Face detection using Zernike moments. Evaluation by quantitative measurement criteria

MOHAMMED SAAIDIA[1], SYLVIE LELANDAIS[1], VINCENT VIGNERON[1],
EL-MOULDI BEDDA[2]
[1] IBISC FRE CNRS 2873
Université d'Evry Val d'Essonne
40, rue du pelvoux, CE 1455. 91020 Evry courcouronnes cedex
FRANCE
{saaidia, vvigne}@iup.univ-evry.fr, s.lelandais@iut.univ-evry.fr
[2] LAS
Université de Annaba
BP 12 Annaba
ALGERIE
mouldi_bedda@yahoo.fr

Abstract: Face detection using Zernike moments is presented in this work. Face detection is obtained in two steps. In the first step the image to be treated is presented to a program which calculates its Zernike moments vector with appropriate parameters. In the second step, this reduced vector is used as an entry on the input layer of a neural network, beforehand trained with similar vectors. The neural network provides on its output layer a vector of co-ordinates in (R,θ) representing pixels surrounding the face contained in the treated image. The Zernike moments are used here for there properties of orthogonality and rotational invariability. The experimental results of the application of our method on images of the XM2VTS database are presented using a new objective criteria for measuring performances.

Key words: Image processing, Face detection, Geometrical moments, Zernike moments, Neural network.

1 Introduction:

Face detection in an image or in a sequence of images, is being of great importance in the applications which treat aspects related to the Human-Machine communication. Thus, for face recognition (identity check), analysis of expressions or to take movements of face parts into account (gesture communication); localization of face in image or in video acquired by various peripherals (cameras, scanner, infra-red...) is necessary to the achievement of these operations. Several ways were explored by the researchers. Classification, according to Hjelmas and Low [1], allows distinguishing two principal approaches. The global approach which consists in entirely seeking the face and the components approach which consists in finding the face through the localization and the regrouping of its components (eyes, nose...). According to one or the other of these approaches, each developed method exploits one or more characteristics of face like colour, shape, movement, ...
Method presented in this work exploits geometrical characteristics of the face in a global way. It carries out the operation of face localization through the use of a neural network trained beforehand, on the basis of vectors of geometrical moments, to deliver on its output layer a number of pixels representing a probable contour of the required element.
Geometrical moments are known for their capacity to compress the geometrical information, contained in the image treated, in a rather reduced vector of parameters through the projection of the image on an orthogonal basis [2]. Here in our present work we are particularly interested by the geometrical Zernike moments since in more of their orthogonality, they allow, through simple transformations, to obtain a complete and orthogonal base invariant in rotation, translation and scale [1], [3]. This characteristics makes them very adapted to the training of classifiers who often need, on their input layer, feature vectors reduced in size but rather representative of the element subject to the classification. To measure method performances we propose a new criteria which permit to appreciate more objectively the results obtained though an automatic computation of detection rate.
In section 2 we will explain the formulation of geometrical moments of Zernike and the possibilities of their computation in a fast and

effective way, in section 3 we will introduce the relationship between face detection and Zernike moments and then we will give the description of the method suggested and the way of its implementation. In Section 4 we propose a new quantitative criteria to measure our method performances. Section 5 shows the results and section 6 the conclusion.

2 Formulation and implementation of Zernike moments :

Zernike moments form part of the general theory of the geometrical moments. They were introduced initially by F. Zernike [4]. At the difference of the general geometrical moments, those of Zernike are built on a set of orthogonal polynomials. These polynomials are the basic elements of the construction of an orthogonal base given by the relation **(1)**

$$\mathrm{V}_{\mathrm{n,m}}(\mathrm{x},\mathrm{y}) = \mathrm{V}_{\mathrm{n,m}}(\rho,\theta) = \mathrm{R}_{n,\mathrm{m}}(\rho).e^{\mathrm{j.m.}\theta} \quad \textbf{(1)}$$

where :

