Assessing the Uniqueness and Permanence of Facial Actions for Use in Biometric Applications

Lanthao Benedikt, Darren Cosker, Paul L. Rosin, and David Marshall

***Abstract*—Although the human face is commonly used as a physiological biometric, very little work has been done to exploit the idiosyncrasies of facial motions for person identification. In this paper, we investigate the *uniqueness* and *permanence* of facial actions to determine whether these can be used as a behavioral biometric. Experiments are carried out using 3-D video data of participants performing a set of very short verbal and nonverbal facial actions. The data have been collected over long time intervals to assess the variability of the subjects' emotional and physical conditions. Quantitative evaluations are performed for both the identification and the verification problems; the results indicate that emotional expressions (e.g., smile and disgust) are not sufficiently reliable for identity recognition in real-life situations, whereas speech-related facial movements show promising potential.**

***Index Terms*—Active appearance model, biometrics, dynamic time warping, facial motion, pattern recognition, 3-D facial recognition.**

I. Introduction

THE FACE is the primary means for humans to recognize each other in everyday life and therefore represents the most natural choice of biometric. Early works such as Eigenfaces [1] and methods based on support vector machine [2] have inspired a number of 2-D face recognition solutions which are now employed in various commercial and forensic applications (e.g., Neven Vision, Viisage, Cognitec) [32]. Despite such wide acceptance, the face is commonly considered a weak biometric because it can display many different appearances due to rigid motions (e.g., head poses), nonrigid deformations (e.g., face expressions), and also because of its sensitivity to illumination, aging effect, and artifices such as makeup and facial hair [3].

With recent advances in 3-D imaging technologies, the head pose and illumination problems have been successfully overcome thanks to the development of many novel 3-D algorithms [4]. However, the facial expression variability still remains an open issue. While Chang *et al.* [5] suggest using expression-invariant facial regions for identity recognition (e.g., the region around the nose), another solution is to employ model-based algorithms such as the 2-D Active Appearance Model [6] and the 3-D Morphable Model [7] which can accurately interpret unseen facial poses thanks to a prior learning process [20]. More recently, a novel research trend opens new horizons by considering facial expressions as an additional source of information rather than a problem to overcome. Pioneer works have reported very promising results [8]–[10]; however, there was little emphasis on examining in depth the characteristics of facial dynamics and their suitability for biometric applications: Are facial actions stable over long time intervals? How much do they vary with the subject's emotional and physical conditions? Are they sufficiently discriminative across a large population?

This paper proposes to investigate these questions and is organized as follows. Section II reviews the existing public databases and explains our choices for a new set of very short verbal and nonverbal facial actions. Section III describes in detail the data recording conditions and apparatus. Our experiments employ exclusively naive users and seek to model as closely as possible a real-life scenario. Section IV outlines the architecture of an identity recognition system using facial actions; the data processing tasks and an accurate feature extraction method are described, and then an efficient pattern recognition algorithm derived from dynamic time warping (DTW) [23] is proposed. Section V assesses the use of facial actions in biometrics considering both the face identification and the face verification problems, in light of which improvements are proposed.

II. Choices of Facial Actions

In principle, any facial actions can be considered for person recognition. However, choosing those which exhibit high biometric power is similar to choosing strong computer login passwords over weak ones. Furthermore, *very short* facial actions help reduce the processing effort so that genuine users can gain access quickly to the secure services.

The majority of related works have been carried out by researchers working on speech analysis [8]–[10], who usually employ popular face databases for lipreading and speech synthesis such as the M2VTS [11] and the XM2VTSDB [12] databases which include audio-video data of continuously uttered digits from "0" to "9," and long spoken phrases, e.g., "Joe took father's green shoe bench out." In addition, commonly used is the DAVID database [13] where nine participants wearing blue lipstick utter isolated digits. These databases are not

Manuscript received November 30, 2008. First published March 25, 2010; current version published April 14, 2010. This paper was recommended by Guest Editor K. W. Bowyer.

L. Benedikt, P. L. Rosin, and D. Marshall are with the School of Computer Science, Cardiff University, CF24 3AA Cardiff, U.K. (e-mail: lanthao99@googlemail.com; paul.rosin@cs.cardiff.ac.uk; dave.marshall@cs.cardiff.ac.uk).

D. Cosker is with the School of Computer Science, Cardiff University, CF24 3AA Cardiff, U.K., and also with the School of Computer Science, University of Bath, BA2 7AY Bath, U.K. (e-mail: dpc@cs.bath.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2010.2041656

1083-4427/$26.00 © 2010 IEEE

TABLE I
ENGLISH LANGUAGE PHONEME TO VISEME MAPPING [34]

Viseme	Phoneme
/p/	P
	B
	M
	EM
/f/	F
	V
/t/	T
	D
	S
	Z
	TH
	DH
	DX
/w/	W
	WH
	R
/ch/	CH
	JH
	SH
	ZH
/ey/	EH
	EY
	AE
	AW
/k/	K
	G
	N
	L
	NX
	HH
	Y
	EL
	EN
/iy/	IY
	IH
/aa/	AA
/ah/	AH
	AX
	AY
/er/	ER
/ao/	AO
	OY
	IX
	OW
/uh/	UH
	UW
/sp/	SIL
	SP

suitable for biometrics because the use of physical markers is inconvenient, and while long phrases are necessary for speech analysis, they require intensive processing effort, particularly when 3-D dynamic data are employed. Although one might consider using only short segments of the long phrases, such "cut out" syllables are inevitably plagued by coarticulation that unnecessarily complicates the problem at this early stage. For these reasons, we decided to collect a new set of *very short* and *isolated* facial actions. Two directions have been explored and are described in the following sections.

A. Verbal Facial Actions

With regard to speech-related facial motions, one might be tempted to see a direct link with research on speech analysis where the distinctiveness of lip motions has been studied to a certain extent [30], [35]. However, lipreading and biometrics are two different problems. While lipreading aims to recognize phonemes from lip shapes (visemes) and must be speaker independent, we seek on the contrary to recognize the visemic dissimilarities across speakers uttering the same phoneme. Another field where there has also been a great deal of interest in characterizing facial motions is Orthodontics and Craniofacial research. However, these works often use a very small number of subjects (between 5 and 30) and examine limited speech postures [17], which is inconclusive for biometrics.

Our objective here is to assess: 1) the repeatability of viseme production over time for any speaker and 2) the distinctiveness of lip motions across speakers. To this end, we will examine a set of words chosen among the visemes of the English language as depicted in Table I.

Consonants: The *point of articulation* plays a great role in the strength of a consonant. While bilabial consonants (involving the lips, e.g., /p/, /b/, /m/) and labiodental consonants (upper teeth and lower lip, e.g., /f/, /v/) are informative because their visible variations can be easily captured by the camera, consonants involving internal speech organs such as the palate and the tongue (e.g., /t/, /k/, /r/, etc.) are expected to be poor because their variations are hidden.

Vowels: Vowels typically form the nuclei of syllables, while consonants form the onsets (precede the nuclei) and the coda (follow the nuclei). A codaless syllable of the form V, CV, CCV—where V stands for vowel and C for consonant—is called an open syllable, and a syllable that has a coda, e.g., VC, CVC, CVCC is called a closed syllable. The vowel of an open syllable is typically long, whereas that of a closed syllable is usually short. Two adjacent vowels in a syllable of the form CVVC usually behave as a long vowel. The r-controlled (vowel followed by "r") and the l-e-controlled (consonant-l-e at end of syllable) are neither long nor short, but allow trailing as in a long vowel [33].

In this paper, a comparative evaluation will be performed for different types of syllables, using the following words: "***puppy***" (CVC/CV, short vowel/long vowel), "***baby***" (CV/CV, long vowel/long vowel), "***mushroom***" (CVCC/CVVC, short vowel/long vowel), "***password***" (CVCC/CVC, irregular case: long vowel/r-controlled vowel), "***ice cream***" (VCV/CCVVC, long vowel/long vowel), "***bubble***" (CVC/CV, short vowel/l-e-controlled), "***cardiff***" (CVC/CVC, r-controlled vowel/short vowel), "***bob***" (CVC, short vowel), and ***rope*** (CVC, irregular case: long vowel).

B. Nonverbal Facial Actions

Previous works have also considered emotional expressions for identity recognition; however, these are only limited to the study of the smile dynamics [37], and the expression of pain [36]. Standard databases of nonverbal expressions do exist, but they are aimed for behavioral studies rather than for biometrics, and therefore, do not include repetitions of the same expression over time [38], [39].

In this paper, we propose to collect a selected set of nonverbal facial actions. Unlike visual speech which involves mainly the lip motions, emotional expressions require movements of the forehead, the eyebrows, the eyes, the nose, and more. Thus, we choose to study a number of expressions that involve muscle activations in various facial regions, and in order to improve the repeatability of the performances, we impose further constraints inspired from the Facial Action Coding System (FACS) [16], e.g., surprise AU1 (inner-brow raiser) + AU2 (outer-brow raiser), disgust AU9 (nose wrinkler), joy AU12 (lip corner puller), sadness AU15 (lip corner depressor). Interestingly, research in Orthodontics and Craniofacial has found evidence that facial actions involving maximal muscle stretches are more reproducible compared to moderate expressions [17]. Therefore, we focus on facial expressions of maximum intensity only.

