FROG: Face Recognition via Multimedia Messaging

Prospero C. NAVAL, Jr., Riza Theresa B. BATISTA,
Hazel Jean M. MERCADO, Usman P. MOTI, Jr.
Computer Vision & Machine Intelligence Group
Department of Computer Science, University of the Philippines
Diliman, Quezon City 1101 Philippines

ABSTRACT

The challenging task of face recognition is addressed with the aim of developing an identification system that processes queries coming from camera-equipped mobile phones. Queries are in the form of colored images transmitted via the Multimedia Messaging Services (MMS) technology. Face localization is performed on each query image through the methods of skin color detection, region growing, and ellipse analysis. A localized face region is represented as a set of features extracted through Kernel Principal Components Analysis. The features are submitted to Support Vector Machines for pattern classification. Recognition rates comparable with traditional systems demonstrate that it is possible to effectively implement a remote face recognition engine which takes as input the images captured by mobile phone cameras. Such an application, not limited by distance restrictions and readily accesible to any authorized person anywhere, exhibits strong potential in biometrics and security systems

Keywords

Biometrics, Security Systems, Face Recognition, Face Detection, Kernel Principal Component Analysis, Skin Color Detection, Multimedia Messaging Services (MMS)

1. INTRODUCTION

There has been much interest in recent years in biometric technologies brought about by the heightened concern for national security. Law enforcement organizations face the major challenge of identifying criminals while at the same time respecting civil liberties, often relying on computer systems that process enduring physical or behavioral characteristics. Examples of these biometric features are fingerprint, iris, retina, handwriting, voice and face.

Among the biometrics, face recognition enjoys the advantage of being the least obstrusive since it is a) passive, requiring no special electromagnetic illumination and b) people are identified by others through their faces and therefore are generally comfortable with systems that recognize in the same way as humans [8]. Contact biometric technologies require the subject to interact with a sensor while passive biometrics do not require any action from the subject. As a non-contact biometric, face recognition is the most passive and bothers the subject least.

Face recognition systems are widely used especially in high-security venues such as airports. A typical face identification system consists of facial information of known criminals in a watch list stored in a database together with cameras that scan for individuals whose facial data match those in the database. An alarm is set off when a match is found. A major disadvantage of such a system, however is that it is confined to one place or building and is not readily accessible by other authorized users.

The proposed implementation, referred to by the authors as Face Recognition On the Go (FROG), is a fully-automated face recognition system that is accessible to any authorized user regardless of his location. It aims to overcome the limitations of traditional identification systems by providing a remotely accessible face recognition engine. Using a camera-equipped and Multimedia Messaging Services (MMS)-capable mobile phone, any authorized user can send a query to the FROG engine anytime and from anywhere and receive results quickly.

Face Recognition On the Go has a Mobile Framework into which the two major modules, namely, Face Localization and Face Recognition, are integrated. The Mobile Framework provides a means for transmitting an image captured by a mobile phone to the remote engine. A query image will be first submitted to the Face Localization module which performs skin color detection, region growing, and ellipse analysis. The Face Recognition module takes as input the isolated face region and performs Kernel Principal Component Analysis on the image to produce a set of features. Support Vector Machines (SVM) classifiers previously trained on features of multiple face samples of different persons are then used for classifying these features. The result will be transmitted to the mobile phone which sent the query.

2. MOBILE FRAMEWORK

The system is designed such that data transmission is enabled between mobile phones and the remote server where the Face Localization and Face Recognition modules reside.

2.1 Submission of Queries

An authorized user of the system sends a query image to the engine through the Multimedia Messaging Services (MMS) feature of his mobile phone. With Multimedia Messaging, mobile phone users can incorporate images in their messages and send them to an email address. On the server side, a daemon waits for incoming queries in the form of multimedia messages. After verifying that the sender is authorized, the daemon extracts the image attached to each message and

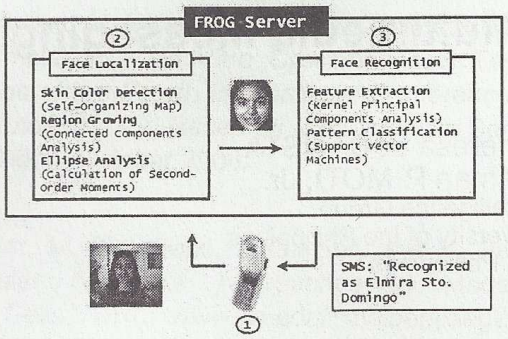

Figure 1: System Overview of the FROG Project

submits it to the major modules.