$$\begin{cases} \mathrm{R}_{\mathrm{n,m}}(\rho) = \displaystyle\sum_{k=|m|}^{n} \frac{(-1)^{(n-k)/2}.(n+k)!}{(\frac{n-k}{2})!(\frac{k+m}{2})!(\frac{k-m}{2})!}.\rho^{k} \\ \rho=\sqrt{x^2+y^2} \quad \text{and} \quad \theta=arctg(y/x) \end{cases} \quad \textbf{(2)}$$

with: $\mathbf{n} \geq 0$, $\mathbf{m} \neq 0$, $|\mathbf{m}| < \mathbf{n}$, $\mathbf{n} - |\mathbf{m}| < \mathbf{n}$ and (**n-k**) even.

$\mathrm{R}_{\mathrm{n,m}}(\rho)$ is the orthogonal radial polynomial, **n** is the order of the moment and **m** the factor of repetition (the smoothness of the required details) at this order. ρ and θ are respectively the radius and the angle of treated point of the function.

The orthogonality of this base being assured only inside the unit circle, the bi-dimensional function to be projected must be remapped inside it. In the case of a discrete function, image for example, this transformation between the relative co-ordinates (**i**, **j**) of the initial image pixels and ($\mathbf{x_j}$, $\mathbf{y_i}$) of the remapped one can be carried out through **(3)**.

$$x_j=c+\frac{j.(d-c)}{(N-1)} \quad \text{and} \quad y_i=d-\frac{i.(d-c)}{(M-1)} \quad \textbf{(3)}$$

With (**M**,**N**) dimensions of the function to be projected, **i** and **j** indices of the point to be remapped and (**c**,**d**) couple of parameters allowing to remap the function inside, completely (c=-1√2 and d=-c) or partially (c=-1 and d=1), of the unit circle.

The projection of a numerical function on the basis functions of (1) gives the Zernike moments $Z_{n,m}$ according to **(4)**:

$$Z_{n,m} = \frac{n+1}{\pi} \sum_{x_j^2+y_i^2 \leq 1}\sum f(x_j,y_i).\mathrm{V}^{*}_{\mathrm{n,m}}(x_j,y_i) \quad \textbf{(4)}$$

* : denotes the complex conjugate of the function

where : $f(x_j,y_i)$ is the numerical function to be projected, x_j and y_i were defined on **(3),** n and m are respectively the Order and repetition on the Zernike moments.

However, traditional formulation of Zernike moments is very easy to implement but its computational cost is very high and so less adapted for treatments which requires fast execution. This major handicap pushed the researchers to try to find a formulation more suitable to enhance the speed computation [5], [6], [7], [8]. The proposed algorithm in [8] have the advantage of preserving accuracy of computation on the same level as the traditional formulation. To do so, no quantification on the parameters angle and radius (ρ,θ), was introduced. The safeguarding of the precision ensures maintenance of the orthogonality of the build base. We adopted This last formulation in our work. To lead to this form of representation, the preceding equations are rewritten and reorganized as the equation **(5)** shows it.

$$\begin{aligned} Z_{n,m} &= \frac{n+1}{\pi} \sum_{x^2+y^2 \leq 1}\sum \left(\sum_{k=|m|}^{n} \beta_{n,m,k}.\rho^{k} \right) e^{-j.m.\theta}.f(x_j,y_i) \\ &= \frac{n+1}{\pi} \sum_{k=|m|}^{n} \beta_{n,m,k}.\left(\sum_{x^2+y^2 \leq 1}\sum e^{-j.m.\theta}.\rho^{k}.f(x_j,y_i) \right) \\ &= \frac{n+1}{\pi} \sum_{k=|m|}^{n} \beta_{n,m,k}.\mathrm{X}_{m,k} \end{aligned} \quad \textbf{(5)}$$

where: n: order, m:repetition, ρ and θ are respectively the radius and the angle of treated point of the function.

