Geometric Approach for Human Emotion Recognition using Facial Expression

S. S. Bavkar
Assistant Professor
VPCOE Baramati

J. S. Rangole
Assistant Professor
VPCOE Baramati

V. U. Deshmukh
Assistant Professor
VPCOE Baramati

ABSTRACT
Paper contains emotion recognition system based on facial expression using Geometric approach. A human emotion recognition system consists of three steps: face detection, facial feature extraction and facial expression classification. In this paper, we used an anthropometric model to detect facial feature points. The detected feature points are group into two class static points and dynamic points. The distance between static points and dynamic points is used as a feature vector. Distance changes as we track these points in image sequence from neutral state to corresponding emotion. These distance vectors are used for input to classifier. SVM (Support Vector Machine) and RBFNN (Radial Basis Function Neural Network) used as classifier. Experimental results shows that the proposed approach is an effective method to recognize human emotions through facial expression with an emotion average recognition rate 91 % for experiment purpose the Cohn Kanade databases is used.

Keywords
Geometric Method, Anthropometric model, SVM, RBFNN and LK Tracker

1. INTRODUCTION
Recently there has been a growing interest in improving the interaction between humans and computers. It is argued that to achieve effective human-computer intelligent interaction, there is a need for the computer to interact naturally with the user, similar to the way humans interact. Humans interact with each other mostly through speech, but also through body gestures to emphasize a certain part of speech and/or display of emotions. Emotions are displayed by visual, vocal and other physiological means. There is more and more evidence appearing that shows that emotional skills are part of what is called 'intelligence'. One of the most important ways for humans to display emotions is through facial expressions. Mehrabian [1] points out that 7% of human communication information are communicated by linguistic language (verbal part), 38% by paralanguage (vocal part) and 55% by facial expression. Therefore, facial expressions are the most important information for emotional perception in face to face communication. The emotion recognition system can be broadly classified in two methods: Appearance based (Texture) and Geometric based. Texture-based method model local texture around a given feature point [2][3], for example the pixel values in a small region around a mouth corner. Geometric-based methods regard all facial feature points as a shape [4], which is learned from a set of labeled faces and try to find the proper shape for any unknown face.

The remainder of this paper is organized as follows: section II, specify the related work, and then in section III facial feature point localization system is presented. Our facial expression recognition system is presented in section IV, where we specify the feature extraction method and the recognition approach. In section V, we present the obtained experimental results. Finally, conclusion is presented in section VI.

2. RELATED WORK
A Neural Network (NN) is employed to perform facial expression recognition in [5]. The features used can be either the geometric positions of a set of fiducial points on a face or a set of multi-scale and multi-orientation Gabor wavelet coefficients extracted from the facial image at the fiducial points. The recognition is performed by a two layer perceptron NN. A Convolutional NN was used in [6]. The system developed is robust to face location changes and scale variations. Feature extraction and facial expression classification were performed using neuron groups, having as input a feature map and properly adjusting the weights of the neurons for correct classification. A method that performs facial expression recognition is presented in [7].

There are several works in the field of facial expression recognition using feature points. The added value of our work is the modelization of muscle contraction using the variation of muscle distances relative to the neutral state. The previous studies on facial feature modelization are based on:

1) Points displacements [8]
2) Facial points coordinate [9],
3) Distances between points but it is based on the deformation of the facial contour [10],
4) Deformation of the shape from the neutral state but not on the contraction of facial muscles [11].

3. FACIAL FEATURE POINTS LOCALIZATION
For facial feature point localization, a anthropometric model is developed on detected face. Corners of the eyes, corners of the eyebrows, corners and outer mid points of the lips, corners of the nostrils, tip of the nose, and the edge of the face are localized as facial feature points. After detecting the facial feature points the first step is to track these points throughout the image sequence. Currently, facial feature point's localization is usually carried out by manually labeling the points.

3.1 Face Detection
Face detection is the first step in our facial expression recognition system to localize the face in the image. A real-time face detector proposed in [12] which represents an adapted version of the original Viola-Jones face detector (Fig. 1-a). The next step in the automatic facial point localization is to determine the coarse position for each point. To achieve this, we develop a fully automatic method using an anthropometric model.

