A Fusion Process Based on Belief Theory for Classification of Facial Basic Emotions

Z.Hammal, A. Caplier, M. Rombaut
LIS_INPG
46 avenue Félix Viallet
Grenoble,France.
firtsname.name@lis.inpg.fr, zakia_hammal@yahoo.fr

Abstract - *This paper presents a system of facial expressions classification based on a data fusion process using the belief theory. The considered expressions correspond to the six universal emotions (joy, surprise, disgust, sadness, anger, fear) as well as to the neutral expression. Since some of the six basic emotions are difficult to simulate by non-actor people, the performances of the classification system are evaluated only for four expressions (joy, surprise, disgust and neutral). The proposed algorithm is based on the analysis of characteristic distances measuring the deformations of facial features, which are computed on skeletons of expression. The skeletons are the result of a contour segmentation process of facial permanent features (mouth, eyes and eyebrows). The considered distances are used to develop an expert system for classification. The performances and the limits of the recognition system and its ability to deal with different databases are highlighted thanks to the analysis of a great number of results on three different databases: the Hammal-Caplier database, the Cohn-Kanade database and the Cottrel database.*

Keywords*: Classification, facial expression, data fusion, transferable belief model.*

1. Introduction

With the extension of the domains of application involving human computer interactions (HCI), keyboards and mice are not well suited as efficient mean of interaction between man and machine. The user interface for computer systems is evolving into an intelligent multi-modal interface by taking into account the user behavior, his speech and his facial expressions in order to make machines use as natural as possible [3].

The face is the most expressive and communicative part of a human. As indicated by Mehrabian [4], in face-to-face human communication, only 7% of the communicative message is due to linguistic language, 38% is due to paralanguage like intonation, while 55% of it is transferred by facial expressions.

From a physiological point of view, a facial expression results from the deformations of some facial features, these deformations being caused by an emotion. In his seminal work, Eckman [5] shows that there are six universal emotions (*joy, surprise, disgust, sadness, anger* and *fear)*, which could be associated to 6 facial expressions.

Most of expression information is contained in the deformation of the principal permanent facial features (eyes, eyebrows and mouth). This observation is also validated from a psychological point of view by various works [5, 6]. The latter shows that information about a facial expression is not contained in the deformation of a particular feature but in the combination of deformations of several features. In his work, Bassili [6] shows that facial expressions could be characterized by using only the temporal evolution of some points defined on the permanent facial features.

Several methods have been proposed for facial expression synthesis and recognition based on the analysis of facial features deformations. Cohn et al [7] recognize Action Units (AUs) and some combinations of AUs in facial image sequences. After a manual marking of facial feature points around the contours of the eyebrows, the eyes, the nose and the mouth in the first image of the sequence, they use Lucas-Kanade's points tracking algorithm [8] to track automatically the points in the remaining of the sequence. They use Hidden Markov Model to combine the obtained AUs in order to classify the six facial expressions and the *neutral* expression.

Pantic and Rothkrantz [9] use point based models made of two 2D facial views, the frontal view and the side view. After automatic segmentation of facial features (eyes, eyebrows and mouth), they code some characteristic points (such as eyes corners, mouth corners…) into AUs using a set of rules and use Ekman Facial Actions Coding System to recognize the six facial expressions.

Cohen and al [10] develop a system using a model based on non-rigid face tracking algorithm to extract local motion features. These motion features are the inputs of a Bayesian network classifier used to recognize the six facial expressions and the *neutral* expression.

Based on the work of Bassili, we adopt an approach based on the analysis of the permanent facial features (eyes, eyebrows and mouth). Since such analysis of facial expressions uses a whole set of sensors to characterize an expression, this requires a classification process based on data fusion.

To combine the values of the whole set of data, most of the time the authors apply classification techniques based on probabilistic theory such as Bayesian network or Hidden Markov Model [7, 9]. These approaches are very efficient if a large learning data set can be used which is not our case because of the difficulty to realize good simulations of all the considered facial expressions.

Moreover in most of cases, the existing expression analyzers perform a classification of the examined expression into one of the basic emotion categories proposed by Ekman and Friesen [11]. This approach of expressions classification has two main limitations. First, since human people are not "binary", "pure" facial expressions are rarely produced. People show a mixture of facial expressions. Therefore, the classification of an expression into a single emotion category is not realistic. The automated facial expression analyzer should be able to recognize mixed emotion categories.

In addition, the system must be able to take into account the fact that each person has his/her own maximal intensity of displaying a particular facial expression and so must deal with different expressions intensities.

For all these reasons, the possibility theory or transferable belief model (TBM) [12] appear to be well suited to the facial expression recognition problem. The suitability of our approach is also discussed and demonstrated in [13] where we compare it with the Bayesian Theory and the Hidden Markov Models. The results of this comparison show that the best classification rates are those of the Belief Theory. Moreover, in addition to its capacity of generalization, the use of the Belief Theory emphasizes the fact that some expressions are not always dissociable and allows to recognize a mixture of facial expressions contrary to the Hidden Markov Models and the Bayesian Theory. This model also facilitates the integration of a priori knowledge about the problem. It can deal with uncertain and imprecise data, which could be the case with data measures resulting from an automatic video based segmentation algorithm.

