Subject Identification Across large expression variations using 3D facial landmarks

SK Rahatul Jannat, Diego Fabiano, Shaun Canavan and Tempestt Neal

University of South Florida, Tampa FL, 33620, USA
{jannat, dfabiano, scanavan, tjneal}@usf.edu

Abstract. In this work, we propose to use 3D facial landmarks for the task of subject identification, over a range of expressed emotion. Landmarks are detected, using a Temporal Deformable Shape Model and used to train a Support Vector Machine (SVM), Random Forest (RF), and Long Short-term Memory (LSTM) neural network for subject identification. As we are interested in subject identification with large variations in expression, we conducted experiments on 3 emotion-based databases, namely the BU-4DFE, BP4D, and BP4D+ 3D/4D face databases. We show that our proposed method outperforms current state of the art methods for subject identification on BU-4DFE and BP4D. To the best of our knowledge, this is the first work to investigate subject identification on the BP4D+, resulting in a baseline for the community.

1 Introduction

Broadly, face recognition can be categorized as holistic, hybrid matching, or feature-based [38]. Holistic approaches look at the global similarity of the face such as a 3D morphable model (3DMM) [2]; hybrid matching make use of either multiple methods [14] or multiple modalities [17]; feature-based methods look at local features of the face to find similarities [40]. The work proposed in this paper can be categorized as feature-based. Due to its non-intrusive nature and wide applicability in security and defense related fields, face recognition has been actively researched by many groups in recent decades.

Since some of the earlier methods for face recognition [31], [37], to more recent works within the past 10 years [7], [35] 2D face recognition has been an actively researched field. With the recent advances in deep neural networks, we have seen significant jumps in performance [12], [18], [22], [24], [28], [33]. Liu et al. [21] proposed the angular softmax that allows convolutional neural networks (CNN) the ability to learn angularly discriminative features. This was proposed to handle the problem where face features are shown to have a smaller intra-class distance compared to inter-class distance. Recently, Tuan et al. [30] proposed regressing 3D morphable model shape and texture parameters from a 2D image using a CNN. Using this approach, they were able to obtain a sufficient amount of training data for their network showing promising results. Zhu et al. [41] proposed a high-fidelity pose and expression normalization method that made use of a 3DMM to generate natural, frontal facing, neutral face images. Using this

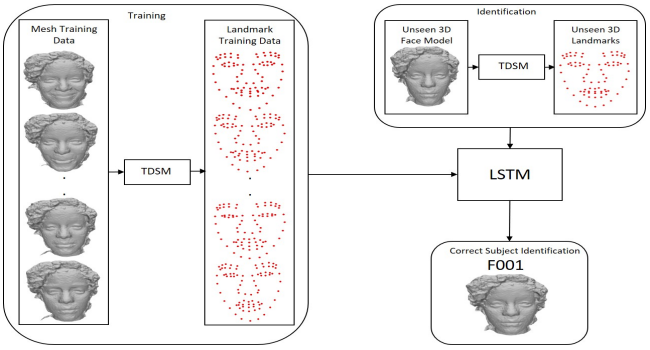

Fig. 1. Overview of proposed method. Example is showing an unseen 3D mesh model of subject 'F001' from BP4D+ [39], who is correctly identified based on training a LSTM [13] from 3D facial data detected from a TDSM.

method, they achieved promising results in both constrained and unconstrained environments (i.e. wild settings). Although performance has been increasing and groups have been actively working on 2D subject identification, there are still some challenges such as pose and lighting. 3D faces can help to minimize these challenges [25], and in recent years, this research has made significant strides [11], [12], [26] due to the development of powerful, high-fidelity 3D sensors.

Echeagaray-Patron et al. [11] proposed a method for 3D face recognition where conformal mapping is used to map the original face surfaces onto a Riemannian manifold. From the conformal and isometric invariants that they compute, comparisons are then made. This method was shown to have invariance to both expression and pose. Lei et al. [20] proposed the Angular Radial Signature for 3D face recognition. This signature is extracted from the semi-rigid regions of the face, followed by mid-level features being extracted from the signature by Kernel Principal Component Analysis. These features were then used to train a support vector machine showing promising results when comparing neutral vs. non-neutral faces. Berretti et al. [1] proposed the use of 3D Weighted Walkthroughs with iso-geodesic facial strips for the task of 3D face recognition. They achieved promising results on the FRGC v2.0 [23] and SHREC08 [10] 3D facial datasets. Using multistage hybrid alignment algorithms and an annotated face model, Kakadiaris et al. [15] used a deformable model framework to show robustness to facial expressions when performing 3D face recognition.

