Real Time Facial Expression Recognition Based On Hierarchal SVM

Nehal O. Sadek[1], Noha A. Hikal[2], Fayez W. Zaki[3]

[1,2] Department of Information technology, Faculty of Computers and Information Sciences, Mansoura University, Mansoura, Egypt.
[3]Departement of communications engineering, Faculty of engineering, Mansoura University, Mansoura, Egypt.
nehal_osama12@yahoo.com, nmhikal_5@yahoo.com, fwzaki@yahoo.com

Abstract

Facial expression analysis is one of the essential medium of behavior interpretation and emotion modeling. In this paper, the proposed approach presents a real time approach for emotion recognition through facial expression in live video. The proposed approach employs an automatic facial feature tracker to perform face localization and feature extraction. The facial feature displacements in the video stream are used as input to Hierarchal a Support Vector Machine for classifying emotions. Our approach aim to recognize facial expressions in few msec with high accuracy rate combines two steps (classifications, post processing) by using hierarchical SVM .The proposed approach has achieved a tradeoff between accuracy and result rate. A processing time of 5-6 ms per 100 frames was achieved with accuracy of around 90%.

Keywords: *Face expression, Anthropometric model, Hierarchal Support Vector Machine.*

1. Introduction

Human Computer Interaction (HCI) nowadays focused on development a natural human interaction with the computers based on the normal human to human behavior interaction. Recent psychological research has shown that facial expressions are the most expressive way in which humans display emotion, so it is one the most active research area in the field of HCI. Facial expression recognition is a sort of visual learning process. Applications include video conferencing, forensics, virtual reality, computer games; machine vision etc., Facial Expression Recognition (FER) from video is an essential research topic in computer vision, impacting important applications in areas such as human–computer interaction and data-driven animation. In general, facial expressions are divided by psychologists into six basic categories: anger, disgust, fear, happiness, sadness, and surprise [1].

In this paper our approach proposed automatic method for dynamic system unlike other approaches [2, 3, 4, 5, 6)]. For classification step our approach proposed hierarchal SVM that fast and intuitive SVM [7]. The proposed hierarchal SVM classification consists of two levels. First level maps the incoming expression to category that contains two expressions which look similar at mouth movement. Second level classify between this two expressions to recognize the incoming expression by taking more features. The proposed approach can recognize six facial expressions compared with [6, 4].

This paper is organized as follows: section 2 and 3 present briefly the recent techniques which are commonly used for face detection and dynamic facial feature extraction, respectively. Section 4 introduces the optical flow point tracking technique and how it is employed to track the dynamic features of face. Section 5 discusses the idea of facial expression recognition based on the distance vectors. Section 6 describes the proposed hierarchical SVM for facial expression recognition. Section 7 introduces discussion of the experimental simulation results and comparison with previous work. Finally a conclusion and suggestions for future work is presented in section 8.

2. Face Detection

Automatic systems for facial expression recognition usually take the form of a sequential configuration of processing blocks, which adheres to a classical pattern recognition model [8]. The main blocks are: image acquisition, pre-processing, feature extraction, classification, and post-processing. The pre-processing blocks: convert frame to gray scale for detection face. After detecting face apply sobel derivative to get maximum gradient projection curve for finding eyes line. For detect center of mouth converting mouth region from RGB to HSV in order to separate color from intensity.

Face detection is the first step in the proposed approach system, which consist to delimit the face area with a rectangle. This step applies only on the first frame then in the other frames tracking points that will define. The proposed approach use adaptive version of Viola-Jones face detector that based on the Haar-like features Viola & Jones [9]. Three major contributions of the algorithm: Feature extraction, Classification using boosting and Multi-scale detection algorithm .At Feature extraction rectangular features are used, with a new image representation their calculation is very fast. A Classifier training using a slight variation of a method called AdaBoost [9].

This approach minimizes computation time while achieving high detection accuracy. The approach was used to construct a face detection system which is approximately 15 faster than any previous approach [9].as shown at figure 1-a, 1-b.