2.2 Sending out of Responses

When the Face Localization and Face Recognition modules have finished processing the image in the query, the server sends out the identities to the sender's mobile phone, if several matches are found. The response is sent through Short Messaging Services (SMS). In this implementation, the authors utilized Kannel, an open-source SMS gateway that requires a GSM modem attached to the server. In our system, the GSM modem used is the one built into the Nokia 6210 mobile phone. A slight modification can easily transform the system to return to the sending phone not only the identities of the top ranking matches but also their corresponding mugshot images, thus making it more useful for law enforcement purposes.

3. FACE LOCALIZATION

Face Localization is a necessary preliminary step in our fully-automated face recognition system. It isolates the face region from the image enclosing it in a bounding box before submitting it to the pattern classifier. Our system's face localization module relies on the following processing: skin color detection followed by region growing and ellipse analysis.

3.1 Skin Color Detection

Skin color detection was chosen as the basis for face localization because it is computationally less demanding than the more popular view-based and neural-network based approaches. One criticism, though, of color-based approaches is their difficulty to robustly detect skin colors in the presence of complex background and different lighting conditions [3]. Still, the authors demonstrate that using a Self Organizing Map (SOM), it is possible to construct a skin detector that produces results comparable to other methods [1]

The Self-Organizing Map is an unsupervised neural network frequently used to find meaningful patterns from data. It consists of nodes or neurons which, after training, become tuned to the various classes of patterns in a topologically ordered fashion. Our skin color detector is based on the work of Brown, Craw, and Lewthwaite [2].

The Normalized RG color space, known to have reduced sensitivity to illumination, was chosen for our implementation since it has given good results in skin color detection experiments. Training pixels, both skin and non-skin, were gathered from colored images. The normalized R and G values of these pixels were computed and arranged in a random list of 15,000 skin and 15,000 non-skin vectors.

After randomly initializing the codebook vectors of the neurons of the Self Organizing Map, the training vectors are sequentially presented to the network. For each vector, a search for the best-matching neuron is made. After all the training vectors have been presented, the network is calibrated by the presentation of the same set of pixels, this time, labelled as skin or non-skin. Each neuron is assigned a label of 'skin' or 'non-skin', depending on the type for which it fired the most [2]

The feature map is now ready to accept pixels for classification. Normalized RG pixels in an input image are presented to the map and classified with the label of the winning neuron. The result is a binary image whose 1's and 0's correspond to skin and non-skin pixels in the original image.

3.2 Region Growing

The detected skin colored pixels are grouped into connected components through a queue-based method that grows a set of input seed regions, where each seed is grown as far as possible [6].

Since the design of this particular implementation is for use in security systems, it is assumed that there is only one face in the query image. Therefore, only the largest connected component of skin-colored pixels need to be extracted, and this will be considered as the candidate face region; components that correspond to the subject's other body parts or to skin-colored objects in the image will be discarded. We have the following algorithm that isolates the face region from the other skin-colored body parts and artifacts in the image.

- The binary image produced by the skin detector is scanned for a skin pixel. This pixel becomes the seed of a new region, is marked as visited and marked with a new region number.
- Each of the eight neighbor pixels of the seed is checked. If a neighbor is also a skin pixel and is not yet visited, it is enqueued, labelled with the corresponding region number, and marked as visited.
- After checking all of the eight neighbors, the queue is dequeued once. The dequeued pixel is now the new seed. Repeat the second step.
- When the queue becomes empty, i.e., all the previously enqueued pixels have already been dequeued, the image is scanned again for a skin pixel that is not yet visited. This is now the seed of a new region.
- The steps above are repeated until all skin pixels have been labelled with a corresponding region number.

Figure 2: Original Image

Figure 3: Skin Pixels as Detected by SOM

The connected component with the greatest number of skin-colored pixels is taken as the candidate face region and is thus isolated. All the other skin-colored regions are discarded.

3.3 Ellipse Analysis

Since the human face is elliptical, ellipse analysis on the largest component of skin colored pixels is needed. The best-fit ellipse is then used in the determination of the bounding box which enclose the isolated face region.

