Online Quality measurement of face localization obtained by neural networks trained with Zernike moments feature vectors

M. Saaidia[1]

S. Lelandais[2]

M. Ramdani[3]

[1]*Département d'Electronique*
Université de Tébessa, Algérie
Saaidia.m@gmail.com

[2]*IBISC FRE CNRS 2873*
Univ. Evry Val d'Essonne, FRANCE
s.lelandais@iut.univ-evry.fr

[3]*Univ. de Annaba*
Annaba, Algérie
mes_ramdani@yahoo.com

Abstract

Quality measurement of face localization using neural networks is presented in this communication. First, neural network was trained with Zernike moments feature parameters vectors. Coordinate vectors of pixels surrounding faces in images were used as target vectors on the supervised training procedure. Thus, trained neural network provides on its output layer a coordinate's vector (ρ, θ) representing pixels surrounding the face contained in treated image. In second stage, another neural network, trained using TSL color space of images, is used to give a measure quantifying the quality of the localization obtained in the first stage. Experiments of the proposed method were carried out on the XM2VTS database.

Key Words- face localization, quality measurement, neural network, TSL

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), face expressions analysis 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 color, shape,   movement, ...
In this proposed work we present a continuity part of face localization method introduced in papers [2] and [3].

Localization methods are generally presented and evaluated offline according to an experimental database. Here we propose to add a module to the localization system to realize an online measurement of the performed localization. The method concerned by this measure was proposed in [2] and [3]. This global localization method exploits geometrical characteristics of the face to build feature vectors which are used to train neural network to recognize the region of the face in the image. The neural network (trained beforehand) uses the feature vector produced in the first step to output a coordinate's vector for pixels of the face's probable contour contained in the treated image. Offline measurement of method performances was done according to a quantitative measurement criterion [2]. Online quality measurement, presented here, is applied to *the part of the image which was recognized as a face.*

In section 2 and section 3, we will summarize localization method and results obtained for Zernike moments. Section 4 develops the proposed way to online quality measurement implementation and experimental results. Section 5 concludes.

2- Face localization method [3]:

Geometrical moments, particularly Zernike ones, were used for their capacity to compress the geometrical information, contained in the treated image, in a rather reduced parameters vector by projection of the image on an orthogonal basis [4]. This compression characteristic makes them very adapted to the training of classifiers such as neural networks, which often need, on their input layer, feature vectors reduced in size but rather representative of the element subject to the classification. Zernike moments were particularly used for face recognition [5,6] and target recognition in general [7].

2.1. Zernike moments formulation

Zernike moments are part of the geometrical moment's general theory. They were introduced initially by F. Zernike. Zernike moments are built on a set of

orthogonal polynomials which allow construction of orthogonal base given by Eq. (1).

$$V_{n,m}(x, y) = V_{n,m}(\rho, \theta) = R_{n,m}(\rho) \cdot e^{j.m.\theta} \qquad (1)$$

where:

$$\begin{cases} R_{n,m}(\rho) = \sum_{k=|m|}^{n} \dfrac{(-1)^{(n-k)/2} \cdot (n+k)!}{(\frac{n-k}{2})!(\frac{k+m}{2})!(\frac{k-m}{2})!} \cdot \rho^k \\ \rho = \sqrt{x^2 + y^2} \qquad \text{and} \qquad \theta = arctg(y/x) \end{cases}$$

with $n \geq 0, m \neq 0, m < n, n - m < n$ and $n - k$ even.

$R_{n,m}(\rho)$ are the orthogonal radial polynomials, n is the order of the moment and m the repetition factor (the smoothness of the required details) at this order. ρ and θ are respectively the radius and the angle of function's treated point.

This base being orthogonal only inside the unit circle, the image to be projected must be mapped according to Eq.(2) which gives relations between the relative coordinates (i, j) of the initial image pixels and the new pixels coordinates (x_i, y_j) of the mapped one.

$$x_j = c + \frac{j.(d-c)}{(Q-1)} \quad \text{and} \quad y_i = d - \frac{i.(d-c)}{(P-1)} \qquad (2)$$

where (P, Q) are dimensions of the image to be projected, i and j are indices of the point to be mapped and (c, d) defines couple of parameters allowing to map the function inside the unit circle (completely: $c = -1/\sqrt{2}$ and $d = -c$ or partially: $c = -1$ and $d = 1$).

