A scaled down version with half the size of the original block.
- A mirrored image of the block.

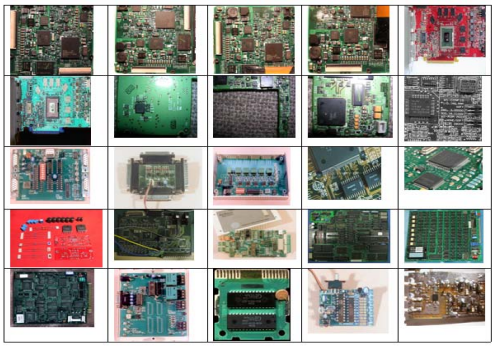

Figure 7. Sample PCB Images from the Library.

1. FFT feature vectors: The "dc" component of the FFT, standard deviation of the magnitude of the FFT, average of the FFT coefficients less than the coefficients mean, average of the FFT coefficients less than the 1.5 multiplied by the coefficients mean and average of the FFT coefficients less than the 2 multiplied by the coefficients mean.
2. FFT feature vectors with invariant moments in the FFT domain: The coefficients computed in step 1 in addition to the invariant moments of the spectrum magnitude coefficients of the FFT transform.
3. DWT feature vectors: Mean of the approximation coefficients, standard deviation of the approximation coefficients, mean of the horizontal detail coefficients, standard deviation of the horizontal detail coefficients, mean of the vertical detail coefficients, standard deviation of the vertical detail coefficients, mean of the diagonal coefficients and standard deviation of the diagonal detail coefficients.
4. DWT feature vectors with invariant moments in the DWT domain: The coefficients computed in step 3 in addition to the invariant moments of the approximation coefficients.
5. Image invariant moments combined with DWT feature vectors with invariant moments in the DWT domain: The invariant moments of the input image, invariant moments of the DWT first level approximation coefficients and the invariant moments of the DWT second level approximation coefficients.

For each of the above methods three sets of images were selected to run the system: a set of 9 images, a set of 24

images and the full library of images with 80 images. The target image was a block extracted from one of the images in its original form, then it was rotated and scaled to create another two target blocks.

The first method failed to match the target extracted block with the image it was extracted from as shown in Figure 8.

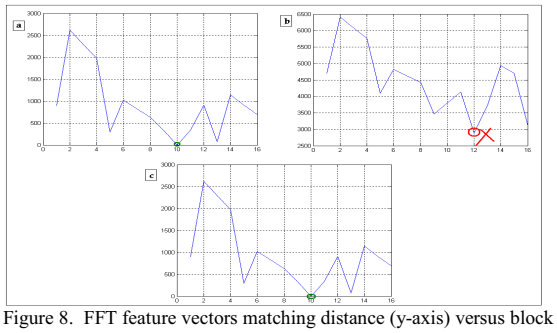

Figure 8. FFT feature vectors matching distance (y-axis) versus block number (x-axis).
(a) Original block matching chart, (b) Half-sized block matching chart block, (c) Mirrored block matching chart.

The second method matched the target block within the image it was extracted from but it failed with a library of 9 images as shown in Figure 9.

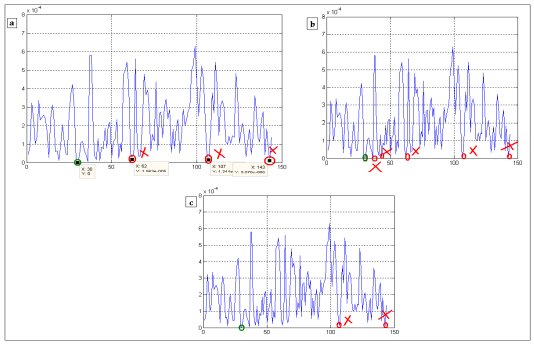

Figure 9. Nine images experiment using FFT and Invariant Moments feature vectors matching distance (y-axis) versus block number (x-axis).
(a) Original block matching chart, (b) Half-sized block matching chart block, (c) Mirrored block matching chart.

The third method failed to match the block within the image it was extracted form as shown in Figure 9.

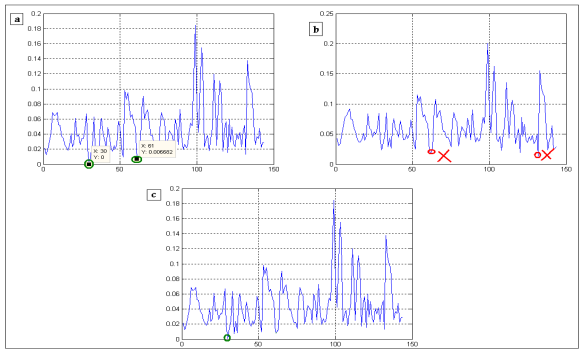

Figure 10. DWT feature vectors matching distance (y-axis) versus block number (x-axis).
(a) Original block matching chart, (b) Half-sized block matching chart block, (c) Mirrored block matching chart.