Anthropometry is a biological science that deals with the measurement of the human body and its different parts. Data obtained from anthropometric measurement informs a range of enterprises that depend on knowledge of the distribution of measurements across human populations. After carefully performing anthropometric measurement on Cohn Kanade database [13], we have been able to build an anthropometric model of human face that can be used in localizing facial feature points from face images. The landmarks points that have been used in our face anthropometric model for facial feature localization are represented in Fig. 1-e. It has been observed from the statistics of proportion evolved during our initial observation that location of these points (P1 to P38) can be obtained from: The distance D measured between eyes axis EA and mouth axis MA (Fig. 4-c); Face symmetry axis SA as reference for x position of points.
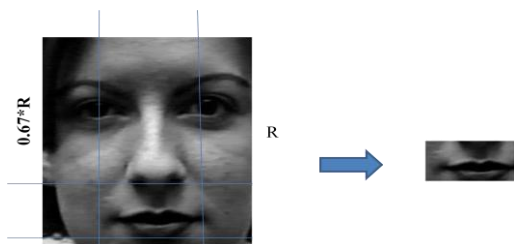
The facial feature point's localization has been found from proportional constants in (Table I) using distance between eyes axis, mouth axis, symmetry axis and the face box center YC as the principle parameter of measurement. To localize feature points on face we have to find out MA, SA, EA and center of face box.

(a) Face Detection
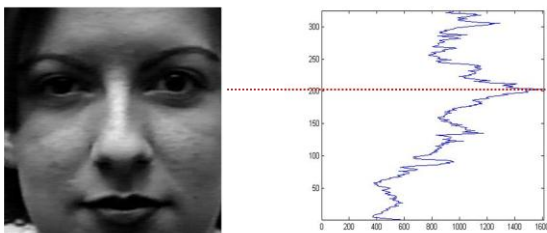
(b) Eye Axis localization

(c) Mouth Axis Localization

(d) Points Localization using Anthropometric model

Fig. 1: Facial feature points detection outline

3.2 Main Axis Localization

The key step of our anthropometric model realization is facial features axis localization. Fig. 4-c shows the three main axis which passed by facial features: two horizontal axis for mouth (MA) and eyes (EA) and vertical axis passed by nose which give the symmetry of the face (SA). YC is the face box center.

3.1.1 *Eyes axis localization*

It is determined by the maximum of the projection curve which has a high gradient. First, we calculate the gradient of the image I (corresponds to the face rectangle extracted by the Viola-Jones detector):

$$\nabla I_x = \frac{\delta I}{\delta x}\vec{I} \qquad (1)$$

∇I_x corresponds to the differences in the x direction. The spacing between points in each direction is assumed to be one. Computing the absolute gradient value in each line is given by:

$$HI_x(x) = \sum_{y=1}^{n} \nabla I_x(x, y) \qquad (2)$$

Then, we find the maximum value which corresponds to the line contains eyes (Fig. 1-b). This line corresponds to many transitions: skin to sclera, sclera to iris, iris to pupil and the same thing for the other side (high gradient).

Table 1. Proportion of facial feature points positions measured from subjects of different geographical territories

Point	X position	Y position
P1	SA-0.91*D	YC-0.91*D
P2	SA-0.58*D	YC-1.17*D
P3	SA-0.26*D	YC-1.3*D
P4	SA+0.26*D	YC-1.3*D
P5	SA+0.58*D	YC-1.17*D
P6	SA+0.91*D	YC-0.91*D
P7	SA+1.04*D	YC
P8	SA+0.71*D	YC+0.91*D
P9	SA+0.52*D	YC+1.17*D
P10	SA+0.13*D	YC+1.3*D
P11	SA-0.13*D	YC+1.3*D
P12	SA-0.52*D	YC+1.17*D
P13	SA-0.71*D	YC+0.91*D
P14	SA-1.04*D	YC
P15	SA-0.06*D	YC-0.26*D
P16	SA+0.06*D	YC-0.26*D
P17	SA-0.78*D	YC-0.52*D
P18	SA-0.58*D	YC-0.58*D
P19	SA-0.26*D	YC-0.52*D
P20	SA+0.26*D	YC-0.52*D
P21	SA+0.58*D	YC-0.58*D
P22	SA+0.78*D	YC-0.52*D
P23	SA-0.26*D	YC+0.32*D
P24	SA+0.26*D	YC+0.32*D
P25	SA-0.65*D	MA
P26	SA-0.32*D	MA-0.1*D
P27	SA	MA-0.15*D
P28	SA+0.32*D	MA-0.1*D
P29	SA+0.45*D	MA
P30	SA+0.32*D	MA+0.1*D
P31	SA+0.19*D	MA+0.13*D
P32	SA	MA+0.15*D
P33	SA-0.19*D	MA+0.13*D
P34	SA-0.32*D	MA+0.1*D
P35	SA-0.74*D	YC-0.26*D
P36	SA-0.58*D	YC-0.39*D
P37	SA+0.58*D	YC-0.39*D
P38	SA+0.74*D	YC-0.26*D

