Pattern Recognition 45 (2012) 1265–1280

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Pain monitoring: A dynamic and context-sensitive system

Zakia Hammal [a,*], Miriam Kunz [b]

[a] Brain, Music, and Sound Laboratory (BRAMS), BRAMS - Suite 0-120, Pavillon 1420 Boul. Mont Royal, Université de Montréal, C.P. 6128—Station Centre Ville, Montréal (Que.), Canada H3C 3J7

[b] Centre de Recherche, Institut Universitaire de Gériatrie de Montréal, 4565, Chemin Queen-Mary, Montréal (Que.), Canada H3W 1W5

ARTICLE INFO

Article history:
Received 28 February 2010
Received in revised form
8 August 2011
Accepted 24 September 2011
Available online 5 October 2011

Keywords:
Pain expression
Spontaneous facial expressions
Context based recognition
Transferable Belief Model
Classification

ABSTRACT

The current paper presents an automatic and context sensitive system for the dynamic recognition of pain expression among the six basic facial expressions and neutral on acted and spontaneous sequences. A machine learning approach based on the Transferable Belief Model, successfully used previously to categorize the six basic facial expressions in static images [2,61], is extended in the current paper for the automatic and dynamic recognition of pain expression from video sequences in a hospital context application. The originality of the proposed method is the use of the dynamic information for the recognition of pain expression and the combination of different sensors, permanent facial features behavior, transient features behavior, and the context of the study, using the same fusion model. Experimental results, on 2-alternative forced choices and, for the first time, on 8-alternative forced choices (i.e. pain expression is classified among seven other facial expressions), show good classification rates even in the case of spontaneous pain sequences. The mean classification rates on acted and spontaneous data reach 81.2% and 84.5% for the 2-alternative and 8-alternative forced choices, respectively. Moreover, the system performances compare favorably to the human observer rates (76%), and lead to the same doubt states in the case of blend expressions.

1. Introduction

To be more efficient, the next generation of computer systems needs to integrate the ability to detect and interpret human affective states, and to be capable of adapting its responses accordingly. Several research areas could benefit from such systems: interactive games with the ability to adapt to the user's intentions; interactive teaching systems that permit teachers to be aware of student stress and inattention; and medical tools for automatic diagnosis and monitoring. In this paper we will focus on the development of automatic pain assessment tools, which could be of considerable value in certain circumstances where patients are not able to communicate pain verbally (e.g. newborns, individuals with severe cognitive impairments like autism). Because of the potential of such systems in everyday applications, the past 20 years has witnessed an increasing number of efforts for automatic recognition of human facial expressions. However, until the beginning of the year 2000 most of the affect recognition methods attempted to recognize a small set of prototypic and acted expressions like happiness, surprise, fear, disgust, anger, and sadness [16,22,20]. Given the limited potential of such expressions in real applications, the last years have seen a shift within the affect recognition community towards non-prototypic (e.g. attentiveness [12], fatigue [13,14], frustration [64], and pain [8,15,51]) and spontaneous affect recognition [6–10].

1.1. Context of the study

In the current paper, we are interested in the automatic recognition of pain expression in a medical context application, where videotaped adult subjects are not able to verbally express there own painful state. We focused on automatic recognition of pain facial expression due to its potential medical significance. For example, it can be used as a pain assessment tool for individuals who are not able to communicate pain verbally (e.g. newborns [67], individuals with severe cognitive impairments like autism [18,19]). Automatic recognition of pain from facial expression provides an alternative to common signs of pain, such as verbal reports, and provides information that may be more useful and valid in certain circumstances (for example schizophrenic patients) [17]. For this, we propose a new contribution for the automatic recognition of pain based on facial information in a medical context application.

Significant efforts have been made in human behavior studies to identify reliable and valid facial indicators of pain [56,1]. The standard method requires manual labeling of facial Action Units

* Corresponding author. Tel.: +1 514 343 6111x4966.
E-mail address: zakia_hammal@yahoo.fr (Z. Hammal).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.09.014

(AUs) [53] or other observational measurements by highly trained observers [53]. However, the main limitation of these methods is the time required for coding, which is accomplished manually by human experts. It takes over 100 h of training to become a proficient coder, and it takes approximately 2 h for such a person to code each minute of video [53]. This limitation makes the standard method ([1] for example) non-suitable for the real-time situations described previously. An emerging alternative is a fully automatic facial image analysis using computer vision.

1.2. Automatic recognition of pain expression in computer Vision

In the current paper we focus our work on the automatic recognition of the facial expression of pain on adult subjects. Few models have been developed for the automatic recognition of pain expression and none have tried to recognize pain expression among the other six basic facial expressions and neutral (i.e. a 8-alternative forced choice).

Monwar et al. [51] proposed a video analysis technique for the recognition of the facial expression of acted pain sequences in a 2-alternative forced choice classification (pain vs. painless). The face is first detected using skin color. Then, the centers of the eyes, the inner end points of the eyebrows, and the two corners of the mouth are detected using an iterative threshold method [54,55]. The extracted features are then transformed into a set of positions and shape parameters (21 features in total). Finally, these features are used as inputs to an artificial neural network trained to classify faces expressing pain or not. The authors reported several problems on face detection (false alarm) due to the use of a skin color detection method. They also reported problems with aged people. However, they did not explain the limitation and precision of their feature detection method and its consequences on the expression classification process. The authors reported a recognition rate between 45.7% and 87.99% according to the used number of hidden layer units. However they did not report the exact number of subjects used for the training and for the tests of their neural network.

Littlewort et al. [8] proposed an automatic system to discriminate real from posed facial expression of pain in video sequences. Twenty-six participants were videotaped under three experimental conditions: baseline, posed pain, and real pain [15]. The pain condition consisted of cold pressor pain induced by immersing an arm in cold water. The faces were automatically detected using the boosting based method proposed by Fasel et al. [57]. The detected faces were then rescaled, aligned, and passed through a bank of Gabor filters, leading to a collection of features. The obtained features (72 features in total) are then inputted in 20 linear support vector machines (SVMs) trained to detect the presence of 20 Action Units at each frame. After the detection of the Action Units in all frames, the videos are broken down into overlapping segments of 500 frames. For each segment, five statistical parameters are measured to characterize each one of the detected AUs [8]. The obtained statistical parameters are inputted into a nonlinear SVM trained to dissociate baseline and faces expressing pain (2-forced choice classification); the SVM is further trained to dissociate real and posed expressions of pain. However, the authors only report the classification of only real vs. faked pain expressions without reporting the classification performances of pain expression from the baseline.

In Ashraf et al. [15], 21 adult patients with rotator cuff injury were video-recorded while a physiotherapist manipulated their affected and unaffected shoulder. The authors proposed automatic pain expression recognition in a 2-alternative forced choice classification. The Active Appearance Models (AAMs) were used to extract shape and appearance parameters from the digitized face images (100 features in total). SVMs were used with several representations from the AAM to recognize painful and painless faces. In order to recognize video images as a pain sequence or not, the output scores of individual frames are summed together to give a cumulative score for the entire sequence. A decision threshold is then used to determine if the current sequence corresponds to pain expression or not. A leave-one-subject-out validation leads to a Hit rate of 81%.

The few proposed methods for the recognition of facial expression of pain have several limitations. First, these models are based only on static classification or, at best, a combination of static classification results. None of them take explicitly into account the temporal dynamic of the facial features and their asynchronous deformation during a sequence of expression [7]. Second, automatic facial expression recognition implies face and facial features segmentation where noisy and partial data should be expected. Therefore, facial expression analyzers should take into account noisy and partial data and should generate their conclusions with a confidence that reflects the certainty of face and facial features tracking [20]. None of the already proposed methods reported the uncertainty of the used features nor explicitly modeled it in the classification process. Third, the proposed methods focused on a 2-alternative forced choice classification (pain vs. no pain or real vs. acted pain), making it difficult to evaluate whether the proposed models can dissociate pain from the other basic facial expressions. In addition, these models skipped the problem of doubt between multiple expressions. Fourth, the existing models are context-independent. Few attempts have been made towards context dependent interpretation of the observed facial expression [13,30–32]. 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 undergoing task, the other people involved, and the identity and natural expressiveness of the individual [6]. Unless mistaken, none of the already vision-based models introduce one of the possible context variables for the recognition of facial expressions. The proposed model aims at overcoming these four limitations.

1.3. Proposed contribution

In the current paper we are interested in pain expression recognition from the six basic facial expressions (happiness, surprise, disgust, anger, fear, and sadness) and neutral. The classification is made in video sequences based on a dynamic fusion process of facial information in a medical context application (i.e. pain monitoring). To this end, we extended our previous model using a static approach (proposed for the recognition of the six basic facial expressions plus neutral [2]) and applied a new temporal and multi-cues modeling process to the task of automatic and dynamic recognition of pain facial expression.

Facial expressions (such as pain) induce the contraction of facial muscles, which result in temporally deformed facial features such eyes, eyebrows, lips, and skin texture, often revealed by wrinkles. Moreover, in a recent study, Roy et al. [25,26] have made a finer and less biased analysis of the importance of facial features for the discrimination of the basic facial expressions including 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. Therefore, in addition to the permanent facial features (like eyes, eyebrows, and mouth), nasal root wrinkles are introduced as a refinement sensor for the automatic recognition of pain expression. Our choice has been driven by the behavioral experiments presented in [25], which reported the used features among the most important features for the classification of the six basic facial expression and neutral, and reported by our own achievements [61] as the necessary features for the classification of the six basic facial expressions and neutral.

The temporal dynamics of facial behavior represent a critical factor for categorization of complex behaviors like pain, shame, and amusement [4]. A recent study by Cohn et al. showed as well that spontaneous smiles, in contrast to posed smiles, can have multiple apices (see Fig. 8) and are accompanied by other facial feature deformations that appear either simultaneously with the mouth corners rising or following it within 1 s [7]. Another study by Valstar et al. showed also that the order and timing of spontaneous and deliberately displayed eyebrow actions are different [10]. Based on these findings, facial expression can be described as the result of progressive deformations of a set of facial features appearing at different times (asynchronously) without any defined appearance order [46]. Using only aspects of temporal dynamics of facial expression such as the speed of a facial point displacement or the persistence of facial parameters over time [20] or including the information at time $t-1$ [65] is not sufficient. Compared to the already proposed models, and in addition to the information at time $t-1$, the asynchronous facial feature behavior is explicitly modeled in the current paper, by taking into account at each time (each frame) all the available information (facial feature states) from the beginning until the end of the sequence, to recognize the current facial expression.

In addition to the visual cues and their dynamic behavior, without context, even humans may misunderstand the observed facial expression. For example, as reported by [20] smiling in the context of downward head pitch communicates embarrassment rather than joy. Then, an important related issue that should be addressed in the recognition of all emotions is how to take into account the context information. Indeed, the context dependency remains an unexplored area for automatic facial expression recognition given the difficulty of modeling this variable and of its introduction in the classification process [20]. The current paper introduces the concept of context application as new information, which bias the classifier toward the most relevant expression (i.e. the context tell us which expressions are more likely to occur). A main advantage of the dynamic fusion process proposed in the current paper is the ability to introduce the context as a new refinement sensor in the Transferable Belief Model (TBM) based model.

