Spontaneous Pain Expression Recognition in Video Sequences

Zakia Hammal, Miriam Kunz, Martin Arguin and Frédéric Gosselin
Département de psychologie, Université de Montréal, Canada
zakia_hammal@yahoo.fr, miriam.kunz@uni-bamberg.de, martin.arguin@umontreal.ca, frederic.gosselin@umontreal.ca

Automatic recognition of Pain expression has potential medical significance. In this paper we present results of the application of an automatic facial expression recognition system on sequences of spontaneous Pain expression. Twenty participants were videotaped while undergoing thermal heat stimulation at non-painful and painful intensities. Pain was induced experimentally by use of a Peltier-based, computerized thermal stimulator with a 3 x 3 cm^2 contact probe. Our aim is to automatically recognize the videos where Pain was induced. We chose a machine learning approach, previously used successfully to categorize the six basic facial expressions in posed datasets [1, 2] based on the Transferable Belief Model. For this paper, we extended this model to the recognition of sequences of spontaneous Pain expression. The originality of the proposed method is the use of the dynamic information for the recognition of spontaneous Pain expression and the combination of different sensors: facial features behavior, transient features and the context of the expression study. Experimental results show good classification rates for spontaneous Pain sequences especially when we use the contextual information. Moreover the system behaviour compares favourably to the human observer in the other case, which opens promising perspectives for the future development of the proposed system.

Keywords: *Spontaneous facial expressions, Pain, Transferable Belief Model, Classification*

1. INTRODUCTION

The interpretation of facial expressions, and particularly expressions of emotion, is critical to everyday social interactions [3]. The study of human facial expressions has an impact in several areas of life such as art, social interaction, medicine, security and human-computer interaction (HCI). Other applications of automated systems for facial expressions recognition is in affect-related research like cognitive psychology, psychiatry, and neuropsychology, where such systems can improve research quality by improving the reliability of measurements. Most importantly, it can speed up the currently tedious, manual task of processing data on human affective behavior, notably using the Facial Action Coding System (FACS) developed by Ekman [4, 5, 6]. For all these applications an automatic facial expressions classification system is necessary.

Most of the past work on automatic facial expression analysis has been dedicated to the analysis of posed facial expression and was not always applicable in real-life situations [16, 22]. Indeed, spontaneous facial expressions are often characterized by subtle changes of facial features while the acted facial expressions are characterized by exaggerated changes of facial features [7]. Therefore, the focus of the research in the field has started to shift to automatic analysis of spontaneously displayed facial expressions [7, 8, 9, 10, 11, 6]. Moreover, from the overall state of the art in the field few efforts have been made towards detection of deliberately non-basic affective states such as attentiveness [12], fatigue [13, 14], and pain [8, 15].

Extending our approach [2], we applied machine learning to the task of automatic recognition of spontaneous Pain expressions involving subjects undergoing thermal heat simulation of painful intensities [16]. We focused on Pain because of its potential medical significance, for example, as a pain assessment tool in individuals who are not able to communicate Pain verbally (e.g. newborns, individuals with pronounced cognitive impairments [17, 18, 19]).

This paper reports on a method of dynamic, multi-cue and context depend recognition of spontaneous Pain sequences and aims at making new contributions to the already proposed models.

The first step for automatic facial expression recognition implies face and facial features segmentation. But if we consider the state of the art in face detection and facial features localization as well as tracking, noisy and partial data should be expected [20]. Therefore, facial expression analyzer should be able to deal with noisy and partial data and to generate conclusions so that the associated certainty varies with the certainty of face and facial points localization and tracking [20, 6]. The Transferable Belief Model (TBM) [33] is well adapted to deal with these considerations and is then chosen as analyzer. It facilitates the integration of a priori knowledge and can deal with uncertain and imprecise data. The TBM has been used in several applications such as image processing, geoscience, medicine, robotics and defense [21, 22] and more recently in the analysis and the recognition of human facial and body behavior [1, 2, 23, 24].

In the case of Pain expression, in addition to the permanent facial features behaviour (like eyebrows and mouth), one important part in the automated system was (brow lower) [8]. Moreover, the mechanisms used by the human visual system, remains the best automatic facial expression recognition system. In a recent study, Roy et al. [25, 26] have made a finer and less biased analysis [27] of the importance of facial features for the discrimination of the basic facial expressions including the facial expression of Pain. Their findings showed that Nasal root wrinkles are one of the prominent facial features that drive the human observer for the recognition of Pain expression. Based on these findings nasal root wrinkles are analyzed in combination with the permanent facial features for spontaneous Pain recognition.

In addition to the static facial feature information, it is important to note that in daily life facial expressions are not static, but are the result of dynamic and progressive combinations of facial features deformations. Bassili and more recently Ambadar [44] have shown that facial expressions can be more accurately recognized from image sequences than from single images. Moreover, it has been shown that temporal dynamics of facial behavior represent a critical factor for distinction between spontaneous and posed facial behavior [7, 28, 4, 6], as well as for categorization of complex behaviors like Pain, shame, and amusement [4]. Pain sequences shall be analyzed to take into account the temporal dynamics of the facial features and the decision shall be taken over the whole sequence.

Another limitation of the existing models is their context-independent classification. A few attempts have been made towards context dependent interpretation of the observed facial expression [29, 13, 30, 31, 32]. However facial expressions are accordingly displayed in a particular context, such as the location (outdoor, indoor), the situation (driving a car or being treated in a hospital), the task undergoing, the other people involved, the identity and personality of the expresser [20, 6]. However, to the best of our knowledge, no vision-based model takes into account the context of the application for spontaneous expression recognition. Here we present a model that integrates a context variable in order to refine the recognition process.

The proposed work is a new development of the previously proposed system [2] applied to the dynamic recognition of spontaneous Pain expression in spontaneous videos. To summarize, the originality of the proposed method is twofold: (1) the fusion of different sensors: the permanent facial features deformation (such as eyes, eyebrows and mouth), the transient features information (nasal root wrinkles) and the context of the expression production (e.g. medical context); (2) the use of the dynamic behavior of these sensors for the recognition of sequences of spontaneous Pain expression.

The fusion process of all these pieces of information is based on the TBM which is well-adapted to design a fusion approach where various independent sensors or sources of information collaborate together to provide a more reliable decision [33, 34]. Moreover, most of the already proposed models map facial expressions directly into the basic facial expressions proposed by Ekman and Friesen [35] and are not able to model the doubt between several facial expressions

in the recognition process. This property is important considering that "binary" or "pure" facial expressions are rarely perceived (people usually display mixtures of facial expressions [50]. The TBM based model has proven its ability to deal with all these considerations for the recognition of the basic facial expressions [2, 23]. Here we demonstrate the suitability of this model also for the recognition of spontaneous Pain expression sequences.

The remainder of the paper is organized as follows: first, we describe the facial expression databases we used to train and evaluate the system; second, we describe briefly the main features of our automatic facial expression system and describe the fusion process using the TBM; third, we present the model of temporal classification and the final fusion and decision process of Pain expressions sequences, introducing a context variable; finally we present the classification results both on spontaneous and acted Pain expressions, emphasizing not only on the good performances but also on the quality of the information extracted from the video sequence.

2. FACIAL EXPRESSION DATA
Experiments in this paper are run across two databases: a spontaneous Pain expression database and an acted facial expressions database (STOIC).

2.1 Spontaneous Pain Expression Database
In most facial expression databases, facial expressions are acted. These acted facial expressions differ in appearance and timing from spontaneously occurring facial expressions [7]. Here, we describe the creation of a spontaneous Pain expression database during a study conducted in a lab at the Institut Universitaire de Gériatrie de Montréal. Subjects were participating in a study on the relation between Pain catastrophizing and facial responsiveness to Pain in healthy, Pain-free individuals.