To accurately capture the idiosyncrasies of a person, the entire face needs to be analyzed. However, such a holistic approach may be computationally prohibitive; therefore, it is useful to determine which facial regions are the most important to convey identity, so that we can focus only the most informative part of the face.

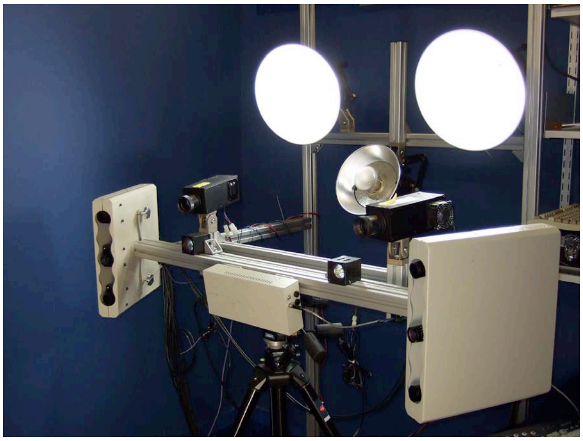

Fig. 1. 3dMDFace Dynamic System.

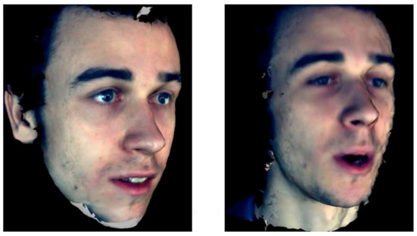

Fig. 2. Stills from two video sequences of the same subject uttering the word "puppy," collected in two recording sessions with different settings to test how well the recognition algorithm performs on data of variable qualities. The 3-D mesh in the first sequence appears more bumpy, and the texture more blurred. In the second sequence, poor head positioning with respect to the camera viewpoint has led to missing texture at the chin.

III. Data Acquisition

To collect data of facial actions for our research, two 3-D video cameras operating at 48 frames per second are employed (see Fig. 1). Although the cameras have been purchased from the same provider (3dMDFace Dynamic System), they are of different generations, hence the 3-D mesh densities and the texture qualities they produce are noticeably different. The recording sessions are carried out in two laboratories with different ambient conditions (day light or dark room), and there is no strict control of the head position; small head movements are allowed as long as we ensure an ear-to-ear coverage of the face. No physical markers are used. Such settings resemble a real-life scenario where faces are also expected to be collected from different data capture devices in different recording conditions. This way, we can evaluate how well our recognition algorithm performs on data of variable qualities which has been collected in moderately constrained environments (see Fig. 2).

The participants are staff and students at the Schools of Computer Science, Engineering and Dentistry, Cardiff University, U.K. All are naive users who have not been trained beforehand; in particular, none are familiar with the FACS Coding System [16]. The participants are asked to perform a number of basic emotions (smile, disgust expression) and a small set of Action Units as described in Section II-B, targeting a maximal muscle stretch; on the other hand, they are also required to utter a set of *isolated words* as described in Section II-A, speaking in a normal and relaxed way.

The recording sessions are scheduled over large time intervals (over a few weeks or months and in some cases, over more than two years) to assess the variability of the subjects' emotional and physical conditions. Due to difficulties to collect 3-D dynamic data on a large scale, our database is unfortunately nonuniform at the moment. We have been able to collect data from 94 participants (61 males and 33 females, 70 are natural English speakers) uttering the word "puppy" over more than two years in some cases. However, only 15 participants were used to study other utterances (e.g., as "password," "mushroom," etc.), and the repetitions have been recorded over a period of one month maximum. The smile dynamic has been studied for about 50 subjects, and smaller tests are conducted on FACS AUs. In reality, we very quickly abandoned the investigation of AUs because during the recording sessions, it became clear that it was very difficult for naive users to produce accurate AUs, let alone to repeat exactly the same performance several times.

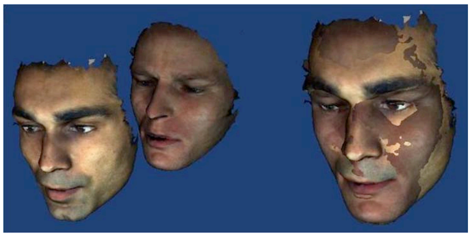

Fig. 3. Alignment of two face scans of different head poses and sizes. The nose root is located, together with the eye directions and surface normals, the alignment is achieved through rotations, translations and scaling, then refined by applying Iterative Closest Point to the nose region.

IV. Feature Extraction

Many techniques have been proposed in previous studies for extracting facial dynamic features. Of the approaches which do not require the use of physical markers, there are, for example, the work of Pamudurthy *et al.* [19] which aims to track the motions of skin pores, and the work of Faraj *et al.* [10] which uses the velocity of lip motions for speaker recognition. These methods are interesting and novel; nevertheless, they still require more research. In this paper, we adopt another feature extraction approach relying on the well-established model-based algorithms used in static face recognition, e.g., the Active Appearance Model [6] and the Morphable Model [7]. The data preprocessing and feature extraction steps are described as follows.

A. Face Normalization

One nontrivial preliminary task in face recognition is to normalize the scans so that they can be compared in the same coordinate frame (i.e., similar head pose and size). To this end, Fidaleo *et al.* [18] use a closed wooden box with a cushioned opening in order to control the head position. Although this device fulfills its purpose, it may raise some hygiene concern when such a system is to be used in public places such as airports. In this paper, the normalization is achieved computationally using a nose-matching technique inspired from a method proposed by Chang *et al.* [5], the result is shown in Fig. 3. The nose root is found using 3-D curvature analysis, then the surface normal at the nose root and the eye direction are computed.

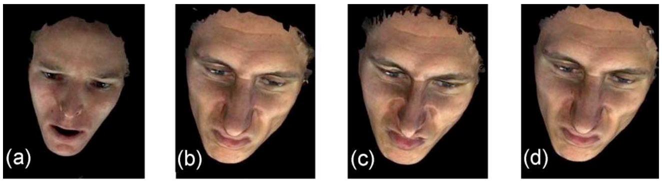

Fig. 4. Frame correspondence. The reference frame (a) is first deformed to the target mesh (b) using thin plate spline warping only. The result is (c) where the cheek folds have not been correctly transferred. To refine this process, we superimpose the meshes (b) and (c) and each vertex of mesh (c) is projected along its surface normal onto the target mesh (b). We then obtain mesh (d) where the muscle folds have been recovered.

These quantities determine the adequate translations, rotations and scaling to align the faces. Further refinements are achieved using the Iterative Closest Point matching algorithm, applied to the most stable regions around the nose root.

In practice, we use one video frame as the reference frame, and register all frames in the database to the latter. The choice of the reference frame is arbitrary, as long as we employ one and only one reference for the entire database.

B. Frame Correspondence and Segmentation

One principal requirement in our feature extraction approach is that all face scans must be in full correspondence (i.e., same number of vertices and identical vertex topology). However, the 3-D data capture system reconstructs each scan independently, resulting in all meshes being uncorrelated. Thus, one additional processing step is needed for establishing vertex correspondence.

The idea consists of choosing one face to be *the* reference frame for the entire face database, and use a thin-plate-spline process [14] to deform this one to each of the frames in the database [15]. This produces a new database where all faces have the same vertex topology as the reference frame, but the facial expression of each frame is identical to that of the corresponding frame in the original database. The reference frame here can be different from that used for 3-D registration described in Section IV-A. Its choice is not arbitrary. A good candidate is a high-quality frame (not too bumpy, accurate inner-mouth details), and presenting a slightly open mouth and eyes because such a facial pose is a good approximation of a "mean" face. The warping alone does not produce highly accurate result, thus the target mesh [Fig. 4(b)] and the warped mesh [Fig. 4(c)] are superimposed, then each vertex of the warped mesh is projected along its surface normal onto the target mesh, as shown in Fig. 4.

The purpose of the warping process is twofold. First, it brings all the frames into correspondence; and second, it also permits the segmentation of the region of interest (ROI). For example, by using a lip shape as a reference template and deforming this one to the lip region of the target frames, we can produce a sequence of lip shapes in full correspondence, as shown in Fig. 5.

C. Model-Based Feature Extraction

After the preprocessing step, the face database is now divided into training set (gallery) and test set (probe). The gallery includes only one video sequence per subject due to limited data availability; the probe contains several repetitions for each subject. The gallery and probe are entirely distinct data segments; the probe includes subjects enrolled in the gallery, as well as "unseen" persons. In this paper, we typically build one model for each facial expression, which includes all subjects enrolled in the gallery.

