Hierarchical dictionary-based technique for face recognition system

Mohammed Saaidia and Messaoud Ramdani

October 6, 2019

Hierarchical dictionary-based technique for face recognition system

Mohammed SAAIDIA[1] and Messaoud RAMDANI[2]

[1]*Electrical Engineering Department, University MCM of Souk-Ahras*
Souk-Ahras 41000, Algeria

mohamed.saaidia@univ-soukahras.dz

[2]*Electronics Department, University Badji-Mokhtar of Annaba*
BP 12, Annaba 23000, Algeria

messaoud.ramdani@univ-annaba.org

Abstract— **Face recognition still the most studied topic in the pattern recognition research field. This is probably due to the multiple useful applications which can be developed for important domains like HMC, Forensic, Security, Entertainment etc. Such deployed research efforts produced a huge number of methods, techniques and algorithms with different characteristics according to their simplicity, efficiency, robustness and speed. In the present work, we investigate the performances of a simplified technique using a hierarchical classification scheme based on a constructed multi parts dictionary of elementary blocs obtained by applying sequential classifier to the whole set of features of the working database.**
The elementary features of the different parts of the constructed dictionary were obtained using the well-known and simplest and effective way to depict the similarities between two compared patterns; named cross-correlation operator applied to the original images and their transformed images known as integral images. Hierarchical clasification scheme is used to overcome the fact that this operator has high consumption time cost.
The proposed strategy was implemented and tested on the images of the well known ORL and YALE database sets. Practical results demonstrate largely recognizable efficiency and speed characteristics.

Keywords— **Face recognition, Cross Correlation, Integral image, K-means classifier, Hierarchical structure.**

I. Introduction

Based on the advancements in both hard and soft technologies in computer sciences and the huge developments in applied mathematics domain, researchers in pattern recognition domain, in general, and face processing, in particular, have succeeded to exploit computer performances and new mathematical models to deploy a large number of methods, techniques, and algorithms to try to solve one of the most challenging problems of pattern recognition field namely the face recognition one.

As a pattern, the “face” was subject to a particular concern by the researchers according to its importance in day-life applications. Face processing concern can be divided into two principal axes; face analysis and face synthesis. The last one was exploited in applications related to entertainment domain like animated cartoons, special effects in movies, etc. The former one can be subdivided into three principal research subjects namely face detection, facial expression recognition, and face recognition.

Face recognition had already crossed laboratory doors and is on route to be standardized through common applications in security domains like face verification or face authentication for institutions, factories, airports, network applications, etc. Scientific literature is full of methods, techniques, and algorithms developed for this type of face processing. Researchers have exploited different types of mathematical models like time processing, spectral processing, scale processing, statistical models and analysis and also took advantages of new computer technologies like speedy processors, parallel processors, DSP (Digital Signal Processors) and so on to develop more and more efficient but also more complex algorithms. However, despite these advancements and enhancements, important challenges still remain to be overcome.

According to literature and especially review papers [1], [2], face recognition methods can be classified into three approaches; holistic approaches, feature-based approaches, and hybrid approaches. The first type of methods is based on the detection and interpretation of the whole face as a unique component. Different types of information, like spatial information, spectral information, scale information, temporal information etc, are then extracted, characterized and classified to take a decision on the processed face identity. The second one deals with the face’s components like eyes, brows, nose, mouth, etc to extract spatial, spectral, scale or temporal information to extract different types of features to be used in characterization and classification phases to take a decision. In such types of methods, one or some components (e.g the eyes and the mouth) can be taken as more valuable against others at the decision phase. By extracting both local and global information, the hybrid approaches try to take advantages of the two former ones to enrich the feature vectors used in classification phase.

In the present work, we try to reuse And exploit the simplest and most effective operator for depicting similarities between two compared patterns; named cross-correlation operator. Indeed, this operator is one of the oldest statistical mean used in signal processing and pattern recognition studies to extract the level of similarity between two studied signals or patterns. However, the time consumption cost involved using this operator still so important that it becomes a real handicap especially for signals or patterns that demand a large amount of processed data, which is the case of image processing in general and face recognition in particular. The following of this paper is subdivided in four sections; firstly, in section 2, we will give an overview of the Cross-correlation operator; its theoretical basis and the different ways in which it was used, then in section 3, we will expose our proposed technique and in section 4, the results recorded applying this technique are given. Finally, section 5 will be dedicated to the conclusions and perspectives of the present work.

