Quantitative analysis on PCA-based statistical 3D face shape modeling

A.Y.A. Maghari, I. Yi Liao & B. Belaton
School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia

ABSTRACT: Principle Component Analysis (PCA)-based statistical 3D face modeling using example faces is a popular technique for modeling 3D faces and has been widely used for 3D face reconstruction and face recognition. The capability of the model to depict a new 3D face depends on the exemplar faces in the training set. Although a few 3D face databases are available to the research community and they have been used for 3D face modeling, there is little work done on rigorous statistical analysis of the models built from these databases. The common factors that are generally concerned are the size of the training set and the different choice of the examples in the training set. In this paper, a case study on USF Human ID 3D database, one of the most popular databases in the field, has been used to study the effect of these factors on the representational power. We found that: 1) the size of the training set increase, the more accurate the model can represent a new face; 2) the increase of the representational power tends to slow down in an exponential manner and achieves saturation when the number of faces is greater than 250. These findings are under assumptions that the 3D faces in the database are randomly chosen and can represent different races and gender with neutral expressions. This analysis can be applied to the database which includes expressions too. A regularized 3D face reconstruction algorithm has also been tested to find out how feature points selection affects the accuracy of the 3D face reconstruction based on the PCA-model.

1 INTRODUCTION

The reconstruction of 3D faces is a very important issue in the fields of computer vision and graphics. The need for 3D face reconstruction has grown in applications like virtual reality simulations, plastic surgery simulations and face recognition (Elyan & Ugail 2007), (Fanany et al. 2002) It has recently received great attention among scholars and researchers (Widanagamaachchi & Dharmaratne 2008). 3D facial reconstruction systems are to recover the three dimensional shape of individuals from their 2D pictures or video sequences. There are many approaches for reconstruction 3D faces from images such as Shape-from-Shading (SFS) (Smith & Hancock 2006), shape from silhouettes (Lee et al. 2003) and shape from motion (Amin & Gillies 2007). There are also learning-based methods, such as neural network (non-statistical). (Nandy & Ben-Arie 2000) and 3D Morphable Model (3DMM) (statistical based) (Volker & Thomas 1999). 3DMM is an analysis-by-synthesis based approach to fit the 3D statistical model to the 2D face image (Widanagamaachchi & Dharmaratne 2008). The presence of 3D scanning technology lead to create a more accurate 3D face model examples (Luximon et al. 2012). Examples based modeling allows more realistically face reconstruction than other methods (Widanagamaachchi & Dharmaratne 2008), (Martin & Yingfeng 2009). However, the quality of face reconstruction using examples is affected by the chosen examples. For example (Kemelmacher-Shlizerman & Basri 2011), and (Jose et al. 2010) urged that learning a generic 3D face model requires large amounts of 3D faces. However, this issue has not been statistically analyzed in terms of representational power of the model, to the best of our knowledge. Although (Iain et al. 2007) has studied the representational power of two example based models, i.e. Active Appearance Model (AAM) and 3DMM, they did not conduct any statistical analysis or testing on the two models.

A PCA-based model with relatively small sample size (100 faces) was used for face recognition and has obtained reasonable results (Volker & Thomas 2003). The generation of synthetic views from 2D input images was not needed. Instead, the recognition was based on the model coefficients which represent intrinsic shape and texture of faces. In case of 3D face reconstruction, a more diverse set of 3D faces would build a more powerful PCA-based model to generate an accurate 3D representation of the 2D face. On the other hand, PCA-based models have not been quantitatively studied in terms of the effect of 3D reconstruction accuracy on face recognition.

Although in some statistical modeling methods both shape and texture are modeled separately using PCA (e.g. 3DMM), it is suggested that shapes are more amenable to PCA based modeling than texture, as texture varies dramatically than the shape. In many situations, the 2D texture can be warped to the 3D
A series of experiments are designed to answer the questions:

1. What is the relationship between the size of training set and the representational power of the model?
2. How many examples will be enough to build a satisfactory model?
3. What is the effect on the representational power if the model is trained with a different sample for the same number of faces?

Finally, the regularization based 3D face reconstruction algorithm was analyzed to find the relationship between the number of feature points and the accuracy of reconstruction.

This paper is organized as follows: section 2 provides a theoretical background. The Experimental design and statistical analysis are explained in Section 3, whereas in Section 4 the findings and discussion are reported. In the last section, the paper is summarized.