2.1.1 Description of stimulus material
In the database videos, Pain was induced experimentally by mean of a Peltier-based, computerized thermal stimulator (Medoc TSA-2001; Medoc Ltd, Ramat Yishai, Israel) with a 3 × 3 cm^2 contact probe. The contact probe was attached to the left lower leg. Baseline temperature was always set to 38°C. 1 non-painful (1 °C below the individual pain threshold) and 2 painful thermal stimuli (2 – 3 °C above the individual pain threshold) were applied in a random order. The temperature increased from baseline with a heating rate of 4°C/s to the pre-set temperatures, remained at a plateau for 5 seconds and returned to baseline with a rate of 4°C/s. ISIs varied between 30-35 seconds.
The faces of subjects were videotaped (see Figure 1). The video camera was placed in front of the subject at a distance of approximately 4 m. Before applying a stimulus, subjects were always instructed to focus on an emotionally neutral picture being positioned next to the camera in order to ensure a frontal view of the face. Subjects were also instructed not to talk during thermal stimulation. To mark the onset of stimulation on the videotape (for further analysis), we switched on a light signal concurrently. The light was visible to the camera but not to the subject (see Figure 1).
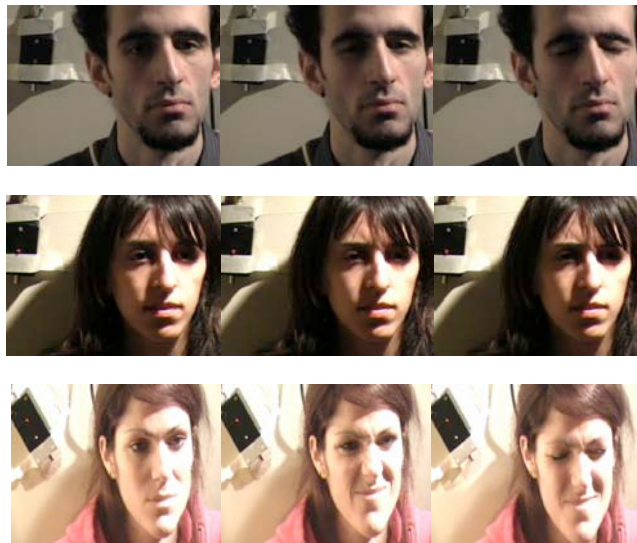
FIGURE 1: Examples from the spontaneous Pain sequences

2.1.2 FACS
Facial expressions displayed on each frame of the obtained videos have been analysed using the Facial Action Coding System (FACS) [35]. This system is based on anatomical analysis of facial muscle movements and distinguishes 44 different action units (AUs). These are the minimal numbered units of facial activity that are anatomically separate and visually distinguishable. The intensity for each action unit was rated on a 5-point scale (A – E) with A being the least intense of the action and E the maximum strength of the action. A FACS coder (qualified by passing an examination given by the developers of the system) identified the frequency of all 44 AUs and the intensity of 42 AUs (AUs 45 and 46 do not allow for intensity coding). A special software designed for analysis of observational data (the Observer Video-Pro (Noldus Information Technology)) was used to segment the videos and to enter the FACS codes into a time-related data-base. Time segments of 5 seconds beginning just after stimulus had reached maximum were selected for scoring.
The database was used for the validation of the proposed model.

2.2 Posed Expression Database
We also used the STOIC dataset (see Figure 2) developed and validated by Roy and collaborators from the Université de Montréal [36]. It is one of the latest facial expressions dataset validated by human observers.
Videos were recorded using a camera located directly in front of the subjects (students from theatrical schools) who were asked to perform the six basic facial expressions (Fear, Happiness, Surprise, Anger, Sadness, and Disgust) as well as Pain and Neutral expressions. Fifteen videos of facial expressions unambiguously recognized as Pain by 25 human observers were employed. The database was divided in two parts: the training set (10 of the Pain videos) and the test set (the remaining five videos).
The training set was used to define the rules for Pain expression recognition. The rules correspond to the facial feature deformations (and then the corresponding characteristic distance states) leading to the maximization of the correlation between the human and the system performances. This process allows defining a combination rules (from all the possible ones) based on human observer validation (see Table 1). This process was also carried out on the basic facial expressions plus Neutral expression but will be presented separately in another paper.

FIGURE 2: Example of Pain sequences from the STOIC database

3. AUTOMATIC SYSTEM

This paper proposes a new development of our model on facial expressions classification [2]. The model is based on the comparison of the permanent facial features (eyes, eyebrows and mouth) deformations to their neutral state using the TBM [33]. It is able to recognize pure expressions plus Neutral as well as doubt between the basic facial expressions (Joy, Surprise, Fear, Disgust, Sadness, Anger). It is also able to deal with all facial feature configurations that does not correspond to any of the cited expressions (Unknown expressions). In the following, the system is generalized for the recognition of Pain expression in video sequences using fusion of visual and contextual information.

3.1 Characteristic distance measurements

The first step in the Hammal *et al.* [2] facial expression model is the extraction of the contours of the permanent facial features (eyes, eyebrows and mouth—see Hammal *et al.* [37]). A specific parametric model is defined for each deformable feature. Several characteristic points are extracted in the image to be processed to initialize each model (for example, eyes corners, mouth corners and brows corners). In order to fit the model with the contours to be extracted, a gradient flow (of luminance and/or chrominance) through the estimated contour is maximized. The chosen models are flexible enough to produce realistic contours for the mouth, the eyes and the eyebrows. More details about this method have already been presented in [37] (see Figure3).

Based on the segmentation results of the first frame a set of characteristic points is selected and tracked in the remaining of the sequence. The algorithm we used to track these facial points is the Lucas-Kanade feature-tracking algorithm [38]. To be the most robust possible the characteristic point positions are re-detected automatically at each eye blink. Indeed the spontaneous expressions are slower than the acted expressions and thus the tracking process of the characteristic points is sufficient. Figure 3 shows an example of the characteristic points tracking.
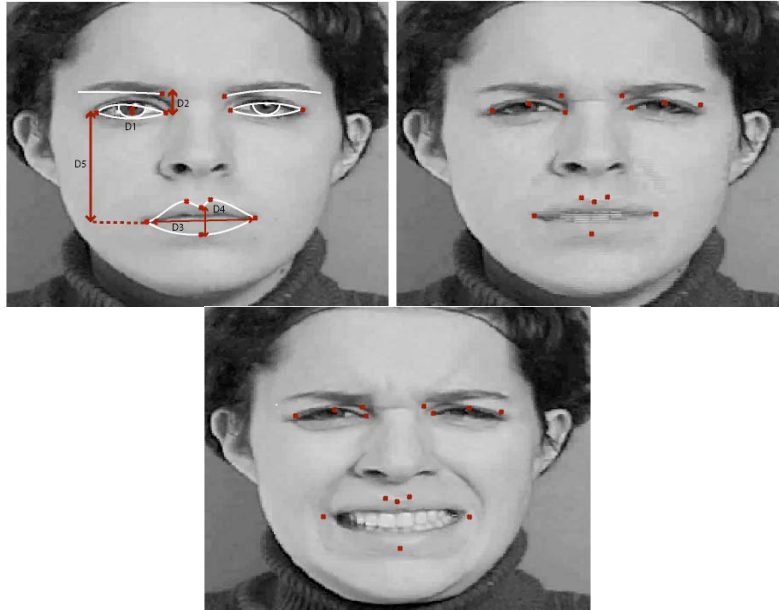
FIGURE 3: Characteristic points tracking and the corresponding characteristic distances.

From the segmentation results, the permanent facial features deformations occurring during facial expressions according to the Neutral state are measured by five characteristic distances D_1, D_2, D_3, D_4 and D_5 (see Figure 3).

In a recent modeling process of a psychophysical experiment modeling the visual cues used by human observer for the classification of the six basic facial expressions Hammal *et al.* proved that these characteristic distances summarize the most important information necessary for the classification process [23].

A numerical to symbolic conversion is then carried out using a fuzzy-like model for each characteristic distance D_i (see TBM section and Hammal *et al.* [2]). It allows the conversion of each numerical value to a belief in five symbolic states reflecting the magnitude of the deformation. S_i if the current distance is roughly equal to its corresponding value in the Neutral expression, C_i^+ (vs. C_i^-) if the current distance is significantly higher (vs. lower) than its corresponding value in the Neutral expression, and $S_i \cup C_i^+$ (vs. $S_i \cup C_i^-$) if the current distance is neither sufficiently higher (vs. lower) to be in C_i^+ (vs. C_i^-), nor sufficiently stable to be in S_i (see Figure 4 for example).

In order to determine the current expression (according to the characteristic distances), a fusion process of the states of the characteristic distances is then performed based on the Transferable Belief Model (see section 4). The TBM has already demonstrated its suitability for the classification of the basic facial expressions [1, 2, 23]. The authors have validated their model on the two well-known benchmark databases (the Cohn-Kanade database [39] and CAFE database [40]) and on their own database (Hammal-Caplier database [41]).

