

# Differential Evolution to Optimize Hidden Markov Models Training: Application to Facial Expression Recognition

Khadoudja Ghanem, Amer Draa, Elvis Vyumvuhore and  
Arsène Simbabawe

MISC Laboratory, Constantine 2 University, Constantine, Algeria

The base system in this paper uses Hidden Markov Models (HMMs) to model dynamic relationships among facial features in facial behavior interpretation and understanding field. The input of HMMs is a new set of derived features from geometrical distances obtained from detected and automatically tracked facial points. Numerical data representation which is in the form of multi-time series is transformed to a symbolic representation in order to reduce dimensionality, extract the most pertinent information and give a meaningful representation to humans. The main problem of the use of HMMs is that the training is generally trapped in local minima, so we used the Differential Evolution (DE) algorithm to offer more diversity and so limit as much as possible the occurrence of stagnation. For this reason, this paper proposes to enhance HMM learning abilities by the use of DE as an optimization tool, instead of the classical Baum and Welch algorithm. Obtained results are compared against the traditional learning approach and significant improvements have been obtained.

*Keywords:* facial expressions, occurrence order, Hidden Markov Model, Baum-Welch, optimization, differential evolution

## 1. Introduction

Analyzing the dynamics of facial features and (or) the changes in the appearance of facial features (eyes, eyebrows and mouth) is a very important step in facial expression understanding and interpretation. Many researchers attempt to study the dynamic facial behavior. Timing, duration, speed and occurrence order of face/body actions are crucial parameters related to dynamic behavior (Ekman, & Rosenberg, 2005).

For instance, facial expression temporal dynamics are essential for recognition of either full expressions (Kotsia & Pitas, 2007; Littlewort & al, 2006), or components of expressions such as facial Action Units (AUs) (Pantic & Patras, 2006; Valstar & Pantic, 2007). They are essential for categorization of complex psychological states like various types of pain and mood (Williams, 2002) and are highly important cues for distinguishing posed from spontaneous facial expressions (Cohn & Schmidt, 2004; Valstar & al, 2006). Timing, duration and speed have been analyzed in several studies (Cohn & Schmidt, 2004; Valstar & al, 2006; Valstar & al 2007). However, little attention has been given to occurrence order (Valstar & al, 2006; Valstar & al 2007).

Several efforts have been recently reported on automatic analysis of facial expression data (Zeng & al, 2009; Sandbach & al, 2012; Gunes & al, 2011). We note from all cited methods that most recent methods employ probabilistic (Hidden Markov Models, Dynamic Bayesian Network), statistical (Support Vector Machine), and ensemble learning techniques (Gentle-Boost), which seem to be particularly suitable for automatic facial expression recognition from face image sequences. Because we want to explicitly study the dynamic behavior, both the HMM (Koelstra & al, 2010; Cohen & al, 2003) and DBN (Tong & al, 2007; Tong & al, 2010) can be used.

The presented work in this paper is a part of a project which aims to construct “An Optimal

Emotional System”, optimal in terms of speed, precise, full, and its performances reaching the 100%. The aim of this part is to reach the 100% of the facial expressions recognition rate.

This paper is an extended version of the work described in (Ghanem<sub>(1)</sub>, 2013), where the authors propose to explicitly analyze facial motion in terms of the order of occurrence of facial features by modeling semantic dependencies among facial features. For example, Joy expression configuration is as follows: Eyes are slightly closed and mouth is opened, its corners are pulled backward to ears. The fact is that it is not known in which order these facial deformations occur when displaying this expression. This is why we aim to discover temporal rules to give a new temporal description of each facial expression.

The evaluated method was Hidden Markov Models (HMM) and four facial expressions which are Joy, Disgust, Anger and Sadness were studied.

In the proposed approach, characteristic facial points were localized on the first frame, then, they were automatically tracked for the rest of the sequence. Five facial distances were calculated for each frame from each video leading to a multivariate facial time series. These time series were cleaned from noise, segmented and annotated to get sequence observations. These sequences were used as input to the HMM to allow it to learn the best structure of each facial expression studied.

The main contributions of this paper are summarized as follows:

- As HMM learning is generally trapped in a local optimal solution, we evaluate the use of the Differential Evolution (DE) optimization algorithm, in order to get a global optimal solution in the learning step of HMMs, instead of the classical Baum and Welch algorithm.
- To improve DE performances, we evaluate three (03) DE variants-based mutation criterion.

The remainder of this paper is structured as follows. We present a brief review of related work in the next section. Section 3 discusses the use of HMM in modeling occurrence order of facial dynamics. In Section 4, we investigate the proposed DE-based training of the HMM, we also evaluate different Schemes of DE. Experimental

results and discussions are presented in Section 5. Finally, Section 6 concludes the paper.

## 2. Related Work

Performance of classification methods is very dependent on the characteristic features used as input to classifiers, or, to local classifiers parameters. Many researchers in facial expression analysis were interested in enhancing classification performance by using statistic methods such as the Adaboost algorithm (Littlewort & al, 2006; Bartlett a al, 2005) and the Principal Component Analysis (PCA) (Lajevardi & Husain, 2009; Dubuisson & al, 2002) to select the most pertinent features to be used to classify expressions.

To optimize local classifiers parameters, a very few researchers were interested in using meta-heuristics. The authors in (Ye & al, 2008) propose a best-step operator (to refine the prey behavior) with the artificial fish-swarm algorithm (AFSA) for RBF neural network training. They applied the proposed algorithm to solve the problem of expression recognition. Their reported results demonstrate that the BAFSA has better global astringency and stability. That research indicates that the new algorithm has some advantages in terms of convergence performance and recognition rate, compared with BP and RBF neural network training. In (Mpipeiris & al, 2008) anatomical correspondence between faces is first established using a generic 3D face model which is deformed elastically to match the facial surfaces. Surface points are then used as a basis for classification according to a set of classification rules, which are discovered by the ant colony algorithm and particle swarm optimization ACO/PSO-based rule discovery algorithm. The performance of that algorithm has been evaluated on the BU-3DFEDB facial expression database, where a total recognition rate of 92.3% has been achieved, but the expressions ‘angry’ and ‘sad’ are not completely discriminated.