1.4. Justification of the fusion based model

The proposed fusion process is based on the Transferable Belief Model (TBM) [43]. The TBM is a formal framework of increasing interest for imprecise and uncertain data analysis ([69,80,81]), information fusion ([70–72]), and pattern recognition ([75–77]). It has been used in several applications such as image processing, geosciences, medicine, robotics, and defense ([82]) and more recently in the analysis and the recognition of human facial and body behavior ([2,61,79]). The TBM based modeling has already proved its suitability for the static classification of the six basic facial expressions and neutral [2] even in the case of partially occluded facial parts [61]. The effectiveness of the TBM based approach for the purpose of facial expression classification has been also showed by its explicit comparison with the Bayesian and HMM's formalisms [2,5]. The TBM based modeling for facial expression recognition is of great interest for two main reasons. Firstly it deals with imprecise and uncertain information, which could be the case with data measures resulting from video based segmentation algorithm. Indeed, the TBM based model allows the explicit modeling of the doubt on the sensor states (i.e. the facial feature states) and generates its conclusions so that the associated certainty varies with the certainty of the facial features detection. This is an important characteristic that has been reported as one of the most important limit of the existing approaches for facial expression recognition (see [6,20]). Secondly, the TBM based model allows an explicit modeling of blend expressions, which result from the combinations and uncertainty between two or more facial expressions [61]. Indeed, human facial expressions are not always binary; blend expressions often appear and lead to doubt between facial expressions even by human observers [50]. Compared to the already reported models, the TBM based model allows dealing with these considerations.

To summarize, the proposed work is a new development of a previously proposed model [2] significantly modified here for the automatic and dynamic recognition of pain expression in video sequences. The originality of the proposed method compared to the already proposed models for the recognition of pain expression lies in: (1) the automatic fusion of different sensors information: the permanent facial feature deformations (i.e., eyes, eyebrows, and mouth), the transient features information (i.e., nasal root wrinkles), and the context of the expression production (i.e. the place like a hospital); (2) dynamic and progressive fusion (i.e. taking into account asynchronous facial feature deformations) of the permanent facial feature behavior rather than combining the result of static classification for the recognition of pain expression in a video sequence; (3) modeling explicitly the imprecision and the uncertainty of the sensors detection in the fusion process based on the dynamic modeling of the TBM [33,34]; and (4) the recognition of pain expression among the other six basic facial expressions as well as neutral.

2. Automatic system

The current paper proposes a new development of a previously proposed model for static classification of the six basic facial expressions and neural [2,61]. In the current paper, the previous model is generalized for the dynamic and progressive recognition of pain expression in video sequences using a dynamic fusion of facial and contextual information. Fig. 1 shows an overview of the proposed model where the dashed square corresponds to the new developments proposed in the current paper. First, in order to retrieve asynchronous information on the permanent facial feature states, a temporal refinement process is introduced (temporal evolution model and the asynchronous information); second, the detection and analysis of the transient facial features (e.g. nasal root wrinkles); and third, the introduction of the contextual information (e.g. place).

The first step of the proposed model is the automatic detection of the facial features (i.e. eyes, eyebrows, mouth, and wrinkles) and the context and then the representation of the corresponding behavior.

2.1. Features detection

2.1.1. Permanent facial features

The first step of the facial expression model [2] is the extraction of the contours of the permanent facial features (eyes, eyebrows, and mouth, a detailed description can be found in [37]). To cope with the illumination variation, a preprocessing stage based on a model of the human retina [37] is applied. The used retinal prefiltering enhances the contours and at the same time, realizes a local correction of the illumination variations (see Fig. 2).

A specific parametric model is then defined for each deformable feature (eyes, eyebrows, and mouth). Several characteristic points are extracted in the image to be processed to initialize each model (for example, eye corners, mouth corners, and eyebrow 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,

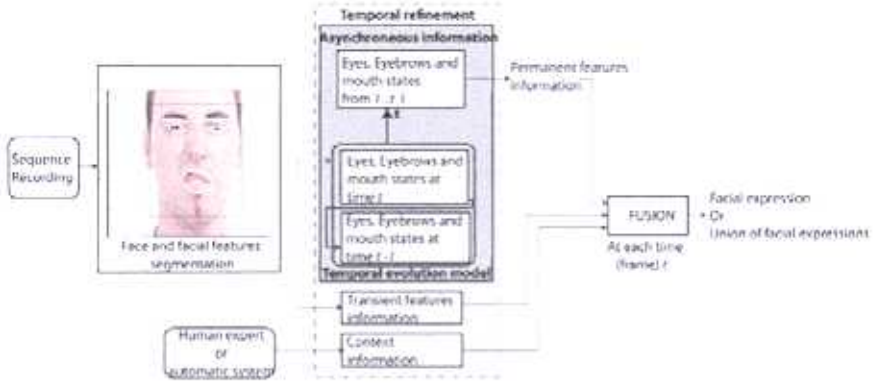

Fig. 1. System overview; dashed square gathers the new developments of the TBM based model for the recognition of pain facial expression compared to [2,61].

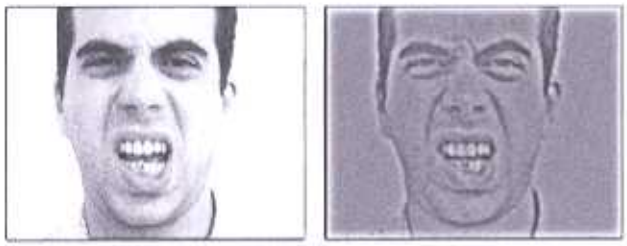

Fig. 2. Example of retinal prefiltering used for luminance correction [37].

the eyes, and the eyebrows (Fig. 3 shows an example of the obtained contours during happiness, surprise, and disgust facial expressions). Intensive tests (a comparison with a manual ground truth) on several databases under various face sizes (corresponding to iris radius between 4 and 13 pixels) presenting a large range of varying conditions show the accuracy and the robustness of the method to spectacles, luminance conditions, ethnicity, and facial expression deformations [37].

In the current paper, based on the segmentation results of the first frame, the set of characteristic points corresponding to the detected facial features (see Fig. 6 left) are 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 [37].

2.1.2. Transient facial features

Based on the eye corner positions (Fig. 6 left) and a set of morphological constraints the transient facial feature area is first located (see Fig. 4). Inside each selected area the Canny edge detector is applied for the automatic detection of the wrinkles appearance. The presence or absence of wrinkles is made by comparing the number of edge points in the current expressive image with the number of edge points of a neutral facial image. If there are about twice more edge points in the current image than in the neutral image (neutral expression), wrinkles are considered to be present. The Canny edge threshold is set by expertise but is kept constant over all the databases and at a high value in order to minimize the risk of errors. In Section 3.2.2 we will describe how the proposed TBM based model allows the system to report doubt instead of taking the chance of making a wrong decision in the case of uncertainty on the appearance of wrinkles.

2.1.3. Context information

Two main issues need to be addressed during the introduction of the context of the application: (1) the context source detection and (2) the use of the context related information. It should be stressed that this work does not focus on how to extract context information, which can be performed by a complementary system from other research fields, but rather shows how the output of such a system can be easily added as a new sensor to define which expressions are more likely to occur (i.e. the context changes the a-priori probabilities). As a knowledge modeling process the proposed TBM based model can easily model several context variables and the corresponding information for facial expression classification (reported by [20] as the most important issue at introducing the context in the classification process). The purpose of the current model is the automatic recognition of pain expression of videotaped people in a medical environment. For this end, several context variables can be defined (such as the place, the task, the answer to a written questionnaire, etc.) allowing the elimination of a set of unexpected facial expressions. Considering a hospital context application (videotaped waiting room, aged people under camera monitoring, injured people), the aim of the proposed work is to automatically alert medical personnel – physician or nurse – that the patient in charge is suffering pain when this one is unable to report his pain by himself (see Fig. 5). The place variable is the most general and suitable context variable for our purpose. The place variable is defined with two values, *medical* or *not medical* (considering the expresser in the medical environment like hospital or not), and allows the support of the expected facial expression of pain (the modeling process of the context variable and the corresponding information are explained in Sections 2.2.2 and 3.2.2). A physician or a nurse using a simple pressure button, which changes the system parameters towards pain monitoring, can assess the medical context. Otherwise the whole set of possible facial expressions are expected without favoring any of them (see Fig. 5). The proposed modeling is only one among several possible modeling to introduce the context information. The proposed model can be easily generalized to incorporate several other context variables according to the studied facial expressions and their context dependency.

The strength of the proposed model is that the context information is easily added as a "refinement sensor" in addition to the permanent and transient facial feature behavior for the classification process. Once all required information is collected, it is incorporated into a fusion architecture based on the Transferable Belief Model.

2.2. Representation of features behavior

2.2.1. Facial feature states

Five important characteristic distances (D_1–D_5) [61] measure the permanent facial feature deformations occurring during each facial expression according to the neutral state (see Fig. 6). Each

Fig. 3. Eyes, eyebrows, and lips segmentation results under three facial expressions [37].

Transient features	Transient feature Regions description
Nasal root wrinkles	Y11 = u (y)-0.7*W; X11 = b (x)+2 Y12 = a (y)+0.4*W; X12 = d (x)-2

Fig. 4. Example of nasal root wrinkles detection. W represents the distance between the eyes corners. The system detected the presence of the nasal root wrinkles in the second and third frames of the second row.

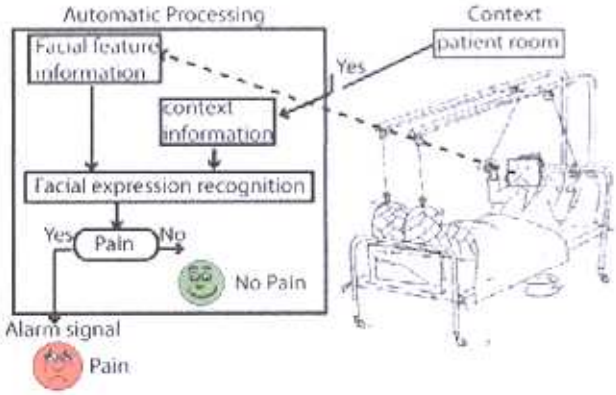

Fig. 5. Overview of the expected application in a hospital context.

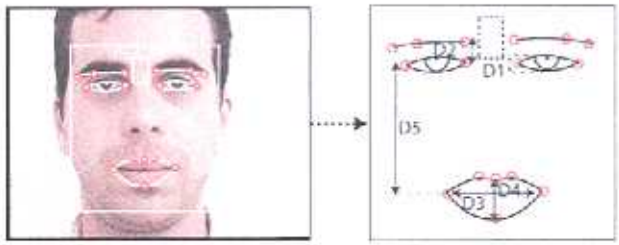

Fig. 6. Segmentation results of the permanent facial features and the corresponding characteristic distances. D_1: eye opening, distance between the upper and lower eyelids, D_2: distance between the interior corner of the eye and the interior corner of the eyebrow, D_3: mouth opening width, distance between left and right mouth corners, D_4: mouth opening height, distance between upper and lower lips; D_5: distance between a corner of the mouth and the corresponding external eye corner [2].

distance is normalized with respect to the distance between the centers of both irises in the analyzed face. This makes the analysis independent of the variability of face dimensions and of the position of the face with respect to the camera. In the present work, we assume that facial expressions and then the corresponding facial feature behaviors are symmetrical. Each characteristic

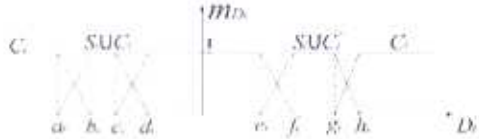

Fig. 7. Model of basic belief assignment based on the characteristic distance D_i. $S_i \cup C_i^+$ (vs. $S_i \cup C_i^-$) means S_i or C_i^+ (vs. S_i or C_i^-) and corresponds to the doubt between S_i and C_i^+ (vs. S_i and C_i^-). For each value of D_i, the sum of the pieces of evidence of the states of D_i is equal to 1.

Table 1
Characteristic distance states during pain expression.

	D_1	D_2	D_3	D_4	D_5
Pain	C^-	C^-	C^+	C^+ Or C^-	C^-
neutral	S	S	S	S	S

distance value D_i is then considered as the mean of its corresponding left and right side values. The relative importance of each of the five characteristic distances for the recognition of each one of the basic facial expressions and neutral has been recently examined (a detailed description of the specific contribution of each one of the characteristic distances can be found in [61]). The obtained results highlighted that the five characteristic distances are necessary for the classification of the six facial expressions and neutral, and then also for dissociating them from pain expression.

