Realistic facial expression synthesis of 3D human face based on real data using multivariate tensor methods

Jacey-Lynn Minoi
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia
jacey@fit.unimas.my

Duncan F. Gillies
Department of Computing
Imperial College London
180 Queen's Gate, London, SW7 2RH, U.K.
d.gillies@imperial.ac.uk

Amelia Jati Robert Jupit
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia
rjajati@fit.unimas.my

ABSTRACT
This paper presents a novel approach of facial expression synthesis and animation using real data sets of people acquired by 3D scanners. Three-dimensional faces are generated automatically through an interface provided by the scanners. The acquired raw human face surfaces went through a pre-processing stage using rigid and non-rigid registration methods, and then each of the face surface is synthesized using linear interpolation approaches and multivariate statistical methods. Point-to-point correspondences between face surfaces are required in order to do the reconstruction and synthesis processes. Our experiments focused on dense correspondence, as well as, to use some points or selected landmarks to compute the deformation of facial expressions. The placement of landmarks is based on the Facial Action Coding System (FACS) framework and the movements were analysed according to the motions of the facial features. We have also worked on reconstructing a 3D face surface from a single two-dimensional (2D) face image of a person. After that, we employed tensor-based multivariate statistical methods using geometric 3D face information to reconstruct and animate the different facial expressions.

Keywords
Three-dimensional facial animation, facial expression synthesis, face reconstruction.

1. INTRODUCTION
Facial animation is complex and difficult to achieve realistically. Facial features that contribute the most to facial expressions are the eyelids, eyebrows and mouth. Wrinkles and budges also contribute to the change of facial appearances. Movements or the flow of features can be measured and then used to animate facial expressions. This approach is known as feature-based deformation. [Ste81] used landmark information to deform face shapes and models while [Wat87] used pseudo muscles for face expression animation. Work by [Gue98, Pig98] used facial movement information.

According to [Fas02], the deformation approach does not necessarily require extensive facial movement, which makes the animation process faster and simpler. However, this approach is unreliable in creating exaggerated realistic face shapes and facial expressions.

A large number of facial modelling and facial animation works have employed muscle-based approaches [Ter90, Lee95]. Synthetic facial movements are generated by mimicking the contraction of facial muscles. This can be done by firstly defining the functionality and locations of the facial muscles on the face model and then applying a combination of muscle contractions [Wat87]. The combinations of the muscles are defined by Action Units (AUs) from the Facial Action Coding System (FACS) framework. Using AUs could reduce the amount of work in characterising facial expression data.

Many face animators, [Fox05] for example, imitate facial muscles movements to generate facial expressions. Similar to the prior muscle-based animation problems, this approach only creates a limited set of facial expressions.

Multi-layer models supplement the use of facial muscles. A multi-layer model is built from the anatomical structure of the face, facial muscles, skin, soft tissues and etc. [Ter90, Lee95, Sif05, Wil97]. [Ter90] proposed facial animation by contracting synthetic facial muscles embedded in a face skin model. This approach improves the realism of the synthetic facial expressions; however the use of sophisticated biomedical models requires accurate simulation methods and high computational costs [Lee95]. Furthermore, animating expressions in complex multi-layer structures requires non-linear methods to simulate dynamic deformation of the skin. Failing to create a detailed skin deformation (such as wrinkles) may result in less realistic facial expressions [Ers08].

The geometry warping approach is another method for synthesizing facial expressions. The facial expression information is measured from two images – one with a neutral expression and another with a particular facial expression. The calculated facial movement difference vectors are transferred to a target image of a neutral face [Pig98, Sha01, Par96, Wil90]. The facial movement differences can be controlled by using linear interpolation. The disadvantage of this approach is that the overall shapes of the face, including the geometric shapes, poses and orientations, and facial expression, are calculated and computed together. It is therefore not a perfect solution for generating the in-between facial expressions [Ers08]. Recent work has been undertaken to overcome this weaknesses by using non-linear interpolation or by combining linear and non-linear interpolation.

[Par96] used simple geometric interpolation to synthesize expression on 3D face models. The feature points are manually digitised on each face model. This was followed by automatic expression synthesis where the data of real actors are captured and analysed [Wil90, Ber85, Ess96]. The captured face surfaces are represented using a structured mesh, along with texture information.