The center of gravity of the largest component is determined by

$$\bar{x} = \frac{1}{N} \sum_{(x,y)\in C} x \qquad \bar{y} = \frac{1}{N} \sum_{(x,y)\in C} y \tag{1}$$

where x and y correspond to the x and y coordinates of the pixels, respectively. The orientation Θ of the connected component can be determined by computing the second order moments of the pixels' formation.

$$\Theta = \frac{1}{2} \cdot \arctan\left(\frac{2 \cdot \mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}\right) \tag{2}$$

The values $\mu_{i,j}$ are given by the following formulas [14]:

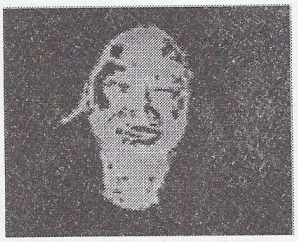

Figure 4: Region Isolated by Ellipse Analysis

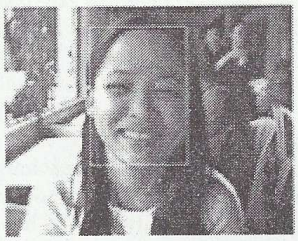

Figure 5: Localized Face Region

$$\begin{aligned}
\mu_x &= \frac{\sum_{(x,y)\in C} x \cdot I(x,y)}{\sum_{(x,y)\in C} I(x,y)} \\
\mu_y &= \frac{\sum_{(x,y)\in C} y \cdot I(x,y)}{\sum_{(x,y)\in C} I(x,y)} \\
\mu_{0,2} &= \frac{\sum_{(x,y)\in C} (y-\mu_y)^2 \cdot I(x,y)}{\sum_{(x,y)\in C} I(x,y)} \\
\mu_{2,0} &= \frac{\sum_{(x,y)\in C} (x-\mu_x)^2 \cdot I(x,y)}{\sum_{(x,y)\in C} I(x,y)} \\
\mu_{1,1} &= \frac{\sum_{(x,y)\in C} (x-\mu_x) \cdot I(x,y)}{\sum_{(x,y)\in C} I(x,y)}
\end{aligned} \tag{3}$$

where C is the largest connected component and $I(x,y)$ is the pixel intensity at coordinate (x,y).

In determining the lengths of the major and minor axes of the best-fit ellipse, the moments of inertia need to be evaluated.

$$\begin{aligned}
I_{min} &= \sum_{(x,y)\in C} [(x-\bar{x}) \cdot \cos\Theta - (y - \bar{y} \cdot \sin\Theta)]^2 \\
I_{max} &= \sum_{(x,y)\in C} [(x-\bar{x}) \cdot \sin\Theta - (y - \bar{y} \cdot \cos\Theta)]^2
\end{aligned} \tag{4}$$

where I_{min} and I_{max} are the least and greatest moment of inertia of an ellipse with orientation Θ, respectively.

Finally, the lengths of the axes are given by

$$a = \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \cdot \left[\frac{(I_{max})^3}{I_{min}}\right]^{\frac{1}{8}} \qquad b = \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \cdot \left[\frac{(I_{min})^3}{I_{max}}\right]^{\frac{1}{8}} \quad (5)$$

where a is the length of the major axis and b is the length of the minor axis. These values become the length and width, respectively, of the bounding box that will enclose the localized face region.

Fig. 2 shows an image before undergoing any processing and the next three figures are the outputs produced by the SOM Skin Dectector (Fig. 3), Ellipse Analysis (Fig. 4), and the face enclosed in a bounding box Fig. 5).

3.4 Face Recognition

Principal Component Analysis has long been used for extracting structure from high dimensional data sets. Kernel Principal Component Analysis [10] has recently been proposed by Scholkopf, Smola and Muller as a nonlinear extension to Principal Component Analysis, drawing inspiration from the "'kernel trick" employed in Support Vector Machines, of implicitly mapping data from input space into a high dimensional feature space and doing what was done in input space in the feature space.

Principal Component Analysis has been found to be an effective method for face recognition [12]. PCA projects the high dimensional image vector into the subspace spanned by the dominant eigenvectors (eigenvectors whose eigenvalues are large) where the variance of the training images is high, making it easy for the classifier to compute for the correct decision surface. Principal Component Analysis can discover linearly embedded manifolds and produce a compact orthonormal basis representation. The manifold structure of the facial recognition task, however, cannot be assumed to be linear since it unlikely that the complications present in this task such as variations in facial expression, illumination, etc. are linear in nature. One promising subspace method that could deal with nonlinearly embedded manifolds is Kernel Principal Component Analysis.

