

# Intelligent Real-Time Facial Expression Recognition from Video Sequences based on Hybrid Feature Tracking Algorithms

Nehal O. Sadek

Information Technology Department,  
Faculty of Computers and  
Information Sciences,  
Mansoura University,  
Mansoura - Egypt

Noha A. Hikal

Assoc. Prof. at  
Information Technology Department,  
Faculty of Computers and  
Information Sciences,  
Mansoura University, Mansoura –  
Egypt

Fayez W. Zaki

Prof. at communications engineering  
Department,  
Faculty of Engineering,  
Mansoura University,  
Mansoura – Egypt

**Abstract**—In this paper, a method for automatic facial expression recognition (FER) from video sequences is introduced. The features are extracted from tracking of facial landmarks. Each landmark component is tracked by appropriate method, results in proposing a hybrid technique that is able to achieve high recognition accuracy with limited feature dimensionality. Moreover, our approach aims to increase the system accuracy by increasing the FER recognition accuracy of the most overlapped expressions while achieving low processing time. Thus, the paper introduces also an intelligent Hierarchical Support Vector Machine (HSVM) to reduce the cross-correlation between the confusing expressions. The proposed system was trained and tested using a standard video sequence dataset for six facial expressions, and compared with previous work. Experimental results show an average of 96% recognition accuracy and average processing time of 93 msec.

**Keywords**—facial expression recognition (FER); Hierarchical Support Vector Machine (HSVM); Human computer Interaction (HCI); Real-time facial expression recognition

## I. INTRODUCTION

Science of HCI focuses on simulating the human interaction with computers as human like human interaction. The researches in that field care mainly about knowing the human behaviors to facilitate the interaction with computers. The researches show that the facial expression recognition (FER) of human is considered the most important way to represent the emotion which reflects essentially the human behavior. FER has many applications such as video conferencing, medical applications, forensics, virtual reality, computer games; machine vision and many more. FER is categorized according to the type of input data into two types; video based FER, and image based FER. This paper considers mainly real-time video based FER, standard dataset for real-time video facial expression are used for testing and evaluating. Moreover, the proposed system is designed to recognize the most six effective expressions of Human face; anger, disgust, fear, happiness, sadness, and surprise [1,2,3].

Most of video-based real-time FER systems depend on tracking the motion and the position of the muscles in the face.

These movements indicate the emotion state of the human from his face. An automatic FER system usually consists of three steps [1,2]: i) Face detection, ii) Facial feature extraction, iii) Facial expression recognition. Each step of them is considered a separate research area and has its own challenges. On the other hand, when speaking on video-based real-time FER, the trade-off between processing time and recognition accuracy becomes the main challenge. A high level of accuracy and a few milliseconds processing time are the main goal of such system. Many FER systems have been developed to satisfy these goals [1,2,3,4]. However, some limitations still exist. There are some constraints that affect the FER system accuracy and time such as; numbers of features, number of recognized emotions, pose of the face...etc. The goal of this paper is to develop a hybrid feature tracking technique that is able to track the most effective face features, regarding the differences in its movements nature, with the corresponding technique while keeping feature vectors dimensionality as minimum as possible. Moreover, an intelligent Hierarchical SVM classification system is deployed to achieve low processing time and high recognition accuracy.

Previous researches had proven the superiority of FER based geometric feature (GF) compared with that based on appearance feature (AF) [1,2]. The processing time was among the important challenges in FER. In [3,5] the authors achieved accuracy rate up to 90% and 91% within 31 msec processing time, while the processing time was reduced to 5 msec using HSVM with the same accuracy rate [4]. In [6,7,8] using geometric feature extraction achieved 78.3%, 88.8% and 89% with low processing time. The low accuracy rate with low processing time is because of using specific face components like mouth and eyes. In [1], the triangles based feature was deployed instead of distances based feature or points based feature with accuracy rate 95% and 97% by implementing MUG and CK+ datasets. In [2] using points, distances and triangle features in CK+ database is 96.37%, 96.58%, and 97.80%, in MMI database is 67.64%, 74.31%, and 77.22%, and in MUG database is 91.41%, 94.13%, and 95.50%, respectively. The systems [1,2] achieve to high accuracy, on the other hand it shows a high processing time.

The rest of this paper is organized as follows: the proposed framework, face detection and landmark points extraction is described in section 2. The implementing of the three FER systems (distances, triangles and hybrid) are presented in section 3. Section 4 describes the intelligent HSVM of the three systems. Experimental results of the three systems are presented in section 5. Finally, conclusion and future work are given in section 6.

## II. THE PROPOSED FRAMEWORK

Most automatic systems for FER usually consist of main sequence processing blocks [2], these blocks are: video acquisition, frames preprocessing, feature extraction, feature tracking and classification. Fig. 1 shows the framework of the proposed systems. The input of the proposed system is video. The frames are extracted from input video. After the frames are extracted from facial expression video, the face detection algorithm is applied on the first frame. After that the facial landmark points are extracted from the detected face. The accuracy of the geometric FER system mainly depends on the landmark point's detection accuracy. Once the landmark points are extracted, the point tracking algorithm is applied on

the video frames. The feature vector is formed from the tracking results of landmark points. Since the scope of this paper is to accelerate real-time video FER process by integrating the most effective algorithms at each stage, so, the proposed approach implemented three systems which have a different feature vector formation: based on distances, based on triangles and hybrid between them. The distances based system achieves low processing time but some expressions have low accuracy. The triangles based system achieves to high accuracy with high processing time. The hybrid system proposed for improving the accuracy of expressions that have lack of accuracy in distances based system and reduce processing time. The hybrid system recognizes happy, anger and disgust expressions by distances and recognizes the other expressions by triangles. The feature vector input as training data with class label at the intelligent HSVM training stage. The feature vector input as test data after the intelligent HSVM is trained and it can recognize the expression of this feature vector. The proposed approach uses tree to represent HSVM and uses depth-first search algorithm for expression searching. In the coming sub sections, each step will be explained and discussed.