$\beta_{n,m,k} = \frac{(-1)^{(n-k)/2}.(n+k)!}{(\frac{n-k}{2})!(\frac{k+m}{2})!(\frac{k-m}{2})!}$ is a term whose calculation depends neither on the image $f(x_j, y_i)$ nor of its co-ordinates (x_j, y_i) and $\chi_{m,k}$ is the general term which we calculate only once for all the repetitions.
Thus, the equation **(5)** reduces the computing of the Zernike moments of any image to the computing of a linear combination of these two last terms.

According to this formulation, we need only $2.(\frac{n}{2}+1)^2+(M^2-1)$ additions and $\frac{n^2.M^2}{2}+2.M^2.n$ multiplications to compile Zernike moments up to order n for an image of dimension (MxM) [8].

3 Zernike moments for face detection :

On consulting various work on the application of Zernike moments, we noticed that this form of representation was never used for the target detection even when the context of work presented required this operation [9], [10], [11] and [12]. However, the definition and the formulation of the Zernike moments as being parameters able to contain geometrical information of a two-dimensional function and to compress them in a vector with reduced length make it possible to claim with their use in purpose of target image detection. Indeed, the geometrical moments are not abstract parameters. Each one of them have a significance related to the statistical characteristics of the bi-dimensional function which they represent such as the surface, the total mass center, the mass centers on horizontal and vertical directions, horizontal and vertical symmetry, ...etc. Thus, a face by its particular shape and its contents geometrically rich by the details of the elements which it contains (eyes, mouth, eyebrows...) could have certain preponderance on the parameters of the Zernike vector representing an image which contains it. This observation encouraged us to build a method of face detection based on the use of Zernike moments vectors as input to a neural network.
Fig. 1 gives the diagram block of the detection system we propose.

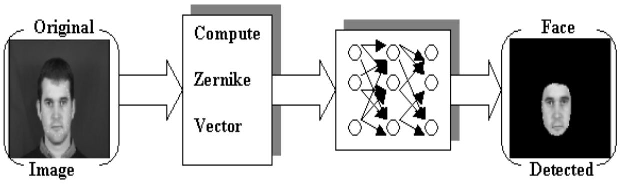

Figure 1 : *General diagram of the system detection*

The operation of face detection is thus done in two steps:

- During the first step, an image is presented to an algorithm which extracts representative Zernike vector.
- At the second phase, a back-propagation neural network, beforehand trained, receives on its input layer the Zernike moments vector. Then the neural network gives on its output layer a set of points representing the probable contour of the face contained in the original image.

The neural network is used to extract statistical information contained in the Zernike moments and in there interactions which are closely related to the area of the required face.
It should be notified here, that we make no assumption on the probable shape of the face subject to detection and no pre-processing operation is required for the image processed.
It is clear that the implementation of our method is mainly based on training phase which we summarize here in four stages:

- Computation of the vectors of Zernike moments for all the images (**N**) in the work database.
- Construction of the training database by randomly pulling up **M** images from the work database (**M**<<**N**) and their corresponding Zernike moments vectors **Zi**.
- Manual delimitation of the face area in each image of the training database by a set of points representing the contour **Ci** of each treated face.
- Training of the neural network on the set of **M** couples (**Zi**,**Ci**).

To test and measure the performances of the network obtained after training operation, we proceed, according to Fig. 1, on the hole (**N**-**M**) images remaining in the work database.

4 Quantitatif measurement criteria :

To give an objective appreciation of the results given by our method we propose a new way to calculate the detection rate based on the relation between the number of pixel correctly and wrongly detected as pixels of the face, the number of all pixels of the treated face and the number of pixels in the image. To do so, all the images of the database were manually segmented in three regions. The first region (white one on the masks of Fig 2) contains the **W** pixels which represents the essential components of the face (brows, eyes, nose, mouth and surrounding pixels). The second region (grey one on the masks of Fig 2) represents the **G** pixels surrounding the first region and belonging to the face. The last region represents all the **B** pixels of the image which are not parts of the face. For the detection system, the first region is one which have to be contained imperatively in the resulting contour and the third one is to be discarded imperatively from it. The second region is optional and has no effect on the calculated result.