The fourth method matched the target block within the image it was extracted from but it failed with a library of 9 images as shown in Figure 10.

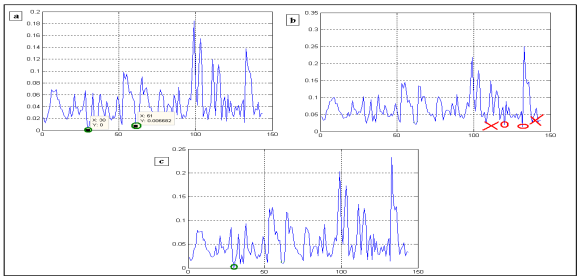

Figure 11. Nine images experiment using DWT and Invariant Moments feature vectors matching distance (y-axis) versus block number (x-axis).
(a) Original block matching chart, (b) Half-sized block matching chart block, (c) Mirrored block matching chart.

The fifth method was able to match the target block irrespective of how many images contained in the library, as well as rotated or half sized target blocks. Figure 12 shows the matching results for a library of 9 images, Figure 13 shows the matching results for a library of 24 images and Figure 14 shows the matching results for a library of 80 images.

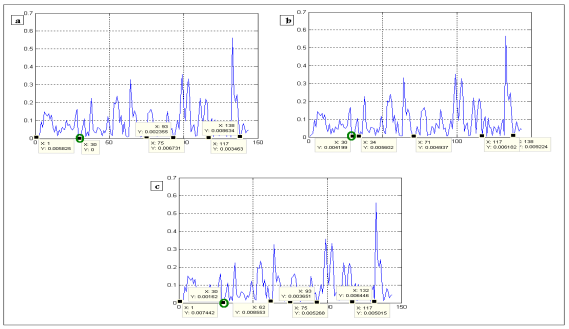

Figure 12. Invariants moments of image combined with DWT and invariant moments of DWT feature vectors matching distance (y-axis) versus block number (x-axis) for a library with 9 image
(a) Original block matching chart, (b) Half-sized block matching chart block, (c) Mirrored block matching chart.

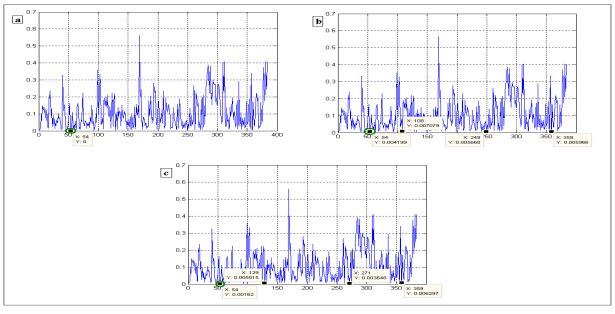

Figure 13. Invariants moments of image combined with DWT and invariant moments of DWT feature vectors matching distance (y-axis) versus block number (x-axis) for a library with 24 image.
(a) Original block matching chart, (b) Half-sized block matching chart block, (c) Mirrored block matching chart.

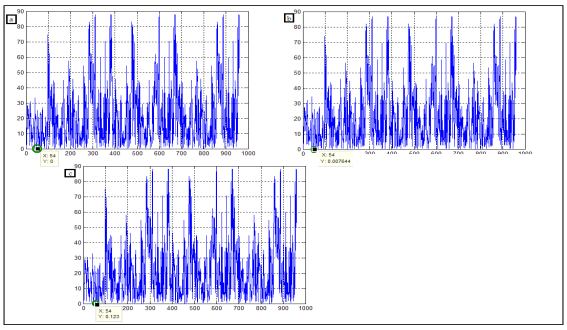

Figure 14. . Invariants moments of image combined with DWT and invariant moments of DWT feature vectors matching distance (y-axis) versus block number (x-axis) for a library with 80 images.
(a) Original block matching chart, (b) Half-sized block matching chart block, (c) Mirrored block matching chart.

IV. CONCLUSIONS

A forensic tool has been presented that allows a software or hardware system to be built that is capable of probing a large database of digital images of circuit boards and compare them for similarities to provide investigation leads for electronic crimes digital forensic science investigations. The developed system has been implemented and proved to be very efficient in detection similarities between a target image and a large image database even when the target image is noisy, scaled or mirrored. The results presented here are highly promising. Therefore, it is possible to say that this proposed program is useful in many other cases. In the near future, we will perform experiments for the real compounding of medicines.