The projection of a numerical function $f(x_j, y_i)$ on the basis functions of Eq.(1) gives the Zernike moments $Z_{n,m}$ according to Eq.(3).

$$Z_{n,m} = \frac{n+1}{\pi} \sum_{x_j^2 + y_i^2 \leq 1} \sum f(x_j, y_i).V^*_{n,m}(x_j, y_i) \qquad (3)$$

where: * denotes the *complex conjugate* of the function.

Traditional formulation of Zernike moments is very easy to implement but its computational time cost is very high. Researchers tried to overcome this major handicap by developing new formulations to enhance the speed computation [8,9]. The proposed algorithm in [9] (which is adopted here) has the advantage to preserve the same accuracy of computation as in the traditional formulation. To lead to this form of representation, previous equations are rewritten and reorganized in Eq.(4) which reduces the computation of Zernike moments for any image to the computation of a linear combination of $\beta_{n,m,k}$ and $X_{m,k}$.

$$Z_{n,m} = \frac{n+1}{\pi} \sum_{x^2 + y^2 \leq 1} \sum \left(\sum_{k=|m|}^{n} \beta_{n,m,k} \cdot \rho^k \right) e^{-j.m.\theta} f(x_j, y_i)$$

$$= \frac{n+1}{\pi} \sum_{k=|m|}^{n} \beta_{n,m,k} \left(\sum_{x^2 + y^2 \leq 1} \sum e^{-j.m.\theta} \cdot \rho^k f(x_j, y_i) \right)$$

$$= \frac{n+1}{\pi} \sum_{k=|m|}^{n} \beta_{n,m,k} \cdot X_{m,k} \qquad (4)$$

where: $\beta_{n,m,k} = \dfrac{(-1)^{(n-k)/2} \cdot (n+k)!}{(\frac{n-k}{2})!(\frac{k+m}{2})!(\frac{k-m}{2})!}$

2.2. Method implementation

Figure 1 gives in a block diagram form the proposed face localization system.
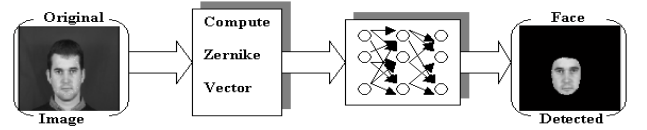

Figure 1: General diagram of face localization system

The implementation of face localization method is mainly based on the training phase which will be achieved in four stages:

(i) Computation of Zernike moments for all the N images of the work database.

(ii) Construction of the training database by randomly taking N_1 images from the work database ($N_1 << N$) and their corresponding Zernike moments vectors Z_i.

(iii) Manual delimitation of the face area in each image of the training database by a set of points $\Omega_i = \{P_1, \cdots, P_d\}$ representing its contour.

(iv) Training of neural networks on the N_1 sets of couples (Z_i, Ω_i).

Neural networks trained with Zernike feature vectors learn to extract statistical information contained in Zernike moments and in their interactions which are closely related to the area of the required face.

To test and measure performances of the network obtained after training operation, we proceed, according to Figure 1, on all $N - N_1$ images remaining in the work database. Face localization procedure will be the same for the two methods compared in this work and will be done in two steps:

- During first step, an image is presented to a program which extracts Zernike vector.

- At the second step, a back-propagation neural network, beforehand trained, receives on its input layer the feature vector which was computed in first step. In response, it gives on its output layer a coordinate's vector for a set of points representing the probable face contour contained in the treated image.

2.3. Offline measurement criterion

To give an objective appreciation of results given by the studied methods, a proposed new way to calculate the detection rate based on the relation between the number of pixels correctly and wrongly detected as pixels of the face and the number of face pixels in each treated image was used. To do so, all testing database images were manually segmented in three regions like it is shown in Figure 2.
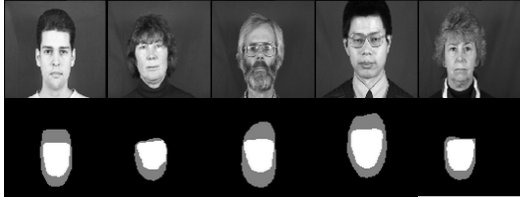

