Image and Vision Computing 31 (2013) 231–245

Contents lists available at SciVerse ScienceDirect

Image and Vision Computing

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

A unified probabilistic framework for automatic 3D facial expression analysis based on a Bayesian belief inference and statistical feature models ☆

Xi Zhao [a,*], Emmanuel Dellandréa [a], Jianhua Zou [b], Liming Chen [a]

[a] Université de Lyon, CNRS, Ecole Centrale Lyon, LIRIS, UMR5205, F-69134, France
[b] School of Electronic and Information Engineering, Xi'an Jiaotong University 710049, China

ARTICLE INFO

Article history:
Received 30 October 2011
Received in revised form 31 July 2012
Accepted 17 October 2012

Keywords:
Bayesian Belief Network
Statistical feature model
3D face
Facial expression recognition
Action units recognition
Automatic landmarking

ABSTRACT

Textured 3D face models capture precise facial surfaces along with the associated textures, making it possible for an accurate description of facial activities. In this paper, we present a unified probabilistic framework based on a novel Bayesian Belief Network (BBN) for 3D facial expression and Action Unit (AU) recognition. The proposed BBN performs Bayesian inference based on Statistical Feature Models (SFM) and Gibbs–Boltzmann distribution and feature a hybrid approach in fusing both geometric and appearance features along with morphological ones. When combined with our previously developed morphable partial face model (SFAM), the proposed BBN has the capacity of conducting fully automatic facial expression analysis. We conducted extensive experiments on the two public databases, namely the BU-3DFE dataset and the Bosphorus dataset. When using manually labeled landmarks, the proposed framework achieved an average recognition rate of 94.2% and 85.6% for the 7 and 16 AU on face data from the Bosphorus dataset respectively, and 89.2% for the six universal expressions on the BU-3DFE dataset. Using the landmarks automatically located by SFAM, the proposed BBN still achieved an average recognition rate of 84.9% for the six prototypical facial expressions. These experimental results demonstrate the effectiveness of the proposed approach and its robustness in landmark localization errors.

Published by Elsevier B.V.

1. Introduction

Facial expression is one of the most naturally pre-eminent means in human communication. Its automatic analysis and recognition have many potential applications, including human–computer interaction, security, interactive games, computer-based learning, entertainment, etc. Over the last two decades, Facial Expression Recognition (FER) has been the subject of extensive research from several research communities, ranging from computer vision, psychology to human–computer interaction.

The sources of facial expressions are multiple. They occur in human verbal and non-verbal communication, during their mental states, e.g. felt emotions, conviction, etc., or physiological activities such as pain, tiredness, etc. Facial expressions are characterized by contractions of facial muscles, leading to temporally deformed facial features. These temporal facial deformations account for displacement of intransient features (e.g. eye lids, eye brows, nose, lips, cheeks) as well as occurrence of transient features like skin texture changes and facial surface changes, often revealed by wrinkles, furrows and bulges in both texture and geometry modalities. For instance, an exaggerated smile leads to swelling of the cheeks, wide opening of the mouth and significant displacement of mouth corners, thus deforming globally both the face morphology, texture and shape. On the other hand, a subtle surprise may be simply materialized by raising of the eyebrows and results in very local changes both in the face morphology and appearance. Reliable analysis of facial expressions thus requires accurate measurements of facial feature deformations, both global and local, in terms of morphology, texture and shape.

The current research on machine-based analysis of facial expressions features two main streams of approaches: judgment-based approaches [1] centered on the messages conveyed by facial expressions, and sign-based approaches, [2,3], targeting the recognition of facial muscle actions coding visually discernible facial motions and deformations. Most of the existing efforts on FER are judgment-based approaches and directly map facial expressions into the six basic emotions, namely happiness, sadness, fear, disgust, surprise and anger, due to their universal properties, their marked reference representation in our affective lives, and the availability of the relevant training and test material. Unfortunately, these six prototypical emotion categories are only a subset of thousands of possible facial displays, most of them having subtle changes in discrete facial features such as pulling the lip corners or raising the eyebrows. As far as sign-based approaches are concerned, the most commonly used facial muscle action descriptors are the Action Units (AUs) introduced by Ekman et al. in 1978 in the

☆ This paper has been recommended for acceptance by Lijun Yin.

* Corresponding author.
E-mail addresses: zhaoxi1@hotmail.com (X. Zhao), emmanuel.dellandrea@ec-lyon.fr (E. Dellandréa), jhzou@sei.xjtu.edu.cn (J. Zou), liming.chen@ec-lyon.fr (L. Chen).

0262-8856/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.imavis.2012.10.001

Facial Action Coding System (FACS) [2]. These AUs can be further interpreted, by employing facial expression dictionaries such as EMFACS and FACSAID, in complex and subtle facial expression categories such as depression or pain [3].

1.1. The background

The extensive research on FER has accumulated significant lessons and results over the last two decades. Detailed surveys of previous work can be found in [4–8,1,9].

However, most of the existing research on FER deals only with the six basic emotions and works with nearly frontal faces in 2D images, being static or dynamic in sequence [1]. The typical facial features are either geometric features such as the shapes of the facial components (eyes, mouth, etc.) and the location of facial salient points (corner of the eyes, mouth, etc.) [10,11] or appearance features describing the facial texture, including wrinkles, bulges and furrows [12,13]. However, these methods suffer from the problems of head pose changes and illumination variations in 2D images.

With 3D imaging systems readily available, FER in 3D, which is the focus of the current paper, has recently emerged as a major solution to these unsolved issues [14,9]. The release of the three public 3D datasets, namely the BU-3DFE, BU-4DFE and the Bosphorus datasets [15,16], has further fostered research efforts in this direction. As a textured 3D face scan is theoretically insensitive to lighting conditions and head pose while capturing the accurate facial surface and texture, one expects more reliable, view-independent and illumination robust solutions to FER. However, most of the current studies in 3D are judgment-based approaches dealing with the six prototypic emotions. Most of them are geometric-based approaches, making use of various geometric features (e. g., fiducial facial points [17–19], curvatures [20], 3D face parametrization [21]) to best account for variations in 3D face morphology.

One of the early works in 3D FER is Wang et al. [20] which computed the histogram of the principal curvatures, surface principal directions and steepness of the surface around face regions extracted from 64 manually labeled landmarks. Soyel and Demirel [17] retrieved six distances between facial landmarks, describing the openness of the eyes, the height of the eyebrows, the openness of the mouth and its width, the stretching of the lips and the openness of the jaw. Such distance-like features are further explored by Tang and Huang [22], where fewer than 30 'best' features were automatically selected from a candidate pool (all distances between 83 manually labeled landmarks in the BU-3DFE dataset). In addition to these distance-based features, Hao and Huang [18] also extracted the slopes of the line segments connecting the 83 feature points. Although distance-feature based approaches are computationally efficient, they ignore appearance information and need a predefined set of facial landmarks to compute morphology deformations. Thus, their performance highly relies on the accuracy of facial landmark location. This probably explains that all these works only made use of the 83 manually labeled landmarks defined in the BU-3DFE dataset. Furthermore, these approaches can also have difficulties in dealing with non-exaggerated and non-prototypical facial expressions as they ignore appearance features.

In the literature there also exist model-based approaches which typically fit a deformable face model to an input 3D face model [21,23,24]. For instance, Ramanathan et al. proposed in [21] a Morphable Expression Model (MEM) and used the fitted model parameters for facial expression recognition. Mpiperis et al. [23] first applied an elastically deformable model algorithm to fit a prototypic facial surface model and then made use of bilinear models for both face and facial expression recognition. Unfortunately, all these approaches require a dense registration of point clouds, to build point-to-point correspondences. This is quite computationally expensive when dealing with thousands of 3D vertex in a facial mesh. Moreover, fitting algorithms used so far also require the existence of some landmarks for initialization.

1.2. The proposed approach

In this paper, we present a unified framework based on a Bayesian Belief Network (BBN) to the recognition of 3D facial expressions, including both the six prototypical expressions and action units (AUs). The proposed BBN encompasses a subject node, an expression node and a set of feature nodes describing multiple evidences as provided by both geometry and appearance features. These features are extracted from three different facial modalities and thereby account for both facial intransient and transient deformations. The three facial modalities are namely facial morphology as defined by the configuration of a predefined set of 3D landmarks, texture and shape as defined by intensity and range values of local patches in the vicinity of each landmark respectively. For each expression state associated with the expression node, a set of Statistical Feature Models (SFM) is trained for each type of features to enable Bayesian inference for FER on the proposed BBN. In this work, we make use of Gibbs–Boltzmann distributions and convert the initial Bayesian inference for FER from the probability domain into the energy domain so that the similarities between an instance of a feature type and those synthesized by the associated SFMs represent the beliefs of a feature node on different expression states. These belief sets are further fused and the highest one in the fused set is recognized as the hidden facial expression. When combined with our previously developed morphable partial face model (SFAM) [25,26], the proposed BBN achieves fully automatic FER. We demonstrated the effectiveness of the proposed approach using the two public datasets, namely Borphorus and BU-3DFE datasets, for the recognition of six basic facial expressions as well as action units.

The flowchart of the proposed system is shown in Fig. 1 and encompasses 4 main stages: offline SFAM construction, offline BBN training, online landmarking and feature extraction and online facial expression recognition. Specifically, SFAM is trained for automatic landmarking using a small set of 3D face models with all expressions. For facial expression recognition, a set of SFMs are trained for all combinations of types of feature and facial expressions. During online recognition, a 3D face scan is first landmarked by the SFAM, then 15 types of features in the vicinity of these landmarks are extracted and used as evidence by the BBN for belief inference. The recognized facial expression is thus the one having the highest a posteriori probability given all the geometric and appearance evidence. Alternatively, this BBN based inference may also be applied to manually labeled landmarks whenever they are available, thus skipping the automatic landmarking stage.

The contributions of this paper can be summarized as follows:

1. The use of Bayesian Belief Net (BBN) for FER which fuses multiple evidence as revealed by both geometric and appearance features extracted from three different modalities of 3D face scans;
2. Bayesian inference using Statistical Feature Models (SFMs) and Gibbs–Bolzmann distribution;
3. Facial Expression Recognition using automatically located landmarks and its comparison with manually labeled ones;

Early versions of this work presenting step results appeared in [25,27]. The reminder of the paper is organized as follows. The BBN for FER is presented in Section 2. We then describe both the geometric and appearance features in Section 3. The Statistical Facial Feature Model (SFAM) used for automatically landmarking is shortly introduced and discussed in Section 4. Experimental evaluations are given in Section 5. Section 6 concludes the paper.

2. Bayesian Belief Net

In this section, we first give a short introduction to BBN and then present the proposed BBN for 3D facial expression recognition and the associated belief inference.

