HUMAN FACIAL EXPRESSION RECOGNITION USING A 3D MORPHABLE MODEL

S.Ramanathan, Ashraf Kassim, Y.V. Venkatesh and Wu Sin Wah

Dept. of Electrical and Computer Engg, National University of Singapore, Singapore 119260.

ABSTRACT

We propose a novel approach to the detection and classification of human facial expressions using a morphable 3D model. We acquire the various expressions of an individual using a face scanner that produces textured 3D meshes using stereoscopic reconstruction. A **Morphable Expression Model (MEM)**, that incorporates emotion-dependent face variations in terms of morphing parameters, is then computed by establishing correspondence among the emotive faces. These morphing parameters are used for emotion recognition and classification. We demonstrate that the different facial expressions correspond to distinct clusters in the **expression space**.

Index Terms— Pattern Recognition, Facial Expression Classification, Morphable Expression Model, Clustering methods.

1. INTRODUCTION

Automated recognition of faces and facial expressions is one of the most attractive applications of pattern recognition and image analysis. Despite significant progress in image-based face and expression recognition, no automated system has been shown to work satisfactorily outside a controlled environment. Since the face image is a function of *illumination*, *viewpoint*, *pose*, *color* etc., recognition has turned out to be difficult [1]. As the human face is an active 3D object, shape information appears better suited to describe the face than its projection as a 2D intensity image. Also, the 3D face shape is insensitive to pose, illumination or color; and recognition rates have been found to increase significantly when shape information is combined with gray-level information.

Still, the number of works that deal with 3D face and expression recognition are relatively few [2]. Curvature measurements computed from the 3D face shape acquired using range scanners have been used for recognizing faces [3]. However, the inability to handle non-rigid face deformations associated with expressions is a major limitation of a majority of these algorithms. Range scanners produce the face shape as an ordered point-cloud containing thousands of textured points in 3D space. The motion of critical face feature points extracted from the face shape needs to be tracked to understand facial expressions. The feature points are either manually marked [4, 5] or, alternatively, a dense point-to-point correspondence is established between the reference face and the novel test face, using optical flow-like methods [6]. Furthermore, feature points in a novel face can be moved appropriately to synthesize various emotions. Morphable models, typically employed for this purpose, express new objects as a linear combination of prototypical objects of the same class. Blanz and Vetter use morphable 3D models for (i) face shape reconstruction from images [7] (ii) face recognition [6] and (iii) animating faces in images and video [5].

We propose a novel morphable 3D face model to identify and classify human facial expressions. Out of the six *basic emotions* defined by Ekman [8], namely, *happiness, sadness, anger, surprise, fear* and *disgust*, we present results for the first three along with the reference *neutral* expression. However, the proposed framework can be extended to handle the other emotions as well. The expressive faces of an individual are first acquired using a 3D face scanner. The Morphable Expression Model (MEM) is then computed in order to obtain the various expressions as deformations from the *neutral* face. We use Shelton's [9] correspondence computation algorithm to compute face correspondences and obtain expression-related face deformations as a set of vector differences. The MEM can be used to represent an expressive face uniquely in terms of Morphing Parameters (MPs) in the *expression space*. We find that the different expressions form distinct clusters in the *expression space*. We demonstrate the usefulness of the MEM in identification and classification of facial expressions.

The paper is organized as follows. Section 2 outlines related work available in the literature. Section 3 details the MEM synthesis and expression recognition using the MEM. Conclusions with directions for future work are presented in Section 4.

2. RELATED WORK

Image-based facial expression analysis [10] consists of 3 steps - face detection, face feature extraction and emotion classification. Matsuno *et al.* [11] suggested face detection and recognition of the facial expression in an image using a 2D Potential Net. The energy of the net is employed to identify probable faces in the image and the net is projected onto the emotion space to infer the expression of a novel face. Recognizing facial expressions using a neural network-based sys-

tem is discussed in [12] where the face image is subjected to a grayscale transformation followed by frequency analysis to track the shape of the eyes, nose, mouth and cheeks. A feature vector computed from the contours is fed into the neural network which outputs the corresponding expression.

Over the years, optical flow methods have been employed for emotion detection. In [13], expression recognition from videos is performed by tracking face regions using optical flow to identify the beginning and end of an expression. In [5], an optical-flow based dense point-to-point correspondence is computed between exemplar faces for 3D morphable model synthesis. Another work on expression synthesis using 3D range images is described in [4]. The position of the facial feature points is quantified relative to an object-centered coordinate system invariant to pose. An image warping technique is applied then to both the range and texture images to generate expressive faces from the *neutral*.