This paper tackles the specific problem of facial expression classification which is an application of the framework of symbolic classification using numerical data get from different sensors.

2. Global overview of the system

Fig.1 presents an overview of the proposed system. It is divided into 5 different steps:

1. In the segmentation step, facial features contours are extracted with the segmentation algorithms described in [1, 2]. This leads to "facial skeletons".
2. In the data extraction step, distances related to the deformation of facial features contours are computed.
3. In the analysis step, facial deformations are encoded and symbolic states are related to the measured distances.
4. In the classification step, the transferable belief model (TBM) [12] is used for the facial expression recognition purpose. With the use of TBM, it is possible to add a supplementary expression called *unknown* expression which is associated to all the expressions which do not correspond to any of the 7 expressions (*joy, surprise, disgust, fear, anger, sadness* and *neutral*)). This yields to the definition of a reject class.
5. In the post-processing step, a separation of doubt states is proposed.

Section 3 describes the low level data used for the complete recognition process i.e. the different numerical distances computed on face skeletons and the low-level information used in the post-processing step. Section 4 describes the logic structure of the fusion process. Section 5 shows how to take into account uncertainty and inaccuracy by using the transferable belief functions. Finally, classification rates evaluated on three different databases are presented in section 6.

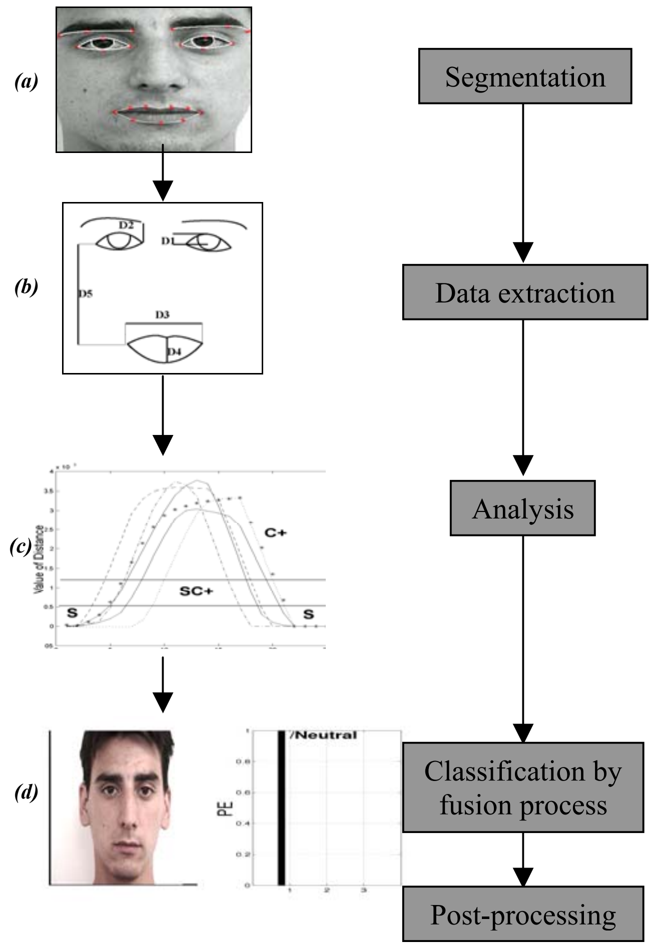

Fig.1 - Overview of the classification system

3. Data extraction for classification and post-processing

Facial expressions are characterized by a set of deformations occurring on permanent facial features such as eyes, eyebrows and mouth. The assumption that human beings are able to recognize a facial expression by the analysis of these deformations has been partly validated by a recognition rate of 60% obtained by an experimentation carried out in psychology. 60 subjects (30 males and 30 females) were asked to recognize a given facial expression in viewing only the extracted contours of eyes, eyebrows and mouth.

The aim of our classification method is to proceed in a similar way. After the segmentation step, skeletons of expressions are available. An example of skeleton is given in (Fig.1.b); it is made of the extracted contours of eyes, eyebrows and mouth. Five normalized distances are defined on each skeleton:

- D_1: eye opening,
- D_2: distance between the interior corner of the eye and the interior corner of the eyebrow,
- D_3: mouth opening in width,
- D_4: mouth opening in height,

- D_5: distance between each corner of the mouth and the external corner of the corresponding eye.

A fusion process involving all the distances D_i yields to a classification in term of a single expression or in term of a mixture of expressions. A state of doubt between two expressions (the system is sure that it is one of the two expressions but is not able to know which one) can appear. In order to solve the doubt state, a post processing based on the analysis of the presence or the absence of transient wrinkles in the nasal root and based on the mouth shape is added. Transient wrinkles in the nasal root appear in case of *disgust* and *anger* [14], while the mouth has a very typical shape for *disgust*, *joy* and *surprise* [14]. Since the extraction of these two indices is not always reliable, they are not directly used for the transferable belief model but to solve the doubt between two expressions.