Segmenting face models into smaller regions is also employed with the aim of synthesizing only the relevant parts of the face contributing to an expression. [Jos03] applied this approach on 3D face models. On the other hand, [Bla03] employed a morphable technique to animate facial expressions on existing 2D images and videos. The advantage of the morphable modelling approach is that it can work on faces without acquiring examples of facial expression data of a person. [Vla05] mapped facial movements from a recorded video to a target face using an optical flow-based tracker to estimate 3D shape movements. In addition, they used a multilinear model to manage the face attributes separately. Theoretically, by using multilinear models on a larger collection of faces with different expressions, faces with any expressions can be generated. However, the collection used in Vlasic's work is limited in size. Nevertheless, the advantage of this technique is that visible facial markers or special face-spanning equipment is not required.

In order to simulate a realistic facial expression, a larger collection of facial expression examples is required. When using muscle information, accurate muscle descriptions or templates are needed to produce visually correct facial movements. Using fewer facial expression resources means that expressions may appear artificial and unrealistic.

Methods to optimise the animation computation may also be needed to allow real-time facial animations. It should be noted that facial animation field has grown into a complicated and broad subject. Facial animation applications are extensively used in various areas, including movie industries, computer games, medicine and telecommunication.

The remainder of this work is organised as follows. In Section 2, we start with a description of the data set of 3D face scans from which our synthesis of facial expression model is built.

In Section 3, we briefly present the pre-processing technique used on raw 3D face surfaces, followed by a description on the algorithm for synthesizing and animating facial expressions using linear interpolation based on landmark placements, is given in Section 4.

In Section 5, we describe PCA and LDA approaches used in this work and then introduce our idea of applying tensor model to those two approaches on generating a variety of facial expressions that can be applied in differing degrees. Following that, we look at reconstructing 3D faces from 2D photographic images of faces with only neutral expression and then map facial expression onto the reconstructed face surface.

Section 6 then describes all the experiments that were carried out in the study, and present the results of the synthesis on 3D face data. Finally, in Section 7, we conclude the paper, summarising its main contributions and describing possible future work.

2. FACE DATA SETS USED

In our experiment, we have used four face data sets of real human faces: the Notre Dame 3D face data set, the Imperial College 3D face data set, SUNY Binghampton 3D face data set, and the FERET 2D face data set.

The Notre Dame data set was acquired at the Computer Vision Research Lab at the University of Notre Dame (see web page, http://www.nd.edu/~cvrl/CVRL/Data_Sets.html). A total of 150 subjects participated in the image acquisition sessions, giving a total of 300 three-dimensional face surfaces. The 3D data was captured using a Minolta Vivid camera [Kon] which uses a structured light sensor to scan surfaces. The captured faces are only frontal faces of neutral facial expression.

In the Imperial College face data set was acquired at the Department of Computer Science at Imperial College London [Pap05]. The 3D face surfaces are captured using a VisionRT stereo camera [Vis]. It contains a set of 60 individual face surfaces of which we used two expressions, one where the subject is frowning and another smiling – totalling to 120 three-dimensional face surfaces. Each face is associated with greyscale texture image.

In the SUNY Binghampton data sets [Wan06], the 3D face surfaces consists of 7 different emotional facial expressions, namely anger, disgust, happiness, sadness, surprise, fear and neutral. Each of the emotional expression contains 4-levels of expression intensity ranging from low to high. In total, there are 2,500 faces from 100 subjects.

FERET face data set is a well-known 2D standard face image data set normally used for evaluating face recognition performance. See web page, http://www.itl.nist.gov/iad/humanid/feret/feret_master.html).

3. PRE-PROCESSING

The obtained 3D face surfaces using current technologies would require pre-processing procedure before they can be further analysed. A raw surface may have holes or spikes caused by acquisition and measurement errors. Raw surfaces also have different alignment and surface areas in addition to having a different number of points in the surface representation. The first step in any technique using statistical shape modelling is to normalise the surface with respect to the orientation and the surface area. The second step is then to establish point-to-point correspondence between the surfaces in the input database.

As a result of these steps, extraneous portions of a surface are removed so that all the surfaces cover the same features and are represented by the same number of corresponding vertices. The dense correspondence is established so that, ideally, each point on a surface represents the same anatomical position on all the other surfaces. Thus, each vertex is a landmark point once the correspondence is established. In our experiments, correspondence between 3D face models was established using the method proposed by [Pap05]. Pre-processing is applied to each raw face surface to regularise the position, scale and surface tessellation, and to establish a correspondence between faces.

4. SYNTHESIZING EXPRESSIONS USING LINEAR INTERPOLATION BASED ON LANDMARK PLACEMENTS

In this section, we describe the use of features points that gives parameters of facial expression and deformation. Using these parameters, one can generate any possible facial expressions according to the selected landmark points.

