IET Biometrics

Research Article

Study on using individual differences in facial expressions for a face recognition system immune to spoofing attacks

ISSN 2047-4938
Received on 13th September 2015
Revised on 8th February 2016
Accepted on 9th February 2016
doi: 10.1049/iet-bmt.2015.0078
www.ietdl.org

Mihai Gavrilescu ✉

Department of Telecommunications,' Politehnica' University of Bucharest, 1-3 Iuliu Maniu Blvd, 061071 Bucharest 6, Romania
✉ *E-mail: mike.gavrilescu@gmail.com*

Abstract: Although current face recognition systems in biometrics field are accurate enough to be used as substitutes for passwords or keys, most of them are prone to face spoofing attacks. Different techniques for face spoofing identification have been researched but most of them introduce additional sensors and are not cost or computationally efficient. In this study, the authors study the possibility of using individual differences in facial expressions for improving a face recognition system and make it immune to spoofing attacks. The authors develop a soft biometric neural-network-based system for video-based face recognition by analysing patterns in individual facial expressions on multiple frames. Results show that such a system is possible and has accuracies higher than 85%. Used alongside with a standard principal component analysis-based face recognition system, the combined method achieved 94.5% accuracy on Honda/UCSD Video Database and 92.9% on Youtube Faces DB, comparable with state-of-the-art. When tested against photo spoofing attacks on three public anti-spoofing databases the proposed method was immune. In terms of video spoofing, the error rate for the authors' proposed method was 1% surpassing state-of-the-art methods.

1 Introduction

The field of biometrics has a complex history in person recognition and most research revolve around using face for individual recognition. This research is based on two important ideas. The first one refers to the fact that it offers people an easy way of identifying themselves, without needing to remember password or possess keys. Second, this soft biometric system should be more accurate than other identification systems as facial features of an individual do not change, cannot be forgotten, or lost.

With the increase in reliability of such soft biometric systems now achieving over 95%, attention is turned to reduce the susceptibility of current face recognition methods to spoofing attacks. Compared with other biometric means of personal identification, face is easier to spoof using a photo of the actual face or, for more complex systems, a three-dimensional (3D) face model. Along with the increase of social media exposure for many people nowadays, exposing themselves in photographs happens often and such spoofing attacks are increasing. Various sensors can be used together with face recognition systems to defend against successful attacks, but these are costly and harder to use, hence other soft biometric approaches are needed.

Starting from the fact that facial expressions are ways humans communicate with others and react, for a computer to achieve human-like performance it will have to know what our feelings are and if are expressed by us and not someone else. This paper turns attention to facial expressions researching the possibility of using individual differences in facial expressions in a neural-network-based soft biometric face recognition system and investigating the improvements it brings to standard face recognition systems and if it makes them less susceptible to spoofing attacks.

2 Related work

Most research in face recognition aims recognising faces from images, but there are studies that also treat video-based face recognition systems. Huang *et al.* [1] discuss in the lack of adequate database for video-based face recognition, specifically in terms of datasets and comparisons. They propose a new dataset named COX Face DB containing large scale still and video face databases in three scenarios: video-to-still (V2S), still-to-video (S2V), and video-to-video. They also propose a new point-to-set correlation learning method for V2S/S2V face recognition and show it offers good accuracy. Li and Hua [2] discuss in the problem of pose variations which typically affects real-world face recognition system and propose a hierarchical-PEP model for decomposing the face image in face parts at different levels of dimensionality building pose invariant face representations. Verified on two public databases (including Youtube Faces DB [3]) it showed better results than state-of-the-art. Xilin Chen *et al*. [4] analyse several clues in a person's face, body, and motion for determining his identity and use video-dictionaries to describe pose, temporal, and illumination information. Their method showed up to 94.87% accuracy on Honda/UCSD dataset [5]. Taigman *et al*. [6] show in a 3D face modelling for piecewise transformation, face representation being divided in nine-layer deep neural-network. Tested on Youtube Faces DB, it showed human-like performances, surpassing state-of-the-art with an accuracy of 91.4%.

There are also various methods for fighting against spoofing attacks. Maatta *et al.* [7] show a way of increasing a system's reliability when facing printed faces spoofing attacks, making the supposition that printed face contains printing defects of poor quality, hence introducing texture and gradient analysis of facial image it can determine if the face is real or printed. Compared with state-of-the-art, the method offered promising results. Komulainen *et al.* [8] propose treating face spoofing as a set of attack-specific sub-problems that can be solved using a combination of complementary methods. Tested on NUAA Photograph Imposter DB [9] and CASIA Face Anti-Spoofing DB [10] a low error rate of 3% was obtained. Consequently, Bharadwaj *et al.* [11] presented another method for defeating spoofing attacks on printed images using Eulerian motion magnification, enhancing facial expressions of subjects from video frames. The method was tested on REPLAY-ATTACK DB [12] and offered promising error rate of only 1.25%.