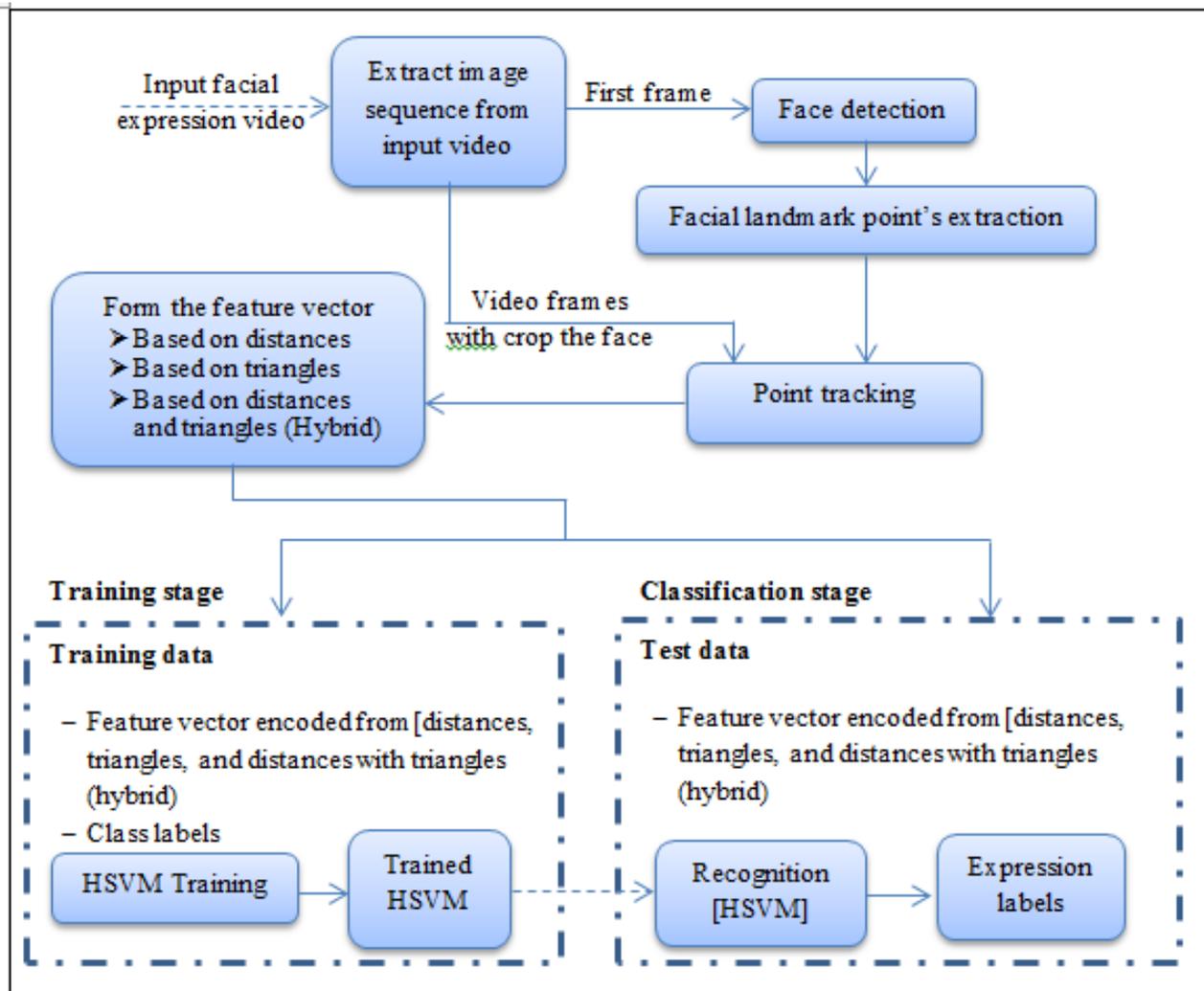


Fig. 1. The steps of the proposed FER system

A. Face Detection and Feature Point's Extraction

Face detection is the first step in any FER system. This step is applied only to the first frame then the face will be tracked through the follow frames using defined specific tracking points. The proposed approach employs an adaptive version of Viola-Jones (VJ) face detector which is based on the Haar-like features [3]. This approach suits mainly real time video applications, since it is approximately 15 times faster than most recent approaches.