3.2 Transient features

In the current modeling, adding to the permanent facial features deformation, transient features (Nasal root wrinkles) are also used for the classification process. The choice of the Nasal root wrinkles is due to their appearance in the Pain expression [42]. Moreover, more recently Roy and collaborators [25,26] have made a finer and less biased analysis of the importance of facial features for the discrimination of the basic facial expressions as well as Pain expression for human observer. The experiment revealed the precise effective filters for the categorization of the six basic expressions as well as Neutral and Pain. Their results prove that the nasal root wrinkles correspond to one of the most important visual cues used by human observer for Pain expression recognition.

Based on the eyes characteristic points (inner eyes corners) the nasal root area is selected for wrinkles detection (see Figure 4). In the selected area the Nasal root wrinkles detection is based on the Canny edge detector. The presence or absence of wrinkles is decided by

comparing the number of edge points in the nasal root in the current expressive image with the number of edge points in the nasal root of a Neutral facial image. If there are about twice more edge points in the current image than in the reference image, wrinkles are considered to be present. Figure 4 shows an example of nasal root detection. The Canny edge threshold is set by expertise but is kept constant over all the databases. We take a high threshold to minimize the risk of errors. Then based on TBM modeling (see section 4.2) the system will keep the doubt instead of taking the risk of making a wrong decision.
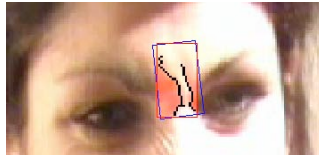

FIGURE 4: Example of nasal root wrinkles detection

3.3 The contextual information

As reported by some researchers [6, 20] a largely unexplored area for facial expressions recognition is that of context dependency. Without context, even human may misunderstand the observed facial expression. Yet, with the exception of a few studies investigated the influence of context on affect recognition, all existing approaches to machine analysis of human affect are context insensitive [6].

Then, an important related issue that should be addressed in all affect recognition is how to make use of information about the context.

The aim of the proposed work is to prove the suitability of the TBM to easily adding one or more context variables in the model of facial expressions classification.

In the case of our application -Pain recognition- several contextual variables can be defined: the place, the task, the answer to a writing question, etc. These contextual variables allow reducing the set of the expected facial expressions. In the current paper only the place is used. It consists of the place where the expression is done. This place variable can take two values medical or not (considering that the expresser is in the hospital or not) and where the aim is to identify if the videotaped expression is painful or not.

The context variable is introduced for Pain recognition in the present study but it can be easily generalized for the other expressions defining a set of rules conditions for each one of them. For example, as reported by [20] Smiling in the context of downward head pitch, communicates embarrassment rather than Joy (the results for the basic facial expressions will be presented separately in another paper).

The context variable is then added to the permanent facial features and transient features as a refinement "sensor" for the classification process. Once all the required information is collected, a fusion architecture based on the TBM is done (see section 6).

4. FUSION PROCESS BY THE TRANSFERABLE BELIEF MODEL

In a realistic interaction environment, a facial expression analyzer should be able to deal with noisy and partial data and to generate its conclusion with confidence that reflects uncertainty of output of face and face point localization and tracking [6]. The Transferable Belief Model (TBM) is then chosen as analyzer.

The TBM is a model of representation of partial knowledge [43, 44] and can be understood as a generalization of probability theory. It can deal with imprecise and uncertain information explicitly defining doubt states and provides a number of tools for the combination of this information [33, 34]. It considers the definition of the frame of discernment $\Omega = \{H_1, ..., H_N\}$ of N exclusive and exhaustive hypotheses characterizing some situations. It means that the solution to the problem is unique and is one of the hypotheses of Ω. The TBM is well adapted to design a fusion approach where various independent sensors or sources of information collaborate together to provide a more reliable decision.

The TBM has already proved its suitability for the classification of the basic facial expressions [1, 2]. It has also proved its ability to deal with partially occluded facial parts, optimizing all the

available information to take the best possible decision and its performances compare favorably to those of human observer in experimental conditions [23].

Based on these considerations, the proposed model for the classification of spontaneous Pain expression is based on the TBM fusion process of all the information resulting from the characteristic distance states, with the addition of the nasal root wrinkles and the context information (medical context).

4.1 The basic belief assignment of the characteristic distances:

Using the TBM approach requires the definition of the Basic Belief Assignment (BBA) associated to each independent source of information.

The BBA $m_{Di}^{\Omega_{Di}}$ of each characteristic distance state D_i is defined as:

$$m_{Di}^{\Omega_{Di}} : 2^{\Omega_{Di}} \rightarrow [0,\ 1]$$
$$A^{\Omega_{Di}} \rightarrow m_{Di}^{Di}(A), \quad \sum_{A \in 2^{\Omega_{Di}}} m_{Di}^{\Omega_{Di}} = 1 \qquad (1)$$

where $\Omega_{Di} = \{C_i^+, C_i^-, S_i\}$, the power set $2^{\Omega_{Di}} = \{\{C_i^+\}, \{C_i^-\}, \{S_i\}, \{S_i, C_i^+\}, \{S_i, C_i^-\}, \{S_i, C_i^+, C_i^-\}\}$ the frame of discernment, $\{S_i, C_i^+\}$ (vs. $\{S_i, C_i^-\}$) the doubt state between C_i^+ (vs. C_i^-) and S_i. $m_{Di}^{\Omega_{Di}}(A)$ is the belief in the proposition $A \in 2^{\Omega_{Di}}$ without favoring any of propositions of A in case of doubt proposition. This is the main difference with the Bayesian model, which implies equiprobability of the propositions of A. A is called focal element of $m_{Di}^{\Omega_{Di}}(A)$ whenever $m_{Di}^{\Omega_{Di}}(A) > 0$. Total ignorance is represented by $m_{Di}^{\Omega_{Di}}(\Omega_{Di}) = 1$. To simplify, $\{C_i^+\}$ is noted C^+ and $\{S_i, C_i^-\}$ is noted $S \cup C^+$ (i.e. S or C^+).

The piece of evidence $m_{Di}^{\Omega_{Di}}$ associated with each symbolic state given that the value of the characteristic distance D_i is obtained by the function depicted in Figure 5. The threshold values {a, b, c, d, e, f, g, h} have been derived by statistical analysis on the Hammal-Caplier database (Hammal-Caplier database [41]) for every characteristic distance. Details can be found in Hammal et al. [2].
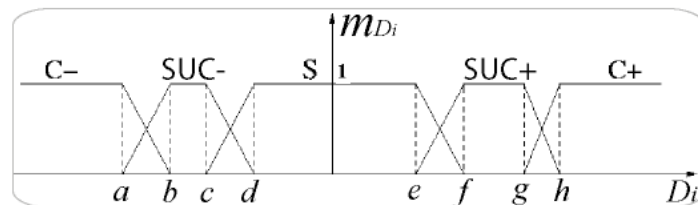

FIGURE 5: Model of basic belief assignment based on characteristic distance D_i. For each value of D_i, the sum of the pieces of evidence of the states of D_i is equal to 1.

4.2 The basic belief assignment of the transient features

The BBA $m_{TF}^{\Omega_{TF}}$ of the nasal root wrinkles is defined as:

$$m_{TF}^{\Omega_{TF}} : 2^{\Omega_{TF}} \rightarrow [0,\ 1]$$
$$B^{\Omega_{TF}} \rightarrow m_{TF}^{\Omega_{TF}}(B), \quad \sum_{B \in 2^{\Omega_{TF}}} m_{TF}^{\Omega_{TF}} = 1 \qquad (2)$$

where $\Omega_{TF} = \{P, A\}$, the power set $2^{\Omega_{TF}} = \{\{P\}, \{A\}, \{P, A\}\}$ the frame of discernment, P means that the nasal root wrinkles are present and A means that they are absent. From the frame of discernment only the states P (we are sure that the wrinkles are present) and the state $P \cup A$ (we don't know) are considered (the notation are simplified like the section 4.1).

The piece of evidence $m_{TF}^{\Omega_{TF}} = 1$ is associated with each symbolic state given that the presence or the absence of the transient features. The detection threshold has been derived by statistical analysis on the Hammal-Caplier and the STOIC databases [41, 36].