The main issues in video-based face recognition are pose, expression and illumination invariance. Current paper considers the possibility of using individual facial expressions to improve video-based face recognition methods, specifically by resolving pose and expression invariance. A similar study was conducted in [13] where differences on human face are analysed in five local entities (eyes, nose, mouth, eyebrows, and rest of face) using a genetically-inspired learning method selecting meaningful facial features in a genetic programming-based approach. In contrast, current paper aims analysing a set of specific microexpressions using facial action coding system (FACS) developed by Ekman and Friesen [14] with the advantage that these microexpressions cannot be mimicked and were shown to provide authenticity for person identification [15]. Compared with the study in [13], we use pattern recognition techniques on a behavioural map (BM) containing facial evolutions of microexpressions in each frame, and this correlation we expect to translate in better accuracy. Also, because microexpressions cannot be mimicked, we test the proposed methods in spoofing attack cases verifying how well it deals in such scenarios. Theoretical details, proposed architecture, and experimental results are presented in following chapters.

3 Psychological grounds

As previously mentioned, we aim creating a reliable soft biometric face recognition system taking into account individual facial expressions. The method is based on FACS [14] stating that for each emotion a corresponding facial muscular activity is triggered and studying these muscle activities we can determine the exact emotion of an individual. Depending on facial muscles and how they are linked, FACS defines each expression as a sum of action units (AUs) that can be *additive* (triggering an AU depends on the trigger of another AU, the AU combination forming an AU cluster) or *non-additive* (triggering an AU is independent of other AUs).

FACS [14] identifies 44 AUs on face and neck muscles. However, not all AUs are effective in case of face recognition through individual facial expression analysis. We consider a more recent work of Cohn *et al.* [15] showing that individual differences in facial expressions are stable over time and offer the possibility of person identification. They collected video recorded interviews of 85 men and women emphasising the spontaneous facial expressions and their individual identification tests offered <80% accuracy. Their results, although showing facial behaviour alone cannot offer satisfactory results, could improve and offer more reliability and robustness when used alongside a face recognition system. Reliability comes from analysing microexpressions which are harder to mimic or capture as they are spreading on more consecutive frames, hence spoofing attacks using photos or 3D face models will be harder to succeed. In [15], a set of AUs that can be successfully used for person identification cases as they provide more stability over time is determined as below:

Eye: AU5 (*upper lid raiser*), AU7 (*lid tightener*)
Brow: AU1 (*inner brow raiser*), AU2 (*outer brow raiser*), AU4 (*brow lowerer*)
Cheek: AU6 (*cheek raiser*), AU14 (*dimpler*)
Mouth: AU10 (*upper lip raiser*), AU12 (*lip corner puller*), AU15 (*lip corner depressor*), AU18 (*lip puckerer*), AU20 (*lip stretcher*), AU24 (*lip pressor*)

We observe four types of AUs, each corresponding to different face areas and therefore we divide the face in four components and analyse each of them in particular as well as the interaction between them.

4 Proposed architecture

As previously shown, we analyse only AUs shown relevant in face recognition tasks, hence only 13 out of the 44 AUs. The architecture has two main blocks: individual expression recognition (IER) block and standard principal component analysis -based face recognition (PCA-FR) block. The novelty is the IER block, and this is typically divided in the following sub-blocks:

Face detection, face segmentation, AU detection, AU classification.
BM building.
Artificial neural-network (ANN) for pattern recognition.

Each of these sub-blocks is detailed in the following subchapters.

4.1 Face detection, face segmentation, AU detection, and AU classification

In literature, there are various algorithms employed for face detection, but the most used approach is the Viola–Jones method [17] which uses a specific algorithm to distinguish facial patterns from non-facial patterns, dividing the image in smaller and smaller areas until only the face is selected. Viola–Jones uses Haar-cascade classifiers fetched with the image, each classifier restricting more the area of the image until the last classifier returns the actual face. This method is fast and was successfully used in other research papers, such as that of Bartlett and coworkers [18] and Zhan *et al.* [19], as well as OpenCV library. Since we need facial detection performed fast and real-time with no hard computations, we use Viola–Jones object detection method and, as mentioned in [20], we establish a trade-off between the number of classifier stages (45 stages), number of features in each stage (10,000 features) and threshold for each stage. Viola–Jones algorithm offers a detection rate of 98% when tested on 1000 faces.

As previously described, face is segmented in four components and each of them is analysed separately as shown in Fig. 1. We have two symmetric components for brow, eye, and cheek and a single component for mouth. We design a multi-state face model with the below classifiers in order to determine the presence of the corresponding AUs, track the AU evolution frame by frame and compute the AU intensity level:

Mouth component: Lip feature point tracking method (as used in [21]).
Eye component: Haar-Cascade classifiers (as used in [22]).

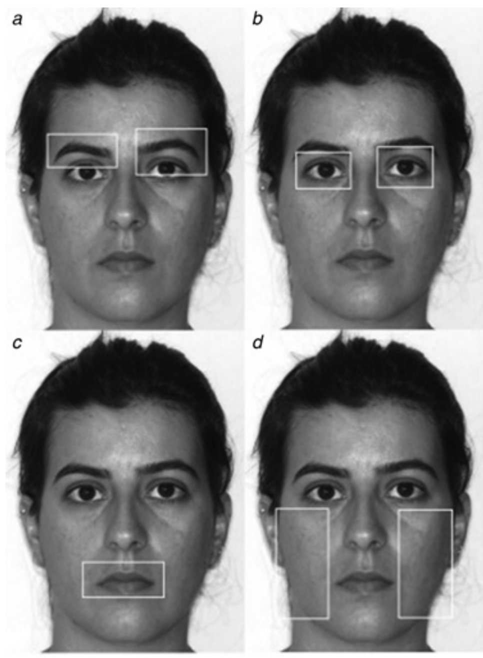

Fig. 1 *Face is segmented into four components and each of them is analysed separately as*

a Brow
b Eye
c Mouth
d Cheek

Brow component: Haar-Cascade classifiers (as used in [22]).
Cheek component: Modified Viola–Jones algorithm for cheek detection and tracking (as used in [23]).