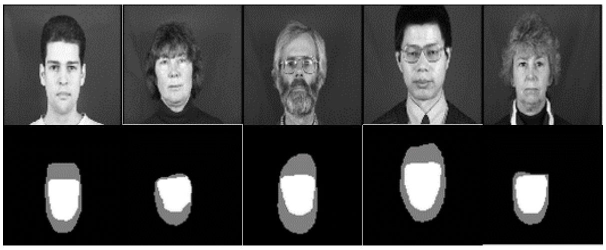

Figure 2: Examples of regions definition. Top: original image, Bottom: mask of regions

We so define two types of rates; Good detection rate (**Gdr**) and Quality detection rate (**Qdr**)

$$Gdr = \frac{W1}{W}.100 \text{ and } Qdr = (\frac{W1}{W} - \frac{B1}{B}).100 \quad \textbf{(6)}$$

Were $W1$ and $B1$ are respectively the number of pixels correctly and wrongly detected as belonging to the face. The **Gdr** measures at which point the essential parts of the face are detected.

The **Qdr** gives a more strict measure of face detection taking in account the parts of images that wrongly detected as belonging to the face as a decreasing factor. Fig. 3 gives examples on the difference between **Gdr** and **Qdr** values on some images of the evaluate database.

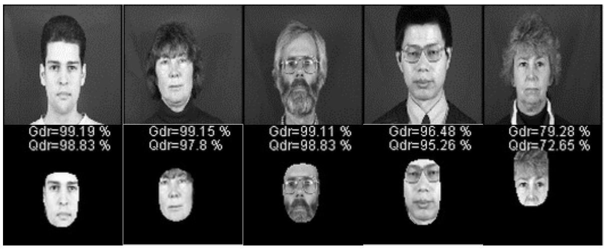

Figure 3: Top: original images. Bottom: faces detected with different values of Gdr, Qdr

5 Experimental results :

In order to check the validity of our proposed method experimental studies were carried out on the XM2VTS images database [13]. It contains **4** recordings of **295** subjects taken over a period of 4 months with rotating head shot in vertical and horizontal directions. Images are **colour** and in **ppm** format.

In our experiences we first brought some transformations to the original images like the change to **gif** format (more compressed) and the use of luminance information only (grey scale images) to compute the Zernike moments.

To obtain the training database we take randomly **15** images of different people, each one with **3** different recordings, so that gives us **45** couples (**Zi**,**Ci**) examples for training the neural network.

To have a precise and rather general idea on the performances of the method, we carried out the construction of **40** training databases always by random taking of the examples. For each database, the network was trained (3 times successively to validate the result) then tested on the whole of the remaining images. For each test, we compute the **Gdr**, the **Qdr** and the Standard deviation (**Std**).

Our experiences aimed at the study of the behavior of the method with respect to the training database, the Zernike vectors parameters m and n and the complexity of the neural network.

All the results presented in this section were obtained on the network output layer in the form of 30 pairs (R,θ) representing the co-ordinates of 30 pixels surrounding the region of the image detected as containing a face.

5.1 General results :

In first, we present on Fig. 4 an example of results obtained for each one of the 295 images representing part of the validation database.

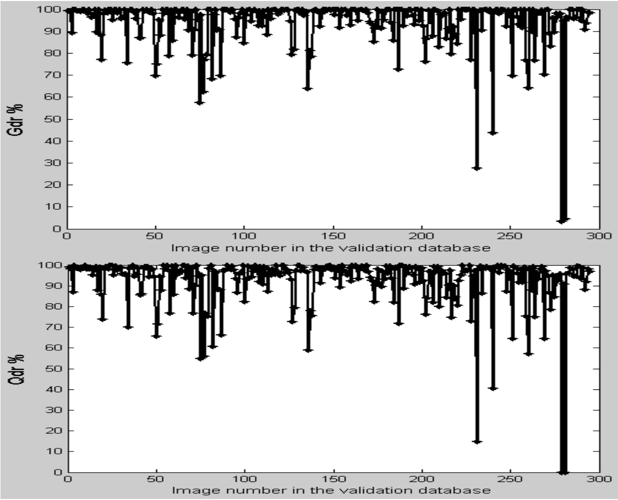

Figure 4: Gdr (Top) and Qdr (Bottom) recorded for 295 images of the validation database

These rates correspond to Zernike moments vectors obtained with order n=10 and repetition m=5. The network was trained on one of the forty training databases constructed randomly.