II. PROCESSING TOOLS

In this section we are going to expose the two principal processing tools used in our proposed technic, namely the cross-correlation operator and the integral image computing artifice.

A. *Cross-Correlation operator*

Cross-Correlation operator belongs to the operators firstly developed by mathematicians to study the statistical characteristics of a time-series recorded data [3], [4]. Equations (1) and (2) give, respectively, the continuous and discrete versions of this operator applied to the one-dimensional functions *f* and *g.*

$$f(t) * g(t) = \int_{-\infty}^{+\infty} f^*(\tau).\, g(\tau + t)\, d\tau \quad (1)$$

$$f(n) * g(n) = \sum_{m \to -\infty}^{+\infty} f^*(m).\, g(m + n) \quad (2)$$

$* \; denotes\ the\ complexe\ conjugate$

Because images are considered as two dimensional, we give in the equation (3), the cross-correlation operator applied to two-dimensional discrete functions *f1* and *f2*.

$$f(n1, n2) * g(n1, n2) = \sum_{m1 \to -\infty}^{+\infty} \sum_{m2 \to -\infty}^{+\infty} f^*(m1, m2).\, g(m1 + n1, m2 + n2) \quad (3)$$

Acording to equation (3), it's clear that the huge quantity of processed data will make the processing time-cost so high that using such operator becames unpractical especially for real-time applications.

Despite that indisputable fact, researchers have used and still largely using this operator in different types of data processing and especially for pattern recognition problems. Recently, it was used by Gubta et all. [5] for speech recognition and also by Farahani et all. [6] for the construction of a robust speech recognition system. It was also exploited by Hamici [7] and Chadha et all. [8] to elebarate efficient algorithms for speaker recognition. For biomedical signals' analysis, we can find researchers who exploited this type of operator like Morgado et all. [9] and Olav et all. [10] for electrocardiograms, and Jensen et all. [11] for blood velocity estimation, etc. For two-dimensional pattern recognition, cross-correlation operator had been largely investigated for almost all the applied problems in this domain. For object detection, William [12] had used a modified version of the primary operator applied to detection of projection views of biological macromolecules in electron micrographs. Wang et all. [13] have applied the operator on adjacent A-scans to extract reflection hyperbola which were used to detect underground objects. Al-Bakry et all. [14] applied the operator in the frequency domain between the input image and the weights of neural networks for fast face detection. Greche et all [15] and Kapoor et all. [16] elaborated efficient algorithms for facial expression recognition respectively based on histogram of oriented gradient and normalized cross-correlation, for the first, and using concepts of cross-correlation and Mahanalobis distance for the second ones.

For face recognition also, cross-correlation operator was and still be used for this challenging problem in images and video. Recently, Parchami et all. [17] and Rosales et all. [18] have separately proposed a cross-correlation-based system to detect individuals in video. The first one was applied to video over a distributed network of cameras and the second one to frames randomly choosen from a well known database. Fattah et all.[19] repported simple and practical implementation with a high degree of face recognition accuracy by appliying such operator to DCT transformed images to recognize faces on standard databases.

B. *Integral Image*

firstly introduced by Crow [20] under the name of "Summed-area tables", the "integral image" was used in several researches especially for the construction of the well known Viola-Johns [21] algorithm for face detection and face's features detection. The inherent process involved by integral image allow to highlight the important features present in a particular region and to smooth the closest ones. This important fact represents a fast manner to calculate the different patterns that will be used to characterize the original image. Thus, the processed image I(x',y') will be firstly transformed to a new image $I_i(x,y)$ called integral image according to equation (4).

$$I_I(x, y) = \sum_{x'=1}^{x} \sum_{y'=1}^{y} I(x', y') \quad (4)$$

Each one of the Ii(x,y) values represents the algebric sum of all the pixels at its left and top in the processed image I(x',y'). Mathematically speaking, it's the double integral of the original image for x'€ [1,x] and y' € [1,y].

The recurences given in equations (5) and (6) represents a fast algorithm to obtain Ii(x,y) [31].

