DYNAMIC FACIAL EXPRESSION UNDERSTANDING BASED ON TEMPORAL MODELLING OF TRANSFERABLE BELIEF MODEL

Zakia Hammal
Laboratory of Images and Signals
46, Avenue Felix Viallet, Grenoble, France

Keywords: facial expression classification, dynamic modelling, transferable belief model, facial feature behavior.

Abstract: In the present contribution a novel approach for dynamic facial expressions classification is presented and discussed. The presented approach is based on the use of the Transferable Belief Model applied to static facial expression classification studied in previous developments. The system is able to recognize pure facial expressions, i.e., *Joy*, *Surprise*, *Disgust* and *Neutral* as well as their mixtures. Additionally, this approach is able to deal with all facial feature configurations that does not correspond to any of the cited expression, i.e., *Unknown* expressions. The major improvement of this former work consists in the introduction of the temporal evolution of the facial feature behavior. Initially, the temporal information is introduced to improve the robustness of the frame-by-frame classification by the correction of errors due to the automatic segmentation process. In addition since a facial expression is the result of a dynamic and progressive combination of facial features behavior, which is not always synchronous, a frame-by-frame classification is not sufficient. To overcome this constraint, we propose the introduction of the temporal information inside the TBM fusion framework. The recognition is accomplished by combining all facial feature behaviors between the *beginning* and the *end* of an expression sequence independently to their chronological order. Then the final decision is taken on the whole sequence and consequently, the recognition becomes more robust and accurate. Experimental results on the Hammal_Caplier database demonstrate the improvement on the frame-by-frame classification and the ability to recognize entire facial expression sequences.

1 INTRODUCTION

Developing automatic and accurate facial expression recognition systems is a challenging issue due to their applications in many fields like medicine, security or human-machine interaction. In computer vision, significant amount of research on facial expression classification have led to many systems based on different approaches. They can be divided into three categories: optical flow analysis from facial actions (Yacoob and Davis, 1996), (Essa and Pentland, 1997), model based techniques (Zhang et al., 1998), and finally, feature based methods (Pantic and Rothkrantz, 2000), (Tian et al., 2001). In our opinion, these approaches present two limitations. A first limitation is that these methods perform a classification of the examined expression into one of the basic emotion categories proposed by Ekman and Friesen (Ekman and Friesen, 1978). However, ”binary” and ”pure” facial expressions are rarely found, i.e., generally speaking, people show mixtures of facial expressions. A second limitation is related to the recognition, which is only based on static information without considering the temporal behaviors. Indeed, Bassili (Bassili, 1978) shows that facial expressions can be characterized using some points defined on the permanent facial features and that the temporal evolution of these points yields more accurate results than from single static images. In this work we propose a synergetic combination of techniques, which are intended to overcome some of the limitations of the existing approaches, e.g., in our previous works (Hammal et al., 2005a) we have proposed a system based on the Transferable Belief Model that overcomes the first limitation. This fusion framework allows to recognize mixtures of expression as well as an *Unknown* state which corresponds to all the facial features configurations which do not belong to the set of studied expressions (*Joy, Surprise, Disgust and Neutral*).
To overcome the second limitation Zhang *et al* (Zhang and Qiang, 2005) propose a system which takes into account the temporal information but it does not ex-

Hammal Z. (2006).
DYNAMIC FACIAL EXPRESSION UNDERSTANDING BASED ON TEMPORAL MODELLING OF TRANSFERABLE BELIEF MODEL.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 93-100
DOI: 10.5220/0001377600930100

plicitly fuse the dynamic facial features behavior. They only takes into account the information provided by the last frame in the classification process. In this work we present an extension of our previous development with the introduction of the temporal information. This paper is articulated as follows. In section 2 and 3, we present a brief description of our previous system as well as the Transferable Belief Model. In section 4 we propose a model of temporal analysis incorporated into the Transferable Belief Model. In section 5 this model is used to overcome errors due to the automatic facial feature segmentation process and consequently to improve the frame-by-frame facial expression classification. Section 6 presents the fusion of all the facial feature behaviors in order to obtain a dynamic recognition of facial expressions sequences. Finally, in section 7 we present experimental results on facial expression sequences with a detailed description of the facial features behaviors.