A numerical to symbolic conversion is carried out using a fuzzy-like model for each characteristic distance D_i (see Fig. 7). It allows the conversion of each numerical value to a belief in five symbolic states reflecting the magnitude of the deformation according to the neutral state. 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 or C_i^+ (vs. S_i or 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. This conversion process has been validated on the three-benchmark databases (the Cohn–Kanade database [39], the CAFE database [40], and the Hammal–Caplier database [41]) for the recognition of the six basic facial expressions plus neutral [2,61]. Based on the same mapping, the analysis of the behavior of each characteristic distance (i.e. the permanent facial feature behavior) during pain expression sequences led to the definition of the rules (Table 1). This mapping has been obtained from the pain expression sequences validated by human observers on the STOIC database [36]. Different rules have been evaluated (i.e., with the corresponding inference model). From all these models the most efficient rules (leading to the best performances) have been selected. The rules correspond to the facial feature deformations leading to the maximization of the correlation between the human and the system performances (see Table 1). In the current paper we reported the behavior of the characteristic distances for

Table 2
Transient feature and context states during pain expression.

	Pain	Anger	Disgust	Happiness	Surprise	Fear	Sadness	Neutral
Nasal root wrinkles	*P*	*P*	*P*	*A*	*A*	*A*	*A*	*A*
Context	*MC*	*MC* Or *NMC*	*MC* Or *NMC*	*MC* Or *NMC*	*MC* Or *NMC*	*MC* Or *NMC*	*MC* Or *NMC*	*MC* Or *NMC*

only pain expression. Their behavior (and then the corresponding characteristic distance states) during the six basic facial expressions has been already reported in previous publications (the interested reader can refer to [2,61] for the detailed description of the rules of the six basic facial expressions).

In addition to the permanent facial features, the analysis of the behavior of the nasal root wrinkles (i.e., the number of pixels detected as wrinkles in each selected area) during the six basic facial expressions and pain compared to the neutral state is also made. Their behavior shows that they can have three characteristic states: *P* (meaning present) if the number of pixels detected as wrinkles is significantly higher then their number in the neutral state; *A* (meaning absent) if their number is roughly equal to its corresponding value in the neutral state; *P* or *A* if their number is neither sufficiently higher, to be in *P*, nor sufficiently stable to be in *A*. The nasal root states during pain expression are reported in Table 2.

2.2.2. Context variable states

The context information is used to bias the classifier toward the most relevant expression. In the current application, the context is medical and pain expression is the more relevant expression. The context information is coded with two states: *MC* if the videotaped sequence is in a medical context, for example a "hospital," and *NMC* if not. In the current modeling, the context information allows refining the already obtained classification results (based on the facial sensors) where pain is already identified alone or where the system hesitates between it and another expression, discarding irrelevant facial expressions (see Section 4). In the context of pain monitoring (videotaped waiting room, injured people), we deliberately choose to take a risk of having a false alarm in the case of doubt between pain and another expression (by biasing the classifier toward the most relevant expression, i.e. pain) rather than missing a painful state. In this case an alert is sent and a human observer (physician or nurse) can then confirm or not this information. The proposed modeling is the first attempt to introduce context information. This modeling can be however easily generalized to several other context variables (i.e., MC and NMC variables can be replaced by several other variables) as a selection sensors of relevant expressions (i.e., different expressions according to each context variable) according to the applications (e.g., context-based recognition of different kind of smiles) by a simple addition of the corresponding rules in Table 2.

At this point, pain expression is modeled by a set of rules according to the sensor states (see Tables 1 and 2). It should be stressed, however, that humans are not all or nothing, be it for the production or the recognition of emotional facial expressions [50]. Pain expression as all facial expressions may include a blend of expressions, which makes human observers often hesitating between several expressions (see Section 4.2.2.2). Moreover, automatic facial features segmentation can lead to measurement errors on the characteristic distance states as well as on the transient feature states. For these reasons, an all-or-nothing system based only on logical rules is not sufficient to reliably recognize facial expressions. These issues can be directly tackled by the TBM [33] modeling process of all the knowledge resulting from the characteristic distance states, the nasal root wrinkles, and the context information.

3. Knowledge representation based on the Transferable Belief Model

In information fusion, it is usually important to take into account the reliability of the different sources in the evidence aggregation process [70]. Indeed, in a realistic interaction environment, a facial expression analyzer should be able to deal with noisy and partial data and to generate its conclusions confidently with respect to the data. Compared to the already reported model for facial expression recognition the TBM is a model for reasoning that is well suited to represent imprecise and uncertain information using belief functions [43]. The TBM can be considered as an extension of Bayes's theorem in which probability measures are replaced by belief functions, and where no prior knowledge is assumed [75]. The TBM deals with imprecise and uncertain information by the explicit definition of doubt associated to the sensors information [2]. It provides a theoretical framework for the combination of various independent sensors or sources of information to obtain a more reliable decision [33,34]. In the following, we will introduce the main concepts of this theory and some essential notions that will be used in the proposed approach.

3.1. General concepts

The TBM modeling requires the definition of the basic belief assignment (BBA) associated to each independent source of information.

3.1.1. Basic belief assignment

Let x be a variable taking values in a finite set $\Omega=\{H_1,\ldots,H_N\}$ of N exclusive and exhaustive hypotheses called frame of discernment and characterizing some situations. This means that the solution is unique and is one of the hypotheses of Ω. In our case, the hypotheses correspond to one of the 8 facial expressions and the space of discernment is

$$\Omega=\{happiness(E_1),\ surprise(E_2),\ disgust(E_3),\ fear(E_4),\ anger(E_5),\ sadness\ (E_6),\ pain(E_7),\ neutral(E_8)\}.$$

The belief related to the variable x is modeled by a basic belief assignment function (BBA) that assigns a piece of evidence (the equivalent of the probability in the Bayesian formalism) to the proposition A of the power set: $2^{\Omega}=\{A/A\subseteq\Omega\}=\{\phi,\{H_1\},\{H_2\},\ldots,\{H_N\},\{H_1,H_2\},\ldots,\Omega\}$, such as

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

where $m^{\Omega}(A)$ is the part of belief on the proposition $A\in 2^{\Omega}$ without favoring any of the propositions of A. This is the main difference with the Bayesian model, which implies equiprobability of the propositions of A. The subsets (for instance $\{H_1\}$) correspond to the logical propositions (for instance H_1 is true). In the following, the logical propositions are used instead of the subset, which means that the notation $\{H_1\}$ will be simplified to H_1 and $\{H_1,H_2\}$ to $H_1\cup H_2$ (i.e. $H_1 or H_2$). A is called the focal element of $m^{\Omega}(A)$ whenever $m^{\Omega}(A)>0$. The total ignorance is represented by the BBA $m^{\Omega}(\Omega)=1$. Moreover,

with TBM, the BBAs are not required to be normalized, i.e. we may have $m^{\Omega}(\phi) \neq 0$ (ϕ) being the conflict between the sensors).

3.1.2. Dempster's rule of combination

The main feature of TBM is the set of powerful combination methods to fuse information from different sensors. The usual way of combining the beliefs resulting from each source of information (in the current case the visual cue and context states) is the Dempster's rule of combination (conjunctive sum operation, see Eq. (10) ([33,34]):

$$m^{\Omega} = \oplus\, m^{\Omega}_{Cues} \tag{2}$$

In the current application $Cues \in \{D_i, TF, CT\}$, $1 \leq i \leq 5$.

For example, if we consider two BBAs m^{Ω}_1 and m^{Ω}_2 of the same frame of discernment 2^{Ω}. The joint belief m^{Ω}_{12} is given using the conjunctive combination (orthogonal sum) as

$$m^{\Omega}_{12}(A) = m^{\Omega}_1(B) \oplus m^{\Omega}_2(C) = \sum_{B \cap C = A} m^{\Omega}_1(B)\, m^{\Omega}_2(C) \tag{3}$$

where A, B, and C correspond to hypotheses or subset of hypotheses.

3.1.3. Decision process

The decision is the ultimate step and consists in making a choice between various hypotheses (i.e. E_e) and their possible combinations. However making a choice means taking a risk, except if the result of the combination is perfectly reliable: (i.e. $m(E_e)=1$). As this is not always the case, several classical criteria can be used: the plausibility, which favors the single hypotheses in the case of mixture of expressions, and the belief, which favors the mixture of hypotheses [33]. However, in some applications it is necessary to make a decision and choose the most reliable single hypothesis. To do so, Smets proposed the use of the Pignistic transformation defined as

$$BetP : \Omega \rightarrow [0,1]$$

$$I \rightarrow BetP(I) = \sum_{H \subseteq \Omega, I \in H} \frac{m^{\Omega}(H)}{(1-m^{\Omega}(\phi))Card(H)} \quad \forall I \in \Omega \tag{4}$$

where $BetP(I)$ corresponds to the Pignistic probability of each of the hypothesis I of H, ϕ corresponds to the conflict between the sensors, and $Card(H)$ corresponds to the number of elements (hypothesis) of H.

3.2. TBM based modeling

The first step of the TBM modeling process is the definition of the BBA of each sensor (characteristic distance states, transient feature states, and context state) and then the combination of these BBAs leading to the definition of the BBAs of the facial expressions.

3.2.1. Basic belief assignment of the transient feature states

In this case the BBA of the nasal root wrinkle states is defined as $m^{\Omega_{TF}}_{TF}$ (see Eq. (1)) where $\Omega_{TF} = \{P, A\}$. From the power set 2^{Ω} only the state P (we are sure that the wrinkles are present) and the state $P \cup A$ ("we don't know", corresponding to the doubt on their state) are considered leading to the subset $2^{\Omega_{TF}} = \{P, P \cup A\}$. The wrinkles detection threshold has been derived by statistical analysis on the Hammal–Caplier and the STOIC databases [41,36]. It is defined as the mean of the ratio of the number of the wrinkle pixels when the nasal root wrinkles are present (ground truth) compared to their number in the case of neutral state. Thus, if the number of wrinkle pixels is higher than the threshold, the system is sure that the nasal root wrinkles are present ($m^{\Omega_{TF}}_{TF}(P) = 1$); otherwise the system keeps the doubt instead of taking the chance of making a wrong decision ($m^{\Omega_{TF}}_{TF}(P \cup A) = 1$). Indeed, due to the difficulty of defining a reliable threshold for the intermediate values between the two classes P and $P \cup A$, binary beliefs are chosen with a high and precise threshold for the detection of the transient features. This choice can lead to some miss detections of the transient features. These miss detections will eventually lead to keeping the doubt between several expected facial expressions instead of taking the risk of making a wrong decision by eliminating one expected expression because of a false detection of the transient feature. The nasal root wrinkles are used for pain identification as reported in Table 2. Their presence allows a refinement of the classification eliminating the expressions happiness, surprise, fear, sadness, and neutral, reducing the number of the possible expressions to 3 rather than 8 as follows:

- if the nasal roots are present the current expression is pain or anger or disgust (without favoring any of them, see Table 2) and the corresponding piece of evidence is computed as

$$m^{\Omega_{TF}}_{TF}(P) = m^{\Omega_{TF}}_{TF}(pain \cup anger \cup disgust) = 1$$

- if they are absent the current expression is one of the 8 expressions and the corresponding piece of evidence is computed as

$$m^{\Omega_{TF}}_{TF}(P \cup A) = m^{\Omega_{TF}}_{TF}(pain \cup anger \cup disgust \cup happy \cup surprise \cup fear \cup sadness \cup neutral) = 1$$