In our experiment, we selected two sets of identifiable landmark points, whereby set 1 has thirty-three landmarks and set 2 has forty-three landmark points. Figure 1(a) illustrates thirty-three landmarks and they are placed along the eyebrows, the corners of the eyes and crowfeet, the glabella, the upper part and the tip of the nose, the mouth and areas around the mouth and lips, the chin and the cheeks. Figure 1(b) shows set 2 landmark points. The chosen landmark points are based on the landmarks used in craniofacial anthropometry and muscle-based landmarks in the FACS framework.

The main source that contributes to expression variation is facial muscular movements. We employed a registration framework [Pap05], based on the selected facial landmarks, to create expressions on surface model of the face.

The step is similar to the geometric warping approach, whereby we calculate the facial movement difference vectors from two face surfaces – one face with a neutral expression and another with a particular facial expression – using landmark points. Then, we use linear interpolation to transfer the generated facial expression to a neutral face.

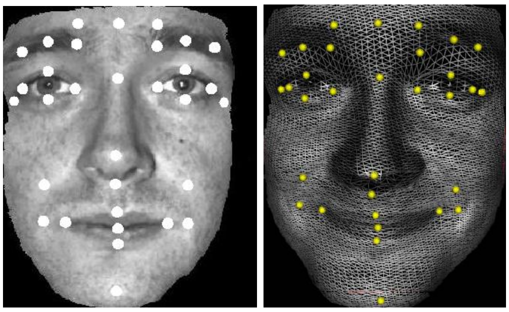

Figure 1(a). A set of 33 selected points.

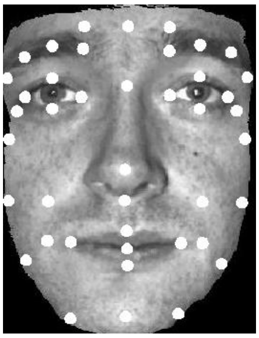

Figure 1(b). A set of 43 selected points.

5. SYNTHESIZING EXPRESSIONS USING MULTIVARIATE STATISTICAL METHODS

In this section, we present another approach to generate and synthesize facial expressions using multivariate statistical method based on PCA and mLDA (Maximum uncertainty Linear Discriminant Analysis). And then we introduce a tensor-based multivariate statistical method to construct new face shapes with a range of different face variations.

Active Shape Model (ASM) [Coo95] is a commonly used approach to build statistical shape models of the human face. The modelling of anatomical face structures are from statistical information found in a training set. Unlike the multivariate statistical method and tensor-based multivariate statistical approaches, new face shapes are created based on statistical information from pre-defined classes of specific features or face variations found in the training set.

Our idea of using tensor model on the multivariate statistical method is to use all the face features with a variety of facial variants simultaneously rather than separating them into two numbers of classes. The advantage with this method is that it is practical to generate a variety of face shapes applied in different degrees. Additionally, the transition between face shapes is also continuous and natural.

Following that, we applied the tensor-based method to the reconstructed 3D faces from 2D photographic images and synthesize facial expressions.

Revisiting PCA and LDA

Principal Component Analysis (PCA) is one of the most successful methods to reduce the dimensionality of the original space with a minimum loss of information by finding the projection directions that maximise the total scatters across all data. However, in the covariance structure of PCA, the first principal component with the largest eigenvalue does not necessarily represent the important discriminant directions to separate sample groups. Therefore, we employ the idea of using the discriminant weights given by separating hyperplanes to select among the principal components the most discriminant ones. Linear Discriminant Analysis (LDA) is computed to separate samples of distinct groups by maximising the ratio of the determinant of the between-class separability to the determinant of the within-class variability. The performance will degrade if there are only a limited number of total training samples N compared to the dimension of the feature space n. This critical issue is the singularity and instability of the within-class scatter matrix. In order to avoid these problems, we propose to use Maximum uncertainty Linear Discriminant Analysis (mLDA) approach. The idea of mLDA is to regularise the eigenvalues. The details of the mLDA method can be found in [Tho06].

Multivariate statistical method

The multivariate statistical method is essentially a two-stage approach, the first stage is to characterise a type of variation and the second stage is to reconstruct faces. The method is used to find the most significant direction of change between two classes, and to reconstruct and visualize intermediate data between two classes. This method is based on Principal Component Analysis (PCA) and Maximum uncertainly Linear Discriminant Analysis (mLDA) separating hyper-plane. This technique was first applied by [Kit06] to extract and characterise the most discriminant changes between two groups of 2D probed images.