In (Ilbeygi & Shah-Hosseini, 2012), a Fuzzy Inference System (FIS) for emotion recognition from facial expressions is proposed. A genetic algorithm for parameter-tuning of the membership functions having big influence on the final precision of the fuzzy inference system is used. Experimental results report an average preci-

sion rate of 93.96% for emotion recognition of six basic emotions from RaFD dataset.

To optimize feature selection for input classifiers, the authors in (Bhanu & al, 2004) propose a genetic programming-based technique, which learns to discover composite operators and features that evolve from combinations of primitive image processing operations. In the mentioned approach, the output of the learned composite operator is a feature vector that is used for facial expression recognition. The experimental results show that the proposed approach can find good composite operators to effectively extract useful features. Obtained results are about 72% with genetic programming and 63,3% without genetic programming on the Japanese Dataset.

In (Yu & Bhanu; 2006), a learning method can discover the features automatically in a genetic programming-based approach that uses Gabor wavelet representation for primitive features and linear/nonlinear operators to synthesize new features. These new features are used to train a support vector machine classifier which is itself used for recognizing facial expressions. The obtained results are reported to be about 80,95% on the Japanese Dataset with a feature vector dimension equal to 35. In (Lajvardi & Hussain, 2009), a wrapper approach to feature selection from image sequences applied to the facial expression classification problem is introduced. The pre-processing phase automatically scans image sequences and detects frames with maximum intensity of facial expression. The features are generated using the log-Gabor filters. A global optimization genetic algorithm (GA) is adopted to select a sub-set of features based on minimization of the classification error. The wrapper approach is compared with two previously known filter-based feature selection methods: MID-mRMR and MIQ-mRMR. The features are classified using the naive Bayesian (NB) classifier. The average classification rates are: 79% (MIQ-mRMR), 78% (wrapper) and 64% (MID-mRMR) on CK dataset. The results from the filter methods did not appear to be significantly affected by the size of the feature subset.

In (Ruihu & Bin, 2011), to build a surveillance system, facial expression and body gesture features are extracted. Particle Swarm Optimization algorithm is used to select feature subset and parameters optimization. The selected features are trained or tested for the cascaded Support Vector Machine to obtain a high-

accuracy classifier. The accuracy of anxiety recognition is found to be about 91,46% with PSO/SVM, about 85,54% with PCA-SVM and about 83,27% with PCA-RBF on real video sequence captured by cameras control.

In (Anisha & al, 2011), a lip-contour model was developed to represent the boundary of the lip, and the parameters of the model were adapted using an artificial bee colony (ABC) algorithm to match it with the boundary contour of the lip. An SVM classifier was then employed to classify the emotion of the subject from the parameter set of the subject's lip-contour. The experiment was performed on 50 subjects, and the average case accuracy in emotion classification was found to be 86%.

In (Amir & MdJan, 2012), a hybrid model which combines a special genetic algorithm called Queen Bee Algorithm (QBA) with a fuzzy rule-based system is used to make better performance and parameter optimization to improve the accuracy of facial expression recognition. Therefore, the QBA as a training technique sets the fuzzy membership functions under the adverse conditions.

### 3. The HMM for Modeling Occurrence Order of Facial Temporal Dynamics

A specific contribution in the field of facial expression analysis is to introduce the occurrence order of facial feature deformations in the description of each facial expression. The aim is to study this parameter by using Hidden Markov Models as a dynamic descriptor tool as well as a classifier. The main three steps of a classification process are as follows:

#### 3.1. Facial Features Extraction

Image sequences from different facial expression databases were used. Eighteen facial points for each participant's video were manually pointed, and then automatically tracked for the rest of the sequence. Lucas-Kanade algorithm (Lucas & Kanade, 1981) was used to track the points as illustrated in Figure 1. The main limitation inherent in this tracker is that the tracking becomes less precise as the number of sequence frames increases, because of the accumulation of tracking errors over time. As we are only

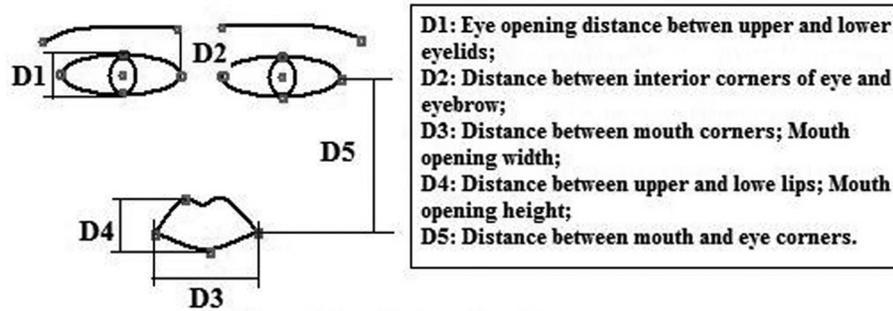


Figure 1. Characteristic distances.

interested in the first frames, where a facial distance transits from the neutral state to the onset one, the errors were negligible. After tracking, each frame of the video data is represented as a vector of the 18 facial points' 2D coordinates.

For each frame, geometric features  $D1 - D5$  (Figure 1) are calculated on the basis of the positions of facial points. Therefore, each frame of the captured video was represented as a 5-dimensional vector, so each video segment containing  $n$  frames is represented as a  $5 \times n$ -dimensional vector. This leads to multivariate (5) time series.

Eye blinking is a natural reaction; it occurs each 300 to 400 milliseconds. Generally, the eyelid automatically closes 10 to 30 times per minute. The adjusted seasonality for time series associated to eyelids movement reflects real eyelids movement without blinking. The Moving Averages (Makridakis & al, 1998), one of the first mathematical methods developed for seasonal factor estimates, is used to adjust seasonality of  $D1$  time series.

Another problem with the obtained time series concerns mouth time series which can be disturbed by speaking. Blinking and speaking are

usually manipulators and can disturb time series values. The solution to this problem is not discussed in this paper. However, to test the proposed method we have chosen to remove manually the speech sections from the studied sequences.

Interesting information is not directly accessible from the raw data. So, a common prerequisite task is changing data representation, for both information extraction and dimensionality reduction.