Figure 2: Examples of regions definition. Top: original image, Bottom: mask of regions

First region (white one on the masks of Figure 2) contains the W pixels which represent the essential components of the face (brows, eyes, nose, mouth and surrounding pixels). The second region (grey one) contains pixels surrounding the first region and belonging to the face. The last region contains all the B pixels of the image which do not belong to the face. For the detection system, the first region is one which has to be contained imperatively in the resulting contour and the third one has to be imperatively discarded from it. The second region is optional and has no effect on the computed results. We define two types of measures; Good detection rate (Gdr) and Quality detection rate (Qdr).

$$Gdr = \frac{W_1}{W}.100 \quad \text{and} \quad Qdr = \left(\frac{W_1}{W} - \frac{B_1}{A-B} \right)100 \quad (5)$$

where W_1 and B_1 are respectively the number of pixels *correctly* and *wrongly* detected as belonging to the face and A is the number of all pixels of the image.
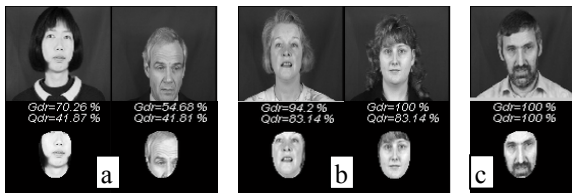

Figure 3. *Top: original images. Bottom:* complementary relation between Gdr and Qdr measures

On Figure 3.a we have the same bad Qdr (about 41%), with two different Gdr(54% and 70%). On Figure 3.b it's the same situation for a good Qdr (about 83%) with two different Gdr values (94% and 100%). To finish, we give on Figure 3.c an example of a face perfectly detected with Gdr and Qdr at 100%. Thus, to have a correct appreciation of recorded results, each one of Gdr and Qdr has to be computed. Best results are obtained when they are both closest to 100% with minimum difference between them.

3. Experimental results

In order to check the validity of the proposed method studied here, experimental studies were carried out on the *XM2VTS* images database [10]. This extended database contains 4 recordings of 295 subjects taken over a period of 4 months with rotating head shot in vertical and horizontal directions. Images are colored and in ppm format. In our experiments we brought some transformations to original images like change to GIF format (more compressed) and the use of luminance information only (grey scale images) to compute the Zernike moments vectors.

To obtain a training database we take randomly 15 subjects with their first 3 different recordings, so that gives 45 examples of couples (Z_i, Ω_i) to train neural networks. To have a precise and rather general idea on method performances, we carried out the construction of 20 training databases always by randomly taking examples from the first 3 recordings of the database. For each test, we compute the average values of Gdr and Qdr and their Standard deviations (Std) σ. Neural networks trained and used in our experiments have 60 neurons on their output layers so they provide 30 coordinate pairs (ρ, θ) for 30 pixels. This number was experimentally chosen to be sufficient to surround face region in the treated image.
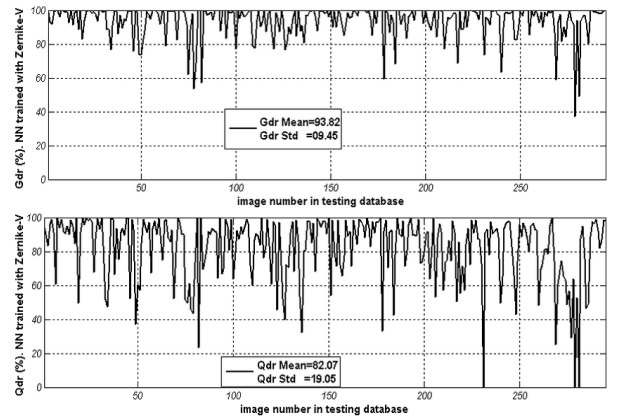

Figure 4: Gdr and Qdr curves for the images of the fourth Database recording. Neural network trained with Zernike moments

Our experiments aimed at the study of behavior of the method according to the training database, training vectors dimension and the neural network complexity.

- First, we present in Figure 4 an example of results given by trained neural network applied to the 295 images of the fourth database recording.

These rates are the best ones, according to different training databases with feature vector dimension equal to 22. The resulting rates show that only few faces were incorrectly detected. Most of the images were correctly treated indicating good generalization performances.