The detected face is represented by a number of static and dynamic points. The static points are the points that are not located on the face components (e.g., eyes, eyebrows, nose, and mouth). These points are located on the face border, in addition to two points are located on the nose, total of 8 blue points are shown in Fig. 2. The dynamic points that are located on the face component (eyebrows, eyes, nose, and mouth) are strongly related to FER accuracy. Thus, these points are more important than static points. So, they must be located carefully, shouldn't be determined by face ratio. Since this paper integrates the most powerful algorithms to serve high FER accuracy, the next subsections will introduce briefly the employed algorithms for extracting these dynamic points.

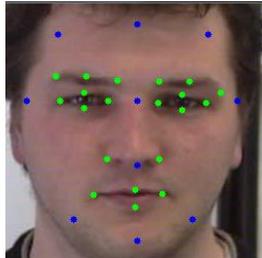


Fig. 2. The static landmark points (blue points) and the feature landmark points (green points)

1) *Eyes and Eyebrows Points' Extraction:* The proposed approach employs the enhanced VJ algorithm [9]. The face is divided into three sub portions: upper left half, upper right half and lower half show in fig. 3. The VJ eyes detection was applied on upper left half and upper right half. Face image division based on physical approximation of location of eyes and mouth on face. This algorithm increases the accuracy of VJ techniques and decreases processing time.

The idea of this algorithm that used for eyes and eyebrows points detection is based on color segmentation, since that the skin pixels would have high red intensities compared to eye and eyebrows pixels.

This algorithm is based on RGB as shown at fig.4 [10]:

- Complement the eye ROI red channel
- Calculate the exponential operator for each pixel in the image  $I=CR$ :
- $I_{i,j}=\exp[I_{i,j},K]$  (1) [10]

Where k is the value of  $\frac{\ln 255}{255}$

- binarizing the image I' as following:
- $Bw=\begin{cases} 1, & \text{if } I' > \bar{I} + z\sigma \\ 0, & \text{otherwise} \end{cases}$  (2) [10]

Where  $\bar{I}$  is the average,  $\sigma$  is the standard deviation of pixels' intensity of the image I', and Z is a constant equal to 0.9

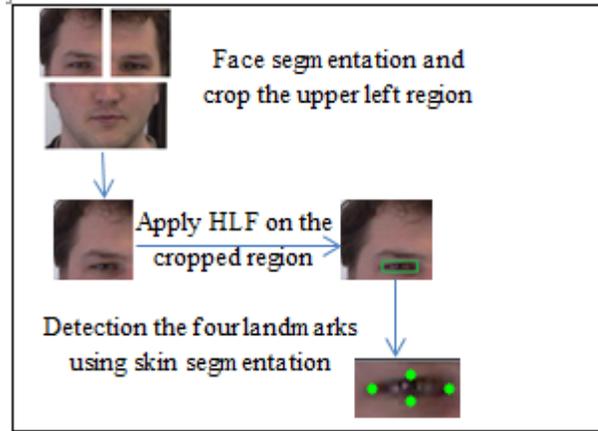


Fig. 3. the steps for extract eyes landmark points

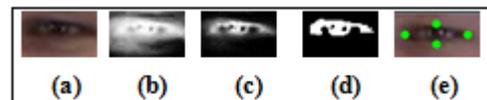


Fig. 4. The steps of detecting eyes landmark points

Determine the eyebrows ROI that above the eyes ROI. After that determine the mask on eye ROI and apply the same algorithm that used for eye. These steps of eyebrows are shown in fig. 5.

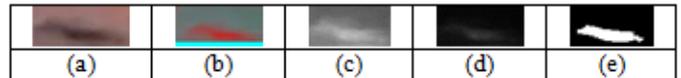


Fig. 5. The steps of detecting eyebrows landmark points

This method detects four points on each eye and three points on each eyebrow as shown in Fig. 2. The algorithm of eyes and eyebrows points' detection is considered the most effective technique for locating these dynamic points [10].

2) *Nose Points' Extraction:* Locating the nose dynamic points is the next step after detecting the eyes points. The nose as a ROI is specified as the vertical part which it's top is between the eyes that have been detected by enhanced VJ [3]. The nose bounds by three lines: two vertical lines and horizontal line which passes on the nose holes are determined by the highest gradient H of the Sobel projection curve that proposed in [3]. Finally, two points of the intersection of the three lines (vertical lines and horizontal line) are the landmark points of nose. This method detects two points at any frontal face position as shown in Fig. 6.

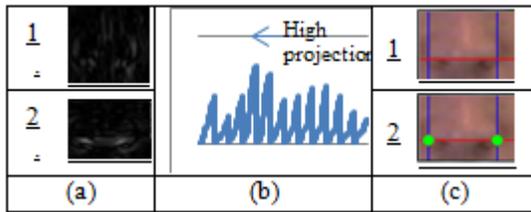


Fig. 6. Description of the steps of detecting nose feature points:

- (a) 1.Gradient Image  $\nabla I_x$  2. Gradient Image  $\nabla I_y$
- (b) Curve of gradient projection
- (c) Two points of the intersection of three lines (vertical lines, horizontal line)

3) *Mouth Points' Extraction:* Detecting the mouth dynamic points depends mainly on the static points of face border and the pre-located nose points. Firstly, the mouth box as a ROI is determined as a ratio of the side face box 'R'. The nose is located at the top of the mouth region, it is at  $0.67 * R$  from the face top border and has a width equal to  $0.25 * R$ . This part is the bounding box of the face with a width  $0.1 * R$  [3]. Once the mouth ROI is determined, a lip map is employed to detect the mouth landmarks points [11] as following:

$$\text{Lipmab} = \left\{ \left( \frac{r}{r+g} \right) * \left( 1 - \frac{g}{r+g} \right) \right\}^2 \quad (3) \quad [11]$$

Where r, g and b are the RGB component after normalization

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} \quad (4) \quad [11]$$

The four mouth landmarks are determined as shown in Fig. 7.