3.2.2. Basic belief assignment of the context variable states

In this case the BBA of the contextual variable is defined as $m^{\Omega_{CT}}_{CT}$ (see Eq. (1)) where $\Omega_{CT} = \{MC, NMC\}$, MC means medical context (the expresser is in a medical context and it is more likely that the expected expression corresponds to pain) and NMC means not medical context, $MC \cup NMC$ means that the context of the expresser is unknown. From the power set $2^{\Omega_{CT}}$ only the state MC (the current context is medical) and the state $MC \cup NMC$ (context unknown) are taken into account, leading to the subset $2^{\Omega_{CT}} = \{MC, MC \cup NMC\}$. In the current modeling the context variable is discrete, then, if the context is medical, the piece of evidence of the state MC is equal to 1 ($m^{\Omega_{CT}}_{CT}(MC) = 1$); otherwise the piece of evidence of the state $MC \cup NMC$ is equal to 1 ($m^{\Omega_{CT}}_{CT}(MC \cup NMC) = 1$). The context information is used for pain selection as reported in Table 2. The piece of evidence of the corresponding expressions is computed as follows:

1. if the context of the application is medical, and then the aim is to know if the current videotaped expresser is in pain or not, the piece of evidence of the corresponding expression is

$$m^{\Omega_{CT}}_{CT}(MC) = m^{\Omega_{CT}}_{CT}(pain) = 1$$

2. otherwise the piece of evidence of the corresponding expressions is

$$m^{\Omega_{CT}}_{CT}(MC \cup NMC) = m^{\Omega_{CT}}_{CT}(pain \cup anger \cup disgust \cup happy \cup surprise \cup fear \cup sadness \cup neutral) = 1$$

3.2.3. Basic belief assignment of the characteristic distance states

The BBA of each characteristic distance states D_i is defined as $m^{\Omega_{D_i}}_{D_i}$ (see Eq. (1)), where $\Omega_{D_i} = \{C_i^+, C_i^-, S_i\}$. $2^{\Omega_{D_i}} = \{C_i^+, C_i^-, S_i, S_i \cup C_i^+, S_i \cup C_i^-, C_i^+ \cup C_i^-, S_i \cup C_i^+ \cup C_i^-\}$ is the set of possible focal elements. The states $C_i^+ \cup C_i^-$ and $S_i \cup C_i^+ \cup C_i^-$ are not allowed, leading to the subset $2^{\Omega_{D_i}} = \{C_i^+, C_i^-, S_i, S_i \cup C_i^+, S_i \cup C_i^-\}$. The piece of evidence $m^{\Omega_{D_i}}_{D_i}$ associated with each symbolic state, given the

value of the characteristic distance D_i, is obtained by the function depicted in Fig. 7. The normalized threshold values $\{a_i, b_i, c_i, d_i, e_i, f_i, g_i, h_i\}$ have been derived by statistical analysis on the Hammal–Caplier database [41] for each characteristic distance. More precisely, the obtained thresholds have been computed based on the normalized average of minimum, maximum, and neutral values for each characteristic distance D_i for all the facial expressions and all the subjects. The generalization of these thresholds has been already proved on several benchmark databases such as the Cohn–Kanade database [39] and the CAFFE database [40] for the recognition of the six basic facial expressions and neutral. The interested reader can find a detailed description of their computing in [2,61]. The same thresholds without re-setting are used here.

As reported in the introduction, temporal dynamics of human facial behavior is a critical factor for the interpretation of the facial expressions [7,44] and is moreover essential for the 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 introducing a dynamic TBM based model.

3.3. Temporal refinement

The temporal information is introduced at two levels: first, by taking into account at each time t the previous information (facial feature states) at time $t-1$ to prevent some miss classifications that can appear due to wrong distance states estimation (temporal evolution in Fig. 1); second, by combining at each time all the previous information from the beginning until the end of the sequence to make a decision (asynchronous information in Fig. 1) taking into account the dynamic and asynchronous behavior of the facial features during the whole sequence. This is the main difference with the previous works, which take into account only the last information at time $(t-1)$ or combine the information of a predefined number of frames [8,20,65].

3.3.1. Temporal evolution model

The temporal evolution model applies a temporal constraint on the evolution of the BBAs. The proposed model allows predicting the BBAs of the permanent facial feature states at each time t according to their BBAs at time $t-1$. In other words, the model predicts the BBAs $\dot{m}^{\Omega_{D_i}}_{D_i,t}$ of each characteristic distance at time t according to the corresponding BBAs $m^{\Omega_{D_i}}_{D_i,t-1}$ at time $t-1$. It is assumed that the two BBAs are close because the information between two consecutive frames is strongly related. Moreover, in the case of every day life application, it has been shown that spontaneous facial expressions are slower in time than posed facial expressions [10,44].

The temporal evolution model needs first the definition of the conditional pieces of evidence, gathered in a "transition matrix" for each characteristic distance. The predicted BBA is computed at each time t by the combination of a transition matrix $\overline{M}(D_i)$ and the computed BBAs at time $t-1$ in the following way:

$$\dot{m}^{\Omega_{D_i}}_{D_i,t} = \overline{M}(D_i)\, m^{\Omega_{D_i}}_{D_i,t-1} \tag{5}$$

where

$$\dot{m}^{\Omega_{D_i}}_{D_i,t} = \begin{pmatrix} m_{D_i,t}(C^+) \\ m_{D_i,t}(C^-) \\ m_{D_i,t}(S) \\ m_{D_i,t}(S\cup C^+) \\ m_{D_i,t}(S\cup C^-) \end{pmatrix}, \quad m^{\Omega_{D_i}}_{D_i,t-1} \begin{pmatrix} m_{D_i,t-1}(C^+) \\ m_{D_i,t-1}(C^-) \\ m_{D_i,t-1}(S) \\ m_{D_i,t-1}(S\cup C^+) \\ m_{D_i,t-1}(S\cup C^-) \end{pmatrix}$$

The transition matrix $\overline{M}(D_i)$ is composed of a distribution of conditional pieces of evidence [47]. These pieces of evidence are associated to the transitions 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 denoted as $m^{\Omega_{D_i}}_{D_i}[A](B)$ (A and $B \in \{S, C^+, C^-, S\cup C^+, S\cup C^-\}$). For example, for a considered distance D_i, $m^{\Omega_{D_i}}_{D_i}[S](C^+)$ corresponds to the piece of evidence (the belief) $m^{\Omega_{D_i}}_{D_i}(C^+)$ associated with the state C^+ at time t for the distance D_i given the fact that D_i was in the state S at time $t-1$. The conditional piece of evidences for each characteristic distance D_i $(1 \le i \le 5)$ is gathered in a specific *transition matrix* $\overline{M}(D_i)$ as

$$\overline{M}(D_i) = \begin{pmatrix} m^{\Omega_{D_i}}_{D_i}[S](S) & m^{\Omega_{D_i}}_{D_i}[C^+](S) & m^{\Omega_{D_i}}_{D_i}[C^-](S) & m^{\Omega_{D_i}}_{D_i}[S\cup C^+](S) & m^{\Omega_{D_i}}_{D_i}[S\cup C^-](S) \\ m^{\Omega_{D_i}}_{D_i}[S](C^+) & m^{\Omega_{D_i}}_{D_i}[C^+](C^+) & m^{\Omega_{D_i}}_{D_i}[C^-](C^+) & m^{\Omega_{D_i}}_{D_i}[S\cup C^+](C^+) & m^{\Omega_{D_i}}_{D_i}[S\cup C^-](C^+) \\ m^{\Omega_{D_i}}_{D_i}[S](C^-) & m^{\Omega_{D_i}}_{D_i}[C^+](C^-) & m^{\Omega_{D_i}}_{D_i}[C^-](C^-) & m^{\Omega_{D_i}}_{D_i}[S\cup C^+](C^-) & m^{\Omega_{D_i}}_{D_i}[S\cup C^-](C^-) \\ m^{\Omega_{D_i}}_{D_i}[S](S\cup C^+) & m^{\Omega_{D_i}}_{D_i}[C^+](S\cup C^+) & m^{\Omega_{D_i}}_{D_i}[C^-](S\cup C^+) & m^{\Omega_{D_i}}_{D_i}[S\cup C^+](S\cup C^+) & m^{\Omega_{D_i}}_{D_i}[S\cup C^-](S\cup C^+) \\ m^{\Omega_{D_i}}_{D_i}[S](S\cup C^-) & m^{\Omega_{D_i}}_{D_i}[C^+](S\cup C^-) & m^{\Omega_{D_i}}_{D_i}[C^-](S\cup C^-) & m^{\Omega_{D_i}}_{D_i}[S\cup C^+](S\cup C^-) & m^{\Omega_{D_i}}_{D_i}[S\cup C^-](S\cup C^-) \end{pmatrix}$$

where the sum of all the conditional pieces of evidence belonging to the same column is equal to 1 and the matrix dimension is $[5\times 5]$.

The piece of evidence of all the transitions are learned using the Hammal–Caplier database and have already been validated for the three expressions: happiness, surprise, and disgust [46]. More precisely, for each expression, for each subject, and each sequence the conditional basic belief assignments are defined between each two consecutive frames $(t-1, t)$ from the beginning to the end of the sequence as $(m_{D_i,1} \to m_{D_i,2})$, $(m_{D_i,2} \to m_{D_i,3})$, ..., $(m_{D_i,t-1} \to m_{D_i,t})$; $2 \le t \le N$, where N corresponds to the total number of frames per sequence.

A conditional matrix $\overline{M}^e_j(D_i)$ is thus defined for each distance D_i, for each subject j and for each expression e by

$$m^{\Omega_{D_i}}_{D_i,t} = \overline{M}^e_j(D_i)\, m^{\Omega_{D_i}}_{D_i,t-1} \tag{6}$$

Eq. (6) is defined for each transition between $(t-1, t)$. To obtain the transitions on the whole sequence, the $m^{\Omega_{D_i}}_{D_i,t}$ and $m^{\Omega_{D_i}}_{D_i,t-1}$ are concatenated as:

$$\begin{pmatrix} m^{\Omega_{D_i}}_{D_i,2}(S) & \dots & m^{\Omega_{D_i}}_{D_i,N}(S) \\ m^{\Omega_{D_i}}_{D_i,2}(C^+) & & m^{\Omega_{D_i}}_{D_i,N}(C^+) \\ \vdots & \dots & \vdots \\ m^{\Omega_{D_i}}_{D_i,2}(S\cup C^-) & \dots & m^{\Omega_{D_i}}_{D_i,N}(S\cup C^-) \end{pmatrix}^{e}_{[5\times N]}$$

$$= \overline{M}_j^e(D_i) \underbrace{\begin{pmatrix} m_{D_i,1}^{\Omega_{D_i}}(S) & \dots & m_{D_i,N-1}^{\Omega_{D_i}}(S) \\ m_{D_i,1}^{\Omega_{D_i}}(C^+) & \dots & m_{D_i,N-1}^{\Omega_{D_i}}(C^+) \\ & \dots & \\ & \dots & \\ m_{D_i,1}^{\Omega_{D_i}}(S \cup C^-) & \dots & m_{D_i,N-1}^{\Omega_{D_i}}(S \cup C^-) \end{pmatrix}^T}_{[5 \times N-1]} \quad j$$

$$\overline{M}_{j(D_i,2,\dots,N)}^e = \overline{M}_j^e(D_i)\overline{M}_{j(D_i,1,\dots,N-1)}^e \tag{7}$$

where $M_{j(D_i,2,\dots,N)}^e$ (respectively $M_{j(D_i,1,\dots,N-1)}^e$) corresponds to the BBAs of the distance D_i, for the subject j from the frame 2 until N (respectively 1 until $N-1$) for the expression e. For example, the element $m_{D_i,2}^{\Omega_{D_i}}(S)$ in $M_{j(D_i,2,\dots,N)}^e$ corresponds to the piece of evidence of the state S for the subject j in the frame 2 for the expression e.