- To study training database influence, experiments were carried out on the 40 training databases randomly constructed.
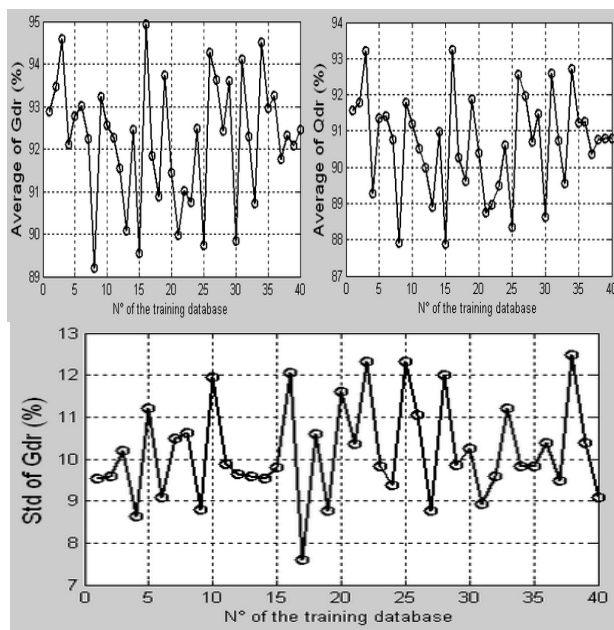

Figure 5: Gdr, Qdr and Std variations related to the Training databases (1,2, ..., 10). Top: Gdr and Qdr averages, Bottom: Std

Results given on Figure 5 were obtained by training, testing and measuring performances of a neural network for each one of the fourty training databases with the same fixed parameters. Input, hidden and output layers have respectively 6, 10 and 60 neurons with a *sigmoid* activation function for the hidden layer and *linear* activation function for the output layer. "Resilient propagation" was used as neural training function.

According to Gdr, Qdr and Std reported by curves in Figure 6, we can see that training with Zernike feature vectors gives Gdr averages that are greater than 90% (up to 95%) for almost the totality of the training databases and Qdr averages are about 85%. In the same way, low Std

values show good generalization performances on images of the testing database (no more than 13%).

4. Online quality measurement:

Our goal would be to find a way how to know that the localized area represents a face. Thus, It would be very useful to instantly (online) calculate the quality of detection.
Figure 6 gives the proposed general diagram to obtain a value measuring the quality of the localized face. The bloc named "Quality Measurement" represents a neural network which was trained on a set of vectors each one representing pixels characteristics of a trained image in the TSL color space.
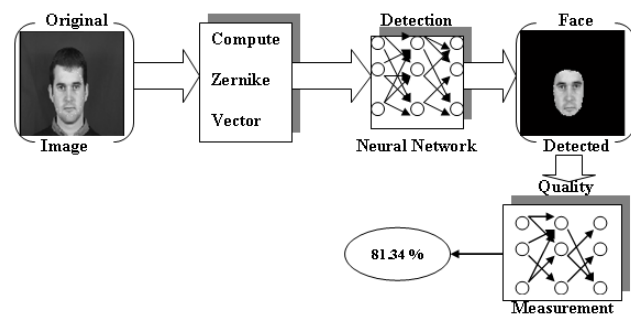

Figure 6: General diagram of online quality measurement on the face localization system

The method exposed here is based on the skin color characteristics. Indeed, Skin color has proven to be a useful and robust cue for face detection, localization and tracking [11]. Skin color modeling problem was widely studied for computer graphics, video signal transmission and compression, … These studies have led to the emergence of several color spaces like RGB, HSI, YCrCb, YUV, TSL, …
TSL color space is a transformation of the normalized RGB. Compared to other color spaces, TSL appears more adapted for skin modeling [12].
(Eq.6) gives TSL color space formulation :

$$S = \left[9/5(r'^2 + g'^2\right]^{1/2}$$
$$T = \begin{cases} \arctan(r'/g')/2\pi + 1/4, \dots g' > 0 \\ \arctan(r'/g')/2\pi + 3/4, \dots g' < 0 \\ 0, \dots g' = 0 \end{cases} \quad (6)$$
$$L = 0.299R + 0.587G + 0.114B$$

Where r' and g' are the normalized components of R and G in the RGB color space.