2 PREVIOUS WORK

In this section, we describe how a video sequence of face images is analyzed for the recognition of facial expressions (see (Hammal et al., 2005a) for a complete description). In order to recognize the four expressions *Joy, Surprise, Disgust, Neutral* and additionally, the *Unknown* expressions. First, the contours of facial features are automatically extracted in every frame using the algorithms described in (Hammal et al., 2006) (see Figure.1.left). Then, based on the segmentation results, five characteristic distances (D1, D2, D3, D4 and D5) are defined and estimated as shown in Figure.1.right

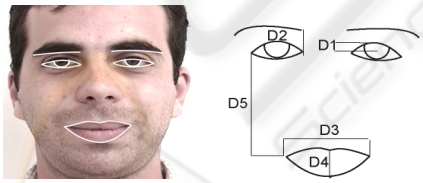

Figure 1: left: Facial features segmentation; right: facial skeleton and characteristic distances.

Finally, five symbolic states are associated with the distances, which are: 1) S if the current distance is roughly equal to its corresponding value in the *Neutral* expression, 2) C^+ (resp.C^-) if the current distance is significantly higher (resp. lower) than its corresponding value in the *Neutral* expression, 3) SC^+ (resp. SC^-) if the current distance is neither sufficiently higher (resp. lower) to be in C^+ (resp. C^-) and neither sufficiently stable to be in S, 4) SC^+ and 5) SC^-, which correspond to a doubt states respectively. The numerical/symbolic conversion was carried out using a model for each distance. Finally a fusion process, based on Transferable Belief Model (Smets, 1998) has been used for the classification of facial expressions based on static information. The suitability of this model for facial expression classification has been proved in comparison with two classical classifiers (HMM's and Bayesian model) (Hammal et al., 2005b). As mentioned before, the major improvement on this former work consists in the introduction of the temporal information in order to take into account the evolution of the facial features behavior during a facial expression sequence.

3 TRANSFERABLE BELIEF MODEL AND MODELLING PROCESS

Originally introduced by Dempster and Shafer and revisited by Smets (Smets, 1998), the Transferable Belief Model (TBM) can be seen as a generalization of the probability theory. It requires the definition of a set $\Omega = \{E_1, E_2, ..., E_N\}$ of N exclusive and exhaustive assumptions. We also consider the power set 2^Ω that denotes the set of all subsets of Ω. To each element A of 2^Ω is associated an elementary piece of evidence $m(A)$ which indicates all the confidence that one can have in this proposal. The function m is defined as:

$$m : 2^\Omega \rightarrow [0,1] \quad A \rightarrow m(A), \sum_{A \subseteq 2^\Omega} m(A) = 1 \tag{1}$$

In our application, the assumptions E_i correspond to the four studied facial expressions: *Joy* (E_1), *Surprise* (E_2), *Disgust* (E_3) and *Neutral* (E_4); 2^Ω corresponds to single expressions or combinations of expressions, that is, $2^\Omega = \{E_1, E_2, E_3, ..., E_1E_2, E_2E_3, ...\}$, and A is one of its elements.

Modelling process. The facial expression classification is based on the TBM fusion process of all the characteristic distances states. The modelling process aims at computing the state of every distance Di and at associating it a piece of evidence. Let's define the basic belief assignment (BBA) m for each Di as:

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

With $2^{\Omega'} = \{C^+, C^-, S, SC^+, SC^-\}$. This gives a set of pieces of evidence for all the characteristic distances. Then a set of rules (defined in table 3) is applied to combine the pieces of evidence of all distances states in order to obtain the pieces of evidence of each facial expression.

4 CONDITIONAL PIECES OF EVIDENCE AND PREDICTED BBA

Based on automatic facial feature segmentation, the static classification results show that some miss classifications are due to wrong distances states estimation caused by segmentation errors. The segmentation leads to two type of errors: motion intensity: the induced distance value is much higher than its real value; motion direction: the distance evolution is in the opposite of its real evolution (C^- instead of C^+). To solve the motion intensity errors, a smoothing process is applied on the values of the characteristic distances states between each two consecutive frames. The smoothing prevents casual jumps corresponding to a higher distance variation than the mean plus/minus the standard deviation of the corresponding distance variation (see Figure 2). These values have been learned after a manual segmentation results obtained on Hammal_Caplier database [1]. Figure 2 presents a smoothing results of the values of two characteristic distances states (D5 and D1) during a *Joy* and *Surprise* sequences. The second segmentation errors correspond to the