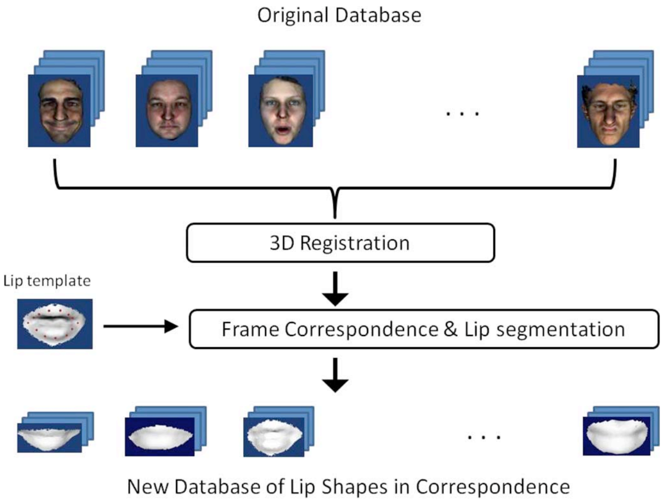
Original Database
. . .
3D Registration
Lip template
Frame Correspondence & Lip segmentation
. . .
New Database of Lip Shapes in Correspondence

Fig. 5. Warping process also permits the segmentation of the ROI when using different facial templates. Here, the lip segmentation is realized by using a lip template.

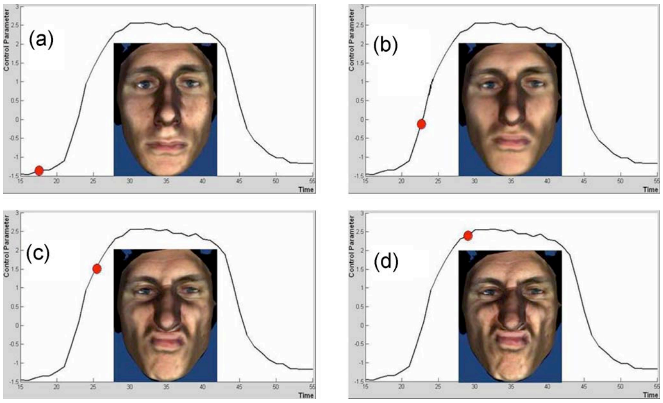

Fig. 6. Subject performing a FACS Action Unit 9 (nose wrinkler) and corresponding temporal evolution of the first Eigencoefficient.

Our statistical model of combined 3-D shape and texture is built in similar fashion to the 2-D Active Appearance Model [6], but the shape model includes the xyz coordinates of all vertices as in the Morphable Model [7]. Fig. 6 shows the temporal variations of the first Eigencoefficient and the corresponding face expressions of a subject performing the FACS Action Unit 9 (nose wrinkler).

The facial dynamics can be extracted from the variations of the 3-D shape alone, independently of the texture information. Therefore, for the remainder of this paper, we build a simple 3-D Active Shape Model (3-D ASM) instead of the complete Active Appearance Model. For a more efficient face recognition (i.e., dimensionality reduction of the feature space and noise removal), we retain only 90% of the shape variations, which

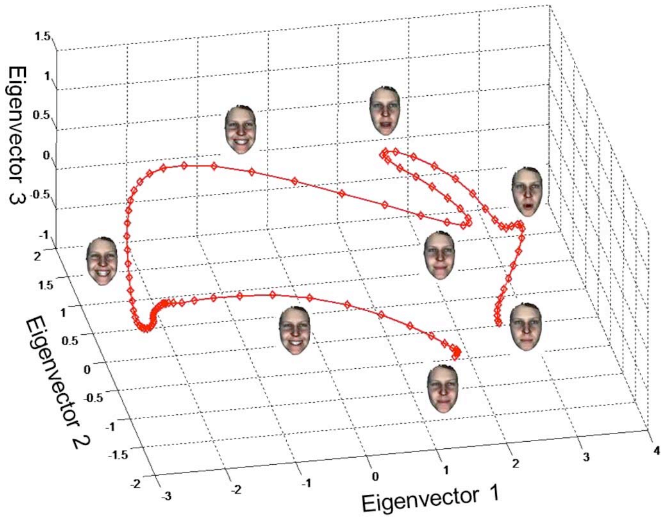

Fig. 7. "Puppy" and smile dynamics of a subject plotted in the subspace spanned by the first three Eigenvectors.

involves keeping the p lead Eigenvectors. Let $\mathbf{s}_k$ be the kth shape in the gallery, it can be approximated as

$$\mathbf{s}_k \approx |\mathbf{s}| + \Phi \mathbf{v}_k \tag{1}$$

where $|\mathbf{s}|$ is the mean shape, Φ is the matrix of the p lead Eigenvectors, and $\mathbf{v}_k = \{v_k^1, \ldots, v_k^p\}^{\mathrm{T}}$ is the $p \times 1$ column vector of Eigencoefficients corresponding to the shape $\mathbf{s}_k$. Inverting this equation, we can extract the Eigencoefficient variations for a sequence of N shapes as follows, where k is the frame number:

$$\mathbf{v}_k \approx \Phi^{\mathrm{T}} \left(\mathbf{s}_k - |\mathbf{s}|\right), \quad k \in \{1, \ldots, N\}. \tag{2}$$

Fig. 7 shows the lip dynamics of a subject saying "puppy" and smiles. The temporal variations of the first three Eigencoefficients $(v_k^1, v_k^2, v_k^3, k \in \{1, \ldots, N\})$ are depicted.

D. Template Matching Algorithms

Well-established algorithms such as hidden Markov models (HMM) and Gaussian mixture models are the preferred pattern matching techniques employed in the previous works [8], [10]. However, it is unclear whether these statistical approaches would perform efficiently in the present context where we use only very short data sequences, and do not have many repetitions per subject to train an accurate HMM.

For this reason, in a related work [26], we survey a number of pattern matching techniques commonly used in behavioral biometrics (e.g., voice, signature authentication), and evaluate which of these performs the best when applied to facial dynamic matching. Well-known methods such as the Fréchet distance [21], correlation coefficients [22], DTW [23], continuous DTW (CDTW) [24], derivative DTW (DDTW) [25], and HMM [8] have been examined, in light of which we proposed an improved algorithm weighted derivative DTW (WDTW) [26] as follows.

Suppose we have two feature vectors which are, respectively, depicted by two sequences of discrete data points C_1 and C_2, as shown in Fig. 8. For each point P_i of C_1, we compute its closest point Q_j on C_2 with respect to the distance $d^i = w_0 \times d_0^i + w_1 \times d_1^i + w_2 \times d_2^i$, where d_0^i is the Euclidean distance between P_i and Q_j, d_1^i is the difference between the local first derivatives, and d_2^i the difference between the local second derivatives; w_0, w_1, and w_2 are weights. The distance D_{WDTW} between C_1 and C_2 is defined as the sum of all pairwise distances d^i, normalized by the sequence length N. Thus, the similarity between two feature vectors can be computed as $S = 1/D_{WDTW}$.

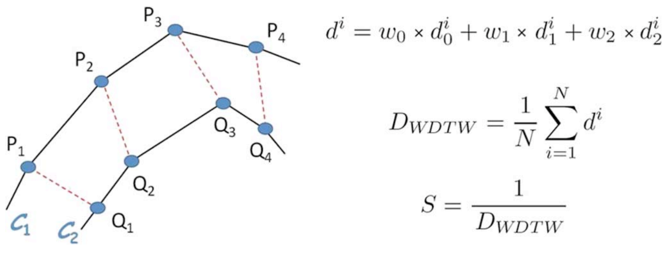
P1
P2
P3
P4
Q1
Q2
Q3
Q4
C1
C2
$d^i = w_0 \times d_0^i + w_1 \times d_1^i + w_2 \times d_2^i$
$D_{WDTW} = \frac{1}{N}\sum_{i=1}^{N} d^i$
$S = \frac{1}{D_{WDTW}}$

Fig. 8. Weighted derivative DTW algorithm (WDTW).

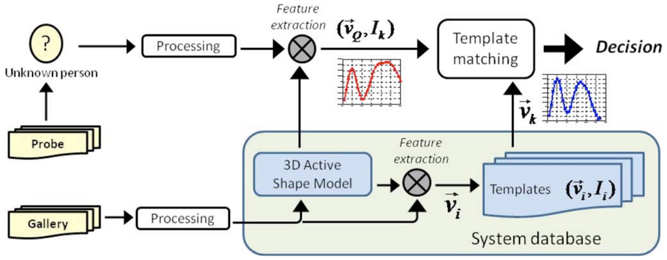

Fig. 9. Architecture of the face recognition prototype. Given an biometric feature vector v_Q of an unknown person and a claimed identity I_k, determine if the person is a genuine user or an impostor. Typically, v_Q is matched against v_k, the biometric template of I_k.

The classic DTW algorithm [23] determines the warping path based on distances between data points, which can yield pathologic results in some particular cases. For instance, let us consider two data points P_4 and Q_3 in Fig. 8 that are close, but one belongs to a rising slope and the other to a falling slope. DTW considers a mapping between P_4 and Q_3 optimal, although it would be inaccurate to map a rising trend to a falling trend. By taking into account the derivatives, WDTW correctly maps P_3 to Q_3, and P_4 to Q_4.