As reported above (see section 3.2), we take a high threshold to minimize the risk of errors. Then if the number of wrinkles pixels is higher than the threshold, the system is sure that the

nasal root wrinkles are present ($m_{TF}^{\Omega_{TF}}(P) = 1$) otherwise the system keeps the doubt instead of taking the risk of making a wrong decision and then ($m_{TF}^{\Omega_{TF}}(P \cup A) = 1$).

The nasal root wrinkles are used for the Pain identification as reported in the Table 2. The detection of the nasal root wrinkles allows a refinement of the classification eliminating the expressions: Happy, Surprise, Fear, Sadness and Neutral reducing the number of the possible expressions to 3 rather than 8. Indeed, in addition to Pain expression, the Nasal root wrinkles can also be present in the case of Anger and Disgust expressions.

To summarize:

- If the nasal roots are present: the current expression is Pain or Anger or Disgust (without favoring any of them) and the corresponding piece of evidence is computed as:

$$m_{TF}^{\Omega_{TF}}(P) = m_{TF}^{\Omega_{TF}}(Pain \cup Anger \cup Disgust) = 1$$

- If they are absent: the current expression is one of the 8 expressions with the piece of evidence:

$$m_{TF}^{\Omega_{TF}}(P \cup A) = m_{TF}^{\Omega_{TF}}(Pain \cup Anger \cup Disgust \cup Happy \cup Surprise \cup Fear \cup Sadness \cup Neutral) = 1$$

	Pain	Anger	Disgust	Happy	Surprise	Fear	Sadness	Neutral
Nasal root wrinkles	$P \cup A$	$P \cup A$	$P \cup A$	A	A	A	A	A

TABLE 1: Rules table based on the nasal root wrinkles for Pain detection

4.3 The basic belief assignment of the context variable

The BBA $m_{CT}^{\Omega_{CT}}$ of the contextual variable is defined as:

$$m_{CT}^{\Omega_{CT}} : 2^{\Omega_{CT}} \to [0, 1]$$

$$C^{\Omega_{CT}} \to m_{CT}^{\Omega_{CT}}(C), \quad \sum_{C \in 2^{CT}} m_{CT}^{\Omega_{CT}} = 1 \qquad (3)$$

where $\Omega_{CT} = \{MC, NMC\}$, the power set $2^{\Omega_{CT}} = \{\{MC\}, \{NMC\}, \{MC, NMC\}\}$ the frame of discernment, MC means medical context (the expresser is in a medical context then it is more likely that the expected expression corresponds to Pain) and NMC means not medical context, $MC \cup NMC$ means that we don't have any idea of the context of the expresser (then the expected expression can be one of the 8). From the frame of discernment only the states MC (we are sure) and the state $MC \cup NMC$ (we don't know) are taken into account (the notation are simplified like the section 4.1).

The piece of evidence $m_{CT}^{\Omega_{CT}}$ associated with each symbolic state given that the context of the sequence acquisition is medical or not. It corresponds to the answer to the question: are we trying to know if the expresser is painful or not? -. Then if this is the case the piece of evidence of the state MC is equal to 1 ($m_{CT}^{\Omega_{CT}}(MC) = 1$), otherwise the piece of evidence of the state $MC \cup NMC$ is equal to 1 ($m_{CT}^{\Omega_{CT}}(MC \cup NMC) = 1$).

The piece of evidence of the corresponding expressions is computed as:

   - If we are in a medical context, trying to know if the expresser is painful or not, the pieces of evidence of the corresponding expressions is:

$$m_{CT}^{\Omega_{CT}}(MC) = m_{CT}^{\Omega_{CT}}(Pain) = 1$$

   - Otherwise the pieces of evidence of the corresponding expressions is:

$$m_{CT}^{\Omega_{CT}}(MC \cup NMC) = m_{CT}^{\Omega_{CT}}(Pain \cup Anger \cup Disgust \cup Happy \cup Surprise \cup Fear \cup Sadness \cup Neutral) = 1$$

In the case of our application the variable state corresponding to the context is defined manually according to the fact that the expresser is in a medical context and we are trying to know if its current state is painful or not.

However, the context variable allows only refining the already obtained classification results (based on the facial sensors) where Pain is already recognized or where the system hesitates between Pain and another expression (see section results).

Moreover as reported above we can have several contextual variables according to the expressions and the context information we want to model. We are working on that development of our model for a contextual recognition of the six basic facial expressions.

In order to prove the refinement process based on the use of the context variable, two simulations are presented for the recognition of spontaneous Pain expressions sequences: with the use of the context variable and without the use of the context variable.

5. TEMPORAL INFORMATION FOR FACIAL EXPRESSION CLASSIFICATION

Temporal dynamics of human facial behaviour is a critical factor for the interpretation of the facial expressions [7, 44] and is moreover essential for categorization of complex psychological states like various types of Pain and Mood [45].

In the following we take into account the dynamic behavior of the permanent facial features for the classification process of sequences of facial expressions introducing their temporal behavior pattern.

The described model has been proposed for the 3 facial expressions Joy, Disgust and Surprise [46] and has been generalized for the 8 facial expressions (Joy, Disgust, Surprise, Sadness, Fear, Anger and Pain as well as Neutral). However as we are interested in the Pain expression classification only the results related to Pain and especially spontaneous Pain sequences are reported in this paper. To our knowledge this is the one of the first tentative to explicitly model the dynamic behavior for the recognition of sequences of spontaneous Pain expressions.

The temporal information is introduced at two levels: first by taking into account at each time t the information at time t-1; second by combining all these information from the beginning until the end of the sequence to take the decision.

5.1 Basic belief assignment prediction of the characteristic distance states

The main idea is to define an evolution model for the permanent facial features behavior and then the corresponding characteristic distance states. The model predicts the basic belief assignment m_t at time t according to the basic belief assignment m_{t-1} at time $t-1$ (it is assumed that the two BBAs are close because the information between two consecutive frames is strongly related). Indeed, spontaneous facial expressions are slower in time than posed facial expressions [cohnsmile04, valstarpantic06].

The temporal model consists in defining the conditional pieces of evidence, gathered in a "transition matrix" for each characteristic distance. The predicted basic belief assignment $\hat{m}_{D_j,t}$, t $(1 \leq j \leq 5)$ defined on $2^{\Omega_{D_j}}$ consists in predicting the pieces of evidence $m_{D_j,t}$ of the characteristic distance states at time t according to their pieces of evidence $m_{D_j,t-1}$ at time $t-1$. The predicted basic belief assignment is computed at each time t by the combination of a transition matrix $M(D_j)$ and the computed basic belief assignment at time $t-1$ in the following way:

$$\hat{m}_{D_j,t} = \overline{M}(D_j) * m_{D_j,t-1} \qquad (4)$$

Where:

$$\hat{m}_{D_j,t} = \begin{pmatrix} m_{D_j,t}(C^+) \\ m_{D_j,t}(C^-) \\ m_{D_j,t}(S) \\ m_{D_j,t}(S \cup C^+) \\ m_{D_j,t}(S \cup C^-) \end{pmatrix} \qquad m_{D_j,t-1} = \begin{pmatrix} m_{D_j,t-1}(C^+) \\ m_{D_j,t-1}(C^-) \\ m_{D_j,t-1}(S) \\ m_{D_j,t-1}(S \cup C^+) \\ m_{D_j,t-1}(S \cup C^-) \end{pmatrix}$$

The elements of the matrix $M(D_j)$ are the fraction of the mass to be in a given state at time t knowing the state at time $t-1$. This matrix corresponds to the temporal evolution model; it is defined for each characteristic distance independently of the subject and of the expression.

Temporal evolution model

The temporal evolution model corresponds to the transition matrix composed of a distribution of conditional pieces of evidence [47]. These pieces of evidence correspond to the pieces of transition from each proposition *A* of the frame of discernment at time *t*−1 (previous frame) to each one of the possible propositions *B* at time *t* (current frame) and is noted $m_{D_j}[A](B)$ (*A* and $B \in \{S, C^+, C^-, S \cup C^+, S \cup C^-\}$). For example, for a considered distance D_j, $m_{D_j}[S](C^+)$ corresponds to the piece of evidence (the belief) $m_{D_j}(C^+)$ at time *t* such as $m_{D_j}(S) = 1$ at time *t* − 1.