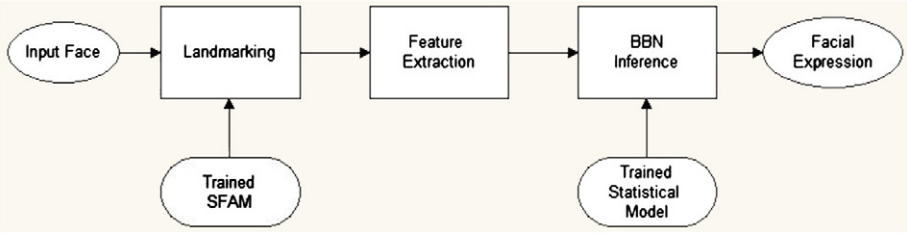

Fig. 1. The flowchart of the proposed 3D facial expression recognition system.

2.1. Overview of BBN

Bayesian Belief Net [28] is a probabilistic graphical model with the topology of directed acyclic graph (DAG). It is thus a graph formed by a set of nodes and directed edges without cycles. Nodes represent a set of random variables and directed edges represent their conditional dependencies.

In a given BBN, the 'belief' of variables ($X=(x_1,x_2,\ldots,x_n)$) on a node X describes its probability of states given the evidence e (observations) on its connected neighbor nodes. We divide these nodes into parents (those nodes pointing directly to X via an edge) and children (those nodes pointed directly from X via an edge) so as to compute the belief as:

$$P(X|e) \propto P(e^c|X)P(X|e^p) \tag{1}$$

where e^p is evidence on all parents and e^c evidence on all children. The first term can be rewritten as:

$$P(e^c|X) = P(e_1^c, e_2^c, \ldots, e_{Nc}^c|X) = \prod_{l=1}^{Nc} P(e_l^c|X) \tag{2}$$

where e_l^c is the evidence or observation of the lth child node, N_c is the number of children, $P(e_l^c|X)$ is the probability of evidence knowing the X state.

2.2. BBN for expression recognition

Our BBN is constructed as shown in Fig. 2. The node X represents the random variable having different states, each of which corresponds to a facial expression to be recognized. It can thus be one of the states corresponding to the six universal expressions or one of the facial AUs. The node S is X's parent whose states are human subjects displaying the facial expression in X. It thus has as many states as the number of subjects. X's children $F_1, F_2, \ldots, F_{N_f}$ represent the different types of facial features that can be extracted and used as evidence. In linking the X node directly to each feature node F_{N_i} as in Fig. 2, we simply state by assumption that all the types of features as represented by those nodes are conditionally independent each other with respect to the X node.

This BBN-based Bayesian formulation of FER is interesting as it says that the recognition of a facial expression should take into account not only the identity of a subject S_i but also the prior probability of facial expressions for a given subject $P(X|S_i)$. This statement is consistent with our intuition that the prior distribution of facial expressions is subject dependent, widely depending on the character of a person, e.g., a joyful person would mostly display a smiling face while a sad person in nature would exhibit sadness more frequently. While such a statement is very simple and intuitive, the computation of such a prior distribution $P(X|S_i)$, is not easy as it assumes plenty of additional observation data made available for each subject. In this work, we carried out FER experiments on two public datasets, namely BU-3DFE and Bosphorus, without any knowledge regarding to prior distributions $P(X|S_i)$. As a result, we assume in this work uniform distributions for $P(X|S_i)$ which becomes a simple constant and the parent node S can be ignored in the subsequent inference for 3D FER.

By introducing Eq. (2) into Eq. (1) and making use of the conditional independence of the children nodes of X while omitting the parent term, $P(X|e^p)$, which is a constant value C according to our previous assumption of uniform distributions, Eq. (1) can be rewritten as follows:

$$p(X|e_\kappa) \propto C\prod_{l=1}^{Nc} P(e_l^c|X) \propto \prod_{l=1}^{Nc} P(e_l^c|X) \tag{3}$$

where e_l^c is the observation of the lth child, N_c is the number of children, $P(e_l^c|X)$ is the conditional probability of evidence knowing the state X whereas e_κ refers to observations from a given 3D face scan κ. Thus, the belief for each expression state is computed from e_κ and the state holding the highest belief is considered as the most probable expression or Action Unit (AU) of the given 3D face scan κ.

2.3. Belief computation for BBN

In order to estimate $P(e_l^c|X)$ in Eq. (3), we adopt a statistical feature model (SFM) associated with the type of features within the lth child node in assuming that $P(e_l^c|X)$ follows the Gibbs–Boltzmann distribution. Accordingly, the probability $P(e_l^c|X)$ can be computed as the

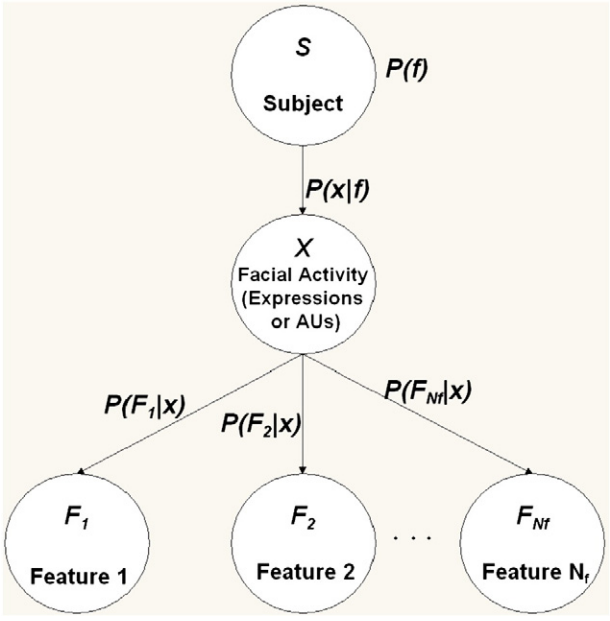

Fig. 2. The proposed Bayesian Belief Net for 3D FER in its general form. The node X represents facial expressions or AU states; the node S represents subject identity; The nodes F represent facial features. It depicts that facial activity X is subject dependent (S) and conditions the various features F_i that one can observe. However, without any specific priori knowledge of facial activities of the subjects as it is generally the case, e.g., those in the two public datasets, namely BU-3DFE and Bosphorus, we have assumed a uniform distribution of facial activities for each subject and simply ignored the parent node S in the BBN inference.

matching score between an instance F_l^x from the type of features in lth node with its equivalent $\hat{F}_l^x$ that is synthesized from the SFM associated with the type of features in lth node for the expression $X=x$. Specifically, $P\left(e_l^c|X\right)\propto e^{A_l Q_l^x}$, where Q_l^x is the match quality and A_l is a normalizing constant. Inserting the Gibbs distribution into Eq. (3) and taking the logarithm gives:

$$\log P(X|e_\kappa) = \log\left(\prod_{l=1}^{Nc} P\left(e_l^c|X\right)\right) + c = \sum_{l=1}^{N_c} A_l Q_l^x + c. \tag{4}$$

In this way, the Gibbs distribution converts the belief inference from the probability domain to the energy domain. The Q_l^x is computed as the normalized cross-correlation between evidence F_l^x and its instance $\hat{F}_l^x$. Finally, the belief for each state of facial expressions of the node X is computed as in Eq. (4) and the facial expression is recognized as in Eq. (5). is to discover the highest belief among all expression states, the constant c can be omitted.

$$X = \underset{X}{\operatorname{argmax}}\, P(X|e_\kappa) = \underset{X}{\operatorname{argmax}} \sum_{l=1}^{N_c} A_l \left\langle \frac{F_l^x}{||F_l^x||}, \frac{\hat{F}_l^x}{||\hat{F}_l^x||} \right\rangle \tag{5}$$

where X is a state of facial expressions, thus one of the six basic facial expressions in case of a judgment-based approach, i.e. anger, disgust, fear, happiness, sadness, surprise, or a state of AUs in the case of a forced-choice classification of a single AU.

2.4. SFM training and BBN testing algorithms

Given a type of features in a node in the BBN, a Statistical Feature Model (SFM) is built for each expression x to be recognized. Specifically, given a training set for a type of features F_l as represented by the lth node, we divide it into N_e (number of expressions or AUs) subsets. For each subset corresponding to the facial expression x, Principle Component Analysis (PCA) is applied to learn the major modes which explain the 95% variation of the type of features under the xth expression or AU [29]:

$$F_l^x = \bar{F}_l^x + P_l^x b_l^x \tag{6}$$

where l indexes the type of features represented by the lth node in the BBN, x designates a particular state of facial expressions to be recognized, $\bar{F}_l^x$ denotes the mean feature in the xth subset, P_l^x denotes a matrix composed of eigen-features, b_l^x denotes the control parameter vector. Each parameter ($b_{l_j}^x$) is supposed to follow a Gaussian distribution with zero mean and the standard deviation $\sigma_{l_j}^x$. The training algorithm is depicted in Algorithm 1.

In Algorithm 1, $j\in 1..N$, $l\in 1..N_f$, $x\in X$ which is the set of facial expressions to be recognized, $N_e=|X|$, N_f equals 15 in this work. The computational complexity is $O(N)$ for feature extraction and $O(N_f\times N_e)$ for SFM learning.

Algorithm 1. Training Algorithm for building statistical feature models (SFMs)

Input: N_e sets of textured face scans, corresponding to N_e different expressions or AUs. Each textured face scan contains a 3D face mesh, its texture and 19 landmarks (N face scans in total).

Output: $N_e * N_f$ learnt statistical feature models.

1. For each face scan j, extract N_f types of features from the local regions around the landmarks and concatenate them into N_f feature vectors F_l^j respectively.
2. For each type (l) of feature vectors:
   For each facial expression (x):
   Apply PCA to learn the statistical feature model: $F_l^x = \bar{F}_l^x + P_l^x b_l^x$.

Feature instances $\hat{F}_l^x$ can be generated from Eq. (6) using F_l to estimate the best parameter b_l^x by $b_l^x = P_l^{xT}\left(F_l - \bar{F}_l^x\right)$. The detailed steps can be found in steps 2a–2c in Algorithm 2. The belief or matching quality is computed in step 2d. Once computed the quality matrix Q, we convert it to a belief vector Q^x of size $N_e\times 1$ by adding beliefs to each column. Then, the expression or AU is recognized as the one having the highest value in the belief vector according to Eq. (5).

In Algorithm 2, $\langle\cdot,\cdot\rangle$ denotes the inner product and $||\cdot||$ denotes the L_2 norm.

Once all the features extracted, the computation complexity is $O(N_eN_f)$ for belief computation and $O(N_e)$ for the recognition of facial expressions. In our experiments, it takes on average around 0.24 s to compute beliefs for each child node using a desktop PC with Intel Core2 E4400 at 2.00 GHz CPU and takes less than 4 s to classify a testing data on the six universal expressions.