### 3.2. Data Representation

What characterizes a temporal nature is the existence of an order (temporal) in the data. This criterion would greatly benefit our analysis of facial expression dynamics. Indeed, in order to have a meaningful representation, we proposed to build symbolic representations of time series by incorporating "the occurrence order parameter". The new constructed information elements which are in a higher level are more comprehensible for humans.

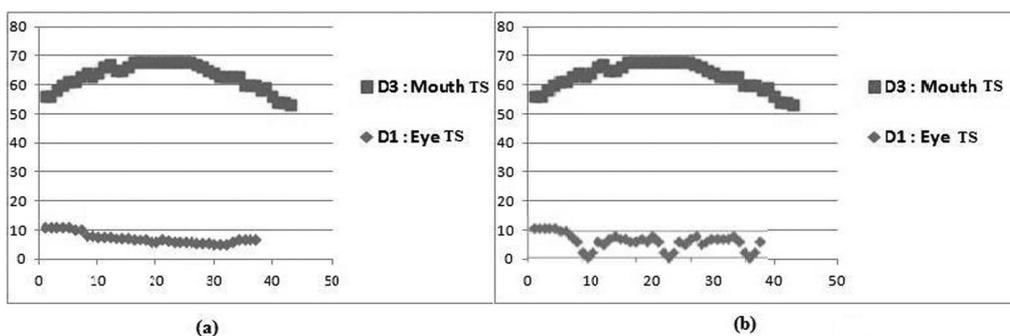


Figure 2. Example of TS adjustment seasonality. (a): Before; (b): After.

Each video is characterized by 5-Time series and each time series counts more than 200 (time) samples (each value is for a particular frame of the video). To reduce time series dimensionality, time series segmentation is performed. Time series segmentation can be seen as the task of detecting transition points. Transition points correspond to the frames where the distance increases or decreases with regard to the first frame.

Many time series segmentation methods have been developed in the literature (Arlot, Celisse, 2010; Gu'edon, 2008; Lavielle, 2009). Most of these methods are based on dynamic programming; they aim at reducing the number of the obtained segments. However, in our context, we were only interested in the detection of the first and second segments (just to know if the time series increases or decreases). For this purpose, the first transition point from each time series is detected. A transition point which corresponds to the start of a given emotion indicates the transition from neutral state to an onset one.

To detect transition points, a very simple algorithm based on a mobile means was used (Ghanem, 2013). A label determining whether the time series increases or decreases is assigned to each time series and the rank of these points is saved.

After segmentation, all time series associated with each video are sorted to detect the order of the changed time series, according to the saved ranks. It leads to a set of  $2 \times 5$  features to describe each video. (2=the label assigned to each time series and its rank, 5=facial distance time series). Then, the annotation process takes place; each time series (distance) has two discrete states which represent the "decrease/increase" states of the distance. If distances from D1 to D5 decrease, the values 1,2,3,4,5 are associated to each distance consecutively, and, if the distances from D1 to D5 increase, the values 6,7,8,9,10 are associated to each distance consecutively. The number of observations is 10 (1..5:D1..D5 decrease; 6..10:D1..D5 increase) . These new features represent all possible observations which can be emitted from each HMM state.

### 3.3. Classification Using HMM

In order to classify a studied facial expression, we choose to explore the HMM classifier.

#### 3.3.1. Formal Definition of an HMM

Formally, an HMM is characterized by the following:

N: the number of hidden states in the model

M: the number of distinct observation symbols

A: The state transition probability distribution

B: The observation symbol probability distribution

$\pi$ : The initial state distribution

Given appropriate values of N, M, A, B and  $\pi$ , the HMM can be used as a generator to give an observation sequence:  $O = O_1 O_2 \dots O_T$ .

The compact notation:  $\lambda = (A, B, \pi)$  is used to indicate the complete parameter set of the model.

#### 3.3.2. HMM Structure

A Hidden Markov Model HMM is a stochastic process for sequential data. Its performance depends heavily on the availability of an adequate amount of representative training data to estimate its parameters, and in some cases its topology.

In our work, we are interested in video sequences which begin by a neutral state. From each state the possible observations correspond to any of the 5 changed distances (which can increase or decrease). As we only consider the beginning of each facial expression and not the entire expression, we aim at investigating the order of changed facial distances when the state changes from neutral state to the onset one. We have chosen to study the three first changed distances for each facial expression. These distances form the basis for three possible states of the hidden variables.

On the other hand, with spontaneous smile, AU12 (lip corner puller) is usually first activated to express a slight emotion; then, with the increasing of emotion intensity, AU6 (cheek

raiser) is activated in an average of 0.4 second after the activation of AU12 (Schmidt, Cohn, 2001). This means that if the expression intensity is low, the number of activated units is less than when the intensity is high (HMMs are well suited when the used examples are of different length). Consequently, the type of our constructed HMM must have a Left-to-Right topology; the three states are sequentially connected such that the model is only allowed to stay in its current state or change to the next state.

The parameters used in the proposed HMM can be summarized as follows: The number of states is 3 ( $N=3$ , consequently the dimensions of A are  $A(3,3)$ ) (neutral, first changed distance, second changed distance), the number of observations is 10 ( $M=10$ , consequently, the dimensions of Bare  $B(3,10)$ ), (1..5:D1..D5 decrease; 6..10:D1..D5 increase), and the initial state probabilities are set to 1 for the neutral state and 0 for the others. Besides, initial transition probabilities are generated randomly.

To study the four facial expressions which are Joy, Sadness, Anger and Disgust, we associate one HMM to each facial expression.

HMM profiles are usually trained with a set of sequences that are known to belong to a single category. Consequently, this set of sequences is often redundant in the sense that many sequences would have the same values taken by the different features describing the category in question. Due to the variability among individuals, many facial expressions (e.g. Disgust or Anger) can be displayed in different ways. So, many sets of sequences can be associated to each category.

In order to reduce the run-time of HMM training, while providing accurate discovery and discrimination of different category behaviors, an unsupervised clustering process was performed (Ghanem<sub>(2)</sub>, 2013). To preserve training data from any information loss, we have associated a weight factor to each cluster. This weight factor is the frequency of observing instances with similar behavior (characteristics).