Motivated by the above works, we propose to use 3D facial landmarks for subject identification across large variations in expression. We track the facial landmarks using a Temporal Deformable Shape Model (TDSM) [6]. See Fig. 1 for an overview of the proposed approach. The rest of the paper is organized as follows. Section 2 gives a brief overview of the TDSM algorithm, Section 3 details our experimental design and results, and we conclude in Section 4.

2 Temporal Deformable Shape Model

The Temporal Deformable Shape Model (TDSM) models the shape variation of 3D facial data. Given a sequence of data (i.e. 4D), it also models the implicit constraints on shape that are imposed (e.g. small changes in motion and shape). To construct a TDSM, a training set of 3D facial landmarks is required. First, the 3D facial landmarks are aligned using a modified version of Procrustes analysis [5]. Given a training set of size L 3D faces, where each face has N facial landmarks (aligned with Procrustes analysis), a parameterized model S is constructed, $S = F_1^1, ..., F_N^1, ..., F_1^m, ..., F_N^m$. F_i^m is the i^{th} landmarks of the m^{th} 3D face in the training set, where $F_i^m = (x_i^m, y_i^m, z_i^m)$ and $1 \leq m \leq L$. From this model, principal component analysis (PCA), is then applied to learn the modes of variation, V, of the training data.

Given the parameterized model, S, and the modes of variation, V, to detect 3D facial landmarks, an offline weight vector, w, is constructed that allows for new face shapes to be constructed, by a linear combination of landmarks as $S = \bar{s} + Vw$ where $\bar{s}$ is the average face shape. These constructed face shapes are constrained to be within the range $-2\sqrt{\lambda_i} \leq w_i \leq 2\sqrt{\lambda_i}$, where w_i is the i^{th} weight in the range, and λ_i is the i^{th} eigenvalue from PCA. This constraint is imposed to make sure the new face shape is a 3D face.

To fit (i.e. detect landmarks) to a new input mesh, an offline table of weights (w) is constructed with a uniform amount of variance. The Procrustes distance, D, is then computed between each face shape (referred to as an instance of the TDSM) and the new input mesh. The smallest distance is considered the best detected landmarks. Note that this is not meant to be an exhaustive overview of a TDSM, therefore we refer the reader to the original work [6] for more details.

3 Experimental Design and Results

Using a TDSM, we detected 83 facial landmarks on 3 publicly available 3D emotion-based face databases: BU4DFE [34], BP4D [36], and BP4D+ [39]. From these facial landmarks, we then conducted subject identification experiments, where the landmarks are used as training data for 3 machine learning classifiers. Using these 83 facial landmarks we have also reduced the dimensionality of the 3D faces from over 30,000 3D vertices, while still retaining important features for subject identification. This allows us to reduce storage requirements, as well as processing time of the 3D face, which can be limitations of 3D face recognition [3], [16]. An overview of the databases and the experimental design is detailed in the following subsections.

3.1 3D face databases

One of the main goals of this work is to show subject identification across large variations in expression. Considering this, we evaluated 3 large, state-of-the-art 3D emotion-based face databases with a total of 282 subjects across all 3.

BU-4DFE [34]: Consists of 101 subjects displaying 6 prototypical facial expressions plus neutral. The dataset has 58 females and 43 males, including a variety of racial ancestries. The age range of the BU-4DFE is 18-45 years of age.

BP4D [36]: Consists of 41 subjects displaying 8 expressions plus neutral. It consists of 23 females and 18 males; 11 Asian, 4 Hispanic, 6 African-American, and 20 Euro-American ethnicities are represented. The age range of the BP4D is 18-29 years of age. This database was developed to explore spatiotemporal features in facial expressions. Due to its large variation in expression, it is a natural fit for our subject identification study.

BP4D+ [39]: Consists of 140 subjects (82 females and 58 males) ages 18-66. This data corpus consists of ethnic and racial ancestries that include African American, Caucasian, and Asian each with highly varied emotions. These emotions are elicited through tasks designed to elicit dynamic emotions in the subjects such as disgust, sadness, pain, and surprise resulting in a challenging dataset. Like the BP4D database, this dataset was also designed to study emotion classification. Its diversity and number of subjects, as well as large variations in expressions, make it a natural fit for our study.