2 THEORITICAL BACKGROUND

2.1 Modeling shape using PCA

The 3D face shape model is a linear combination of eigenvectors obtained by applying PCA to training set of 3D shape faces. The linear combination is controlled by shape parameters called \( \alpha \), where shape vectors are given as follows:

\[
 s_j = (x_{i1}, y_{i1}, z_{i1}, \ldots, x_{in}, y_{in}, z_{im})^T
\]

where \( s_j \) has the dimension \( n \times 3 \), \( n \) is the number of vertices and \( i = 1, \ldots, m \) (number of face shapes).

All vertices must be fully corresponded using their semantic position before applying PCA to get a more compact shape representation of 3D shape face. Let

\[
 s_0 = \frac{1}{m} \sum_{i=1}^{m} s_i
\]

where \( s_0 \) be the average face shape of \( m \) exemplar face shapes and \( S = \{ s_1, s_2, \ldots, s_m \} \in \mathbb{R}^{3 \times m} \).

The covariance matrix of the face shapes is defined as

\[
 C = \sum_{i=1}^{m} (s_i - s_0)(s_i - s_0)^T
\]

The eigenvectors \( e_i \) and the corresponding eigenvalues \( \lambda_i \) of the covariance matrix \( C \) are such that

\[
 C e_i = \lambda_i e_i \quad \text{where} \quad i = 1, \ldots, m.
\]

After PCA modeling, every shape of the \( m \) face shapes can be decomposed into the form

\[
 s_j = s_0 + \sum_{i=1}^{m} \alpha_i e_i
\]

where \( e_i \) represent the \( i \)-th eigenvector of the covariance matrix \( C \) and \( \alpha_i \) is the coefficient of the shape eigenvector \( e_i \).

The coefficient of a face shape \( s_j \) can be calculated using the following equation

\[
 \alpha = E^T(s_i - s_0)
\]

where \( E = [e_1, e_2, \ldots, e_m] \) are the eigenvectors of the covariance matrix \( C \). The projected new face shape can be represented by applying Equation 1.

2.2 Representational power of the model

In this study, we define the representational power (RP) as the Euclidean Distance between the reconstructed shape vector and the true shape vector divided by the number of points in the shape vector.

In Cartesian coordinates, if \( p = (p_1, p_2, \ldots, p_n) \) and \( q = (q_1, q_2, \ldots, q_n) \) are two points in Euclidean \( n \)-space, then the Euclidean Distance from \( p \) to \( q \) is given by:

\[
 d = \sum_{i=1}^{n} \sqrt{(p_i - q_i)^2}
\]

Let \( s \) be a shape face belongs to the testing data set and \( s_r \) be the face shape that is represented by PCA-model using Equations 2 and 1 then:

a. Calculate the coefficient of the testing face shape \( s \) using Equation 2.

b. Apply Equation 1 to represent \( s_r \) using the PCA-model.

c. Determine the Euclidean Distance between the true shape face \( s \) and the reconstructed shape face \( s_r \).

RP is the Euclidean Distance divided by shape dimension \( n \times 3 \).

2.3 Reconstruction based on regularization

Since all points of different shape faces in the database are fully corresponded, PCA is used to obtain a more compact shape representation of face with the primary components (Dalong et al. 2005).
Let $t$ be the number of points that can be selected from the 3D face shape in the testing set, $s_f = (p_1, p_2, \ldots, p_t) \in \mathbb{R}^t$ be the set of selected points on the 3D face shape ($p_i$ can be $x$, $y$ or $z$ of any vertex on the 3D face shape, whereas every vertex has 3 axis $x$, $y$ and $z$), $s_{f0} \in \mathbb{R}^t$ is the $t$ corresponding points on $s_0$ (the average face shape) and $E_f \in \mathbb{R}^{t \times m}$ is the $t$ corresponding columns on $E \in \mathbb{R}^{3n \times m}$ (the matrix of row eigenvectors). Then the coefficient $\alpha$ of a new 3D face shape can be derived as

$$\alpha = (E_f^T E_f + \lambda \Lambda^{-1})^{-1} E_f^T (s_f - s_{f0})$$

where $\Lambda$ is a diagonal $m \times m$ matrix with diagonal elements being the eigenvalues and $\lambda$ is the weighting factor. Then we apply $\alpha$ to Equation 1 to obtain the whole 3D face shape.

