Facial emotion recognition with the hidden Markov model

Mouheb Lahbiri, Ahmed Fnaiech, Moez Bouchouicha*, Mounir Sayadi and Philippe Gorce*

SICISI unit, ESSTT, University of Tunis, 5 Av. Taha Hussein, 1008, Tunis, Tunisia.
*Labo Handibio, University of Toulon-Var, France
E-mails: mouheb.lahbiri@hotmail.fr, ahmedfnaiech@hotmail.com, moez@univ-tln.fr, mounir.sayadi@esstt.rnu.tn, gorce@univ-tln.fr

Abstract—this paper presents a simple algorithm for an automatic recognition of facial expressions. First we extract feature points, and then we define distances using these features. The variation of these distances is used to characterize the transition from one emotion to another. Our approach is based on the use of a hidden Markov model whose states can recognize facial expressions.

Key words: facial recognition, feature points, HMM, segmentation.

I. INTRODUCTION

In the daily life, each person meets many events that can act directly on his feelings and on his psychological state or more precisely on his internal state. The expressions of a person's internal state form some emotions. These emotions appear on the face under shape of facial expressions where the main organs of the face play a role to construct these expressions. The eyes, the mouth, the lids and the eyebrows are the elements that permit to draw expressive faces.

We can distinguish six primary emotion types according to studies generally admitted by the psychologist Paul Ekman [1]; surprise, happiness, anger, disgust, fear and sadness without forgetting the neutral state. Other types of secondary emotions exist as the jealousy, the pride, the nostalgia, the shame... But, we insist in this work on the six primary emotions because they are the simplest to be identified.

Emotion recognition is easy for every human; that means one can know a person's psychological state after a mental analysis of the expressions that appears on his face. Nowadays, with regard to the scientific and technological development, it could be very interesting to realize systems that can automatically recognize facial expressions and deduce person's emotions.

We can use such systems of recognition in many applications; Humanoid robots, video games, automotive industry, MMI, communication with handicap….

We can mention an example where we can apply this system (automotive industry): A system based on a camera which is placed directly ahead of the driver to analyze his facial expressions. If the system detects a negative psychological state (Anger, sadness, fear), a decision is made: alert, stopping the engine car, etc., to avoid accident.

In recent years, several researchers have studied facial emotions recognition. Several methods have been implemented. In this paper, we use hidden Markov Models to describe facial emotions.

To detect emotions using the hidden Markov model, several approaches have been elaborated to extract facial features. The global approaches as the use of the methods of space reduction (PCA, LDA, ICA…) [2], [3], [4], while creating sequences of observation by quantifications of vectors [5]. And, features extraction that can achieve with the help of local approaches as the active model of appearance (AAM) [6], [7] or filters of Gabor [8].

In the continuation, we will enumerate the different stages of realization of an algorithm to recognize facial emotions based on a hidden Markov Model.

We achieve the local facial feature extraction (distances between the characteristic points of face). These distances vary from an emotion to another.

Dealing with these variations, we create sequences of observation that help us to estimate a hidden Markov model for each emotion. This approach allows us to recognize human’s emotions.

II. FACIAL FEATURE EXTRACTION AND PRESENTATION OF THE HIDDEN MARKOV MODEL (HMM).

For each system of recognition, the input signal is a face image that cannot to be used directly for the classification. It is due to its big dimension and also, especially for our theme, to the absence of the clear discriminative elements leading to a fast and accurate classification.

To build our recognition system, we follow the steps in the picture below (Fig 1) where N represents the number of the HMMs models to estimate, and L represents the size of the feature vector to extract.

Fig 1: Description of the facial recognition system with the hidden Markov model
II.1. Feature Extraction

Feature extraction is the first step to guarantee the efficiency of our recognition system. It is a transformation or a creation of a basis of observations appearing in the image.