The initial training set of 3D face data consisting of N training examples on n variables is managed by dividing the training data into two groups or classes, C_1 and C_2. The training datasets can be projected from the original vector space of N by n to a lower dimensional space using a full rank PCA transformation. The principal component space forms an $n \times m$ transformation matrix, where $m = N - 1$. This step may or may not be necessary to overcome the singularity of the within class covariance matrix. If $N \geq n$, then PCA transformation is not required. It is possible that after PCA dimensionality reduction, the within-class scatter matrix S_w may still be less than full rank. If so, the mLDA approach is used to ensure that the scatter matrix S_w is non-singular. Ordering the eigenvectors is not necessary for this process. As there are only two classes, $g = 2$ and the resulting mLDA is a unidimensional vector of length m (the linear discriminant eigenvector has dimension $m \times 1$).

Back projecting into the original data space, the most discriminant feature is a n x 1 vector. The final step in the first stage is to calculate the mean of each group and the corresponding variances on the unidimensional space. This is a very fast computation because we are dealing with one-dimensional data from two-group classification problem. To reconstruct these discriminant points (based on standard deviations and means) on the original space, we do the inverse steps. Once the classifier has been constructed, we then extract and project the discriminant vector. This can be done simply by converting the discriminant to its corresponding n x 1 dimensional face vector.

The final stage of the PCA+mLDA method is the reconstruction step. If we project the most discriminant vector found for the two classes into the original data space, we will obtain a n x 1 vector. Moving a point in the original data space in this direction will change the point from an example of one class to a maximum likelihood estimate of that point in the other class. Assuming that the spread of each class follows a Gaussian distribution, the limits of variation can be set to $\pm 3sd_i$, where sd_i is the standard deviation of each class i. By moving along the n x 1 dimensional most discriminant features based on the mean of each group and the corresponding standard deviations of each group, the face shapes according to the class variant can be reconstructed in the original face domain. As stated before, since there is only one dimensional data from a two group classification problem, the computation is very fast. Figure 2 shows the geometric overview of the two-class multivariate statistical approach. This method is then extended by using tensor model to 3D face shapes to allow multiple numbers of classes.

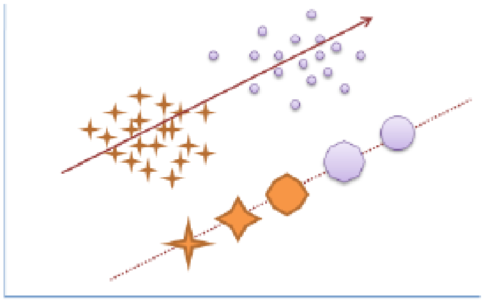

Figure 2. The geometric overview of the multivariate method between two classes.

Tensor-based multivariate statistical method

A tensor is a multidimensional matrix or mode-n matrix and is useful for the description of higher order quantities. The N^{th} order tensor is written as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, where the I_N represents the mode-n space. A tensor is then flattened (see [Lat00] for flattening details) into a matrix form, A_n, along any dimension n where $n = I_1, I_2, \ldots, I_N$.

Starting with a dataset of 3D face surfaces, we organise the data in a tensor model according to N varieties of face shapes. In our experiment, the training set is arranged into a tensor explicitly accounts for facial expression variation, where the core tensor manages the interaction between the indices of the 9-mode matrices, ($I_{subject}$ x I_{anger} x $I_{disgust}$ x I_{fear} x I_{happy} x I_{sad} x $I_{surprise}$ x $I_{neutral}$ x I_{points}). Next, we perform matrix unfolding to retrieve a set of the basis matrices for all the 9-mode matrices. Then, we compute the left singular value matrices using Singular Value Decomposition (SVD) method to obtain U_N matrices. Each of the U_N matrix can be thought of as the principal components in each mode, and they may not necessarily be of the same dimension as the tensor. The generalised N-mode SVD can be written as follows, and can be interpreted as a standard linear decomposition of the data ensemble.

$$A_n = U_n \bullet D_n (U^{n-1} \circ \ldots \circ U^1 \circ U^N \circ \ldots \circ U^{n+2} \circ U^{n+1})^T$$

The '○' is denoted as Kronecker product [Rao71] and it is applied to compute the product of the matrices. All the computed I sets of non-zero eigenvalues are extracted and stored differently according to the features of interest and are not ordered.

Since we have I-group classification, there are $g_i = g_{i-1} + (i - 1)$ number of discriminant vectors, where $g_1 = 0, i \in \{1, 2, \ldots I\}$ is the number of group classification based on the feature of interest and each represent the most expressive features. The resulting mLDA now has multiple coefficient vectors in a dimension of {g x $(m$ x $1)$} depending on the choice of facial expression transformations. Next, is to compute the second stage of multivariate statistical method. Based on the feature of interest, we then determine the most discriminant vectors that best characterise the particular change in facial features.