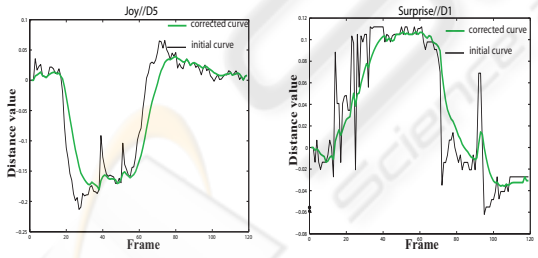

Figure 2: Effect of the smoothing process on the distances states variation of D5 and D1.

motion direction of the characteristic distances. Having no mean to automatically recognize this kind of errors at the segmentation level, a higher information level has to be used to correct them. For this, a temporal constraint on the basic belief assignment of each characteristic distances states is introduced based on conditional pieces of evidence.

Conditional pieces of evidence. Based on the basic belief assignment of the characteristic distances states obtained at each time, the pieces of evidence corresponding to the transitions from each state k at time $t-1$ (previous frame) to each one of the other possible states r at time t (current frame) are computed on the results of a manual segmentation of Hammal_Caplier training database (ham,) where each expression is expressed with different speed evolution (for example we have slow *Joy* as well as quick *Joy*). This computation leads to the definition of a transition matrix M_{trans} whose elements correspond to the conditional pieces of evidence noted $m(r/k)$ (see table 1).

Table 1: Transition Matrix; the elements are the conditional pieces of evidence for S and C^+ (similarly for C^-,SC^+ and SC^-).

	t				
$t-1$	S	C^+	.	.	.
S	$m(s/s)$	$m(c^+/s)$	.	.	.
C^+	$m(s/c^+)$	$m(c^+/c^+)$	.	.	.
C^-	$m(s/c^-)$	$m(c^+/c^-)$	.	.	.
SC^+	$m(s/sc^+)$	$m(c^+/sc^+)$	.	.	.
SC^-	$m(s/sc^-)$	$m(c^+/sc^-)$	.	.	.

In the scope of our application we have a set of five states $\Omega' = \{C^+, C^-, S, SC^+, SC^-\}$. The associated set of conditional pieces of evidence is reported in Table 1 and corresponds to the elements of the transition matrix. The conditional pieces of evidence associated with the distances states are computed as:

$$\begin{pmatrix} m(S) \\ m(C^+) \\ m(C^-) \\ m(SC^+) \\ m(SC^-) \end{pmatrix}_t = M_{trans} * \begin{pmatrix} m(S) \\ m(C^+) \\ m(C^-) \\ m(SC^+) \\ m(SC^-) \end{pmatrix}_{t-1} \quad (3)$$

$$M_t = M_{trans} * M_{t-1} \quad (4)$$

The transition matrix is learned on our training database. Once defined, it prevents motion direction errors (see section 4) resulting of segmentation errors. However, in real expression sequences subjects tend to remain longer in the same singleton state than in doubt states. This leads to overweighted transitions for the diagonal elements of M_{trans} (for example $m(C^+/C^+)$ is more than 90%). To avoid this prediction bias, a *discounting* is applied to the

[1] Hammal_Caplier database is used for our experiments (19 subjects and 4 expressions *Joy, Surprise, Disgust and Neutral*) (ham,). 11 subjects are used for the training and 8 subjects for the test. Each video recording starts with neutral state, reaches the expression and comes back to the neutral state. The sequences have been acquired during 5 seconds at 25 images/second.

transition matrix. This allows to remove incorrect transitions while preventing the overweighting of other transitions. This process avoids the definition of a specific evolution model and as a consequence allows to handle with free facial features motion.

