3D Facial Gestures in Biometrics: from Feasibility Study to Application

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

Abstract— **The human face has been so far mainly seen as a physiological biometric and very little work has been done to exploit the idiosyncrasies of facial gestures for person identification. This study proposes a markerless method to capture 3D facial motions, and investigates a number of pattern matching techniques which operate accurately on *very short* facial actions. Qualitative and quantitative evaluations are performed for both the face identification and the face verification problems. The emphasis is placed on designing a system which is not only accurate but also usable in a real-life scenario.**

I. INTRODUCTION

The face is the primary means for humans to recognise each other in everyday life and therefore represents the most natural choice of biometric technology. Early works such as Eigenfaces [1], Fisherfaces [2] and Active Appearance Model [3] have inspired a number of 2D face recognition solutions which are now deployed in various commercial and forensic applications. For example, Toshiba's new generation of laptop Satellite U400 will soon introduce a face verification to substitute for a password. In another utilisation, London airports will experiment this summer with a security system which scans airline passengers' faces and match them against biometric data stored on passport chips.

Despite such wide acceptance, 2D face recognition suffers a number of drawbacks such as its sensitivity to lighting conditions and head-pose variations [4]. These limitations have motivated the development of 3D techniques which constitute the current state-of-the-art in the field [5]. However, both 2D and 3D algorithms still share one common drawback which is their sensitivity to face expressions. While Chang et al. [6] proposed an expression-invariant face recognition technique, other studies consider facial gestures as an additional source of information rather than a problem to overcome and analyse the idiosyncracies of facial dynamics through lipreading [7][8][9], the bio-mechanical characteristics of facial tissues [10][11], but also manifold analysis [12]. Promising preliminary results have been reported, which are furthermore confirmed by Perceptual Studies suggesting that humans use facial motions as a cue for identity recognition [13], and these complex sequences of muscle activations are almost impossible to spoof [10].

In this study, we further investigate the use of facial gestures for identity recognition. The following aspects are discussed: 1) We evaluate the existing standard face databases and propose new facial actions which are *very short*, and offer strong biometric power in terms of *reproducibility* and *distinctiveness*. 2) We investigate a number of well-established pattern matching techniques used in similar problems (e.g. signature authentication, gait, voice recognition) and adapt these algorithms to the face recognition problem. 3) Finally, we outline the structure of an identification system using 3D facial gestures which is not only *accurate* but also *usable* in a real-life scenario.

L. Benedikt, Paul L. Rosin and D. Marshall are with the School of Computer Science, Cardiff University, 5 The Parade, Roath, Cardiff, CF24 3AA, United Kingdom l.truong-benedikt@cs.cardiff.ac.uk paul.rosin@cs.cardiff.ac.uk dave.marshall@cs.cardiff.ac.uk

D. Cosker is with the School of Computer Science, University of Bath, Bath BA2 7AY, United Kingdom dpc@cs.bath.ac.uk

II. CHOICES OF FACIAL ACTIONS

In principle, any facial gestures can be employed for person recognition. However, choosing those which exhibit high *distinctiveness* and *reproducibility* is similar to choosing strong passwords over weak ones. Besides, using *very short* facial actions helps reduce the processing cost so that genuine users can gain access quickly to the secure services.

The majority of researches on facial motion analysis employ standard datasets such as the 2D Cohn-Kanade database of non-verbal face expressions [14], or the M2VTS [16] and XM2VTSDB [17] databases which include audio-video data of continuously uttered digits from '0' to '9', and spoken phrases such as *"Joe took father's green shoe bench out"*. Also commonly used is the DAVID database [18] where 9 participants wearing blue lipstick utter isolated digits. Such use of physical markers is inconvenient in a real-life scenario. Besides, although long phrases are necessary for speech synthesis, they are not optimal for recognition because computationally too expensive to process, especially in 3D. For these reasons, we choose to collect a new dataset. A literature review of Orthodontics and Craniofacial researches helps identify the most distinctive and reproducible facial gestures [19]. Two directions are investigated:

Non-verbal facial actions: a number of facial action units (AU) based on the Facial Action Coding System (FACS) [15] are considered, for example AU1+2 (brow raiser), AU5 (upper lid raiser), AU9 (nose wrinkler) and AU12 (lip corner puller). Gestures involving maximal muscle stretches (e.g. maximal smile) are preferred since these allow more accurate reproduction compared to moderate expressions [19].