For each component, classifiers employed determine the presence/absence of AUs as well as a level indicating how pronounced is the AU, as follows: A – trace (between 15 and 30 classification score), B – slight (30–50), C – marked and pronounced (50–75), D – severe or extreme (75–85), E – maximum (over 85), O – AU is not present (under 15).

In pre-processing step, each classifier is trained on Cohn–Kanade database [24] and tested on MMI database [25] and offers over 90% accuracy in cross-database test for each classified AU. We employ these two databases because they are widely used in facial expression recognition research and have samples with pose and illumination invariance. Tested on 30 positive sessions and 160 negative sessions, we obtained for each AU a classification rate (successfully classified samples divided by total number of samples) of over 94% and a precision (successfully classified samples divided by the sum of successfully classified positive samples and number of false positive) of over 92%.

Besides AU classification, we also determine the below distances:

Vertical distance between eyes (E) and brows (EB) (E–B distance): Used to determine in what degree brows are outlying eyes (useful for emotions like sadness or anger).
Vertical distance from mouth peripherals (MP) to eyes (E) (MP–E distance): For a sad face the distance from mouth corners to eyes is higher than for a happy face.
Vertical distance from cheek internal extremities (CIE) to mouth centre (MC) (CIE–MC distance): Used with previous distances to increase accuracy.

All these distances are determined using standard Euclidean distance algorithm.

4.2 BM building

For each of the 13 AUs classified we determined a classification score that is inputted in the BM alongside with the three distances. BM therefore contains a row for each frame, each row having the following structure: (A1A, A2B, A4A, A5E, A6O, A7O, etc., 0.3, 4.3, 0.6), where A1A represents that AU1 was classified with level A intensity, and 0.3, 4.3, and 0.6 represent the E–B, MP–E, and CIE–MC distances.

BM will then be fetched to next level where an ANN checks for patterns in BM and outputs if the face is recognised or not. The ANN is presented in next subchapter.

4.3 ANN for pattern recognition

The ANN used is a multi-layer perceptron with three layers: input layer, one hidden layer, and the output layer. We use as input 15 consecutive frames from the BM, knowing that microexpressions associated with AU movement typically last an average of 500 ms and also for avoiding ANN overfitting. We therefore have 240 input nodes (16 values from BM for each of the 15 frames) that are normalised in [0, 1] interval (the three distances are normalised by scaling between 0 and 1 and the levels have the following values: A – 0.2, B – 0.4, C – 0.6, D – 0.8, E – 0.9, and O – 0). The output layer contains only one node with a score in [0, 1] interval for the face recognition task.

If we denote $X=\{x_p\}$, $p=1,2\dots p$ a N-dimensional set of input vectors, such as $x=[x_1, x_2\dots x_N]^{\mathrm{T}}$, $Y=\{y_p\}$, $p=1, 2\dots p$ a K-dimensional set of output vectors such as $y=[y_1, y_2, \dots, y_k]^{\mathrm{T}}$, $\boldsymbol{W}^{\mathrm{H}}$ (matrix of weights between input and hidden nodes), $\boldsymbol{W}^{\mathrm{O}}$ (matrix of weights between hidden nodes and output nodes), and f_{1a} and f_{2a} activation functions, the expression form can be written as below (1)

$$y_k = f_{2a}\sum_{l=0}^{L} w_{lk}^{O} f_{1a}\left[\sum_{n=0}^{N} w_{nl}^{H} x_n\right] \quad \text{with } k=1,2\dots K \tag{1}$$

Since we are implementing a bottom-up architecture with no feedback loops we chose backpropagation as a training method, as it offers better performance, convergence and is faster in pattern recognition tasks [26]. Being a feed-forward neural-network, input data is fetched to input layer neurons and passed to the first hidden layer. Each hidden layer computes a weighted sum of its inputs and fetches the sum through its activation function to the output layer. The difference between the outputs and the expected values is used to determine the average absolute relative error (AARE) (2)

$$\mathrm{AARE} = \frac{1}{N}\sum_{n}^{1}\left|\left(\frac{y_p - y_e}{y_e}\right)\right| \tag{2}$$

After AARE is computed, the weight matrices $\boldsymbol{W}^{\mathrm{H}}$ and $\boldsymbol{W}^{\mathrm{O}}$ are tuned and the procedure is repeated for several iterations, until AARE is minimised.

Through trial and error method, we determined the optimal number of 270 nodes in the hidden layer in order to properly learn the pattern without high execution times. The activation function in input layer (f_{1a}) and the activation function of output layer (f_{2a}) determined as optimal were both log sigmoid functions ('logsig'). For learning the weights and biases of the neural network we used the gradient descent algorithm, for which we obtained the lowest AARE of 0.005. The optimal learning rate is 0.015 and the optimal momentum 0.03. We made use of 120,000 training epochs to avoid over-training, and it took ∼5 h to complete on an Intel i7 processor. In terms of initial weights, for distributing the active regions of each neuron in the layer evenly we used Nguyen–Widrow weights initialisation method.