Discounting. The *discounting* consists in defining a parameter $\alpha \in [0, 1]$ which allows to compute the new pieces of evidence of each state A_i as:

$$m^{'}(A_i) = \alpha m(A_i) \quad (5)$$

$$m^{'}(A_i \cup \overline{A_i}) = 1 - \alpha(1 - m(A_i \cup \overline{A_i})) \quad (6)$$

In the scope of our application it consists in discounting the diagonal pieces of evidence of the transition matrix M_{trans} and to redistribute the subtracted value only on the doubt states. The other transition states do not change. For example for the first row of M_{trans} the discounting consists in computing:

$$m^{'}(S/S) = \alpha m(S/S)$$
$$m^{'}(S/SC^{+}) = 1 - \alpha(1 - m(S/SC^{+}))$$
$$m^{'}(S/SC^{-}) = 1 - \alpha(1 - m(S/SC^{-}))$$

Predicted basic belief assignment. The predicted (*pred*) basic belief assignment at each time t is computed by the combination of the transition matrix and the basic belief assignment computed (*comp*) at time $t-1$ in the following way:

$$M_{pred,t} = M_{trans} M_{comp,t-1} \quad (7)$$

Where $M_{pred,t} = \{m(C^{+}),\ m(C^{-}),\ m(S),\ m(SC^{+}),\ m(SC^{-})\}_t$ is the predicted basic belief assignment of the characteristic distances states at time t, M_{trans} the transition matrix and $M_{comp,t-1} = \{m(C^{+}),\ m(C^{-}),\ m(S),\ m(SC^{+}),\ m(SC^{-})\}_{t-1}$ the computed basic belief assignment at time $t-1$.

Table 2: Confusion Matrix: combination between the pieces of evidence of the predicted and of the estimated states.

	$Comp$				
$Pred$	S	C^{+}	C^{-}	SC^{+}	SC^{-}
S	S	ϕ	ϕ	S	S
C^{+}	ϕ	C^{+}	ϕ	C^{+}	ϕ
C^{-}	ϕ	ϕ	C^{-}	ϕ	C^{-}
SC^{+}	S	C^{+}	ϕ	SC^{+}	S
SC^{-}	S	ϕ	C^{-}	S	SC^{-}

Table 3: Theoretical table of D_i states for each expression.

	D_1	D_2	D_3	D_4	D_5
$JoyE_1$	C^{-}	S/C^{-}	C^{+}	C^{+}	C^{-}
$SurpriseE_2$	C^{+}	C^{+}	C^{-}	C^{+}	C^{+}
$DisgustE_3$	C^{-}	C^{-}	S/C^{-}	C^{+}	S/C^{+}
$NeutralE_4$	S	S	S	S	S

5 COMBINATION OF THE COMPUTED AND PREDICTED PIECES OF EVIDENCE

At each time t, the dynamic correction of the pieces of evidence associated with each characteristic distance state combines two processes: computation of the BBA of the characteristic distances states at time t and their temporal prediction based on the previous frame at time $t-1$ using the transition matrix. This combination (*comb*) allows to overcome the segmentation errors corresponding to motion direction as described in section 4 and, as a consequence, to improve the classification results. The new basic belief assignment $M_{comb,t}$ is defined by the combination of the predicted $M_{pred,t}$ and the computed $M_{comp,t}$ basic belief assignments of the characteristic distances states. The combination is based on the *conjunctive combination* (orthogonal sum) as:

$$M_{comb,t}(A) = M_{pred,t}(B) \otimes M_{comp,t}(C) = \sum_{B \cap C = A} M_{pred,t}(B).M_{comp,t}(C) \quad (8)$$

where A, B and C are characteristic distances states.

Then at each time t the confusion matrix, described in Table 2, is applied to define the news pieces of evidence of all the states using the conjunctive combination between the predicted and the computed pieces of evidence. For example for C^{+} the combination is made as:

$m_{pred,comp}(C^{+}) = m_{pred}(C^{+})m_{comp}(C^{+}) + m_{pred}(C^{+})m_{comp}(SC^{+}) + m_{pred}(SC^{+})m_{comp}(C^{+})$

The combination leads sometimes to a conflict, noted ϕ, between the predicted and the computed pieces of evidence ($m_{pred,comp}(\phi) \neq 0$). This is mainly due to the fact that the prediction follows a (learned) model that can fail in an atypical case of facial feature behavior. So in this case, the computed results are chosen to form the basic belief assignment associated with the distances states at time t (ie. we choose the results obtained from the segmentation process). This choice allows to handle better free facial features deformations.