Verbal facial actions: the spoken word 'Puppy' was chosen among others not only because it involves the most distinctive and repeatable muscle activations [19], but also because most of the variations occur at the visible parts of the face, which are easier to capture than variations involving the teeth and tongue. Besides, this speech posture gives a good representation of facial movements in the lower third of the face while leaving stable the upper thirds for an accurate alignment of the sequential frames.

978-1-4244-2730-7/08/$25.00 ©2008 IEEE

III. DATA ACQUISITION

A 3D motion analysis image capture system operating at 48 frames per second (3dMDFace^{TM} Dynamic System) is used to capture 3D video data from 50 participants (31 males and 19 females, 33 are natural English speakers). All face scans have an ear-to-ear coverage and no physical markers are used. Each participant is asked to utter the word 'Puppy' several times in a normal and relaxed way, and to perform an AU12-based maximal smile. Smaller tests are conducted for other AUs due to limited data availability.

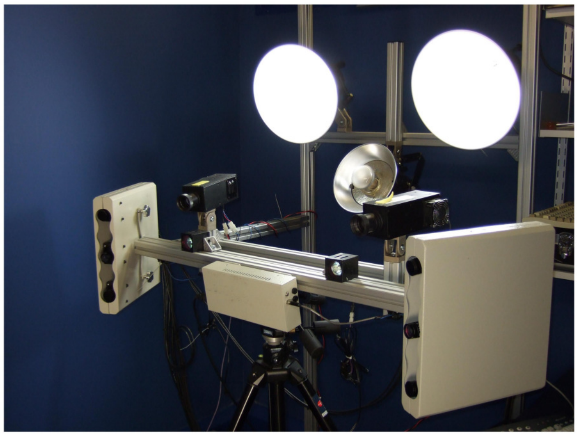


Fig. 1. Photographic representation of the 3dMDFace^{TM} Dynamic System

One non-trivial preliminary task in face recognition is to normalise the face scans so that they can be compared in the same frame space (i.e. same head pose and size). To this end, Fidaleo et al. [12] 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 study, we use a 3D video camera to capture image sequences in a semi-controlled mode where small head-pose variations are allowed as long as we ensure an ear-to-ear coverage of the face. The normalisation is computationally achieved using nose-matching as follows: the nose root is located using 3D curvature analysis and the faces are aligned through translations, rotations and scaling. Further refinements are achieved by applying the Iterative Closest Point matching algorithm at the most stable regions around the nose root [21].

IV. FEATURE EXTRACTION

Many novel techniques have been proposed for extracting facial dynamics. Among non-invasive approaches which do not require the use of physical markers, we can find for example the work of Pamudurthy et al. [11] which aims to track the motions of skin pores, or the work of Faraj et al. [9] for extracting the velocity of lip motions. However, we believe that the well-established methods used in static face recognition [3][5] can be extended for 3D dynamic data and in a previous work [20], we proposed a feature extraction technique which consists of two main phases as follows:

- *Bring all frames into correspondence* (i.e. same number of vertices and topology) *:* a face template is warped to each frame using a thin-plate-splines process, which gives a good initial superimposition of the two meshes. Then each vertex of the warped template is projected along its normal onto the frame so that the former conforms to the shape of the latter. Details of this pre-processing step can be found in [20].
- *Extraction of the facial dynamics:* a combined model of 3D shape and texture is built in similar fashion to the Active Appearance Model [3], but all vertices are included as with the Morphable Model [5]. Fig. 2 depicts the temporal variations of the first Eigen-coefficient and the corresponding face expressions of a subject performing AU9 (nose wrinkler).