V. Experiments and Results

This paper aims to investigate the use of facial actions for person recognition rather than to compare algorithms. Therefore, for the remainder of this paper, we will employ only the WDTW which performs the best in the present context [26]. A summary of the comparative algorithm evaluation is presented in the next section.

A. Choice of Pattern Matching Technique

The comparative evaluation of the pattern matching algorithms described in Section IV-D is carried out on the face verification problem, the architecture of the recognition system designed for the experiment is shown in Fig. 9. The gallery and probe are distinct data segments of 94 participants uttering

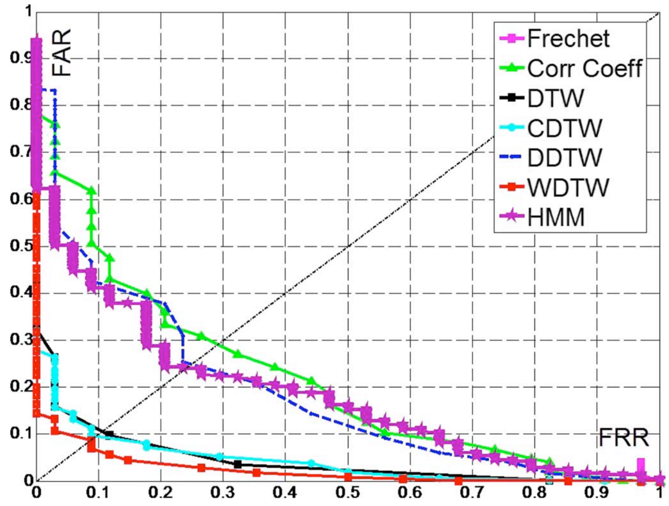

Fig. 10. Performance comparison of several pattern matching algorithms. The recognition process is based on matching the dynamics of the first Eigencoefficient. The facial action used in this experiment is the lip motions of the "puppy" utterance.

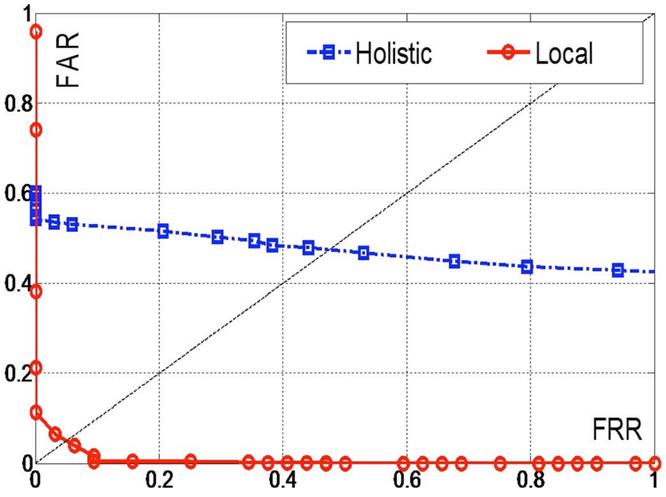

Fig. 11. Holistic versus local feature extraction. We can observe that using the lip motions alone yields better recognition rate than using the entire face. This is principally due to the effect of unwanted facial actions such as blinking which degrade the recognition performance.

the word "puppy," recorded over variable time intervals. In this experiment, only the lip motions are employed in the recognition process.

The results of the comparative study is shown in Fig. 10. The definitions of False-Accept-Rate (FAR) and False-Reject-Rate (FRR) conform to [27]. We observe that, as expected, HMM does not yield a high performance in this situation where we use very short data sequences. WDTW performs slightly better than DTW [23] and CDTW [24].

B. Level of Details

Before we study the use of facial dynamics for person recognition, one clarification seems necessary: Is facial dynamic a physiological or a behavioral trait? This question has been examined by Mason *et al.* [31] who placed the emphasis on the temporal variation: a feature is considered a behavioral biometric if it is a function of time. However, in a case such as facial dynamics, the physiological characteristic is unavoidably embedded, and its contribution is all the more important in the context of a holistic feature extraction. Does it imply that a holistic face recognition is more accurate than using lip motions alone?

Holistic Versus Local Feature Extraction: Intuitively, we might be tempted to think that a holistic approach is more precise because it provides further information about the physical aspects of the face. However, in Section III, we observed that the recorded facial actions were often plagued by unwanted actions such as blinking, which can degrade the recognition performance. For example, the system may issue a false reject if we attempt to match two "puppy" dynamics from the same subject with and without blinking. One possible solution to this problem is to adopt a local feature extraction approach, leaving out the eye region.

The performances of the holistic approach versus using the lip motions alone are shown in Fig. 11. The experiment is carried out on a data set of 28 subjects. The result shows that unwanted actions, e.g., blinking have a negative impact on the recognition performance, with an equal error rate (EER) EER $\approx 50\%$, whereas using lip motions alone yields a better performance, with an EER $\approx 8\%$. Moreover, another advantage of the local feature extraction is that it minimizes the amount of data to process and improves the processing speed.

TABLE II
NUMBER OF EIGENCOEFFICIENTS REQUIRED TO RETAIN A GIVEN PERCENTAGE OF THE VARIATIONS OF THE LIP DYNAMICS. EXAMPLE OF A "PUPPY" UTTERANCE TESTED ON 94 SUBJECTS

% of Variations	50%	80%	90%	95%	98%	100%
Eigencoefficients	1	6	14	183	580	2833

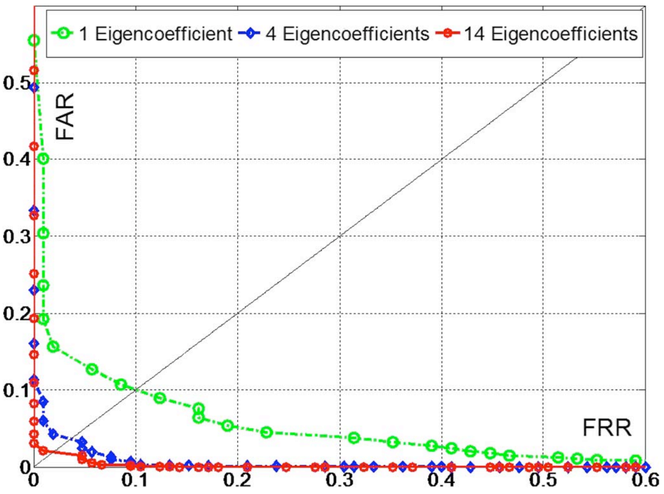

Fig. 12. ROC curves. Performance of the verification system for different number of Eigencoefficients used in the recognition process.

Using Higher Order Eigencoefficients: Facial motions result from complex muscle activations and exhibit many degrees of freedom, e.g., lip opening, lip protrusion, asymmetries, etc. An example is shown in Table II: A data set of 94 participants saying "puppy" is used to build a 3-D ASM, and we compute the number of Eigencoefficients which are necessary to retain a certain percentage of the variations. By manually varying

TABLE III

SIMILARITY BETWEEN DATA SEQUENCES$S = 1/D_{WDTW}$. COLUMNS: REFERENCE TEMPLATES. LINES: UNKNOWN USERS

	Abi	Avril	Bahvna	Chris	Emily	George	Hash	Jamie	James	Kim	Luke	Matt	Vedran	Vitaly
Abi	**15.78**	4.15	5.19	4.78	3.21	1.95	1.43	1.82	2.58	1.36	1.95	2.21	3.17	3.26
Avril	5.18	**14.05**	3.89	7.06	3.96	2.29	1.67	1.74	2.66	2.04	2.38	2.34	3.34	2.96
Bahvna	4.61	2.64	**12.87**	3.43	3.28	1.44	1.97	1.40	2.43	1.09	1.39	1.44	5.60	2.04
Chris	5.22	5.22	3.99	**13.86**	4.13	2.37	1.97	1.78	2.59	2.15	2.31	2.92	6.74	3.90
Emily	2.72	2.43	2.14	2.73	**8.87**	1.22	0.82	0.90	1.36	1.13	1.15	0.95	2.86	1.45
George	1.22	1.34	1.15	1.54	1.08	**5.76**	2.23	1.27	1.04	1.39	1.14	3.34	1.35	2.38
Hash	1.61	1.50	2.38	1.43	0.99	2.12	**22.32**	1.79	2.01	0.95	1.69	3.23	2.46	2.10
Jamie	1.57	1.37	1.32	1.62	0.98	1.28	1.72	**7.49**	1.29	0.85	1.55	2.75	1.56	1.51
James	2.17	2.32	2.21	1.77	1.64	1.22	1.53	1.42	**14.15**	0.90	2.23	2.09	2.18	1.42
Kim	1.00	1.26	0.97	1.72	1.18	1.30	1.04	0.83	0.89	**8.24**	0.92	1.20	1.20	1.45
Luke	2.39	2.84	1.85	2.23	1.47	1.50	1.72	1.94	2.46	1.25	**12.06**	3.29	2.20	1.89
Matt	1.78	1.76	1.49	1.62	1.03	2.05	3.36	1.93	2.01	1.03	2.48	**16.26**	1.66	2.23
Vedran	3.70	3.10	4.96	4.34	3.23	2.19	2.93	1.63	2.80	1.42	2.33	2.91	**23.51**	3.10
Vitaly	2.90	2.42	2.26	3.14	2.00	2.52	2.02	1.48	1.61	1.59	1.63	3.26	2.57	**14.63**
Laura	0.74	0.94	0.82	1.19	1.05	0.31	0.40	0.46	0.62	0.50	0.61	0.45	0.94	0.47
Melanie	1.12	1.08	1.00	1.35	1.52	1.65	0.81	0.67	0.83	1.76	0.62	0.94	1.07	1.68

the control parameters of the 3-D ASM, we observe that the first Eigencoefficient corresponds to the lip opening. The higher order Eigencoefficients correspond to other modes of variations such as the lip stretch, asymmetries, and also physiological information of the lip shapes.