3.4.1 Kernel Eigenfaces

Kernel PCA uses a nonlinear kernel function $\Phi(\mathbf{x})$ to project input data $\mathbf{x}$ into feature space F which is nonlinearly related to the input space and performs PCA in feature space F . Although the dimensionality of the kernel feature space F is much higher than that of the input space, kernel methods do not suffer from the curse of dimensionality because computation in feature space is carried out implicitly.

The Kernel PCA algorithm is as follows [10]:

Given a set of N vectors $\mathbf{x}_1, ..., \mathbf{x}_N$, and an inner product kernel $k(\mathbf{x}_i, \mathbf{x}_j)$ we construct the $N \times N$ kernel matrix $K^{'}$ whose ijth element is $k(\mathbf{x}_i, \mathbf{x}_j)$. We center the mapped data points in feature space using the following equation:

$$K_{ij} = K^{'}_{ij} - \frac{1}{N}\sum_{m=1}^{N} 1_{im}K^{'}_{mj} - \frac{1}{N}\sum_{m=1}^{N} K^{'}_{in}1_{nj} + \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N} 1_{im}K^{'}_{in}1_{nj} \quad (6)$$

where $\mathbf{1}$ is an $N \times N$ matrix whose entries are all 1's.

We then solve the eigenvalue problem

$$\mathbf{Ka} = \lambda\mathbf{a} \quad (7)$$

for positive eigenvalues j_j with these eigenvalues sorted in decreasing order ($\lambda_j \geq \lambda_{j+1}$) and normalize the eigenvector coefficients

$$\mathbf{a}^{(n)} \cdot \mathbf{a}^{(n)} = \frac{1}{\lambda_n} \qquad n = 1, \ldots, p \quad (p \leq N) \quad (8)$$

To extract the nth kernel principal component $q^{(n)}$ of a test image $\mathbf{x}_t$ we use the following formula:

$$q^{(n)} = \sum_{i=1}^{N} \mathbf{a}_i^{(n)} k(\mathbf{x}_i, \mathbf{x}_j) \quad (9)$$

3.4.2 Support Vector Machines

In pattern classification using Support Vector Machines (SVM), examples are mapped to a higher dimensional feature space and a decision hyperplane in feature space that separates the data points is computed. The SVM algorithm [13] computes for the separating hyperplane whose margin of separation between positive and negative examples is maximized.

4. EXPERIMENTAL RESULTS

We determined FROG's face recognition performance using two face datasets: the ATT Face Database and another database that we built ourselves. Performance of our system on the grayscale ATT Face Database allows us to test our algorithms on a well-known database and compare our results with published data. Experiments with our own database will show how the system performs when handling colored images captured by the mobile phone cameras.

4.1 Experiments on the ATT Face Dataset

The ATT Face Dataset consists of a total of 400 grayscale images, with 10 different images for each of the 40 persons in the database. Each image has a resolution of 92 x 112 pixels.

A partition of 50/50 was used for training and testing. Each training set contains 200 labelled images of which 5 and 195 are positive and negative examples, respectively. A 10-fold cross validation was performed.

Feature extraction was performed on the training set of 200 labelled examples using Kernel Principal Components Analysis. In forming the input vector, we down-sampled the images by 4 resulting in an image resolution of 23 x 28 pixels.

The pixels are then normalized by dividing each gray level value by 256.

The kernel function employed is the polynomial kernel $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d$. The values of d were varied from 2 to 6. For each image, the top n kernel principal components $q^{(n)}$ corresponding to the kernel were then extracted.

The principal components were submitted to a nonlinear Support Vector Machines classifier for processing. The SVM classifier was trained using a polynomial kernel of degree 2.

The authors used the publicly-available SVMLight pattern classifier code written by Thorsten Joachims which implements Vapnik's SVM algorithm for pattern classification [4]. The results of the experiments for the different values of polynomial kernels are shown in Fig 6. The best performer on the ATT Face Dataset was found to be the 2nd degree polynomial Kernel PCA [7].