3. Facial Feature Extraction

Generally, feature extraction aims to reduce the dimensionality of the input image while preserving the most essential data into it. Features that should be extracted for real time facial expression recognition must represent highly the motions corresponds to all muscles activation points during the expression. Therefore, spatial feature is the most appreciated one for this application.

The most regions that contain information to recognize facial expression are mouth, eyes, eyebrows and nose but mouth is the most important region [10]. The facial features are based on anthropometric model that based on spatial features not transforms [11]. 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. The proposed approach build an anthropometric model of human face that can be used in localizing facial feature points from face images. This technique is faster than [4, 6, 5, 10, 12].

3.1 Detect eyes line

The employed technique detects the horizontal axis of eyes by the maximum of the projection curve which has a high gradient H.

Calculate the gradient of image (I) that contains detected face:

$$\nabla I_x = \frac{\delta i}{\delta x} I \qquad (1) \qquad [13]$$

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

$$H\,Ix(x) = \sum_{y=1}^{n} \nabla Ix(x, y) \qquad (2) \qquad [13]$$

Then, the employed technique find the maximum value which corresponds to the line contains eyes. [13] as shown at figure 1-c.

3.2 Detect mouth line

To locate the mouth axis, the employed technique 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 X 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 width of face [14].

Then the employed technique converts color system of mouth from RGB to HSV in order to segment mouth area [15]. The center of mouth height corresponds to the mouth line (ML) [14] as shown at figure 1-c.

3.3 Determine center lines

The horizontal line obtained by dividing face height by two this line called the center of face height (CHL). The vertical line divides the frontal face in two equal sides. In other words, it is the line passed by the nose. This symmetry axis obtained by dividing face width by two this line called the center of face width (CWL) [14] as shown at figure 1-c.

3.4 Feature point detection

Data obtained from anthropometric measurement of determined axis and D. The distance D measured between eye line and mouth line [14]. The employed technique determine 38 points, 10 for mouth, 4 for eyes, 6 for eyebrows, 4 for nose and 14 for face edge as shown at figure 1-d.

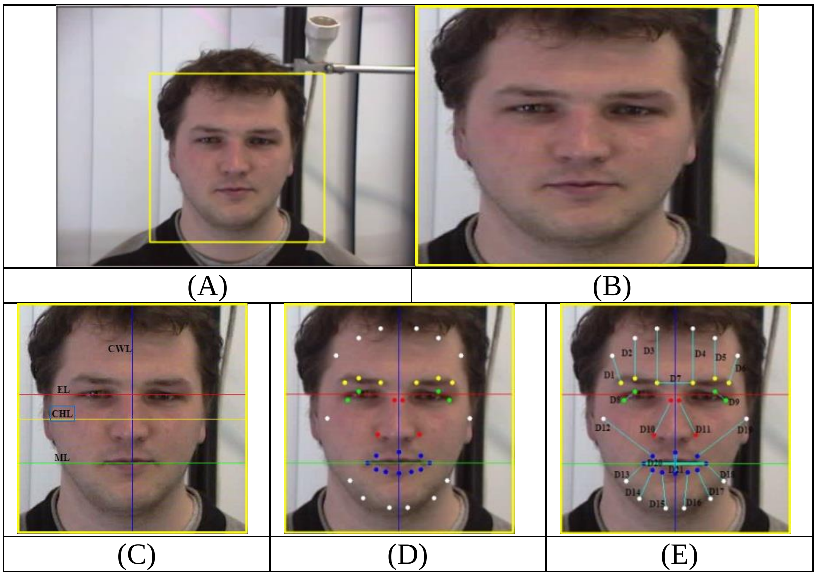


Figure 1. Steps of face feature detection

4. Optical Flow Tracking

After detecting points on the first frame the employed technique use Pyramidal Implementation of the Lucas Kanade Feature Tracker algorithm to tracking points [16]. This algorithm widely used with almost optimal performances which suppose that the brightness of every point of a moving or static object does not change in time. Consider an image point u = $[u_x \; u_y]^T$ on the first image I where u_x and u_y are the two pixel coordinates of point u. The goal of feature tracking is to find the location v = u+d = $[u_x + d_x \; u_y + d_y]^T$ on the second image J such as I(u) and J(v) are "similar" where v is the new pixel of u .