The emotions must appear on the face in the form of facial expressions, where the principal bodies of the face play a part to build these expressions. Therefore, the eyes, the mouth, the eyelids, the eyebrows are elements which make it possible to characterize the expressions of the face. [9]

![Fig 2: The particular points of face](image)

During the transition from the neutral state to another state (emotion), movements are carried out naturally on the muscles of the face, therefore the distance between the end of a principal body of the face and another end of the same body or of another, will undergo a change. This change must be an increase or a reduction.

The facial movements are involuntary movements. They involve a variation in the sites of the facial points of end of the bodies this inevitably generates a variation of the distances between these points.

With the naked eye, we can estimate the variations of the distances and directly deduce the emotion produced on the face from each person. But our concept seeks to create or model a computing system which must recognize the emotion.

As we stated at the beginning, the modeling of the recognition system consists in extracting a set of feature that can appear on the face with each emotion. These features will be used in the classification to recognize these emotions. Therefore, to guarantee a very high rate of recognition, it is necessary to well choose these features and the classifier.

From figure 2, we can count 16 points of main organ extremities affecting the production of emotion on the face. These organs are the mouth, eyes and eyebrows. These points are: the interior corners (F and I) and outsides (E and G) of the two eyes, the ends of the higher eyelids (G and K) and of the lower eyelids (H and L), the interior corners (B and C) and outsides (A and D) of the eyebrows, the left (M) and right (N) corners of the mouth and the upper (OH) and lower (P) ends of the lips.

![Fig 3: Definition of the D_i distances](image)

Between these points, we can count 120 distances (AB, AC, OM, FG…) that we can adopt as feature when they change an emotion to another.

We can only choose a set of five distances (D1=KL=GH, D2=CI=BF, D3=MN, D4=OP and D5 = AE=JN) that are submitted to considerable variations at the time of each emotion to facilitate the modeling of the recognition system. These distances are represented below in the picture.

These distances are feature distances that we will use in our process of classification of the emotion. Therefore, each emotion is characterized by a vector $D_i$ that includes the five distances [10].


The hidden Markov model (HMM) is a statistical method, among the most popular, to deal with the problems of recognition or classification.

It is often used in the areas of natural language processing, voice recognition, pattern recognition and analysis of biological sequences.

It represents a good method to study the evolution of the sequences of observations. This survey is based on the calculation of the probabilities of progress of these sequences to take a decision on the nature or the class.

Each HMM $\lambda$ is defined by the five following parameters [12]: $\lambda = (S, O, A, B, \pi)$

- A model $S$ formed of N states
  $S = \{S_1, S_2, \ldots, S_N\}$

- The distinct observations according to M symbols
  $O = \{O_1, O_2, \ldots, O_M\}$

- The probabilities of transitions between states:
  $A_{N \times N} = \{a_{ij}\}$

  Where $a_{ij} = P(s_{t+1} = S_j|s_t = S_i)$ ; $1 \leq i, j \leq N$

- Probabilities of observations (Probabilities of emission)
  $B_{M \times N} = \{b_j(O_i)\}$

  Where $b_j(O_i) = P(O_t = o_i|S_t = S_j)$

  $1 \leq j \leq N$

- The probabilities of initial state
  $\pi = \{\pi_i\}$

  Where $\pi_i = P(S_1 = S_i)$

  $1 \leq i \leq N$
• **Evaluation problem**

Given the HMM \( \lambda = (A, B, \pi) \) and the observation sequence \( O = \{O_1, O_2, \ldots, O_M\} \), calculate the probability that model \( \lambda \) has generated sequence \( O \).

• **Decoding problem**

Given the HMM \( \lambda = (A, B, \pi) \) and the observation sequence \( O = \{O_1, O_2, \ldots, O_M\} \); calculate the most likely sequence of hidden states \( S_i \) that produced this observation sequence \( O \).

• **Learning problem.**

Given some training observation sequences \( O = \{O_1, O_2, \ldots, O_M\} \) and general structure of HMM (numbers of hidden and visible states), determine HMM parameters \( \lambda = (A, B, \pi) \) that best fit training data.