3 EXPERIMENTAL DESIGN & STATISTICAL ANALYSIS

3.1 USF human ID database

A case study is conducted using USF Raw 3D Face Data Set. This database includes shape and texture information of 100 3D faces obtained by using Cyberware head and face 3D color scanner. The 3D faces are aligned with each other as explained by (Volker & Thomas 1999). They developed the 3D morphable model (3DMM) which has been widely used in many facial reconstruction systems. A basic assumption is that any 3D facial surface can be practically represented by a convex combination of shape and texture vectors of 100 3D face examples, where the shape and texture vectors are given as follows:

$$S_i = (x_{i1}, y_{i1}, z_{i1}, \ldots, x_{i75972}, y_{i75972}, z_{i75972})^T$$

$$T_i = (R_{i1}, G_{i1}, B_{i1}, \ldots, R_{i75972}, G_{i75972}, B_{i75972})^T$$

Information from each of these exemplar heads was saved as shape and texture vectors ($S_i$, $T_i$) where $i = 1, \ldots, 100$ (Martin & Yingfeng 2009). For each face shape there are 75972 vertices saved in a text file, one line per vertex and 3 points each line.

This study uses only the shape vectors for training and testing the models.

3.2 Representational power analysis

RP of 3D PCA-based models is analyzed using USF Human ID 3D database. A series of experiments are designed to find the relationship between the size of the training data set and RP of the trained model. On the other hand, the effect of different training set for the same number of faces has been analyzed.

3.2.1 Size of training set

We divided the current 100 shape faces into different sets in term of the number of samples. For example PCA15 is a shape model that has 15 shape faces as training data, PCA18 is a model with 18 training face shapes and so on until PCA93 with 93 training face shapes. The models with training data between 15 and 70 face shapes were tested with 30 testing shape faces. The other models from 73 to 93 were tested with the remaining faces out of 100 faces. For example the PCA80 is tested with 20 shape faces.

The testing face shape $s$ is projected on the model by calculating the shape coefficient vector $\alpha$ using Equation 2. The projected new face shape $sr$ can be obtained by applying $\alpha$ to Equation 1. Figure 5 in the appendix shows a 3D face shape represented by the model PCA80. RP can then be calculated for all testing face shapes according to the following equation

$$RP = \frac{\sqrt{\sum_{i=1}^{3 \times 75972}(s_i - s_{sr})^2}}{3 \times 75972}$$

Then the mean of the RP "RP-mean" for all test faces is used to represent the RP of the model. Figure 1 shows the relationship between the sample size and RP-mean. It shows that there is exponential relationship between the sample size and the RP-mean. An exponential regression model is applied to fit the curve in Figure 1, as follows

$$y = b_0 \times e^{b_1 x}$$

where $y$ is the RP-mean, $x$ is the sample size and $b_0$ and $b_1$ are the regression factors. Thus,

$$\ln(y) = \ln(b_0) + b_1 x$$

The linear relationship between the sample size $x$ and the natural logarithm $\ln(y)$ of the RP-mean is shown in Figure 2. Two variable linear regression was run using MS Excel to find the two factors $\ln(b_0)$ and $b_1$. The regression result is shown in Table 1.

$$b_0 = \exp(\ln(b_0)) = \exp(-5.1425) = 0.0057849$$

The exponential relation is presented as

$$y = 0.0057849 \times e^{-0.0104k}$$

where $y$ is the representational power (RP-mean) and $x$ is the sample size. Figure 3 shows the functional relationship.
Figure 2. The relation between the sample size and the natural logarithm of representational power.

Table 1. Regression results.

<table>
<thead>
<tr>
<th>Regression statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln(b0)</td>
<td>-5.1525</td>
</tr>
<tr>
<td>b1</td>
<td>-0.01046</td>
</tr>
<tr>
<td>Multiple R</td>
<td>0.974219</td>
</tr>
<tr>
<td>R Square</td>
<td>0.9491019</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.141481178</td>
</tr>
<tr>
<td>Observations</td>
<td>32</td>
</tr>
<tr>
<td>SSE</td>
<td>0.0998032</td>
</tr>
<tr>
<td>SST</td>
<td>1.960840125</td>
</tr>
<tr>
<td>SSR</td>
<td>1.861036941</td>
</tr>
</tbody>
</table>

Figure 3. The functional relation between the assumed sample size and the RP.

3.2.2 Different sets of training data
Four pairs of different models with the same sample size but different training data were trained to see if the variations of the data set influence the representational power of the model.

Table 2 shows the comparative results between each pairs of models with significant $\alpha = 0.05$. We made 4 comparisons and applied t-Test. Two cases showed significant differences in the RP-mean of the models. For example the two learning models PCA40 and PCA40R have been trained with 40 faces from different sets. As illustrated in Table 2, the t-Test value corresponding to the two learning models is 0.01322 which is less than $\alpha = 0.05$. This means that there is a statistically significant difference between the RPs of the two models.