3.2.2 Mouth axis localization

To locate the mouth axis, we first define a Region of Interest (ROI) of the mouth to be the horizontal strip whose top is at 0.67*R from the face bounding box top and has a width equal to 0.25*R. This strip is located around the median of the bounding box of the face with a width of 0.1*R, where R is the side face box. Intensity information is used to locate mouth axis, minimum intensity at mouth.

3.2.3 Symmetry axis localization

It is a vertical line which divides the frontal face in two equal sides. To locate the symmetry axis, we first define a ROI of the nose. Since knowing the location of eyes axis (EA) and mouth axis (MA), we define the nose region to be the vertical strip whose top is the eyes axis and has a height equal to D. This strip is located around the median of the bounding box of the face with a width of 10% of the face windows width. Analysis of the gray level vertical projection of the ROI of the nose shows that the maximum of the projection curves corresponds to symmetry axis

Experimental results show that the extraction of the facial feature points using anthropometric model gives a good location independently to skin color and illumination changes (Fig. 2).

4. FEATURE POINTS TRACKING

To track the localized feature Lucas Kanade Optical flow tracker is used [14]. Optical flow is defined as an apparent motion of image brightness. Two main assumptions can be made Su & Hsieh (2007):

1. Brightness $I(x, y, t)$ smoothly depends on coordinates x, y in greater part of the image.
2. Brightness of every point of a moving or static object does not change in time [14].

Fig. 2: Facial feature point's localization

Let some object in the image the image or some point of an object, move and after time *dt* the object displacement is *(dx , dy)*. Using Taylor series for brightness *I(x, y, t)*

$$I(x+dx, y+dy, t+dt) = I(x,y,t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + ... \quad (3)$$

Then according to assumption 2:

$$I(x+dx, y+dy, t+dt) = I(x,y,t) \quad (4)$$

And

$$\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + ... = 0 \quad (5)$$

Usually above equation called optical flow constraint equation, where

$$\frac{dx}{dt} = u \qquad \frac{dy}{dt} = v$$

Are component of optical flow field in *x* and *y* coordinates respectively.

Calculate optical flow returns to calculate for each point in the image the following equation:

$$-\frac{\partial I}{\partial t} = u.\frac{\partial I}{\partial x} + v.\frac{\partial I}{\partial y} \quad (6)$$

However, the above equation cannot determine with a single way the optical flow. The indetermination of optical flow due to the absence of global constraint in precedent equations, only gradients which are local measures are taken into account. Lucas and Kanade have added new constraints to ensure the uniqueness of the solution. The method of Lucas and Kanade consists to find point location in next image by applying a calculation of least squares to minimize constraint. They define a pre-neighborliness, and they optimize the above equation to give solution of the following system for n points:

$$\begin{bmatrix} \frac{\partial I}{\partial x}(p_1) & \frac{\partial I}{\partial y}(p_1) \\ . & . \\ . & . \\ . & . \\ \frac{\partial I}{\partial x}(p_i) & \frac{\partial I}{\partial y}(p_i) \\ . & . \\ . & . \\ . & . \\ \frac{\partial I}{\partial x}(p_n) & \frac{\partial I}{\partial y}(p_n) \end{bmatrix} . \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -\frac{\partial I}{\partial x}(p_1) \\ . \\ . \\ . \\ -\frac{\partial I}{\partial x}(p_i) \\ . \\ . \\ . \\ -\frac{\partial I}{\partial x}(p_n) \end{bmatrix} \quad (7)$$

RECOGNITION

Our approach uses a feature based representation of facial data for SVM and RBFNN classifier. It classifies single images taken from an image sequence with respect to six basic emotions of Ekman [15] happy, fear, disgust, anger, sadness, surprise and neutral state. Our work is based on the facial features deformation compared to the neutral state.

