Expressive Maps for 3D Facial Expression Recognition

Omar Ocegueda   Tianhong Fang   Shishir K. Shah   Ioannis A. Kakadiaris
Computational Biomedicine Lab, Univ. of Houston, Houston, TX, USA

Abstract

We present a semi-automatic 3D Facial Expression Recognition system based on geometric facial information. In this approach, the 3D facial meshes are first fitted to an Annotated Face Model (AFM). Then, the Expressive Maps are computed, which indicate the parts of the face that are most expressive according to a particular geometric feature (e.g., vertex coordinates, normals, and local curvature). The Expressive Maps provide a way to analyze the geometric features in terms of their discriminative information and their distribution along the face and allow the reduction of the dimensionality of the input space to 2.5% *of the original size. Using the selected features a simple linear classifier was trained and yielded a very competitive average recognition rate of* 90.4% *when evaluated using ten-fold cross validation on the publicly available BU-3DFE database.*

1. Introduction

Emotion analysis based on facial expressions has received a lot of attention because of its potential applications in Human Computer Interaction. In recent years, the interest on the development of automatic and semi-automatic 3D facial expression recognition (3D-FER) systems has increased due to the availability of 3D databases and more affordable 3D sensors. It has been shown that the 3D shape of the human face contains very rich information and provides invariance against the factors that have severely hindered similar analysis of 2D images, such as the variations in head pose and illumination conditions. These properties have been successfully exploited in the past to develop efficient face recognition (FR) and FER systems [6][16][15]. Kakadiaris *et al.* [6] have successfully used an adaptive deformable model approach [8] for FR. In this work, we explore the possibility of incorporating diverse dense geometric features for FER that result from model-based mesh fitting algorithms (e.g., the vertex coordinates, vertex normals, and local curvature). One inherent difficulty is the resulting high dimensionality of the mesh representation that may contain a high number of vertices. However, the expressive information is concentrated along specific regions of the face (e.g., the mouth), which may vary depending on the specific geometric features under consideration (e.g., the coordinate or normal vectors at each vector). This idea motivates the application of the feature selection framework proposed by Ocegueda *et al.* [11] to define the *Expressive Maps* which result in a dimensionality reduction technique for dense geometric features and provide a way to analyze them in terms of their discriminative information and their distribution along the face. Our contributions are: (i) Development of a method for mesh fitting under landmark constraints for facial meshes with strong facial expressions using only five facial landmarks; (ii) Extension of the feature selection framework proposed by Ocegueda *et al.*[11] for FER, from which we define the *expressive maps*, which indicate the most important areas of the face for FER; (iii) A systematic evaluation of the discriminatory power of different dense geometric features for FER.

This paper is organized as follows. Previous work in this field is discussed in Section 2. In Section 3, our mesh fitting algorithm under landmark constraints is presented. In Section 4, the dense geometric features that are used in this study are defined. The *Expressive Maps* are introduced in Section 5. In Section 6, the experimental results are presented.

2. Previous Work

Currently, the semi-automatic methods for 3D-FER use a large set of manual annotations to generate features. The methods for FER differ mainly in the type of features and the classification method. One of the first attempts in this regard was the work by Wang *et al.* [17]. In their method, each vertex of the 3D scan was classified into one of 12 surface labels based on local curvature information. To overcome the need of a dense correspondence to obtain consistent feature dimensions, they used 64 manually annotated fiducial points to divide the face into seven expressive regions and concatenated the normalized histograms of the surface labels from each region to form the descriptor. They reported an 83.6% average recognition performance using Linear Discriminant Analysis (LDA) on the BU-3DFE database [18]. In their initial work, Soyel and Demirel [14] presented a back-propagation neural network classi-