The presence or absence of wrinkles in the nasal root (Fig 2.a) is detected by using a Canny edge detector [15]. We compare the number of detected edge points in the nasal root for a reference frame (a frame with a *neutral* expression i.e without any wrinkles in the nasal root) with the number of edge points in the nasal root for the current frame. If there are almost twice more edge points in the current frame nasal root than in the reference frame nasal root, the presence of transient wrinkles is validated. In addition to wrinkles information, mouth shape can be used (Fig.2.b, 2.c). According to the expression, the ratio between length and width of the mouth is larger or smaller than its corresponding value in the *neutral* expression.

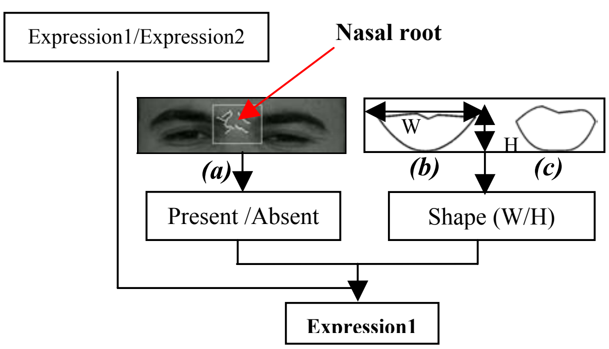

Fig 2 - Post-processing; (a) wrinkles in the nasal root; examples of mouth shapes in case of: (b) *joy*, (c) *disgust*.

4. Rules based expert system

The rules are defined by an expertise of the skeleton distances transformations for each expression w.r.t to the same distances for the *neutral* expression. The *neutral* expression represents the reference expression.

4.1 Definition of the symbolic states associated to the measures

The analysis of the numerical values of all the distances D_i for the 4 expressions contained in an expertise database shows that, for each of the 4 facial expressions, each D_i can be either higher, either lower either equal to its corresponding distance for the neutral expression. We associate to each distance D_i one of the three possible following symbolic states:

- The neutral state S if the current value D_i is close to its value for the *neutral* expression;
- A C^+ state if the current value D_i is significantly higher than its value for the *neutral* expression;
- A state C^- if the current value D_i is significantly lower than its value for the *neutral* expression.

This yields to 3 symbolic states, $\{S, C^+, C^-\}$ to be identified for each distance D_i for all the expressions.

The analysis of the evolution of the D_i curves on the expertise database shows that a similar evolution can be associated to each distance D_i for a given emotion whatever the subject. Such a study was impossible for the expressions (*fear, sadness, anger*) because of a lack of significant data.

The graph of Fig.3 presents the evolution of D_2 (distance between the interior corner of the eye and the interior corner of the eyebrow) and D_5 (distance between one mouth corner and the external corner of the corresponding eye) for several persons and a given expression. In each case, the considered video sequence starts with a *neutral* expression, goes towards a given emotion and returns to the *neutral* expression.

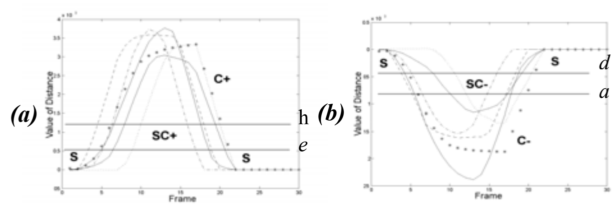

Fig.3 - Temporal evolution of distances D_2 in case of *surprise* (a) and D_5 in case of *joy* (b). Thresholds h, e, d and a are defined in §5.2.

The graph of Fig.3a shows that, D_2 systematically increases in case of *surprise* since people fully open the eyes. On Fig.3b, D_5 always decreases in the case of *joy*. Indeed, the mouth opens and thus mouth corners become closer to the eyes.

4.2 Expert system

The analysis of the states for the 5 distances associated to each of the 4 expressions (*joy, surprise, disgust* and *neutral*) allow us to exhibit for each expression a specific combination of these states. Table 1 shows the resulting states combinations. For example, in case of *joy* (E_1), the mouth is opening (C^+ state for D_3 and D_4), the corners of the mouth are going back toward the tears (C^- state for D_5). The eyebrows are slackened or the inner eyebrows are lowered (S/C^- state for D_2) and the eyes become slightly closed (C^- state for D_1).

Let's note that in some cases, two different states are possible for a given distance (see D_2 for *joy* or *fear*, D_4 for *anger* for example). This could produce a total doubt between two expressions as a result of the classification process. For example, the classifier is not always able to distinguish *disgust* and *joy* because both expressions could be described by the same combination of states in some cases.

The proposed combinations of symbolic states associated to each D_i for the 4 expressions (*joy, surprise, disgust and neutral*) are compared to the MPEG-4 description of the deformations undergone by facial features for such expressions [16]. As a result, we find that the proposed combinations are compliant with MPEG-4 description and give even some extensions.

Table 1 also gives the combination of D_i states to be associated to the 3 expressions *(fear*, *sadness* and *anger)*. These combinations result from the MPEG-4 description of the facial features deformations for such expressions.