Fig. 3: Facial features point's detection and tracking in image sequence.

5.1 Coding

In fact, the human facial expressions originate from the movements of facial muscles beneath the skin. Thus, we represent each facial muscle by a pair of key points, namely dynamic point and fixed point. As shown in Fig. 4-a, the dynamic points can be moved during an expression, while Fig. 4-b shows the fixed points which cannot be moved during a facial expression (face edge, nose root and outer corners of the eyes).

Further, each facial muscle is represented by a distance (Fig. 4-d), as: Eyebrows motions are described by the distances from D_1 to D_7, Eyes motions are described by the distances D_8 and D_9, Nose motions are described by the distances D_{10} and D_{11}, Mouth motions are described by the distances from D_{12} to D_{21}. These distances are calculated by Euclidean distance formula.

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \qquad (8)$$

The used method to encode a facial expression takes into consideration all distances D_i variations of each muscle during the sequence. If $D_T = (d_1, d_2... d_i... d_{21})$ is the vector parameters extracted from a video sequence at the moment T.

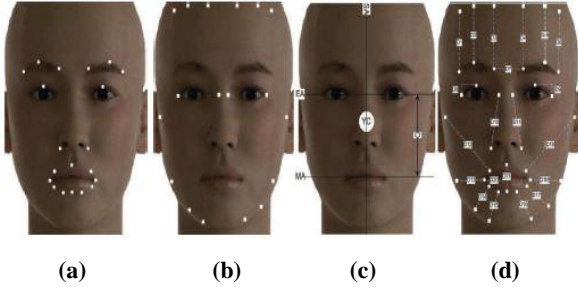**(a)**   **(b)**   **(c)**   **(d)**

Fig. 4: (a) Dynamic Points, (b) Static Points, (c) Principal Axis, (d) Facial Distances

ΔD is the distance variation from the first image (a neutral expression), where:

$$\Delta D = \left(\frac{D_1}{D_{01}}, ... \frac{D_i}{D_{0i}}, ... \frac{D_{21}}{D_{021}} \right) \qquad (9)$$

Where D_{0i} is the ith distance of the neutral state, with i ϵ [1, 21]. After extracting distance feature vector from static and dynamic points, they are normalized in between 0 to 1.

5.2 Support Vector Machine

After the extraction of the necessary information from the facial expression, we have trained a statistical classifier Support vector machine SVM [16]. Support Vector Machine makes binary decisions.

$$f(\vec{x}) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x} + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x} + b \leq -1 \end{cases}$$

There are a number of methods for making multiclass decisions with a set of binary classifiers. The simplest strategy is to train 1 versus all remaining, but this method gives poor results. To overcome this we adapted 1 versus 1 strategy. We trained test emotion with all possible combination of other pair. We get such 15 pair for six basic emotions. Then the polling method is used the class having maximum voting, the test emotion belongs to that class. In general, the RBF kernel is a reasonable first choice. This kernel nonlinearly maps samples into a higher dimensional space so it, unlike the linear kernel, can handle the case when the relation between

class labels and attributes is nonlinear. In addition, the sigmoid kernel behaves like RBF for certain parameters. The second reason is the number of hyper parameters which increase the complexity of model selection. The polynomial kernel has more hyper parameters than the RBF kernel. The rbf kernel is as follow.