978-1-4673-0063-6/11/$26.00 ©2011 IEEE

fier using only six descriptive Euclidean distances between 11 key landmarks and reported an average recognition rate of 91.3 %. Tang and Huang [16] presented a method for feature selection based on maximizing the average relative entropy of marginalized class-conditional feature distributions. Given a set of 83 manually annotated landmarks from the BU-3DFE database, they applied their feature selection algorithm to a complete pool of C_2^{83} normalized Euclidean distances between the landmarks and used a regularized AdaBoost algorithm with LDA as the weak classifier. In 2009, Soyel and Demirel [15] proposed a method based on the distances between each pair of the of 83 landmarks provided in the BU-3DFE database, yielding a set of feature vectors of dimension 3,403. One way to achieve higher FER rates in a fully automatic scenario, using the ideas of the semi-automatic methods, is to minimize user intervention such that the set of manually annotated landmarks can be realistically substituted by an automatic landmark detector [12]. Being sparse in nature, distances between facial landmarks are likely to be very sensitive to errors of an automatic landmark detector. Thus, systems that use a dense set of facial features are expected to be more robust to errors made in the landmark detection phase. In their initial work, Mpiperis *et al.* [9] presented a model-based framework for establishing correspondence among 3D point clouds of different faces. They used the 83 manually annotated landmarks to guide the fitting of their facial model and used the coordinates of the mesh vertices as features in their FER system. In order to reduce the dimensionality, they sequentially applied Principal Component Analysis (PCA) and LDA. Finally, they used particle swarm optimization (PSO) to discover a set of rules for FER from the resulting five-dimensional features. They reported an average recognition rate of 92.3% in BU-3DFE. In a later work [10], they reported a bilinear model for automatic simultaneous identity and expression recognition with an average expression recognition performance of 90.5% when evaluated on BU-3DFE (including all intensity levels of expression). To the best of our knowledge, this is the best performance reported on BU-3DFE using an automatic FER algorithm. Recently, Huang *et al.* [4] proposed to fit a parts-based facial model to the 3D face scans in order to obtain a one-to-one correspondence between the vertices of the scans. Their fitting algorithm is similar to the algorithm proposed by Schneider *et al.* [13], but it is enhanced with additional processing to obtain smoother fitted meshes. They reported an average performance of 83.0 % on BU-3DFE using LDA on the feature vectors extracted from their fitted meshes. For a complete review on the state of the art on 3D FER, the interested reader is refered to the survey by Fang *et al.* [3].

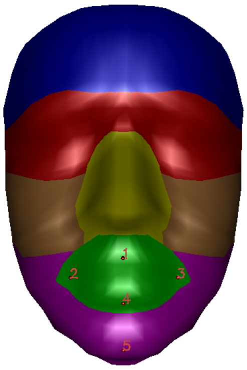


Figure 1: Depiction of the 5 landmarks used in this work: 1) mid upper lip, 2) right mouth corner, 3) left mouth corner, 4) mid lower lip, and 5) chin tip.

3. Mesh fitting under landmark constraints

Our framework is based on the deformable model approach proposed by Kakadiaris *et al.* [6]. This approach has been proven to be very effective for face recognition purposes. According to that approach, a 3D facial mesh is first rigidly aligned with an Annotated Face Model (AFM) using Iterative Closest Point (ICP). Then, the AFM is elastically deformed to fit the aligned input scan. As a result, different 3D facial meshes can be brought into correspondence and a direct comparison is made possible. While this approach has shown its effectiveness in 3D face recognition, it does not handle datasets with extreme facial deformations very well. Also, although the shape of the fitted model may be visually correct, it is not semantically correct (e.g., some of the vertices in the mouth area are identified as part of the chin as illustrated in Fig. 2). To overcome this limitation, we enhanced the AFM with five facial landmarks (Fig. 1). We use these facial landmarks to guide the fitting process by first deforming the AFM using the Thin-Plate Spline (TPS) model proposed by Bookstein [1]. In other words, the TPS warping serves as an initialization to the fitting algorithm. The warping is constrained by the landmarks; hence, it is able to deform the model to roughly match the target surface. Then, the elastic deformation will take over and drive the vertices of the AFM towards the target while maintaining its overall smoothness. The fitting pipeline is presented in Algorithm 1. The application of the TPS model for 3D mesh registration was first proposed by Schneider *et al.* [13]. However, instead of elastically deforming the reference model after the TPS warping step, they proposed a *re-sampling process*. In the re-sampling process, each vertex of the reference mesh is projected onto the input mesh by intersecting its normal with the input surface. The main limitation of that method is that the re-sampled mesh is not smooth and it requires a post-processing step to repair the