This modeling way was mainly used for face detection [13], [14] and [15].

4.1. General results:

In our case we'll use TSL color space to identify the skin pixels contained in the area detected as the face in the image. The identification process will be done by the trained back propagation neural network on the skin color characteristics vector of the localized image to be tested. The measure designed here by Qual represents the ratio of the pixels identified as skin pixels to those representing the total number of pixels of the region detected as a face.

Applying this procedure to the XM2VTS images gives results which a sample is exposed in Figure 7 as a dashed-line curve. The line curve represents the Gdr calculated on the same images. This will permit a comparison between the offline measure and the online one.
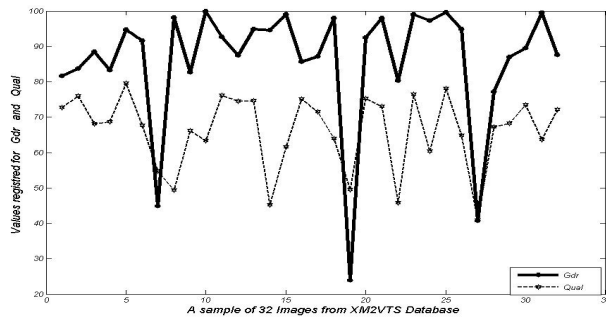

Figure 7.Comparison between Offline and Online face localization quality measurement. Applied method on a sample of 32 images

First remark to do is that all Qual values are less than the correspondent ones of Gdr. This is a normal result since the online measure is evaluated according to a rectangular window while Gdr values are calculated relatively to the real pixels of the face. Also some components of the face like eyes are not taken in account by the TSL analyze.
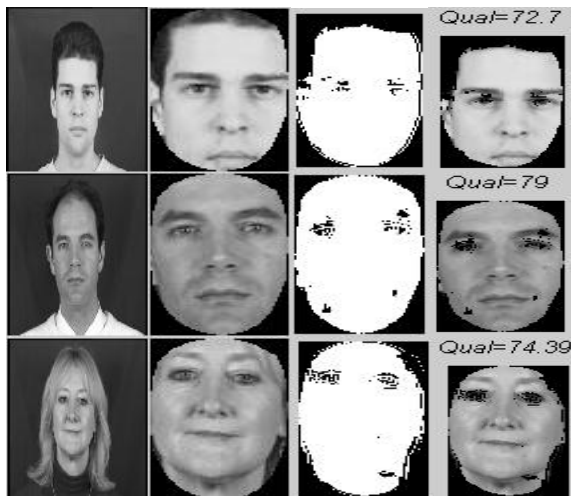

Figure 8. Examples of online Quality measurement applied to images of XM2VTS database

The second remark is that the curves evolution is very similar which validates the proposed online measure.
Figure 8 presents examples obtained by the application of the proposed online measure to some images from the XM2VTS database. The four images present the four steps of the treatment process. In right the original image, next the face is localized. The third bicolor image is produced by the online measurement procedure. White pixels are ones which are considered by the trained neural network as skin pixels in TSL color model. For the three images Gdr was respectively 81.58, 99.66 and 87.12. All results are given in percent.

4.2. Particular results:

Online measure on images of the XM2VTS database gives particular result. Although, on the sample presented on figure 7 we can see that for few images the difference between the online measure Qual and the offline one Qdr is obvious. For example, in the case of image N°8 the Gdr is about 98.08 while the Qual measure gives only 49.37. This was caused by a particular sign present on the face like beard or moustache. Colors characteristics of these components don't agree with skin ones. Glasses also will cause bad quality measurement. On Figure 9 we give examples of these types.
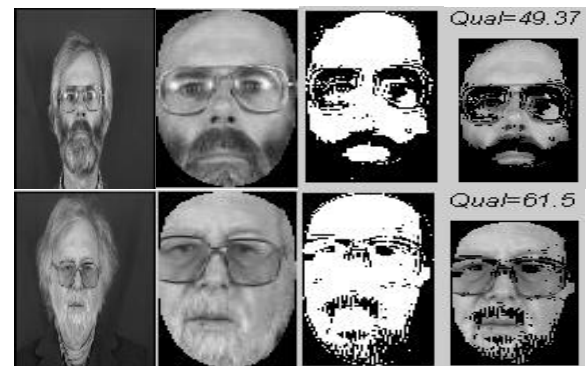