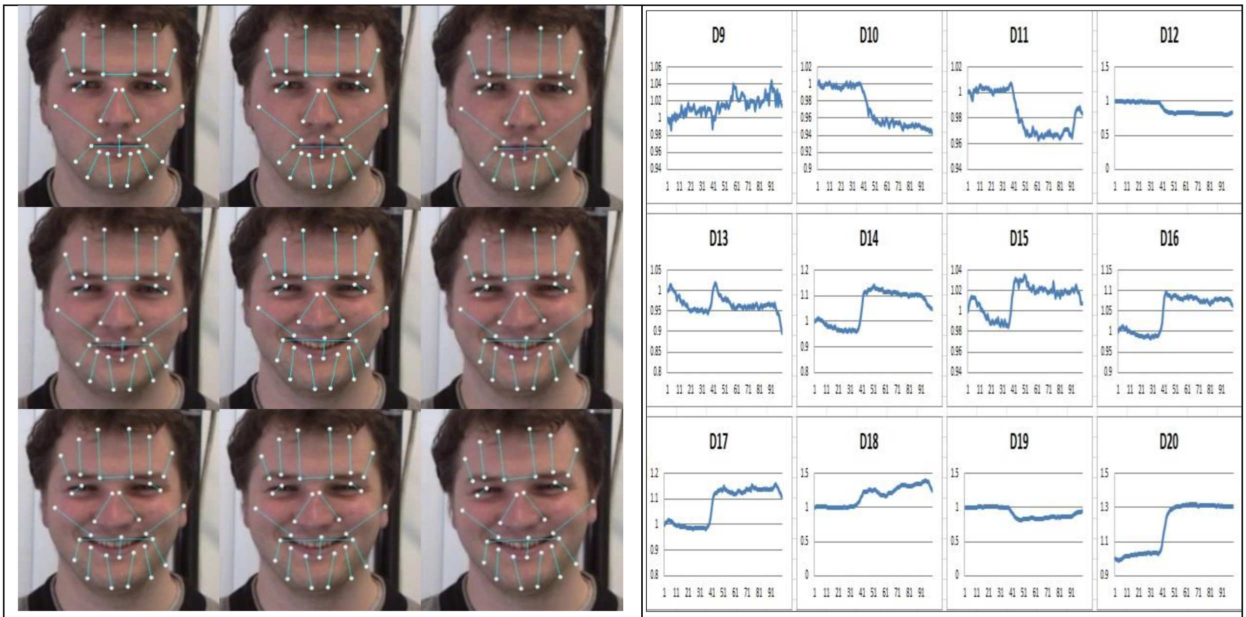
The vector d = $[d_x \; d_y]^T$ is the image velocity at x also known as the optical flow at x, where x is a generic image point. Because of the aperture problem, it is essential to define the notion of similarity in a 2D neighborhood sense. Let w_x and w_y two integers. Defining the image velocity d as being the vector that minimizes the residual function ε defined as follows [16]:

$$\varepsilon(\mathrm{d}) = \varepsilon(\mathrm{d_x}, \mathrm{d_y}) = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} \left(\mathrm{I}(\mathrm{x}, \mathrm{y}) - \mathrm{J}(\mathrm{x} + \mathrm{d_x}, \mathrm{y} + \mathrm{d_y})\right) 2 \qquad (3) \quad [16]$$

5. Facial Expression Recognition

The employed technique relies on the distance vector retrieved from 2D distribution of facial feature points to classify universal facial expression. The employed technique defines 21 distances on face [14] as shown at figure 1-e. Eyebrows are described by 7 distances from D1 to D7. Eyes are described by 2 distances D8 and D9. Nose is described by 2 distances D10 and D11.Mouth is described by 10 distances from D12 to D21. The distance vector contains all distances D1 to D21 per 100 frame.

For data normalization, each distance divided by it's in first frame that represent the nature expression for example D21 at first frame $D21_0$, at third frame $D21_2$ and normalized distance D21 at third frame $nD21_2$; So $nD21= D21_2/ D21_0$ figure 2-b.