In many studies, clusters with small weight are considered as noisy instances, so they are removed. In this work, each cluster describes a specific behavior associated to the considered

category, it cannot be considered as noisy sequences. In this case, only a human expert can remove these clusters to improve classification accuracy.

When clustering sequences, both Euclidean distance and Dynamic Time warping (DTW) were used. Euclidean distance has been used because it is not time consuming. However, it cannot be used when produced sequences are not of the same length. In this case, we used DTW because it offers better accuracy than Euclidean distance in such situation, since it produces meaningful clusters. As a result, each cluster of each category represents a specific configuration (description) of the studied category (facial expression) leading to a set of different descriptions which can be formulated in the form of temporal rules to describe each facial expression.

Finally, we can deduce the following HMM parameters:  $N=3$  states,  $M=10$  observations, "A" dimensions are  $A(3,3)$ , "B" dimensions are  $B(3,10)$  and  $\pi$  dimensions are  $\pi(3,1)$ .

### 3.3.3. Classification

After the representation process, each video is represented by a simple sequence of three values instead of 5D time series. As instance "8 51" is a new constructed sequence which represents Joy expression. According to this configuration, three distances have changed, they are ordered from the first changed distance to the third one. The first changed distance is D3 and it increases (8), the second changed distance is D5 and it decreases (5) and the third changed distance is D1 and it decreases (1). All these new sequences are presented to each HMM as inputs to classify.

In the classification step, given a test sequence  $O$  and multiple HMM structures  $\lambda_i$ , (in our case we have 4 structures, one for each class of facial expressions, namely Joy, Disgust, Sadness and Anger), and in order to score how well a given model matches a given observation sequence, we compute the probability of the observation sequence given at each time, one of the four structures,  $P(O/\lambda_i)$ . The test sequence  $O$  is assigned to the HMM which have the maximum likelihood.

## 4. HMM Learning With Differential Evolution (DE)

In order to improve HMM classification rates, we explore in this paper the use of an optimization method to train HMMs instead of the classical Baum and Welch algorithm.

### 4.1. Problem Formulation

Training HMMs is a crucial problem for most applications using them since it is supposed to allow optimal adaptation of the model parameters to the observed training data, in order to create best models for real phenomena. The fact is that in literature there is no optimal way of estimating the model parameters. An iterative procedure called Baum and Welch method (or, equivalently, Expectation Maximization (EM)) is widely used to estimate HMM parameters, but learning is generally trapped in local optimal solutions. Instead of using the Baum and Welch approach that suffers from stagnation, we use in this paper a DE approach to train the system for setting the A, B and  $\pi$  parameters.

### 4.2. Representation of Solution

In this type of algorithms, a solution is represented by an “Individual”, and the population is the whole set of individuals. To choose the better individual between two individuals, an objective function, called *fitness*, is computed.

Each individual represents an instance of the HMM parameters which are A: the transition matrix and B the observation matrix. The initial matrix  $\pi$  is not optimized here because we suppose that each expression starts from the neutral state, so the initial probabilities are always (1,0,0). A constraint which is the sum of each row of each vector is equal to “1” and is associated to both vectors A and B,

### 4.3. Differential Evolution

Differential Evolution is an evolutionary algorithm introduced by Price and Storn (Storn, Price, 1997), it is designed for global optimization problems over continuous domains. It follows the general procedure of an evolutionary

algorithm: an initial population of individuals is created by random selection and evaluated; then the algorithm enters a loop of generating offspring, evaluating offspring, and selecting solutions to create the next generation, till a stop condition is met. Generating offspring consists of two operations: differential mutation and differential crossover.

The mutation operation creates a new individual  $v_i$  called the *mutant* by adding the weighted difference between two individuals (vectors) chosen randomly from the population to a third one, also chosen randomly, as shown in equation (1) below. In the equation,  $i, i_1, i_2$  and  $i_3 \in \{1, \dots, NP\}$  are mutually different indices,  $NP$  is the size of the population, and  $F$  is a real positive scaling factor of the difference  $d_i = x_{i2} - x_{i3}$ .

$$v_i = x_{i1} + F * (x_{i2} - x_{i3}) \quad (1)$$

The crossover operator implements a discrete recombination of the mutant,  $v_i$ , and the parent vector,  $x_i$ , to produce the offspring  $u_i$ . This operator is formulated as shown in equation (2) below, where:  $U_j(0, 1)$  is a random number in the interval  $[0, 1]$ ,  $CR \in [0, 1]$  is the crossover rate, and  $k \in \{1, 2, \dots, d\}$  is a random parameter index, chosen once for each individual  $i$  to make sure that at least one parameter is always selected from the mutant. The most popular values for  $CR$  are in the range  $[0.4, 1]$ .

$$u_i(j) = \begin{cases} v_i(j), & \text{if } U_j(0, 1) \leq CR \text{ or } j = k, \\ x_i(j) & \text{otherwise} \end{cases} \quad (2)$$

After applying differential crossover, the obtained individual, called *trial vector*, is compared to the parent individual, also called *target vector*. A greedy selection takes place at this point, i.e. the fittest of them will become the target vector.

In the case of our work, the mutation and crossover operators are applied on matrices instead of vectors, but the principle of each remains the same. Indeed, the mutant individual is defined as:

$$\begin{aligned} \text{Mutant individual} &= \{A\_mutant, B\_mutant\} \\ \begin{cases} A\_mutant = A_1 + F * (A_2 - A_3) \\ B\_mutant = B_1 + F * (B_2 - B_3) \end{cases} \end{aligned} \quad (3)$$

In equation (3),  $A_1, A_2, A_3$  are randomly chosen individuals representing potential state transition probability distribution matrices, and  $B_1, B_2, B_3$  are randomly chosen observation symbol probability distribution matrices which represent candidate solutions.

The trial individual is obtained after the application of a crossover operator on the mutant individual and the parent (target) individual, it is defined as:

Trial individual =  $\{A\_trial, B\_trial\}$  where:

$$A\_trial(i, :) = \begin{cases} A\_mutant(i, :), & \text{if } rand < CR \\ A\_target(i, :), & \text{otherwise} \end{cases} \quad (4)$$

$$B\_trial(i, :) = \begin{cases} B\_mutant(i, :), & \text{if } rand < CR \\ B\_target(i, :), & \text{otherwise} \end{cases} \quad (5)$$

Finally the selection operator is defined as the individual which has the best fitness when comparing between target and trial individuals.