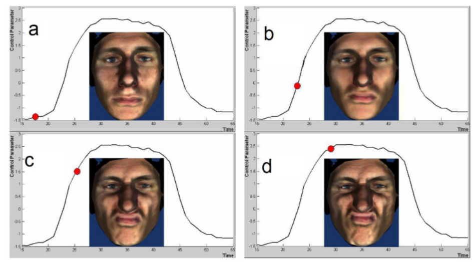


Fig. 2. Face expression sequence of a subject performing AU9 (Nose Wrinkler) and the temporal evolution of the first Eigen-coefficient.

1) Nota Bene 1: for the remaining of this study, we use the 3D shape information alone (i.e. no texture) in order to reduce the processing cost. In addition, we retain only 90% of the variations in order to reduce the dimensionality of the feature space. This involves keeping the p highest Eigenvectors. Thus, any shape $\overrightarrow{\mathbf{s}}_k$ can be approximated as:

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

where $|\overrightarrow{\mathbf{s}}|$ is the mean shape, Φ is the matrix of p Eigenvectors, and $\overrightarrow{\mathbf{v}}_k$ is the vector of Eigen-coefficients corresponding to the shape $\overrightarrow{\mathbf{s}}_k$. Inverting this equation, we can extract the Eigen-coefficient variations for a sequence of D shapes as follows, where k is the frame number:

$$\overrightarrow{\mathbf{v}}_k \approx \Phi^T(\overrightarrow{\mathbf{s}}_k - |\overrightarrow{\mathbf{s}}|), \quad k \in \{1, ..., D\} \tag{2}$$

2) Nota Bene 2: the facial motion analysis can be either holistic (complete face), or localised (only the region where the motions occur is analysed). For the remaining of this study, we adopt the latter approach and analyse the lip region only. This is justified by the following considerations:

- This reduces the amount of data to process. For example, a face template includes $\approx$ 25,000 vertices compared to $\approx$ 1,000 vertices for a lip template.
- This permits to cancel the perturbation due to unwanted actions such as blinking and gaze movements.
- The holistic approach accounts for both the physiological biometric trait and the behavioural peculiarities of facial gestures. The localised approach permits to isolate the latter aspect to a certain degree.

V. PATTERN MATCHING ALGORITHMS

Pattern matching algorithms such as Hidden Markov Models (HMM) and Gaussian Mixture Models are the most commonly employed techniques in facial motion analysis [7][9]. Although these are well-established methods, it is however unclear if they would perform accurately in the present context where we use only very short data sequences in order to minimise the processing cost.

In a related work [26], we conduct a survey of pattern matching techniques used in behavioural biometrics (e.g. gait, voice, signature, key-stroke), and evaluate the best known methods such as the Fréchet distance [21], Correlation Coefficients [22], Dynamic Time Warping (DTW) [23], Continuous DTW (CDTW) [24], Derivative DTW (DDTW) [25], and HMM [7]. In addition, we proposed an improved algorithm Weighted Hybrid Derivative DTW (WDTW). A face verification prototype inspired from [29] is designed for the evaluation as seen in Fig. 3, the comparison result is shown in Fig. 4. The definitions of False-Accept-Rate (FAR) and False-Reject-Rate (FRR) are conformed to [27].

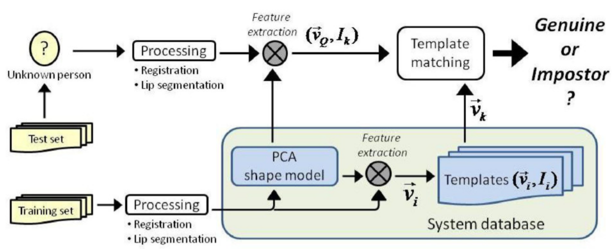


Fig. 3. Architecture of the face recognition prototype. Given an biometric feature vector v_Q of an unknown person and a claimed identity I_k, determine if the person is a genuine user or an impostor. Typically, v_Q is matched against v_k, the biometric template of I_k.