The expression E_8 is added as the *unknown* expression or class of reject. It represents all the expressions, which do not correspond to any of the descriptions of Table 1.

	D_1	D_2	D_3	D_4	D_5
Joy E_1	C^-	S / C^-	C^+	C^+	C^-
Surprise E_2	C^+	C^+	C^-	C^+	C^+
Disgust E_3	C^-	C^-	S / C^+	C^+	S / C^-
Anger E_4	C^+	C^-	S	S / C^-	S
Sadness E_5	C^-	C^+	S	S	S
Fear E_6	C^+	S / C^+	S	S	S
Neutral E_7	S	S	S	S	S

Table 1 – Theoretical table of D_i states for each expression.

5. Fusion process using transferable belief functions

Since human expressions could be variable according to the individual, a logic system is not sufficient to make a reliable recognition of expression. The use of the transferable belief functions to model the doubt between states of the parameters as well as doubt between expressions is proposed.

5.1 The Transferable Belief Model

Initially introduced by Dempster [17], the belief theory was taken again by Shafer [18]. Based on this work Smets has enriched this theory and called it TBM (Transferable Belief Model) [19]. It requires the definition of a set $\Omega = \{E1, E2, \ldots, EN\}$ made up of N exclusive assumptions. In our application, the assumptions Ei correspond to the seven facial expressions : *joy* (E1), *surprise* (E2), *disgust* (E3), *sadness* (E4), *anger* (E5), *fear* (E6) and *neutral* (E7).

5.2 Modeling the measurements states

The modelling process aims at computing the state of every distance D_i and at associating a piece of evidence. Let define the basic belief assignment (BBA) as:

$$m_{Di}: \quad 2^{\Omega'} \rightarrow [0,1] \qquad (2)$$
$$B \rightarrow m_{Di}(B)$$

With $\Omega' = \{C^+, C^-, S\}$, $2^{\Omega} = \{C^+, C^-, S, SC^+, SC^-, C^+C^-, C^+SC^-\}$, where $S \cup C^+$ (noted SC^+) states the doubt between S and C^+, $S \cup C^-$ (noted SC^-) states the doubt between S and C^- and $m_{Di}(B)$ is the piece of evidence (PE) of each state B.

A numerical/symbolic conversion is carried out (Fig.3), which associates to each value of D_i one of the symbols of $2^{\Omega'}$. To carry out this conversion, a model is defined for each distance in using the states of $2^{\Omega'}$ (Fig.4). The symbols C^+C^- and C^+SC^- are impossible.

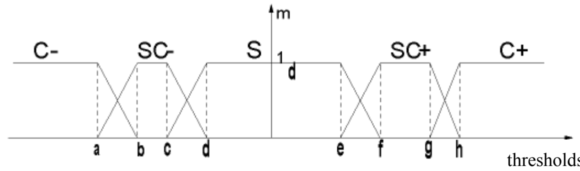

Fig.4 - Model.

m is the piece of evidence (PE) associated to each possible state in 2^{Ω} and the thresholds *(a... h)* are the limit values of D_i corresponding to each state or subset of states.

Thresholds have been evaluated by the analysis of a set of curves similar to those presented on Fig.3.

Thresholds for the states (C^+, C^- and S)

For each distance D_i, the 2^{Ω} threshold ***h*** of the state C^+ (Fig.4) corresponds to the means of the maximum values of *Di* for all the subjects and all expressions of the expertise database. Similarly, the threshold ***a*** of the state C^- (Fig.4) corresponds to the mean of the minimal values of D_i for all the subjects and all the expressions.

On the other hand, the threshold ***d*** and ***e*** of the state S (Fig.4) are calculated for all images of *neutral* expression. For each distance D_i, the threshold ***d*** corresponds to the mean of the maximum values of D_i for all the images in a *neutral* expression and ***e*** corresponds to the mean of the minimal values of D_i for the same images.

Thresholds for states *(SC^+ and SC^-)*

The states SC^+ and SC^- are associated to the states of doubt when the value of D_i is higher than the one of *neutral* state S but is not enough great to be in the state C^+ nor enough small to be in the state C^-.

The median of the maximum values of each distance for all the subjects and all the expressions of the expertise database is computed. The thresholds of the intermediate states (***f***, ***g***, ***b*** and ***c***) are defined by [mean+/-median] of each state (C^+, C^-, S). Once the thresholds have been defined, the numerical/symbolic conversion can be done. This consists in assigning to each distance a state with a PE according to the value of measurement.

5.3 Data fusion process

We have several sources of information (D_i) to which we associate PEs. Our goal is to obtain a PE, which takes into account all the available information. The BBA is obtained using the rule of conjunctive combination or orthogonal sum. In the case of two distances D_1 and D_2, the orthogonal sum is defined in the following way:

$$m = m_{D_1} \oplus m_{D_2} \qquad (3)$$

$$m(A) = \sum_{B \cap C = A} m_{D1}(B) m_{D2}(C) \qquad (4)$$

Where A, B and C are expressions or subsets of expressions. This leads to obtain propositions whose number of elements is lower than the initial ones and to associate them a piece of evidence. Indeed A is a subset of *B* and *C* because $A = B \cap C$. The final PE is thus more accurate.