In our work, the objective function (fitness) represents the maximum likelihood and is defined

as:

$$F = MAX[P(O|\lambda)] \quad (6)$$

$O$  is an observed sequence, and  $\lambda$  represents all numeric parameters defining a considered HMM ( $\lambda = (A, B, \pi)$ ).

#### 4.4. The Proposed Hybrid EM-DE Algorithm

In this work, we propose a new sequential hybrid EM-DE learning algorithm where DE initial individuals are locally optimized with the classical EM Algorithm. This hybrid can be formulated as in Algorithm 1 below.

In Algorithm 1, target is the current individual of the population to be evolved and the trial individual is that generated after mutation and crossover operations to be a candidate substitute of the target one using a greedy selection.

When speaking about the algorithm complexity, we have to note that the complexity of DE is linear, when associated to Baum & Welch algorithm, we get the same complexity of BW, it means  $O(2m^2n)$  where  $m$  is the number of states and  $n$  is the number of observations.

#### Algorithm 1: hybrid EM-DE learning algorithm.

```

G=1// Initialize number of iterations;
1- Choose the DE parameters: NP, F and CR;
2- PopG < -Initialize population randomly;
3- PopG = EM(PopG)//estimate the HMM parameters with Baum and Welch
algorithm;
4- While (G<Number of iterations) do
  For every individual:
    a. Choose auxiliary parents: Xr1,G; Xr2,G; Xr3,G;
    b. Create mutant individuals muti,G using mutation
operator(Equation3);
    c. Add a new individual called trial by applying differential crossover
operator to mutant individual muti,G (Equations 4,5);
    d. greedy selection between trial and target individuals (Equation 6);
  End for
    e. G < -G+1;
5- Endwhile

```

#### 4.5. Different Variants of DE

In this work we evaluate other variants of DE. Each scheme concerns the mutation operator. The different evaluated schemes are:

- a- DE/best/1: The perturbed parent is the current best member added with a weighted difference vector:

$$\begin{aligned} \text{Mutant individual} &= \{A\_mutant, B\_mutant\} \\ \begin{cases} A\_mutant = A_{best} + F * (A_2 - A_3) \\ B\_mutant = B_{best} + F * (B_2 - B_3) \end{cases} \end{aligned} \quad (7)$$

- b- DE/current-to-best/1: The perturbed parent is the target member added with a weighted difference vector:

$$\begin{aligned} \text{Mutant individual} &= \{A\_mutant, B\_mutant\} \\ \begin{cases} A\_mutant = A_1 + F * (A_{best} - A_1) \\ \quad \quad \quad + F * (A_2 - A_3) \\ B\_mutant = B_1 + F * (B_{best} - B_1) \\ \quad \quad \quad + F * (B_2 - B_3) \end{cases} \end{aligned} \quad (8)$$

- c- DE/self-mutation: The perturbed parent is a random member added with a weighted difference vector, but with only two parents where the mutant individual participates in the difference vector:

$$\begin{aligned} \text{Mutant individual} &= \{A\_mutant, B\_mutant\} \\ \begin{cases} A\_mutant = A_1 + F * (A_2 - A_1) \\ B\_mutant = B_1 + F * (B_2 - B_1) \end{cases} \end{aligned} \quad (9)$$

### 5. Experimental Results and Discussion

In order to evaluate performances of the proposed method, we conducted a series of experiments following different scenarios on three universal Datasets.

#### 5.1. Facial Expression Datasets

To validate our approach, we conducted our experiments using the Cohn-Kanade (2000), MMI (2005) and Semaine (2012) databases. The Cohn-Kanade database consists of 100 university students aged between 18 and 30, of whom 65% are females, 15% are African-American and 3% are Asian or Latino.

For our experiments, we used 231 image sequences from the database. The sequences come from 96 subjects, with 1 – 4 posed expressions per subject (Joy, Anger, Sadness and Disgust). The MMI posed database is a resource for Action unit (AU) and basic emotion recognition from face video. It consists of five parts; each part is recorded in a different condition. The second set of the recorded data is posed displays of the six basic facial expressions: Joy, Sadness, Anger, Disgust, Fear and Surprise. In total, 238 videos of 28 subjects are recorded. All expressions are recorded twice. For our experiments, we have used 100 image sequences from the database. Each video contains more than 100 frames. In our experiments we are interested in all frames displaying the expressions from the neutral state, the beginning of the expression (onset) until the apex. We have used 20 frames from each video, so about 2000 images. The two datasets contain recordings of subjects displaying posed facial expressions in frontal or near-frontal head poses and under controlled lighting conditions.

The Semaine spontaneous dataset contains displays of spontaneous expressions recorded in a natural environment. For our experiments, we have used 80 image sequences. The sequences come from 10 subjects. After removing the speech sections from each image sequence, we extract at least three sub-sequences from each image sequence leading to more than 240 image sequences. Most of these sequences display joy expression (smile displaying).

#### 5.2. Our Experiments

We conducted our experiments in two steps. In the first step we aimed at evaluating the real performance of the proposed approach before optimization using different scenarios. First, we split posed studied datasets into three parts, two parts were used as training sets and the last one as a test set. In the second experiment, and in order to test the ability of generalizing to novel conditions, we used one dataset in training and another in the test (cross datasets). In order to collect a set of examples that is as sparse as possible, which spans the problem space completely, we implemented a last experiment where we used two thirds of two datasets

in the training step and one third of the two datasets in the test step.

All our experiments are Subject-Independent. The obtained results are presented in Table 1 below.

Many important conclusions were drawn from Table 1 and are discussed in detail in (Ghanem, 2013). The most important one is that HMM are well suited in situations where it is important to model occurrence order of facial features deformations, and many description rules can be extracted from Transition and Observation (A, B) Matrices learned by EM algorithm. We have shown that it is possible to describe the dynamic of a produced facial expression using one, two or three temporal rules. Also, it has been shown that only ‘first order’ (one rule) is not sufficient to recognize expressions because many facial expressions can start in the same way when speaking about the order of facial features deformations. This is why the second and third orders (when present) are considered to recognize facial expressions. This is exactly what was concluded in (Ghanem, Caplier, 2011).