For each characteristic distance *Dj* (1≤*j*≤5), all the conditional pieces of evidence are gathered in the corresponding *transition matrix* $M(D_j)$ as:

$$\overline{M}(D_j) = \begin{pmatrix} m_{D_j}[S](S) & m_{D_j}[C^+](S) & m_{D_j}[C^-](S) & m_{D_j}[S \cup C^+](S) & m_{D_j}[S \cup C^-](S) \\ m_{D_j}[S](C^+) & m_{D_j}[C^+](C^+) & m_{D_j}[C^-](C^+) & m_{D_j}[S \cup C^+](C^+) & m_{D_j}[S \cup C^-](C^+) \\ m_{D_j}[S](C^-) & m_{D_j}[C^+](C^-) & m_{D_j}[C^-](C^-) & m_{D_j}[S \cup C^+](C^-) & m_{D_j}[S \cup C^-](C^-) \\ m_{D_j}[S](S \cup C^+) & m_{D_j}[C^+](S \cup C^+) & m_{D_j}[C^-](S \cup C^+) & m_{D_j}[S \cup C^+](S \cup C^+) & m_{D_j}[S \cup C^-](S \cup C^+) \\ m_{D_j}[S](S \cup C^-) & m_{D_j}[C^+](S \cup C^-) & m_{D_j}[C^-](S \cup C^-) & m_{D_j}[S \cup C^+](S \cup C^-) & m_{D_j}[S \cup C^-](S \cup C^-) \end{pmatrix}$$

where the sum of all the conditional pieces of evidence belonging to the same column is equal to 1 and the matrix dimensions is [5*X*5].

The piece of evidence of the transitions are learned using the Hammal-Caplier database and have already been validated for the three expressions Joy, Surprise and Disgust [46] and generalized for the six basic facial expressions plus Neutral plus Pain. In this paper only Pain results are reported. Tested on spontaneous sequences the proposed model proves its generalization and robustness for news data.

The conditional basic belief assignment can be defined between each two consecutive frames (*t*−1, *t*) (($m_{D_j,1} \rightarrow m_{D_j,2}$), ($m_{D_j,2} \rightarrow m_{D_j,3}$), …, ($m_{D_j,t-1} \rightarrow m_{D_j,t}$); $1 \leq t \leq N$, N the total number of frames per sequence).

Then it exists a transition matrix $\overline{M}_i^e(D_j)$ for each distance D_j, for each subject *i* and for each expression *e* noted $\overline{M}_i^e(D_j)$ such as:

$$m_{D_j,t} = \overline{M}_i^e(D_j) * m_{D_j,t-1} \quad (5)$$

Equation 5 is defined for one transition between (*t*−1, *t*). To obtain the transitions on the whole sequence, the $m_{D_j,t}$ and $m_{D_j,t-1}$ are concatenated in $M_{i(D_j,2..N)}^e$ (resp. $M_{i(D_j,1..N-1)}^e$) corresponding to the BBA of the considered distance D_j, for the subject *i* from the frame 2 until N (resp. 1 until N-1) for the expression *e* such as:

$$M_{i(D_j,2..N)}^e = \overline{M}_i^e(D_j) * M_{i(D_j,1..N-1)}^e \quad (6)$$

or more detailed:

$$\begin{pmatrix} m_{D_j,2}(S) & ... & m_{D_j,N}(S) \\ m_{D_j,2}(C^+) & ... & m_{D_j,N}(C^+) \\ . & ... & . \\ . & ... & . \\ .m_{D_j,2}(S \cup C^-) & ... & m_{D_j,N}(S \cup C^-) \end{pmatrix}^e_{i~[5XN]} = \overline{M}^e_i(D_j) * \begin{pmatrix} m_{D_j,1}(S) & ... & m_{D_j,N-1}(S) \\ m_{D_j,1}(C^+) & ... & m_{D_j,N-1}(C^+) \\ . & ... & . \\ . & ... & . \\ m_{D_j,1}(S \cup C^-) & ... & m_{D_j,N-1}(S \cup C^-) \end{pmatrix}^e_{i~[5XN-1]}$$

It has to be noted that the elements of each column of these matrices correspond to the BBA associated to the characteristic distance (D_j) for the considered subject i and the expression e in the current frame. For example, for the proposition S in $M^e_{i(D_j,2..N)}$, $m_{D_j,2}(S)$ corresponds to the piece of evidence of the state S for the subject i in the frame 2 for the expression e.

To define the transition matrix (see equation 8) for each distance independently of the subjects and independently of the studied expressions, the matrices $M^e_{i(D_j,2..N)}$ (resp. $M^e_{i(D_j,2..N-1)}$) are concatenated over all the subjects i ($1 \leq i \leq 11$) and over all the expressions e ($e \in$ {*Joy, Surprise, Disgust, Fear, Sadness, Anger, Pain*}) to define the transition matrix $M(D_j)$ ($1 \leq j \leq 5$) for each distance Dj as:

$$\overline{M}(D_j) = M^{ha,sp,dg,fe,sa,an}_{1..11(D_j,2..N)} * M^{ha,sp,dg,fe,sa,an^{-1}}_{1..11(D_j,1..N-1)} \tag{7}$$

where ha: Happy, sp: Surprise, dg: Disgust, fe: Fear, sa: Sadness, an: Anger.

Equation 7 leads to an over determined system. The matrix $\overline{M}(D_j)$ is then computed as its optimum solution in the least square sense [46].

The transition matrix has been already validated for the three expressions (Joy, Disgust and Surprise [46]). This paper describes its generalization for Spontaneous Pain expression.

5.2 Sequence expressions classification

A facial expression is the result of progressive deformations of a set of facial features appearing at different times and without any defined appearance order (asynchronously) [46]. Spontaneous facial expression is characterized by a *beginning*, one or more *apexes* and an *end* [7].

In each expression sequence, the *beginning* is detected as the first frame where at least one of the permanent facial features (and then the corresponding characteristic distances) is no more in the stable state S (Neutral); the *end* is detected as the first frame where all the permanent facial features (and then the corresponding characteristic distance states) have come back to the stable state S. However there is no way to detect the apexes of one expression sequence. The proposed method deals with this consideration. The recognition of Sequence of Pain expression is done taking into account all the available information (previously facial features deformation) between each pair of *beginning* and *end* frames. Then the BBAs of the characteristic distance states at each frame in this interval are computed as described in section 4.1 and combined based on the rules table (see Table 1) to define the expression corresponding to all these deformations.

Processing

Once the *beginning* of the expression has been detected, the analysis of the distance states is made inside an increasing temporal window Δt. The size of the window Δt increases progressively at each time from the *beginning* until the *end* of the expression. Then, at each time t, the whole set of the previous information (the past states of the characteristic distances and then the corresponding facial features behavior) is taken into account to classify the current expression sequence.

This allows to explicitly deal with the dynamic of the facial expression and more importantly with asynchronous facial features deformations. Once the *beginning* is detected, the classification consists in defining at each time *t* the basic belief assignment BBA of the characteristic distance states defined on $\{C^+, C^-, S \cup C^+, S \cup C^-\}$ according to their past basic belief assignments from the *beginning* until the current frame. To do this, at each time *t*, inside the current window Δ*t* and for each characteristic distance, a criterion has to be used to select its corresponding state according to its previous behavior. The selection is made according to the number of appearance of each symbolic state in $\{C^+, C^-, S \cup C^+, S \cup C^-\}$ noted $\underset{\Delta t}{Nb}(state)$ (see Equations 9 for example) and their integral (sum) of plausibility noted $\underset{\Delta t}{Pl}(state)$ computed inside the temporal window Δ*t* (see Equation 10 for example). These rules have been already validated for the 3 expressions Joy, Disgust and Surprise in [46] and are generalized for the Pain expression in this paper.

For instance, for a characteristic distance D_j and for the *state* = C^+:

$$K_t(C^+) = \begin{cases} 1 & if \quad m_{D_j}(C^+) \neq 0 \\ 0 & otherwise \end{cases} \quad 1 \leq t \leq \Delta t \; ; \; (8)$$

$$\underset{\Delta t}{Nb}(C^+) = \sum_{t=1}^{\Delta t} K_t(C^+) \; ; \quad (9) \qquad \underset{\Delta t}{Pl}(C^+) = \sum_{t=1}^{\Delta t} (m_{D_j}(C^+) + m_{D_j}(S \cup C^+)) \quad (10)$$

where K_t indicates the occurring or not of a symbolic state at time *t*.
From these two parameters $\underset{\Delta t}{Nb}(state)$ and $\underset{\Delta t}{Pl}(state)$, some rules are used to choose the distance states at each time *t* inside the temporal window Δ*t* [46] as:
  • If only one singleton state appears 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
At the *beginning*, all the distances are in the stable state *S* and change only if one of the other states appear in the increasing window. In this case the corresponding state is chosen according to the rules defined above.