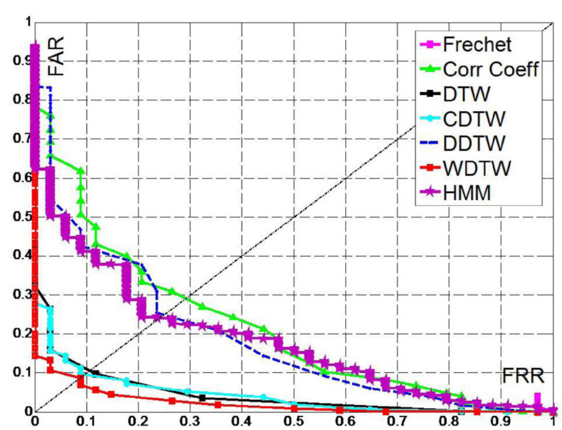
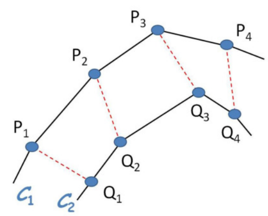


Fig. 4. Performance comparison of several pattern matching algorithms. The recognition process is based on the comparisons of the 'Puppy' utterances. Only the first Eigen-coefficient is used in this experiment.

For the remaining of this study, we will employ only WDTW because our actual goal is to evaluate the use of facial gestures in identity recognition rather than to compare algorithms. The principle of WDTW can be summarised 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. 5. For each point P_i of C_1, we compute its closest point Q_j on C_2 with respect to the distance $d^i = w_0 * d_0^i + w_1 * d_1^i + w_2 * d_2^i$, 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 pair-wise distances d^i, normalised by the sequence length N [26]. Thus, the similarity between two feature vectors can be computed as $S = 1/D_{WDTW}$



$$d^i = w_0 * d_0^i + w_1 * d_1^i + w_2 * d_2^i$$

$$D_{WDTW} = \frac{1}{N}\sum_{i=1}^{N} d^i$$

$$S = \frac{1}{D_{WDTW}}$$

Fig. 5. Weighted Hybrid Derivative DTW algorithm (WDTW).

VI. EXPERIMENT AND DISCUSSION

A. Assessment of the Chosen Facial Actions

1) Reproducibility: during the recording sessions, we observe that it is very challenging for non-actors to produce accurate AUs, let alone to repeat identical performances. Fig. 2 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) [22]. In another example, although the participants were asked to perform a maximal AU12 smile, co-occurrences of AU6 (cheek raiser) and AU25 (lips part) 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.

On the contrary, the 'Puppy' utterance appears more consistent. Fig. 6(a) shows performances of the same subject recorded over 10 months. Although the dynamics are not identical, they exhibit similarities such as comparable onsets and offsets, durations and signal magnitudes. This seems to confirm results in Craniofacial research indicating that the 'rest positions' are the most reproducible [19].

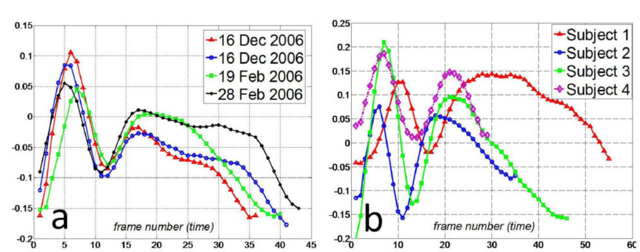


Fig. 6. Dynamic of the 'Puppy' utterance (first Eigen-coefficient). a- Performances of the same subject over 10 months (intra-subject variations). b- Performances of different subjects (inter-subject variations).