2.5. Discussion

Eq. (5) outputs a single state of the node X, i.e. one of the six basic facial expressions or a single AU while maximizing the posterior probability given the different types of features as input evidence. In this sense, the proposed BBN using Eq. (5) for FER is a judgment-based approach rather than a sign-based one. In the latter case, one needs first to recognize simultaneous facial muscle actions, e.g., AUs, occurring in combination during a facial expression, then proceed to the interpretation of these AUs into complex or subtle facial expression categories. Meanwhile, the proposed BBN can be slightly adapted to fit this case, using a threshold in Eq. (5), to recognize the AUs occurring at the same time:

$$P(X|e) > \xi. \tag{7}$$

Algorithm 2. Algorithm for belief computation

Input: A textured face scan for a given subject s, which contains its 3D face mesh, its texture and the 19 landmarks.

Output: A matrix of beliefs Q with size $N_e\times N_f$.

1. Extract N_f types of features from 19 local regions and concatenate them into N_f feature vectors F_l^x respectively.
2. For each vector corresponding to the type of features (l):
   For each facial expression (x):
   a. Apply the trained SFM to the input feature F_l^x to estimate the control parameter b_l^s by $b_l^s = P_l^x\left(F_l^x - \bar{F}_l^x\right)$,
   b. Limit the range of control parameters to increase the separability among classes by applying the function f

$$\hat{b}_{l_j}^s = f\left(b_{l_j}^s\right), f(t) = \begin{cases} b_{l_j}^s, abs\left(b_{l_j}^s\right) < 0.5\sigma_{l_j}^x \\ 0.5\sigma_{l_j}^x, abs\left(b_{l_j}^s\right) \geq 0.5\sigma_{l_j}^x \end{cases}.$$

   c. Generate the instance $\hat{F}_l^x$ for the feature F_l^x by $\hat{F}_l^x = \bar{F}_l^x + P_l^x \hat{b}_l^s$,
   d. Compute the matching quality Q_l^x of the feature F_l^x on the facial expression x by $Q_l^x = \left\langle \frac{\hat{F}_l^x}{||\hat{F}_l||}, \frac{F_l^x}{||F_l^x||} \right\rangle$,

Graphical models have been already used in 2D facial expression analysis. A dynamic Bayesian Network was developed in [30] to model the dynamic and semantic relationships among facial action units. This network was extended to a more sophisticated one in [31] which performs a joint analysis of head pose and action units. A Bayesian Belief Network aiming at modeling the relationship between expressions and facial action units was also proposed in [32] for FER. However, the proposed BBN differs from them in three aspects. Firstly, the proposed BBN carries out FER in 3D whereas all the aforementioned works, e.g. [30,32], uses BBN to perform FER in 2D. Secondly, the structure of our BBN is different. In

[30], the learnt structure of the Bayesian Network explores the dynamic relationship among AUs. In [32], the structure of BBN describes the relationship between AUs and the six universal expressions. In contrast, the proposed BBN concentrates on describing the causal relationship among subject, facial activity (Expressions and AU) and facial features. Thirdly, we propose a novel method for parameter computation based on SFM which is different from the aforementioned works.

3. Extraction of features

In order to retrieve evidence for exaggerated facial displays as well as subtle ones, we propose to make use of both geometric features, thereby accounting for deformations of intransient facial features e.g., eyelids, eyebrows, the mouth, as well as appearance features in terms of both texture and shape, aiming to characterize occurrence of transient features (e.g., wrinkles, furrows, bulges). Savran et al. [33,34] have shown recently that geometric and appearance features carry complementary information and their joint use does indeed improve 3D FER. Therefore, we have not sought to reinvent the wheel each time but instead to capitalize on the best practice in the existing works on FER in terms of features sensitive to facial expression. The originality of our approach is rather its hybrid nature, as we not only explore both geometric and appearance features but also combine global features with local ones. Specifically, we extract 15 types of geometric and appearance features from three modalities as in Fig. 3 to take full advantage of the wealth of information contained in textured 3D face scans. In this section, we explain in detail the features extracted from each modality. The extensive experimental evaluation described in Section 5 will further highlight the relevance of each modality in FER, in particular in terms of discriminating power.

3.1. Morphology features

Morphology features aim at capturing global 3D facial geometry changes during facial displays. In this work, we made use of a set of predefined facial landmarks whose configuration is used to define morphology features. Specifically, as illustrated in Fig. 3, 19 anthropometric landmarks were used. They encompass the corners of the eyes and mouth, the nose tip, etc. They were chosen because they were automatically located by SFAM in a previous work [26] and enable a further study in Section 5 comparing manually labeled landmarks with automatically located ones on FER. Alternatively, more landmarks can be used whenever available.

We first define the configuration of the landmarks by vector S in staking the 3D coordinates of all the landmarks. In our case, the eye centers are interpolated by averaging corresponding eye corners when using automatically located landmarks.

$$S = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N)^T \tag{8}$$

where N is the number of landmarks, 19 in this work, which are located on different facial components sensitive to deformation when a facial expression occurs, e.g., the mouth corners, the eye corners, the eyebrow corners, etc.

Several works [17–19] showed the effectiveness of simple distances between landmarks in FER. We thus use the vector S to compute a new feature vector L, which is formed by all the distances pictorially shown as green lines in Fig. 3(a). These distances are empirically chosen to best describe the configuration relationships among facial components under different facial expressions. For the purpose of comparison, we also used the correlation-based feature subset selection (CFS) [35] to automatically select a subset of distances from all the 381 distances that we can compute between 19 landmarks, as shown in Fig. 3(b). This feature selection method was chosen because it evaluates the value of a subset of distances by considering the individual predictive ability of each along with the degree of redundancy between them. Accordingly, distances that are highly correlated with an expression class while having low intercorrelation are preferred. We evaluated these two sets of distances on FER using 3D face scans of 60 subjects from the BU-3DFE dataset which display the two highest intensity levels of the six prototypical facial expressions (see Section 1). The distances automatically selected by CFS achieved a recognition rate of 73.3% while the empirically selected distances as depicted in green lines in Fig. 3 achieved a recognition rate of 75.3%. Therefore, the empirically selected distances as staked in the feature vector L are used in the subsequent experiments.

We further extract a landmark displacement feature vector D, which measures the displacement of each landmark when an expression occurs from a neutral display. This feature vector is thus very informative since it directly measures the shape difference between the face displaying a facial expression and a neutral one. However, it also imposes the constraint of one neutral face from each subject being available for comparison and therefore is subject biased. To remove this constraint for subject independent FER, we use a mean vector of landmark locations computed from all training neutral faces instead of using the landmark locations on a neutral face of a given subject. Thus, D is computed by subtracting from S the mean of landmark locations from the training neutral faces ($\bar{S}_{\text{neutral}}$) as in Eq. (9). They are illustrated in red lines in Fig. 3.

$$D = S - \bar{S}_{\text{neutral}} \tag{9}$$

3.2. Local texture features

Many existing works on 2D FER [12,13] have shown the effectiveness of texture-based features. As illustrated in Fig. 3, we also extract several local texture features to account for appearance features. We first form the raw vector of texture T by simply stacking the intensity values from the remeshed grids centered at each landmark:

$$G = (g_1, g_2, \ldots, g_m)^T \tag{10}$$

where m is the number of intensity values in all the remeshed grids associated with the landmarks.

To further capture fine facial texture details, we also make use of Local Binary Pattern (LBP) operator which is a simple and powerful texture descriptor widely used in 2D face analysis [36]. Specifically, we extract from local texture patches Multi-Scale LBP [37], namely $LBP^{U^2}_{(16,1)}t$, $LBP^{U^2}_{(16,2)}t, \ldots, LBP^{U^2}_{(16,5)}t$, thus at scales from 1 to 5, to characterize both quite noticeable facial displays as well as subtle ones around each landmark. $LBP^{U^2}_{(P,R)}t$ feature labels each pixel in the image with a number of binary values ($P = 16$ in our case) calculated by thresholding the sampled neighborhood (16 black dots in Fig. 4) with each pixel (the white dot in Fig. 4). R is the Scale, 1–5 in our case, which designates the radius of the circle for neighborhood sampling, shown as the circle in Fig. 4. Superscript u^2 indicates that the LBP relates to uniform patterns with a U value of at most 2 [38]. The LBP values using each (P,R) pair extracted from the local grids around the landmarks are then concatenated into a new feature vector.

Alternatively, one can also use Gabor features on several scales and orientations which have been shown very effective in FER but computationally quite expensive [12,34].

3.3. Local shape features

In order to account for local deformations of a facial surface, sometime without provoking significant texture changes as it is the case for subtle facial expressions, we also derive several local shape features. The very simple one is the depth vector Z by staking all the depth values from all the local grids associated with the landmarks:

$$Z = (z_1, z_2, \ldots, z_m)^T. \tag{11}$$

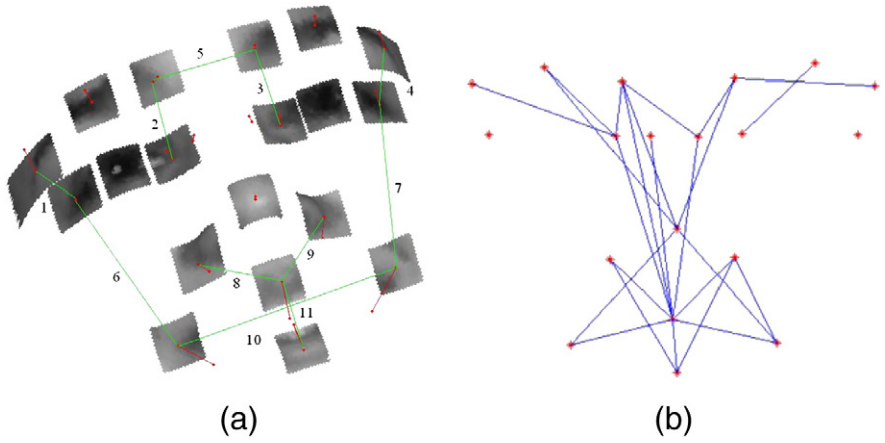

Fig. 3. (a) Extraction of the features from three modalities: global morphology features defined by the configuration of the landmarks, local texture features and local shape features extracted from remeshed local grids centered at each landmark. Local patches in the figure correspond to the remeshed grids formed by local shape and rendered by local texture. The green lines correspond to the manually selected distances between landmarks. (b) Automatically selected distances between landmarks by the correlation-based feature subset selection method.

To better characterize local surface properties, we also compute Multi-scale LBP values on the local range grids associated with the landmarks, namely, $LBP^{U2}_{(16,1)}r$, $LBP^{U2}_{(16,2)}r, \ldots, LBP^{U2}_{(16,5)}r$. The extracted range LBP values on local grids using each (P,R) pair are also concatenated into a vector respectively.

The description of local surface properties is further enhanced by shape index [39] that we compute on all points on the local grids. They are concatenated into a vector *SI*. Recall that curvature-based features prove to be very useful in 3D FER [20,9,34] and shape index is widely used in 3D face recognition [40,41].