To define the transition matrix for each distance independently of the subjects and independently of the studied expressions, the matrices $M_{j(D_i,2,\dots,N)}^e$ ($M_{j(D_i,1,\dots,N-1)}^e$) are concatenated over all the subjects j ($1 \le j \le 11$) and over all the expressions e ($e \in \{happiness, surprise, disgus, fear, sadness, anger\}$) to define the transition matrix $\overline{M}(D_i)$ ($1 \le i \le 5$) for each distance D_i as

$$\overline{M}(D_i) = M_{1,\dots,11(D_i,2,\dots,N)}^{ha,sp,dg,fe,sa,an} M_{1,\dots,11(D_i,1,\dots,N-1)}^{ha,sp,dg,fe,sa,an\ +} \tag{8}$$

where ha: happiness, sp: surprise, dg: disgust, fe: fear, sa: sadness, and an: anger.

Eq. (8) leads to an over-determined system. The matrix $\overline{M}(D_i)$ is then computed as its optimum solution in the least squares sense [46]. In the current paper, the transition matrix proves its generalization and robustness to new data such as spontaneous pain expression sequences. The transition matrix has been computed once and is used as described in the following.

At each time t, the temporal evolution processing associated with each characteristic distance consists in combining the BBAs of the characteristic distance states $m_{D_i,t}^{\Omega_{D_i}}$ at time t and their temporal prediction $\dot{m}_{D_i,t}^{\Omega_{D_i}}$ based on the previous frame at time $t-1$ using the transition matrix (see Eq. (9)). The combination of the predicted and computed BBAs is based on the conjunctive combination (see Eq. (2)) as

$$m_{D_i,t}^{\hat{\Omega}_{D_i}} = \dot{m}_{D_i,t}^{\Omega_{D_i}} \oplus m_{D_i,t}^{\Omega_{D_i}} \tag{9}$$

At each time t the BBAs of each characteristic distance D_i is then re-initialized by the resulting BBA as

$$m_{D_i,t}^{\Omega_{D_i}} \leftarrow m_{D_i,t}^{\hat{\Omega}_{D_i}} \tag{9}$$

The combination leads sometimes to a conflict between the predicted and the computed pieces of evidence. This is mainly due to segmentation errors, so in this case, the BBA obtained by prediction $\dot{m}_{D_i,t}^{\Omega_{D_i}}$ is chosen to form the BBA associated to the corresponding distance states at time t.

3.3.2. Asynchronous behaviors modeling

Facial expression is the result of progressive deformations of a set of facial features appearing at different times (asynchronously) without any defined appearance order [46]. The followed modeling deals with these considerations taking into account, in addition to the information at time $t-1$ (reported in the previous section), at each time (each frame) all the available information (BBAs of facial feature states) from the beginning until the end of the sequence to make a decision.

Spontaneous facial expressions are characterized by a *beginning*, one or more apices, and an *end* [7] (see Fig. 8). The *beginning* corresponds to the first frame where one or more facial features begin to move. The *end* corresponds to the frame where all the facial features come back to their neutral state (stable). In each expression sequence, the proposed model detects the *beginning* as the first frame where at least one of the permanent facial features (and then the corresponding characteristic distances) is no longer in the stable state S, and the *end* as the first frame (after the beginning 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 in the proposed framework, and this is still an open issue in the community, to detect in real time the apices of one expression sequence. The proposed method deals with this consideration taking into account all the available information (previously facial features deformation) between each pair of *beginning* and *end* frames to make a decision. The BBAs of the characteristic distance states are then redefined on the whole sequence and combined based on the rules table (see Table 1) to define the expression corresponding to all these deformations.

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 (see Fig. 9). At each time t (each frame) inside the

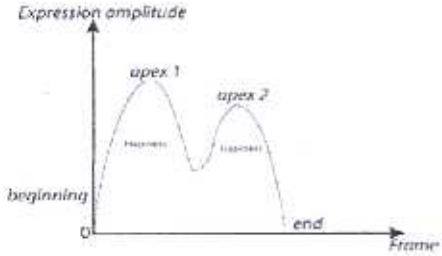

Fig. 8. Example of temporal evolution of spontaneous happiness expression.

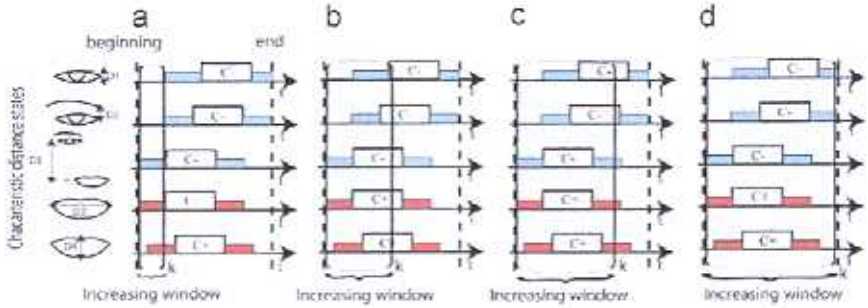

Fig. 9. (a)–(d) show examples of the increasing temporal window during a sequence of pain expression; each row displays the temporal evolution of each distance states; it can be observed that with the temporal evolution each feature evolves asynchronously but the increasing window (gray window) includes all its behavior from the beginning until the end of the sequence.

temporal window, the whole set of the previous information (the past states of the characteristic distances (computed as reported in Section 3.3.1 and then the corresponding facial feature behavior) is taken into account to classify the current expression sequence. This processing allows explicitly dealing with the dynamic of the facial expressions and more importantly with asynchronous aspect of the facial feature deformations [7]. Fig. 9 gives an example of asynchronous facial feature deformations during pain expression sequence. In this case, combining the classification results coming from the obtained information at each time separately (frame by frame) can lead to a wrong decision. Thus, in the proposed model, once the *beginning* of the expression is detected, the dynamic classification consists in computing at each time t (k in Fig. 9) the BBAs of the characteristic distance states defined on $\{C^+, C^-, S\cup C^+, S\cup C^-\}$ according to their behavior from the *beginning* until the current frame. To do that, at each time t (k in Fig. 9), inside the current increasing window Δt and for each characteristic distance, the whole of its behavior from the beginning is used to select the associated state. The selection is made according to the "integral sum of plausibility" of each state divided by its number of appearances. The number of appearances of each symbolic state is noted as $\underset{\Delta t}{Nb}(state)$, $state \in \{C^+, C^-, S\cup C^+, S\cup C^-\}$ (see Eqs. (11) and (12) for an example) and their integral (sum) of plausibility is noted $\underset{\Delta t}{Pl}(state)$ and is computed inside the temporal window Δt (see Eq. (12) for an example). The plausibility is one of the many decision criteria, which could be used in the TBM modeling framework [48,49]. It is used to choose from a set of hypothesis, or a subset of hypothesis, the most plausible one. In the current paper, the plausibility is used to reinforce the first condition (Eq. (11)) in the case where two states appear the same number of times for a given characteristic distance (see below).

For instance, for a characteristic distance D_j and for the $state = C^+$ the selection parameter is defined as

$$K_t(C^+) = \begin{cases} 1 & \text{if } m_{D_j}(C^+) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \quad 1 \le t \le \Delta t \tag{10}$$

$$\underset{\Delta t}{Nb}(C^+) = \sum_{t=1}^{\Delta t} K_t(C^+) \tag{11}$$

$$\underset{\Delta t}{Pl}(C^+) = \sum_{t=1}^{\Delta t} (m_{D_j}(C^+) + m_{D_j}(S\cup C^+)) \tag{12}$$

where K_t indicates the occurring, or not, of a symbolic state at time t.

From the two parameters $\underset{\Delta t}{Nb}(state)$ and $\underset{\Delta t}{Pl}(state)$, the distance states at each time t inside the temporal window Δt are selected as:

$$\underset{\Delta t}{state}(D_j) = \max(\underset{\Delta t}{Pl}(state)/\underset{\Delta t}{Nb}(state)) \tag{13}$$

$$1 \le j \le 5, \quad state \in \{C^+, C^-, S\cup C^+, S\cup C^-\}$$

Fig. 10 shows three examples of the temporal evolution of the states of the characteristic distance D_3 inside the increasing window and the corresponding selected states according to Eq. (13).

The piece of evidence associated to each chosen state for each characteristic distance corresponds to its maximum piece of evidence inside the current temporal increasing window as

$$m_{\underset{\Delta t}{State}D_i\Delta t} = \max(m_{D_i, \ldots, t}), \quad 1 \le i \le 5 \tag{14}$$

Finally, at each time t (see time k in Fig. 9a–d) between the *beginning* and the *end* of the expression sequence, once the BBAs of all the characteristic distances are defined (the selected states and the corresponding pieces of evidence), the current expression or the subset of possible expressions is selected (according to the permanent facial features behavior) according to the rules table (see Table 1). The obtained results are then fused to the information resulting from the nasal root wrinkles and the context variable to give the best possible decision (see Fig. 1). The main feature of this classification is that the recognition of the current expression can be made at each time t (each frame) of the sequence:

1. taking into account all the past BBAs of the characteristic distance states (and then the whole dynamic of the corresponding facial features) from the *beginning* until the current frame [46];
2. 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 by a dynamic refinement process;
3. when reaching the end (the current frame is then the last frame of the sequence), the decision depends on the whole set of past BBAs of the characteristic distance states and it gives the classification of the entire expression sequence.

3.4. Dynamic fusion process

The main feature of the TBM is the powerful combination operator [33,48] that integrates information from different sensors. In the current paper the sensors are the characteristic distance states, the nasal root wrinkles, and the context variable. The fusion requires the definition of the fused information on the same frame of discernment (Fig. 11).

From the rule tables (see Tables 1 and 2), and in order to take into account all the available information, the facial expression classification is based on the fusion process of all the used feature states (D_i, TF, CT, see Fig. 11). The BBAs of the characteristic distance states $m_{D_i}^{\Omega_D}$, the transient feature states $m_{TF}^{\Omega_{TF}}$, and the context states $m_{CT}^{\Omega_{CT}}$ 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 needs to be identified from the set of 8 possible facial expressions). From the rule tables (Tables 1 and 2) and the BBAs $m_{D_i}^{\Omega_D}$, $m_{TF}^{\Omega_{TF}}$, and $m_{CT}^{\Omega_{CT}}$ of all feature states, a set of BBAs on facial expressions $m_{D_i}^{\Omega}$, m_{TF}^{Ω}, and m_{CT}^{Ω} is first derived for each characteristic distance D_i, transient feature TF, and context

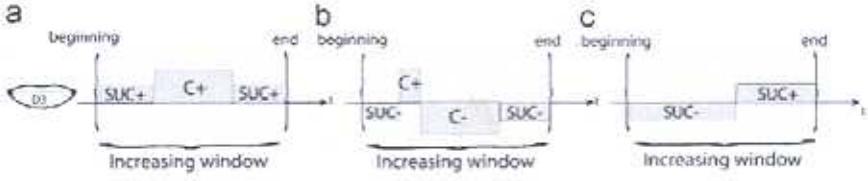

Fig. 10. (a)–(c) show examples of the selection process of the characteristic distance states inside the increasing temporal window using Eq. (13). (a) the state C^+ is chosen, (b) the state C^- is chosen, and (c) the state $S\cup C^-$ is chosen.

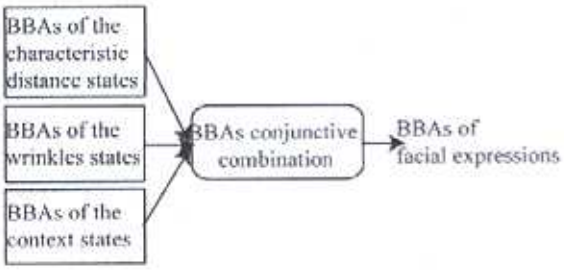

Fig. 11. Block diagram of the BBAs fusion process.

variable CT. In order to combine all this information, a fusion process of their BBAs is performed using the conjunctive combination rule [33,34] described in Section 3.1.2 (see Eq. (2)) and it results in m^{Ω}, the BBA of the corresponding expression or subset of expressions. For example, considering two BBAs $m_D^{\Omega}(E_1 \cup E_3)$ and $m_{Tr}^{\Omega}(E_3)$ of the same frame of discernment 2^{Ω}, their combination leads to the definition of the expressions E_3 corresponding to $(E_1 \cup E_3) \cap (E_1)$ with the piece of evidence $m^{\Omega}(E_1) = m_D^{\Omega}(E_1 \cup E_3) m_{Tr}^{\Omega}(E_1)$.