In a more explicit way, if one takes two basic belief assignments:

$$\begin{array}{ll} m_{D1}(E_1 \cup E_3) & m_{D2}(E_1) \\ m_{D1}(E_1) & m_{D2}(E_2) \\ m_{D1}(E_2) & m_{D2}(E_1 \cup E_2) \end{array}$$

their combination gives the results of Table 2.

D_1 / D_2	E_1	E_2	$E_1 \cup E_3$
$E_2 \cup E_3$	$\varnothing$	E_2	E_3
E_1	E_1	$\varnothing$	E_1
E_2	$\varnothing$	E_2	$\varnothing$

Table 2 - Example of combination of PEs of two distances. $\varnothing$ is the empty set.

The peace of evidence of each expression by the combination of results of the two distances is calculated by:

$m_{D12}(E_1) = m_{D1}(E_1).\ m_{D2}(E_1) + m_{D1}(E_1)\ m_{D2}(E_1 \cup E_3)$
$m_{D12}(E_2) = m_{D1}(E_2 \cup E_3).\ m_{D2}(E_2) + m_{D1}(E_2).m_{D2}(E_2)$.
$m_{D12}(E_3) = m_{D1}(E_2 \cup E_3).m_{D2}(E_1 \cup E_3)$.
$m_{D12}(\varnothing) = m_{D1}(E_2 \cup E_3).m_{D2}(E_1) + m_{D1}(E_1).m_{D2}(E_2) + m_{D1}(E_2).m_{D2}(E_1) + m_{D1}(E_2).\ m_{D2}(E_1 \cup E_3)$.

Conflicts can appear in case of incoherence sources noted $\varnothing$. In the scope of the presented application, the conflict corresponds to a configuration of distance states, which does not appear in Table 1. It comes from the fact that Ω is not exhaustive. The added expression *unknown* or class of reject E_8 represents all these conflict states (Table 2).

The decision is the ultimate step of the data processing sequence. It consists in making a choice between various assumptions E_i and their possible combinations. Making a choice means taking a risk, except if the result of the combination is perfectly reliable: $m(E_i) = 1$. As it is not always the case, instead of using the plausibility which favors the single hypothesis in the case of mixture of expressions, the chosen assumption is the one with the maximum value of PE.

6. Results

6.1 Expertise database

The recognition system has been defined to recognize the 7 basic universal emotions but it has been fully tested only for the 4 expressions (*joy, surprise, disgust* and *neutral)*. For the expertise step, 1170 frames (13 subjects and 4 expressions) have been considered. All the frames of the expertise database are segmented and the 5 distances defined on Fig.1 are computed and used in order to define the thresholds of §5.2 and to establish Table 1.

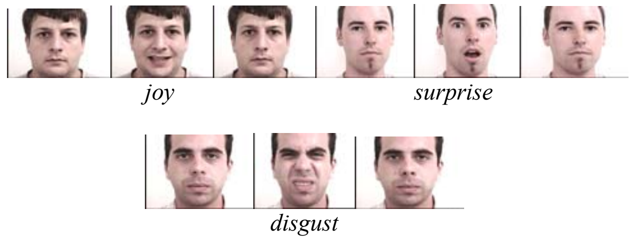

Fig.5 - Examples of expressions.
Each record starts and finishes by a neutral state. The sequences have been acquired during 5s at 25 images/second.

6.2 Test database

In order to evaluate the robustness to different variations (gender, ethnicity, difference of expressions…) the system is tested on three test databases: Hammal-Caplier test database [20] (630 frames for 7 subjects and 4 expressions), Cohn-Kanade database [21] (144 frames for the 4 expressions) and Cottrel database [22] (24 frames for 4 expressions). In these two last databases we only have two images for each expression: the neutral state and the caricatured expression itself.

6.3 Classification rates

Table 3 presents the classification rates for the frames of the Hammal-Caplier test database. The right expressions are given in column and the expressions recognized by the system correspond to the lines. Expressions E_1 (*joy*) and E_7 (*neutral*) yield to good classification rates. On the contrary, the classification rate E_3 *(disgust*) is lower. This is due to individual variability (Fig.6.a) and to the difficulty for a non-actor people to simulate this expression (Fig.6.b). For E_1, there is a high rate of total doubt between E_1 and E_3 the system is sure that it is one of the two expressions but is not able to know which one. This has to be related to the definition of Table 1 with two possible different states for a given distance.

Another state of confusion is the one between E_2 (*surprise*) and E_6 (*fear*). Both expressions are difficult to distinguish using only the information of contours of the facial features. These expressions are also difficult to be distinguished by human observers (Fig.7.a). It is thus preferable that the system keeps this doubt instead of taking the risk of making a wrong decision.

In the Hammal-Caplier database, the *unknown* state often appears for intermediate frames where the person is neither in a *neutral* state, nor in a particular expression (Fig.7.b).