The resulting rates show that only few faces were incorrectly detected. Most of the images were correctly treated which indicates a very good generalization performances. In Table 1 we see that 90% of images have Gdr and Qdr greater than 80%.

Table1: Distribution of performances reported in Fig. 4. (Nbi: Number of images).

	Nbi	(Nbi/295) %
Gdr<70%	**12**	**04.06**
Gdr<80%	**28**	**09.49**
Qdr<70%	**14**	**04.74**
Qdr<80%	**32**	**10.84**

On Fig. 5 we give some examples of good detected faces from the working database. We chose some images with certain face variability in terms of position, colour, pose and gender.

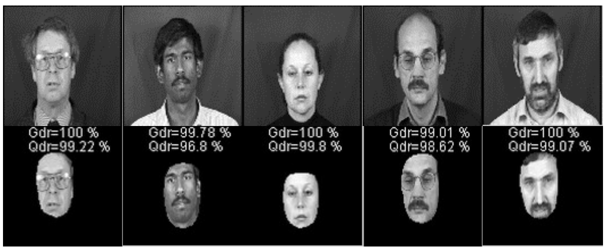

Figure 5 : Examples of face detection. Top: original image. Bottom : face detected.

The difference between Gdr and Qdr is about an average of 2% to 3%. However, this reduced difference is very significant in terms of result accuracy.

5.2 Training database influence :

We want to know the influence that training database has on the quality of detection, especially when the choice of the examples is done randomly. Fig. 6 gives the results of tests validation for each of the 40 training databases. These tests were carried out on all the databases sequentially with the same parameters of Zernike vectors (order **m** equal to **10** and repetition **n** equal to **5**) as well as the same parameters of network (**10** neurons on hidden layer and **60** neurons on output layer) and this for three successive tests for each database. We give the average of quality detection on all the images of the validation database according to the training on each one of the **40** databases. The general form shows that the quality of detection is subject to an average variation up to **6**% of difference relating to the representativeness of the training database compared to the work database in entirely.

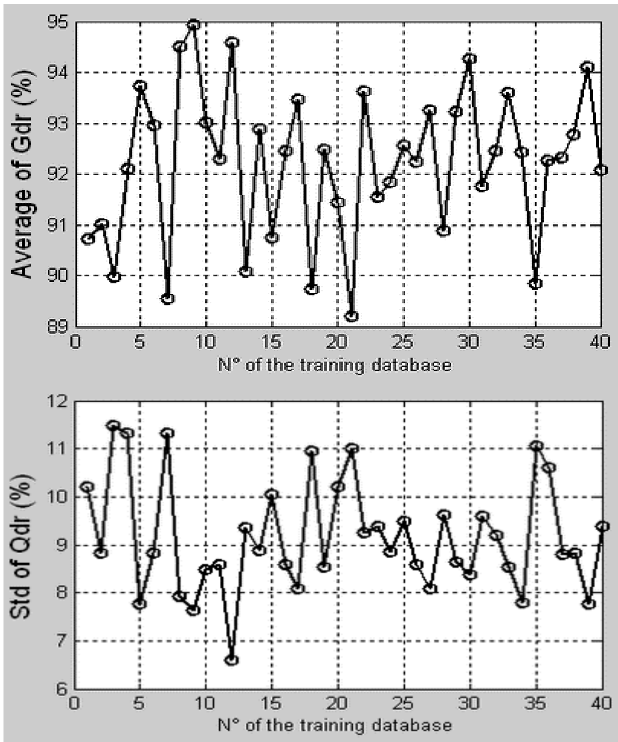

Figure 6: Qdr and Std variations related to the training database. Top: Qdr average Bottom: Std of Qdr

In the same way, we can notice that the variation of the standard deviation is inversely proportional to the average of quality detection what shows that when the training database is rather representative, the most of faces are correctly detected. We can see this clearly on the performances distribution represented on Table 2 for the two extremely cases of the training databases (N° 21 and N° 9) represented on Fig. 6.
The individually comparison between the two extremely training databases (up to 12% of difference for rates inferior to 80%) shows the importance of the choice of the training database to obtain a good detection system.