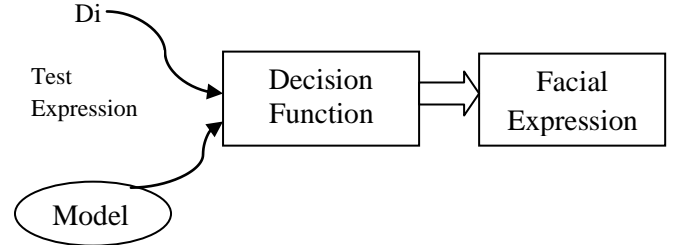$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|}{\sigma} \right)$$

Di

Test Expression → Decision Function → Facial Expression

Model

Fig. 5: Facial expression recognition system

5.3 Radial Basis Function Neural Network

The basic architecture for a RBF is a 3-layer network; input layer is simply a fan-out layer and does no processing. The second or hidden layer performs a non-linear mapping from the input space into a (usually) higher dimensional space in which the patterns become linearly separable. The final layer therefore performs a simple weighted sum with a linear output. If the RBF network is used for function approximation (matching a real number) then this output is fine. However, if pattern classification is required, then a hard-limiter or sigmoid function could be placed on the output neurons to give 0 or 1 output values. The unique feature of the RBF network is the process performed in the hidden layer [17]. The idea is that the patterns in the input space form clusters. If the centres of these clusters are known, then the distance from the cluster centre can be measured. Furthermore, this distance measure is made non-linear, so that if a pattern is in an area that is close to a cluster centre it gives a value close to 1. Beyond this area, the value drops dramatically. The notion is that this area is radically symmetrical around the cluster centre, so that the non-linear function becomes known as the radial-basis function. The most commonly used radial-basis function is:

Where σ is the spread parameter of the Gaussian functions, r is the distance from the cluster centre. The distance measured

$$\varphi(r) = \exp\left(-\frac{r^2}{2\sigma^2} \right)$$

from the cluster centre is usually the Euclidean distance. For each neuron in the hidden layer, the weights represent the co-ordinates of the centre of the cluster. Therefore, when that neuron receives an input pattern, *x,* the distance is found using the following equation.

Number of neurons is 250 and RBF spread value is 250 used in this system.

$$r_j = \sqrt{\sum_{i=1}^{n} (x_i - w_{ij})^2}$$

6. RESULT AND DISCUSSION

For the evaluation of work, we have used Cohn-Kanade [13] databases. The systems will implement on MATLAB 7.5 version. Fig. 5 describes our facial expression recognition system. Performance of system is checked using SVM and RBFNN classifier.

The experiment performed using cross validation in order to compute the accuracy. In Cohn Kanade database contain 50 emotions for each class, so initially we used 40 emotions from each class for training and remaining 10 emotions for testing. Then cross validation is used to increase the database size. Totally 50 emotions from each class are used as testing purpose. Table 2 and Table 3 represents confusion matrix of different emotions using SVM and RBFNN classifier respectively on Cohn-Kanade database. This matrix shows the effectiveness of classification methods with the SVM. For our classifier, we achieve the recognition rate averagely 91%. Fig. 6 shows the recognition rate for different expression using SVM classifier.

Table 2. Emotion confusion matrix using SVM classifier on Cohn kanade database

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	47	0	2	0	4	0
Disgust	0	41	1	5	2	1
Fear	1	1	47	0	0	0
Happy	0	6	0	45	0	0
Sad	2	2	0	0	44	0
Surprise	0	0	0	0	0	49
Average Recognition rate=91%						

Table 3. Emotion confusion matrix using RBFNN classifier on Cohn kanade database

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	47	0	2	0	5	0
Disgust	1	42	1	5	2	1
Fear	1	0	45	1	0	0
Happy	0	5	0	44	0	0
Sad	1	3	2	0	43	0
Surprise	0	0	0	0	0	49
Average Recognition rate=90%						
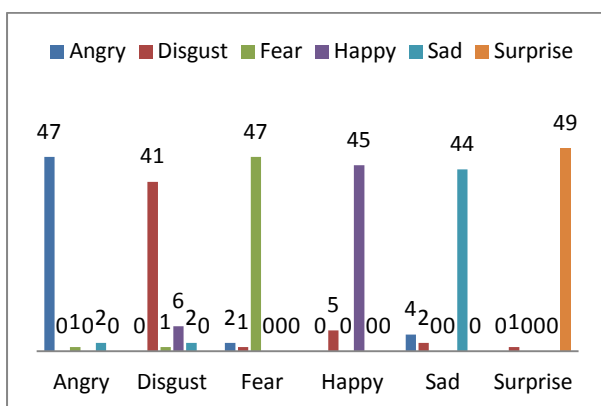
Fig. 6: Recognition rate for different expression using SVM Classifier