The row *others* of the Table 3 gathers the states of mixture of more than two expressions.

Syst\Exp	E_1	E_2	E_3	E_7
E_1 joy	76,36%	0	9,48	3%
E_2 surprise	0	12%	0	0
E_3 disgust	0	0	43,10%	2%
$E_1 \cup E_3$	10,90%	0	8,62%	0
$E_2 \cup E_6$	0	72,44%	0	0
E_7 neutral	6,66%	0,78%	15,51	88%
E_8unknown	6,06%	11,8%	12,06%	0
others	0.02%	2.08%	11.32%	7%
Total	**87,26%**	**84,44%**	**51,72%**	**88%**

Table 3 - Classification rates on the Hammal-Caplier database.

In order to be able to choose between *joy* and *disgust* in case of doubt, we add a post-processing state, which takes into account information about transient features and mouth shape (§2). Nasal root wrinkles (Fig2) are characteristic for *disgust*. This is used to solve the problem of doubt between *joy* and *disgust*. In the case of absence of transient features, we use the ratio between length and width of the mouth (Fig.2). Our analysis shows that this ratio is larger than its value for the *neutral* expression in the case of *joy* and lower in the case of *disgust*. With the proposed post-processing step to make a distinction between *joy* and *disgust* in case of doubt, the recognition rate for E_1 (*joy*) increases by 15% and $E_1 \cup E_3$ (*joy-disgust*) decreases by 17% (2% of false detection of *disgust*). We increase by 19% for E_3 (*disgust*) and $E_1 \cup E_3$ (*joy-disgust*) decreases by 11% (5% of false detection of *joy*).

(a) (b)

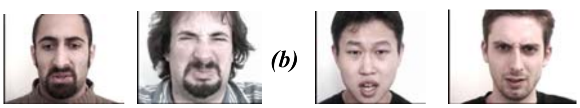

Fig.6 - Examples of *disgust* expressions. (a): individual variability; (b): poor simulations.

(a) (b)

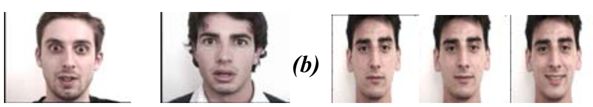

Fig.7 - (a): Examples of mixture between *fear* and *surprise* during a sequence of a simulated *surprise*. (b): Example of *unknown* state: 3 consecutive frames from *neutral* to *joy* during a sequence of a simulated *joy*.

Given the fact that the state of doubt *joy-disgust* is related to the rules defined in the Table 1 and that the state of doubt *surprise-fear* is related to the fact that *surprise* is a mixture of *fear* and *surprise,* they are not due to classification errors of the proposed system. It is thus possible to consider them as a good classification and to associate them to the corresponding expression, which allows us to add their respecting rates leading to the results of the last row of Table 3.

Table 4 (on the right of E_7 column) presents the results obtained on frames of the Cohn-Kanade database. 30 frames have been chosen for *joy*, 25 for *surprise* and 17 for *disgust*. The classification rates for this database are comparable with those of Table 3.

In Table 4 (on the left of E_7 column) are presented the classification rates obtained on the Cottrel database. In the same way, associating the expression and the corresponding mixture of expressions, the system gives good classification rates.

Syst\Exp	E_1	E_2	E_3	E_7	E_1	E_2	E_3
E_1joy	64,51%	0	0	0	62,50%	0	0
E_2surprise	0	16%	0	0	0	25%	0
E_3disgust	0	0	52,94%	0	0	0	75%
$E_1 \cup E_3$	32,25%	0	47,05%	0	37,50%	0	0
$E_2 \cup E_6$	0	84%	0	0	0	75%	0
E_7neutral	0	0	0	0	0	0	0
E_8unkno	3,22%	0	0	0	0	0	25%
others	0	0	0.01%	0	0	0	0
Total	**96,76**	**100%**	**99,99%**	**100%**	**100%**	**100%**	**75%**

Table 4 - Classification rates: on the right the Cohn-Kanade database, on the left the Cottrel database. The column E_7 is the same for both databases.

Fig.8 presents examples of classification of the 4 expressions (*disgust*, *joy, surprise* and *neutral*). The results are presented in three columns per line, each column corresponds to *neutral* state *(a),* beginning of the expression *(b)* and the apex of the expression *(c).*
Rows 1 and 2 of Fig.8 show the ability of the system to recognize *disgust* at different intensities.

Fig.8.b.1 shows the *unknown* state that corresponds to intermediate state between *neutral* and *disgust* expression. Rows 3 and 4 of Fig.8 shows some results in case of *joy* expression. Rows 5 and 6 of Fig.8 shows results of classification in case of simulated surprise.

Row 5 of Fig.8 shows the difficulty of the system to separate *surprise* and *fear*. However the system is sure at 100% that it is one of the both expressions and not any other. This incapacity to distinguish these two expressions is confirmed by human observer. The only manner to separate them is to add information about the context or information about another modality such as speech signal for example.

Fig.9 presents some other results obtained of the two other considered database.