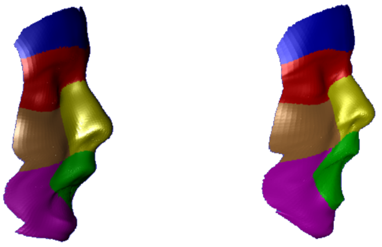


Figure 2: Left: Fitting result using the deformable model proposed by Kakadiaris *et al.* [6]. Right: Fitting result using the proposed framework.

regions where the projections of the reference vertices do not exist. A similar approach was recently proposed by Huang *et al.* [4]. In their work, they reported similar difficulties in their re-sampling process and proposed several improvements to achieve a smoother mesh fitting result of their parts-based facial model.

Algorithm 1 Mesh fitting using landmark constraints

```
Require: The AFM, $\mathbf{X}$, annotated with $K$ facial landmarks
Require: A 3D facial scan, $\mathbf{Y}$ annotated with the corresponding $K$ facial landmarks (in our implementation $K = 5$)
  1. Rigidly align $\mathbf{Y}$ to $\mathbf{X}$ using ICP (refined alignment step)
  2. Deform $\mathbf{X}$ so that its landmarks exactly match the corresponding landmarks of $\mathbf{Y}$ using the TPS model proposed by Bookstein [1]
  3. Deform $\mathbf{X}$ using the elastically adaptive deformable model proposed by Kakadiaris et al. [6]
Output: Deformed AFM that fits the input scan
```

4. Geometric data representation

One of the most important properties of the AFM is its UV parametrization, which allows the representation of the 3D mesh as a three-channel image: each vertex v of the AFM is mapped to a pixel (i_v, j_v) of an image with an arbitrary resolution. The first, second and third channels correspond to the x, y and z coordinates of the vertices, respectively, which defines the "geometry image" of the mesh. From the geometry image, any other geometric feature set can be easily computed. In particular, the "normal image" is constructed by computing the normal vector at each vertex. The main advantage of representing the 3D mesh as a multichannel image is that it is possible to apply any image processing technique directly to this representation of the mesh. Kakadiaris *et al.* [6] extracted the full Walsh wavelet packet decomposition from each band (x, y and z) of the geometry and normal images to obtain a total of 512 wavelet packets (256 corresponding to the geometry image and 256 corresponding to the normal image) from which they obtained a compact representation of the 3D mesh, which resulted in a very efficient 3D FR system. In this work, we additionally study the *local curvature* map in order to illustrate the generality of our framework, which can be easily extended to incorporate useful features (either geometric or textural) to further improve the recognition performance.

5. Expressive Maps

Aside from the rich information that is contained in a dense set of features (e.g., geometry and normal maps) there is the well known problem of the *curse of dimensionality*. As pointed out recently by Le *et al.* [7], extracting the geometrical shape features at some important locations may reduce the dimensionality while still keeping the representation robust. The geometry, normal and local curvature images are a dense set of geometric features with very high dimensionality ($256 \times 256 \times 3 = 196,608$ coefficients for the normal and geometry images, and $256 \times 256 = 65,536$ coefficients for the curvature image). Therefore, a dimensionality reduction technique must be applied to the set of features. Ocegueda *et al.* [11] proposed a framework for feature selection in scenarios in which the discriminative information is distributed along smooth regions of a lattice. In their work, the *discriminative maps* are computed from the wavelet transform of the geometry and normal images obtaining the most discriminative areas of the face for identity recognition. Notice that their framework cannot be directly applied for FER because the areas of the fitted elastically adaptive deformable model are not semantically correct when strong facial expressions are present in the data. We addressed this limitation using the facial landmarks to guide the fitting process as described in Section 3. Ocegueda *et al.* [11] cast the problem of feature selection as a binary labeling problem where the labels $\{0, 1\}$, assigned to each vertex, represent *discriminative* and *non-discriminative* vertices, respectively. After modeling the binary field as a Markov Random Field, they showed that the posterior marginal probabilities can be used for feature scoring yielding a very efficient feature selection algorithm. The posterior marginal probabilities $\{p_v(b)|v \in V\}$ defined over the set of vertices V, represent the probability of each vertex v being discriminative ($p_v(1)$) or non-discriminative ($p_v(0)$) and can be estimated by minimizing the following Gibbs Energy Function:

$$U(p) = \sum_{v \in V} |p_v - \widehat{g}_v|^2 + \lambda \sum_{<u,v>} |p_v - p_u|^2, \quad (1)$$

where $\widehat{g}_v$ is the normalized likelihood of the training set at vertex v. The likelihood is computed, assuming a Gaussian distribution of the features at each vertex, as:

$$g_v(b) = \begin{cases} \prod_{i=1}^{N} \phi(x_v^i; \widehat{\mu}_v, \widehat{\Sigma}_v) & b = 0 \\ \prod_k \prod_{i:y^i=k} \phi(x_v^i; \widehat{\mu}_v^k, \widehat{\Sigma}_v^k) & b = 1 \end{cases}, \qquad (2)$$

where $\phi(x, \mu, \Sigma)$ is the normal density function with parameters μ, Σ evaluated at the feature vector x (e.g., ϕ is a multivariate Normal density function of dimension 3 in the case of normal and geometry images but ϕ is a univariate Normal in the case of the curvature map). The estimators $\widehat{\mu}_v$ and $\widehat{\Sigma}_v$ are the *between-class* maximum likelihood estimators of the mean and variance at vertex v, respectively, while $\widehat{\mu}_v^k$ and $\widehat{\Sigma}_v^k$ are the *within-class* maximum likelihood estimators of the mean and the variance at vertex v, considering only samples of class k (in our context, $k \in \{1, 2, ..., 6\}$ which represents each of the six facial expressions under consideration). The parameter $\lambda \geq 0$ is a regularization parameter that controls the smoothness of the final probability field. As proposed by Ocegueda *et al.*[11] we obtain the minimizer of this energy function using a Gauss-Seidel iterative alorithm with the following update step for each vertex $v \in V$:

$$p_v(b) = \frac{\widehat{g}_v(b) + \lambda \sum_{u \in N_v} p_u(b)}{1 + \lambda \sharp N_v}, \qquad (3)$$

where $\sharp$ denotes the *cardinality* operator and N_v denotes the set of neighboring vertices of v ($\sharp N_v$ denotes the *number of neighbors of* v). After computing the marginal probabilities, feature selection can be performed by simply applying a threshold on the minimum probability for a vertex to be considered *expressive*. In our experiments, the proportion γ of vertices to be eliminated as *non-expressive*, was selected [11].

6. Experiments

In the context of this paper, the function $\Delta(v) = \widehat{p}_v(1)$ defines the "expressive map", which measures the probability of each vertex v being discriminative for FER. Figure 3 depicts the expressive maps using three different geometric features in the wavelet domain, the geometry, normal, and local curvature images. An advantage of using *Expressive Maps* over other dimensionality reduction techniques such as *projection pursuit* techniques [5] is that the selection is easily interpretable in the context of our problem. For example, notice that in the case of the normal map, only the coefficients corresponding to the mouth area are consistently marked as expressive. After the features were selected, a linear classifier using logistic regression was trained, as proposed by Fan *et al.* [2]. For consistency with most published semi-automatic FER systems, we used only the two most expressive samples from each subject in our experiments. In order to reduce the variations related to subject identity, the neutral expression mesh of each subject was used as reference, subtracting its feature vectors from those of each expressive mesh. Figure 5 depicts the average out-of-sample error, obtained using ten-fold cross-validation, as a function of the hyper-parameters. In each fold, 90 subjects were randomly selected for training and the remaining 10 subjects were used for testing. Notice that the regularization parameter consistently presents a higher (positive) impact on the quality of the selected features especially for compact signatures (high values of γ) which indicates the effectiveness of the expressive maps as a feature selection technique using the three geometric features under study. The confusion matrices obtained from the ten-fold cross-validation evaluation using the geometry, normal and curvature images are summarized in Tables 1,2 and 3, respectively. Finally, in Table 4, we summarize the signature size and the recognition performance obtained with each geometric feature.