The more Eigencoefficients we employ, the more details of the subjects' idiosyncrasies are accounted for and the biometric trait becomes more distinctive. However, the higher order Eigencoefficients also contain not only "noise" but also the intrasubject variations, which degrade the recognition performance. Fig. 12 shows the effect of using increasing numbers of Eigencoefficients; the performance of the recognition system improves when more modes of variation are taken into account. Experimentally, we observed that keeping 90% of the variations (i.e., using 14 Eigencoefficients in the present experiment) is a good choice. There is no significant improvement on the recognition rate when more than 14 Eigencoefficients are used, and the system performance degrades beyond 100 Eigencoefficients.

C. Identification Problem

While the face verification problem is a *one-to-one* matching between a probe and a claimed identity, the face identification problem is a *one-to-many* comparison without any initial knowledge of the user's identity. The system can operate in two modes. *Watch list:* are you in my database? and *basic identification:* you are in my database, can I find you? Both scenarios can be formally stated as follows [28]: given an input feature vector $\vec{\mathbf{v}}_Q$ of an unknown person, determine the identity $I_k, k \in \{1, 2, \ldots, N, N+1\}$ where $I_1, I_2, \ldots, I_N$ are the identities enrolled in the system and I_{N+1} indicates the reject case where no suitable identity can be determined for the user

$$\vec{\mathbf{v}}_Q \in \begin{cases} I_k, & \text{if } \max_k S(\vec{\mathbf{v}}_Q, \vec{\mathbf{v}}_{I_k}) \geq t, k = 1, 2, \ldots, N \\ I_{N+1}, & \text{otherwise} \end{cases} \tag{3}$$

where $\vec{\mathbf{v}}_{I_k}$ is the biometric template in the system database corresponding to identity I_k, S is a similarity measure, and t is a threshold. Table III shows a subset of the matching results: the lines correspond to the probes, which are matched against the gallery templates listed in the columns. The quantity computed is the similarity between two feature vectors $S = 1/D_{WDTW}$. The system is capable of recognizing the correct user since the similarity between utterances from the same subject is always greater than that between different subjects. In practice, the correct answer does not always correspond to the best match, therefore the performance of the system is measured by plotting the cumulative match curve (CMC) [27], as shown in Fig. 13.

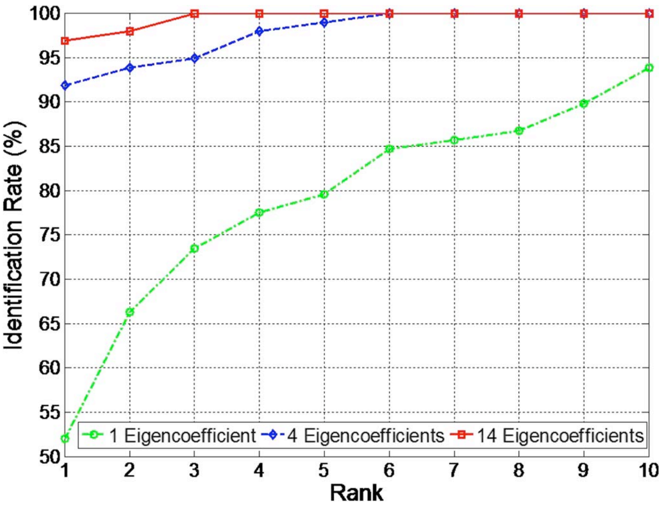

Fig. 13. CMC of the identification system. Percentage of correct answers found in the top m best match, for several values of m.

When the system is used in the *Watch list* mode, a threshold needs to be determined such that users who are unknown to the system are rejected. With reference to the example shown in Table III, *Laura* and *Melanie* are not enrolled in the gallery. Applying equation (3), any threshold $t > 2.00$ permits to correctly reject *Laura* and *Melanie* as unknown to the system watch list.

D. Permanence and Uniqueness of Facial Actions

The discriminatory power of any biometric feature depends essentially on its permanence (i.e., small intrasubject variations) and its uniqueness (i.e., large intersubject variations). As

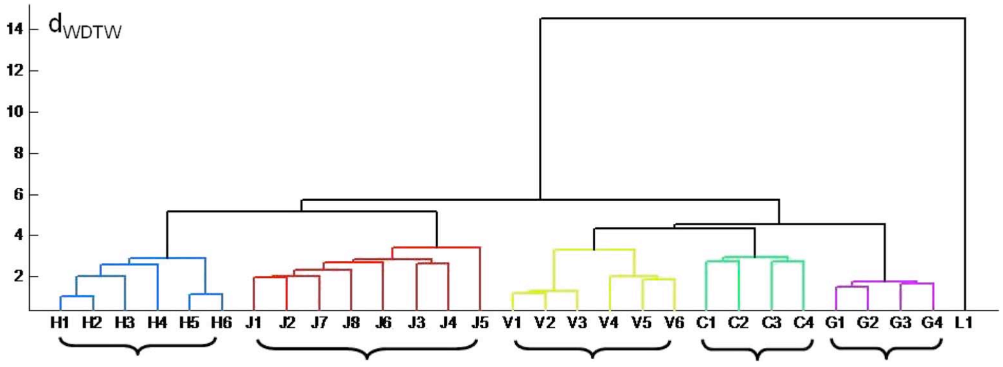

Fig. 14. Clustering of the puppy utterances of a number of subjects recorded over a large period of time.

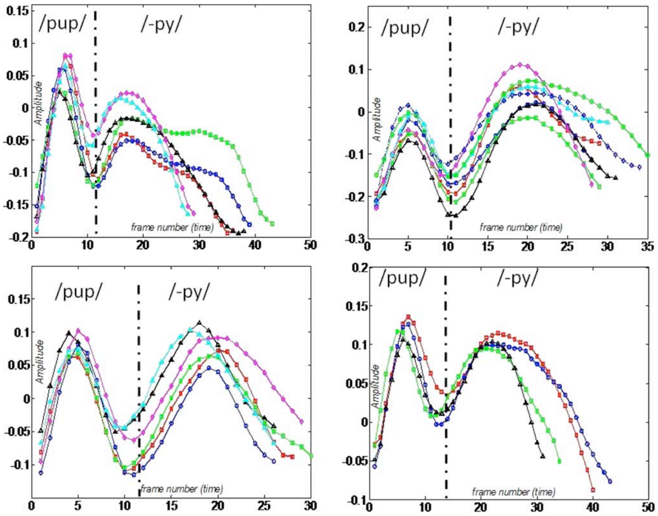

Fig. 15. Reproducibility of the "puppy" utterances from four subjects recorded over several month intervals. Variations of the first Eigencoefficient is shown, which corresponds to the lip opening. The closed syllable /pup/ appears more stable compared to the open syllable /-py/. However, the coda of the /pup/ syllable seems affected by coarticulation.

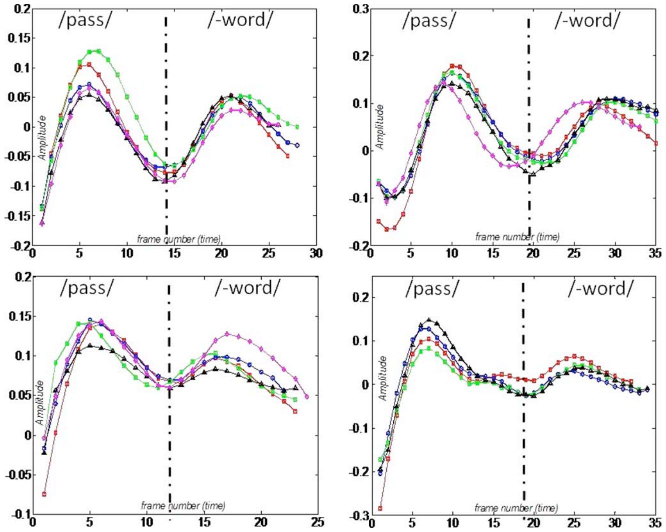

Fig. 16. Dynamics of the "password" utterances of four subjects recorded over several month intervals. Variations of the first Eigencoefficient which corresponds to the lip opening.

a behavioral trait, facial dynamic is not expected to be strictly reproducible over time. Thus, our next experiment aims to measure to which extent facial dynamics can be considered permanent, and to assess whether the intersubject distinctiveness is sufficiently large compared to the intrasubject fluctuations.