3.2 Experimental design

To conduct our experiments, we detected 83 facial landmarks on the 3D data using a TDSM. Given 3D facial landmarks, we then translated them so that the centroid of the face is located at the origin in 3D space to align the data. The translated 3D facial features were then used for subject identification. Each of the 3D facial landmarks (x, y, z coordinates) are inserted into a new feature vector. For all 83 landmarks, this gives us a feature vector of size $83 \times 3 = 249$. This feature vector is used to train classifiers for subject identification. To ensure our results were not classifier specific, we trained a support vector machine (SVM) [32], random forest (RF) [4], and Long short-term memory (LSTM) neural network [13]. Our network consists of one short-term memory layer with a look back of two faces (estimated landmarks), followed by 0.5 dropout, and a fully connected layer for classification. The softmax activation function was used, along with the RMSprop [29] optimizer with a learning rate of 0.0001.

For each classifier, each subject's identity was used as the class (each 3D face is labeled with a subject id). Accurate results on an SVM, RF, and LSTM show the robustness of the 3D facial landmarks to multiple machine learning classifiers. We conducted one-to-many subjection identification, where all subjects were in both the training and testing sets. These sets were split based on time (i.e. different sections of the sequences available in the datasets) so consecutive (i.e., similar) frames did not appear in both sets.

3.3 Subject identification results

We achieved an average subject identification accuracy of 99.9%, on random forest and support vector machine, and 99.93% for an LSTM, across all databases. As can be seen in Table 1, an SVM, RF, and LSTM can accurately identify

Table 1. Subject identification accuracies for the 3 tested datasets and classifiers.

	BU4DFE	BP4D	BP4D+
SVM	99.9%	99.9%	99.9%
RF	100%	99.9%	99.8%
LSTM	100%	99.9%	99.9%

Table 2. Subject identification accuracies(percentage) for faces with simulated occlusion. Key: TR: Top Right; TL: Top Left; LR: Lower Right; LL: Lower Left.

	BP4D				BP4D+			
	TR	TL	LR	LL	TR	TL	LR	LL
RF	**99.7**	**99.7**	**99.7**	**99.7**	**99.3**	**99.3**	**99.6**	**99.5**
SVM	95.1	96.8	93.4	87.5	98.8	99.1	97.5	94.8

subjects from the BU4DFE, BP4D, and BP4D+ datasets achieving a max accuracy of 100% on BU4DFE, and a minimum accuracy of 99.8% on BP4D+. All three of the tested classifiers achieved consistent results across all three datasets, showing these results are not classifier dependent. As each of the datasets contain large variations in expression, these results show the detected 3D landmarks have robustness to expression changes for the task of subject identification.

3.4 Subject identification with occluded faces

Along with subject identification using all 83 landmarks, we also tested on a smaller number of facial landmarks to simulate occluded faces. For these experiments, we split the 3D facial landmarks (i.e. face) into 4 quadrants (Fig. 2) and detected a smaller number of landmarks (top right: 23; top left:23; lower right: 20; lower left: 17) using a TDSM. We then ran the same experiments for each quadrant. As shown in Section 3, the results are not classifier specific, as the random forest, SVM, and LSTM network have similar results. Due to this we only used a random forest and support vector machine for these experiments.

When testing on simulated occluded faces on BU4DFE, both the random forest and SVM achieved 99.9% accuracy in all four quadrants, showing robustness to occlusion. Testing on BP4D, the random forest achieved an average accuracy of 99.7% across the four quadrants, and SVM achieved an average accuracy of 93.2% across the four quadrants. On BP4D+, random forest and SVM achieved an average accuracy of 99.4% and 97.5%, respectively across the four quadrants. These results detail the expressive power of the detected 3D facial landmarks to reliably identify subjects under extreme conditions. See Table 2 for individual quadrant accuracies for BP4D and BP4D+ (BU4DFE not shown as all quadrants had same accuracy of 99.9% for both classifiers).