III. CREATION OF THE OBSERVATION SEQUENCES AND HIDDEN MARKOV MODEL TRAINING

In this part we explain in detail the approach of emotions recognition, by classifying the five feature distances previously indicated. Classification is done with the model of hidden Markov. This approach consists in defining a HMM for each emotion by using the five feature distances. Each model obtained at the end of the training represents a recognition of the emotion. Therefore, to apply the classification of the facial emotions with the model of hidden Markov, it is necessary to find an average which enables us to extract a sequence from observation, starting from these five distances for each emotion.

III.1. Creation of the observation sequences

The modeling of HMM models to be estimated consists in creating a sequence or several sequences of observations, during the analysis of the variations of the five distances compared to the same distances from the neutral state. When the sequences are created using the feature distances, we follow the stages below for defining well the specific sequences of each emotion.

- Detection and normalization of the face
- Detection of the feature points
- Calculation of the five distances
- Creation of the observation sequences

To create the sequences of observation, it is necessary to calculate the five distances on such a number of image of faces for each emotion. Measurements obtained help us to know nature of variation which each distance compared to the same distance from the neutral state. The calculation of the distances requires the detection of the nine points between which the five distances are. After detection, we use an algorithm to calculate these distances.

To know how vary the five distances of those that are at the neutral state, for each emotion for create the sequences of observation, we extract the coordinates of pixel by hand \((x,y)\) of a point for a set of 151 pictures of the only one person face; 25 for each emotion and a picture of the face to the neutral state. Then, we calculate the five distances with the equation of the Euclidian distance:

\[
D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}
\]

The observations are the changes that act on the state of the system all along our interval of time or along our spatial settings.

In this work, the observations are the variations that act on each distance among the five distances compared to the neutral state, during the production of each emotion.

Fig 4 : An excerpt of basis of the data of pictures of a face with the six emotions [13]

That means the creation of each sequence for each emotion is carried out, after observing the changes in the variation made on each distance. The analysis of the values of these distances on the examples (sequences of expressions) enabled us to note that, compared to the neutral state of the face, these distances can have more or less important values according to the expression. We define [14]:

- A **C**+ state if the value of \( D_i \) is bigger than its value in the neutral state.
- A **C**- state if the value of \( D_i \) is lower than its value in the neutral state.
- A state **S** for \( D_i \) if its value is near to its value in the neutral state.

we can analyze the variation of each distance during the production of the emotion; If it decreases, increases or remains the same one compared to the neutral state. For this reason we observe that, for some distances, the increase or the decrease is considerable at the time when, for others, the variation (increase or decrease) is very low.

For the state **S** where the variation \( V \) will be very low, we need to choose a threshold \( \pm h \) which separates this state from the states **C**+ and **C**-. That means, we define the field of variation \([h, +h]\). If it is exceeded, we will have the **C**+ state if \( V < -h \) is or the **C**- state if \( V > h \). If we choose to work with a threshold \( h=10\% \), then the creation of the observation sequences will be developed using the following equations:
We set the variable $\xi_t(i,j)$, the probability of being in the state $S_i$ at the time $t$ and in the state $S_j$ at the time $t+1$ given the observation sequence $O$ and the model $\lambda$, and also we define the variable $\gamma_t(i)$, the probability that we pass in $S_i$ at time $t$ while generating $O$ with $\lambda$.

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})}{Pr(O|\lambda)}$$  \hspace{1cm} (8)

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$  \hspace{1cm} (9)

Where $\alpha_t(i)$ is the Forward variable and $\beta_t(i)$ is the backward variable:

$$\alpha_1 = \pi_i b_i(O_1)$$  \hspace{1cm} (10)

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i)a_{ij}b_j(O_{t+1})$$  \hspace{1cm} (11)

Using these sequences of observation indicated in the table 1, for each emotion a hidden Markov model $\lambda_{emotion}$. That means, we apply an algorithm of training that adjusts the parameters of this model to maximize the probability $P(O|\lambda)$ from the sequence of observation corresponding.

Using the sequences of observation indicated in the table 1, for each emotion a hidden Markov model $\lambda_{emotion}$. That means, we apply an algorithm of training that adjusts the parameters of this model to maximize the probability $P(O|\lambda)$ from the sequence of observation corresponding.