The piece of evidence associated to each chosen state corresponds to its maximum peace of evidence inside the current temporal increasing window. Finally at time *t* between the *beginning* and the *end* of the expression sequence, once the basic belief assignments of all the characteristic distances are defined, the corresponding expression is selected according to the rules table (see Table 1). Then it is fused to the information produced by the nasal root wrinkles and the context variable to give the best possible decision.

The salient character of this classification is that a decision can be made at each time *t* taking into account all the past basic belief assignment of the characteristic distance states (and then the whole dynamic of the corresponding facial features) from the *beginning* until the current frame [46]. At the beginning of the sequence all the expressions are in the set of possible expressions and, during the sequence, this set is progressively reduced. When reaching the *end* (the current frame is then the last frame of the sequence), the decision depends explicitly on all the past basic belief assignments of the characteristic distance states and gives the classification on the entire expression sequence.

6. FUSION PROCESS

The main feature of the TBM is the powerful combination operator [33, 48] that integrates information from different sensors. In the current case the sensors are the characteristic distance states, the nasal root wrinkles and the contextual variable. However it requires the definition of the fused information on the same frame of discernment.
The fusion process is done in three successive steps at each time (frame) of the sequence: first the fusion of all the characteristic distance states; then combination of the obtained results to

the nasal root wrinkles and finally combination to the context of the application to refine the classification results.

6.1 Fusion of the characteristic distances information

Based on the facial feature deformations associated with Pain expression, this latter is characterized by a set of characteristic distance states according to the rule displayed in Table 1 (see [2] for the rules of the six basic facial expressions). This mapping has been obtained from the Pain expression sequences validated by human observers on the STOIC database [36]. From these rules and in order to take into account all the available information, the facial expression classification is first based on the TBM fusion process of all the D_i states.

The BBAs $m_{D_i}^{\Omega_{Di}}$ of the states of the characteristic distances are defined on different frames of discernment. For the fusion process, it is necessary to redefine the BBAs on the same frame of discernment 2^{Ω}, where

$\Omega = \{Happiness(E_1), Surprise(E_2), Disgust(E_3), Fear(E_4), Anger(E_5), Sadness(E_6), Pain(E_7), Neutral(E_8)\}$

the set of expressions (Pain need to be identified from the set of 8 facial expressions).

	D_1	D_2	D_3	D_4	D_5
Pain	C^-	C^-	C^+	$C^+ \cup C^-$	C^-
Neutral	S	S	S	S	S

TABLE 2: Rules table for Pain recognition

From the rules table and the BBAs of the states of the characteristic distances $m_{D_i}^{\Omega_{Di}}$, a set of BBAs on facial expressions $m_{D_i}^{\Omega_{Di}}$ is derived for each characteristic distance D_i. In order to combine all this information, a fusion process of the BBAs $m_{D_i}^{\Omega}$ of all the states of the characteristic distances is performed using the conjunctive combination rule [33, 34] (see equation 11) and results in m^{Ω} the BBA of the corresponding expressions:

$$m^{\Omega} = \oplus m_{D_i}^{\Omega} \qquad (11)$$

For example, if we consider two characteristic distances D_i and D_j with two BBAs $m_{D_i}^{\Omega}$ and $m_{D_j}^{\Omega}$ derived on the same frame of discernment, the joint BBA $m_{D_{i,j}}$ is given using the conjunctive combination (orthogonal sum) as:

$$m_{D_{i,j}}^{\Omega}(A) = (m_{D_i}^{\Omega} \oplus m_{D_j}^{\Omega})(A) = \sum_{E \cap F = A} m_{D_i}^{\Omega}(E) * m_{D_j}^{\Omega}(F) \qquad (12)$$

where A, E and F denote propositions and $E \cap F$ denotes the conjunction (intersection) between the propositions E and F. This leads to propositions with a lower number of elements and with more accurate pieces of evidence.

The results of the characteristic distances combination are then refined by their combination with the nasal root wrinkles and the context information.

6.2 Fusion of the characteristic distances results with nasal root wrinkles and the contextual information

From the BBAs of the nasal root wrinkles states $m_{TF}^{\Omega_{TF}}$ and the context variable $m_{CT}^{\Omega_{CT}}$ a set of BBAs on facial expression is also derived $m_{TF}^{\Omega}, m_{CT}^{\Omega}$ respectively. The results of the combination of the characteristic distances are then combined by the conjunctive combination with those of the nasal root wrinkles as reported by the following equation:

$$m_{D,TF}^{\Omega}(G) = (m_D^{\Omega} \oplus m_{TF}^{\Omega})(G) = \sum_{A \cap B = G} m_D^{\Omega}(A) * m_{TF}^{\Omega}(B) \qquad (13)$$

where G, A and B denote propositions and $A \cap B$ denotes the conjunction (intersection) between the propositions A and B.

Finally the combination of the characteristic distances and the nasal root wrinkles are combined by the conjunctive combination with those of the context variable as reported by the following equation:

$$m_{D,TF,CT}^{\Omega}(H) = (m_{D,TF}^{\Omega} \oplus m_{CT}^{\Omega})(H) = \sum_{G \cap C = H} m_{D,TF}^{\Omega}(G) * m_{CT}^{\Omega}(C) \qquad (14)$$

where H, G and C denote propositions and $G \cap C$ denotes the conjunction (intersection) between the propositions G and C.

7. DECISION PROCESS

The decision is the ultimate step of the classification process. It consists in making a choice between various hypotheses E_e and their possible combinations. Making a decision is associated with a risk except if the result is sure ($m(E_e) = 1$). Several decision criteria can be used [33, 34].
In this paper the decision was made using the pignistic probability BetP [49] as:

$$BetP : \Omega \rightarrow [0,\ 1]\ I \rightarrow BetP(I) = \sum_{H \subseteq \Omega, I \in H} \frac{m^{\Omega}(H)}{(1 - m^{\Omega}(\phi)) * Card(H)}, \forall\ I \in \Omega \qquad (15)$$
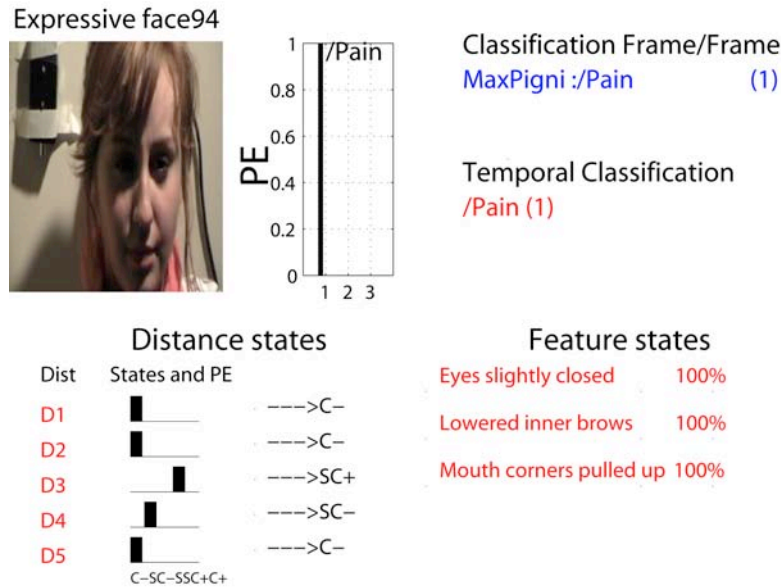
where ϕ corresponds to the conflict between the sensors.