REFERENCES

[1] AAFS. (1996) "What is Forensic Science?" [online], American Academy of Forensic Sciences, http://www.aafs.org/

[2] Eoghan Casey , Handbook of Computer Crime Investigation: Forensic Tools and Technology, Second Edition, Academic Press, 2003.

[3] Department of Justice, computer crime and intellectual property section, [Online]. Available: http://www.usdoj.gov/criminal/cybercrime/searching.html

[4] Meyers, M. & M. Rogers. (2004) "Computer Forensics: The Need for Standardization and Certification.". International Journal of Digital Evidence, 3(2 12] Hannan, M. (2004).

[5] Marcus K. Rogers and Kate Seigfried, The future of computer forensics, Computer and Security, 2004.

[6] Digital Forensics: Defining a Research Agenda System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on, pp. 1 - 6, Jan 2009.

[7] Vaughn, R. and Dampier, D., "Digital Forensics--State of the Science and Foundational Research Activity System Sciences," *HICSS 2007 40th Annual Hawaii International Conference*, pp. 263 – 263, Jan. 2007.

[8] R. Sasaki, Y. Ashino, T. Masubuchi, "A Trial for Systematization of Digital Forensics and Proposal on its Required Technologies", JSSM Vol. 20.2, 2006.09 (pp 46-61)

[9] Brian Carrier, File System Forensic Analysis, Addison Wesley, March 2005.

[10] A. Materka and M. Strzelecki , "Texture Analysis Methods – A Review," *Technical Report*, University of Lodz, Institute of Electronics, 1998.

[11] Sander Janssen, *Text and Photograph Classification*, Masters Thesis, University of Nijmegen, November, 2002.

[12] Agaian, S.S.; Panetta, K. and Grigoryan, A.M, "Transform-based image enhancement algorithms with performance measure," *Image Processing, IEEE Transactions,* Vol. 10, nn. 3, pp. 367-382, March 2001.

[13] Blair Silver, Sos S. Agaian and Karen A. Panetta, "Contrast Entropy Based Image Enhancement and Logarithmic Transform Coefficient Histogram Shifting," *Proceedings, IEEE ICASSP 2005*. March 2005.

[14] Eric Wharton, Sos Agaian, and Karen Panetta, "A Logarithmic Measure of Image Enhancement," *SPIE Defense and Security Symposium*, April 2006.

[15] Eric Wharton, Sos Agaian, and Karen Panetta, "Comparative Study of Logarithmic Enhancement Algorithms with Performance Measure," *Proceedings of Electronic Imaging,* January 20.

[16] M. Teague, "Image analysis via the general theory of moments", J.Opt.SocAmer, Vol7O, No 8, pp.920-930, 1980.

[17] Whoi-Yul Kim, Yong-Sung Kim, "A region-based shape descriptor using BP Wavelet Zernike moments" Signal Processing: Image Communication, Volume Neural Networks Neural Networks 16,Pages 95-102, 2000,

[18] Shen D, Horace H S Ip." Discriminative Wavelet Shape Descriptors for Recognition of 2-D Patterns"[J]. Pattern Recognition, 1999,32(2):151-165.

[19] Ye-Zheng Chun; Lin-Hong Ji; Face detection based on SCNN and wavelet invariant moment in color image Wavelet Analysis and Pattern Recognition, 2007. ICWAPR '07. International Conference on Volume 2, 2-4 Nov. 2007 Page(s):783 – 787

[20] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, 2nd Edition, Prentic Hall, 2002.

[21] Rafael C. Gonzales, Richard E. Woods and Steven L. Eddins, *Digital Image Processing using Matlab*, Prentic Hall, 2004.

[22] Y. Li, "Reforming the theory of invariant moments for pattern recognition", Pattern Recognition Vol25, No.7, pp.723-730, 1992.

[23] Chee-Way Chong, P. Raveendran, R. Mukundan, "Translation and scale uinvariants of Legendre moments", Pattern Recognition, Vol37, ppl 19-129,

[24] Ziliang Ping, Haiping Ren, Jian Zou, "Generic orthogonal moments:Jacobi-Fourier moments for invariant image description", Patternrecognition, Vol40, ppl245-1254, 2007

[25] N.K.Kamila, S.Mahapatra, S.Nanda, "Invariance image analysis using modified Zernike moments", Pattern Recognition Letters, Vol26, pp747-753, 14 32.5 3214 21.2 2268