While [5] and [4] deal with expression synthesis, this is one of the first works that primarily focuses on expression classification in 3D. We demonstrate that the various expressions can be distinctly clustered in the *expression space* using the MEM.

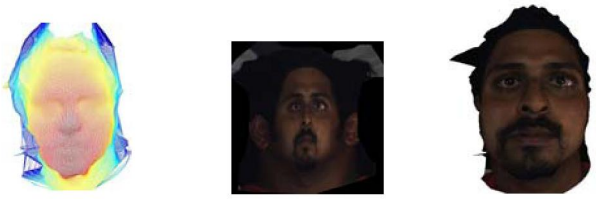
3. MEM SYNTHESIS

To demonstrate the similarity among faces of different individuals showing the same expression as well as the differences among the various expressions shown by the same face, we create a vector space using the MEM. This section describes the process of (i) data acquisition, (ii) correspondence computation (iii) synthesizing the MEM for each individual in the face database and (iv) recognizing the various expressions of the individual using the MEM.

3.1. Data acquisition

We have created a database of 3D faces using the $Geometrix^{TM}$ *FaceVision 600* scanner that generates an ear-to-ear triangular mesh representation of the human face. The *neutral*, *happy*, *sad* and *angry* faces of a person are obtained as textured 3D meshes computed using stereoscopic reconstruction. The scanned face data is shown in Fig 1. The textured 3D triangular mesh can be visualized as follows. It consists of a set of vertices $\boldsymbol{V}$, n in number, and each vertex is associated with positional (x_i, y_i, z_i) and color (R_i, G_i, B_i) attributes, i=1...n. The triangle vertex incidence matrix $\boldsymbol{T}$=$[T_1\ T_2 \ldots T_m]^T$ describes the relation between each triangle and its constituent vertices. Typically, m and n are variable for each face mesh with a maximum possible resolution of 100000 triangles per mesh.

Semi-automatic pre-processing of the textured faces is performed to eliminate surface spikes and non-face regions to retain only the face region. The trimmed faces are automatically normalized in size and orientation by aligning the key face feature points, like the eye-centers and nose-tip. The trimmed faces contain about 6000-20000 vertices but due to computational constraints, we reduce all face meshes to about 2000 vertices.



(a) (b) (c)

Fig. 1. (a) 3D face mesh generated using stereoscopic reconstruction. (b) Registered texture image. (c) Textured 3D triangular face mesh with 17541 vertices and 34871 triangles.

3.2. Computing face correspondences

The most crucial step involved in MEM synthesis is the identification of corresponding points in the expressive faces of a given individual. As the topology of the face meshes is variable, the optical flow algorithm used in [5, 6, 7] cannot be applied for correspondence computation. Instead, we adapt Shelton's algorithm [9] applicable to any pair of triangular surfaces for MEM synthesis. Given any two triangular meshes **A** and **B**, Shelton proposed to obtain the correspondence mapping C from **A** to **B** by minimizing the energy function

$$E(C) = E_{sim}(C) + \alpha E_{str}(C) + \beta E_{smooth}(C) \qquad (1)$$

Here, C can be considered as an update of **A** such that $C(\mathbf{A}) = \mathbf{B}$. Also, for C to be a *good* correspondence mapping (i) $C(\mathbf{A})$ needs to be *close* to **B** as given by the similarity measure E_{sim}, (ii) C should distort **A** minimally as given by the structural energy E_{str} and (iii) $C(\mathbf{A})$ should be a smooth surface modeled using E_{smooth}. Starting from **A** the correspondence computation algorithm iteratively refines C by minimizing the energy function $E(C)$ at every step to finally match to **B**. If meshes **A** and **B** are represented by matrices A and B containing $n_{\mathbf{A}}$ and $n_{\mathbf{B}}$ vertices respectively, embedded in D dimensional space, then A and $C(\mathbf{A})$ are $n_{\mathbf{A}} \times D$ while B is $n_{\mathbf{B}} \times D$.