	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	(92.5±7.5)%	(4.0±6.2)%	(3.5±3.9)%	0	0	0
Disgust	(9.0±6.6)%	(85.0±6.7)%	(4.5±4.1)%	(1.5±2.3)%	0	0
Fear	(7.5±5.6)%	(6.5±6.3)%	(75.5±8.8)%	(10.5±9.3)%	0	0
Joy	0	(1.5±2.3)%	(8.0±6.8)%	(90.5±6.5)%	0	0
Sadness	0	0	0	0	100%	0
Surprise	0	0	0	(0.5±1.5)%	(1.5±2.3)%	(98.0±3.3)%

Table 1: Confusion matrix using the vertex coordinates as features.

	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	(93.0±7.1)%	(5.0±6.3)%	(2.0±2.4)%	0	0	0
Disgust	(7.5±7.5)%	(80.5±10.3)%	(8.0±8.4)%	(3.5±4.5)%	0	(0.5±1.5)%
Fear	(6.5±6.3)%	(9.0±8.6)%	(74.0±11.3)%	(10.5±7.5)%	0	0
Joy	0	(1.0±2.0)%	(8.0±7.4)%	(91.0±8.0)%	0	0
Sadness	0	0	(1.0±2.0)%	0	(99.0±2.0)%	0
Surprise	0	0	(1.0±2.0)%	(0.5±1.5)%	(1.5±2.3)%	(97.0±4.0)%

Table 2: Confusion matrix using the vertex normals as features.

	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	(89.0±9.7)%	(6.0±5.8)%	(3.5±3.9)%	(1.5±2.3)%	0	0
Disgust	(7.5±5.6)%	(80.0±13.2)%	(8.5±12.1)%	(4.0±4.3)%	0	0
Fear	(8.0±5.1)%	(5.0±5.0)%	(74.0±7.7)%	(13.0±8.4)%	0	0
Joy	(0.5±1.5)%	(2.0±2.4)%	(8.0±7.1)%	(89.5±8.2)%	0	0
Sadness	0	0	0	0	(99.0±3.0)%	(1.0±3.0)%
Surprise	0	0	0	0	(1.5±3.2)%	(98.5±3.2)%

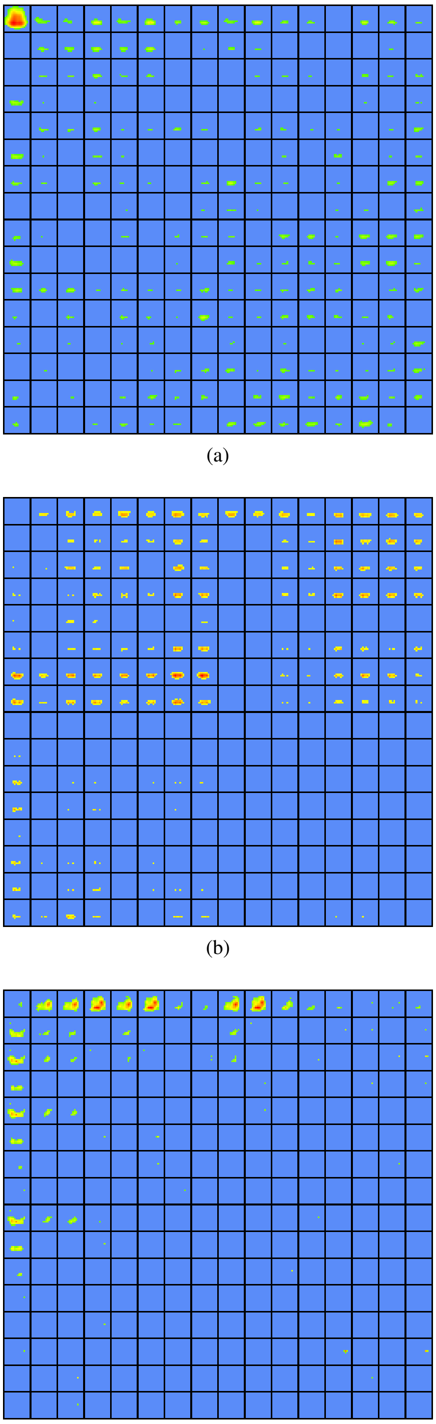
Table 3: Confustion matrix using the local curvature as features.