3.5 Comparisons to state of the art

We compared our proposed method to the current state of the art on BU-4DFE [34] and BP4D [36] (see Table 3 for both). To the best of our knowledge this

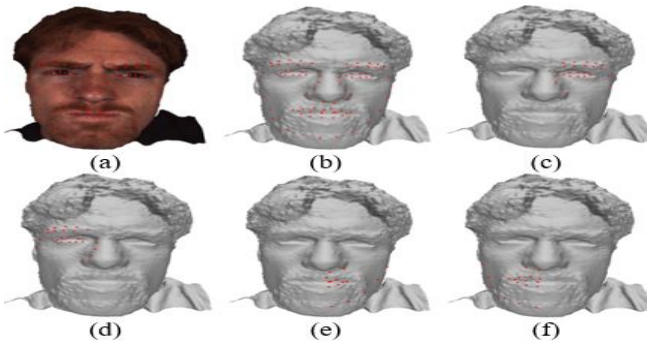

Fig. 2. Detected landmarks (BP4D [36]) used for subject ID (original 3D mesh shown only for display purposes). (a) 83 landmarks with texture (note: texture is shown for display purposes only showing robustness to facial hair); (b) 83 landmarks; (c) top left quadrant; (d) top right quadrant; (e) lower left quadrant; and (f) lower right quadrant.

Table 3. State-of-the-art comparisons.

Method	BU4DFE	BP4D
Proposed Method (RF)	**100%**	**99.9%**
Proposed Method (LSTM)	**100%**	**99.9%**
Proposed Method (SVM)	99.9%	**99.9%**
Sun et al. [27]	98.61%	N/A
Fernandes et al. [19]	96.71%	N/A
Canavan et al. [8]	92.7%	93.4%

is the first study to perform subject identification on BP4D+ [39]; therefore, we did not have any works to compare against resulting in a baseline for the community. In these comparisons, it is important to note that Canavan et al [8] used 1800 and 2400 frames from BU-4DFE and BP4D, respectively, for their experiments. We used all data in both datasets (60402 and 367474 respectively). The work from Sun et al. [27] also requires both spatial and temporal information to achieve their results of 98.61%, and while our approach can incorporate temporal information (e.g. LSTM), it can also identify a subject based on one frame of data, which is useful when temporal information is not available.

4 Conclusion

We have shown 3D facial landmarks can be used for subject identification across large variations in expression. We validated our approach on three 3D emotion-based face databases (BU4DFE [34], BP4D [36], and BP4D+ [39]), using a random forest, support vector machine, and long short-term neural network. The proposed method outperforms current state of the art on 2 publicly available 3D face databases achieving a max identification accuracy of 100% on BU-4DFE and 99.9% on BP4D. To the best of our knowledge, this is the first work to report

subject identification results on the BP4D+. We have also shown the detected landmarks can be used for subject identification in the presence of facial occlusion (simulated). We will further investigate this robustness to expression and occlusion in future work, by investigating other state-of-the-art 3D face emotion datasets such as 4DFab [9], which was also designed with biometrics studies in mind, as well as large variations in expression.

We are also interested in emotion-invariant multimodal subject identification. In this paper, we have shown that 3D landmarks are invariant to large expression changes for the task of subject identification. Since facial expressions are often physiological responses to emotion, emotion-invariant identification can have a broad range of applications such as medicine and healthcare (e.g., identifying individuals despite expressions of pain). Multimodal approaches are generally more accurate due to the fusion of heterogeneous data, each contributing identifying information. Considering this, we hypothesize a multimodal approach will significantly advance research on emotion-invariant subject identification while yielding new insight on the impact of emotion on novel modalities such as smartphone sensor data (e.g., accelerometer and touch measurements) and other unconstrained and transparently acquired data. Such approaches will be valuable for continuous subject identification.

References

1. Berretti, S., Del Bimbo, A., Pala, P.: 3d face recognition using isogeodesic stripes. IEEE Transactions PAMI **32**(12), 2162–2177 (2010)
2. Blanz, V., Vetter, T.: Face recognitoin based on fittign a 3d morphable model. IEEE Transactions on PAMI **32**(12), 1063–1074 (2003)
3. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. CVIU **101**(1), 1–15 (2006)
4. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
5. de Bruijne, M., et al.: Adapting active shape models for 3d segmentation of tubular structures in medical images. In: BICIPMI
6. Canavan, S., Zhang, X., Yin, L.: Fitting and tracking 3d/4d facial data using a temporal deformable shape model. In: ICME (2013)
7. Canavan, S., et al.: Evaluation of multi-frame fusion based face classification under shadow. In: ICPR. pp. 1265–1268 (2010)
8. Canavan, S., et al.: Landmark local on 3d/4d range data using a shape index-based stat shape model with global and local constraints. CVIU **139**, 136–148 (2015)
9. Cheng, S., et al.: 4dfab: A large scale 4d database for facial expression analysis and biometric applications. In: CVPR. pp. 5117–5126 (2018)
10. Daoudi, M., et al.: Shrec 2008-shape retrieval contest of 3d face scans (2008)
11. Echeagaray-Patron, B., Kober, V., Karnaukhov, V., Kuznetsov, V.: A method of face recognition using 3d facial surfaces. J of CTE **62**(6), 648–652 (2017)
12. Emambakhsh, M., Evans, A.: Nasal patches and curves for expression-robust 3d face recognition. IEEE Transactions on PAMI **39**(5), 995–1007 (2016)
13. Hochreiter, S., Schmidhuber, J.: Lstm. Neural computation **9**(8), 1735–1780 (1997)
14. Huang, J., Heisele, B., Blanz, V.: Component-based face recognition with 3d morphable models. ICAVPA (2003)