2) Distinctiveness: a good biometric trait is expected to differ significantly across individuals such that the inter-subject variations is much stronger compared to the intra-subject variations. The 'Puppy' utterance satisfies this requirement, as shown in Fig. 6(b).

3) Universality: FACS action units are not natural human face expressions and a number of them appear very challenging to perform by naive users in a real-life scenario, e.g. AU1 (inner brow raiser) and AU2 (outer brow raiser).

B. Improving the Biometric Power of Facial Gestures

1) Minimising the Intra-Subject Variations: facial gesture is a behavioural biometric trait, and as such is very sensitive to the subject's emotional condition. Fig. 7(a) shows six performances from the same subject captured over different recording sessions. It is clear that the performance of the recognition system will suffer if we mistakenly choose an outlier as a reference template (see Fig. 3). One way to overcome such problem consists of recording several performances and then compute the 'principal curve' which best approximates the population [28], as shown in Fig. 7(b). Table I shows the distances between the performances and the principal curve. We observe that the principal curve represents a more accurate reference biometric feature.

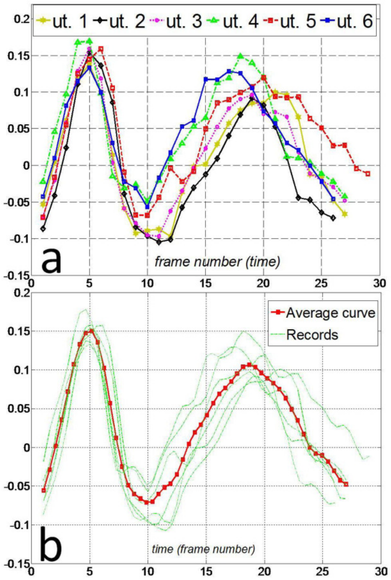


Fig. 7. Intra-subject variations and their principal curve.

2) Using Higher-Order Eigen-coefficients: facial motions result from complex muscle activations and exhibit many degrees of freedom. Fig. 8 shows the 'Puppy' facial dynamics of four different subjects in the space formed by the 3 highest Eigenvectors, and Table II shows the number of Eigenvectors required in order to retain a given percentage of the variations. Fig. 4 shows the recognition rate obtained when using only the first Eigen-coefficient which depicts essentially the lip opening dynamics. In the present experiment, we include additional higher-order Eigen-coefficients and observe their effect on the system performance, as shown in Fig. 9.

TABLE I

DISTANCES D_{WDTW} BETWEEN SIX 'PUPPY' UTTERANCES FROM THE SAME SUBJECT AND THEIR DISTANCES TO THE AVERAGE CURVE.

	ut. 1	ut. 2	ut. 3	ut. 4	ut. 5	ut. 6	Avg
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

TABLE II

PUPPY UTTERANCE: THE NUMBER OF EIGENVECTORS REQUIRED TO RETAIN A GIVEN PERCENTAGE OF THE VARIATIONS.

% of Variations	50%	80%	90%	95%	98%	100%
Required Eigenvectors	1	4	14	99	577	1998

The more Eigen-coefficients we employ, the more details of the subjects' idiosyncrasies are accounted for. Therefore, the biometric trait becomes more distinctive and the False-Accept-Rate decreases. At the same time, the False-Reject-Rate increases because the additional Eigen-coefficients include also 'noise' and the intra-subject variations. However, the overall accuracy of the system improves when more Eigen-coefficients are used, as shown in Fig. 10. The best performance is observed at EER=2%, using 14 Eigen-coefficients. However, using more Eigen-coefficients means increase the processing cost and we observe that beyond 14 Eigen-coefficients (90% of the variations), there is no significant improvement. For example, increasing the percentage of variations from 90% to 95% yields a performance gain of less than 1%, at the cost of computing 99 Eigen-coefficients instead of only 14. Therefore, $p = 14$ Eigen-coefficients is a good compromise between speed and accuracy.