4.4 Overall architecture

Overall architecture is presented in Fig. 2. We can observe that besides IER block, another PCA-FR system is implemented similarly as in [16] using eigenfaces where images must have the same size and be normalised to line up with the eyes and mouth of the subjects. We do not use a more complex method because we want to determine the improvement brought by IER method to standard PCA-FR system.

The image is first fetched to the face detection block responsible for detecting the face via Viola–Jones method. The detected face features are sent in parallel to IER block and the standard PCA-FR block.

In IER block, the detected face is segmented in four components and each AU present on the four components is detected using its corresponding classifier trained to offer over 90% accuracy. Each classifier outputs the presence/absence and intensity level of the AU for each frame and this information is included as a row in BM. Similarly, E–B, MP–E, and CIE–MC distances are calculated and added at the end of each row in BM. A visualisation example on how BM is built is shown in Fig. 3. We can see for each of the six frames AUs presence and intensity levels alongside with distances being determined and inserted in BM, which will contain a row for each frame.

When BM contains 15 rows (corresponding to the first 15 frames), they will be fetched to the ANN previously trained via backpropagation which will output a score for the analysed face. When a new row appears in BM, last 14 rows (fetched in the previous step to the ANN) and this new row are fetched to the ANN which will again output a score; hence for each frame starting with the 15th frame a score is computed by the ANN and fetched to the decisional layer. Decisional layer is a decisional tree receiving input from standard PCA-FR and IER blocks and outputting results for face recognition task.

The testbed was implemented in C++ with OpenCV library, Intel i7 processor, and 8 GB of RAM memory. Video frames were considered at 30 FPS.