15. Kakadiaris, I., et al.: 3d face recognition. In: BMVC (2006)
16. Kakadiaris, I., et al.: 3d face recognition in the presence of facial exp: An annotated deformable model approach. IEEE Transactions on PAMI **29**(4), 640–649 (2007)
17. Kakadiaris, I.o.: Multimodal face recognition: Combination of geometry with physiological information. In: CVPR. vol. 2, pp. 1022–1029 (2005)
18. Kemelmacher-Shlizerman, I., et al.: In: CVPR. pp. 4873–4882 (2016)
19. Lawrence, S., B., J.: 3d and 4d face recognition: a comprehensive review. Recent Patents on Engineering **8**(2), 112–119 (2014)
20. Lei, Y., et al.: An efficient 3d face recognition approach using local geometrical signatures. Pattern Recognition **47**(2), 509–524 (2014)
21. Liu, W., et al.: Sphereface: Deep hypersphere embedding for face recognition. In: CVPR. pp. 212–220 (2017)
22. Parkhi, O., et al.: Deep face recognition. In: BMVC (2015)
23. Phillips, P., et al.: Overview of the face rec grand challenge. In: CVPR (2005)
24. Saragih, J., Lucey, S., Cohn, J.: Deformable model fitting by regularized landmark mean-shift. International Journal of Computer Vision **91**(2), 200–215 (2011)
25. Singh, S., Prasad, S.: Techniques and challenges of face recognition: A critical review. Procedia computer science **143**, 536–543 (2018)
26. Soltanpour, S., Boufama, B., Wu, Q.J.: A survey of local feature methods for 3d face recognition. Pattern Recognition **72**, 391–406 (2017)
27. Sun, Y., Yin, L.: 3d spatio-temporal face recognition using dynamic range model sequences. In: CVPRW (2008)
28. Sun, Y., et al.: Deep learning face representation by joint identification-verification. In: Advances in neural information processing systems. pp. 1988–1996 (2014)
29. Tieleman, T., Hinton, G.: Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA **4**(2), 26–31 (2012)
30. Tuan Tran, A., et al.: Regressing robust and discriminative 3d morphable models with a very deep neural network. In: CVPR. pp. 5163–5172 (2017)
31. Turk, M., Pentland, A.: Face rec using eigenfaces. In: CVPR. pp. 586–591 (1991)
32. Vapnik, V.: The support vector method of function estimation. In: Nonlinear Modeling, pp. 55–85 (1998)
33. Wen, Y., et al.: A discriminative feature learning approach for deep face recognition. In: ECCV. pp. 499–515 (2016)
34. Yin, L., et al.: A high-red 3d dynamic facial expression database. In: FG (2008)
35. Zhang, L., et al.: Sparse representation or collaborative representation: Which helps face recognition? In: ICCV. pp. 471–478 (2011)
36. Zhang, X., et al.: Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image and Vision Computing **32**(10), 692–706 (2014)
37. Zhao, W., et al.: Discriminant analysis of principal components for face recognition. In: Face Recognition, pp. 73–85 (1998)
38. Zhao, W., et al.: Face recognition: a literature survey. ACM Computing Surveys **35**(4), 399–458 (2003)
39. Zheng, Z., et al.: Multimodal spontaneous emotion corpus for human behavior analysis. In: CVPR (2016)
40. Zhong, C., et al.: Robust 3d face rec using learned vis codebook. In: CVPR (2007)
41. Zhu, X., et al.: High-fidelity pose and expression normalization for face recognition in the wild. In: CVPR. pp. 787–796 (2015)