In the second step, we aimed at evaluating the performance of the proposed approach after the optimization using different DE schemes. In order to fix DE parameters (F, CR, NP), we have studied the influence of these parameters on the classification rates. We have found that the value of CR has no significant influence on the process in the contrast to F, where we noted that the variation of this parameter has significant influence on the results. As a result, we have found that the best values for these parameters are NP=20, F=0.4, CR=0.6 and the number of iterations It=1000. The obtained results are given as the mean of 30 runs for each scheme and are summarized in Tables 2 and 3.

It is possible to draw many conclusions from Tables 2 and 3 with both MMI and CK databases after optimization. HMM training with the DE algorithm (DE/rand/1) gives better rates than training with Baum and Welch (EM) algorithm for all studied expressions. The different schemes of DE algorithm give better rates too, except with DE/best/1 scheme (with MMI database). The scheme DE/best/1 is the worst one for both databases, it gives the lowest rates

Training Dataset	Test Dataset	Joy	Anger	Sadness	Disgust	Average Rates
CK	CK	100%	100%	100%	81,81%	95.45%
CK	MMI	76.47%	47.06%	30.43%	58.62%	53.15%
CK	SEMAINE	91.89%	33.33%	90%	100%	78.81%
MMI	MMI	100%	80%	81.82%	70.59%	83.10%
MMI	CK	100%	56.58%	92.59%	91.35%	85.13%
MMI	SEMAINE	94.60%	75%	80%	100%	87.4%
CK+MMI	SEMAINE	94.60%	91.67%	100%	100%	<b>96.57%</b>

Table 1. Cross database testing.

Algorithm	Classification Rates				Average Rates
	Anger	Disgust	Joy	Sadness	
<b>DE/Self</b>	<b>100%</b>	<b>82,57%</b>	<b>100%</b>	<b>100%</b>	<b>95,64%</b>
DE/rand/1	97,91%	81,81%	100%	100%	94,93%
DE/current-to-best/1	90%	82,11%	100%	100%	93,03%
<b>Baum and Welch</b>	<b>100%</b>	<b>80,29%</b>	<b>100%</b>	<b>100%</b>	<b>95,07%</b>
DE/best/1	91,66%	83,03%	100%	97,77%	93,11%

Table 2. Classification rates of the different DE schemes + EM for Cohn & Kanade database.

Algorithm	Classification Rates				Average Rates
	Anger	Disgust	Joy	Sadness	
<b>DE/Self</b>	<b>100 %</b>	<b>75,29 %</b>	<b>100 %</b>	<b>90,29 %</b>	<b>91,40 %</b>
DE/rand/1	100%	72,15%	100%	86,96%	89,78%
DE/current-to-best/1	100%	70,98%	99,33%	88,48%	89,69%
<b>Baum and Welch</b>	<b>95,55 %</b>	<b>71,36 %</b>	<b>100 %</b>	<b>86,05 %</b>	<b>88 %</b>
DE/best/1	97,77%	68,23%	98%	85,14%	87,29%

Table 3. Classification rates of the different DE schemes + EM for MMI database.

when compared to EM and the other variants of DE. The scheme DE/Self is the algorithm which gives the best model structure, i. e. the highest rates for all studied expressions.

For all DE schemes, in addition to the Basic HMM (EM), Disgust is the most difficult expression to recognize, it has the lowest rates. This can be explained by the fact that there are several configurations of displaying this facial expression which leads to a confusion with other facial expressions (e.g Anger or Sadness).

In the last experiment, in the second step, we tested the Semaine database. As seen in Figure 3, there is a stability of the obtained results for the classification of each facial expression across different runs when HMM training is done by the ‘DE/self’ scheme (Figure 3.b) in contrast to the case where the HMM training is done by the EM algorithm (Figure 3.a).

A comparison of the recognition rates achieved for each facial expression with the state of the art (Fanelli & al, 2010; Khan & al, 2012; Zhao, Pietik¨ Ainen, 2007; Shan & al, 2009; Aleksi, Katsaggelos, 2005; Yeasin & al, 2004; Kotsia

& al, 2008), is not really possible. The fact is that the lack of a standardized training and testing database do not allow us to compare any proposed method with others (in term of our results being higher or lower than others). Besides rules and protocols on how to use images or videos in training and testing, a certified benchmark that contains images and video sequences (at different resolutions) of people displaying expressions under different conditions (lighting, occlusions, head rotations, etc) is required. Moreover, each method in the literature has its own constraints, it proposes a full automatic system or a semi automatic system, also it can study three, four, five, six or seven classes. Indeed, when training on a higher number of classes, obtained accuracy rates in testing step, are reduced.

However, our goal in this work is to prove the effectiveness of optimization methods to improve classification process. According to this context, and when considering obtained results before and after optimization, we can see that obtained results after optimization are very interesting.

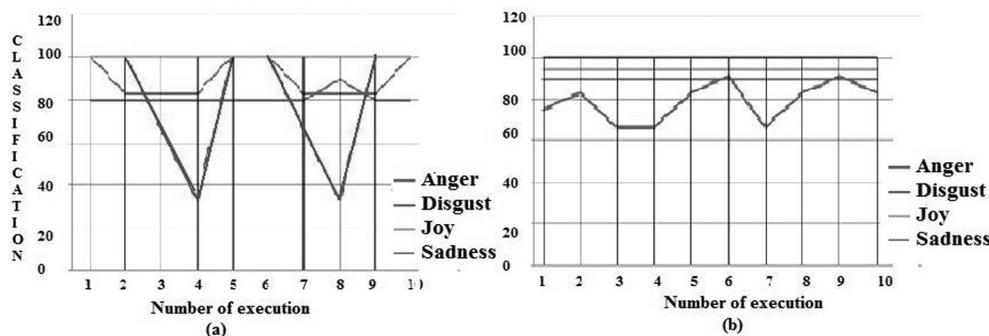


Figure 3. Classification results for Semaine dataset: (a): HMM learning with EM; (b): HMM learning with EM-DE/self.