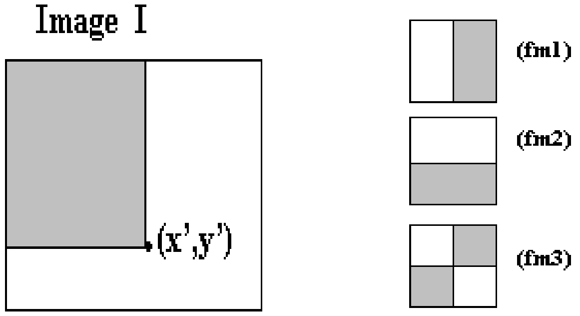
$$\mathrm{S(x, y) = s(x, y - 1) + I(x, y)} \quad (5)$$

$$\mathrm{Ii(x, y) = Ii(x - 1, y) + s(x, y)} \quad (6)$$

$With \quad s(x, -1) = 0 \;\; and \;\; I_i(-1, y) = 0$

Figure 1.a gives a symbolic scheme for compiling Ii(x,y). Compiled Ii(x,y) was exploited in different ways by researchers according to their investigated methods, the studied problems and the expected results.

In our present paper, we were inspired by the work of Viola-Jones especially the way to make choices on the patterns used to scan the processed image. These patterns are same as the basis of Haar functions used by Papageorgiou et al. [22]. Figure 1.b presents those retained in our work for the characterization of the original image.



(a) (b)

Figure 1. (a) Symbolic scheme for integralimage calculation (b) Elementary patterns retained for image charactérization

These elementary patterns were used with two different resolutions for each one; fm1 (1x2 and 6x6), fm2 (2x1 and 6x6) and fm3 (2x2 and 6x6).

Thus, for each pixel of the original image, 06 characteristic features (03 for each resolution) representing its interaction with its neighborhood are produced. Figure 2 gives an example of applying these elementary patterns to an image from the YALE database.

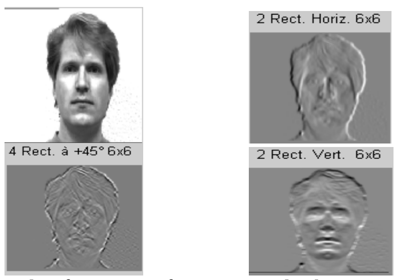


Figure 2. Example of an image from YALE database processed thrue the three retained elementary patterns at a resolution of 6x6 pixels.

III. PROPOSED PROCESSING METHODOLOGY

Our proposed approch is based on the construction of a Global Multi-Parts Dictionnary (GMPD) that handles the elementary unique blocs of the whole images of the working database. This choice was dictated by the fact that for real applications in both face identification and face authentication, the database remains unchanged for a set of persons. And even if we need to add new people to the database, the extension remains possible through the addition of new characteristic blocks specific to the added persons.

A. Global Multi-Parts Dictionary construction

The Multi-Parts (MP) characteristic of our dictionary is related to the fact that all the images of the working database are initially subdivided in separated blocks or regions and for each block we will dedicate a separate sub-dictionary. In Figure 3 we gave the first two stages in the construction of the GMPD. Thus, the face is initially detected then subdivided in blocks according to two manners; longitudinal blocks (row3) or concentric regions (row4). In the following section, we will repport and compare the results obtained for each one of the subdivision strategies.

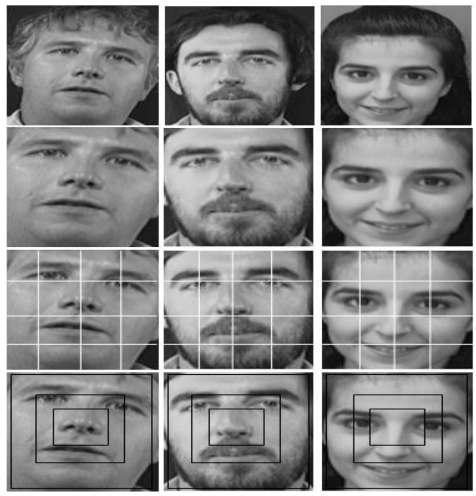


Figure 3: Example of sequential phase of dictionary construction. row1: orogonal image, row2: detected faces, row3: longitudinal subdivided regions on detected faces, row4: concentric subdivided regions on detected faces

Conducted experiments of the proposed method directly on the images had demonstrate insufficient performances. Thus, we was directed to try to enrich the elementary block patterns by introducing integral image information which had the ability to highlight the shape and proximity relationships of the pixels of the same block or area. So, both original images and their corresponding integral images were subdevided in the same way.

The third step was dedicated to the creation of the subdictionaries by subdividing each one of the initially obtained blocks at the primary subdivision step.