The decision is the ultimate step of the classification process and it allows choosing between various hypotheses E_r and their possible combinations. The Pignistic probability, which only deals with singleton expressions, is then used as decision criteria (see Section 3.1.3 Eq. (3)). The decision is made using the maximum of Pignistic probability BetP [49].

4. Experimental results

4.1. Facial expression data

Despite the great efforts in the computer vision community to build and make accessible a large number of databases of the six basic facial expressions, there is still a lack of accessible non-prototypic facial expressions databases including pain expression. In the current paper the performances of the proposed model are evaluated across two new acquired pain expression databases: an acted facial expression database (STOIC database) [36] and a spontaneous pain expression database [68]. The databases were recently recorded and ground-truthed by two-research groups from Université de Montréal and the Institut Universitaire de Gériatrie de Montréal, experts on facial expression perception [36], especially on pain expression [68]. A short description of the two databases acquisition and validation is presented in the following sections.

4.1.1. Acted pain expression database

The STOIC database (see Fig. 12) developed and validated by Roy et al. from the Université de Montréal [36] is used in the current paper. The STOIC database is one of the first dynamic facial expression databases validated by human observers (compared to the usual FACS validation process) independently of the undergoing facial expressions.

4.1.1.1. Acted data acquisition and validation. Forty students from theatrical schools aged between 20 and 45 years old were recorded. In all 7000 videos were recorded using a camera located directly in front of the subjects, who were asked to perform the six basic facial expressions as well as pain and neutral expressions [36]. Videos were captured and then aligned on a template that fitted in a box centered on the face with a resolution of 256 × 256 pixels [36]. Each video begins with the neutral state and evolves until the apex of the undergoing facial expression (see Fig. 12). Thirty-five participants (20 females and 15 males) from Montréal were recruited for the

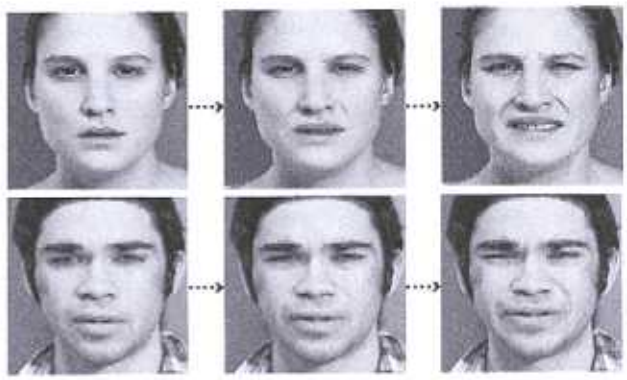

Fig. 12. Examples of pain sequences from the STOIC database [36].

validation of the recorded expressions as one of the six basic facial expressions, pain, and neutral independently of the undergoing facial expression [36]. Eighty-five videos were unambiguously recognized as the six basic facial expressions, pain, and neutral expression [36]. Fifteen videos (seven male and eight female actors) from the eighty-five videos were recognized as pain expression. Fig. 12 shows two examples of the obtained pain sequences. A detailed description of the database and a free access can be found in [http://www.mapageweb.umontreal.ca/gosselif/STOIC.rar]. The obtained pain sequences were divided into two parts: the training set (10 of the pain videos) and the test set (the remaining five pain videos). The training set was used to define the rules for pain expression recognition. The rules correspond to the facial feature deformations (and, as explained in Section 2.2.1, 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 rule based on human observer validation (see Tables 1 and 2).

4.1.2. Spontaneous pain expression database

In addition to the acted database and in order to better test the model generalization, with less controlled pain sequences, a spontaneous database was also used. The database was acquired by a research group at the Institut Universitaire de Gériatrie de Montréal working on the relation between facial expressions and the neural networks of pain. The database was recorded during one of their studies and is not freely accessible. A detailed description of the database can be found in [68].

4.1.2.1. Spontaneous data acquisition and validation. The subjects participated in a study on the relation between pain catastrophizing and facial responsiveness of pain in healthy, pain-free individuals [68]. Forty subjects (20 females and 20 males) were recorded in total. The subjects knew that they were being recorded and that they were taking part in a pain induced experiment. Pain was induced experimentally by means 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. One non-painful (1 °C below the individual pain threshold) and two 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 s and returned to baseline with a rate of 4 °C/s. ISIs varied between 30 and 35 s [68]. The faces of subjects were videotaped during the experiments. The video camera was placed in front of the subject at a distance of approximately 4 m. For subject's confidentiality rights we cannot show the images of this database

(the interested reader can directly contact Kunz et al. [68]). 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), the light signal was switched alternatively [68]. The light was visible to the camera but not to the subject. Videos were captured at a resolution of 720 × 536 pixels, out of which the face area spanned an average of approximately 320 × 470 pixels. Each video begins with the neutral state and evolves until the apex of the undergoing facial expression (in the case of pain expressions) or remains at neutral state in the case of neutral videos. Facial expressions displayed on each frame of the obtained videos have been validated as pain expression using the Facial Action Coding System (FACS) [35]. 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 database [68]. The 5 s segments selected for scoring began immediately after the stimulus had reached its maximum. Twenty FACS validated subjects (and the corresponding sequences) are used for the validation process in the current paper. The data correspond to 40 sequences (20 subjects × 2 (neutral + pain) expressions). A large variability was noted between the acted and spontaneous pain expression samples. Compared to a human observer, the proposed model gives comparable classification rates on these data (see Section 4.2.2.2).

4.2. Simulation results

The simulation results were obtained on the STOIC database (85 video sequences containing the 6 basic facial expressions, pain, and neutral) and the spontaneous database (40 video sequences containing 20 neutral and 20 pain sequences). A total of 125 videos in which 35 different subjects expressed pain: the 15 video sequences (15 subjects) from the STOIC database (validated by human observers, see Section 4.1.1) and the 20 spontaneous pain sequences (20 subjects, FACS validated, see Section 4.1.2). The number of sequences used for the validation process is in fact a higher number than in all the validation databases used by the already proposed models for pain expression recognition (see Section 1.2). The generalization performances of the proposed model for pain expression recognition were performed, first by establishing the rules only on one part of the subjects of the STOIC database and the validation of the model on the whole database (new subjects) and, secondly, by evaluating the model performances on the spontaneous database (20 new subjects recorded in another experimental condition). The proposed results show, first, the performances of the proposed model for the recognition of pain expression in the case of 2-alternative forced choices (pain vs. neutral) and 8-alternative forced choices (the 6 basic facial expressions, pain, and neutral); second, they prove the refinement role of the context information; third, they show the comparison between these performances and those obtained by a human observer in similar conditions, demonstrating the usefulness of the model in a real-life application. Two simulations are reported: first, a 2-alternative forced choice classification between pain and neutral to compare the obtained performances to the 2-alternative forced choice classification systems previously developed (see Section 1.2); second, an 8-alternative forced choice classification between pain and the 6 basic facial expressions plus neutral.

4.2.1. Human behavior: experimental material and methods

Human visual system remains far ahead of everything that has been done in automatic facial expression recognition. Thus, to better evaluate the model performances, a comparison with the human observer performances on the same data is made. Two behavioral experiments were made for the classification of pain expression by human observers. First, human observers were asked to discriminate between the six basic facial expressions as well as pain and neutral expressions on the STOIC database. Second, the same participants were asked to discriminate pain expression from neutral expression on the spontaneous database. Indeed, due to the lack of spontaneous data for the six basic facial expressions, spontaneous pain expression could not be identified among the six other expressions by human observers. In addition, the spontaneous pain data could not be presented with the STOIC videos for a discrimination task because the spontaneous pain expressions were acquired in a completely different environment, which would have made the task trivial as human observers may have classified differences in the dataset, rather than the expression. In the first behavioral experiment the stimuli corresponded to 85 videos (8 facial expressions × 15 subjects) randomly interleaved in 4 separate blocks. In the second experiment the stimuli corresponded to 40 videos (2 facial expressions (pain and neutral) × 20 subjects) randomly interleaved in 4 separate blocks. Twelve naïve observers (6 males and 6 females) from Montreal, undergraduate and graduate students with normal or corrected-to-normal vision, took part in the two experiments. They had no specific training in facial expression recognition and were not clinicians. They were required to respond as accurately as possible in a forced-choice discrimination task, by pressing the appropriate keyboard keys. The classification performances of human observers are compared to the model performances in the same conditions and are reported in Sections 4.2.2.1 and 4.2.2.2.

4.2.2. Model performances for pain expression recognition

The classification performances are reported in two conditions: 2-forced choice classification and 8-forced choice classification on the STOIC and spontaneous databases. The validation of the proposed model on spontaneous data allows investigating how the proposed model performed on 20 new subjects expressing "spontaneous" pain expressions, which included head movement with both in-plane and out-of-plane rotations. Thus these spontaneous data allows evaluating the ability of the model to generalize to a new database presenting less controlled pain expression sequences. For example, compared to the acted data (STOIC database) even if the subjects had exaggerated their expression, which is very unlikely given that pain was physically induced, subjects displayed greater variability than the actors, which challenges even more the robustness of the automatic recognition system. In the following, the obtained classification performances are first reported on each database separately (i.e. STOIC and spontaneous). The classification performances (C) (i.e. the percentage agreement with the ground truth), recall (R), precision (P), and F-measure (F), which combine evenly recall and precision as $F=2(recall \times precion)/(recall+precision)$, are computed to evaluate the performances of the system in detail. Notably, precision gives an estimation of the amount of false alarms introduced when classifying either pain expression or one of the six basic facial expressions and neutral among 8-alternative choices. Figs. 13 and 14 show the result plots.

4.2.2.1. Performances of the proposed model on 2-alternative forced choice classification. In the first simulation the system performed a 2-alternative forced choice classification between neutral and pain facial expressions. The obtained results are $C=92.3\%$, $R=90\%$, $P=86\%$, and $F=86\%$. Fig. 13 summarizes the obtained results. For all the used criteria, the obtained results on acted data are the highest and are higher than the performances reported in the

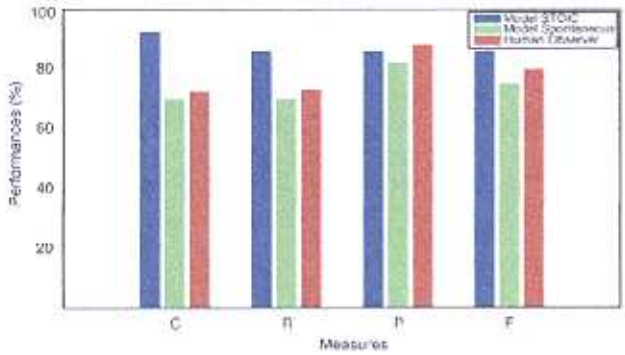

Fig. 13. Classification performances (%) in the case of 2-alternative forced choices for the model and human observers.