Feature type	Signature size	Average accuracy
Geometry	4,917 coeff. (2.5%)	90.4%
Normal	3,825 coeff. (1.9%)	89.4%
Curvature	1,639 coeff. (2.5%)	88.3%

Table 4: Signature size and average FER performance obtained using different geometric features. The percentage of the original size is shown in parenthesis in column 2.

7. Conclusion

In this paper, we have presented a 3D-FER system based on geometric facial information. Specifically, we proposed



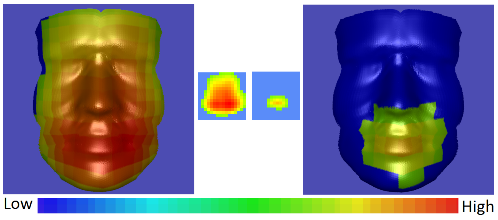
(a)

(b)

(c)

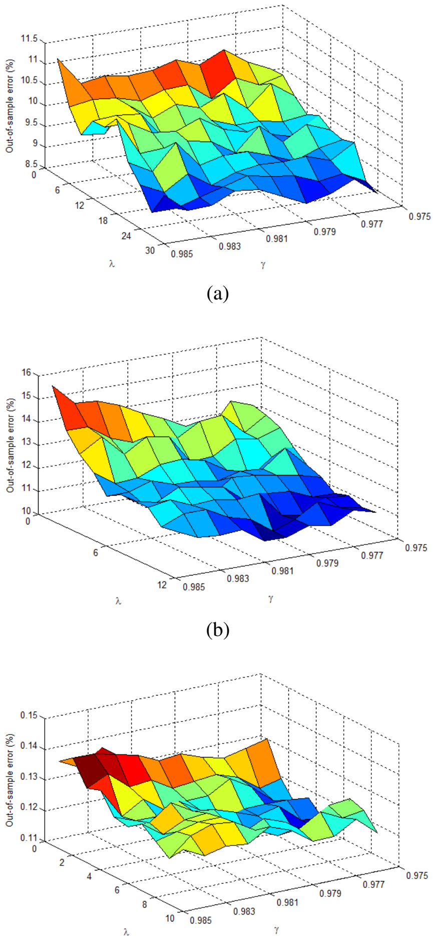
Figure 3: Depiction of the *Expressive maps* according to three different geometric features: (a) Geometry images, (b) Normal images, and (c) Local curvature images.

an extension of the mesh fitting algorithm developed by Kakadiaris *et al.* [6] for facial meshes with strong facial ex-



(a)

Figure 4: Depiction of an *Expressive map* computed in the wavelet domain mapped back to the 3D mesh. The mapping from two wavelet packets computed from the geometry images is depicted (Fig. 3 (a)). Note that in most of the wavelet packets the coefficients located along the mouth area are marked as "expressive".