Variations Over Long Time Intervals: Six participants, code-named H, J, V, C, G, and L have been recorded uttering the word "puppy" several times. All participants are natural English speakers with the exception of V. Recordings of H and G were taken over more than two years, J were recorded over three months and over one month for C. V was recorded on the same day, with only 15 min between the two recording sessions and with three repetitions per session. L is used as an unseen subject.

The first utterance of five subjects H1, J1, V1, C1, and G1 are employed to build a 3-D ASM which is then used to estimate the feature parameters of all the utterances. The pairwise distances of the facial dynamics are computed and the *dendrogram* was formed, as shown in Fig. 14.

We can observe that the reproducibility of facial actions is very person dependent. For example, H's and G's utterances appear more stable over two years compared to V's within the same day. In addition, there is a greater similarity between utterances taken in the same session, e.g., {H1 and H2}, {H5 and H6}, {J1 and J2}, {J3 and J4}, {V1,V2 and V3}, {V4,V5 and V6}, etc. The distinctiveness across speakers is more significant than the intrasubject variations, thus the clusters are accurately formed. L is an unseen subject who presents a very important dissimilarity to all of the subjects in the gallery. These observations indicate that facial dynamics of even very short facial actions provides sufficient discriminatory power to be used for person recognition.

Examination of Different Speech Postures: Fig. 15 shows the dynamics of four different subjects uttering the word "puppy" several times. A careful examination shows that the closed syllable (short vowel) /pup/ appears more stable compared to the open syllable (long vowel) /-py/. In particular, the onsets of the syllable /pup/ are nearly identical. This observation holds for all the 94 subjects recorded in our database. Let us now compare the reproducibility of different types of syllables on other utterances.

As shown in Fig. 16, the repetitions of the word "password" (CVCC/CVC, long vowel/r-controlled vowel) seem to indicate

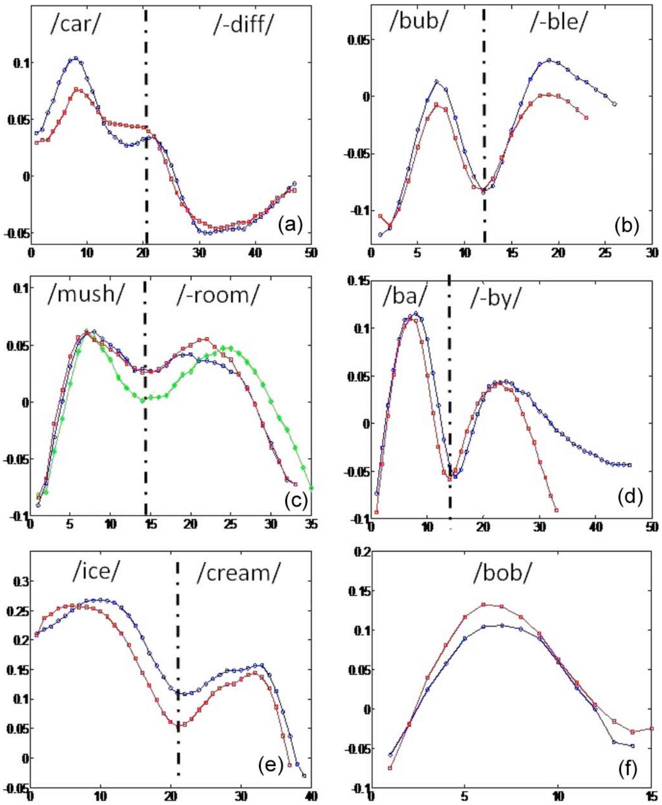

Fig. 17. Dynamics of several speech postures: (a) "cardiff," (b) "bubble," (c) "mushroom," (d) "baby," (e) "ice cream," and (f) "bob." Short vowels seem to be more stable over time, compared to long vowels, r-controlled, and l-e-controlled vowels. Variations of the first Eigencoefficient is shown.

that closed syllables are repeatable over time. This can also be observed in the dynamics of other utterances shown in Fig. 17. The closed syllables, e.g., /diff/ in "cardiff," /bub/ in "bubble" and /bob/ appear reproducible across the repetitions, while the open syllables (long vowels) such as /-by/ in "baby," and the two syllables in "mushroom" and in "ice cream" present fluctuations in their dynamics, with regard to both the durations and the trends of the trajectories. The r-controlled vowel /car/ in "cardiff," and the l-e-controlled vowel in "bubble" also show significant fluctuations.

These observations indicate that there is a hierarchy in the reproducibility of syllables, which means that some types of speech-related facial movements are more suitable for biometric compared to others. However, our experiments are not sufficient to establish any formal rule at the moment. In particular, we can see that the r-controlled vowels /car/ in "cardiff" and /word/ in "password" do not exhibit the same reproducibility. In addition, the open syllable /ba/ in "baby" seems very stable contrary to our predictions. Therefore, further investigations are needed to fully understand the speaking patterns. In particular, we need to examine if the place of the syllable in the word and the stress on the syllable can influence its behavior. We are collecting data in order to investigate this question in a future work.

For the time being, we can already compare the recognition performance when using the facial dynamics of different utterances, the recognition performances are shown in Fig. 18 where the EER is plotted as a function of the number of Eigencoefficients used.

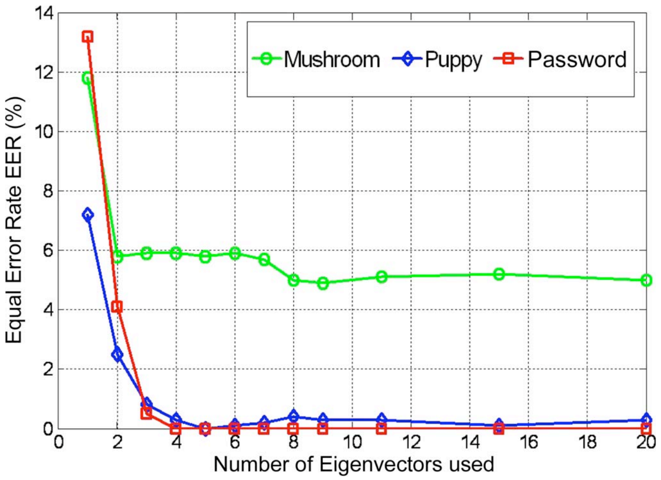

Fig. 18. EER of a face verification system using the dynamics of different words. There exists a hierarchy in the biometric power of speech-related facial actions.

We observe that using the "mushroom" utterance does not allow us to recognize the subjects' identities as accurately as using the "puppy" and "password" utterances. This supports the idea that not all speech-related facial actions are equivalent with regard to their biometric power. The recognition performance comparison conforms to our prediction. The utterance "password" which has two short vowels is expected to be more reproducible/permanent, the utterance "puppy" is the combination of a short vowel followed by a long one is less stable, and finally, the word "mushroom" which has two long vowels is the less permanent, hence has a lower recognition performance. On the other hand, the EER decreases very quickly at the beginning when the number of Eigencoefficients increases. However, beyond eight Eigencoefficients, the recognition performance remains constant and even slightly increases due to the presence of noise.

E. Nonverbal Facial Actions

Fig. 19 shows three sequences of the same subject performing a maximal smile, the videos were captured over several months, and Fig. 20 shows the temporal variations of the three lead Eigencoefficients corresponding to the smile sequences. Considering the large variety of smile patterns that a person can produce, we wish to standardize the performance and required the subject to perform an AU12 maximal smile (closed lips, no significant facial movements in other facial regions than the lips). Although clear instructions have been given, significant intrasubject variations and inaccuracies can still be noticed across the repetitions, which result in the large fluctuations of the trajectories of the Eigencoefficients.

Examinations of the smile dynamics from 50 different participants show similar intrasubject variations and many involuntary facial actions such as blinking, eyebrow movements, and facial asymmetries. For these reasons, the holistic feature extraction does not seem adequate (see Section V-B1). However, the local feature extraction approach also shows very poor recognition result because the lip motions alone are not

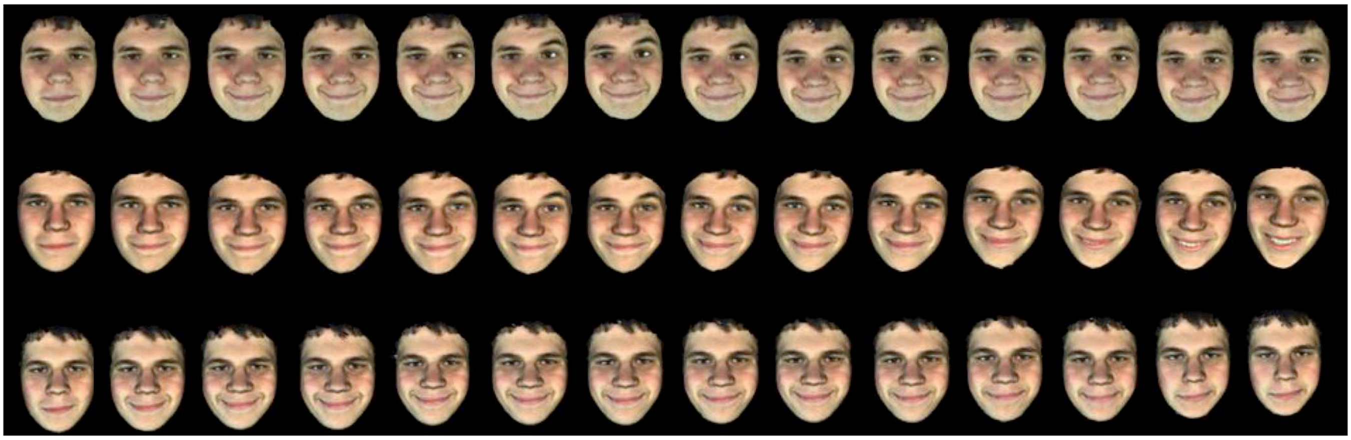

Fig. 19. Three video sequences of a naive user performing an AU12-standardized maximal smile. The data were recorded over three months to assess the variability of the subject's emotional and physical conditions. Significant variations and inaccuracies can be observed, e.g., lip opening, eyebrow movements, and facial asymmetries. The expression appears very unstable.