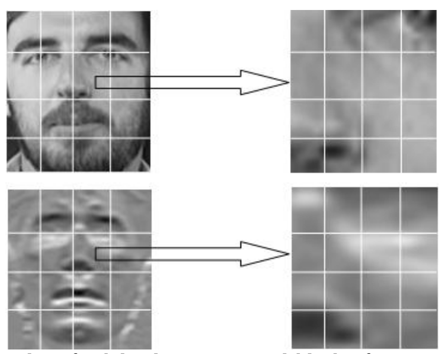


Figure 4: Examples of subdividing a principal block of an original image from ORL database and an integral image from the YALE database

The last step of the construction phase is dedicated to the compilation of the inter-similarities matrix for each subblock of the primary part of the dictionary. Thus, for each subblock we compile the cross-correlation between all the images' subblocks of the same person then between those of different persons. This procedure allow the calculation of the thresholding cuve which will be used by the hierarchical classifier to construct the subdictionary of each subblock, then the dictionary of each face's part. The Figure 5 gives a schematic representation of the GMPD.

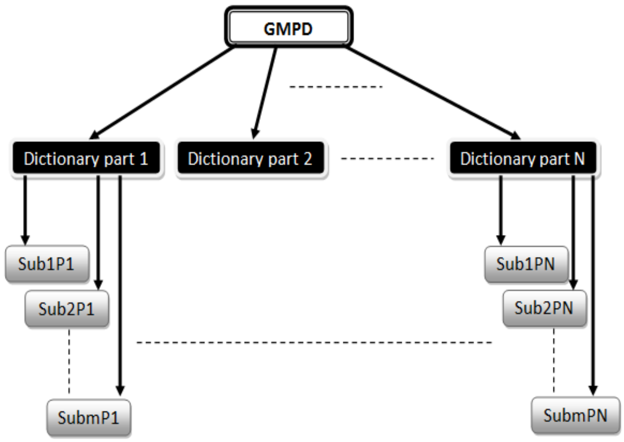


Figure 5: structure of the constructed GMPD

At the root of the GMPD, each image of each person of the working database have a hierarchical code that represents the numbring of the different subblocks of the different parts in the correspending subdictionaries.

The construction of the GMPD needs a heavy compilation process caused by the huge quantity of information processed but fortunately it have to be done offline and only once for the working database unless if there is an update to perform. It's also important to indicate that the number of elementatry subblocks retained in the GMPD is drasticaly reduced compared to the initial number of subblocks representing the whole number of images in the working database.

B. *Global Multi-Parts Dictionary using*

Having the GMPD of a working database, we can perform tests or make real online exploitation according to the procedure described hereafter:

Given a recorded image or taking an online photo of a person

- Firstly, we make detection of the face in the corresponding image.
- We compile the integral image of the deteted face.
- Then we make partitions of the detected face and its integral image according to the same way (i.e the same manner of partitioning and with the same number of partitions) that it was done during the construction of the GMPD.
- Each one of the image partitions is then subdivided according to the same way that it was done during the construction of the subdictionaries.
- For each subblock, we compile the similarity matrix applying correlation operator.
- For each image partition, we compile its corresponding sub-code then the global code of the processed image by concatinating the different subcodes.
- The obtained code is then used to identify the processed personne by comparaison with the recorded codes at the root of the GMPD.

IV. Experimental results

The proposed technique was implemented in different ways and tested on two different standard database YALE and ORL. Images in both database presents personnes with expressions variability; luminance exposer differences, male and female gender and head position variability. Yale database is a recording set of 15 persons with 10 images each. ORL database is a recorded set of 40 persones with 10 images each.

The fellowing results were recorded for only one type of subdividing images; longitudinal subblocks. Faces in images were detected and partitioned in 16 partitions each. Each partition was then subdivided in 16 subblocks and then in 32 subblocks. 10%, 20%, 30% and 40% of the images (and consequently of the subblocks) were sequentially keeped for testing porpouse (2 images for each person). All recorded results are the means of repeating the same experiment with same parameters at least 10 times.

A. *Obtained results on YALE database*

On the Table 1, we give the results of using the constructed GMPDs to perform tests on the remaing images.