Fig. 7. Example of mouth landmark points detection steps

Finally, after detecting static and dynamic points on the first video frame, pyramidal implementation of the Kanade Lucas Tracker (KLT) algorithm is used for locating these points through the upcoming frames [1]. The example of tracking the distances is shown in fig. 8.

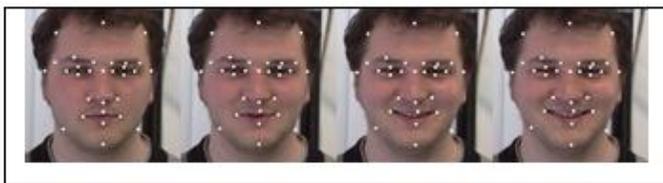


Fig. 8. Samples of point tracking of image sequence of happy expression

### III. REAL-TIME HYBRID FEATURE TRACKING ALGORITHMS FROM VIDEO SEQUENCES

Regarding the slightly changes in the dynamic points movements for different expressions, in addition to the confusion that may occur between two or more facial expression, because of common movements. It is required to

employ different feature tracking algorithms corresponding to the nature of each dynamic point. In this paper, the proposed approach implements a hybrid feature tracking algorithms based on distances feature vector and triangles feature [1,3].

#### A. FER System Based on Distances Feature Vector

This technique relies on defining 43 universal distances on the face, named distance vectors. These vectors are derived from 2D distribution between two types of facial landmark points and are known by being effective to track the facial expressions through video sequences [3]. The distance vectors are shown in Fig. 9. For data normalization, each distance is divided by those in the first frame which represent the nature expression as in [3]. This technique achieves low processing time because of feature vector dimensionality.

However, it fails to achieve high recognition accuracy for some expressions.

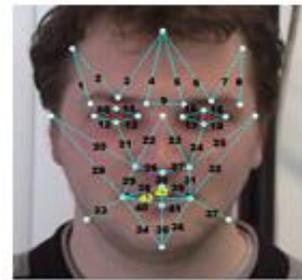


Fig. 9. The FER system distances

#### B. FER System Based on Triangles Feature Vector

This technique tracks the movements of feature points through a triangle shape, three facial points are tracked at a time more rather than tacking one or two facial landmarks [1]. The tracking information of the facial points and the relationship between them can be captured well by tracking three points at a time.

Triangle components in each frame are subtracted from the triangle components in the first frame of the video. Each triangle has four components that are saved at feature vector, named 'a', 'b', 'α' and 'β', the changes in their values corresponding to the first frame are considered the feature vector that will be used in FER. Fig. 10 shows the mathematical representation of the difference between triangle components between two frames

FER based on this technique proposed in [1] employs 52 landmark points. The feature vector is too long. This leads to an increase in the processing and classification time. In this paper, specific 30 landmark points are supposed to be tracked by triangle vectors. These points were chosen to easily identify expressions [12]. There would be a total of  $30! / (3!(30-3)!) = 4060$  unique triangles. If the feature vector is formed by all these triangles it will be very long. So the Adaboost algorithm was used to choose the best triangles that represent expressions. Fig.11 shows the 4060 triangles and the Adaboost 80 triangles.

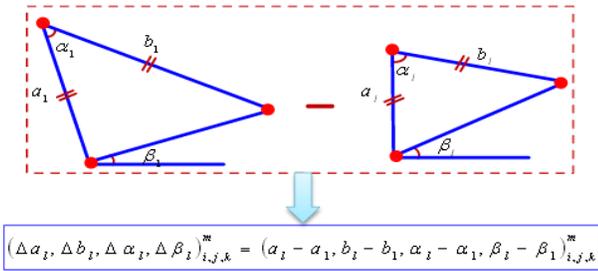


Fig. 10. Difference in components of two triangles used as features [1]

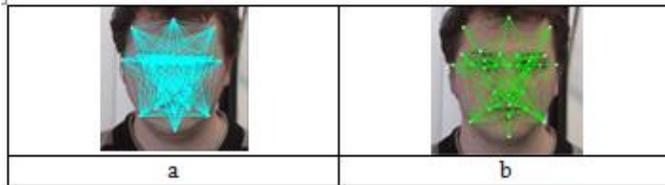


Fig. 11. a) all unique triangles b) the Adaboost triangles

Although the adaboost choose the best triangles, the feature vector still long and the processing time would be high. So, the proposed approach implements a hybrid system that used triangles in recognizing the expressions which have low accuracy in distances based system and used the distances for other expressions.