Table2: Rates distribution for the two extremely cases of training databases (N°21 and N°9)

training database	Qdr	Nbi	(Nbi/295) %
9	<70%	**10**	**03.39**
	<80%	**22**	**07.45**
21	<70%	**32**	**10.84**
	<80%	**54**	**18.30**

5.3 Zernike parameters influence :

We also wanted to see the variation effect of parameters **n** and **m** on the detection rate. These parameters have a direct influence on the complexity of the neural network and computing time of the algorithm in the sense that increasing **m**, **n** or the two increases the Zernike moments vector dimension which controls the dimension of the neural network input layer.
On Fig. 7 we give the variation curves for Qdr average and its standard deviation according to the order **n** which goes from **5** to **23** with a step of **2** and with **m=n-3** and a hidden layer with **10** neurons.
The average curves, computed for the ten neural networks obtained by training on the ten first databases, shows an oscillation in the values for different couples (**n**,**m**) but the general tendency is a decreasing due to the increasing of the complexity of the neural network. The standard deviation increases considerably with the increasing of the values of **m** and **n** which thus indicates the degradation of the measure generalization.
According to averages and standard deviation we obtain best results with n=9 and m=6. However, we found that the choice of n and m can not be dissociated from the choice of the training database in its diversity and its size and from the number of neurons in the hidden layer.

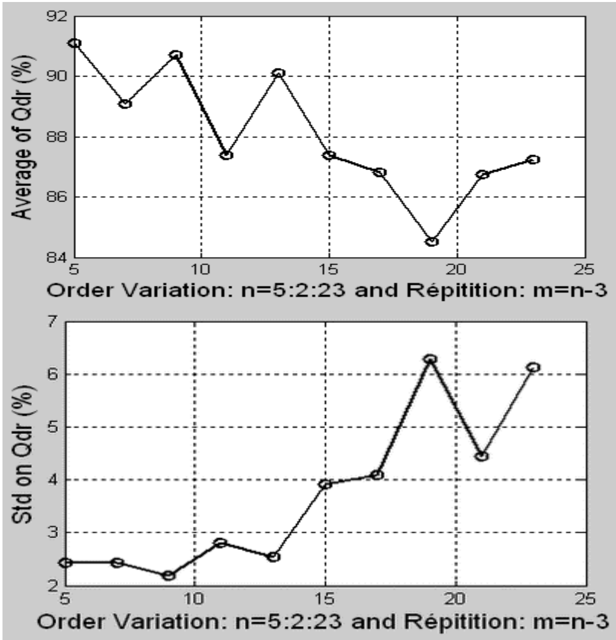

Figure 7 : Qdr and std variations related to the parameters **n** and **m** (**n**=5,7,..., 23 and **m**=**n**- 3). Top: Qdr average. Bottom: Std

5.4 Neural network complexity influence:

The neural network being the last element in the chain of detection system, we wanted to study the influence of the number of neurons in the hidden layer on the quality of detection. So we trained five neural networks with a variable Number of Neurons in the Hidden Layer (NNHL).
In Table 3 we give the results obtained in terms of average and standard deviation (Std) of the good and quality detection rate on the first 10 training databases. Neural networks with 10 and 15 neurons allow obtaining better average results. Beyond this value of NNHL, the complexity of the network decreases his ability to converge for most of the training databases. However, the best individually result obtained (**Gdr=95.29 %** and **Qdr=93.69 %**) was with NNHL equal to **25** on the tenth training database.
It should be also noticed that the average of detections remain very close since the total variation does not exceed the **1%** but the individual variation (for only one training database) recorded is about **4.87 %** on the **Gdr**