Tensor-based multivariate statistical method is used on the SUNY Binghampton 3D face dataset and the reconstructed 3D faces from 2D images of neutral faces.

3D Face Reconstruction

The reconstruction method has four distinct steps, which are: (1) 3D-2D alignment, (2) texture mapping,

(3) illumination adjustment, and (4) shape estimation. The detailed methods can be found in [Has07]. First, the 2D image, which is to be reconstructed, is landmarked manually by hand using a set of uniquely identifiable points with the same landmarks already known in the 3D statistical shape model. This landmarking is used to establish correspondence between the 2D image and the 3D model, which then allows mapping of the 2D image texture to the 3D face surface. The objective of the shape estimation is to optimize the match between the projection of the 3D shape model and the original 2D image. This is carried out by adjusting the shape parameters with the most discriminant vector. Figure 3 illustrates the 2D-to-3D reconstruction approach.

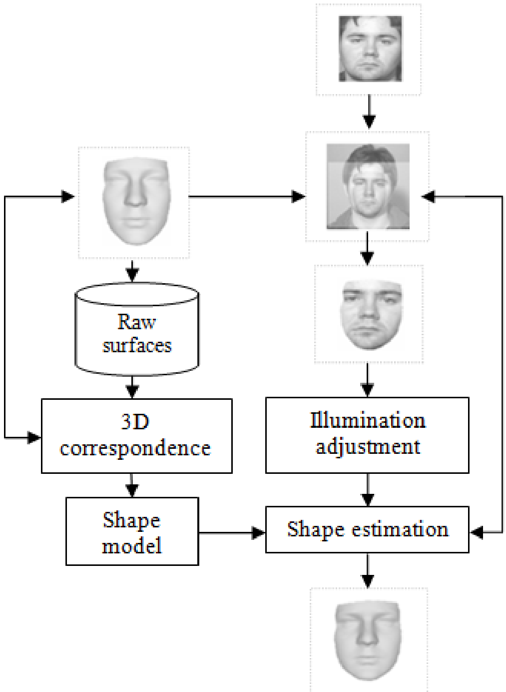

Figure 3. The overall process of 3D face reconstruction (adapted from [Has07]).

Here, we look at reconstructing 3D faces from 2D photographic images of faces with no facial expression and then generating realistic human facial expressions. In this case, frontal images from FERET and Notre Dame Face databases are used for evaluating facial expression synthesis on unseen subjects.

6. EXPERIMENTS AND RESULTS

Three facial expression synthesizing and animation experiments with three different face data sets were performed.

In the first experiment, we selected two sets of identifiable landmark points and then employ linear interpolation method. Figure 4 compares the results of the synthesis of a smile on 3D human faces using 33 and 43 landmark points. The results show an obvious geometric distortion to the overall shapes of the face (see the distortion on the nose at the bottom row of the figure), and using only 33 landmarks do not reconstructed a noticeable smile.

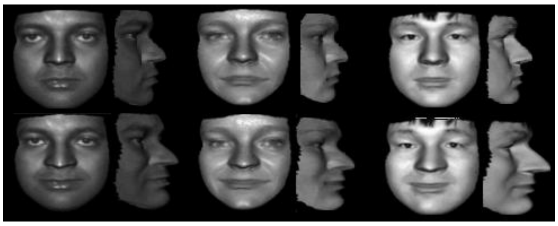

Figure 4. Synthesis of a smile using 33 landmarks (top row) and 43 landmarks (bottom row).

Figure 5 shows an example of an unrealistic face shape with a reconstructed smile using higher number of landmark points, which are placed mainly around the cheek. This shows that the number and the placement of landmark points on a face are critical to produce a realistic facial expression.

The disadvantage of using the linear interpolation approach is that the overall geometric shape of the face and the facial expression are computed together. Furthermore, there is no way to generate the in-between facial expression to allow a smooth synthesis and animation of facial expressions.

Figure 5. (a) Original smile; (b) Reconstructed caricature smile.

In the second experiment, we employed tensor-based multivariate statistical method to the SUNY Binghampton face data set. We compared the output of the synthesis with ASM.

The results of the reconstructed facial expressions using the most expressive features captured using ASM is as illustrated in Figure 6. The reconstructed faces are restricted by limiting the change in each principal component to $\pm 3\sqrt{\{\lambda_i\}}$, where λ_i are the corresponding largest eigenvalues. The first mode describes the vertical stretch along the centre of the face. The second mode models the variations in the horizontal direction. The third mode captures variation around the mouth and cheek to create expression changes from distorted frowning expression to a caricature smile.