Table 4: Rules table for the chosen states inside an increasing window Δ_t (/ : not used).

	$\sum_{\Delta t} C^+$	$\sum_{\Delta t} C^-$	$\sum_{\Delta t} SC^+$	$\sum_{\Delta t} SC^-$	$\underset{\Delta t}{PL}(C^+) - \underset{\Delta t}{PL}(C^-)$	$\underset{\Delta t}{PL}(SC^+) - \underset{\Delta t}{PL}(SC^-)$
C^-	0	> 0	/	/	/	/
	> 0	> 0	/	/	< 0	/
C^+	> 0	0	/	/	/	/
	> 0	> 0	/	/	> 0	/
SC^+	0	0	0	> 0	/	/
	0	0	> 0	> 0	/	> 0
SC^-	0	0	> 0	> 0	/	/
	0	0	> 0	> 0	/	< 0

Facial expressions classification. We combine the new basic belief assignment of all the characteristic distances states to classify the facial expressions. Table 3 describes a specific combination of the characteristic distances states for each of the studied facial expression. It has been defined by human expertise (Hammal et al., 2005a) but also corresponds to the MPEG-4 norm for facial expressions description. These rules are then used for the classification process. The piece of evidence of each facial expression corresponds to the combination (*orthogonal sum*) of the pieces of evidence of the associated distances states (Hammal et al., 2005a). Conflicts can appear in case of incoherent sources. In the scope of our work, it corresponds to all the configurations of distances states which do not appear in Table 3. The added expression *Unknown* E_5 represents all these conflict states. Figure 3 presents an example of the frame-by-frame classification results on a *Joy* sequence. We can see the sensitivity of the proposed method to the evolution of the facial features behaviors under different intensities. We give selected frames to show our results. In each frame we have two plots: the current image on the left and the corresponding expressions and their pieces of evidence on the right. The first frame corresponds to the second frame of the sequence. It is recognized as *Neutral* expression at 100%. The second frame corresponds to a transition state when the subject is neither in the *Neutral* state, neither in the studied expression. The system classifies it as 77% *Neutral* and 23% *Unknown*. In the third frame the system starts to recognize the expression. Finally the last frame corresponds to the apex of the *Joy* expression.

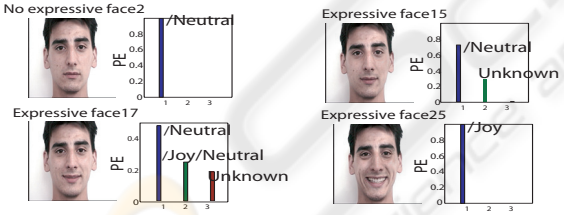

Figure 3: Results of the frame-by-frame classification before the distances correction. In each frame, left: current frame; right: all the facial expressions with a not null piece of evidence.

6 DYNAMIC RECOGNITION OF FACIAL EXPRESSION SEQUENCES

Facial expression is the result of a set of dynamic and asynchronous facial features deformations. Each facial expression is characterized by a *beginning* an *apex* (it corresponds to the maximum of intensity reached by the facial expression associated with a particular distances states configuration) and an *end*. Contrary to the frame-by-frame classification described in the previous section, here we present a method to recognize a sequence of facial expressions between each pair of *beginning* and *end*. In each expression sequence, the *beginning* is detected as the first frame where all the characteristic distances are not any more in the stable state S; the *end* is detected as the first frame where all the states distances have come back to the stable state S. In the scope of our work the *apex* is not identified. We propose two methods for facial expression sequence recognition: the first one combines the frame-by-frame classification results inside a temporal window between the *beginning* and the *end*; the second one is based on the analysis of the dynamic evolution of all the characteristic distances inside a increasing temporal window between each pair of *beginning* and *end*.