Another advantage of retaining a limited number of Eigen-coefficients $p = 14$ is that this allows to design a compact

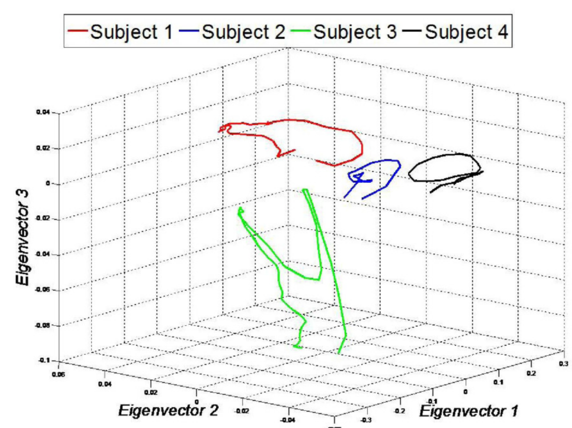


Fig. 8. 'Puppy' facial dynamics of several subjects in the Eigenspace.

recognition system. In fact, we do not store image sequences in the system database, but only an m-by-p feature matrix where m is the length of the feature vector and p is the number of Eigen-coefficients. Typically, a person utters the word 'Puppy' between less than 1 second to 2 seconds. Using a 48fps camera to capture the performance, m varies between 30 and 80 frames. Such a light-weight solution is suitable for small storage capacity such as a passport chip, for example.

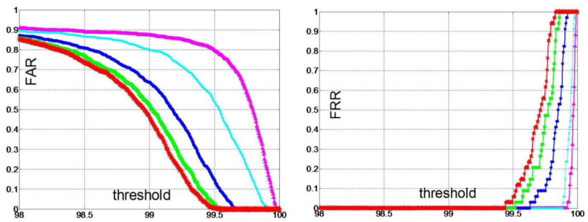


Fig. 9. Using increasing number of Eigen-coefficients in the recognition process improves the False-Accept-Rate, but degrades the False-Reject-Rate.

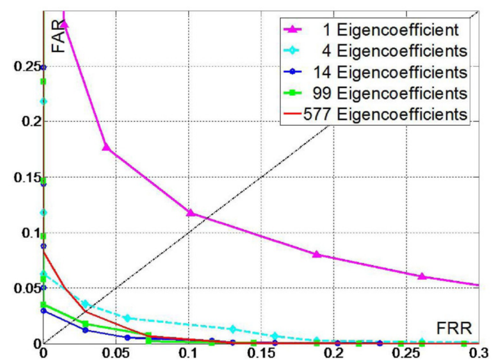


Fig. 10. ROC curves: performance of the verification system for different number of Eigen-coefficients used in the recognition process.

C. The Face Identification Problem

We have previously summarised in Section V the experiment and results from a related study [26] where we examined the face verification problem. This latter is a *one-to-one* comparison of an unknown user against a claimed identity. On the other hand, the face identification problem is a *one-to-many* comparison where the system performance depends on its ability to recognise a person without an initial knowledge of her 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 [29]: given an input feature vector $\overrightarrow{\mathbf{v}}_Q$ of an unknown person, determine the identity $I_k, k \in \{1, 2, ..., N, N+1\}$ where $I_1, I_2, ..., I_N$ 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. Thus

$$\overrightarrow{\mathbf{v}}_Q \in \begin{cases} I_k, & \text{if } \max S(\overrightarrow{\mathbf{v}}_Q, \overrightarrow{\mathbf{v}}_{I_k}) \geq t, k = 1, 2, ..., N \\ I_{N+1}, & \text{otherwise} \end{cases}$$

where $\overrightarrow{\mathbf{v}}_{I_k}$ is the biometric template in the system database corresponding to identity I_k, S is a similarity measure, and t is a threshold. In practice, the correct answer does not always correspond to the best match, therefore the performance measure is the ratio of queries in which the correct answer can be found in the m best matches [27].