IET Biom., 2016, Vol. 5, Iss. 3, pp. 236–242

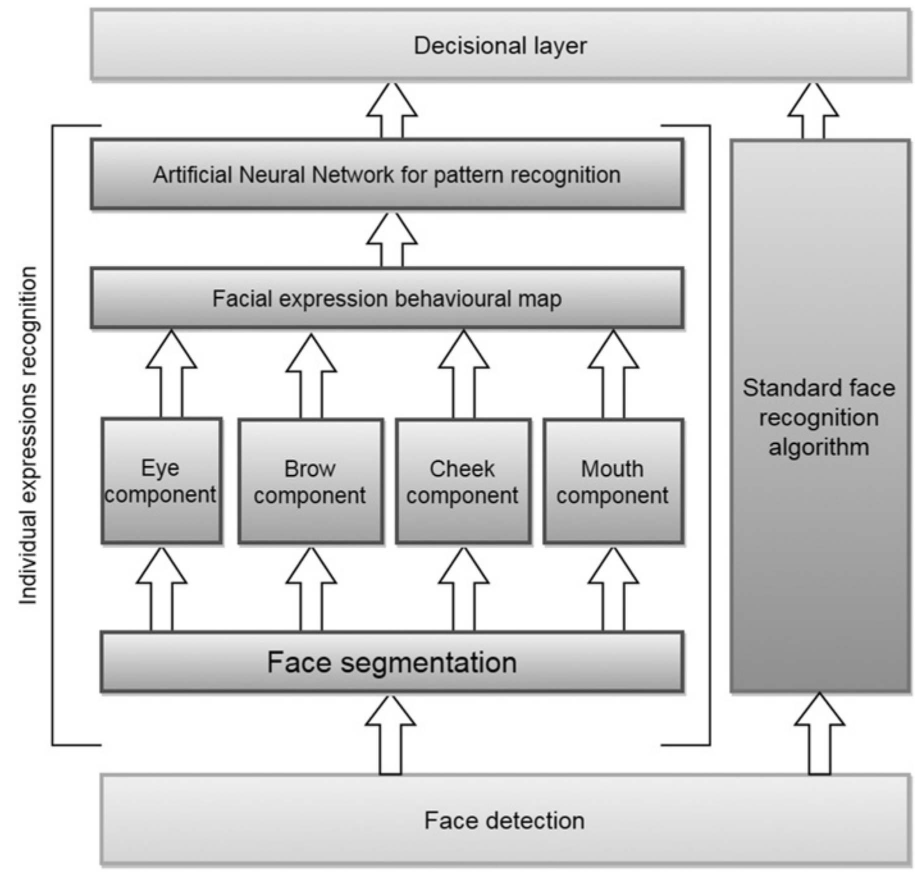

Fig. 2 *Overall architecture*

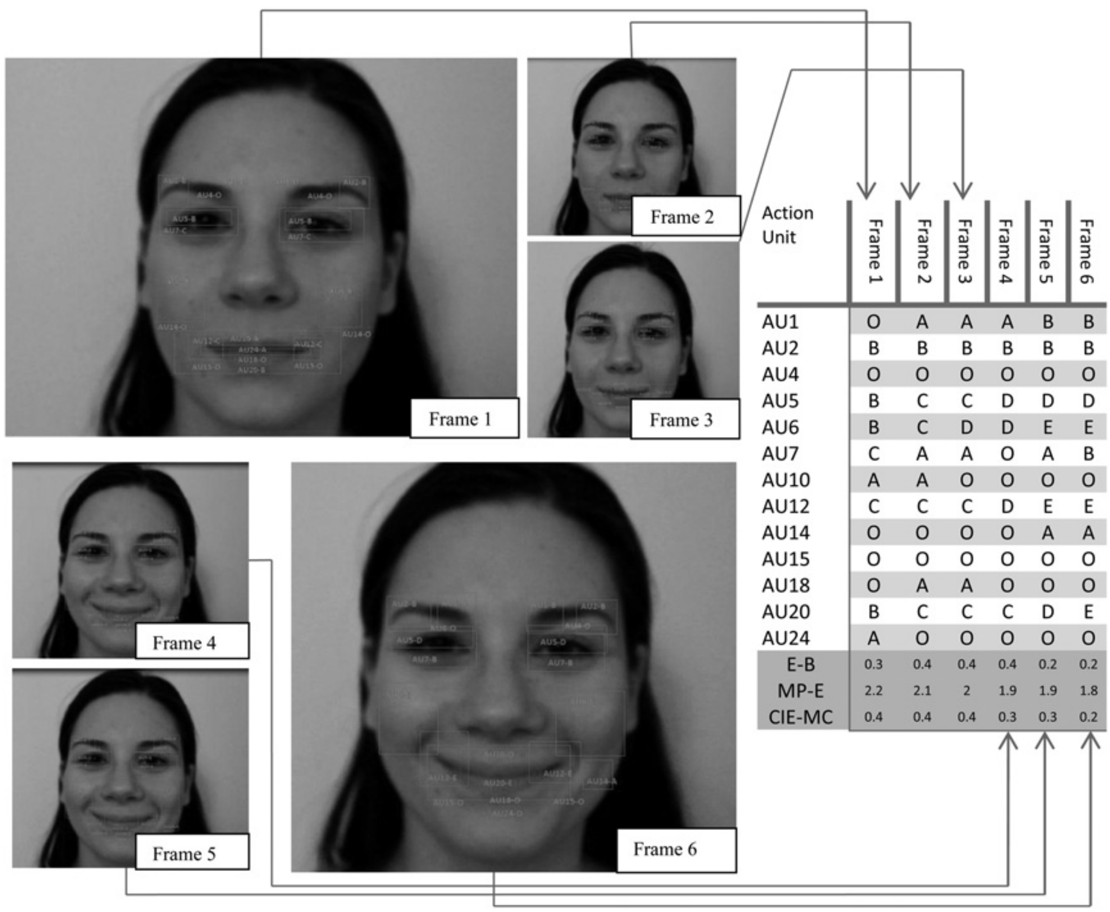

Fig. 3 *BM building*

5 Experimental results

For testing the proposed architecture, we conducted three tests: face recognition on our own dataset, face recognition on standard datasets, and spoofing attack tests. All are detailed in the following subchapters.

5.1 Own dataset tests

We conducted this test on our own database for checking the accuracy of IER block when used standalone and when used with the standard PCA-FR system, and also to see for which emotion face recognition is more accurate. For this, we built our own database containing recording of frontal facial expressions of 64 subjects when watching four emotion inducing videos for four emotions (sadness, fear, happiness, and anger). Each subject was recorded five times for each of four emotions and ten times in random scenario (no emotion induced). Subjects involved in experiment were 32 males and 32 females with ages between 18 and 35, participating in accordance with Helsinki Ethical Declaration. Emotion inducing videos were collected from LIRIS-ACCEDE database [27]. Since our system needs longer videos for training and testing, we combined 8–12 s videos for same emotion in 1 min emotion-inducing video. All tests conducted were performed in an intra-subject methodology. ANN was trained on 20 reactions collected for one of the 64 subjects when emotions were induced, and tested on the remaining 10 reactions in random scenarios for same subject and also on videos from other subjects as negative samples. Test was repeated for all 64 subjects.

First, we wanted to determine person recognition rates when only IER block is used, hence we are identifying individuals solely based on facial expressions. Results are presented in Table 1. In 110 tests with 10 positives (tested subject is the trained subject) precision in recognition was 91.7% and it took 6 s on average to recognise the person. Increasing the number of negatives (tested subject is different from the trained subject) recognition rate drops by 1% and average time increases with 5 s. This can be explained by the negative tests which are introducing more computational work trying to find patterns in negative faces.

We conducted a new test by stopping the clock at 6 s, considering that any computation lasting more is a negative one. Results improved, hence for the proposed method 6 s is the optimal time-out before accepting a face as positive. This is caused by the ANN used for pattern recognition and shows that here is room for improvement; maybe by considering other types of pattern recognition methods.

Next we tested the accuracy of IER block when certain emotion is induced. We trained the system on 16 of the recorded reactions of a subject when watching emotional inducing video and tested on remaining 4 as positive samples and samples from other subject as negative samples. Results are detailed in Table 2.

We observed that when certain emotions are induced, recognition rate increases with 7% for happiness, 6% for sadness, and 4% for fear and anger. Highest precision improvement is observed for happiness, showing that smile can be very good for person identification via IER. Average time taken in case of happiness was 13 s. Moreover, no test performed has recognition time higher than 17 s.