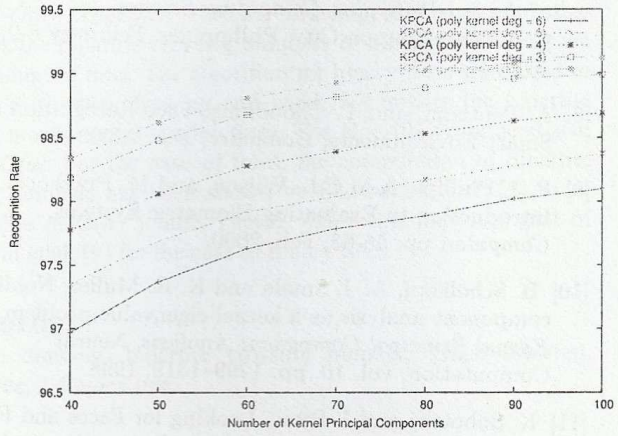

Figure 6: ATT Dataset Recognition Rate for Increasing Dimensionality of Principal Subspace for Different Polynomial Kernel Degree Values

4.2 Experiments on the FROG Face Dataset

The authors built their own face dataset consisting of 10 face images for each of the 40 persons in the database. Each of the raw images was captured using a Canon Powershot A60 Digital camera and saved in jpeg format with a size of 320 × 240 pixels and a resolution of 72 pixels per inch. To produce the face image each raw image was submitted to the Face Localization module. The result is converted to pgm (portable graymap) format of size 92 × 112 pixels. The face images for each individual in the database have different facial expressions and minimal differences (within 15 degrees) in face rotation.

5. FACE RECOGNITION

Training was done using the same settings that were used on the ATT Face Dataset. Tests were done using as queries the images captured by Nokia 7250 and Nokia 3650 mobile phone cameras. For comparison, images from the Canon

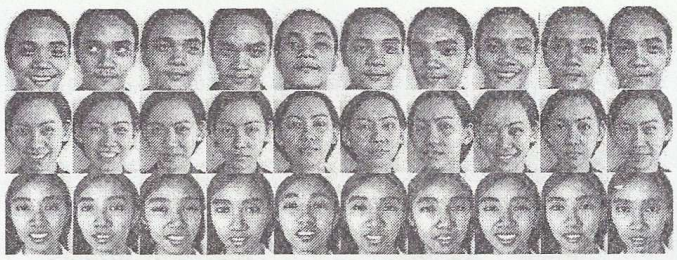

Figure 7: Sample Images for 3 Persons in the FROG Face Dataset

Digital Camera which was used to capture the training images were also tested for recogntion accuracy. The results are shown in Fig. 8.

Face recognition algorithms are known to be sensitive to illumination variations. Images taken at different locations or on different days result in significant reduction in recognition performance. Studies reveal that lighting changes could degrade accuracy by 9 to 11 % [9].

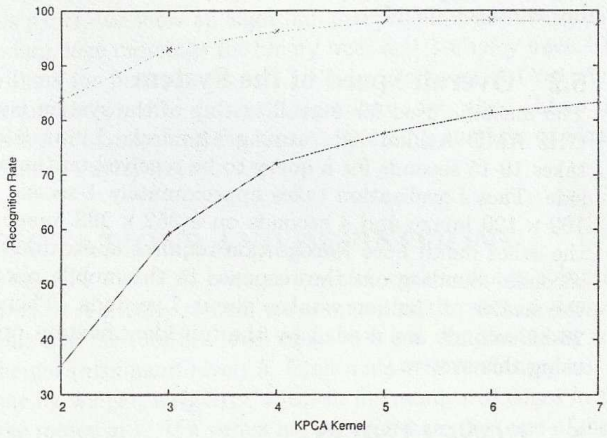

Figure 8: Recognition Rate of Digital Camera (Upper Curve) and Mobile Phone (Lower Curve) for Different Polynomial Kernel Degree Values (Number of Principal Components is Fixed at 100)

5.1 Factors Affecting Recognition Performance

Two major factors in image collection using a mobile phone strongly affect recognition performance: a) the distance between the camera and the subject and b) the difference between the time of acquisition of the training images and the time of acquisition of the query image.