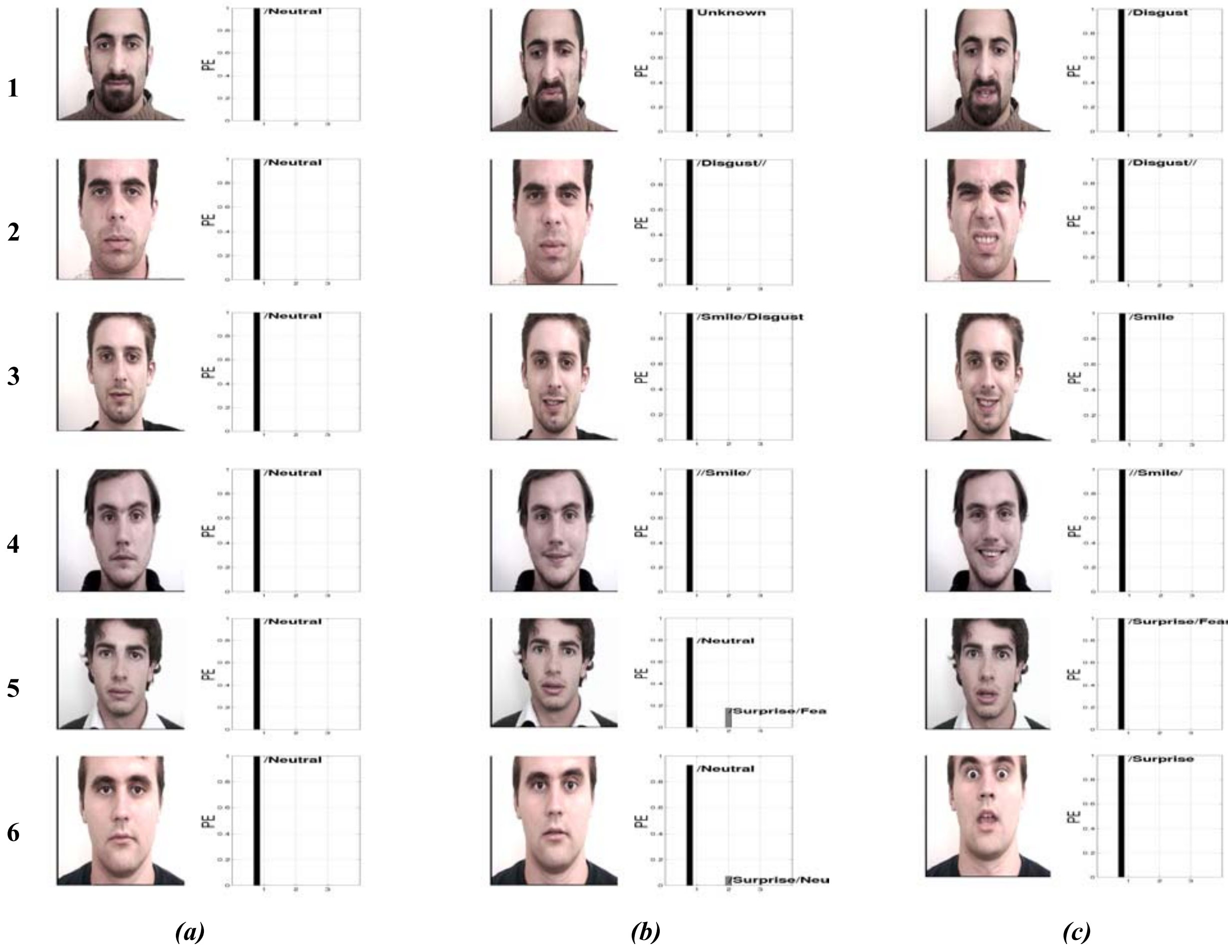

Fig.8 Examples of expressions ; (a) : *neutral* expression, (b) : intermediate state of expression and (c) : apex of *expression.*
The bar graph presents the recognized expression and its associated piece of evidence.

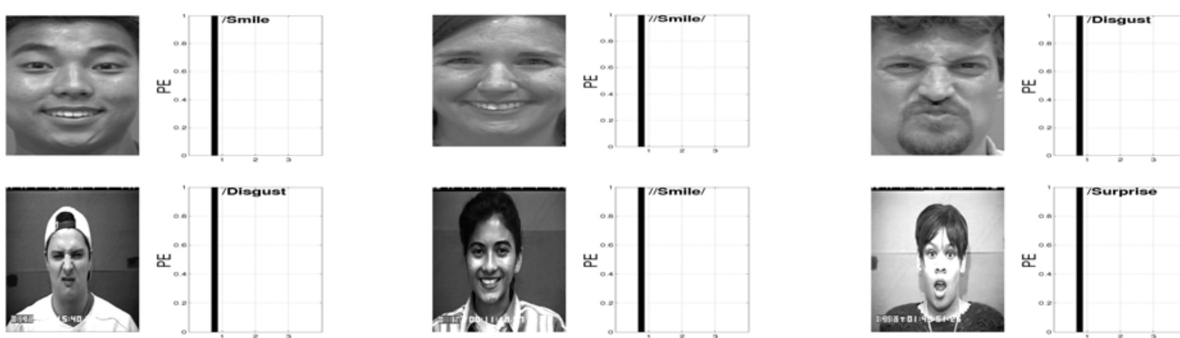

Fig.9 Other examples of expressions classification : first row, images from the Cottrel database; second row, images from the Cohn-Kanade database .
The bar graph presents the recognized expression and its associated piece of evidence.