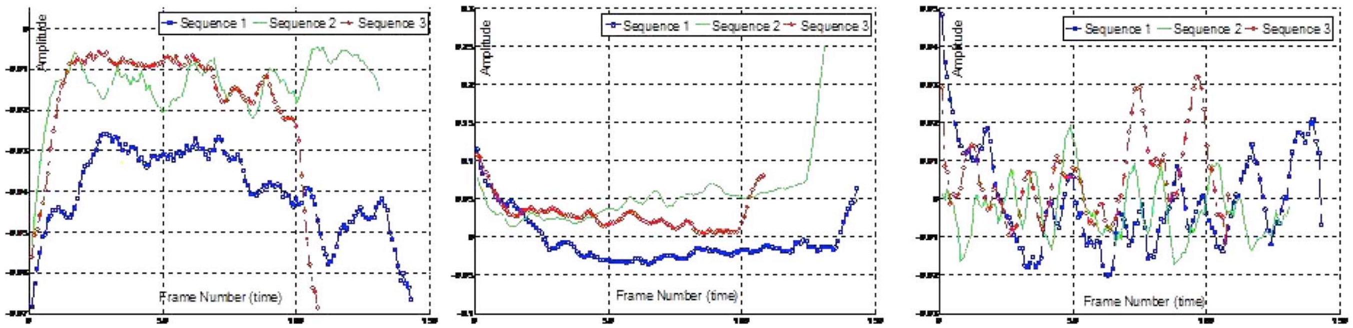

Fig. 20. Smile dynamics corresponding to the three sequences shown in Fig. 19. Variations of the three lead Eigencoefficients. In this experiment, we adopt the holistic feature extraction approach in order to capture the idiosyncrasies of the entire face.

sufficiently discriminative in the example of the smile. Other experiments have been carried out on other facial expressions, e.g., disgust and sadness, where similar poor performance has been observed (≈25% error rate). Thus, we believe that nonverbal facial expressions are not suitable for biometrics. In addition to the limitations previously mentioned, there exists also another disadvantage to the use of emotional expressions for person identification. Even in the case of subjects who produce accurate expressions, the holistic feature extraction approach is still problematic because it is computationally prohibitive and not suitable for a real-time application such as face recognition.

Finally, regarding the use of FACS Action Units for person identification, we have observed during the recording sessions that it is very difficult for naive users to produce accurate AUs, let alone to repeat identical performances. Fig. 6 shows a subject performing AU9 (nose wrinkler). Perceptual evaluation done by a FACS-trained psychologist has found this sequence to include not only AU9 but also AU4 (brow lowerer), AU10 (upper lip raiser) and AU17 (chin raiser) [21]. In another example, although the participants were asked to perform a maximal AU12 smile, cooccurrences of AU6 (cheek raiser), AU25 (lips part), AU45 (blinking) and AU61-64 (eye movements) can also be observed. This raises the question whether AU-based face expressions are suitable in a real-life scenario where users are not familiar with the FACS coding system.

F. Minimizing the Intrasubject Variations

The experiments described in the previous sections have shown that speech-related facial actions are the most suitable for biometrics. However, significant intrasubject variations can be observed across the repetitions, in particular for long vowels. Fig. 21(a) shows six performances from the same subject uttering the word "puppy" captured over several recording sessions. It is clear that the recognition performance will suffer if we mistakenly choose an outlier as a reference template (i.e., the template used in the gallery). One possible solution to overcome such problem consists of recording several performances and then computing the "principal curve" which best approximates the population, as shown in Fig. 21(b). Methods to compute such a curve can be found in [28]. Table IV shows the distances d_{WDTW} between the utterances and the principal curve. We observe that the latter represents an adequate reference template, because it is close to *all* of the utterances.

In this paper, the idea of computing the principal curve is not relevant because we can use only one utterance per subject in the gallery. Therefore, it is possible that a number of outliers were chosen as the reference templates, which may have degraded the recognition performance. However, in a near future, we plan to obtain more repetitions per subject, and we will be able to see if computing the principal curve can improve the recognition rate.

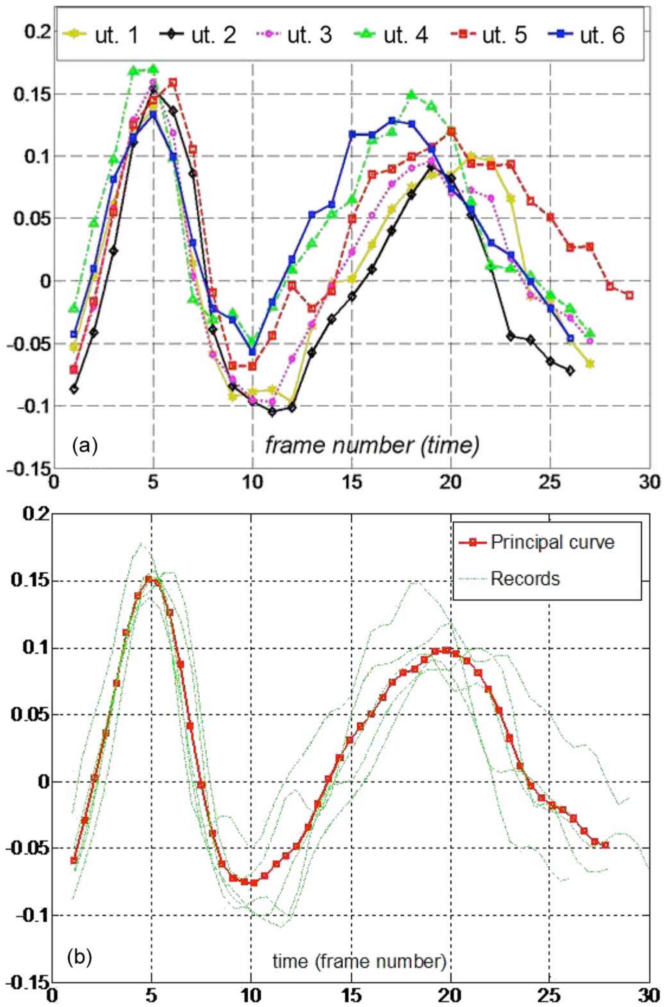

Fig. 21. Intrasubject variations and their principal curve.

TABLE IV
DISTANCES D_{WDTW} BETWEEN SIX "PUPPY" UTTERANCES FROM THE SAME SUBJECT AND THEIR DISTANCES TO THE PRINCIPAL CURVE PC

	ut. 1	ut. 2	ut. 3	ut. 4	ut. 5	ut. 6	PC
ut. 1	0	0.010	0.005	0.026	0.016	0.016	0.010
ut. 2	0.010	0	0.008	0.031	0.019	0.020	0.014
ut. 3	0.004	0.008	0	0.022	0.011	0.014	0.008
ut. 4	0.026	0.031	0.022	0	0.011	0.009	0.020
ut. 5	0.016	0.019	0.011	0.012	0	0.011	0.012
ut. 6	0.016	0.020	0.014	0.009	0.010	0	0.009

VI. CONCLUSION AND FUTURE WORK

The purpose of this paper is to investigate the feasibility of using facial actions for biometric identification. Although previous works have been carried out on using lip motions for speaker recognition, there was very little emphasis on studying in depth the permanence and distinctiveness of facial motions. Furthermore, the problem has never been evaluated in an environment resembling a real-life scenario. In this paper, we review the face databases and the recognition methods used in the previous works, in light of which we propose several improvements.

First, we explain our choice of a new and meaningful set of *very short* verbal and nonverbal facial actions, which allow comparison of the biometric power of different types of facial movements. The evaluation is carried out on 3-D dynamic data which has been recorded over long time intervals to assess the variability of the subjects' emotional and physical conditions. The use of very short facial actions is essential in order to minimize the processing effort, which is very important for a real-time application such as face recognition.