(A) (B)

Figure 2. Point tracking, distances changes of happy expression

6. Hierarchal SVM for Facial Expression Recognition

SVM used for classification and regression analysis. Since, it exhibits; high classification accuracy for small training sets, good generalization, and high performance specially with highly correlated data [17].

The proposed approach employs a hierarchal SVM based on RBF kernel [18]. Since there are some expressions looks similar in few distances The proposed technique divides the different expressions into categories, each category consists of two expressions depending on common distances (mouth width and mouth height (D20, D21)) in each of them.

The hierarchal SVM contain two levels, figure 3. The first level, SVM0 can classify the incoming distance vectors, based on mouth movement; into new categories each has two expressions. The two expressions are the most similar ones. Now, at the second level; the recognition zone is highly delimited into two expressions. So at this level, the task is easier since the classifier chooses between only two out of six expressions. Each category has a SVM classifier that can decide which one of the two expressions is correct based on certain distance features corresponds to each category. The number of classification distances vector at second level according to different categories is shown in table 1. Algorithm 1 summarizes the proposed technique.

The labeled distance vector of each example expression supplied is used as input to an SVM classifier, resulting in a model of the training data, which is subsequently used to dynamically classify unseen feature displacements.

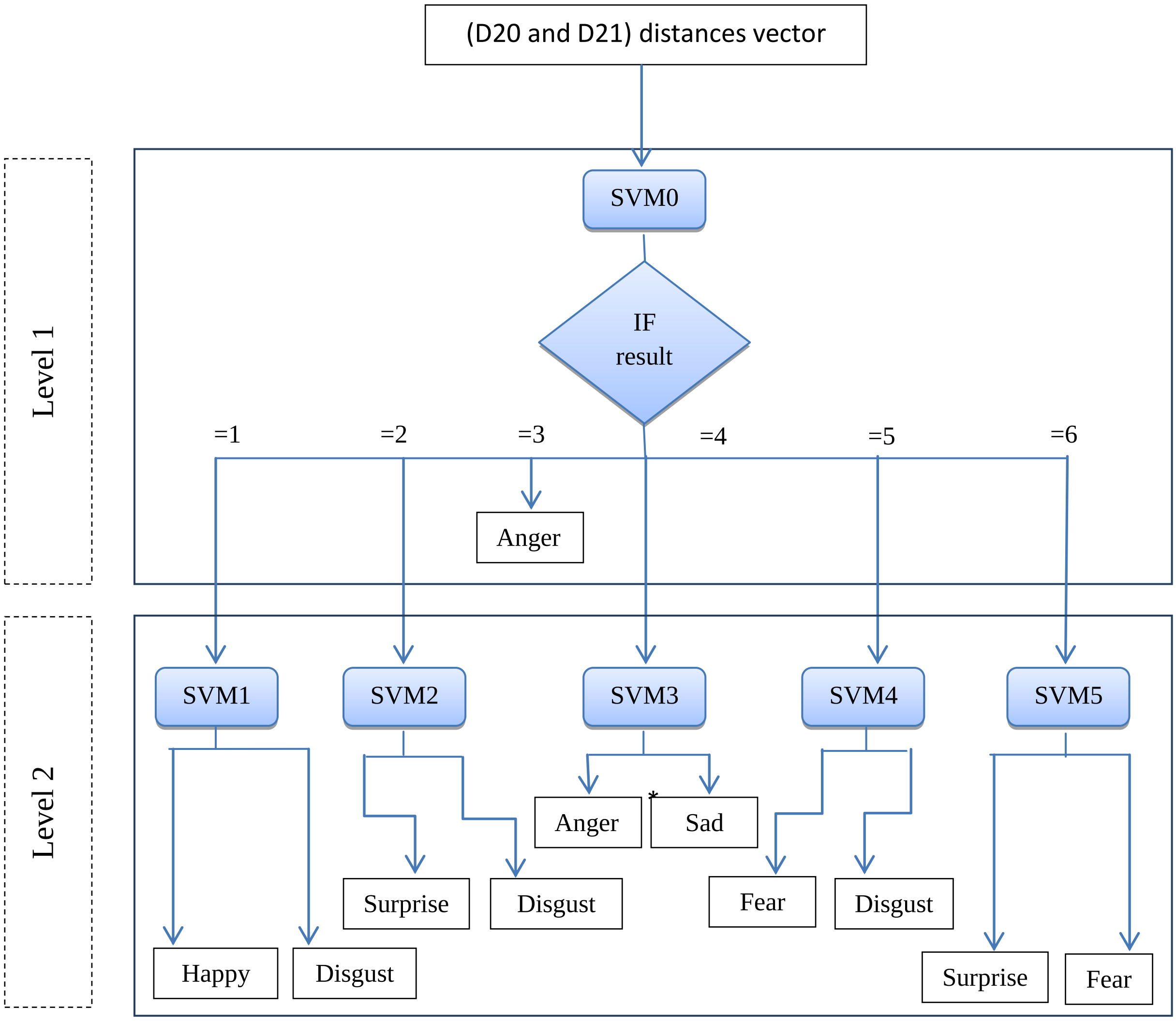


Figure 3. Hierarchical SVM steps

Table 1. SVM classifiers

	Expressions	Vector Distances	Number of distances
SVM0	All expressions	D20,D21	2
SVM1	Happy, Disgust	D12:D21	10
SVM2	Surprise, Disgust	D1:D7	7
SVM3	Anger, sad	D1:D21	21
SVM4	Fear, Disgust	D1:D21	21
SVM5	Surprise, Fear	D1:D21	21

Algorithm 1: Proposed hierarchal SVM for facial expression recognition:

Hierarchical SVM classification
Input distance vector to SVM0
Switch response0
Case 1: Input distances D12:D21 to SVM1
    Switch response1
    Case 1:Happy
    Case 2:Disgust
    End
Case 2: Input distances D1:D7 to SVM2
    Switch response2
    Case 1:Surprise
    Case 2:Disgust
    End
Case 3: Anger
Case 4: Input distances D1:D21 to SVM3
    Switch response3
    Case 1:Anger
    Case 2:Sad
    End
Case 5: Input distances D1:D21 to SVM4
    Switch response4
    Case 1:Fear
    Case 2:Disgust
    End
Case 6: Input distances D12:D21 to SVM5
    Switch response5
    Case 1:Surprise