## 6. Conclusion and Future Work

In this paper, a new HMM approach based on differential evolution for facial expression recognition has been proposed. It consists of improving the performance of basic HMM through applying the DE algorithm for its training. The proposed approach has been validated by using it for two datasets, namely: Cohn & Kanade and the MMI datasets with different configurations. Four basic expressions were considered: anger, disgust, joy and sadness.

Different mutation strategies for the proposed DE-based HMM were compared against each other and against the state-of-the-art Baum and Welch training algorithm. The obtained results proved superiority of the proposed DE-HMM hybrid, where we found that, on the whole, the DE/Self training algorithm gave 95,64% as an average success rate, then comes the Baum and Welch training algorithm with a success rate of 95,07% for the Cohn & Kanade database. For the MMI database, more superiority of the DE/Self was observed; it gave a success rate of 91,40% compared to the Baum and Welch algorithm with a rate of 88%.

As our future work, we plan to train and test the model on spontaneous datasets to deduce new temporal rules, in order to distinguish between posed and spontaneous expressions. Also, a comparison between the DE and other metaheuristics for HMM training would be a good path of research.

## References

- [1] S. P. ALEKSIC, K. A. KATSAGGELOS, Automatic Facial Expression Recognition Using Facial Animation Parameters And Multi-Stream Hmms. *IEEE Trans. Signal Processing, Supplement on Secure Media*, 2005.
- [2] J. AMIR, N. MDJAN, A Modified Genetic Model Based on the Queen Bee Algorithm for Facial Expression Classification. *Journal of Computational and Theoretical Nanoscience*, **9**(8), pp. 1109–1114, 2012.
- [3] H. ANISHA, R. PRATYUSHA, C. ARUNA, Emotion Recognition from the Lip-contour of a Subject Using Artificial Bee Colony Optimization Algorithm. *2nd Swarm, Evolutionary and Memetic Computing Conference, Part I, 7076*, pp. 610–617, 2011.
- [4] S. ARLOT, A. CELISSE, Segmentation of the Mean of Heteroscedastic Data via Cross-Validation. *Statistics and Computing*, pp. 1–20, 2010.
- [5] M. BARTLETT, G. LITTLEWORT, M. FRANK, C. LAINSCSEK, I. FASEL, J. MOVELLAN, Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior, *Computer Vision and Pattern Recognition. IEEE Computer Society Conference*, **2**, pp. 568–573, 2005.
- [6] B. BHANU, J. YU, X. TAN, Y. LIN, Feature Synthesis Using Genetic Programming for Face Expression Recognition. *Genetic and Evolutionary Computation*, pp. 896–907, 2004.
- [7] L. H. BURSTYN, B. BALLOT, C. H. DIEDERIK, *Dictionary of Scientific Biography*, **1**, pp. 628, 1973.
- [8] C. SHAN, S. GONG, P. W. MCOWAN, Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study, *Image and Vision Computing*, **27**, pp. 803–816, 2009.
- [9] I. COHEN, N. SEBE, A. GARG, L. CHEN, T. S. HUANG, Facial expression recognition from video sequences: temporal and static modeling. *In Computer Vision and Image Understanding*, **91**, pp. 160–187, 2003.
- [10] J. F. COHN, K. L. SCHMIDT, The timing of facial motion in posed and spontaneous smiles. *J. WAVELETS, Multi-resolution and Information Processing*, **2**(2), pp. 121–132, 2004.
- [11] S. DUBUISSON, F. DAVOINE, M. MASSON, A solution for facial expression representation and recognition. *Signal Processing: Image Communication*, **17**(9), pp. 657–673, 2002.
- [12] P. EKMAN, E. L. ROSENBERG, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System*, 2nd edition, Oxford University Press, 2005.
- [13] G. FANELLI, A. YAO, P. L. NOEL, J. GALL, L. VAN GOOL, Hough Forest-based Facial Expression Recognition from Video Sequences, *Workshop on Sign, Gesture and Activity (SGA'10)*, 2010, LNCS 6553, pp. 195–206. ©Springer
- [14] K. GHANEM, A. CAPLIER, Occurrence order detection of face features deformations in posed expressions, *International Conference on Informatics Engineering & Information Science, PART 2, CCIS:252*, (2011), pp. 56–66.
- [15] K. GHANEM<sup>(1)</sup>, Hidden Markov Models for Modeling Occurrence Order of Facial Temporal Dynamics. *International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS)*, pp. 174–185, 2013.
- [16] K. GHANEM<sup>(2)</sup>, Towards more accurate clustering method by using dynamic time warping. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, **3**(2), 2013.
- [17] H. GUNES, B. SCHULLER, M. PANTIC, R. COWIE, Emotion representation, analysis and synthesis in continuous space: a survey. *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 827–834, 2011.