### C. Hybrid System Based on Triangles Feature Vector

The proposed hybrid approach depends on employing distance vectors in recognizing the expressions that are defined clearly. Moreover, it uses the triangles for recognizing the expressions that are not recognized accurately in the distances system. The hybrid technique is used for tracking certain dynamic points for confusing expressions. For example, fear and sadness expressions have a low recognition accuracy rate when employing distances based method. Thus, the triangles based method is implemented to enhance their recognition accuracy. However, the fear expression usually confuses with surprise expression, also sadness. Therefore, fear and surprise expressions are trained in ADaboost technique to select triangles features of this expressions, this Adaboost selects 35 triangles. Also, sadness, anger and disgust expressions trained in ADaboost technique to select triangles features for these expressions, this Adaboost selects 41 triangles. Fig. 12 shows the superiority of triangle technique to detect the variations of eye height at fear expression. Fig. 13 shows the varying angles of eye triangle at fear expression. The first angle changes from 108 ° to 120 ° and the second angle change from 38° to 50° beside the line varying of the triangle that represent eye height. These two figures show that tracking three points as and their relationship between them is more efficient than tracking distances between two points.



Fig. 12. Distance varying of eye in the fear expression

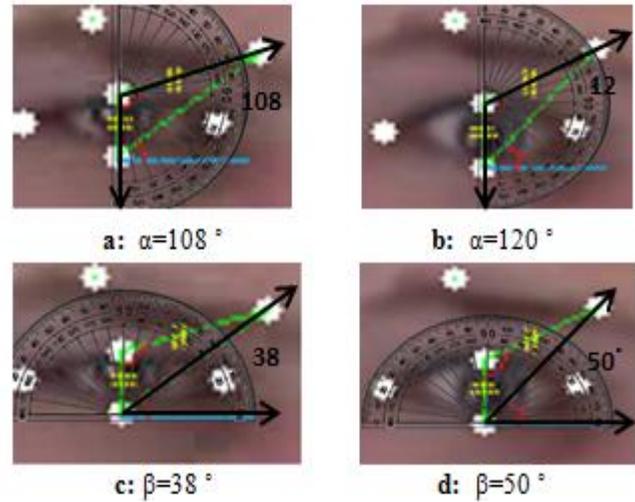


Fig. 13. The triangle angles (  $\alpha$  ,  $\beta$  ) varying of eye in the fear expression

## IV. INTELLIGENT RECOGNITION TECHNIQUE BASED ON HIERARCHICAL SVM CLASSIFICATION

SVM is used for classification and regression analysis. SVMs exhibit high classification accuracy for small training sets and good generalization performance on data that are very variable and difficult to separate [3]. The proposed approach generates hierarchal SVM using RBF kernel [13]. The HSVM framework of the three systems is similar to that shown in Fig. 7. **Notice that** the input feature vectors differ.

The proposed HSVM recognition strategy is processed as an uninformed artificial intelligent search problem. The search tree consists of six SVM that are trained to differentiate accurately among the most common six expressions. The search process based on depth-first algorithm [13] , it starts from the root node and goes through the parent nodes, it ends immediately as soon as the goal achieved, no need to go through all the leaves. Therefore, it reduces the processing time.

As shown in Fig. 14, each SVM examines different input features corresponding to the expressions that are going to be tested. In Fig. 7 the SVM input features are labeled from “A” to “F”, each label indicates specific input features as explained in Table I. Since, for each expression there are governing features that help accurately and quickly to recognize this expression. For example, the root node SVM1 is feed by “A” input features, which are considered a common feature for several expressions. If the SVM1 output result is close to happy or disgust features it leads to SVM2. Otherwise it leads to SVM3. Now, SVM2 node tests the input features “B”, which are considered more specific features and are used to differentiate between two conflicting expressions. Upon SVM2 test result, it decides if the expression is happy or disgust. The process continues, each SVM node examines more details about a pair of confusing expressions. The search process ends as soon as the goal is achieved.

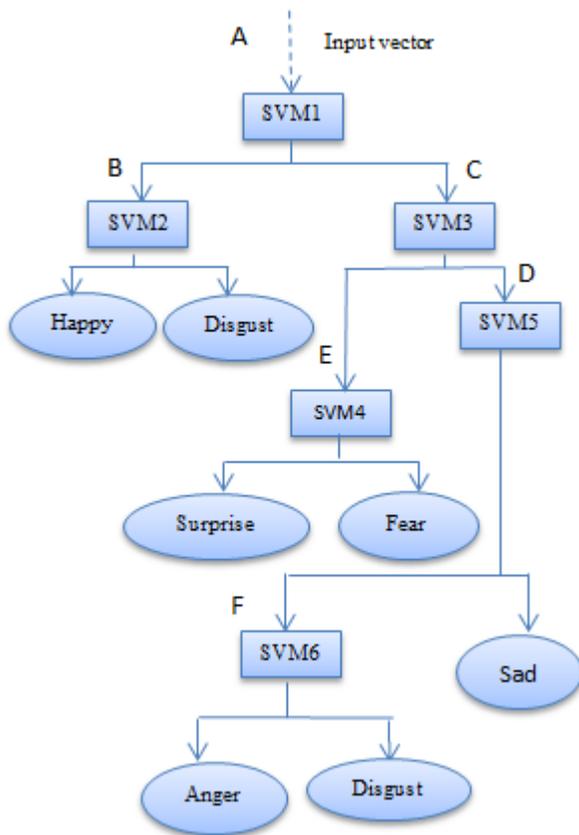


Fig. 14. The HSVM Framework

Table I shows the input vectors of HSVM in the three systems that based on distances, triangles and the proposed system that based on both. The distances from D1 to D43 are shown in fig. 9. The triangles of a face component (eyes, eyebrows, nose, mouth) are triangles that have a landmark point on this component i.e. the triangles of mouth are all triangles that have a landmark point or more on the mouth. At the triangles based system, two triangles were first inputs to the SVM1 are shown in fig.15 .the all face triangles are triangles that adaboost was chosen. The triangles that input to SVM4 and SVM5 in the hybrid system were chosen by adaboost algorithm i.e. the 35 triangles that input to SVM4 were chosen by adaboost algorithm when training it by surprise and fear expressions. The distances in the hybrid system are shown in fig.9.