(a)

(b)

(c)

Figure 5: Depiction of the out-of-sample error as a function of the hyper-parameters obtained using different geometric features: (a) Geometry images, (b) Normal images, and (c) Local curvature images. The regularization parameter in the GMPM model [11] improves the feature selection, especially for compact signatures.

pressions which requires only five facial landmarks. Using our mesh fitting algorithm we obtained a dense correspondence between the 3D facial meshes from which we are able to compute several geometric features which are already in one-to-one correspondence. The UV-parametrization of the AFM [6] is used to represent the facial features of the 3D mesh as a multi-channel image, from which the Walsh wavelet packet decomposition is computed. In order to reduce the dimensionality of the dense set of wavelet coefficients we proposed to use the *expressive maps* obtained by applying the feature selection framework proposed by Ocegueda *et al.* [11] (which is only possible after the 3D meshes are semantically consistent). This allows us to reduce the dimensionality of the original feature space to 2.5% of its original size. An additional advantage of the *expressive maps* is that they are directly interpretable in the context of our problem. For example, we observed that the information located along the mouth area is consistently marked as "expressive".

Acknowledgments

This research was funded in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL) and by the University of Houston (UH) Eckhard Pfeiffer Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, the U.S. Government, or UH.

References

[1] F. L. Bookstein. Principal warps: thin plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–85, June 1989. 2, 3

[2] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, August 2008. 4

[3] T. Fang, X. Zhao, O. Ocegueda, S. Shah, and I. Kakadiaris. 3D Facial Expression Recognition: A perspective on promises and challenges. In *Proc.* 9th *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–8, Santa Barbara, CA, March 2011. 2

[4] Y. Huang, X. Zhang, Y. Fan, L. Yin, L. Seversky, T. Lei, and W. Dong. Reshaping 3D facial scans for facial appearance modeling and 3D facial expression analysis. In *Proc.* 9th *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 422–429, Santa Barbara, CA, March 2011. 2, 3

[5] P. J. Huber. Projection pursuit. *Annals of Statistics*, 13(2):435–475, 1985. 4

[6] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, April 2007. 1, 2, 3, 5, 6

[7] V. Le, H. Tang, and T. Huang. Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. In *Proc.* 9th *IEEE Conference on Automatic Face and Gesture Recognition*, pages 414–421, Santa Barbara, CA, March 21-25 2011. 3

[8] D. Metaxas and I. Kakadiaris. Elastically adaptive deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1310–1321, 2002. 1

[9] I. Mpiperis, S. Malassiotis, V. Petridis, and M. Strintzis. 3D facial expression recognition using swarm intelligence. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2133–2136, Las Vegas, NV, Mar. 31 - Apr. 4 2008. 2

[10] I. Mpiperis, S. Malassiotis, and M. Strintzis. Bilinear models for 3D face and facial expression recognition. *IEEE Transactions on Information Forensics and Security*, 3(3):498–511, 2008. 2

[11] O. Ocegueda, S. K. Shah, and I. A. Kakadiaris. Which parts of the face give out your identity? In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 641–648, Colorado Springs, CO, June 2011. 1, 3, 4, 5, 6

[12] P. Perakis, G. Passalis, T. Theoharis, G. Toderici, and I. Kakadiaris. Partial matching of interpose 3D facial data for face recognition. In *Proc.* 3rd *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 439–446, Arlington, VA, September 28-30 2009. 2

[13] D. C. Schneider and P. Eisert. Algorithms for automatic and robust registration of 3D head scans. *Journal of Virtual Reality and Broadcasting*, 7(7), October 2010. 2

[14] H. Soyel and H. Demirel. Facial expression recognition using 3D facial feature distances. In *Proc. International Conference on Image Analysis and Recognition*, volume 4633, pages 831–838, Montreal, Canada, Aug. 22-24 2007. 1

[15] H. Soyel and H. Demirel. Optimal feature selection for 3D facial expression recognition with geometrically localized facial features. In *Proc. Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control*, pages 1–4, Famagusta, Cyprus, September 2009. 1, 2

[16] H. Tang and T. Huang. 3D facial expression recognition based on automatically selected features. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, Anchorage, AK, June 24-26 2008. 1, 2

[17] J. Wang, L. Yin, X. Wei, and Y. Sun. 3D facial expression recognition based on primitive surface feature distribution. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1399–1406, New York, NY, Jun. 17-22 2006. 1

[18] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In *Proc.* 8th *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–6, Amsterdam, The Netherlands, September 2008. 1