To summarize, 15 types of features are extracted from the three modalities of a textured 3D face model and used in the BBN as evidence in the children nodes. Table 1 summarizes these types of features along with their dimension.

4. Statistical facial feature model for automatic landmarking

As we aim to perform fully automatic 3D FER, we briefly present the Statistical Facial Feature Model (SFAM), first proposed in [25,26], which is used to locate landmarks on 3D faces with expressions. The use of automatically located landmarks also gives rise to a comparison study in Section 5 with the state of the art on 3D FER which mostly uses manually labeled facial landmarks.

4.1. Model building

In order to efficiently learn variations on the global morphology, local texture and local shape among training faces, a preprocessing stage is first performed to exclude variations introduced by global factors like head pose or face scale. Local grids are then used to remesh local regions centered at 19 landmarks (shown in Fig. 5(a)). Intensity and range data are extracted from these grids, as in Fig. 5(b),(c). This process ensures that the same number of points is sampled from all training faces and that they are matched point-to-point.

SFAM is then learnt by applying PCA respectively to the three types of features from training faces, preserving 95% of variations for each one. The resulting model is given in Eq. (12)

$$s = \bar{s} + P_s b_s, g = \bar{g} + P_g b_g, z = \bar{z} + P_z b_z \tag{12}$$

where $\bar{s}$, $\bar{g}$, and $\bar{z}$ are respectively the mean morphology, mean intensity and mean range vectors while P_s, P_g, and P_z are their learnt variation components respectively obtained from PCA. b_s, b_g, and b_z are the corresponding sets of control parameters.

Partial face instances, corresponding to local face regions with texture and shape configured by their morphology vector, can be synthesized by a linear combination of these components, as shown in Fig. 6. This SFAM is built from face scans displaying the six universal expressions. Thus, each learnt variation mode is a mixture from all those expressions. This facilitates the landmarking process on faces with different expressions.

4.2. Automatic landmarking

Automatic landmarking on an input 3D face scan can be considered as a SFAM fitting process. The fitting process is to maximize an objective function:

$$f(b_s) = \alpha \sum_{i=1}^{N} F_{gi}(s_i) + \beta \sum_{i=1}^{N} F_{zi}(s_i) - \sum_{j=1}^{k} \frac{b_j^2}{\lambda_j} \tag{13}$$

where N is the number of local regions, F_{gi} and F_{zi} are the aforementioned normalized cross correlation (Eq. (5)) between the synthesized local instances of the texture and range patches and their corresponding local patches on an input scan, α and β are weight constants, k is the number of retained landmark configuration modes and λ_j denotes the

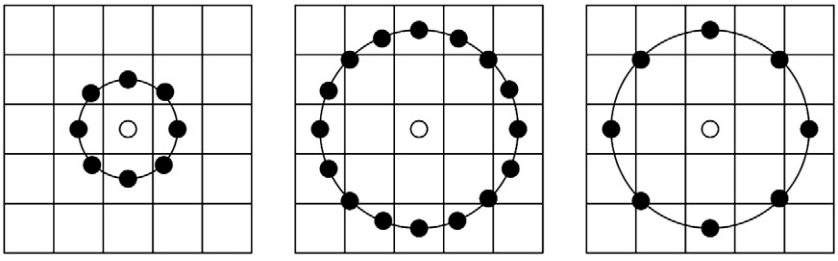

Fig. 4. LBP Operator. The circular (8,1), (16,2), and (8,2) neighborhoods. The pixel values are bilinearly interpolated whenever the sampling point is not at the center of a pixel.

Table 1
15 types of features and their textual descriptions.

Symbol	Textual description	Dimension
D	Person independent point displacement of 19 landmarks	11
Z	Range values extracted from 19 local patches	4275
G	Intensity values extracted from 19 local patches	4275
L	Selected distances between 19 landmarks	57
SI	Shape index extracted from 19 local patches	4275
$LBP^{u^2,t}_{(16,1)}$	LBP feature extracted at scale 1 from 19 local texture maps	4275
$LBP^{u^2,t}_{(16,2)}$	LBP feature extracted at scale 2 from 19 local texture maps	4275
$LBP^{u^2,t}_{(16,3)}$	LBP feature extracted at scale 3 from 19 local texture maps	4275
$LBP^{u^2,t}_{(16,4)}$	LBP feature extracted at scale 4 from 19 local texture maps	4275
$LBP^{u^2,t}_{(16,5)}$	LBP feature extracted at scale 5 from 19 local texture maps	4275
$LBP^{u^2,r}_{(16,1)}$	LBP feature extracted at scale 1 from 19 local range maps	4275
$LBP^{u^2,r}_{(16,2)}$	LBP feature extracted at scale 2 from 19 local range maps	4275
$LBP^{u^2,r}_{(16,3)}$	LBP feature extracted at scale 3 from 19 local range maps	4275
$LBP^{u^2,r}_{(16,4)}$	LBP feature extracted at scale 4 from 19 local range maps	4275
$LBP^{u^2,r}_{(16,5)}$	LBP feature extracted at scale 5 from 19 local range maps	4275

corresponding eigenvalue in the landmark configuration model. b_j denotes the control parameter that generates the landmark configuration s given the statistical model. The values of α and β are fixed and are computed as the ratios of $\sum_{i=1}^{N} F_{gi}$ over $\sum_{j=1}^{k} \frac{b_j^2}{\lambda_j}$, and $\sum_{i=1}^{N} F_{zi}$ over $\sum_{j=1}^{k} \frac{b_j^2}{\lambda_j}$, respectively, during the off-line training. During the fitting process, the first two terms F_{gi} and F_{zi} are computed as two response meshes (Fig. 7), describing respectively the similarity between the local texture and its corresponding instance from SFAM, and the similarity between local shape and its corresponding instance. High values imply high chance to locate the landmark, since the corresponding local texture and range match the texture and range instances by SFAM given the landmark configuration s. The third term represents the Mahalanobis distance, introduced to limit the generation of unplausible configurations since high absolute values of b_j results in the outlier of synthesized configurations s in Eq. (12).

The fitting algorithm as described in Algorithm 3 encompasses five steps to locate the landmarks $\hat{S}$. The optimization in step 2 and 5 is processed by Nelder–Meade simplex algorithm [42] for its robustness to local minimum. More details can be found in [26].

Algorithm 3. SFAM fitting for landmarking input

Input: A textured 3D scan and a trained SFAM.
Output: Optimized morphology parameters.

1. Given a 3D face, its head pose is first compensated using ICP algorithm.
2. The morphology parameters b_s are optimized to minimize the distance between corresponding morphology instances and their closest points on the input face. In this process, only landmarks located on the rigid facial parts are involved, such as those in the eyebrow, eye and nose regions.
3. Synthesize texture and shape instances based on the optimized morphology from the previous step.
4. Correlation meshes are computed over 19 facial regions by cross-correlating the texture and shape instances with local texture and shape samples obtained from a neighborhood around potential landmark locations.
5. Morphology parameters are optimized to maximize Eq. (13), i.e. the sum of values on two correlation meshes on all 19 regions while minimizing the distance associated with the landmarks configuration defined by the control parameters.

5. Discussion

By combining BBN with SFAM, a fully automatic 3D facial expression recognition system can be realized. Recall that it consists of four main stages, as shown in Fig. 1: offline SFAM construction, offline BBN training, online landmarking and feature extraction, and finally online facial expression/AU recognition. Offline, SFAM is trained using a set of textured 3D faces having different facial expressions. As described in Section 2, a set of statistical feature models (SFMs) are also trained offline for each class of facial expression to be recognized and each feature used as evidence in BBN. During online recognition, faces are first landmarked by SFAM, and then the 15 types of features are extracted and used as evidence by BBN for Bayesian belief inference for the recognition of facial expression as represented by the node X. The output of the system is thus class of facial expressions whose corresponding state has the highest belief among different expressions or AUs.

As compared to other 3D model-based approaches requiring computationally costly dense registration of point clouds [43,21,23], SFAM is a morphable face model built on partial faces, which only requires a sparse correspondence in the registration. Thus, it is computationally much cheaper. Moreover, as each instance is only composed of local patches, SFAM is able to deal with partially occluded faces by simply masking the occluded regions in the fitting process [26], as well as hard facial expressions like opening of the mouth as shown in Section 5.

6. Experimental results

Extensive experimental evaluations have been conducted on two public databases, namely the BU-3DFE dataset for the recognition of the six universal expressions using both automatically and manually located landmarks and the Bosphorus dataset for the recognition of 16 AUs featuring ample facial deformations as well as subtle ones. These experimental evaluations have also compared the proposed BBN with two other classifiers, namely SVM and Sparse Representation Classifier (SRC). In this section, we first introduce the two used datasets and the general experimental setup, and then compare the automatic landmarking results with the manual landmarks in terms of accuracy. Subsequently, we describe the experimental results

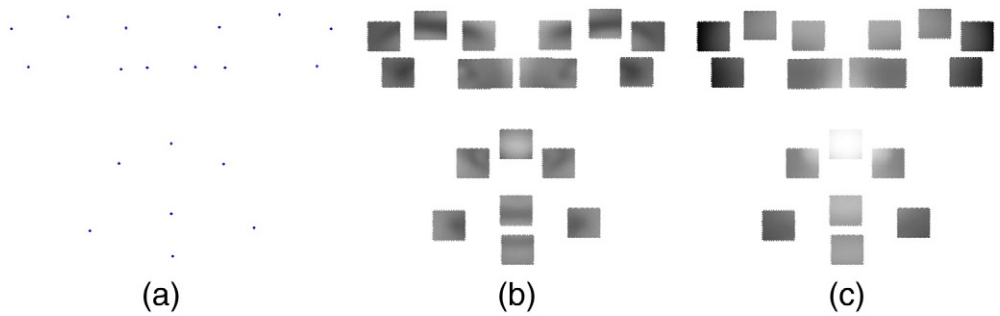

Fig. 5. Three types of features extracted from a textured 3D face scan for building the SFAM. a: 3D morphology as defined by the locations of 19 landmarks, b: intensity values on the remeshed local regions around each landmark, and c: range values on the remeshed local regions around each landmark.

Fig. 6. Instances of the SFAM at the $\pm 3\sigma_{it}$ ending corresponding to $b_{s_1}, b_{g_1}, b_{z_1}$. The first morphology mode explains blend variations in terms of the face size and expression; the first texture mode explains skin color variations; the first range mode explains local curvature variations.

respectively on the BU-3DFE and the Bosphorus datasets. These results are further discussed in comparison with literature.