Dynamic classification based on the expressions evolution. In the first method the dynamic information is introduced as a temporal integration of the classification results on each frame between the *beginning* and the *end*. For each expression E_i we take into account the number of times it appears and the corre-

sponding piece of evidence m_k. Then the final weight $Sum(E_i)$ associated with each E_i is computed as:

$$Sum(E_i) = Nbf * \sum_{k=beginning}^{k=end} (m_k(E_i)) \quad (9)$$

$$Nbf = (end - beginning) - \sum_{k=beginning}^{k=end} (m_k(E_i) = 0) \quad (10)$$

with $E_i \in \{$*Joy* (E_1), *Surprise* (E_2), *Disgust* (E_3), *Neutral* (E_4), *Unknown* $(E_5)\}$. The chosen expression is the one with the maximum value of $Sum(E_i)$. However based on the results of the frame-by-frame classification (which takes into account at each time only the current distances states) this method cannot handle with asynchronous facial features behaviors. To overcome this limitation, we introduce a second method based on the combination of all the characteristic distances states inside a facial expression sequence.

Dynamic classification based on the characteristic distances evolution. A facial expression is the result of progressive deformations of a set of facial features which can appear at different time and without any defined appearance order (asynchronously). The proposed method allows to deal with all these considerations. It is based on the dynamic analysis of the characteristic distances states during the studied expression. Once the *beginning* of the expression detected, the analysis of the distances states is made inside an increasing temporal window. The size of the window increases progressively between the *beginning* and the *end* which allows at each time to take into account the whole set of previous information to class the current expression sequence.
At each time t, inside the current increasing window, and for all the characteristic distances, the most plausible states are selected for the expression sequence classification from the beginning until current time t. The states selection is made following several steps. First the sum of the occurring states and their plausibility are computed inside the increasing window (Δt). Then for all the distances states $\{C^+, C^-, SC^+, SC^-\}$ we compute the number of times they appear $\sum_{\Delta t} state$ and their plausibility $\underset{\Delta t}{PL}(state) = \sum_{(\Delta t, state \cap states \neq \phi)} m(states)$.
For example for C^+:

$$\sum_{\Delta t} C^+ = \sum_{\Delta t} (state) == (C^+)$$

$$\underset{\Delta t}{PL}(C^+) = \sum_{\Delta t} m(C^+) + m(SC^+)$$

Secondly, based on these values the basic belief assignment of the characteristic distances states is defined. The study of the whole set of the distances states configuration under the studied expressions on the training database allows us to define the rules table described in Table.4. The rows correspond to the rules associated with each distance state and the columns to the required conditions obtained inside the increasing window. For example the two first rows correspond to the rules associated with C^-:

- The number of occurrence of C^- in the increasing window is different to zero and the number of occurrence of C^+ is equal to zero.
- The plausibility of C^- is higher than the one of C^+.

The rules table is defined according to three rules:

- If only one singleton state appear inside the increasing window, this one is chosen to be the state of the studied characteristic distance,
- If two singleton states appear, the most plausible state between them is chosen,
- If only doubt states appear, the most plausible one between them is chosen.

The piece of evidence of each chosen state corresponds to its maximum value inside the current increasing window. Finally at each time *t* (between the beginning and the end of the expression sequence) , once the basic belief assignment of all the characteristic distances is computed, the classification is then carried out based on the rules table (Table 3) as described in section 5.

7 EXPERIMENTAL RESULTS

In this section we present the results of the proposed methods for facial expressions classification and facial features behaviors analysis. First we present the results of facial features behaviors analysis in each frame of the facial expressions sequences. Then, we compare the results of the proposed frame-by-frame classification with the results obtained on our previous works without the use of the transition matrix (i.e static information). Finally we present results of temporal classification of facial expressions sequences. All the experiments have been realized on our test database (ham,).

7.1 Facial Features Behavior Analysis

Since the proposed approach deals with the face expression evolution as well as the description of the

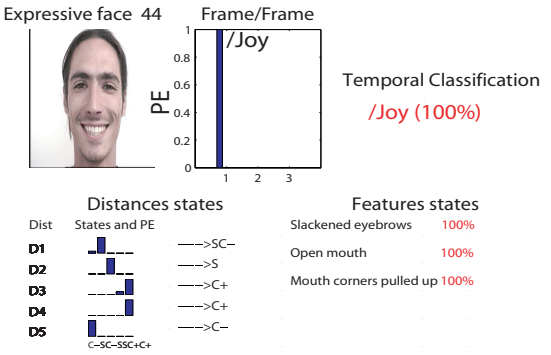

Figure 4: Classification results and facial features behaviors analysis on frame 44 being part of a *Joy sequence*.