Figure 9. Examples of Quality measurement sensitivity to the face particular signs. images N°8 and N°104

4.3. Out of XM2VTS database:

Experiments were also taken on images which are not from the XM2VTS database and in some case they don't contain face but only draws which are similar to face. A sample of results is given in figure 10.
We can see clearly the utility of online measurement in the validation of the results obtained by the localization procedure. On the first and last images a zone is detected as face while online measure determine that almost pixels not agree with the TSL skin pixels color model.
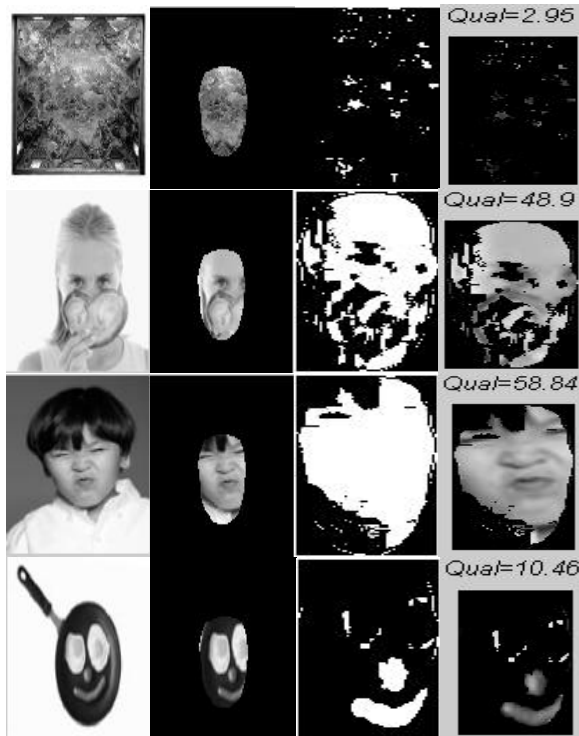
Figure 10. Examples of online Quality measurement applied to images out of XM2VTS database

5. Conclusion

Online measurement quality of face localization using TSL color space and neural networks was presented in this communication. Fist we presented the localization method than we introduced the TSL space color formulation and the way to implement it as the support of the online measurement method.

Recorded results and their comparison to the offline ones demonstrate the validity of the proposed way to realize the online quality measurement. We demonstrate also the sensitivity of this method to the particular signs on the face like beard, moustache and glasses.

Online measurement was also applied to images out of XM2VTS database and containing draws similar to faces. Online quality values permit to identify the localization errors made on these types of images.

References:

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

[2]. M. Saaidia, S. Lelandais, V. Vigneron and M. Bedda, "Face detection using Zernike moments. Evaluation by quantitative measurement creterion", WSEAS Transactions on Signal Processing., Issue 2, Vol. 3, pp 226-232, 2007.

[3]. Saaidia, M. Chaari, A. Lelandais, S. Vigneron, V. Bedda, M, "Face localization by neural networks trained with Zernike moments and Eigenfaces feature vectors. A comparison", AVSS2007, pp. 377-382, London 2007

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

[5]. 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, pp. 890-891, 2003.

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

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

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

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

[10]. K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, ''XM2VTSDB: The Extended M2VTS Database'' Proc. AVBPA'99 Washington DC, 1999.

[11]. V Vezhnevets, V Sazonov, A Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques", Proc. Graphicon, 2003 - graphicon.cs.msu.ru

[12]. Terrillon, J.-C., Shirazi, M. N., Fukamachi, H., and Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images", In Proc. of the nternational Conference on Face and Gesture Recognition, pp. 54–61, 2000

[13]. Kovac, J.; Peer, P.; Solina, F, "Human skin color clustering for face detection", EUROCON 2003. Computer as a Tool. The IEEE Region 8, Volume 2, pp. 144-148, Sept. 2003

[14]. Guerfi, J.-P. Gambotto et S. Lelandais, "Ligne de Partage des Eaux pour l'extraction de visage dans l'espace de couleurs TLS", CORESA, Novembre 2005.

[15]. D Brown, I Craw, J Lewthwait, "A SOM based approach to skin detection with application in real time systems", Proc. of the British Machine Vision Conference, 2001.