6.1. Databases and experimental setup

The BU-3DFE database [15] contains 100 subjects (56% female, 44% male) with a variety of ethnic/racial ancestries. Each subject performs seven expressions in front of the 3D face scanner, i.e. the six universal expressions (happiness, disgust, fear, anger, surprise and sadness) and the neutral. Each of the six universal expressions is displayed with four levels of intensity, from the weakest to the strongest. In our experiments, we have considered the two highest intensity levels: level 3 and level 4. Generally, facial scans in level 4 capture the apex of a facial expression whereas scans in level 3 capture its onset.

We have also used the Bosphorus dataset [44] for AU recognition. It contains 4666 face scans from 105 subjects. This dataset contains not only the six universal facial expressions, but also 3D face scans displaying AUs. However, the number of AUs is not evenly distributed over the subjects. The number of acted AUs per subject[1] varies from 6 to 23. Thus, in the following experiments, subjects have been selected based on the availability of acted AU scans.

All tests on AU and facial expression recognition followed a 10-fold subject-independent cross-validation process. Subjects were randomly separated into 10 groups. In each round, subjects in 9 groups were used for training and the subjects in the remaining

[1] There are usually more than one AU according to the expert FACS scoring, where the acted AUs are often the most dominant AUs in the actual FACS codes.

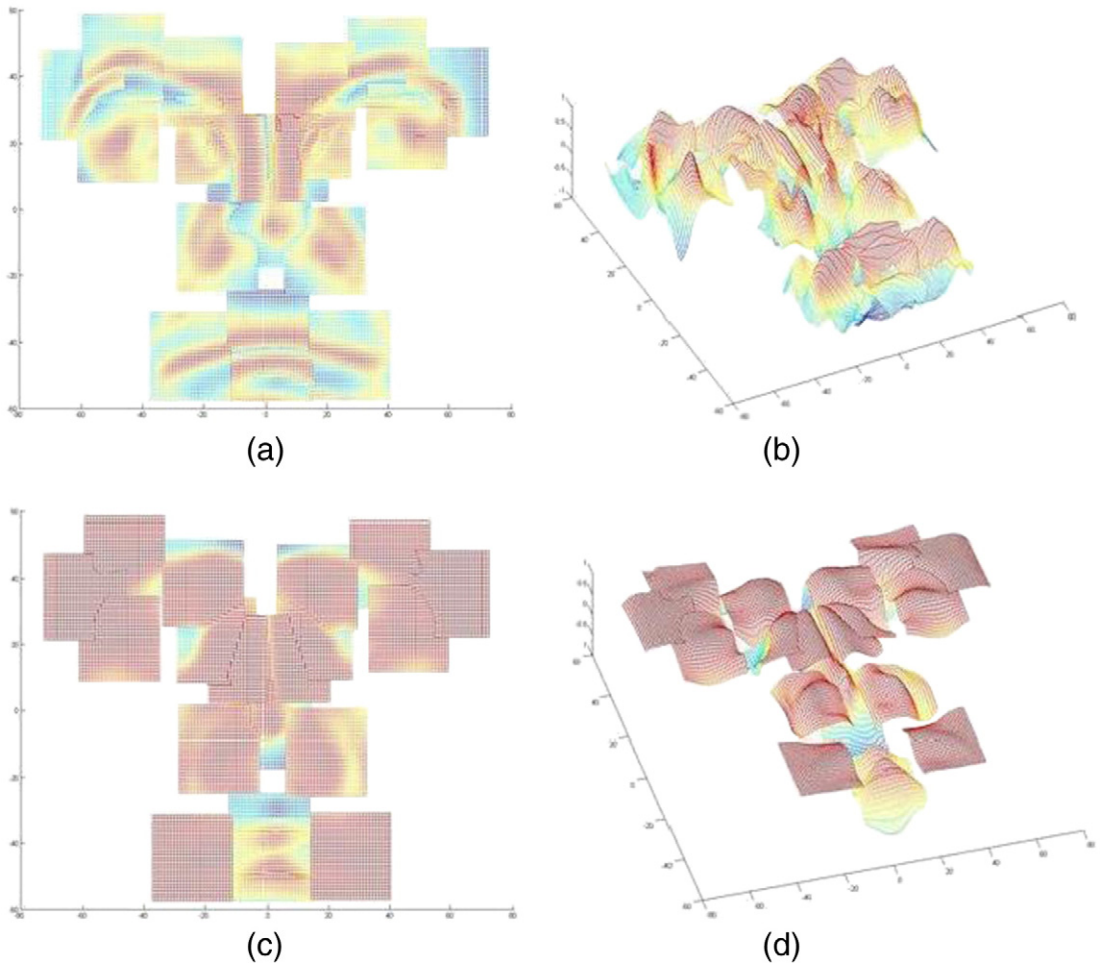

Fig. 7. Correlation meshes from two viewpoints. (a) and (b) are the same response mesh from 2 point of views, describing the similarity of texture instance from SFAM and texture on the given face. (c) and (d) are the correlation mesh describing the similarity of shape instance from SFAM and face shape.

group used for testing. The experiments were conducted 10 times so that 3D face scans of subjects in each group were utilized as test once. This experimental setup guarantees that each subject appears once in the testing set and 9 times in training set and any subject used for testing does not appear in the training set as the partition is based on the subjects rather than individual 3D face scans.

6.2. Results of automatic landmarking

In the experiment on landmarking, we trained our SFAM using 143 face scans from 11 subjects, 6 females and 5 males, out of 100 subjects from the BU-3DFE dataset. These face scans encompass the ones displaying the neutral expression as well as the two highest intensities for each of the six prototypical facial expressions. We then applied SFAM then to locate 19 landmarks on 1157 3D face scans of the other 89 subjects in this dataset. This experimental setup enables comparison of the result with that by [45] which is based on a 3D Point Distribution Model (PDM).

Fig. 8 illustrates several locating examples with facial expression. Table 2 summarizes the mean errors (the first row) of the landmarking algorithm in comparison with the manually labeled landmarks used as ground truth. Table 2 also provides the standard deviations in the second row. For comparison purpose, the third row lists the mean errors of landmarking in [45] which only locates five anthropometric points. As we can see from the table, the mean errors for most landmarks remain within 5 mm, and most of standard deviations are lower than 5 mm. Compared to the results by [45], our approach locate more landmarks with a higher accuracy. This improvement can be explained by the fact that the Point Distribution Model in [45] merely uses landmark locations as features and only takes into account the configuration relationships of landmarks. In contrast, SFAM characterizes each landmark location through its global configuration relationships as well as its local properties in terms of texture and geometric shape, thereby enabling increased landmarking accuracy.

(1: left corner of left eyebrow, 2: middle of left eyebrow, 3: right corner of left eyebrow, 4: left corner of right eyebrow, 5: middle of left eyebrow, 6: right corner of right eyebrow, 7: left corner of left eye, 8: right corner of left eye, 9: left corner of right eye, 10: right corner of right eye, 11: left nose saddle, 12: right nose saddle, 13: left corner of nose, 14: nose tip, 15: right corner of nose, 16: left corner of mouth, 17: middle of upper lip, 18: right corner of mouth, 19: middle of lower lip).

6.3. Results on FER using the BU-3DFE dataset

We first experimented BBN on the BU-3DFE dataset for subject independent recognition of the six universal facial expressions. Even though BU-3DFE database provides 3D face scans from 100 subjects displaying the six prototypical facial expressions with 4 intensity levels, most methods in the literature only make use of face data from 60 subjects displaying the highest intensity levels, i.e. level 3 and 4, in their experiments. To enable comparison with the literature, we also used face data of 60 subjects in our two experiments using either manually labeled or automatically located landmarks. Subjects

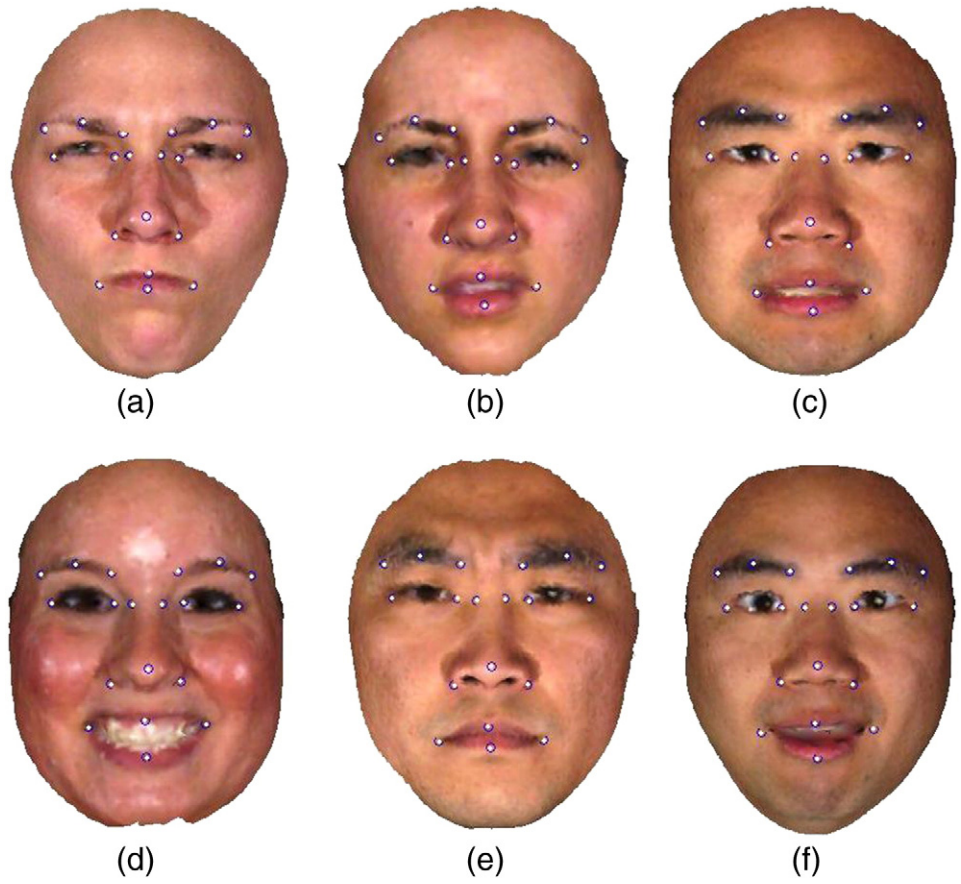

Fig. 8. Automatic landmarking examples from the BU-3DFE dataset with expressions of anger (a), disgust (b), fear (c), joy (d), sadness (e) and surprise (f).

were selected randomly but were not necessarily the same for these two tests, since in the case of using automatically located landmarks, we need to exclude those subjects used for building the aforementioned SFAM in order not to bias the results. As the proposed BBN can be considered as a principled way to fuse different types of features for the purpose of FER, we further studied different fusion schemes both at feature and score level in comparison with BBN.