The three terms in $E(C)$ do not play equal roles in guiding the solution towards an optimal correspondence mapping. Typically, α is set to a large value initially and iteratively reduced to drive E_{sim} to zero. E_{smooth} has relatively less significance in determining the optimal correspondence, and at every step, β is set to a small value which is inversely proportional to our confidence in determining a good correspondence. For MEM synthesis, since the intra-face variations are typically very small (the mean distance between faces is of

the order of 10^{-2}), a large value of α precludes any deformation of the neutral face. The parameters α and β are set to values of the order of 10^{-6} and 10^{-8} respectively during correspondence computation.The variation in each pair of faces (one of them being the *neutral*) is now obtained as a difference matrix C. Each of the expressive faces can be generated by adding the corresponding difference matrix to the *neutral* face. We represent each vertex in the textured 3D face mesh using the 6D vector (x, y, z, r, g, b) to describe its positional and color attributes and therefore, each of the C's are $n \times 6$, where n is the number of vertices in the *neutral* face.

3.3. Creating the MEM

A morphable 3D face model is a vector space of textured 3D shapes [5] where any face can be represented as a linear combination of the exemplar faces constituting the model. Let $\{ S_1, S_2, \ldots, S_m\}$ represent the m textured faces and $\{C_2, \ldots, C_m\}$ denote the difference matrices that generate the $(m-1)$ expressions from the *neutral* face S_1. A warping of the *neutral* face using a convex combination of the differences generates a new face shape S as

$$S = \sum_{i=2}^{m} \zeta_i C_i \tag{2}$$

where $\{\zeta_2, \ldots \zeta_m\}$ are the convex combination coefficients. Since the C_i's are dependent on the choice of the reference, the model is biased relative to the choice of the base face. In order to remove this bias and also the restriction that the linear combination be convex, the center of the model is translated to the origin so that any scaling of a warping is now a valid warping.

The MEM is created as follows. Given the m ($m = 3$ for the results presented) expressions of a person in the form of differences $\{C_1, \ldots, C_m\}$ from the *neutral* face, the mean expression obtained as $\bar{C} = \frac{1}{m}\sum_{i=1}^{m} C_i$ is made the base face from which the other faces (including the *neutral*) will be synthesized. The $(m+1)$ C_i's (including the *neutral*) are now defined relative to the base face, *i.e.*, $C'_i = (C_i - \bar{C})$.

Also, in order to reliably classify the various emotions, each expressive face is represented in terms of an orthonormal basis obtained by performing Principal Component Analysis (PCA) on the covariance matrix MM^T where $M = [C'_1\ C'_2 \ldots C'_{m+1}]$. PCA generates the $m+1$ eigen-expressions that characterize the variations among the various expressions. We find that the leading m eigen-expressions can effectively differentiate the various expressions. If $\{V_1, \ldots, V_m\}$ denote the m eigen-expressions, the MEM is defined as

$$S = \bar{C} + \sum_{i=1}^{m} \zeta_i V_i \tag{3}$$

Each of the $(m+1)$ expressions now corresponds to a point given by the $\zeta_i s$ (that need not sum up to 1) in the m dimensional *expression space* spanned by the orthonormal V_i's.

The morphing vector $\vec{\zeta} = \{\zeta_1, \zeta_2, \ldots, \zeta_m\}$ uniquely characterizes each expression and therefore, can be used for expression *recognition* and *classification*. Once the MEM has been created, expression recognition is straightforward. Given an unknown expression, the projection of the corresponding C_i onto the *expression space* is computed and the novel expression is assigned to the expression class based on the Euclidian distance $\epsilon_k = \|\vec{\zeta} - \vec{\zeta_k}\|$ where $\vec{\zeta_k}$ is the morphing vector corresponding to the k_{th} expression. A block diagram of the MEM synthesis and the recognition scheme is shown in Fig 2.

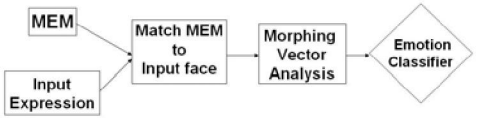


Fig. 2. Expression recognition using MEM

4. RESULTS AND DISCUSSION

The facial deformation associated with each emotion is unique, and therefore, each expression corresponds to a characteristic $\vec{\zeta}$ (Fig 3(a)) in the *expression space*. The proposed approach can distinctly cluster the various emotions in the *expression space* (Fig 3(b)). Also the clusters are adequately separated as indicated by the cluster center positions (Table 1) thereby demonstrating that reliable recognition is possible. However, since the covariance matrices vary for the different clusters, the class-separation boundaries are hypersurfaces (and not hyperplanes). The results for expression recognition for a database of 100 expressive faces (25 persons with 4 expressions per person) are presented in Table 2. An overall recognition accuracy of 97% is achieved using the proposed approach.