Then, we implement a face recognition system using facial dynamics, and explain step by step the data preprocessing techniques, the feature extraction method and propose an accurate pattern matching algorithm derived from DTW. Quantitative evaluations are carried out considering both the face verification and the face identification problems. The results indicate that even *very short* verbal facial actions (e.g., utterance of a two-syllable word) provide sufficient information for person identification. However, while speech-related facial motions show good potential, emotional expressions (e.g., smile and disgust) and the FACS Action Units appear very unstable, particularly when naive users are employed. This indicates that nonverbal expressions may not be suitable for biometrics. In the particular case of speech-related facial motions, we observe that there exists a hierarchy in the reproducibility of different types of utterances. While short vowels seem very repeatable, even over long time periods, long vowels show significant fluctuations. These observations open a new horizon for future works. If we can establish formal rules to identify which types of facial motions are the most reproducible and discriminative across speakers, this will help the users to select strong passwords and the recognition performance will be improved.

ACKNOWLEDGMENT

The authors would like to thank all the members of staff and students of the Schools of Dentistry, Engineering and Computer Science, Cardiff University, U.K., for the volunteer participation. The authors would also like to thank J. MacKenzie for his valuable help to collect the face data used in this paper.

REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[2] P. J. Phillips, "Support vector machines applied to face recognition," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 1999, vol. 2, pp. 803–809.

[3] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 125–143, Jun. 2006.

[4] K. W. Bowyer, K. Chang, and P. J. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Comput. Vis. Image Understand.*, vol. 101, no. 1, pp. 1–15, Jan. 2006.

[5] K. Chang, K. Bowyer, and P. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.

[6] G. Edwards, T. Cootes, and C. J. Taylor, "Face recognition using active appearance models," in *Proc. Eur. Conf. Comput. Vis.*, 1998, vol. 2, pp. 581–595.

[7] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[8] J. Luettin, N. A. Thacker, and S. W. Beet, "Speaker identification by lipreading," in *Proc. Int. Conf. Spoken Language*, 1996, vol. 1, pp. 62–64.

[9] M. J. Roach, J. D. Brand, and J. S. Mason, "Acoustic and facial features for speaker recognition," in *Proc. Int. Conf. Pattern Recog.*, 2000, vol. 3, pp. 258–261.

[10] M. I. Faraj and J. Bigun, "Audio-visual person authentication using lip-motion from orientation maps," *Pattern Recognit. Lett.*, vol. 28, no. 11, pp. 1368–1382, Aug. 2007.

[11] S. Pigeon and L. Vandendrope, "The M2VTS multimodal face database," in *Proc. 1st Int. Conf. Audio- Video-Based Biometric Person Authentication*, 1997, pp. 403–409.

[12] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. 2nd Int. Conf. Audio- Video-Based Biometric Person Authentication*, 1999, pp. 72–77.

[13] C. C. Chibelushi, S. Gandon, J. S. Mason, F. Deravi, and D. Johnston, "Design issues for a digital integrated audio-visual database," in *Proc. IEE Colloq. Integr. Audio-Visual Process. Recog., Synthesis, Commun.*, 1996, pp. 7/1–7/7.

[14] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 12–28, Jun. 1989.

[15] T. J. Hutton, B. F. Buxton, and P. Hammond, "Dense surface point distribution models of the human face," in *Proc. IEEE Workshop Math. Methods Biomed. Image Anal.*, 2001, pp. 153–160.

[16] P. Ekman and W. Friesen, "The facial action coding system: A technique for the measurement of facial action," in *Manual for the Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[17] D. Johnston, D. T. Millett, and A. F. Ayoub, "Are facial expressions reproducible?" *Cleft Palate-Craniofacial J.*, vol. 40, no. 3, pp. 291–296, 2003.

[18] D. Fidaleo and M. Trivedi, "Manifold analysis of facial actions for face recognition," in *Proc. ACM SIGMM Workshop Biometrics Methods Appl.*, 2003, pp. 65–69.

[19] S. Pamudurthy, E. Guan, K. Mueller, and M. Rafailovich, "Dynamic approach for face recognition using digital image skin correlation," in *Audio and Video-Based Biometric Person Authentication*. Berlin, Germany: Springer-Verlag, 2005, pp. 1010–1018.

[20] X. Lu, "3D face recognition across pose and expression," Ph.D. dissertation, Michigan State Univ., East Lansing, MI, 2006.

[21] A. Efrat, Q. Fan, and S. Venkatasubramanian, "Curve matching, time warping, and light fields: New algorithms for computing similarity between curves," *J. Math. Imaging Vis.*, vol. 27, no. 3, pp. 203–216, Apr. 2007.

[22] N. Dhananjaya, "Correlation-based similarity between signals for speaker verification with limited amount of speech data," in *Multimedia Content Representation, Classification and Security*. Berlin, Germany: Springer-Verlag, 2006, pp. 17–25.

[23] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.

[24] M. E. Munich and P. Perona, "Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 108–115.

[25] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," presented at *1st SIAM Int. Conf. Data Mining*, 2001.

[26] L. Benedikt, D. Cosker, P. L. Rosin, and D. Marshall, "Facial dynamics in biometric identification," in *Proc. BMVC*, 2008, vol. 2, pp. 235–241.

[27] R. M. Bolle, J. H. Connell, S. Pankanti, and N. K. Ratha, *The Guide to Biometrics*. Berlin, Germany: Springer-Verlag, 2004.

[28] T. Hastie and W. Stuetzle, "Principal curves," *J. Amer. Stat. Assoc.*, vol. 84, no. 406, pp. 502–516, Jun. 1989.

[29] A. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[30] S. A. Lesner and P. B. Kricos, "Visual vowels and diphthongs perception across speakers," *J. Acad. Rehabil. Audiology*, vol. 14, pp. 252–258, 1981.

[31] J. S. Mason and J. D. Brand, "The role of dynamics in visual speech biometrics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. 4, pp. 4076–4079.

[32] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, Face Recognition Vendor Test (FRVT) 2006. [Online]. Available: http://www.frvt.org/FRVT2006

[33] J. Laver, *Principles of Phonetics*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[34] P. Lucey, T. Martin, and S. Sridharan, "Confusability of phonemes grouped according to their viseme classes in noisy environments," in *Proc. Australian Int. Conf. Speech Sci. Technol.*, 2004, pp. 265–270.

[35] W. M. Weikum, "Visual language discrimination," Ph.D. dissertation, Univ. British Colombia, Vancouver, BC, Canada, 2008.

[36] J. F. Cohn, K. Schmidt, R. Gross, and P. Ekman, "Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification," in *Proc. Int. Conf. Multimodal User Interfaces*, 2002, vol. 116, pp. 491–498.

[37] S. Tulyakov, T. Slowe, Z. Zhi, and V. Govindaraju, "Facial expression biometrics using tracker displacement features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–5.

[38] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," in *Proc. 8th Int. Conf. Autom. Face Gesture Recog.*, 2008, pp. 1–6.

[39] J. F. Cohn and T. Kanade, "Cohn-Kanade AU-coded facial expression database," Pittsburgh Univ., Pittsburgh, PA, 1999.

Lanthao Benedikt received the diplome d'ingenieur degree in electronics and electrical engineering from the Institut National Polytechnique de Grenoble, Grenoble, France, in 1996.

Within the same year, she joined STMicrolelectronics where she worked in the Technical Marketing team of the Wireless Communication Department until 2004. She is currently with the School of Computer Science, Cardiff University, Cardiff, U.K.

Darren Cosker received the Ph.D. degree in computer science from the University of Cardiff, Cardiff, U.K., in 2006.

He is a Royal Academy of Engineering/Engineering and Physical Sciences Research Council Postdoctoral Research Fellow based with the Department of Computer Science, University of Bath, Bath. He is also a Visiting Researcher with the Centre for Vision, Speech and Signal Processing, University of Surrey, Surrey. His research interests are in human motion analysis, modeling, and animation.

Paul L. Rosin received the B.Sc. degree in computer science and microprocessor systems from Strathclyde University, Glasgow, U.K., and the Ph.D. degree in information engineering from City University London, London, U.K., in 1984 and 1988, respectively.

He was a Lecturer with the Department of Information Systems and Computing, Brunel University London, Uxbridge, U.K., a Research Scientist with the Institute for Remote Sensing Applications, Joint Research Centre, Ispra, Italy, and a Lecturer with Curtin University of Technology, Perth, Australia. He is currently a Reader with the School of Computer Science, Cardiff University, Cardiff, U.K. His research interests include the representation, segmentation, and grouping of curves, knowledge-based vision systems, early image representations, low-level image processing, machine vision approaches to remote sensing, methods for evaluation of approximations, algorithms, etc., medical and biological image analysis, mesh processing, and the analysis of shape in art and architecture.

David Marshall received the B.Sc. degree in mathematics from University College, Cardiff, U.K., in 1986, where he is currently working toward the Ph.D. degree in 3-D inspection of manufactured objects, working in collaboration with the Sowerby Research Centre, British Aerospace, Farnborough.

Since 1986, he has been working in the field of computer vision. Since 1989, he has been a Lecturer and is currently a Reader with the School of Computer Science, Cardiff University. His current research interests include articulated modeling of human faces, models of human motion, statistical modeling, high dimensional subspace analysis, audio/video image processing, and data/sensor fusion. He has published over 100 papers and one book in these research areas.