6.3.1. BBN for FER

The first experimental evaluations featured several comparisons for subject independent FER: the use of automatically located landmarks versus manually labeled ones, the proposed BBN as classifier versus the Support Vector Machine (SVM) [46] and the Sparse Representation Classifier (SRC) [47], and the discriminating power of the features extracted from each modality and modality combination. All tests followed a 10-fold cross-validation process as described in Section 1. Face scans in levels 3 and 4 were tested separately and the final recognition rate was obtained by averaging the results from the two intensity levels for all the three classifiers. For all tests using manual landmarks, feature extraction and classifier training were based on manual landmarks. For all tests using automatic landmarks, feature extraction and classifier training was based on automatic landmarks.

For classification tests using SVM, a multi-class SVM (one-against-all) using RBF kernel was trained respectively for each of the 15 types of features extracted from face scans of each of the two facial expression intensities, thus leading to 30 SVMs in total. The output of the SVMs is a set of probabilities describing how likely a face scan belongs to each expression class according to the type of features under test. These probabilities (15 in total per level) were simply added together and the testing face was then labeled by the facial expression class having the maximum probability score. Grid search was used for choosing the best parameters (c, g) over the 10 folder cross-correlation. This process was repeated for each type of features and for each facial expression intensity level of 3D face scans, thus repeated 30 times.

For classification tests using SRC, 30 SRCs were respectively trained following the approach proposed in [47]. The principle of these SRC is to represent a test sample using an overcomplete dictionary whose elements, or atoms, were training faces represented through a given type of features. The sparse coefficients used to describe the test sample according to atoms were obtained via a $l^1 - norm$ minimization by an orthogonal matching pursuit. As in the tests using SVM, parameters in SRC were set empirically to obtain the best performance. The SRC output is a set of distances between the input feature and its six approximations generated from the sparse coefficients associated with each facial expression class. These distances (15 in total per

Table 2
Mean error (1st row) and the corresponding standard deviation (2nd row) of the 19 automatically located landmarks on 3D face scans from 89 subjects in the BU-3DFE dataset, all expressions included. The 3rd row gives the mean errors of the five landmarks automatically located by [45].

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Mean	6.26	4.58	4.87	4.88	4.51	6.07	4.11	2.93	2.90	4.07	3.30	3.27	3.32	4.04	3.62	7.15	4.19	7.52	8.82
Std	3.72	2.82	2.99	2.97	2.77	3.35	1.89	1.40	1.36	2.00	1.70	1.56	1.94	1.99	1.91	4.64	2.34	4.75	7.12
Mean′	–	–	–	–	–	–	20.46	12.11	11.89	19.38	–	–	–	8.83	–	–	–	–	–

expression intensity level) were simply added together and the testing face was labeled by the class having the minimum distance. For tuning SRC training, we selected the best value from candidates [50, 60, 70, 80, 90, 100] as the max number of coefficients in terms of recognition rate also over 10 folder cross validation. The process was repeated for each type of features and for each intensity level, thus repeated 30 times as well. Note that in the case of BBN, this parameter tuning process was not necessary since the unique parameter used in BBN is the boundary of control parameter used in Algorithm 2.

The left three columns depict the recognition rates and standard deviations of the six universal expressions based on manual landmarks, using SVM, SRC, and the proposed BBN, respectively. The right three columns depict the recognition rates and standard deviations based on automatic landmarks. The first three rows are the results using features from each single modality, respectively morphology features (M), local texture features (T), local shape features (S), the following three rows display the results using every combination of two modalities, and the last row shows the results for all features from the three modalities.

Table 3 depicts all the performance figures using different experimental configurations. Several lessons can be drawn from that table. Firstly, the first three rows compare the discriminating power of the feature sets from the three modalities, namely morphology (M), local texture (T) and local shape (S) and the feature set from local shape proves to be the best one for all the three classifiers regardless of whether the landmarks are manually labeled or automatically located. This is consistent with the fact that 3D face scans capture accurate facial surfaces which are sensitive to facial deformations due to facial expressions. Secondly, When fusing feature sets from two different modalities (row M + T,M + S and T + S)for FER, the best combination in terms of the recognition rate is T + S, thus the one combining local texture and local shape features, which displays almost the best performance for all the three classifiers. The combination (M + S) of the morphology and local shape features performs secondly, further highlighting the discriminating power of the shape features in FER. Third, the performance of the three classifiers using only morphology features, which are extracted from the configuration of the landmarks (e.g., distances among 19 landmarks and their displacements), highly depends upon the location accuracy of the landmarks involved. As we can see from the row M, there is a drop on performance of nearly 30% when switching from manually labeled landmarks to automatically located ones. Meanwhile local texture features (row T) and local shape features (row S) display relatively stable performance by all the three classifiers when switching from manual landmarks to automatic ones. However, all three classifiers record some performance drop because of errors in landmark locations when switching from manually labeled landmarks to automatically located ones.

When comparing BBN to SVM and SRC, we can see that SVM performs the best in all the tests using the feature set from one single modality, either morphology (M), texture (T) or shape (S), whereas BBN and SRC are clearly behind. However, the proposed BBN with its Bayesian inference proves to be an efficient fusion engine which improves its recognition accuracy when more evidence is added. Indeed, when fusing feature sets from two different modalities, BBN performs slightly better than SVM in the case of using manually labeled landmarks. When all the feature sets from the three modalities (row M + T + S) are fused, BBN keeps improving and performs slightly better than SVM regardless of whether manually labeled or automatically located landmarks were used.

Table 4 is the confusion matrix of the proposed BBN for the recognition of the 6 prototypical facial expressions using the 15 types of features extracted from 19 landmarks manually labeled (first value in each cell) or automatically located by SFAM (second value in each cell). The average recognition rates are 89.2% and 84.9% respectively. In both case, the best recognized facial expressions are happiness and surprise certainly due to their large deformation on face meshes while fear is the least recognized since its facial display is quite subtle. We also discover that fear and sadness expressions record the most remarkable performance drop when switching from manually labeled landmarks to automatically located ones. This can be explained by the fact that the accuracy of landmark locations is more important for subtle facial expressions than exaggerated ones.

6.3.2. BBN vs. feature level fusion

As the proposed BBN can be considered as a score level fusion method, we also carried out a comparison with a feature level fusion scheme that we develop in this subsection. For this purpose, we made use of the whole BU-3DFE dataset (100 subjects) and extracted 15 types of features from each face scan which, once normalized into [0 1], were further packed into a single feature vector. Then, PCA was applied to reduce the feature vector dimension from 55608 to 650 so that 98% of data variations were preserved. SVM and SRC were utilized to classify the six universal expressions using the subject-independent 10 fold cross-validation. Grid search was used for choosing the best parameters for SVM (linear kernel, c 4, g 0.0014). SRC parameter was empirically chosen to obtain the best performance.

Tested on both the levels 3 and 4 data (each 650×600), the average recognition rate is 59.9% and 64.9% for classifying the six expressions using SVM and SRC respectively. In contrast, with the same experiment setup and raw data (levels 3 and 4, each $55608 * 600$), the proposed BBN achieves a recognition rate of 80.9%. This experiment tends to demonstrate that feature level fusion, in packing together features of different nature into a single feature vector, is not effective for FER and probably explains why most FER techniques in the literature perform fusion rather at score level.

6.3.3. BBN vs. late fusion schemes

In its Bayesian inference using statistical feature models (SFMs), BBN fuses, through a simple sum rule as in Eq. (4), all the matching scores computed between each of 15 types of features, l, extracted from an input 3D face scan with the corresponding SFM associated with each facial expression class x. Alternatively, some other late fusion schemes, at score, rank or decision level, e.g., product, Borda count, plurality voting, max, min, could also be used [48]. In this subsection, we propose to compare BBN with two different late fusion schemes, namely two stage classification and score, rank and decision fusion.

Table 3
Average recognition rates for the six universal expressions with different feature configurations and classifiers on both manual and automatic landmarks. The standard deviations over 10 fold tests are the values in the brackets.

	Manual			Automatic		
	SVM	SRC	BBN	SVM	SRC	BBN
M	83.6% (4.4%)	61.7% (7.9%)	76.9% (8.7%)	54.2% (8.2%)	35.6% (7.7%)	51.1% (9.9%)
T	76.9% (6.8%)	74.7% (6.4%)	75.8% (7.5%)	74.2% (7.2%)	67.5% (7.7%)	67.8% (6.2%)
S	84.3% (5.6%)	81.3% (6.7%)	82.9% (5.9%)	80.8% (6.1%)	74.7% (6.5%)	77.3% (6.3%)
M + T	83.1% (6.1%)	78.3% (8.2%)	84.9% (5.8%)	76.7% (7.8%)	62.5% (6.1%)	67.8% (6.2%)
M + S	86.4% (5.7%)	83.1% (5.6%)	86.5% (5.1%)	80.8% (6.2%)	73.6% (6.7%)	77.5% (6.7%)
T + S	87.2% (4.3%)	83.5% (7.0%)	86.1% (4.5%)	83.1% (6.4%)	79.7% (7.1%)	84.9% (5.4%)
M + T + S	88.1% (4.1%)	85.3% (6.8%)	89.2% (3.6%)	82.8% (8.7%)	77.2% (6.6%)	84.9% (5.9%)

Table 4
Confusion matrix of the subject-independent expression recognition by the proposed BBN with all the 15 types of features. Left value on each cell is the result based on manual landmarks and right value the result based on automatic landmarks.

Input\output	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	86.7/83.3%	2.5/3.3%	1.7/1.7%	0.0/0.0%	9.1/11.7%	0.0/0.0%
Disgust	3.3/3.3%	89.3/86.7%	3.3/6.7%	0.8/0.0%	3.3/0.0%	0.0/3.3%
Fear	1.7/5.1%	6.7/8.5%	79.1/67.8%	6.7/8.5%	5.0/1.7%	0.8/8.5%
Happiness	0.0/0.0%	0.0/1.7%	5.8/3.3%	94.2/93.3%	0.0/0.0%	0.0/1.7%
Sadness	6.7/13.3%	0.8/0.0%	2.5/1.7%	0.0/0.0%	90.0/83.3%	0.0/1.7%
Surprise	0.0/1.8%	1.7/1.8%	2.5/1.8%	0.0/0.0%	0.0/0.0%	95.8/94.6%

Regarding the two stage classification fusion scheme, the basic idea is to feed matching scores output by BBN to another classifier, e.g., SVM or LDA, to improve recognition accuracy. For this purpose, we concatenated, for each face scan, score sets (6 scores per set corresponding to the six prototypical facial expressions to be recognized) from all the 15 types of features into a vector (90×1). Scores of all 360 scans from 60 subjects were combined into one score matrix (360×90) for each level. SVM and LDA were used for classification following the 10-fold cross-validation approach to these matrixes. For each fold, SVM and LDA were thus trained on score vectors of 54 subjects and tested on the remaining ones. The final recognition rate was averaged on recognition rates achieved on 3D facial expression scans of intensity levels 3 and 4. For the experiment denoted "N-sum" in Table 5, we first normalized 6 scores for each feature type into the range between 0 and 1. These 6 scores were then summed up over the 15 types of features and the class of facial expression having the highest score was then declared as the recognized class.