8. CLASSIFICATION OF SPONTANEOUS VERSUS ACTED PAIN

The simulation results were obtained on the 20 spontaneous pain sequences obtained on the experimental condition described in section 2.1 and the 15 videos sequences of the STOIC database (validated by human expert).
Our simulations aimed at, first, proving the generalization of the proposed model for all the facial expressions and its robustness to identify Pain expression, discriminating it from the six basic facial expressions plus Neutral and second, proving the refinement role of the contextual information. Three simulations were carried out: first, the system performed a 2-alternatives choice between Pain and Neutral without the use of the context variable; second, it performed an 8-alternatives choice between Pain and the six basic facial expressions plus Neutral without the use of the context variable and finally it performed the same 8-alternatives choice using the context variable.
Figure 5 presents an example of the information displayed during the analysis of the Pain expression sequences. The interface is divided into five different regions: on top left, the current frame to be analyzed; on top middle the result of the static classification (based only on the information at the current frame, here this is a Pain expression with a Pignistic probability equal to 1); on top right, the result of the dynamic classification which corresponds to the classification of the sequence since the beginning until the current frame 94 (here Pain sequence with a Pignistic probability equal to 1, see section 5); on bottom left, the current states of the characteristic distances and their pieces of evidence; on bottom right, the corresponding facial features deformations.

FRAME 6: User interface displaying the classification information extracted during a sequence. Top left: current frame; top middle, BBAs of the expressions; Top right, classification results (MaxPigni :maximum pignistic probability); bottom left, estimation of the distance states and the corresponding facial features deformations with their pieces of evidence

8.2 Spontaneous pain results

In this section we investigate how the proposed model performed on "spontaneous" Pain sequences. Our investigation includes the generalization to a new database, as well as head movement with both in-plane and out-of-plane rotations. The results are presented on 20 subjects.

In the first simulation the system performed a 2-alternatives choice between Neutral and Pain facial expressions. The classification rates compare favorably (70%) with the already reported classification rates on spontaneous facial expressions. However, this comparison remains difficult to do as their approach differ in several characteristics (the systems were applied on different expressions, they either classify expressions or Action Units and they are tested on different databases). Based on a 2-alternatives choice, the Ignorance state corresponds to the cases where the Pignistic probabilities of Pain and Neutral are equal (0.5). Having only two expressions this state corresponds to total ignorance of the system (30%) (see Table 3).

	Pain	Ignorance
Pain	70	30

TABLE 3: Classification results (%) of spontaneous Pain sequences in the case of two-alternatives choice

In order to know if the system was able to recognize Pain expression and discriminate it from the six basic facial expressions as well as Neutral, a second simulation was carried out where the system performed an 8-alternatives choice between the six basic facial expressions as well as Neutral and Pain. It has to be noted that such an 8-alternatives classification has never been done on spontaneous Pain expressions. The classifications rates are reported in Table 4.

The row (Sadness, Pain + Anger, Pain) corresponds to the cases where the system recognizes at the same time Pain expression with one of the two expressions Sadness or Anger. In these cases the two couple of expressions Pain/Sadness or Pain/Anger are recognized with the same Pignistic probability (Pain=0.5 and Sadness=0.5 or Pain=0.5 and Anger=0.5). Pain is then recognized equiprobably with Sadness or with Anger. Similarly the row (Sadness, Anger, Pain) corresponds to the cases where the three expressions are recognized equiprobably (Pignistic probability equal to 0.33).

To summarize, the system recognizes spontaneous Pain expression but mix (or doubt) with Sadness or Anger at the same time. In these cases the system is sure that the current expression is one of these 2 and never one of the 6 other expressions. Considering hospital

context application (waiting room, older people under camera monitoring), such as information is more than sufficient to alert somebody in charge that the patient is suffering Pain. A human observer (medical doctor or nurse) can then confirm or not this information.

Moreover, the obtained results reflect the ability of the proposed model to deal with the perception of mixture of facial expressions (people usually do not display "pure" facial expressions [50]). The obtained results are comforted by the results of Roy and collaborators where human observers where asked to classify Pain expression in the 8-alternatives choice. Interestingly, the model shows a striking similarity with humans who also misclassify Pain with Sadness or with Anger expressions [25, 26].

In the third simulation, the context variable is added to the 8-alternatives choice to refine the classification results. Results show that the doubt (between Pain, Sadness and Anger) is solved (Table 4). This result proves the usability of the context variable in our application. We obtain a classification rate of 77% which compares favorably to the previous results on automatic recognition of spontaneous expressions.

The ignorance state corresponds to the case where the system recognizes at the same time the 8 facial expressions with the same Pignitic probability (.0125). This case corresponds to the total ignorance of the system.

	Happ	Disgu	Surpris	Fe	Sadnes	Ang	Pa	Neutr	Sadness, Pain +Anger, Pain	Sadness, Anger, Pain	Sadness Anger	Ignoranc
Pain (without context variable)	0	0	0	C	0	0	C	0	62	15	15	8
Pain (wi context variable)	0	0	0	C	0	0	7	0	0	0	15	8

TABLE 4: Classification results (%) of spontaneous Pain sequences in the case of an 8-alternative choice without context variable (first row) and with context variable (second row)

As a conclusion, all these results show the ability of the proposed model to deal with spontaneous Pain sequences and to take advantage of the context information. It also shows its robustness when applied to new sets of data (not used for training).

8.1 Acted pain results
The classification results are also reported on the acted database (the Stoic database validated by human observer). In the first simulation the system performed a 2-alternatives choice between Neutral and Painful. The classification rates are higher than spontaneous Pain expressions (92.3%) while the Ignorance rates decrease significantly (7.7%) (see Table 5).

	Pain	Ignoranc
Pain	92.3	7.7

TABLE 5: Classification results (%) of acted Pain sequences in the case of two-alternatives choice

	Happy	Disgust	Surprise	Fear	Sadness	Anger	Pain	Neutral	Sadness, Pain	Smile, Disgust, Pain
Pain (without context variable)	8	0	0	0	0	0	0	0	15	77
Pain (with context variable)	8	0	0	0	0	0	92	0	0	0

TABLE 6: Classification results (%) of spontaneous Pain sequences in the case of an 8-alternative choice without context variable (first row) and with context variable (second row)

Similarly to the spontaneous Pain simulation, the system was tested on an 8-alternatives choice with and without context variable. The classification rates are reported in Table 6. In this database, the doubt is no longer between the expected expressions (Sadness and Anger) but between Smile and Disgust expressions. These results emphasize the difference between the acted and the spontaneous expressions especially for Pain expression.

As for the spontaneous results, the combination with the context variable leads to solve this doubt leading to a good classification results (92%).

The classification results for acted and spontaneous Pain sequences and especially the difference between the mixed expressions in each case opens promising perspectives for future development of the model aiming at discriminating between acted and real Pain expressions.

CONCLUSION

Here we presented results for the classification of spontaneous Pain expression. The system proves its suitability to deal with spontaneous sequences and gives good classification results, encouraging our future work on other spontaneous facial expressions. Moreover, it allows modeling the doubt between expected expressions revealing the same confusion as the ones obtained by a human observer. The good classification rates in the two databases and the difference between the misclassified expressions gives us indications for our current development of the model and its ability to dissociate between acted and real Pain expressions.