Query images that were taken too close to or too far from the subject cause errors in recognition. When taking the picture of a face, untrained users have the tendency to obtain the image from a distance such that face occupies a large portion of the LCD monitor of the camera. In most cameras, the face image that is captured is distorted since the subject is already too close to the lens. Taking the image from too

far results in grainy images. Fig. 9 shows the considerable variation in appearance of the same subject imaged by the same camera from different distances. It is therefore recommended that users of the system take the image of a subject from an camera-dependent optimal distance (e.g. 2 ft for the Nokia 7250 and Nokia 3650).

Figure 9: Variation of Facial Appearance of the Same Subject with Camera Distance

Changes in the subject's physical appearance attributed to considerable difference between the time of acquisition of the training images and the query image also cause misclassifications. Considerable degradation in performance has been observed for face recognition systems when the the query and training images were taken over 1.5 year apart [9]. An occasional updating of the face database and SVM retraining are recommended.

5.2 Overall Speed of the System

The machine used for overall testing of the system is a 1.8 GHz AMD Athlon XP running Mandrake Linux 9.2. It takes 10-15 seconds for a query to be received on the server side. Face Localization takes approximately 1 second on a 160×120 image and 4 seconds on a 352×288 image. On the other hand, Face Recognition requires approximately 8 seconds. Sending out the response to the mobile phone of the sender of the query takes about 7 seconds. Therefore, 26-34 seconds are needed by the full identification process using this system.

6. CONCLUSION

We have implemented a fully-automated face recognition system called Face Recognition on the Go (FROG) that can accept face images as queries from camera-equipped mobile phones with recognition accuracy comparable to traditional systems. The system has the major advantage of overcoming distance restrictions inherent in current implementations. We hope that its ease of use, availability, and low cost will make face recognition an even more attractive biometric for high security applications.

7. REFERENCES

[1] R.T.B. Batista, H.J.M. Mercado, U.P. Moti, Jr., and P.C. Naval, Jr., Face Recognition on the Go: Security Plus Mobility via Multimedia Messaging (The Face Detection Module), *Proceedings of the 4th Electronics and Communications Engineering Conference*, Cebu, Philippines, Nov. 2003

[2] D.Brown, I. Craw and J. Lewthwaite, A SOM Based Approach to Skin Detection with Application in Real Time Systems, *In Proceedings of the British Machine Vision Conference*, 2001.

[3] C. Garcia and G. Tziritas, Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis, *IEEE Trans. Multimedia*, 1(3):264-277, September 1999.

[4] T. Joachims, Making large-Scale SVM Learning Practical, Advances in Kernel Methods - Support Vector Learning, B. Schlkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

[5] T. Kohonen, J. Hynninen, J. Kangas, and J. Laaksonen, SOM PAK: The Self-Organizing Map program package, *Report A31*, Helsinki University of Technology, Laboratory of Computer and Information Science, Jan. 1996.

[6] M. Levine and S. Shaheen, A Modular Computer Vision System for Image Segmentation, *IEEE PAMI*, 3(5), 540-554, Sept. 1981.

[7] P. C. Naval, Jr., Recognizing Faces Using Kernel Eigenfaces and Support Vector Machines, *Proceedings of the 3rd Philippine Computing Science Conference*,Quezon City, Philippines, February 8-9, 2003.

[8] A. Pentland, and T. Choudhury, Face Recognition for Smart Environments, *Computer*, Feb. 2000.

[9] P. J. Phillips, A.M.C.L. Wilson, and M. Przybocki, An Introduction to Evaluating Biometric Systems, *Computer*, pp. 56-63, Feb. 2000.

[10] B. Scholkopf, A. J. Smola and K, R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, *Kernel Principal Component Analysis*, Neural Computation, vol. 10, pp. 1299–1319, 1998.

[11] K. Sobottka and I. Pitas, Looking for Faces and Facial Features in Color Images, *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, Russian Academy of Sciences, 1996.

[12] M. Turk, and A. Pentland, Eigenfaces for Recognition, *Journal for Cognitive Neuroscience* 3:71-86, 1991.

[13] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 1995.

[14] V. Vezhnevets, Method for Localization of Human Faces in Color-Based Face Detectors and Trackers, *The Third International Conference on Digital Information Processing And Control In Extreme Situations*, Minsk, Belarus, May 28-30, 2002.

[15] Z. Yankun and L. Chongqing., Face Recognition Using Kernel Principal Components Analysis and Genetic Algorithms.