As it can be seen from Table 5, BBN achieved the best recognition rate whereas the two stage score-space based methods using LDA or SVM were several point behind. N-Sum was in between these two performances. This performance difference can be explained by the fact that the proposed BBN is a generative approach for FER. Given a type of features among the 15, we learn a priori knowledge on feature variations for each facial expression through a Statistical Feature Model (SFM). The recognition of an expression is then carried out in explicitly estimating $p(X|e_\kappa)$ as in Eq. (3) or equivalently $logP(X|e_\kappa)$ according to Eq. (4), and to choose the facial expression maximizing this a posteriori probability as in Eq. (5). Therefore, while the proposed BBN gathers as much as possible evidence from the different features for the inference on FER, the different matching scores, as delivered by the 15 types of features through their SFMs and used as input to SVM and LDA, may not be as discriminating as required a discriminative classifier such as SVM or LDA. N-sum proceeds in much a similar way as BBN except the fact that the scores delivered by various types of features were first normalized in the 0.0.1 range before their addition to produce the final score. In doing so, efficient features with high matching scores may have their effect decreased in the final score, thereby leading to a slight performance decrease in comparison with BBN.

We also studied two different voting rules, namely plurality voting and Borda count, on the score sets from BBN, SVM, SRC as used in Table 3, compared with score level fusion schemes using Product, Max, Min and Sum rule. In both plurality voting and Borda count voting, each type of features represents a voter (15 in total), each class of facial expressions represents a candidate (6 in total when using BU-3DFE). In the former voting, each type of features used its highest score to vote for an expression class. The expression which has been voted the most is then recognized. In the latter voting, each type of features gave 6 points to its highest score, 5 points to its second highest one, …, and 1 point to its lowest score. Then the points from 15 voters were summed up and the expression having most points was then recognized. In the test using Max rule, the max score from all 15 features for each expression is compared with others and the highest one is recognized, while in the one using Min rule, the min score from all 15 features for each expression is compared with others and the highest one is recognized. When using product fusion rule, we multiplied scores from 15 types of features together for each expression and selected the highest one as the recognized facial expression class. Table 6 depicts the comparison on different score level fusion methods for SVM, SRC and BBN respectively. Note that the Sum Rule was the one used in Table 3.

It can be seen from Table 6 that the sum rule performs the best for all the three classifiers while the proposed BBN achieves the highest recognition rate. The low recognition rates by the plurality voting rule can be explained by the information loss in the decision process as the recognition decision is only made based on the number of feature voters, thereby discarding the matching scores or likelihoods achieved by each type of features on the six facial expressions. Borda Count Voting alleviates this information loss in affecting a rank number to each matching score by a given type of features on the six facial expressions. In doing so, it enables an improvement of the recognition rates of plurality voting by a number of points up to 25. The other fusion rules, e.g., Max, Min, Product and Sum, take into account, each to some extent, all the matching scores as delivered by the 15 types of features for each facial expression. The Max rule is a kind of "Or" operator as it says that a facial expression is recognized as long as a type of features among the 15 has recognized it in delivering the highest matching score, i.e. likelihood within the Bayesian framework. The Min rule is a kind of "AND" operator as intuitively it says that a facial expression can be recognized as long as all the 15 types of features have recognized it. Finally, the Product rule behaves in a similar way as the Sum rule as in taking logarithm on the Product rule, the latter becomes a Sum rule. However, the Product rule, in taking a direct product of the matching scores, is likely to be more sensitive to outliers in comparison with the Sum rule. All these explain in some way that the Max and Min rules generated comparable performance figures on one side and the Product and Sum rules similar results on the other side, the best recognition rates being achieved by the Sum rule. These results are consistent with the study by Kitller et al. [49] which gives a theoretical support of the superiority of the sum rule in comparison of several other popular rules, i.e. product, min, max, median rule along with majority voting, from a sensitivity analysis.

Table 5
Average recognition rate (RR) for different score weighting strategies.

	LDA	SVM	N-sum	BBN
RR	82.5%	85.1%	86.5%	88.9%

Table 6
Score fusion comparison of SRC, SVM and BBN.

	Plurality voting	Borda count voting	Max rule	Min rule	Product rule	Sum rule
SRC	32.4%	57.8%	75.6%	78.5%	84.8%	85.3%
SVM	37.4%	60.4%	74.9%	79.3%	87.9%	88.1%
BBN	36.7%	60.8%	76.1%	77.6%	88.6%	89.2%

6.4. Results of 3D AU recognition using the Bosphorus dataset

In order to highlight the flexibility and capacity of the proposed BBN in facial expression analysis, we carried out additional experiments on the Bosphorus dataset for recognizing AUs. Due to the data availability, we used BBN to recognize two sets of AUs. 7 facial AUs from 100 subjects were first evaluated and 16 facial AUs from 60 subjects were then analyzed.

The 7 AUs used for the experiment include AU2, AU4, AU9, AU12, AU27, AU28 and AU34. They account for ample facial displays, e.g., AU27, as well as subtle facial expressions, e.g., AU2, AU4. AU34 displays significant facial surface deformation almost without texture variations. All the tests followed a 10-fold subject-independent cross-validation process as described in Section 1. Table 7 shows the overall average recognition rates using different feature sets.

As it can be seen from Table 7, the first three columns show the recognition rates by features from each single modality, namely Morphology (M), local texture (T) and local shape (S). Once more, as we have already discovered from the results in Table 3, we see again that the local shape modality, in capturing facial surface properties, performed the best, which is followed by local texture modality and finally the morphology modality. The following three columns are the results combining features from any two modalities, and the last column shows the result using all the 15 features from the three modalities. Again in line with the findings from the results in Table 3, we can observe that the best combination, in terms of average recognition rate when fusion two modalities, is local shape (S) and texture (T), which is followed by local shape (S) and morphology (M) and finally local texture (T) and morphology (M). The proposed BBN has efficiently fused multiple evidences with its Bayesian inference and displays a recognition rate up to 94% when all types of features were used as evidence. Table 8 shows the confusion matrix of the proposed BBN on recognizing the 7 AUs. It is worth noting that AU27 (Mouth Stretch) featuring opening of the mouth recorded a 100% recognition rate while subtle facial displays such as AU2 (Outer Brow raiser), AU4 (Brow Lowerer) and AU34 (Puff) were recognized in roughly 90% cases with a peak for AU4 recognized in roughly 98% cases.

The second experiment made use of 60 subjects out of 62 having 3D face scans displaying all the 16 AUs in order to keep the data balanced in the experiment. These subjects were only selected based on the data availability of the aforementioned 16 AUs. In this experiment, we defined the states of the node X in BBN as the 16 AUs to be recognized and conducted the test following the 10-fold subject-independent cross-validation process. The results are given in Table 9 in terms of average positive rates and average false-alarm rates for all AUs. Indeed, recognizing each AU_i can be considered as a two-class classification according to the AU_i and the non-AU_i. The positive rate is defined as $PR = \frac{TP}{TP+FN}$ and the false-alarm rate is $FAR = \frac{FP}{TP+FP}$ where TP stands for "True Positive", FN for "False negative" and FP for "False Positive". Among the 16 AUs, seven of them (AU10, AU18, AU22, AU26, AU27, AU2, AU43) have an average PR over 90%, while 4 of them (AU14, AU24, AU7, AU4) have average PR below 80%. Meanwhile, AU24 has the highest FAR, which suggests that it is easily confused with other AUs. This is also the case for AU34 and AU4 having each a FAR above 20%. On the other hand, AU43, AU27, AU22 with a FAR below 5% are relatively clearly identified. Globally, our BBN achieves an overall average PR for all 16 AUs of 85.6% with an overall average FAR of 13.6%.

Table 7
Average recognition rates for 7 action units using different feature sets.

	M	T	S	M+T	M+S	T+S	M+T+S
RR	74.60%	87.01%	91.92%	88.74%	90.19%	93.36%	94.23%

Table 8
Confusion Matrix of the subject-independent AU recognition.

Input\output	AU12	AU27	AU28	AU34	AU9	AU2	AU4
AU12	94.95%	0.00%	0.00%	0.00%	0.00%	1.01%	4.04%
AU27	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AU28	0.00%	0.00%	95.96%	0.00%	0.00%	1.01%	3.03%
AU34	0.00%	0.00%	1.01%	89.90%	0.00%	0.00%	9.09%
AU9	1.01%	0.00%	0.00%	0.00%	89.90%	0.00%	9.09%
AU2	3.03%	0.00%	1.01%	1.01%	0.00%	90.91%	4.04%
AU4	0.00%	0.00%	1.01%	0.00%	1.01%	0.00%	97.98%

6.5. Discussion

Table 10 presents a comparison of our work with the literature in 3D FER. With the exception of our work, all the works listed in the table are 3D geometric feature based approaches using distances among landmarks or deformable models to capture 3D facial surface deformations. With a 94% recognition rate [22], displayed the best performance. However, they made use of a set of distance features computed over the 83 manually labeled feature points provided by the BU-3DFE dataset and require a neutral face from each subject for distance normalization. Furthermore, as our experiments have evidenced previously, there will be significant performance drop when switching from the manual landmarks to less accurate automatic landmarks.

Another remarkable performance is 90.5% recognition rate achieved by [23] without using manually labeled landmarks in the testing phase. Meanwhile, their approach requires costly dense registration and hardly converges without specific treatment for ample facial displays such as wide opening of the mouth. Furthermore, most misclassification in their work occurs in distinguishing anger and sadness, which have very subtle differences in the configuration of the eyebrows. This suggests that in order to distinguish subtle expressions or AUs, one really needs to resort to other features such as appearance ones.

As compared to all these works, the proposed BBN achieves a good balance between recognition accuracy and computational simplicity in following a hybrid approach. It takes full advantage of textured 3D face models in FER and makes use of Bayesian inference to efficiently fuse multiple evidences from both geometric and appearance features. Compared with [23], the building and fitting of the SFAM for morphable partial face models can be easily implemented and does not require dense registration nor specific treatment for ample facial deformations such as wide opening of the mouth. Meanwhile, the proposed BBN proves to be effective in recognizing both the six universal facial expressions and AUs which account for ample facial displays as well as subtle ones.

In [50], 22 AUs are detected automatically by estimating the deformation between the registered face and the reference. Based on the same dataset, they achieve an average PR of 91.1%. In [51], 7 AUs are considered and a AU combination on their own database is performed allowing to achieve a PR of 89.1%. In [31], authors use a Dynamic Bayesian Net to learn the relationship between AUs on 2D Cohn–Kanade database in order to enhance the recognition performance using Gabor features and Ababoost classifier. They achieve an 85.8% PR on 14 AUs. Our approach achieves an average PR of 85.6% for 16 AUs, which achieves a consistent result with the highly optimized 2D method [31].