Figure 7 illustrates the facial expression transformations of the first four largest principal components captured by ASM on the Imperial College data set. The first mode describes the horizontal stretch around the cheek and mouth. The second mode models the variation in the vertical direction. The third mode captures variations around the nose and eyes areas. The fourth mode captures the horizontal variation of the geometric shape of the face.

By examining Figure 6 and Figure 7, we see that the face shapes are not properly grouped according to facial expressions. The changes of face shape are global to the data set which makes it impossible to synthesize individual facial expression. Thus, ASM method is not suitable to capture specific facial expression variations.

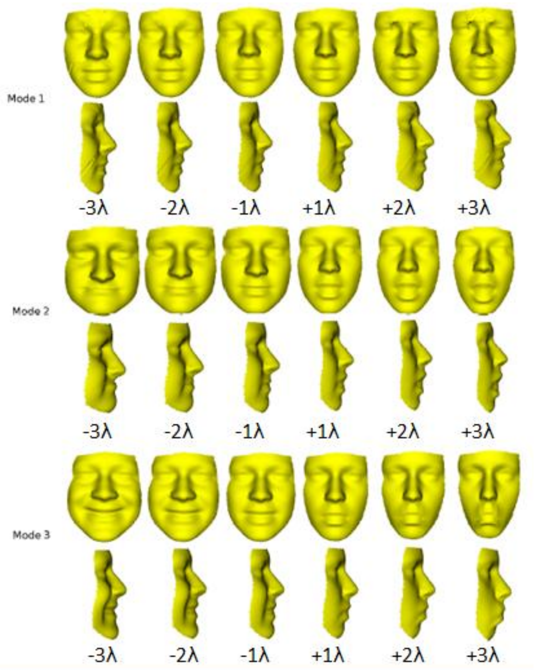

Figure 6. Synthesis of facial expressions reconstruction using the most expressive principal components captured by ASM.

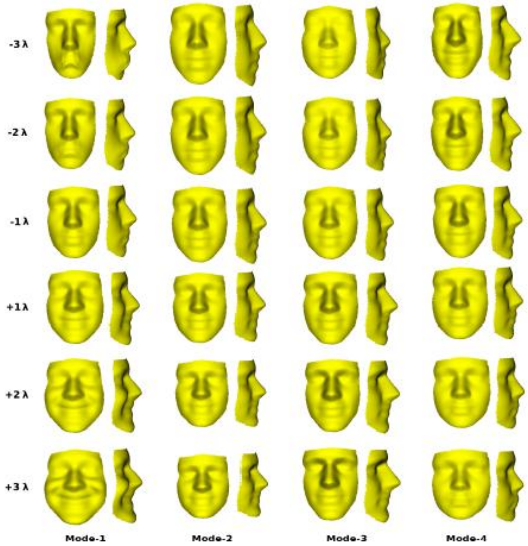

Figure 7. Synthesis of facial expressions reconstruction using the most expressive principal components captured by ASM using Imperial College face data set.

Figure 8 shows the results of the reconstruction for a neutral and an angry face expression using the tensor-based multivariate statistical method. Figure 9 illustrates the synthesis of facial expressions between a surprise and a frowning face.

This method is effective to capture facial expressions variation and it is able to find the most characteristic direction of change involved in an expression. This magnitude of change can be controlled by a single scalar magnitude. We explore the reconstruction and synthesis of face shapes by moving the point from one side of the dividing hyper-plane to the other, respecting the limits of the standard deviation and the measured mean of each sample group.

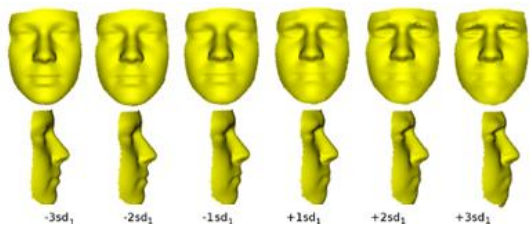

Figure 8. Synthesis from a neutral to an angry expression using most characteristic direction captured by tensor-based multivariate statistical method.

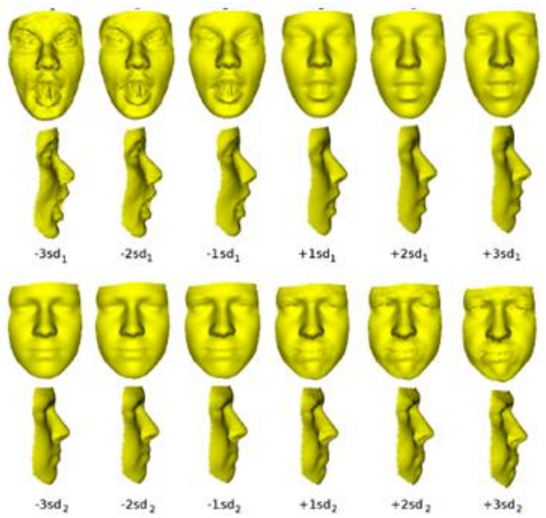

Figure 9. Synthesis of a surprise to a frowning expression using the most characteristic direction captured by tensor-based multivariate statistical method.