    Case 2:Fear
    End
End

7. Testing and Result

A MATLAB simulation program was designed to test and evaluate the proposed technique. The simulations were tested using the standard FEEDTUM database [19]. In the training mode of SVM used 10 distances vector for each expression as training data. In the testing mode used 120 difference distances vector for evaluate system. The time average of point tracking is 111 ms per frame and the average time of recognizing expression is 5 ms per 100 frames, which gives the proposed technique the superiority as a real time video recognition technique. Table 2 presents the accuracy rate of six expressions. Compared with other techniques [10, 14, 20], the proposed system achieved an average of 90.5% accuracy rate, as show at table 3.

Table 2. facial expression recognition by hierarchical SVM

	Anger	Disgust	Happy	Fear	Sad	Surprise
Anger	100%	0	0	0	0	0
Disgust	0	95%	0	0	0	5%
Happy	0	5%	95%	0	0	0
Fear	0	0	0	50%	15%	35%
Sad	0	0	0	0	100%	0
Surprise	0	0	0	0	0	100%

Table 3. Related work

Reference	Number of expressions	Database	Feature extraction	classification	Number of frame per sec	Accuracy rate	Recognition Time msec
[10]	Six	Cohn_Kanade	PLBP features	SVM	30	96%	10
[14]	Three	FEEDTUM	Geometric facial features &Appearance Features	SVM	25	90%	31
[20]	Four	CMU Multi-pie	Geometric facial features	SVM	10	60%	100
our	Six	FEEDTUM	Geometric facial features	Hierarical SVM	25	90.5%	5

8. Conclusion and Future Work

In this paper the proposed approach has presented a fully automated system for facial expression recognition in sequences of images. To detect face the proposed approach used Viola-Jones face detector which is fast and independent on illumination. To extract feature the proposed approach has develop anthropometric method that detect features fast by using the constant ratios of distances between mouth and eyes after detect main axis. Determining distances vector of video frames by using FEEDTUM database. For classification the proposed approach uses hierarchical SVM classifier with RBF kernel for decrease time. The proposed approach achieve to 5 ms for classification because of using two level of SVM first level contain one SVM classifier that take only two distances as input vector. Then pass on second level with decision that limit classification expressions to two overlap expression. The proposed approach achieve to 90% accuracy rate. All expression without fear have good accuracy rate. Fear is overlapping with surprise and sad, so its accuracy rate is 50%. The low accuracy rate of fear because system based on distances and fear expression has tiny features.