7. Conclusion

We have proposed in this paper a unified framework based on a Bayesian Belief Network (BBN) to recognize both facial expressions and AUs. The proposed BBN performs Bayesian inference for 3D FER based on SFM and fuses multiple evidence from both geometric and appearance features. Fully taking advantage of textured 3D face models, the geometric and appearance features are extracted from three modalities which account for deformations of intransient and

Table 9
Average positive rates (PR) and Average false-alarm rates (FAR) of AUs.

	AU2	AU4	AU7	AU9	AU10	AU12	AU14	AU17	AU18	AU22	AU24	AU26	AU27	AU28	AU34	AU43
PR	90.0%	75.0%	78.3%	81.7%	95.0%	85.0%	75.0%	80.0%	91.7%	90.0%	76.7%	91.7%	91.7%	81.7%	88.3%	98.3%
FAR	3.6%	26.2%	13.0%	5.8%	10.9%	19.0%	23.7%	7.7%	14.1%	3.6%	40.3%	12.7%	3.5%	7.5%	20.9%	4.8%

Table 10
Result comparisons between our method and state-of-the-art methods for FER.

Method	Methodology	Subjects	Expressions	Manual landmarks	Results
[17]	Neural network	60	7	Yes (23)	87.9%
[22]	SVM	60	6	Yes (83)	94.7%
[20]	LDA	60	6	Yes (64)	83.6%
[18]	AdaBoost	60	6	Yes (83)	87.1%
Our approach	BBN	60	6	Yes(19)	89.2%
[23]	Bilinear model	100	6	No	90.5%
[19]	Modified PCA	60	6	No	81.7%
Our approach	BBN & SFAM	60	6	No	84.9%

transient facial components. When combined with our previously developed morphable partial face model (SFAM), the proposed BBN achieves fully automatic FER.

Experimented on the two public databases, namely the BU-3DFE dataset for recognition of the six universal facial expressions using both automatically and manually located landmarks and the Bosphorus dataset for recognition of 16 AUs, the proposed BBN proves to be a powerful engine by its Bayesian inference to fuse multiple features from the three modalities and shows its effectiveness for FER. The proposed BBN achieved average recognition rates of 94.2% and 85.6% for 7 AUs and 16 AUs respectively and 89.2% and 84.9% for the six universal expressions using manually and automatic labeled landmarks respectively.

In the future, we want to adapt the proposed BBN for the recognition of multiple AUs as in [34] and envisage also performing joint inference based on the BBN for both 3D face and expression recognition. Furthermore, we also want to better characterize the dynamics of facial displays as we did partly in this paper when measuring displacement of the landmarks.

Acknowledgment

This work was supported in part by the French National Research Agency, Agence Nationale de Recherche (ANR),through the ANR FAR3D project under the grant ANR-07-SESU-004-03, and the ANR 3D Face Analyzer project under the grant ANR 2010 INTB 0301 01.

References

[1] Z. Zeng, M. Pantic, G.I. Roisman, T.S. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Trans. Pattern. Anal. Mach. Intell. 31 (2009) 39–58.
[2] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, Palo Alto, 1978.
[3] P. Ekman, E. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System, second ed. edition Oxford Univ. Press, 2005.
[4] M. Pantic, L. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Trans. Pattern. Anal. Mach. Intell. 22 (2000) 1424–1445.
[5] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. Taylor, Emotion recognition in human-computer interaction, IEEE Signal Process. Mag. 18 (2001) 32–80.
[6] Y. Tian, T. Kanade, J. Cohn, Facial expression analysis, Handbook of Face Recognition, 2003.
[7] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recog. 36 (2003) 259–275.
[8] M. Pantic, A. Pentland, A. Nijholt, T. Huang, Human computing and machine understanding of human behavior: a survey, in: F. Quek, Y. Yang (Eds.), ACM SIGCHI Proceedings Eighth International Conference on Multimodal Interfaces, ACM, New York, 2006, pp. 239–248.
[9] T. Fang, X. Zhao, O. Ocegueda, S.K. Shah, I.A. Kakadiaris, 3D facial expression recognition: a perspective on promises and challenges, in: 7th International Conference on Automatic Face and Gesture Recognition, 2011, pp. 603–610.
[10] Y. Chang, C. Hu, R. Feris, M. Turk, Manifold based analysis of facial expression, Image Vision Comput. 24 (2006) 605–614.
[11] M. Pantic, I. Patras, Dynamics of facial expression: recognition of facial actions and their temporal segments form face profile image sequences, IEEE Trans. Syst. Man Cybern. B 36 (2006) 433–449.
[12] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan, Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior, 2005.
[13] S. Koelstra, M. Pantic, I. Patras, A dynamic texture based approach to recognition of facial actions and their temporal models, IEEE Trans. Pattern. Anal. Mach. Intell. 99 (2010).
[14] K. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition, J. Comput. Vis. Image Underst. 101 (2006) 1–15.
[15] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3D facial expression database for facial behavior research, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2006, pp. 211–216.
[16] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, The First COST 2101 Workshop on Biometrics and Identity Management, 2008.
[17] H. Soyel, H. Demirel, 3D facial expression recognition with geometrically localized facial features, in: Symposium on Computer Science and Information Technology, 2008, pp. 1–4.
[18] H. Tang, T. Huang, 3D facial expression recognition based on properties of line segments connecting facial feature points, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2008, pp. 1–6.
[19] Y.V. Venkatesh, A.A. Kassim, O.V.R. Murthy, A novel approach to classification of facial expressions from 3D-mesh datasets using modified PCA, Pattern Recogn. Lett. 30 (2009) 1128–1137.
[20] J. Wang, L. Yin, X. Wei, Y. Sun, 3D facial expression recognition based on primitive surface feature distribution, in: International Conference on Computer Vision and Pattern Recognition, 2006, pp. 1399–1406.
[21] S. Ramanathan, A. Kassim, Y. Venkatesh, W.S. Wah, Human facial expression recognition using a 3D morphable model, in: IEEE International Conference on Image Processing, 2006, pp. 661–664.
[22] H. Tang, T.S. Huang, 3D facial expression recognition based on automatically selected features, workshop, in: International Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[23] I. Mpiperis, S. Malassiotis, M. Strintzis, Bilinear models for 3-d face and facial expression recognition, IEEE Trans. Inf. Forensics Secur. 3 (2008) 498–511.
[24] M. Rosato, X. Chen, L. Yin, Automatic registration of vertex correspondences for 3D facial expression analysis, in: International Conference on Biometrics: Theory, Applications and Systems, 2008, pp. 1–7.
[25] X. Zhao, E. Dellandréa, L. Chen, A 3D statistical facial feature model and its application on locating facial landmarks, in: ACIVS 2009: Advanced Concepts for Intelligent Vision Systems, 2009, pp. 686–697.
[26] X. Zhao, E. Dellandra, L. Chen, I.A. Kakadiaris, Accurate landmarking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model, IEEE Trans. Syst. Man Cybern. B (2011) 1–12.
[27] X. Zhao, D. Huang, E. Dellandra, L. Chen, Automatic 3D facial expression recognition based on a Bayesian Belief Net and a statistical facial feature model, in: International Conference on Pattern Recognition (ICPR), 2010.
[28] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd Edition Wiley-Interscience, 2000.
[29] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern. Anal. Mach. Intell. 23 (2001) 681–685.
[30] Y. Tong, W. Liao, Q. Ji, Facial action unit recognition by exploiting their dynamic and semantic relationships, IEEE Trans. Pattern. Anal. Mach. Intell. 29 (2007) 1683–1699.
[31] Y. Tong, J. Chen, Q. Ji, A unified probabilistic framework for spontaneous facial action modeling and understanding, IEEE Trans. Pattern. Anal. Mach. Intell. 32 (2010) 258–273.
[32] D. Datcu, L. Rothkrantz, Automatic recognition of facial expressions using Bayesian Belief Networks, in: International Conference on Systems, Man and Cybernetics, 2004.
[33] A. Savran, B. Sankur, M.T. Bilge, Regression-based intensity estimation of facial action units, Image Vision Comput. 30 (10) (October 2012) 774–784.
[34] A. Savran, B. Sankur, M.T. Bilge, Comparative evaluation of 3D versus 2D modality for automatic detection of facial action units, Pattern Recog. 45 (2012) 767–782.
[35] M. Hall, Correlation-based feature selection for machine learning, Technique Report, 1999.
[36] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its applications on facial image, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 41 (2011) 765–781.

[37] C. Shan, T. Gritti, Learning discriminative lbp-histogram bins for facial expression recognition, British Machine Vision Conference, 2008.
[38] C. Chan, J. Kittler, K. Messer, Multi-scale local binary pattern histograms for face recognition, in: Proceedings of International Conference on Advances in Biometrics, 2007, pp. 809–818.
[39] C. Dorai, A. Jain, Cosmos — a representation scheme for 3D free-form objects, IEEE Trans. Pattern. Anal. Mach. Intell. 19 (1997) 1115–1130.
[40] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, L. Chen, 3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching, in: IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, 2010, pp. 1–4.
[41] P. Szeptycki, M. Ardabilian, L. Chen, W. Zeng, D. Gu, D. Samaras, Partial face biometry using shape decomposition on 2D conformal maps of faces, Int. Conf. Pattern Recog. (ICPR) (2010) 1–4.
[42] J. Nelder, R. Mead, A simplex method for function minimization, Comput. J. 7 (1965) 308–313.
[43] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces, in: SIGGRAPH '99: Proceedings of the 26th annual conference on Computer Graphics and Interactive TechniquesACM Press/Addison-Wesley Publishing Co, New York, NY, USA, 1999, pp. 187–194.
[44] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, in: The First COST 2101 Workshop on Biometrics and Identity Management, 2008.
[45] P. Nair, A. Cavallaro, 3-D face detection, landmark localization, and registration using a point distribution model, IEEE Trans. Multimedia 11 (2009) 611–623.
[46] C. Chang, C. Lin, LIBSVM: a library for support vector machines. (Software available at) http://www.csie.ntu.edu.tw/cjlin/libsvm 2001.
[47] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern. Anal. Mach. Intell. 31 (2009) 210–227.
[48] B.G. kberk, H. Dutagaci, A. Ulas, L. Akarun, Representation plurality and fusion for 3-D face recognition, IEEE Trans. Syst. Man Cybern. B 38 (2008) 155–173.
[49] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Learn. 20 (1998) 226–239.
[50] A. Savran, B. Sankur, Automatic detection of facial actions from 3D data, ICCV09: Workshop on Human Computer Interaction, 2009.
[51] Y. Sun, M. Reale, L. Yin, Recognizing partial facial action units based on 3D dynamic range data for facial expression recognition, in: Automatic Face Gesture Recognition, FG '08. 8th IEEE International Conference on, 2008, pp. 1–8.