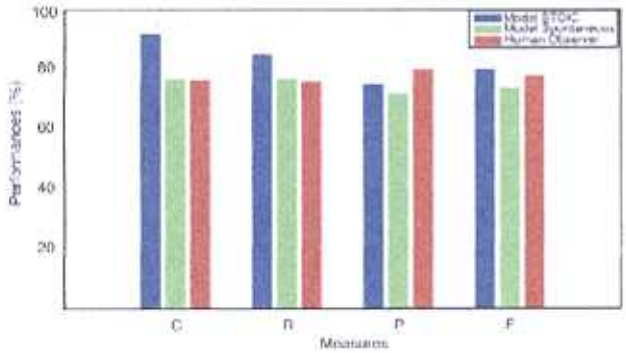

Fig. 14. Classification performances (%) in the case of 8-alternative forced choices for the model and human observers.

state of the art in the same conditions. Indeed, the obtained classification rate is much higher than the one reported by the already proposed models (C between 45.7% and 87.99% [51]) in the same conditions (i.e. pain expression recognition in the case of 2-alternative forced choices on acted data). The dynamic fusion process of the required facial feature information as well as the introduction of the context information (compared to the already reported models) for the recognition of pain expression may explain this result. As expected, the performances obtained on spontaneous data are lower ($C=70\%$, $R=70\%$, $P=82\%$, and $F=75\%$) than those on acted data. However, they compare favorably to the already reported classification performances on spontaneous data in the state of the art ($C=50.4\%$ for [10], $C=76\%$ for [9], and $C=81\%$ for [15]). Yet, this comparison remains difficult to make as the approaches differ in several characteristics (e.g. the models are applied on different expressions, they either classify expressions or only AUs [10,9], they are tested on different databases [15], or they do not always report their classification performances (for example at recognizing pain expression from the 2-alternative choices (painful or baseline) [8]). However, these performances give the reader a measure to refer to in order to evaluate the difficulty of the task in the case of spontaneous data. Moreover, compared to [15] (the most appropriate comparison with the state of the art even if obtained on a different dataset) the obtained performances compare favorably given the number of features used for the classification of only two classes. Indeed, the authors [15] used more than 100 features compared to 7 in our case. Using additional features (such as shape information) should also provide useful information to the proposed model and possibly increase the obtained performances. Moreover compared to [15] the proposed model has been generalized to the classification of pain from the six other expressions plus neutral. Interestingly, the performances obtained on spontaneous data compare favorably to those of human observers (see Section 4.2.2) on the same data ($C=72.6\%$, $R=73\%$, $P=88\%$, and $F=80\%$). In fact, human and model performances are not significantly different (two-way ANOVA, $P>0.01$). Then, even if the obtained performances on spontaneous data are lower than those obtained on acted data they are very encouraging given the fact that the system was not trained on samples of spontaneous pain expression while it is still able to generalize to them. In our point of view this comparison is a more appropriate evaluation of the system performances considering humans as experts in facial expression recognition.

To summarize, the mean performances obtained on the 2-alternative forced choices are very encouraging ($C=81.2\%$, $R=78\%$, $P=84\%$, and $F=80.5\%$). They are much higher than the already proposed models for pain expression recognition on acted data and compare favorably to human observer performances in the case of spontaneous data. However, in every day life situations, in addition to pain expression other facial expressions can be expressed. This is the case for example of the hospital context we introduced previously (see Section 3.1.3). Pain expression needs then to be dissociated from at least the six basic facial expressions and neutral to have a better estimation of the classification performances.

4.2.2.2. Performances of the proposed model on 8-alternative forced choice classification. In order to evaluate the robustness of the proposed model for the recognition of pain expression and to discriminate it from the six basic facial expressions and neutral, a second simulation was carried out where the system performed an 8-alternative forced choice classification. This simulation was carried out on the STOIC database (i.e. on the 85 videos displaying the 6 basic facial expressions, neutral, and pain) and on the spontaneous database (i.e. on the 40 videos displaying either pain or neutral expressions). Fig. 14 summarizes the obtained results. The classification performances obtained in the case of 8-alternative forced choices on the STOIC database are $C=92\%$, $R=85\%$, $P=75\%$, and $F=80\%$. These results are slightly lower than those obtained on the same data for the 2-alternative forced choices (see Section 4.2.2.1). Notably, the precision drops from 86% to 75%. Given the task difficulty (8-alternative choices) the obtained performances remain satisfactory (F-measure$=80\%$). Moreover, similarly to the 2-alternative forced choices, the classification performances on spontaneous database ($C=77\%$, $R=77\%$, $P=72\%$, $F=74\%$) show that the obtained precision (72%) is lower (but satisfactory) than the one obtained on the acted database (82%). However, as pain is more relevant in a medical context (modeled in our case by the introduction of the context in the classification process), recall is more important than precision (i.e. in medical context having a false alarm is better than missing a painful state). Interestingly, the obtained performances reflect this characteristic, leading to a higher recall ($R=85\%$, $R=77\%$) compared to the precision ($P=75\%$, $P=72\%$) in the case of acted as well as spontaneous data respectively. The obtained results (on acted and spontaneous data) are very encouraging given the difficulty of the task due to the large variation between individuals and the subtle facial feature changes displayed during spontaneous expressions. It has to be emphasized as well that the model was not trained on spontaneous data, but only one part of the STOIC database, which shows its ability to generalize to other databases. Unfortunately, there are no available results in the state of the art classifying pain from the 8-alternative forced choices (neither on acted or on spontaneous data) to compare with. The only possible comparison was with the performances obtained by human observers. Based on the

Table 3
Misclassification results (%) of acted and spontaneous pain sequences in the case of an 8-alternative choice with the context variable.

	Happy	Disgust	Surprise	Fear	Sadness	Anger	Pain	Neutral	Sadness, anger, pain	Sadness, anger	Ignorance
Acted pain	8	0	0	0	0	0	92	0	0	0	0
Spontaneous pain	0	0	0	0	0	0	77	0	0	15	8

same data (STOIC database) the obtained results in the case of 8-alternative forced choices compare favorably to the performances of human observers ($C=76\%$, $R=76\%$, $P=80\%$, $F=80\%$). Importantly, both humans and the model show comparable precision levels and their overall performances are not significantly different (two-way ANOVA, $P>0.01$). However, compared to the performances of the proposed context based model, recall is lower than precision in the case human observer performances. We explain these results by our implementation of the dynamic fusion process, which uses the most important visual information for the human observer [25,26] and especially the refinement process due to information conveyed by the context variable (as described in Sections 3.2.3 and 4.3). It is important to stress that people usually do not display "pure" facial expressions [50], which introduces doubts between expressions. Indeed, the 8-alternative forced choice behavioral experiment, on the STOIC database, reported that human observers misclassify pain expression as sadness, disgust, or happiness. In addition to the performances reported in Fig. 14, a refined analysis shows that before the introduction of the context information different kinds of confusions appear according to the used database. Compared to the human observers, the same doubt states appeared between the same expressions on STOIC database (i.e. doubt between sadness, disgust, and happiness). Notably the obtained results are supported by the results of Roy et al. [25,26], where 35 human observers were asked to classify pain expression with the 8-alternative forced choice paradigm (also on the STOIC database). The authors reported that humans also misclassify pain with sadness, disgust, and happiness expression. More interestingly, we obtain the same confusions when we asked naïve human observers to recognize the same data. These results prove the ability of the proposed model to report doubt in the case of confused expressions, where even human observers cannot dissociate between them without context. The same analysis shows as well that in the spontaneous database, these doubts are no longer between the expected expressions: sadness, disgust, and happiness but between sadness and anger expressions. In these cases the system is sure that the current expression is one of these 2 (or 3) and never one of the 6 (or 5) other expressions (their corresponding rates are equal to zero). These results further assess that the model shows a striking similarity with the human observers and that the introduction of the context information allows solving these confusions.

These results are of first interest as they demonstrate the ability of the proposed model to describe a given facial expression with a set of possible responses (in the case of blend expressions) corresponding to doubt that would also be reported by human observers. In some critical applications where the human validation is requested this ability may appear to be very important and useful. For example, considering the application in the hospital context (waiting room, sick people under camera monitoring (see Fig. 5)), such information is more than sufficient to alert somebody in charge that the patient is suffering pain (displaying also the possible doubt with one or two other facial expressions rather than making a wrong decision). Medical personnel – physician or nurse – can then confirm or not confirm this information. The introduction of the context information (as described in Section 2.2.2) solves the doubt state in the case of blends expression, increasing the performances by 15%. Indeed, in the case of pain monitoring application (videotaped waiting room, injured people), in the current modeling the context information is used as additional information and allows the support of the more relevant facial expression (i.e. pain) in the case of doubt. Table 3 summarizes the obtained misclassification results with the context information (detailed misclassification of the six basic facial expression can be found in [2,61]).

To summarize, the proposed model performances on the 8-alternative forced choice classification prove its robustness to recognize pain expression on acted as well as on spontaneous data. It shows its robustness to new situations (i.e. applied to new sets of data not used for training). Comparison between the proposed model and the human behavior allows better quantification the model performances. Despite the fact that humans are more precise than the automatic model, the time required for coding, which is accomplished manually by a human expert, makes the automatic model more suitable for real-time situations.

5. Conclusion

In the current paper we present a new dynamic modeling process for the automatic classification of pain expression. The proposed model is based on a dynamic fusion process of facial and context information for the automatic classification of pain expression from 8-alternative forced choices. The system proves its suitability to deal with spontaneous sequences and the obtained performances compare favorably to those of human observers, leading to similar doubt between confused expressions. These results are encouraging given the context of the application. They prove that an automatic detection of pain expression in video sequences is a feasible task.

The obtained results compare favorably to the already proposed models for the automatic recognition of pain expression in the case of 2-alternative forced choice classification. Moreover, the proposed model shows its robustness and suitability in the case of a challenging 8-alternative forced choice classification. This opens promising perspectives for the development of the model and its application in everyday-life situations. For example, preliminary results on the recognition of spontaneous pain expression intensities [66] proved its potential to be applied in a hospital context application. Even though in the current paper we focus our work on recognizing pain expression on adult subjects it is of interest to test it on the recognition of newborns' pain expression. Finally, the dynamic and multi-cues model has been tested and validated for pain expression. However, the same approach can be easily applied to other facial expressions. For example, the proposed model is the first contribution where the context information is explicitly taken into account for the classification of facial expressions. This modeling can be easily generalized to take into account several other context variables according to the studied facial expression. Moreover, the original version of the proposed model (static recognition based only on the permanent facial features information) has been extensively validated on benchmark databases. The addition of new sensors and the dynamic information enhanced also the classification

performances for the six basic facial expressions (this claim is based on the preliminary results obtained in [58,60]).

Acknowledgments

We would like to thank Pierre Rainville and colleagues of the Center de Recherche de l'Institut Universitaire de Gériatrie de Montréal and of the Otto-Friedrich University Bamberg, Germany, for allowing us to use part of their non-publically available spontaneous pain expression data. The proposed work is in a continuation of the work of Zakia Hammal on the automatic and dynamic analysis of facial expression using the TBM based modeling. The first version of the proposed work appeared in Hammal (2006) [46].