Table 1: Recorded TPRs for different ratios of testing images and using 3 subdividing configurations on YALE database

YALE database	*Remaining test images*	*TPR in %*
*Sub-blocks with **64x64** pixels*	10 %	**88,67**
	20 %	**89,56**
	30 %	**86,74**
	40 %	**81,19**
*Sub-blocks with **64x32** pixels*	10 %	**89,31**
	20 %	**89,98**
	30 %	**87,45**
	40 %	**81,23**
*Sub-blocks with **32x64** pixels*	10 %	**88,23**
	20 %	**90,15**
	30 %	**87,89**
	40 %	**82,73**

According to above parameters and for 16 subblocks in each partition; there was a set of 76800 (2x150x16x16) sublocks for the whole database. Each subblock was 64x64 pixels. 61440 subblocks were processed to construct the GMPD. At the end, only 13056 subblocks were retained to form the different levels of the GMPD.

for 32 subblocks in each partition; there was a set of 153600 (2x150x16x32) sublocks for the whole database.

Each subblock was 64x32 pixels. 122880 subblocks were processed to construct the GMPD. At the end, only 17340 subblocks were retained to form the different levels of the GMPD.

In both cases, the retained subblocks were not uniformly distributed across the different partitions.

According to results reported on Table 1, we can firstly conclude for the validity of the proposed technique. We can also see that for sub-blocks with reduced size, the TPRs are generally better than those obtained for sub-blocks with greater size.

B. Obtained results on ORL database

On the Table 2, we give the results of using the constructed GMPDs of the ORL database to perform tests on the remaing images.

Table 2: Recorded TPRs for different ratios of testing images and using 3 subdividing configurations on ORL database

ORL database	*Remaining test images*	*TPR in %*
*Sub-blocks with **64x64** pixels*	10 %	**92,12**
	20 %	**92,34**
	30 %	**89,11**
	40 %	**86,47**
*Sub-blocks with **64x32** pixels*	10 %	**92,29**
	20 %	**93,57**
	30 %	**93,01**
	40 %	**88,33**
*Sub-blocks with **32x64** pixels*	10 %	**91,12**
	20 %	**93,34**
	30 %	**91,11**
	40 %	**89,73**

Results given on Table 2 demonstrate that the proposed technique gives better results with the images of the ORL database. This will be due to firstly the better quality of the registred images in this database compared to those of the former one and secondly to the larger number of persones and images involved which gives an enriched variation which permits to obtain a more representative GMPD. About the subdividing configurations, the same conclusions as those of the YALE database will be retained.

C. Obtained results on a combined YALE-ORL database

Trying to obtain a greater database with more variabilities we construct a combined working database with the two primary sets of images. Obtained results with the same conditions are recorded on Table 3.

Table 3: Recorded TPRs for different ratios of testing images and using 3 subdividing configurations on Combined database

ORL and YALE database	*Remaining test images*	*TPR in %*
*Sub-blocks with **64x64** pixels*	10 %	**90,66**
	20 %	**93,17**
	30 %	**92,55**
	40 %	**91,34**
*Sub-blocks with **64x32** pixels*	10 %	**91,10**
	20 %	**93,88**
	30 %	**93,23**
	40 %	**91,90**
*Sub-blocks with **32x64** pixels*	10 %	**91,23**
	20 %	**93,09**
	30 %	**92,56**
	40 %	**91,81**

Obtained results in this case are better than those obtained for each individual database. Results appear to be more stable for the different subdividing configurations despite the fact that the best ones are those recorded for the smaller sub-blocks.

V. Conclusion

In the presente paper we have introduced, explained, tested and discussed results of a proposed dictionary-based method for face recognition. The studied technique is based on the construction of a hierarchical Global Multi-Parts Dictionary of elementary sub-blocks obtained from the original images and their corresponding integral images processed thrue the cross-correlation operator as a statistical mesure of inter-blocks similarity.

Recorded results for differents test parameters make proof for the validity of the proposed technique and highlight the sensitive parameters like the sub-block size and the working database.

According to preliminary presented results, it appears clearly that this technique needs more deep investigations especially for its sensitivity to the size of the elementary sub-block, the working database, the similarity measurement criterian etc.

References

[1] Yongzhong Lu, Jingli Zhou, Shengsheng Yu, “A survey of face detection, extraction and recogntion”, International Journal of Computing and Informatics. Vol . 22, 2003

[2] R. S. Ghiassy, O. Arandjelovi, H.Bendaday, and X. Maldague,‗Infrared Face Recognition: A Literature Review‘, International Joint Conference on Neural Networks, 2013

[3] Bracewell, R. "Pentagram Notation for Cross Correlation." The Fourier Transform and Its Applications. New York: McGraw-Hill, pp. 46 and 243, 1965.