In the third experiment, we work on synthesizing realistic facial expressions on a reconstructed 3D real human face given only a single frontal 2D face image. We tested the technique using the 2D FERET face data set. Figure 10 shows the reconstructed 3D face shapes from faces taken from FERET face images and the synthesized facial expressions to the reconstructed 3D face shapes. The fourth column from the left of the figure displays the original faces. As we move from the original to the left side of the figure, a range of smiling expressions is generated. Similarly, when we move to the right side of the figure, a range of frowning expressions is generated. Having texture embedded to the 3D face surfaces makes the expression change smoother and more visible. For example, the raised cheeks and eyebrows, and the opened mouth show a smile.

Figure 11 shows the synthesis and animation of facial expressions on the Notre Dame 3D face data set, given that Notre Dame 3D faces only contain neutral face (the fourth column from the left of the figure).

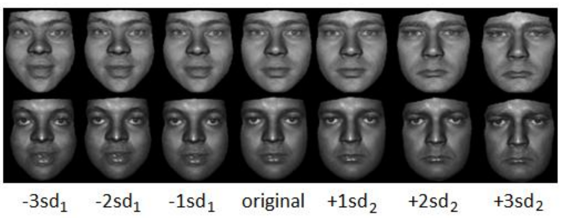

Figure 10. The reconstruction of smiling and frowning expressions using tensor-based multivariate statistical approach.

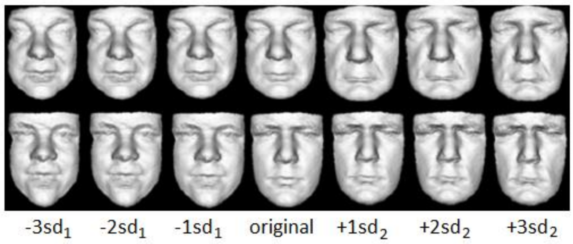

Figure 11. The reconstruction and synthesis of the most characteristic component when using 3D Notre Dame faces along the smile and frown expressions.

Examining Figures 8 to 11, we can clearly see that the tensor-based multivariate statistical approach effectively extract the 3D facial expression changes. In fact, this approach is also able to generate a gradual change on facial expressions that is not explicitly present in the training data sets.

7. CONCLUSIONS

This paper describes 3D facial expression animation using real human faces. We have analysed the placement of landmarks based on FACS for deformation and synthesis of facial expressions. Unfortunately, landmark-dependent may not create realistic facial expressions. We introduce another method, which is the multivariate statistical method, which differs from many other synthesizing and facial expression animation approaches in terms of using the whole face data points instead of selecting feature points or landmarks on the face for shape variations.

This approach could extract facial expression characteristic discriminant information efficiently, providing a gradual transformation on the 3D faces. The strength of this work is the realism of the facial expression generated as we use and extract only facial expressions from real human faces. We could also generate facial expressions at varying intensities for a subject without prior examples of expression.

The concept of using PCA+ mLDA approach to discriminate pattern of interest is not new. However, in the work of real human 3D faces, synthesizing and analysing facial expression is still in its preliminary stage. We have also implemented to using tensor model to extend the two-class problem to several classes.

8. ACKNOWLEDGMENTS

The authors would like to thank Prof. Carlos Thomaz for his technical advices, the referees for their constructive comments which helped to improve this paper, and Universiti Malaysia Sarawak (UNIMAS) for the travelling fund.

9. REFERENCES

[Ste81] Stephen M. Platt and Norman I. Badler. "Animating facial expressions." SIGGRAPH Computer Graphics, 15(3):245–252, 1981.

[Wat87] K. Waters. "A muscle model for animation three-dimensional facial expression." International Conference on Computer Graphics and Interactive Techniques, pages 17–24, 1987.

[Gue98] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin. "Making faces." Annual Conference on Computer Graphics and Interactive Techniques, pages 55–66, 1998.

[Pig98] F. Pighin, J. Hecker, D. Lischinski, and R. Szeliski. "Synthesizing realistic facial expressions from photographs". In SIGGRAPH Computer graphics, pages 75–84, Orlando; FL, 1998. ACM Press.