7. CONCLUSION AND FUTURE SCOPE

In this paper an automatic approach to emotion recognition based on facial expression analysis presented. To detect the face in the image, we have used the face detector of Viola Jones which is fast and robust to illumination condition. For feature point's extraction, we have developed anthropometric models which suppose that the positions of facial feature points are proportional to the vertical distance between eyes and mouth. Then that feature points are tracked throughout the image sequence. The proposed method gives good results, when tested on images from the Cohn-Kanade database, under various illuminations. The tracking of the detected points have been realized with Lucas-Kanade algorithm. Variations distance vector was used as a descriptor of the facial expression. This vector is the input of SVM and RBFNN classifier. Emotion recognition rates about 91 % were achieved in real time.

Proposed combination method for feature extraction does not extract features parameters properly if there are hairs on face area. Therefore in future an attempt can be made to develop hybrid approach for facial feature extraction and recognition accuracy can be further improved using NN approach and hybrid approach such as ANFIS. An attempt can also be made for recognition of other database images or images captured from camera

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] A. Mehrabian, "Communication without Words," Psychology Today, Vo.1.2, No.4, pp 53-56, 1968.

[2] Z. Xin, X. Yanjun, and D. Limin. "Locating facial features with color information". IEEE International Conference on Signal Processing, 2 :889–892, 1998.

[3] Shishir Bashyal, Ganesh K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization". Elsevier Engineering Applications of Artificial Intelligence 21 (2008) 1056–1064.

[4] F. ABDAT, C. MAAOUI and A. PRUSKI, "Human-computer interaction using emotion recognition from facial expression" 2011 UKSim 5th European Symposium on Computer Modeling and Simulation, 978-0-7695-4619, 2011.

[5] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron". in Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara Japan, 14-16 April 1998, pp. 454.459.

[6] B. Fasel, "Multiscale facial expression recognition using Convolutional neural networks". IDIAP, Tech. Rep., 2002.

[7] M. Matsugu, K. Mori, Y.Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a Convolutional neural network". Neural Networks, vol. 16, no. 5-6, pp. 555.559, June-July 2003.

[8] I. Cohen Y. Sun T. Gevers N. Sebe, M.S. Lew and T.S. Huang. "Authentic facial expression analysis". Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition (AFGR), 2004.

[9] J. Bailonson, E. Pontikakisb, I. Maussc, J. Grossd, M. Jabone, C. Hutchersond, C. Nassa, and O. Johnf. Real-time classification of evoked emotions using facial feature tracking and physiological responses. International Human Computer Studies, 66 :303–317, 2008.

[10] Koen van de Sande Roberto Valenti Aitor Azcarate, Felix Hageloh. Automatic facial emotion recognition, 2005.

[11] P.W. Yuille, A.L.and Hallinan and D.S. Cohen. "Feature extraction from faces using deformable templates". International Journal of Computer Vision, 8:99–111,

[12] P. Viola and M. Jones. "Robust real-time object detection". 2nd international workshop on statistical and computational theories of vision - modeling, learning, computing, and sampling". canada, 2001.

[13] T. Kanade, J. F. Cohn, and Y. Tian. "Comprehensive database for facial expression analysis". Fourth IEEE International Conference on Automatic Face and Gesture Recognition Grenoble France, FG'00:46–53, 2000.

[14] J.Y. Bouguet. "Pyramidal implementation of the Lucas kanade feature tracker". Intel Corporation, Microprocessor Research Labs, 2000.

[15] P. Ekman. Emotion in the human face. Cambridge University Press, 1982.

[16] S. Gunn, "Support vector machines for classification and regression". Image Speech and Intelligent System Group, Univ. of Southampton MP-TR-98-05, 1998.

[17] D. T. Lin. (2006). Human facial expression recognition using hybrid network of PCA and RBFN, Lecture Notes on Computer Science 4132, 624–633.