III.2. Hidden Markov model training

The first stage of the training consists in choosing a structure for the hidden Markov model. The choice of the number of states and the choice of the emission function represent the main elements that serve to structure our model. Since the states are hidden and for model a HMM to $N$ states for each emotion $\lambda = \{A, B, \pi\}$, we propose to use a left-to-right model with 4 states

The algorithm of Baum-Welch will be used for the training phase. It is an algorithm of Expectation-Maximization type (EM). It starts from a set of initial values and improves this set of values to each iteration. The initial parameters are chosen arbitrarily.

Fig 5: HMM Model "Left-to-right" with 4 states

The algorithm of Baum-Welch will be used for the training phase. It is an algorithm of Expectation-Maximization type (EM). It starts from a set of initial values and improves this set of values to each iteration. The initial parameters are chosen arbitrarily.
possibilities of sequences with these two parameters when most these sequences are not accessible. If a HMM model obtained after the training will give to the sequence that suits him a probability equal to 1, the probabilities of the progress of the other 242 sequences will immediately be equal to 0. For these reasons, our system will issue null results that lead to non-recognition of emotion. That’s why, we propose as a solution an increase in the size of each sequence, knowing that if the size of a sequence is larger, the probability of flow will be lower. The creation of the sequences of T>5 size gets used by the addition of the same sequence to the initial sequence a certain number of time. The size of the new sequence is a multiple of five.

\[ O_{\text{emotion}} = \{O_1, O_2, O_3, O_4, O_5\} \]

\[ O_{\text{emotion}} = \{O_1, O_2, O_3, O_4, O_5, \ldots, O_1, O_2, O_3, O_4, O_5\} \]

IV. FACIAL EMOTION RECOGNITION WITH THE HIDDEN MARKOV MODEL

For the test phase, we use a set of 25 facial images for each emotion. For each one, we follow all the stages; detection and normalization of facial area, detection of the feature points, measure of the five distances, creating sequences and calculation of the probabilities \( P(O_{\text{test}}|\lambda_{\text{emotion}}) \) with the Forward algorithm.

![Fig 6: Evolution of P during the training for each model](image)

The gotten results are united below in the tables:

**TABLE II: MATRIX OF CONFUSION OF PERCENTAGE OF RECOGNITION OF THE FACIAL EMOTIONS (T=10) (%)**

<table>
<thead>
<tr>
<th></th>
<th>Hap</th>
<th>Sur</th>
<th>Dis</th>
<th>Ang</th>
<th>sad</th>
<th>Fea</th>
<th>NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Sadness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td>RR(^1)</td>
<td>94</td>
<td>RNR(^2)</td>
<td>4.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE III: MATRIX OF CONFUSION OF PERCENTAGE OF RECOGNITION OF THE FACIAL EMOTIONS (T=30) (%)**

<table>
<thead>
<tr>
<th></th>
<th>Hap</th>
<th>Sur</th>
<th>Dis</th>
<th>Ang</th>
<th>sad</th>
<th>Fea</th>
<th>NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Sadness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td>RR(^1)</td>
<td>94,66</td>
<td>RNR(^2)</td>
<td>2.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

V. Conclusion

We presented an emotion recognition system with the hidden Markov model. Our system is characterized by the simplicity of calculations. We can consider that this system is semi-automatic because it uses just the images in the neutral state as references to create the sequences of observations. I.e. A system can know the nature of emotion based only on the characteristic of the facial image in a neutral state without need to store a database containing all the emotions. With the hidden Markov model (HMM), we obtain some results higher than 90%. All that explains the efficiency of this algorithm for the facial emotion recognition. Whereas the simplicity of

\(^1\) Rate of recognition
\(^2\) Rate of non-recognition
calculation represents a very important element to carry out a good recognition.
Increasing the size of observation sequence causes a reduction in the rate of non-recognition. The size of sequence of observation also acts on the variation of the rate of recognition. It is more important if the size of sequence is larger. So, we can improve the obtained result by using further face distances to increase the size of the observation sequences.

REFERENCES