[Fas02] B. Fasel and J. Luettin. "Automatic facial expression analysis: a survey". IDIAP Research Report 99-19, 2002.

[Ter90] D. Terzopoulos and K.Waters. "Physically-based facial modeling, analysis and animation". Journal on Visualisation and Computer Animation, 1(2):73–80, 1990.

[Lee95] Y. Lee, D. Terzopoulos, and K. Walters. "Realistic modeling for facial animation". In SIGGRAPH on Computer Graphics and Interactive Techniques, pages 55–62, New York, NY, USA, 1995.

[Fox05] B. Fox. "Barrett fox character animator". Internet: http://www.barrettfox.com/, 2005.

[Sif05] E. Sifakis, I. Neverov, and R. Fedkiw. "Automatic determination of facial muscle activations from sparse motion capture marker data." ACM Transactions Graphics, 24(3), 2005.

[Wil97] J. Wilhelms and A.V. Gelder. "Anatomically based modeling." Annual Conference on Computer Graphics and Interactive Techniques, 1997.

[Ers08] N. Ersotelos and F. Dong. "Building highly realistic facial modeling and animation: a survey." The Visual Computer, 24(1):13–30, 2008.

[Sha01] A. Shashua and T. Riklin-Raviv. "The quotient image: class-based re-rendering and recognition with varying illuminations." IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2):129–139, 2001.

[Par96] F.I. Parke and K. Waters. "Computer facial animation". A K Peters, 1996.

[Wil90] L. Williams. "Performance-driven facial animation." Annual Conference on Computer Graphics and Interactive Techniques, pages 235–242, 1990.

[Ber85] P. Bergeron and P. Lachapelle. "Controlling facial expressions and body movements in the computer-generated animated short "tony de peltrie"." In ACM SIGGRAPH Advanced Computer Animation seminar notes, 1985.

[Ess96] I. Essa, S. Basu, T. Darrell, and A. Pentland. "Modeling, tracking and interactive animation of faces and heads using input from video." IEEE International Conference on Computer Animation, 00:68, 1996.

[Jos03] P. Joshi, W.C. Tien, and M. Desbrun. "Learning controls for blend shape based realistic facial animation." ACM SIGGRAPH Eurographics Symposium on Computer Animation, pages 187–192, 2003.

[Bla03] V. Blanz, C. Basso, T. Poggio, and T. Vetter. "Reanimating faces in images and video." Annual Conference of the European Association for Computer Graphics, 22(3):641–650, 2003.

[Vla05] D. Vlasic, M. Brand, H. Pfister, and J. Popovic. "Face transfer with multilinear models." ACM Transactions on Graphics, 24(3), 2005.

[Kon] Konica Minolta, Vivid 910. [Online].
Available: http://www.minoltausa.com

[Pap05] T. Papatheodorou and Daniel Rueckert. "Evaluation of 3d face recognition using registration and PCA." In Takeo Kanade, Anil K. Jain, and Nalini K. Ratha, editors, AVBPA, volume 3546 of Lecture Notes in Computer Science, pages 997-1009. Springer, 2005.

[Vis] VisionRT. [Online]. Available: http://www.visionrt.com

[Wan06] J. Wang and L. Yin and X. Wei and Y. Sun. "Facial expression recognition based on primitive surface feature distribution." in the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, 2006.

[Coo95] Cootes, T., Taylor, C.J., Cooper, D.H., Graham, J. "Active shape models - their training and application." Computer Vision and Image Understanding, 61(1): 3859. 1995.

[Tho06] C.E. Thomaz, E.C. Kitani and D.F. Gilles: "A Maximum uncertainty LDA-based approach for limited size problems with applications to face recognition." Journal of the Brazilian Computer Society, Vol 12(2), pg 7-18, 2006.

[Kit06] Kitani, E.C., Thomaz C.E., Gillies, D.F. "A Statistical discriminant model for face interpretation and reconstruction." In: 19th Brazillian Symposium on Computer Graphics and Image Processing. 2006.

[Lat00] L.D. Lathauwer, B.D. Moor, and J. Vandewalle. "A multilinear singular value decomposition". SIAM Journal on Matrix Analysis and Applications, 21(4):1253–1278, 2000.

[Rao71] C.R. Rao and S. Mitra. "Generalised inverse of matrices and its applications". Wiley New York, 1971.

[Has07] S. Hassan Amin and Duncan Gillies. "Quantitative analysis of 3d face reconstruction using annealing based approach". In: IEEE International Conference on Biometrics: Theory, Applications, and Systems, BTAS 2007, pg 1–6, 2007.