In the following experiment, we use the 'Puppy' utterances of 50 subjects to build the database. A distinct set of *unseen* utterances from the same 50 subjects and also from 5 unseen subjects are used for testing. Table III shows a subset of the results, the quantity computed is the similarity between two feature vectors $S = 1/D_{WDTW}$, and Fig. 11 shows the ratios of correct queries found in the m best matches, for several values of m. We observe that when using 14 Eigen-coefficients, the performance of the system is high: a probability of 99% that the genuine user is the best match (value m=1, purple bar).

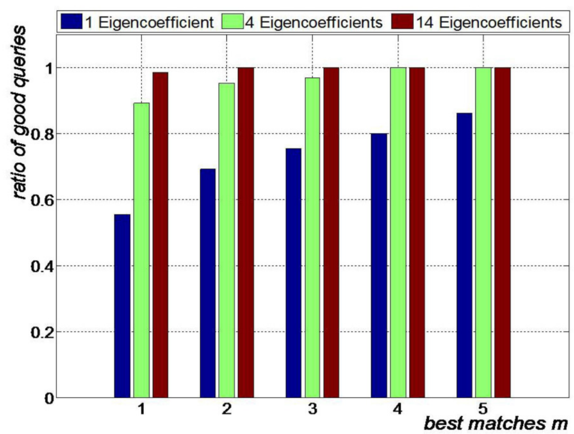


Fig. 11. Performance of the identification System: ratios of correct answers found in the top m best match, for several values of m.

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 (e.g. Laura, Melanie). For example, threshold $t = 2.00$ can be a suitable choice.

VII. CONCLUSION AND FUTURE WORK

In this study, we examine the feasibility of using 3D facial gestures for person recognition. A survey of related works in the field is conducted and new approaches are proposed for designing an accurate and user-friendly solution.

Our observations indicate that there exists a hierarchy in facial gesture reproducibility and distinctiveness, and speech postures are more functional than FACS-based face expressions in a real-life scenario. Moreover, verbal facial gestures offer a better potential because they can be easily combined with other biometrics such as voice and passwording in a multi-modal system in order to improve the security level. We also observe that pattern matching techniques derived from Dynamic Time Warping (DTW) perform very efficiently in the present context where we employ very short sequences of biometric features, and these latter exhibit noticeable fluctuations as it is the case for all behavioural traits. Our preliminary tests have shown promising results in terms of accuracy, usability and compactness.

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

However, much work still needs to be done. First, we will conduct further tests on a larger database and experiment various recognition system architectures. Then, we will study the use of facial gestures in conjunction with voice recognition and passwording in a multi-modal system.

VIII. ACKNOWLEDGMENTS

The authors wish to thank the Department of Dental Health, Cardiff University for sharing their face database.

REFERENCES

[1] M. Turk and A. Pentland. "Eigenfaces for Recognition", *Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86; 1991.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7); 1997.

[3] T. Cootes, G. Edwards, and C. Taylor. "Active appearance models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(6); 2001.

[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", *Computer Vision and Image Understanding*, vol. 101(1)pp. 115; 2006.

[5] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1063-1074; 2003.

[6] K. Chang, K. Bowyer, and P. Flynn. "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression", *IEEE Trans. on Pattern Analysis and Machine Intelligence* 28(10); 2006.

[7] J. Luettin, N.A. Thacher, and S.W. Beet. "Speaker identification by lipreading", *in Proc. of Intl Conf on Spoken Language* ; 1996.

[8] M. J. Roach, J. D. Brand and J. S. Mason. "Acoustic and Facial Features for Speaker Recognition", *International Conference on Pattern Recognition*, vol III: 258-261; 2000.

[9] M. I. Faraj and J. Bigun. "Audio-visual person authentication using lip-motion from orientation maps", *Pattern Recognition Letters* 28(11), pp. 1368-1382, 2007.

[10] E. Y. Zhang, S. J. Kundu, D. B. Goldgof, S. Sarkar and L. V. Tsap. "Elastic Face, An Anatomy-Based Biometrics Beyond Visible Cue", *International Conference on Pattern Recognition*, vol 2, 19-22; 2004.