Third test was to verify the improvements brought by the IER block to the standard PCA-FR system. We trained the system in both cases on 20 videos collected for one of the 64 subjects where

Table 1 IER precision

Number of positives/ number of negatives	Precision, %	Average time, s	Precision, % – clock stopped at 6 s
10/100	91.7	6	91.7
10/200	88.3	11	90.2
50/1000	86.5	8	88.5

Table 2 IER precision – user was emotionally stimulated

Emotion induced by the video clip	Precision, %	Average time, s
sadness	97.2	15
happiness	98.9	13
fear	95.3	16
anger	95.2	17

Table 3 IER + PCA-based FR precision

Number of positives/ number of negatives	Precision, % – only PCA-based FR	Precision, % – IER and PCA-based FR
10/100	90.4	99
10/200	90.2	98.6
50/1000	89.6	97.5

Table 4 IER + PCA-based FR precision – user emotionally stimulated

Emotion induced by the video clip	Precision, % – only PCA-based FR	Precision, % – IER and PCA-based FR	Average time, s
sadness	90.4	98.8	17
happiness	91.3	99.9	15
fear	90.2	97.5	18
anger	89.7	96.5	20

emotions were induced and tested on remaining 10 collected in random scenarios alongside with videos from other subjects. Time-out was kept at 6 s. Results are detailed in Table 3 and we observed that, by adding the IER block, accuracy improved with up to 8%.

Last test performed was testing the entire system on a video recorded reaction of the subject when stimulated with certain emotions. Results are detailed in Table 4 and we see that recognition rate increased with up to 7%, achieving a top 99.9% accuracy when subject is stimulated with happiness emotion.

5.2 Standard datasets tests

For testing the accuracy of our video-based face recognition system and compare it with state-of-the-art we chose two databases typically used for testing in such cases: Honda/UCSD Video Database and Youtube Faces DB.

Honda/UCSD Video Database [5] is a standard database used for evaluating face tracking/recognition algorithms. Since, our method deals with facial expressions affected by pose changes, testing the method on this database will show the accuracy of the system and how well the tracking methods employed are dealing with pose invariation. Honda/UCSD Video Database contains three subsets: Training, Testing, and Occlusion Testing. We trained our system in an intra-subject methodology on the Training subset and tested on the other two subsets and repeated this for all 20 subjects. We obtained 94.5% accuracy when tested on Testing subset and 90.3% accuracy when tested on Occlusion Testing subset. The good results on Occlusion Testing subset are explained by the symmetry of the facial expressions, as our method segments face and it is only necessary to detect an eye, a brow, or a cheek to recognise a face. A comparison with state-of-the-art is presented in Table 5 and we observe that our method offers comparable results with state-of-the-art DFRV-f and KDFRV-f [33] methods.

We continued testing the IER+(PCA-FR) system on the *Youtube Faces DB* [3] as it is a larger database with illumination and pose invariance. Youtube Faces DB contains 3.425 videos of 1.595 different people, an average of 2.15 videos per subject. We train our system with intra-subject methodology on subjects appearing in more than 100 frames and test face recognition on remaining video samples. Results are detailed in Table 6 and we observe that

IET Biom., 2016, Vol. 5, Iss. 3, pp. 236–242

Table 5 Honda/UCSD video database – state-of-the-art

Reference/Year	Method	Accuracy
[28]/2010	Discriminant-analysis of canonical correlations (DCC)	85.47
[29]/2013	Manifold-Manifold distance (MMD)	83.76
[30]/2009	Manifold discriminant analysis (MDA)	88.89
[31]/2006	Affine hull based image set distance (AHISD)	87.18
[31]/2006	Convex hull based image set distance (CHISD)	86.33
[32]/2008	Sparse approximated nearest point (SANP)	92.31
[33]/2015	baseline (no DL)	88.89
[33]/2015	DFRV-f	94.87
[33]/2015	KDFRV-f	94.87
current work/2015	(IER with ANN) + PCA	94.5

Table 6 Youtube faces DB – state-of-the-art

Reference/Year	Method	Accuracy
[34]/2011	Min dist, FPLBP	65.6 ± 1.8
[34]/2011	Min dist, LBP	65.7 ± 1.7
[34]/2011	‖U1'U2‖, FPLBP	64.3 ± 1.6
[34]/2011	‖U1'U2‖, LBP	65.4 ± 2.0
[34]/2011	MBGS L2 mean, FPLBP	72.6 ± 2.0
[34]/2011	MBGS L2 mean, LBP	76.4 ± 1.8
[35]/2013	MBGS + SVM-	78.9 ± 1.9
[36]/2013	APEM-FUSION	79.1 ± 1.5
[4]/2013	STFRD + PMML	79.5 ± 2.5
[37]/2013	VSOF + OSS (Adaboost)	79.7 ± 1.8
[6]/2014	DeepFace-single	91.4 ± 1.1
[3]/2014	DDML (LBP)	81.3 ± 1.6
[3]/2014	DDML (combined)	82.3 ± 1.5
[38]/2014	EigenPEP	84.8 ± 1.4
[39]/2014	LM3L	81.3 ± 1.2
current work/2015	(IER with ANN) + PCA	92.9 ± 1.2

our method surpasses DeepFace-single [6] method with almost 1.5%, achieving a 92.9% accuracy.