References

[1] K.M. Prkachin, P.E. Solomon., The structure, reliability and validity of pain expression: evidence from patients with shoulder pain, Pain 139 (2008) 267–274.
[2] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut, Facial expressions classification: a new approach based on transferable belief model, International Journal of Approximate Reasoning 46 (3) (2007) 542–567.
[4] P. Ekman, E.L. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System, 2nd edition, Oxford University Press, 2005.
[5] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut, Facial expressions recognition based on the belief theory: comparison with different classifiers, in: Proceedings of the 13th International Conference on Image Analysis and Processing (ICIAP), Cagliari, Italy, September 6–8, 2005.
[6] Z. Zeng, M. Pantic, G.I. Roisman, T.S. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (1) (2009).
[7] J.F. Cohn, K.L. Schmidt., The timing of facial motion in posed and spontaneous smiles, Wavelets, Multiresolution & Information Processing 2 (2) (2004) 121–132.
[8] G.C. Littlewort, M.S. Bartlett, L. Kang, Faces of pain: automated measurement of spontaneous facial expressions of genuine and posed pain, in: Proceedings of ICMI, Nagoya, Aichi, Japan, November 12–15, 2007.
[9] M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, J.R. Movellan, Automatic recognition of facial actions in spontaneous expressions, Journal of Multimedia 1 (6) (2006) 22–35.
[10] M.F. Valstar, M. Pantic, Z. Ambadar, J.F. Cohn, Spontaneous vs. posed facial behavior: automatic analysis of brow actions, in: Proceedings of the ACM International Conference on Multimodal Interfaces, 2006, pp. 162–170.
[12] R. El Kaliouby, P. Robinson, Mind reading machines: automated inference of cognitive mental states from video, in: Proceedings of the International Conference on Systems, Man and Cybernetics, 2004, vol. 1, pp. 682–688.
[13] Q. Ji, P. Lan, C. Looney., A probabilistic framework for modeling and real-time monitoring human fatigue, IEEE Transactions on Systems, Man and Cybernetics A 36 (5) (2006) 862–875.
[14] U. Hess, R.E. Kleck, Differentiating emotion elicited and deliberate emotional facial expressions, European Journal of Social Psychology 20 (5) (1990) 369–385.
[15] A.B. Ashraf, S. Lucey, T. Chen, K. Prkachin, P. Solomon, Z. Ambadar, J. Cohn, The painful face: pain expression recognition using active appearance models, in: Proceedings of the ACM International Conference on Multimodal Interfaces (ICMI), 2007, pp. 9–14.
[16] M. Pantic, L.J.M. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1424–1445.
[17] M. Kunz, S. Scharmann, U. Hemmeter, K. Schepelmann, S. Lautenbacher, S. The facial expression of pain in patients with dementia, Pain 133 (1–3) (2007) 221–228.
[18] V.T. Lehr, P.S. Zeskind, J.P. Ofenstein, E. Cepeda, L. Warrier, J.V. Aranda, Neonatal facial coding system scores and spectral characteristics of infant crying during newborn circumcision, Clinical Journal of Pain 23 (5) (2007) 417–424.
[19] A.C. Lints-Martindale, T. Hadjistavropoulos, B. Barber, S.J. Gibso., 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) (2007) 678–689.
[20] M. Pantic, M.S. Bartlett, in: K. Delac, M. Grgic (Eds.), Machine analysis of facial expressions, in: Face Recognition, 2007, I-Tech Education and Publishing, Vienna, Austria, 2007, pp. 377–416.
[22] B. Fasel, J. Luettin, Automatic facial expression analysis: A survey, Pattern Recognition 1 (30) (2003) 259–275.
[25] S. Roy, C. Roy, Z. Hammal, D. Fiset, C. Blais, R. Jemel, F. Gosselin, The use of spatio-temporal information in decoding facial expression of emotions, in: Proceedings of the Vision Science Society (VSS) Conference, 9–14 May 2008, Naples Grand Hotel, Floride.
[26] C. Roy, S. Roy, D. Fiset., Z. Hammal, C. Blais, P. Rainvill, F. Gosselin, Recognizing static and dynamic facial expressions of pain: gaze-tracking and bubbles experiments, in: Proceedings of the Vision Science Society (VSS) Conference, 9–14 May 2008, Naples Grand Hotel, Floride.
[30] K. Forbes-Riley, D. Litman, Predicting emotion in spoken dialogue from multiple knowledge sources, in: Proceedings of the International Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL), 2004.
[31] A. Kapoor, W. Burleson, R.W. Picard, Automatic prediction of frustration, International Journal of Human–Computer Studies 65 (8) (2007) 724–736.
[32] D.J. Litman, K. Forbes Riley, Predicting student emotions in computer human tutoring dialogues, in: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), July 2004, International Conference on Acoustics, Speech, and Signal Processing.
[33] P. Smets, Data fusion in the transferable belief model, in: Proceedings of the International Conference on Information Fusion, July 2000, Paris, France, pp. 21–33.
[34] T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artificial Intelligence 172 (2–3) (2008) 234–264.
[35] P. Ekman, W.V. Friesen., The Facial Action Coding System (FACS): A Technique for the Measurement of Facial Action, Consulting Psychologists Press, Palo Alto, CA, 1978.
[36] S. Roy, C. Roy, I. Fortin, C. Ethier-Majcher, P. Belin, F. Gosselin, A dynamic facial expression database, in: Proceedings of the Vision Sciences Society (VSS) Conference, Sarasota, Florida, 2007.
[37] Z. Hammal, N. Eveno, A. Caplier, P.Y. Coulon, Parametric models for facial features segmentation, Signal Processing 86 (2006) 399–413.
[38] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of the International Joint Conference on Artificial Intelligence, Vancouver, Canada, August 1981, vol. 18, No. 11, pp. 674–680.
[39] Cohn Kanade database, 2000. ⟨http://vasc.ri.cmu.edu/idb/html/face/facial expression⟩.
[40] M. Dailey, G.W. Cottrell & J. Reilly, California Facial Expressions (cafe), University of California, San Diego, 2001, unpublished digital images.
[41] Hammal Caplier database, 2003). ⟨http://viscog.psy.umontreal.ca/zakia/index.htm⟩.
[43] P. Smets, R. Kruse., The transferable belief model, Artificial Intelligence 66 (1994) 191–234.
[44] Z. Ambadar, W.S. Schooler, J.F. Cohn., The importance of facial dynamics in interpreting subtle facial expressions, Psychological Science 16 (5) (2005) 403–410.
[45] A.C. Williams, Facial expression of pain: an evolutionary account, Behavioral and Brain Sciences 25 (4) (2002) 439–455.
[46] Z. Hammal, Facial Features Segmentation: Analysis and Recognition of Facial Expressions Using the Transferable Belief Model, Ph.D. Thesis, IIS Laboratory, Grenoble, France, June 2006.
[47] B. Ristic, P. Smets, Kalman filters for tracking and classification and the transferable belief model, in: Proceedings of the International Conference on Information Fusion ICIF 04, 2004, Stockholm, Sweden.
[48] T. Denoeux, Pattern recognition and information fusion using belief functions: some recent developments, in: Proceedings of ECSQARU 2007, p. 1.
[49] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, International Journal of Approximate Reasoning 38 (2005) 133–147.
[50] A.W. Young, D. Rowland, A.J. Calder, N.L. Etcoff, A. Seth, D.I. Perrett, Facial expression megamix, Cognition 63 (1997) 271–313.
[51] Md. Maruf Monwar, Siamak Rezaei, Pain recognition using artificial neural network, in: Proceedings of the IEEE International Symposium on Signal Processing and Information Technology, 2006.
[53] P. Ekman, W.V. Friesen, J.C. Hager, Facial Action Coding System: Research Nexus, Network Research Information, Salt Lake City, UT, 2002.
[54] J. Yang, R. Stiefelhagen, U. Meier, A. Waibe, Real-time face and facial feature tracking and applications, in: Proceedings of the Auditory–visual Speech Processing (AVSP) Conference, NSW, Australia, 1998.
[55] Y. Tian, L. Brown, Real world real-time automatic recognition of facial expressions, in: Proceedings of the IEEE Workshop on Performance Evaluation of Tracking and Surveillance, Graz, Austria, 2003.
[56] K.D. Craig, K.M. Prkachin, R.V.E. Grunau., The Facial Expression of Pain, in: D.C. Turk, R. M (Eds.), Handbook of Pain Assessment, 2nd edition,Guilford, New York, 2001.
[57] I. Fasel, R. Fortenberry, J.A. Movellan, Generative framework for real time object detection and classification, Computer Vision and Image Understanding 98 (2005) 182–210.
[58] Z. Hammal, C. Massot, Holistic and feature-based information towards dynamic multi-expressions recognition, in: Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP), Angers, France, May 2010.
[60] Z. Hammal, C. Massot, Dynamic facial expression classification based on human visual cues information, in: Proceedings of the Fourth Workshop on Emotion and Computing—Current Research and Future Impact, Paderborn, KI, 15.9.2009.

[61] Z. Hammal, M. Arguin, F. Gosselin, Comparing a novel model based on the Transferable Belief Model with humans during the recognition of partially occluded facial expressions, International Journal of Vision 9 (2) (2009) 22. doi:10.1167/9.2.22. 1–19.

[64] A. Kapoor, W. Burleson, R.W. Picard, Automatic prediction of frustration, International Journal of Human–Computer Studies 65 (8) (2007) 724–736.

[65] Y. Zhang, J. Qiang, Active and dynamic information fusion for facial expression understanding from image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 699–714.

[66] Z. Hammal, Context based recognition of pain expression intensities, in: Proceedings of the Fifth Workshop on Emotion in Human-Computer Interaction—Real World Challenges held at the 23rd BCS HCI Group Conference, Cambridge University, Cambridge, UK, September 2009.

[67] S. Brahnam, C.-F. Chuang, F.Y. Shih, M.-R. Slack, Machine recognition and representation of neonatal facial displays of acute pain, Artificial Intelligence in Medicine 36 (2006) 211–222.

[68] M. Kunz, C. Chatelle, S. Lautenbacher, P. Rainville, The relation between catastrophizing and facial responsiveness to pain, Pain 140 (2008) 127–134.

[69] T. Denoeux, M.-H. Masson, EVCLUS: Evidential clustering of proximity data, IEEE Transactions on Systems, Man and Cybernetics B 34 (1) (2004) 95–109.

[70] A. Arguin, T. Denoeux, Fusion of one-class classifiers in the belief function frame-work, in: Proceedings of the 10th International Conference on Information Fusion, Quebec, Canada, July 2007.

[71] D. Mercier, B. Quost, T. Denoeux, Refined modeling of sensor reliability in the belief function framework using contextual discounting, Information Fusion 9 (2) (2008) 246–258.

[72] B. Quost, T. Denoeux, M.-H. Masson, Pairwise classifier combination using belief functions, Pattern Recognition Letters 28 (5) (2007) 644–653.

[75] T. Denoeux, P. Smets, Classification using belief functions: the relationship between the case-based and model-based approaches, IEEE Transactions on Systems, Man and Cybernetics B 36 (6) (2006) 1395–1406.

[76] Z. Elouedi, K. Mellouli, P. Smets, Belief decision trees: theoretical foundations, International Journal of Approximate Reasoning 28 (2001) 91–124.

[77] S. Trabelsi, Z. Elouedi, K. Mellouli, Pruning belief decision tree methods in averaging and conjunctive approaches, International Journal of Approximate Reasoning 46 (3) (2007) 568–595.

[79] F. Ramasso, M. Rombaut, D. Pellerin, 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 (2007) 944–949.

[80] S. Petit-Renaud, T. Denoeux, Nonparametric regression analysis of uncertain and imprecise data using belief functions, International Journal of Approximate Reasoning 35 (1) (2004) 1–28.

[81] M.-H. Masson, T. Denoeux, Clustering interval-valued data using belief functions, Pattern Recognition Letters 25 (2) (2004) 163–171.

[82] L. Valet, G. Mauris, P. Bolon, A statistic overview of a recent literature in information fusion, in: Procedings of Information Fusion ICIF 00, Paris, France, pp. 22–29.

Zakia Hammal received her PhD degree in cognitive science from the Université Joseph Fourier, Grenoble, France, in 2006. From 2006 to 2010, she was a postdoctoral research associate at the Université de Montréal, Montreal, Canada. She is currently at the Robotics Institute of the Carnegie Mellon University, Pittsburgh, USA. Her main research topics include computer vision, machine learning and visual perception applied to human-machine interaction and emotion recognition. Hammal has published over 30 technical papers in the areas of affective computing, image processing, machine learning and visual perception in prestigious journals and conferences in these fields.

Miriam Kunz received her bachelor's degree (2000) and master's degree (2002) in Psychology from the University of Berlin (FU, Germany). Her doctoral research focused on pain processing in patients with dementia and she received her PhD degree in Psychology from the University of Bamberg in 2006. During her post-doctoral stay (2007–2009) at the University of Montreal (Canada) she studied which brain activation patterns underlie the regulation of facial expressions. At the moment she is working as a researcher at the University of Bamberg (Germany).