[11] S. Pamudurthy, E. Guan, K. Mueller, and M. Rafailovich. "Dynamic approach for face recognition using digital image skin correlation", *Audio- and Video-Based Biometric Person Authentication*; 2005

[12] D. Fidaleo and M. Trivedi. "Manifold analysis of facial gestures for face recognition", *ACM SIGMM Multimedia Biometrics Methods and Application Workshop*, 2003.

[13] A. J. O'Toole, D. A. Roark Dana and H. Abdi. "Recognizing moving faces: A psychological and neural synthesis", *Trends in cognitive sciences* vol. 6, no6, pp. 261-266; 2002.

[14] J.F. Cohn and T. Kanade. "Cohn-Kanade AU-Coded Facial Expression Database", Pittsburgh University, 1999.

[15] P Ekman and W. Friesen. "The Facial Action Coding System: A Technique for the Measurement of Facial Action", *Consulting Psychologists*; 1978.

[16] S. Pigeon and L. Vandendrope. "The M2VTS Multimodal Face Database", *Proc. First International Conf. Audio- and Video-Based Biometric Person Authentication*; 1997.

[17] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. "XM2VTSDB: The extended M2VTS database", 2^{nd} *International Conference on Audio and Video-based Biometric Person Authentication*, Washington D.C., pp. 72-77, 1999.

[18] C. C. Chibelushi, S Gandon, J. S. Mason, F. Deravi, and D Johnston. "Design Issues for a Digital Integrated Audio-Visual Database", *IEE Colloquium on Integrated Audio-Visual Processing for Recognition, Synthesis and Communication*; 1996.

[19] D. Johnston , D. T. Millett and A. F. Ayoub. "Are Facial Expressions Reproducible?", *Cleft Palate-Craniofacial Journal*. 40(3):291-6; 2003.

[20] L. Benedikt, E. Krumhuber, A. Calvert, D. Cosker, P. Rosin and D. Marshall. "Using Dynamic 3D Facial Data to Create 3D Appearance Models of Facial Action Units", *Eurographics/SIGGRAPH Symposium on Computer Animation*; San Diego 2007.

[21] A. Efrat, Q. Fan and S. Venkatasubramanian. "Curve Matching, Time Warping, and Light Fields: New Algorithms for Computing Similarity between Curves", *Journal of Mathematical Imaging and Vision*; 2007.

[22] N. Dhananjaya. "Correlation-Based Similarity Between Signals for Speaker Verification with Limited Amount of Speech Data", *Multimedia Content Representation, Classification and Security*; 2006.

[23] H. Sakoe and S. Chiba. "Dynamic programming algorithm optimization for spoken word recognition", *IEEE Trans. Acoustics, Speech, and Signal Processing*; 1978.

[24] M.E. Munich and P. Perona. "Continuous Dynamic Time Warping for Translation-Invariant Curve Alignment with Applications to Signature Verification", *The Proceedings of the Seventh IEEE International Conference on Computer Vision*; 1999.

[25] E.J. Keogh and M.J. Pazzani. "Derivative dynamic time warping", *First SIAM International Conference on Data Mining*; 2001.

[26] L. Benedikt, D. Cosker, P. L. Rosin and D. Marshall. "Facial Dynamics in Biometric Identification", *In Proc. of British Machine Vision Conference (BMVC)*; 2008.

[27] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki. "An Introduction to Evaluating Biometric Systems", *IEEE Computer, February, pp. 56-63 (Special issue on biometrics)*; 2000.

[28] T. Hastie and W. Stuetzle. "Principal curves", *Journal of the American Statistical Association*, 84:502516; 1989.

[29] A. Jain, A. Ross, and S. Prabhakar. "An Introduction to Biometric Recognition". *IEEE Transactions on circuits and systems for video technology*, Vol. 14, No. 1; 2004.