7 Conclusion

We present a method for the classification of facial expressions based on the analysis of characteristic distances computed on skeleton of expression. The use of the belief theory with the currently definite base of rules makes it possible to emphasize the fact that certain expressions are not always dissociable (*joy and disgust, surprise and fear*). This result is directly related to the modeling of each one of these expressions: knowledge is not sufficient to be able to differentiate them. But, this result can be interesting for man-machine interactions. Indeed, in case of doubt, it will be preferable to consider that the two expressions are possible rather than to choose one of both by having a considerable risk to be mistaken. To improve these results, we can increase the number and the quality of measurements, by taking into account the explicit information about the forms of contours of the skeletons of expression in addition to the characteristic distances and by taking into account the temporal evolution of measurements.

References

[1] N.Eveno, A. Caplier, and P.Y. Coulon "Automatic and Accurate Lip Tracking". *IEEE Trans. On Circuits and Systems for Video Technology, vol.14, pages 706-715 May 2004*.

[2] Z. Hammal, A. Caplier "Eye and Eyebrow Parametric Models for Automatic Segmentation".*IEEE Southwest Symposium on Image Analysis and Interpretation, Lake Tahoe, Nevada, 2004.*

[3] http://www.similar.cc/

[4] A. Mehrabian, "Communication without words", *Psychology* , *vol 4, pages 53-56, 1968*.

[5] P. Ekman, "Facial Expression", *The Handbook of Cognition and Emotion, T. Dalgeleish and M. Power, John Wiley & Sons Ltd. 1999.*

[6] J.N. Bassili. "Facial motion in the perception of faces and of emotional expression". *Journal of Experimental Psychology, vol.4, pages 373-379, 1978.*

[7] J. Cohn, A.J. Zlochower, J.J. Lien , T. Kanade, "Feature-point tracking by optical flow discriminates subtles differences in facial expression". *IEEE International Conference on automatic Face and Gesture Recognition, N°3, pages 396-401, 1998.*

[8] B.D. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision" , *Seventh International Joint Conference on Artificial Intelligence, pages 674-680, 1981*.

[9] M. Pantic, L.J..M Rothkrantz "Expert system for automatic analysis of facial expressions". *ElSEVIER Image and Vision Computing, vol.18, pages 881-905, 2000.*

[10] I. Cohen, F.G. Cozman, N. Sebe, M.C. Cirelo and T.S. Huang "Semisupervised Learning of Classifiers : Theory, Algorithms and their Application to Human-Computer Interaction", *IEEE Trans. On PAMI, vol.26, pages 1553-1567, 2004.*

[11] P. Ekman, W.V. Freisen , "Facial Action Coding System (FACS). *Manual, Consulting Psychologists Press, Palo Alto, 1978.*

[12] P. Smets, Handbook of Defeasible Reasoning and Uncertainty Management Systems: "The transferable belief model for quantified belief representation", v*ol. 1, pages 267-301, publisher Kluwer, Doordrecht, The Netherlands 1998.*

[13] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut, "Facial Expressions Recognition Based on The Belief Theory: Comparison with Different Classifiers", *in Proc. 13-th International Conference on Image Analysis and Processing, ICIAP, 2005*.

[14] M.L Smith, G.W. Cottrell, F. Gosselin, P. Schyns, "Transmitting and decoding facial expressions of emotion. *Vision Science Society Meeting, Sarasota, Florida, 2004.* .

[15] J. Canny, "A Computational Approach to Edge Detection", *IEEE Trans. PAMI, vol.8, N°6, pages 679-698, June, 1986.*

[16] M. Malciu, F. Preteux, "MPEG-4 compliant tracking of facial features in video sequences" *Proceedings EUROIMAGE International Conference on Augmented, Virtual Enviroments and Three-Dimensional Imaging (ICAV3D'01), Mykonos, Greece, pages 108-111, May, 2001.*

[17] A. Dempster. "A generalization of Bayesian inference". *Journal of the royal statistical society, vol.30, pages 205-245, 1968.*

[18] G. Shafer. "A Mathematical Theory of Evidence". *Princeton University Press, 1976.*

[19] PH. Smets "Data Fusion in the Transferable Belief Model"." *Proc. 3rd Intern. Conf. Information Fusion, Paris, France pages 21-33, 2000.*

[20] http://www.lis.inpg.fr/pages_perso/hammal/index.htm.

[21] http://www-2.cs.cmu.edu/~face.

[22] M. Dailey, G.W. Cottrell, & J. Reilly, "California Facial Expressions", *CAFE, unpublished digital images, UCSD Computer Science and Engineering Department, 2001.*