AKNOWLEDGMENT

REFERENCES

[1] Hammal, Z., Caplier, A. and Rombaut, M. (2005) A fusion process based on belief theory for classification of facial basic emotions. *Proceedings of the 8th International Conference on Information fusion, ICIF 05,* 27 June-1 July, Philadelphia, PA, USA.
[2] Hammal, Z., Couvreur, L., Caplier, A., and Rombaut, M. (2007) Facial expressions classification: A new approach based on transferable belief model. *International Journal of Approximate Reasoning* 46(3), 542-567, December.
[3] Ekman, P. and Rosenberg, E. (1997) What the face reveals. Oxford university, New York.
[4] Ekman, P. and Rosenberg, E.L. (2005) What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system. 2nd edition, Oxford University Press.
[5] Roisman, G.I., Tsai, J.L., Chiang, K.S. (2004) The Emotional Integration of Childhood Experience: Physiological, Facial Expressive, and Self-reported Emotional Response During the Adult Attachment Interview, Developmental Psychology, 40(5), pp. 776-789.
[6] Zeng, Z., Pantic, M., Roisman G.I. and Huang, T.S. (2008) A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, accepted for publication.
[7] Cohn, J.F. and Schmidt, K.L. (2004) The timing of facial motion in posed and spontaneous smiles. *Journal of Wavelets, Multiresolution & Information Processing*, 2(2), pp. 121-132.
[8] Gwen, C., Bartlett, M.S., Littlewort, G.C. and Kang, L. (2007) Faces of Pain: Automated Measurement of Spontaneous Facial Expressions of Genuine and Posed Pain. *Proceedings of ICMI 07*, November, pp. 12-15, Nagoya, Aichi, Japan.
[9] Bartlett, M.S., Littlewort, G.C., Frank, M.G., Lainscsek, C., Fasel, I. and Movellan, J.R. (2006) Automatic recognition of facial actions in spontaneous expressions., *Journal of Multimedia.*, 1(6) pp. 22-35.
[10] Valstar, M. F., Pantic, M., Ambadar Z. and Cohn, J.F. (2006) Spontaneous vs. posed facial behavior: automatic analysis of brow actions. *Proceedings of International Conference on Multimodal Interfaces ACM 06,* pp.162–170.
[11] Valstar, M.F., Gunes, H. and Pantic, M. (2007) How to Distinguish Posed from Spontaneous Smiles using Geometric Features, Proceedings of International Conference on Multimodal Interfaces ICMI 07, pp. 38-45, Nagoya, Japan, November.
[12] El Kaliouby, R., Robinson, P. (2004) Mind reading machines: automated inference of cognitive mental states from video. SMC (1), pp. 682-688.
[13] Ji, Q., Lan, P., and Looney, C. (2006) A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE System Man Cybernetic-Part A*, 36(5), pp. 862-875.
[14] Hess, U. and Kleck, R.E. (1990) Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, 20(5), pp. 369-385.

[15] Ashraf, A.B., Lucey, S., Chen, T., Prkachin, K., Solomon, P., Ambadar, Z. and Cohn, J. (2007) The painful face: Pain expression recognition using active appearance models. Proceedings of the International Conference on Multimodal Interfaces ICMI 07, pp. 9–14.

[16] Pantic, M. and Rothkrantz, L.J.M. (2000) Automatic analysis of facial expressions: The state of the art. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 22(12) pp. 1424–1445, December.

[17] Kunz, M., Scharmann, S., Hemmeter, U., Schepelmann K. and Lautenbacher S. (2007) The facial expression of pain in patients with dementia. *Pain*. December, 133(1-3):221-8.

[18] Lehr, V.T., Zeskind, P.S., Ofenstein, J.P., Cepeda, E., Warrier, I. and Aranda J,V. (2007) Neonatal facial coding system scores and spectral characteristics of infant crying during newborn circumcision. *Clin J Pain.* 23(5), pp. 417-24.

[19] Lints-Martindale, A.C., Hadjistavropoulos, T., Barber, B., Gibso, S.J. (2007) A psychophysical investigation of the facial action coding system as an index of pain variability among older adults with and without Alzheimer's disease. *Pain,* 8(8), pp. 678-89.

[20] Pantic, M. and Bartlett, M.S. (2007) Machine Analysis of Facial Expressions, in Face Recognition, K. Delac & M. Grgic, Eds., Vienna, Austria: I-Tech Education and Publishing, pp. 377-416.

[21] Valet, L., Mauris, G., and Bolon, P. (2000) A statistic overview of a recent literature in information fusion. *Procedings of Information Fusion ICIF 00*, pp. 22-29, Paris, France.

[22] Fasel, B. and Luettin, J. (2003) Automatic facial expression analysis: A survey. *Pattern Recognition*, 1 (30), pp. 259–275.

[23] Hammal, Z., Martin, A. and Gosselin, F. (2007) Comparing a Transferable Belief Model Capable of Recognizing Facial Expressions with the Latest Human Data. *Proceedings of the 3 rd International Symposium on Visual Computing ISVC 07*, 26-28 November, Lake Tahoe, Nevada/California.

[24] Ramasso, E., Rombaut,M. and Pellerin, D. (2007) State filtering and change detection using tbm conflict application to human action recognition in athletics videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 17, pp. 944–949.

[25] Roy, S., Roy, C., Hammal, Z., Fiset, D., Blais, C., Jemel, B., Gosselin, F. (2008) The use of Spatio-temporal Information in decoding facial expression of emotions. *Proceedings of Vision Science Society VSS 08*, 9-14 May, Naples Grand Hotel, Floride.

[26] Roy, C., Roy, S., Fiset, D., Hammal, Z., Blais, C., Rainvill, P. and Gosselin, F. (2008) Recognizing static and dynamic facial expressions of Pain: Gaze-tracking and Bubbles experiments. *Proceedings of Vision Science Society VSS 08,* 9-14 May, Naples Grand Hotel, Floride.

[27] Gosselin, F. and Schyns, P.G. (2001) Bubbles: A technique to reveal the use of information in recognition. *Vision Research*, 41, pp. 2261–2271.

[28] Cohn, J.F. (2006) Foundations of Human Computing: Facial Expression and Emotion, *Proceedings of International Conference on Multimodal Interfaces ICMI 06*, pp. 233-238.

[30] Forbes Riley Kand Litman D. (2004) Predicting emotion in spoken dialogue from multiple knowledge sources. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL) Facial Expression Analysis, Proceedings of International Conference on Face and Gesture Recognition, pp. 46-53.

[31] Kapoor, A., Burleson, W. and Picard, R.W. Automatic prediction of frustration. International Journal of Human Computer Studies, 65(8), pp. 724-736.

[32] Litman,D.J. and Forbes Riley,K. (2004), Predicting Student Emotions in Computer Human Tutoring Dialogues. *Proceeding of the 42nd Annual Meeting of the Association for Computational Linguistics* (ACL), July laughter and speech.

[33] Smets. P. (2000) Data fusion in the transferable belief model. *Proceedings of International Conference on Information Fusion* ICIF 00, pp. 21–33, July Paris, France.

[34] Denoeux, T. (2008) Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial. Intelligence. 172(2-3) pp. 234-264.

[35] Ekman, P. and Friesen, W.V. (1978) The facial action coding system (FACS): A technique for the measurement of facial action. *Palo Alto, CA: Consulting Psychologists Press.*

[36] Roy, S., Roy, C., Fortin, I., Either-Majcher, C., Belin, P. and Gosselin, F. (2007) A dynamic facial expression database. Proceedings of Vision Sciences Society, VSS 07, Sarasota, Florida.

[37] Hammal, Z., Eveno, N., Caplier, A., and Coulon, P.Y. (2006) Parametric models for facial features segmentation. Signal processing, 86, pp. 399–413.

[38] Lucas, B.D. and Kanade, T. (1981) An iterative image registration technique with an application to stereo vision. Proceedings of International Joint Conference on Artificial Intelligence, 18(11), pp. 674–680, August Vancouver, Canada.

[39] Cohn Kanade database. (2000). http://vasc.ri.cmu.edu/idb/html/face/facial expression.

[40] Dailey, M., Cottrell, G.W. and Reilly, J. (2001) California facial expressions (cafe). unpublished digital images, University of California, San Diego.

[41] Hammal Caplier database. (2003). http://viscog.psy.umontreal.ca/ zakia/index.htm.

[42] Prkachin, K.M. (1992). The consistency of facial expressions of pain: A comparison across modalities. Pain, 51, pp. 297–306.

[43] Smets, P. and Kruse, R. (1994) The transferable belief model. *Artificial Intelligence*, 66 pp. 191–234.

[44] Ambadar, Z., Schooler, W.S. and Cohn, J.F. (2005) The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16 (5) pp. 403–410.

[45] Williams, A.C. (2002) Facial expression of pain: An evolutionary account. Behavioral and Brain Sciences, 25(4), 439–455.

[46] Hammal. Z., Facial Features Segmentation, Analysis and Recognition of Facial Expressions using the Transferable Belief Model. PhD thesis, LIS laboratory, Grenoble, France, Jun, 2006.

[47] Ristic, B. and Smets. P. (2004) Kalman filters for tracking and classification and the transferable belief model, *Proceedings of International Conference on Information Fusion ICIF 04*, Stockholm, Sweden.

[48] Denoeux, T. (2007) Pattern Recognition and Information Fusion Using Belief Functions: Some Recent Developments. Proceedings of ECSQARU 2007.

[49] Smets, P. (2005) Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38, pp. 133-147.

[50] Young, A.W., Rowland, D., Calder, A.J., Etcoff, Seth, A. and Perrett, D.I (1997) Facial expression megamix. Cognition, 63, pp. 271-313.