permanent facial features behavior (eyes, eyebrows and lips), one of the main goals is to be able to handle the *Unknown* expressions corresponding to all the facial deformations which do not match to any of the studied expressions (in this case even a human observer cannot decide between one of them). Hence in this case the analysis of the individual facial features behaviors is particularly informative because it provides a description of these deformations. This analysis is based on the study of the eyes, the mouth and the eyebrows state. Then the states of the characteristic distances are recovered. Each characteristic distance state is translated into behavior of the corresponding facial features associated with its piece of evidence (see Figure 4). To provide the most accurate possible information, only the facial feature behaviors whose state is singleton (i.e S, C^+ or C^-) are analysed. The doubt states are not used. The set of recognized facial features deformations rules are:

Eyes analysis: Slackened eyes ($D_1 == S$), Open eyes ($D_1 == C^+$), Eyes slightly closed ($D_1 == C^-$),

Eyebrows analysis: Slackened eyebrows ($D_2 == S$), Raised inner eyebrows ($D_2 == C^+$), Lowered inner eyebrows ($D_2 == C^-$),

Mouth analysis: Closed mouth ($D_4 == S$), Open mouth ($D_4 == C^+$), Mouth corners pulled downwards ($D_5 == C^+$), Mouth corners pulled up ($D_5 == C^-$).

Figure 4 presents an example of the information displayed during the analysis of a facial expression sequence and of the corresponding facial features behaviors. This example shows the results on the frame 44 from one *Smile sequence* composed of 120 frames. It corresponds to the apex of the expression. The interface is divided into five different regions: on top left, the current frame to be analysed; on top middle the result of the TBM classification on this frame (here this is a *Joy* expression with a piece of evidence equal to 100%); on top right, the results of the temporal classification which corresponds to the classification of the sequence since the beginning until the current frame (here *Joy sequence)*; on bottom left, the current states of the characteristic distances and their pieces of evidence; on bottom right, the results of the analysis of the facial features behaviors with their pieces of evidence.

7.2 Frame-by-Frame Classification

The aim of this section is to discuss the improvement of the frame-by-frame classification induced by the introduction of the smoothing as well as the temporal information (conditional pieces of evidence). The analysis of the frame-by-frame classification on the facial expression sequences shows that the mean improvement of the results according to our previous works are 6.9%. However, the most interesting results are observed on the sensitivity of the method to transitory states when the facial feature behavior does not correspond to neither the *Neutral* state nor one apex expression. Indeed in the studied sequences the subject begins at neutral state, then reaches the apex of the studied expression and finally comes back to the neutral state. Before and after the apex, the subject exhibits some transitory facial expressions. In these cases, even a human observer cannot classify them or hesitates between different expressions. Our system exhibits a similar behavior as can be seen in Figure3 where top right frame and bottom left frame correspond to transitory states between the *Neutral* and *Joy* expression. However this temporal combination is only based on the information contained in the previous frame. This is not sufficient to recognize the asynchronous temporal behavior of the facial features during a facial expression sequence. This can be overcome by a dynamic recognition of the facial expression sequences.

7.3 Classification in Video Sequence

Figure 5 presents an example of a dynamic classification (based on the characteristic distances evolution) on a *Surprise* sequence. The original sequence has 120 frames where the subject evolves from *Neutral*, reaches *Surprise* and comes back to *Neutral*. We give selected frames to convey our results in Figure 5 which shows the evolution over time of the facial features behaviors and of the facial expressions.

In the first frame the subject is in the *Neutral* state. At this time the system cannot give any temporal classification results. In the second frame we can observe the sensitivity of the system to recover the behavior of the facial features. Based on the states of the characteristic distances, the frame-by-frame classification confidence is 80% that it corresponds to a *Surprise* expression and 20%, to a doubt between *Surprise* and