[4] Papoulis, A. The Fourier Integral and Its Applications. New York: McGraw-Hill, pp. 244–245 and 252-253, 1962.

[5] Pankaj Raibagkar, Anup Palsokar Speech Recognition Using Correlation Technique Anjalika Gupta1, International Journal of Current Trends in Engineering & Research (IJCTER) -ISSN 2455–1392 Volume 3 Issue 6, June 2017 pp. 82 – 89

[6] Farahani1, S.M. Ahadi1 and M.M. Homayounpour, "Improved autocorrelation-based noise speech recognition using kernel-based cross-correlation and over-estimation parameters G". 15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007, copyright by EURASIP

[7] Speaker Recognition Using Spectral Cross-correlation: A Fast Algorithm Journal of Computer Science, Science Publications, issue, April 15, 2006;

[8] Aman Chadha, Divya jyoti and M. Mani Roja, "Text-Independent Speaker Recognition for Low SNR Environments with Encryption" International Journal of Computer Applications (0975 – 8887) Volume 31– No. 10, October 2011 pp. 43-50

[9] Morgado E., Alonso-Atienza F., Santiago-Mozos R., Barquero-Pérez Ó., Silva I., Ramos J., et al. . (2015). Quality estimation of the electrocardiogram using cross-correlation among leads. Biomed. Eng. Online 14:59. 10.1186/s12938-015-0053-1.

[10] Olav E. Liseth Daniel Mo Hakon A. Hjortland "Power-Efficient Cross-Correlation Beat Detection in Electrocardiogram Analysis Using Bit streams" IEEE Transactions on Biomedical Circuits and Systems vol. 4 no. 6 December 2010.

[11] J. A. Jensen I.R. Lacasa "Estimation of blood velocity vectors using transverse ultrasound beam focusing and cross-corcelation" Proc. IEEE Ultrason. Symp pp. 1493-1497 1999.

[12] William V. Nicholson "Object Detection by Correlation Coefficients Using Azimuthally Averaged Reference Projections" IEEE Transactions on Biomedical Engineering vol. 51 no. 11 November 2004.

[13] Jian Wang ; Yi Su, "Underground object detection based on cross correlation and Hough transform", Microwaves, Radar and Remote Sensing Symposium (MRRS), Kiev, Ukraine, August 2011, DOI: 10.1109/MRRS.2011.6053674.

[14] El-Bakry, H.M. & Zhao, Q. , "Fast Pattern Detection Using Normalized Neural Networks and Cross-Correlation in the Frequency Domain", EURASIP Journal on Advances in Signal Processing (2005): Vol. 13, pp. 2054-2060, DOI: 10.1155/ASP.2005.2054

[15] Latifa Greche ; Najia Es-Sbai, "Automatic system for facial expression recognition based histogram of oriented gradient and normalized cross correlation", International Conference on Information Technology for Organizations Development (IT4OD), March-April 2016, Fez Morroco, DOI: 10.1109/IT4OD.2016.7479316.

[16] Supriya Kapoor, Shruti Khanna and Rahul Bhatia, "FACIAL GESTURE RECOGNITION USING CORRELATION AND MAHALANOBIS DISTANCE", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7,Issue 2, pp. 267-272, 2010.

[17] Mostafa Parchami ; Saman Bashbaghi ; Eric Granger, "CNNs with cross-correlation matching for face recognition in video surveillance using a single training sample per person", 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2017, Aug-Sept 2017, Lecce, Italy.

[18] Edward Rosales ; Yun Tie ; Anastasios Venetsanopoulos ; Ling Guan, "Automatic face recognition from video sequences using a template based cross correlation method", 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), May 2013, Regina SK, Canada, DOI: 10.1109/CCECE.2013.6567723

[19] S. A. Fattah M. R. Khan A. Sharin and H. Imtiaz "A face recognition scheme based on spectral domain cross-correlation function " TENCON 2011-2011 IEEE Region 10 Conference 2011 pp. 10-13

[20] F. Crow. Summed-area tables for texture mapping. In Proceedings of SIGRAPH, volume 18(3), pp. 207- 212, 1984.

[21] Michael Jones Paul Viola. Robust real-time object detection. Second Interational Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling, 2001

[22] C. Papageorgiou, M. Oren, and T. Poggio A general frame work for object detection. In Internationnal Conference on Computer Vision, 1998.