5.3 Face spoofing attacks

In terms of face spoofing, we tested the system on REPLAY-ATTACK DB [12], NUAA Photograph Imposter DB [9] and CASIA Face Anti-Spoofing DB [10]. For printed photo spoofing attack for tests conducted on all 3 databases, we have not seen any successful spoofing, emphasising that the proposed method is photo spoofing immune. For video spoofing, we obtained an average error rate on the three databases of 1% (0.5% on REPLAY-ATTACK DB, 1.2% on NUAA Photograph Imposter DB, and 1.3% on CASIA Face Anti-Spoofing DB) surpassing state-of-the-art as seen in Table 7.

Table 7 Spoofing attacks – state-of-the-art

Reference/ Year	Method	Type of Attack	Database	Error, %
[40]/2012	face texture using LBPs	photo/ video	REPLAY-ATTACK DB, NUAA photograph imposter DB, CASIA face anti-spoofing DB	15
[41]/2012	face texture using difference of Gaussians (DoG) filters	photo/ video	CASIA face anti-spoofing DB	15
[8]/2013	context-based using upper body, spoof support detection	photo/ video	NUAA photograph imposter DB, CASIA face anti-spoofing DB	3
[11]/2013	motion detection using histogram of oriented optical flows (HOOF) after Eulerian magnification	photo/ video	REPLAY-ATTACK DB	1.25
current work/2015	(IER with ANN) + PCA	photo/ video	REPLAY-ATTACK DB, NUAA photograph imposter DB, CASIA face anti-spoofing DB	1

6 Conclusions

We implemented a multi-state neural-network-based system for face recognition based on individual facial expressions. We used FACS knowing that FACS microexpressions have the advantage of being harder to mimic and we analysed frame by frame changes in 13 AUs present on the face alongside with three Euclidean distances between certain areas of the face. We showed that individual facial expressions indeed offer information for identifying users, obtaining more than 86% accuracy on our own dataset. Used together with standard PCA-based face recognition system, the method showed comparable results with state-of-the-art, achieving 94.5% accuracy on Honda/UCSD Video Database and 92.9% on Youtube Faces DB. Tested against photo spoofing attacks on three public databases no attack was successful, showing the proposed method is immune to such attacks. For video spoofing, error rate for our proposed method was 1% surpassing state-of-the-art. When tested on a subject's reaction when watching happiness inducing videos, the proposed method offered close to 99% accuracy, hence happiness can successfully be used for face recognition with extremely high accuracy. As a drawback, we have seen that for IER, the ANN needs more than 6 s to identify a pattern, results provided after 6 s being inaccurate. For this, we should consider combining other pattern recognition methods as well as better AU classification to improve the recognition time.

7 References

1 Huang, Z., Shan, S., Wang, R., *et al.*: 'A benchmark and comparative study of video-based face recognition on COX face database', *IEEE Trans.Image Process.*, 2015, **24**, (12), pp. 5967–5981
2 Li, H., Hua, G.: 'Hierarchical-PEP model for real-world face recognition'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2015, 7–12 June 2015, pp. 4055–4064
3 Hu, J., Lu, J., Tan, Y.-P.: 'Discriminative deep metric learning for face verification in the wild'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1875–1882
4 Cui, Z., Li, W., Xu, D., *et al.*: 'Fusing robust face region descriptors via multiple metric learning for face recognition in the wild'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3554–3562
5 Goh, R., Liu, L., Liu, X., *et al.*: 'The CMU face in action (FIA) database', *Lect. Notes Comput. Sci.*, 2005, **3723**, pp. 255–263
6 Taigman, Y., Yang, M., Ranzato, M., *et al.*: 'DeepFace: closing the gap to human-level performance in face verification'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1701–1708
7 Määttä, J., Hadid, A., Pietikainen, M.: 'Face spoofing detection from single images using texture and local shape analysis', *IET Biometrics*, 2012, **1**, (1), pp. 3–10
8 Komulainen, J., Hadid, A., Pietikainen, M.: 'Context based face anti-spoofing'. Proc. of IEEE Sixth Int. Conf. on Biometrics, Theory, Applications and Systems (BTAS), September 2013, pp. 1–8
9 Tan, X., Li, Y., Liu, J., *et al.*: 'Face liveness detection from a single image with sparse low rank bilinear discriminative model'. Proc. Eur. Conf. on Computer Vision (ECCV), 2010, vol. 6316, pp. 504–517
10 Zhang, Z., Yan, J., Liu, S., *et al.*: 'A face antispoofing database with diverse attacks'. Proc. IAPR Int. Conf. on Biometrics (ICB), March/April 2012, pp. 26–31
11 Bharadwaj, S., Dhamecha, T.I., Vatsa, M., *et al.*: 'Computationally efficient face spoofing detection with motion magnification'. Proc. IEEE Int. Conf. on Computer Vision Pattern Recognition Workshops (CVPRW), June 2013, pp. 105–110
12 Chingovska, I., Yang, J., Lei, Z., *et al.*: 'The 2nd competition on counter measures to 2D face spoofing attacks'. Proc. IAPR Int. Conf. on Biometrics (ICB), June 2013, pp. 1–6
13 Bhanu, B., Yu, J.: 'Super-resolution and facial expressions for face recognition in video' (ACM Digital Library), ISBN: 978-0-549-41669-2
14 Ekman, P., Friesen, W.V.: 'Facial action coding system: investigator's guide' (Consulting Psychologists Press, Palo Alto, CA, 1978)
15 Cohn, J.F., Schmidt, K., Gross, R., *et al.*: 'Individual differences in facial expression: stability over time, relation to self-reported emotion, and ability to inform person identification'. Fourth IEEE Int. Conf. on Multimodal Interfaces, 2002, pp. 491–496