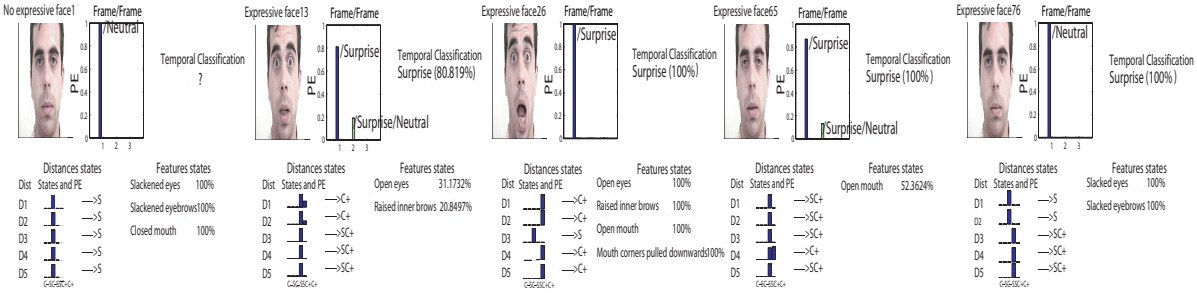

Figure 5: Examples of results obtained on a *Surprise* sequence by the dynamic classification.

Neutral. The temporal classification corresponds to the result of the classification of all the frames from the beginning until the current frame 13. At this time, the temporal classification confidence is 81% that it corresponds to a *Surprise* sequence. The third frame corresponds to the apex. The frame-by-frame results as well as the temporal classification reach a confidence level of 100% on the recognition of a *Surprise* expression. The last two frames give the classification results when the subject comes back to the *Neutral* state. We can observe the evolution of facial features behaviors coming back to slackened states and the frame-by-frame classification which gives a *Neutral* state. However it has to be noticed that the temporal classification does not change and gives the classification over the whole sequence. The same analysis on the remaining part of the Hammal_Caplier test database shows that the proposed approach is able to model the evolution of the expression magnitude at each step of the expression sequences. The obtained mean classification rates on the test database are 70%, however this result is based on few test examples (8 subjects). We are currently testing our method on a larger database (The Japanese Female Facial Expression (JAFFE) Database).

8 CONCLUSION

The presented approach extends the state-of-the-art of automatic facial expression classification by including the temporal information for dynamic expression sequence classification. We focus our attention not only on the nature of the facial features deformation but also on their temporal behavior during a facial expression sequence. Our dynamic facial expression modelling is based on the TBM. Compared to other existing methods, the main advantage of our approach lies in its ability to deal with imprecise data, to model the Unknown expressions and finally to deal with asynchronous facial feature behaviors for the analysis of facial expression sequences.

REFERENCES

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

Bassili, J. (1978). Facial motion in the perception of faces and of emotional expression. *Experimental Psychology - Human Perception and Performance*, 4(3):373–379.

Ekman, P. and Friesen, W. (1978). Facial action coding system. *Consulting Psychologist Press*, 18(11):881–905.

Essa, I. and Pentland, A. (1997). Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. PAMI*, 19(7):757–763.

Hammal, Z., Caplier, A., and Rombaut, M. (2005a). A fusion process based on belief theory for classification of facial basic emotions. *ISIF. Philadelphia, USA*.

Hammal, Z., Couvreur, L., Caplier, A., and Rombaut, M. (2005b). Facial expressions recognition based on the belief theory: Comparison with diferent classifiers. *ICIAP05*, pages 743–752.

Hammal, Z., Eveno, N., Caplier, A., and Coulon, P. (2006). Parametric models for facial features segmentation. *Signal processing*, 86(2):399–413.

Pantic, M. and Rothkrantz, L. (2000). Expert system for automatic analysis of facial expressions. *Image and Vision Computing Journal*, 18(11):881–905.

Smets, P. (1998). *The Transferable Belief Model for Quantified Belief Representation*, volume 1. Handbook of Dfeasible Reasoning and Uncertaintly Management System, Kluwer.

Tian, Y., Kanade, T., and Cohn, G. (2001). Recognizing action units for facial expression analysis. *IEEE Trans. PAMI*, 23(2):97–115.

Yacoob, Y. and Davis, L. (1996). Recognizing human facial expressions from long image sequences using optical flow. *IEEE Trans. PAMI*, 18:636–642.

Zhang, Y. and Qiang, J. (2005). Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Trans. PAMI*, 27(5).

Zhang, Z., Lyons, L., Schuster, M., and Akamatsu, S. (1998). Comparison between geometry-based and gabor wavelets-based facial expression recognition using multi-layer perceptron. *AFGR*, pages 454–459.