Fig. 15. The two triangles that first input in the triangles system

TABLE I. THE CLASSIFIERS OF HSVM AND ITS INPUT DISTANCES FOR CLASSIFICATION PROCESS

| classifier | Distance based vector |                        | Triangle based vector                  |                        | Hybrid based vector          |               | Classification node of classifier |
|------------|-----------------------|------------------------|--|------------------------|------------------------------|---------------|-----------------------------------|
|            | Input distances       | No. of input distances | Input triangles                        | No. of input triangles | Input distances or triangles | No. of inputs |                                   |
| SVM1       | A=D38:D43             | 6                      | A=Two triangles of mouth               | 2                      | A=D38:D43                    | 6             | SVM2,SVM3                         |
| SVM2       | B=D28:D43             | 16                     | B=All mouth triangles                  | 26                     | B=D28:D43                    | 16            | Happy, Disgust                    |
| SVM3       | C=D1:D9               | 9                      | C=Eye brows triangles                  | 20                     | C=D1:D9                      | 9             | SVM4, SVM5                        |
| SVM4       | D=D9:D43              | 35                     | D=All face triangles                   | 80                     | D=35 triangle                | 35            | Surprise, Fear                    |
| SVM5       | E=D20:D43             | 24                     | E=Eye brows triangles & nose triangles | 38                     | E=41 triangle                | 41            | sad, SVM6                         |
| SVM6       | F=D1:D43              | 43                     | F=All face triangles                   | 80                     | F=D1:D43                     | 43            | disgust, Anger                    |

V. EXPERIMENTS AND EVALUATION

A series of experiments and tests were conducted using a C++ computer to simulate the previous techniques. Testing was done based on 'FEEDTUM' common standard video sequences dataset for six different facial expressions. It contains 385 sequences. Each sequence begins from neutral and ending to its emotion. The proposed approach used 20 video in each expression for training and 344 sequences for testing. In the training the sequences were attached by expression labels. In the testing mode the sequence input to the system without labeling. Table II summarizes the proposed technique, the dynamic points that are tracked as a distance vector, triangle or hybrid ones are shown, and also the input features corresponding to each SVM are shown in Table III.

Table IVV shows the FER accuracy for the six facial expressions based on distance vector only, triangle vectors only and the proposed hybrid technique, the average of these accuracy and the average of processing time. The most important factor of the processing time is the feature vector length. There are the positive relation between the feature vector length and the processing time in the recognition.

Regarding the comparable results introduced in Table VIVII, It can be concluded that the hybrid technique has achieved the tradeoff between high recognition accuracy and low processing time which is the main challenge of real-time video sequences FER.

There are an example that shows that the hybrid system is better than distances system and triangles system. For recognizing the fear expression the feature vector path from SVM1 then SVM3 finally SVM4 the total feature vectors length are calculated for distances system, triangles system and hybrid system as following:

The fear path in the HSVM is:

$$SVM1 \rightarrow SVM3 \rightarrow SVM4$$

Calculating the total feature vector length of this path by the numbers of inputs that are shown in table II:

Distance system:

$$6 \text{ distances} + 9 \text{ distances} + 35 \text{ distances} = 6 + 9 + 35 = 50$$

Triangle system:

$$2 \text{ triangles} + 20 \text{ triangles} + 80 \text{ triangles} = 2 * 4 + 20 * 4 + 80 * 4 = 408$$

Hybrid system:

$$6 \text{ distances} + 9 \text{ distances} + 35 \text{ triangles} = 6 + 9 + 35 * 4 = 155$$

**Note that** the number of triangles are multiplied by 4 because each triangle is represented by 4 components as shown in fig.10.

Fig.16 shows the relationship between the feature vector length and the accuracy of fear expression in the three systems.

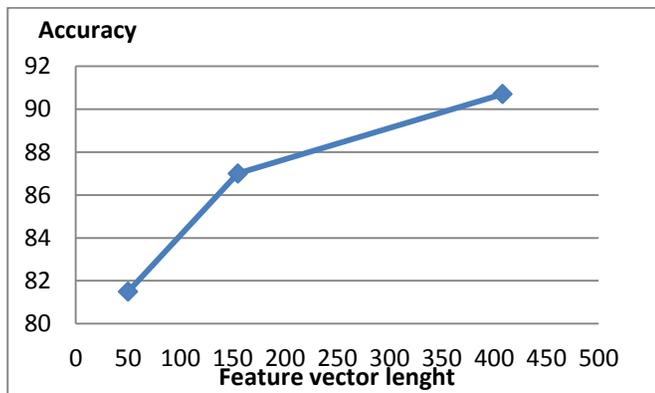


Fig. 16. The relationship between the feature vector lengths of fear expression and its accuracy in the three systems