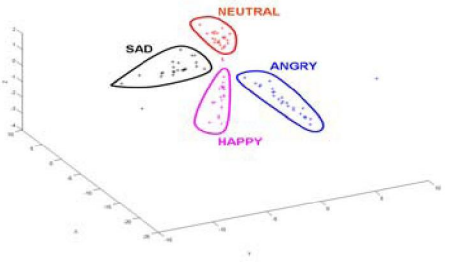
Expression	Cluster Mean position
Neutral	[4.889 1.342 0.485]
Happy	[-0.091 -0.228 -1.965]
Sad	[-0.586 -4.695 0.338]
Angry	[-13.994 0.897 0.172]

Table 1. Cluster-means in *expression space*. The distinctiveness in the principal dimension(s) for each expression ensures reliable recognition.

As expression classification is performed by measuring the deformations from the person's *neutral* face, recognition is performed in a supervised manner. However, since the intra-face expressional variations are found to be consistent over individuals, the proposed approach can be extended to perform unsupervised recognition of *identity* as well as *expression*. Also, while the morphing procedure can effectively

capture variations around the high-contrast regions like the eyes and eyebrows, the deformations around the mouth and the cheeks are not as evident. Also, clusters can be better characterized by measuring deformations of salient face features rather than the entire face. In this context, a region-based warping of the *neutral* face appears promising. The current morphing procedure cannot accurately morph to open-mouth configurations as there are major changes in geometry and color around the mouth region. We intend to extend the current framework to handle *surprise*, *fear* and *disgust* expressions as well in future.

(a)



(b)

Fig. 3. (a) Each emotion is associated with a characteristic $\vec{\zeta}$ in the *expression space*. (b) Expression clusters.

Expression	*Neu*	*Hap*	*Sad*	*Ang*
Neu	24	1	0	0
Hap	0	25	0	0
Sad	0	1	24	0
Ang	0	1	0	24

Table 2. Confusion matrix for expression recognition showing *actual* (rows) and *recognized* expressions (columns).

5. REFERENCES

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face Recognition: A literature survey," Tech. Rep. CAR-TR-948, University of Maryland, 2000.

[2] K. W. Bowyer, K. Chang, and P. J. Flynn, "A survey of approaches to three dimensional face recognition," in *International Conference on Pattern Recognition*, 2004.

[3] G. Gordon, "Face Recognition based on depth maps and surface curvature," *Geometric Methods in Computer Vision*, vol. SPIE, pp. 234 – 247, 1991.

[4] Yumiko Tatsuno, Satoshi Suzuki, Naokazu Yokoya, Hidehiko Iwasa, and Haruo Takemura, "Analysis and synthesis of six primary facial expressions using range images," in *Proceedings of ICPR*, 1996, pp. 489 – 493.

[5] V. Blanz, C. Basso, T. Poggio, and T. Vetter, "Reanimating faces in images and videos," in *Eurographics*, 2003, vol. 22(3), pp. 641 – 650.

[6] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI–25(9), pp. 1063 – 1074, 2003.

[7] V.Blanz and T.Vetter, "A morphable model for the synthesis of 3d faces," in *SIGGRAPH*, 1999, pp. 187 – 194.

[8] P. Ekman, *Emotion in the Human Face*, Cambridge Univ. Press, 1982.

[9] C. R. Shelton, "Morphable surface models," *International Journal of Computer Vision*, vol. vol 38, pp. 75 – 91, 2000.

[10] Maja Pantic and Leon J.M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI–22(12), pp. 1424 – 1445, 2000.

[11] K. Matsuno, Chil-Woo Lee, S. Kimura, and S. Tsuji, "Automatic recognition of human facial expressions," in *Proc. Fifth International Conference on Computer Vision*, 1995, pp. 352 – 359.

[12] P.K. Manglik, U. Misra, Prashant, and H.B. Maringanti, "Facial expression recognition," in *IEEE Conf. on Systems, Man and Cybernetics*, 2004, pp. 2220 – 2224.

[13] Yaser Yacoob and Larry S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI–18(6), pp. 636 – 642, 1996.