16 Martinez, A.M., Kak, A.C.: 'PCA versus LDA', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (2), pp. 228–233
17 Viola, P., Jones, M.: 'Robust real-time object detection'. Technical report, University of Cambridge, 2001
18 Littlewort, G., Bartlett, M., Fasel, C., *et al.*: 'Towards social robots: automatic evaluation of human-robot interaction by face detection and expression classification'. The Annual Conf. on Neural Information Processing Systems (NIPS), 2013
19 Zhan, P.O.C., Li, W., Safaei, F.: 'Facial expression recognition for multiplayer online games'. Proc. Third Australasian Conf. on Interactive Entertainment, 2006, pp. 452–458
20 Viola, P., Jones, M.: 'Rapid object detection using a boosted cascade of simple features'. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2001, pp. 511–518
21 Lien, J.J.-J., Kanade, T., Cohn, J.F., *et al.*: 'Detection, tracking, and classification of action units in facial expression', *J. Robot. Auton. Syst.*, 1999, **31**, (3), pp. 131–146
22 Wilson, P.I., Fernandez, J.: 'Facial feature detection using Haar classifiers', *J. Comput. Sci. Colleges*, 2006, **21**, (4), pp. 127–133
23 Valstar, M., Pantic, M.: 'Fully automatic facial action unit detection and temporal phases of factial actions', *IEEE Trans. Syst. Man Cybern.*, 2012, **42**, (1), pp. 149–149
24 Tian, Y., Kanade, T., Cohn, J.F.: 'Recongnizing action units for facial expression analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (2), pp. 97–115
25 Pantic, M., Valstar, M.F., Rademaker, R., *et al.*: 'Web-based database for facial expression analysis'. Proc. IEEE Int. Conf. on Multimedia and Expo, 2005
26 Xiaoyuang, L., Bin, Q., Lu, W.: 'A new improved BP neural network algorithm'. Second Int. Conf. on Intelligent Computation Technology and Automation, October 2009, pp. 19–22
27 Baveye, Y., Dellandrea, E., Chamaret, C., *et al.*: 'LIRIS-ACCEDE: a video database for affective content analysis', *IEEE Trans. Affective Comput.*, 2015, **6**, (1), pp. 43–55
28 Cevikalp, H., Triggs, B.: 'Face recognition based on image sets'. Proc. IEEE Conf. on Computer Vision Pattern Recognition, June 2010, pp. 2567–2573
29 Lu, J., Wang, G., Moulin, P.: 'Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning'. IEEE Int. Conf. on Computer Vision (ICCV), December 2013, pp. 329–336
30 Wang, R., Chen, X.: 'Manifold discriminant analysis'. Proc. IEEE Conf. on Computer Vision Pattern Recognition, June 2009, pp. 429–436
31 Fan, W., Yeung, D.-Y.: 'Locally linear models on face appearance manifolds with application to dual-subspace based classification'. Proc. IEEE Computer Society Conf. on Computer Vision Pattern Recognition, 2006, vol. 2, pp. 1384–1390
32 Wang, R., Shan, S., Chen, X., *et al.*: 'Manifold-manifold distance with application to face recognition based on image set'. Proc. IEEE Conf. on Computer Vision Pattern Recognition, June 2008, pp. 1–8
33 Chen, Y.-C., Patel, V.M., Phillips, P.J., *et al.*: 'Dictionary-based face and person recognition from unconstrained video', *IEEE Access*, 2015, **3**, pp. 1783–1798
34 Wolf, L., Hassner, T., Maoz, I.: 'Face recognition in unconstrained videos with matched background similarity'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 529–534
35 Wolf, L., Levy, N.: 'The SVM-minus similarity score for video face recognition'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3523–3530
36 Li, H., Hua, G., Lin, Z., *et al.*: 'Probabilistic elastic matching for pose variant face verification'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3499–3506
37 Mendez-Vazquez, H., Martinez-Diaz, Y., Chai, Z.: 'Volume structured ordinal features with background similarity measure for video face recognition'. Int. Conf. on Biometrics (ICB), 2013, pp. 1–6
38 Li, H., Hua, G., Shen, X., *et al.*: 'Eigen-PEP for video face recognition'. The 12th Asian Conf. on Computer Vision (ACCV), 2014, pp. 17–33
39 Hu, J., Lu, J., Yuan, J., *et al.*: 'Large margin multi-metric learning for face and kinship verification in the wild'. The 12th Asian Conf. on Computer Vision (ACCV), 2014, pp. 252–267
40 Chingovska, I., Anjos, A., Marcel, S.: 'On the effectiveness of local binary patterns in face anti-spoofing'. Proc. IEE Int. Conf. on Biometrics Special Interest Group (BIOSIG), September 2012, pp. 1–7
41 Zhang, Z., Yan, J., Liu, S., *et al.*: 'A face anti-spoofing database with diverse attacks'. Proc. IAPR Int. Conf. on Biometrics (ICB), March 2012, pp. 26–31