TABLE II. RECOGNITION ACCURACY OF ALL EXPRESSIONS IN THE THREE TECHNIQUES

| Facial expression              | Feature Tracking algorithm |                       |                     |
|--------------------------------|----------------------------|-----------------------|---------------------|
|                                | Distance based vector      | Triangle based vector | Hybrid based vector |
| Anger                          | 96.3%                      | 100%                  | 100%                |
| Disgust                        | 94.4%                      | 98.1%                 | 94.4%               |
| Happy                          | 100%                       | 100%                  | 100%                |
| Fear                           | 81.5%                      | 90.7%                 | 87%                 |
| Sad                            | 90.7%                      | 94.4%                 | 94.4%               |
| Surprise                       | 100%                       | 100%                  | 100%                |
| <b>Overall Accuracy</b>        | <b>93.8%</b>               | <b>97.2%</b>          | <b>96%</b>          |
| <b>Average Processing time</b> | <b>22 msec</b>             | <b>256 msec</b>       | <b>93 msec</b>      |

Previous published results using FEEDTUM dataset [8,14] had achieved 93.4% and 94.6% accuracy rates, respectively. Moreover, the feature vector dimension in [1] is larger than the proposed feature vector. The proposed approach achieved an average recognition rate of 93.8%, 97.2% and 96% with 22 msec, 256 msec and 93 msec of distance based vector technique, triangle based vector technique and hybrid based vector technique respectively. The recognition accuracy of each expression is represented graphically in Fig. 17. Moreover, the proposed intelligent hierarchal SVM with hybrid based vector succeed in decreasing the classification correlation between the conflicting expressions as shown in Table VIIIXX. Finally, Table XIV introduces the comparison between the related work and the proposed system which shows the superiority of the proposed algorithm in terms of FER accuracy.

TABLE III. THE ACCURACIES RATE OF EXPRESSIONS OF THE HYBRID FOR TECHNIQUE

|          | Anger | Disgust | Happy | Fear | Sad   | Surprise |
|----------|-------|---------|-------|------|-------|----------|
| Anger    | 100%  | 0       | 0     | 0    | 0     | 0        |
| Disgust  | 0     | 94.4%   | 1.9%  | 0    | 0     | 3.7%     |
| Happy    | 0     | 0       | 100%  | 0    | 0     | 0        |
| Fear     | 0     | 0       | 0     | 87%  | 3.7%  | 9.3%     |
| Sad      | 1.9%  | 0       | 0     | 3.7% | 94.4% | 0        |
| Surprise | 0     | 0       | 0     | 0    | 0     | 100%     |

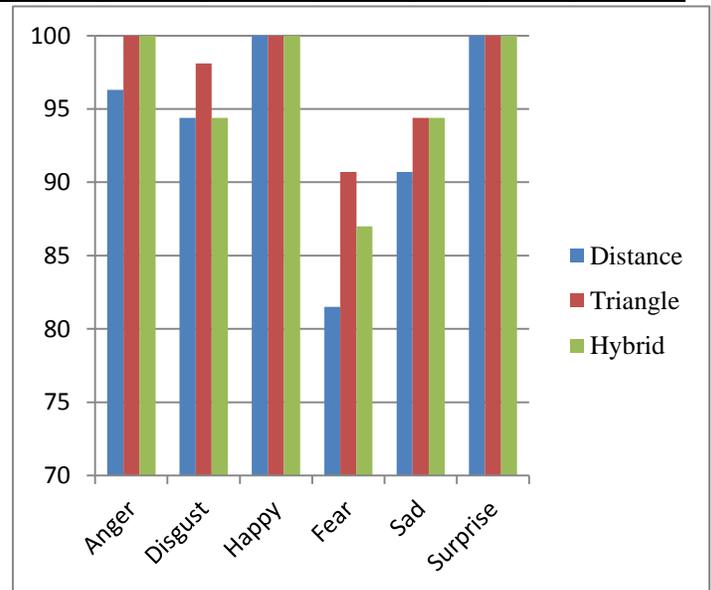


Fig. 17. The recognition accuracy of all expressions in the three techniques

TABLE IV. THE COMPARISON OF THE RELATED WORK AND THE PROPOSED SYSTEMS

| Reference Number       | NO. of expressions | Feature extraction technique              | Classification method | Database     | Accuracy rate                  |
|------------------------|--------------------|---|-----------------------|--------------|--------------------------------|
| [3]                    | Six                | Geometric facial feature                  | SVM                   | FEEDTUM      | 90%                            |
| [7]                    | Six                | SOM                                       | Neural                | FEEDTUM      | 88.8%                          |
|                        |                    |   |                       | Cohn -Kanade | 91%                            |
| [6]                    | Seven              | Geometric and appearance features         | Bézier curves         | FEEDTUM      | 78.3%                          |
| [8]                    | Seven              | Traditional BP<br>LSGA-BP                 | Cubic Bézier curves.  | FEEDTUM      | 89%                            |
|                        |                    |   |                       |              | 93.4%                          |
| [4]                    | Six                | Anthropometric model                      | HSVM                  | FEEDTUM      | 90.5%                          |
| [1]                    | Six                | Geometric Triangles based                 | SVM                   | Cohn -Kanade | 97%                            |
|                        |                    |   |                       | MUG          | 95%                            |
| [2]                    | six                | Geometric points, distances and triangles | SVM                   | Cohn -Kanade | 96.37%,<br>96.58% and<br>97.8% |
|                        |                    |   |                       | MMI          | 67.64%,<br>74.31 and<br>77.22% |
|                        |                    |   |                       | MUG          | 91.41%,<br>94.13% and<br>95.5% |
| [15]                   | seven              | SPTS                                      | SVM                   | Cohn-Kanade  | 66.5%                          |
|                        |                    | Geometry Features                         |                       |              | 84.7%                          |
|                        |                    | Combination                               |                       |              | 88.7%                          |
| [16]                   | Six                | Diffeomorphic Growth Model                | SVM                   | MMI          | 96.5%                          |
| The proposed Technique | Six                | Geometric distances based                 | HSVM                  | FEEDTUM      | 93.8%                          |
|                        |                    | Geometric Triangles based                 |                       |              | 97.2%                          |
|                        |                    | Hybrid                                    |                       |              | 96%                            |