- [18] Y. GU'EDON, Exploring the segmentation space for the assessment of multiple change-point models. Institut National de Recherche en Informatique et en Automatique, Cahier de recherche 6619, 2008.
- [19] M. ILBEYGI, H. SHAH-HOSSEINI, A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. *Engineering Applications of Artificial Intelligence*, **25**, pp. 130–146, 2012.
- [20] KANADE, J. F. COHN, Y. TIAN, Comprehensive database for facial expression analysis. *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46–53, 2000.
- [21] S. KOELSTRA, M. PANTIC, I. PATRAS, A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models. *IEEE transactions on pattern analysis and machine intelligence*, **32**, 2010.
- [22] R. A. KHAN, A. MEYER, H. KONIK, S. BOUAKAZ, Human Vision Inspired Framework For Facial Expressions Recognition. *19th IEEE International Conference on Image Processing*, 2012.
- [23] I. KOTSIA, I. PITAS, Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transactions on Image Processing*, **16**(1), pp. 172–187, 2007.
- [24] I. KOTSIA, S. ZAFEIRIOU, I. PITAS, Texture and Shape Information Fusion for Facial Expression and Facial Action Unit Recognition. *Pattern Recognition*, **41**, pp. 833–851, 2008.
- [25] S. M. LAJEVARDI, Z. M. HUSSAIN, Feature selection for facial expression recognition based on mutual information. *5th IEEE GCC Conference & Exhibition*, pp. 1–5, 2009.
- [26] S. M. LAJEVARDI, Z. M. HUSSAIN, Feature selection for facial expression recognition based on optimization algorithm. *2nd International Workshop on Nonlinear Dynamics and Synchronization*, pp. 182–185, 2009.
- [27] M. LAVIELLE, Detection of Changes using a Penalized Contrast (the DCPC algorithm), <http://www.math.upsud.fr/lavielleprogrammes/lavielle.html>, 2009.
- [28] G. LITTLEWORT, M. BARTLETT, I. FASEL, J. SUSSKIND, J. MOVELLAN, Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, **24**(6), pp. 615–625, 2006.
- [29] B. D LUCAS, T. KANADE, An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings of Imaging Understanding Workshop*, pp. 121–130, 1981.
- [30] S. MAKRIDAKIS, S. C. WHEELWRIGHT, R. J. HYNDMAN, Forecasting: methods and applications, 3rd edn, John Wiley & Sons, New York, 1998.
- [31] G. MCKEOWN, M. VALSTAR, R. COWIE, M. PANTIC, AND M. SCHRODER, The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, **3**, pp. 5–17, 2012.
- [32] I. MPIPERIS, S. MALASSIOTIS, V. PETRIDIS, M. G. STRINTZIS, 3d Facial Expression Recognition Using Swarm Intelligence. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2133–2136, 2008.
- [33] M. PANTIC, I. PATRAS, Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, **36**(2), pp. 433, 2006.
- [34] M. PANTIC, M. F. VALSTAR, R. RADEMAKER L. MAAT, Web-based Database for Facial Expression Analysis. *Proc. IEEE Int'l Conf. Multimedia and Expo*, 2005.
- [35] W. RUIHU, F. BIN, Emotion Fusion Recognition for Intelligent Surveillance with PSO-CSVM. *International Conference on Automation, Communication, Architectonics and Materials*, 225–226(1–2), pp. 51–56, 2011.
- [36] G. SANDBACH, S. P. ZAFEIRIOU, M. PANTIC, L. J. YIN, Static and dynamic 3D facial expression recognition: A comprehensive survey, *IVC*, **30**(10), pp. 683–697, 2012.
- [37] K. SCHMIDT J. COHN, Dynamics of Facial Expression: Normative Characteristics and Individual Differences. *IEEE Int'l Conf. Multimedia and Expo*, pp. 728–731, 2001.
- [38] R. STORN, K. PRICE, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Global Optimization*, **11**(4), pp. 341–359, 1997.
- [39] Y. TONG, W. LIAO, Q. JI, Facial action unit recognition by exploiting their dynamics and semantic relationships. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **29**(10), pp. 1683–1699, 2007.
- [40] Y. TONG, J. CHEN, Q. JI, A Unified Probabilistic Framework for Spontaneous Facial Action Modeling and Understanding. *IEEE transactions on pattern analysis and machine intelligence*, **32**(2), 2010.
- [41] M. F. VALSTAR AND M. PANTIC, Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. *IEEE International Conference on Human-computer interaction*, pp. 118–127, 2007.
- [42] M. F. VALSTAR, M. PANTIC, Z. AMBADAR, & J. F. COHN, Spontaneous versus Posed Facial Behavior: Automatic Analysis of Brow Actions. *ACM Int'l Conf Multimodal Interfaces*, pp. 162–170, 2006.
- [43] M. F. VALSTAR, H. GUNES, & M. PANTIC, How to Distinguish Posed from Spontaneous Smiles Using Geometric Features, *ACM Int'l Conf. Multimodal Interfaces*, pp. 38–45, 2007.

- [44] A. C. WILLIAMS, Facial expression of pain: An evolutionary account, *Behavioral and Brain Sciences*, **25**(4), pp. 439–488, 2002.
- [45] W. YE, W. XIAOJUN, W. SHITONG, Y. JINGYU, Facial Expression Recognition Using RBF Neural Network Based on Improved Artificial Fish Swarm Algorithm, 27th Chinese Control Conference, **4**, pp. 416–420, 2008.
- [46] M. YEASIN, B. BULLOT, R. SHARMA, From Facial Expression to Level of Interest: A Spatio-Temporal Approach. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.
- [47] J. YU, B. BHANU, Evolutionary feature synthesis for facial expression recognition, *Pattern Recognition Letters*, volume **27**, pp. 1289–1298, 2006.
- [48] Z. ZENG, M. PANTIC, G. I. ROISMAN, T. S. HUANG, A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE transactions on pattern analysis and machine intelligence*, **31**(1), 2009.
- [49] G. ZHAO, M. PIETIKÄINEN, Dynamic Texture Recognition Using Local Binary Patterns With an Application to Facial Expressions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **29**, pp. 915–928, 2007.

---

KHADOUJA GHANEM is a lecturer at the Constantine 2 University in Algeria. She received her PhD degree in Computer Science from Mentouri University of Constantine in 2010. Her current research interests include, but are not limited to the areas of human machine interaction and specifically in facial expression recognition from still images and image sequences as well as in the area of clustering, classification, information retrieval and organization techniques, then their application in different fields.

---



---

AMER DRAA is a lecturer at the Constantine 2 University, Algeria. He received his BSc. Master and PhD degrees in Computer Science in 2002, 2004 and 2011 respectively, from Mentouri University, Algeria. His current research interests include complex adaptive systems, meta-heuristics, computer vision and nature-inspired computing.

---



---

ELVIS VYUMVUHORE was a Master student at the Constantine 2 University, Algeria, he received his Master degree from Constantine 2, Algeria in 2013. His research interests include complex adaptive systems, meta-heuristics and nature-inspired computing.

---



---

ARSÈNE SIMBABAWÉ was a Master student at the Constantine 2 University, Algeria, he received his Master degree from Constantine 2, Algeria in 2013. His research interests include complex adaptive systems, meta-heuristics and nature-inspired computing.

---

*Received:* May, 2014  
*Revised:* October, 2014  
*Accepted:* November, 2014

*Contact address:*  
 Khadoudja Ghanem  
 MISC Laboratory  
 Constantine 2 University  
 Constantine  
 Algeria  
 e-mail: gkhadoudja@yahoo.fr