For future work, the proposed approach plan to improve accuracy by using displacements based on elastic bunch graph matching displacement estimation. In addition, combining the head motion measures and steering correlations with facial movement measures is expected to improve the recognition accuracy.

References

[1] E. Jack, C.Blais, C.Scheepers, G. Schyns and R.Caldara, “Cultural Confusions Show that Facial Expressions Are Not Universal”, Current Biology, Canada, September 2009.

[2] Daw-tung Lin, “facial expression classification using pca and hierarchical radial basis function network”, department of computer science and information engineering national taipei university, taiwan, 2006.

[3] G. R. S. Murthy and R.S.Jadon, “Effectiveness of Eigen spaces for Facial Expressions Recognition”, 2009.

[4] Mandeep Kaur , Rajeev Vashisht and Nirvair Neeru, "Recognition of Facial Expressions with Principal Component Analysis and Singular Value Decomposition", 2010.

[5] Xiaoming Zhao and Shiqing Zhang , "Facial Expression Recognition Based on Local Binary Patterns and Kernel Discriminant Isomap" , China, 2011.

[6] Senthil Kittusamy and Venkatesh Chakrapani, "Facial Expressions Recognition Using Eigenspaces", Tamilnadu, India, 2012.

[7] Lee Y, Chang Y, Kim N, Lim J, Seo JB and Lee YK. "Fast and efficient lung disease classification using hierarchical one-against-all support vector machine and cost-sensitive feature selection", korea, 2012.

[8] Claude C. Chibelushi and Fabrice Bourel,"Facial Expression Recognition: A Brief Tutorial Overview"

[9] M.Jones and P.Viola, "Robust Real-time Object Detection", Second International Workshop on Statistical and Computational Theories of Vision – modeling, learning, computing, and sampling, Vancouver, Canada, July 2001.

[10] R.Khan, A.Meyer, H.Konik and S.Bouakaz, "Framework for reliable,real-time facial expression recognition for low resolution images", SienceDirect, France, 2013.

[11] A.Azcarate, F.Hageloh, R.Valenti, "Automatic facial emotion recognition", June 2005.

[12] A. Sumathi and P. Marasamy, "Automatic Recognition and Analysis of Human Faces and Facial Expression by LDA using Wavelet Transform", india, 2012.

[13] F. Abdat, C. Maaoui and A. Pruski, "Real Time Facial Feature Points Tracking with Pyramidal Lucas-Kanade Algorithm", IEEE ROMAN08, The 17th International Symposium on Robot and Human Interactive Communication, Germany, 2008.

[14] A. Pruski, C. Maaoui and F. Abdat, "Human-computer interaction using emotion recognition from facial expression", European Symposium on Computer Modeling and Simulation, Metz, France, 2011.

[15] S.sural, G.qian and S.pramanik, "Segmentation and Histogram Generation using the HSV Color Space for Image Retrieval", ICIP, USA, 2002.

[16] J. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm", Intel Corporation Microprocessor Research Labs, 2000.

[17] S.Maji, A.Berg and J.Malik, "Classification using Intersection Kernel Support Vector Machine is Efficient", IEEE Computer Vision and Pattern Recognition, 2008.

[18] Hichem Sahbi and Donald Geman, "A Hierarchy of Support Vector Machines for Pattern Detection", Journal of Machine Learning Research, USA, 2006.

[19] http://cotesys.mmk.e-technik.tu-muenchen.de/isg/content/feed-database,1September 2013.

[20] saumil srivastava, "real time facial expression recognition using a novel method", noida, india, 2012.