## VI. CONCLUSIONS AND FUTURE WORK

This paper proposed a real time FER from a video sequences based on hybrid feature tracking algorithm and an intelligent HSVM search method. The proposed technique integrated the best well known algorithms in each step. The main goal is to achieve the tradeoff between real time video demands and high recognition accuracy. Hybrid system is implemented to reduce the feature vector dimensionality and intelligent HSVM achieve high accuracy with low processing time. The proposed approach was tested using the standard FEEDTUM database. FEEDTUM database is difficult to be treated because intra-class confusions. This confusion between some expressions makes it very difficult in recognizing even by human. Despite this limitation, the proposed technique showed good results regarding accuracy and processing time. For future work, the proposed approach may be implemented on real life image sequences instead of the FEEDTUM database. For feature selection some other methods will be considered for more advantages. Moreover, using variance angles of human face may be used for training and testing the new approach.

### REFERENCES

- [1] D.Ghimire, J.Lee, Z.Li, S.Jeong, S.Park and H.Choi, "Recognition of Facial Expressions Based on Tracking and Selection of Discriminative Geometric Features", International Journal of Multimedia and Ubiquitous Engineering, Vol.10, NO.3, pp.35-44, 2015.
- [2] D.Ghimire, J.Lee, Z.Li, S.Jeong and S.Park, "Recognition of Facial Expressions Based on Salient Geometric Features and Support Vector Machines", Multimedia Tools and Applications, Vol.1, pp.1-26,2016.
- [3] A. Pruski, C. Maaoui and F. Abdat, "Human-Computer Interaction Using Emotion Recognition from Facial Expression", UKSim 5th European Symposium on Computer Modeling and Simulation, pp. 196-201, 2011.
- [4] N.Sadek, N.Hikal and F.Zaki, "Real time facial expression recognition based on hierarchical SVM", IJICIS, Vol.15, No.3,pp.51-60, 2015
- [5] S. S. Bavkar , J. S. Rangole and V. U. Deshmukh "Geometric Approach for Human Emotion Recognition using Facial Expression", International Journal of Computer Applications, Vol.118, No 14, pp.17-22, 2015
- [6] R.KANDEMIR and G.ÖZMEN, "facial expression classification with haar features, geometric features and cubic bézier curves ", IU-JEEE, Vol.13, No.2,pp.1667-1673, 2013
- [7] M.Su,C.Yang, S.Lin, D.Huang, Y.Hsieh, and P.Wang, "An SOM-based Automatic Facial Expression Recognition System", IJSCAI, Vol.2, No.4, pp.45-57, 2013.
- [8] M.Alsmedi, "Facial Recognition under Expression Variations", The International Arab Journal of Information Technology, Vol.13, No.1,pp.133-141, 2016
- [9] I.Khan, H.Abdullah and M.Zainal, "Efficient eyes and mouth detection algorithm using combination of viola jones and skin color pixel detection", International Journal of Engineering and Applied Sciences, Vol.3, No.4,pp.52-60, 2013.
- [10] J.Moreira, A.Braun, S.Musse , "Eyes and Eyebrows Detection for Performance Driven Animation", SIBGRAPI '10 Proceedings of the 2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images, pp.17-24, 2010.
- [11] A.Atharifard, and S.Ghofrani,"Robust Component-based Face Detection Using Color Feature", World Congress on Engineering, Vol.2, No.8,pp.978-988, 2011.
- [12] M.Valstar , B.Martinez, X.Binefa "Facial Point Detection using Boosted Regression and Graph Models", Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp.2729 – 2736, 2010
- [13] Ch.Lee , Emily Mower , C.Busso , S.Lee and Sh.Narayanan, "Emotion recognition using a hierarchical binary decision tree approach", Speech Communication journal , Vol.53,issue 9-10, pp. 1162–1171, 2011.
- [14] K.Sikka , A.Dhal and M.Bartlett "Exemplar Hidden Markov Models for Classification of Facial Expressions in Videos",CVPRW,pp.15-25,2015

- [15] C.Gacav and B.Benligiray," Greedy Search for Descriptive Spatial Face Features", Proceedings of the 42nd IEEE ICASSP, pp.7-10, 2017.
- [16] Y.Guoand G.Zhao," Dynamic Facial Expression Recognition with Atlas Construction and Sparse Representation", IEEE Trans Image Process, vol.25, No.5, pp1-16,2016