Table 3 : Rates detection according to NNHL

NNHL	5	10	15	20	25
Gdr %	**92,32**	**92,87**	**92,72**	**92,33**	**92,55**
Std %	**01,30**	**01,33**	**01,34**	**01,64**	**01,30**
Qdr %	**90,73**	**91,33**	**91,22**	**90,72**	**90,82**

6 Conclusion :

We presented in this work two contributions: a new approach for face detection in an image and a new performance measuring criteria . The face detection method uses a neural network trained on Zernike moments vectors to localize and delimit the area supposed containing a face. The study of its performances was carried out on XM2VTS database through tests concerning the various parameters of the Zernike formulation and the neural network like the training database, the order **m** and repetition **n** parameters and the neural network complexity. We have shown the importance of the choice of the training database to improve the quality of detection. We have also shown that a judicious choice of **m** and **n** have to be done to ensure a high performances with a low complexity of the detection procedure.

Method performances were computed and presented using the new quantitative criteria proposed in this work which allow an automatic computation and appreciation of the quality detection.

Improvements on the judicious choice of the training database dimension and contents and compromise between Zernike vector parameters (m and n) and number of neurons in network hidden layer (NNHL) have to be made in a future work to enhance performances of the method presented.

The method presented can be used to detect parts of the face like eyes, nose, mouth, ... on images or video. It can be also extended to be used in face expression analysis and recognition.

Acknowledgment:

M. SAAIDIA is supported by a national grant from the ministry MESRS of the Algerian state.

References :

[1] E. Hjelmas et B. K. Low. “Face detection : A survey”, Computer Vision and Image Understanding, vol. 83,No. 3,2001, pp. 236-274

[2] M.R. Teague, “image analysis via the general theory of moments”, J. Opt. Soc. Amer., vol. 70, 1980, pp. 920-930.

[3] C-H. Tech, and R.T. Chin, “On image analysis by moment invariants”, IEEE TPAMI, vol. 10, no. 4, 1988, pp. 496-513.

[4] F. Zernike. Physica. 1934

[5] R. Mukundan and K.R. Ramakrishnan, “Fast computation of Legendre and Zernike moments”, Pattern Recognition, vol. 28, no. 9, 1995, pp. 1433-1442.

[6] S. O. Belkasim, M. Ahmadi, and M. Shridhar, “Efficient algotithme for fast computation of Zernike moments”, IEEE 39th MSCS, vol. 3, 1996, pp. 1401-1404.

[7] J. Gu, H. Z. Shua, C. Toumoulinb, and L. M. Luoa, “A novel algorithme for fast computation of Zernike moments”, Pattern Recognition, vol 35, 2002, pp. 2905-2911.

[8] G. Amayeh, A. Erol, G. Bebis, and M. Nicolescu, “Accurate and efficient computation of high order zernike moments”, First ISVC, Lake Tahoe, NV, USA, 2005, pp. 462–469.

[9] J. Haddadnia, M. Ahmadi, and K. Faez, “An Efficient Feature Extraction Method with Peudo Zernike Moment in RBF Neural Network Based Human Face Recognition System”, EURASIP JASP, vol. 9, 2003, pp. 890-891

[10] S. Alirezaee, M. Ahmadi, H. Aghaeinia, and K. Faez, “A weighted pseudo-zernike feature extractor for face recognition”, IEEE, ICSMC vol. 3, 2005, pp. 2128-2132

[11] A. Ono, “Face recognition with Zernike moments”, Systems and Computers. Japan vol. 34, No 10, 2003, pp. 26-35

[12] C. Rosenberger, A. Rakotomamonjy, B. Emile "Generic target recognition", EUSIPCO vienne, 2004, pp. 1613-1616

[13] www.ee.surrey.ac.uk/Research/VSSP/xm2vtsd