Table 2. Comparative result of model-pairs with different samples.

<table>
<thead>
<tr>
<th>Learning model</th>
<th>RP-mean %</th>
<th>Std%</th>
<th>p-value of t-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA 40</td>
<td>0.371</td>
<td>0.075</td>
<td>0.01322</td>
</tr>
<tr>
<td>PCA 40R</td>
<td>0.343</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>PCA 50</td>
<td>0.32</td>
<td>0.048</td>
<td>0.26031</td>
</tr>
<tr>
<td>PCA 50R</td>
<td>0.313</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>PCA 60</td>
<td>0.289</td>
<td>0.04</td>
<td>0.105</td>
</tr>
<tr>
<td>PCA 60R</td>
<td>0.278</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td>PCA 70</td>
<td>0.272</td>
<td>0.039</td>
<td>0.00857</td>
</tr>
<tr>
<td>PCA 70R</td>
<td>0.251</td>
<td>0.026</td>
<td></td>
</tr>
</tbody>
</table>

3.3 Regularization based reconstruction
The regularized algorithm was categorized as one of the main existing four methods for 3D facial reconstruction (Martin & Yingfeng 2009). The algorithm was used by Dalong (Dalong et al. 2005) to reconstruct 3D face for face recognition purposes. The selected feature points are used to compute the 3D shape coefficients of the eigenvectors using Equation 3. Then, the coefficients are used to reconstruct the 3D face shape using Equation 1. Figure 5 in the appendix shows some examples of reconstructed face shapes using different number of feature points.

3.3.1 Number of feature points
Three models with different dataset sizes 60, 70, 80 face shapes were used to analyze the algorithm. The selected points are between 10 and 500 randomly selected from face shape inside and outside the training set. If the test face is from the training set, any number of selected points greater than or equal the number of training face shapes can reconstruct exact 3D face. If the test face is not from the training set, it hardly reconstructs the exact 3D face. Figure 4 shows the relation between the number of feature points (from faces outside the training data) and the accuracy of the reconstructed face shapes using PCA80 (PCA-based model with 80 training shapes).

The algorithm was analyzed for different values of the weighting factor $\lambda = 0.005, 0.05, 0.1, 1, 10, 20, 50, 100,$ and $200$ as shown in Figure 4. 20 faces outside the training data set are used to test the regularized algorithm. The following steps demonstrate the experiment procedures for each different value of the weighting factor $\lambda$.

a) For each shape we do the following 30 times:
   I. Select randomly a set of feature points.
   II. Apply the reconstruction algorithm and calculate the RP between the original face shape and reconstructed shape.

b) Determine the mean of 50 RP, termed as RP-mean.
Figure 4. The relationship between the number of feature points and the accuracy of the reconstructed face shapes on PCA80 with different values of $\lambda$.

Table 3. ANOVA results with $\lambda = 1$.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Between groups</th>
<th>Within groups</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>5.82E-05</td>
<td>0.000968</td>
<td>0.001027</td>
</tr>
<tr>
<td>DF</td>
<td>49</td>
<td>0.000968</td>
<td>999</td>
</tr>
<tr>
<td>MS</td>
<td>1.19E-06</td>
<td>1.02E-06</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1.165887</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.206254</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F crit</td>
<td>1.367567</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c) Determine the mean of the 20 RP-means (RP-mean Error)
d) RP-mean Error is used to measure the accuracy of the reconstructed shape faces from the set of selected points.

3.3.2 Location of feature points
For each of the 20 face shapes chosen from outside the training set, a number of points were randomly selected 50 times to reconstruct face shapes. The 50 selections of feature points were repeated with each of the weighting factors 0.05, 1, 10, and 100. The reconstruction results of all repeated 50-time selections showed the similarity of RP.

Single factor ANOVA was run using MS Excel to compare the 50 alternatives for each of the selected 30, 40, 50, 60, 80 and 100 feature points. In all cases, the ANOVA results show that there is no significant difference among the 50 feature selections. Table 3 shows sample results using 60 feature points with $\lambda = 1$.

The ANOVA results in Table 3 show that there is no statistically significant difference among the 50 alternatives of 60 points selection, as the F-value is smaller than the tabulated critical value (F crit).

4 FINDINGS AND DISCUSSION

4.1 Representational Power
The experimental results in Figure 1 show that there is an exponential relationship between the sample size and the RP mean. To determine this relation we applied linear regression on the sample size and the natural logarithm of the RP mean. Based on this relationship, a minimum number of 250 shape faces is expected to build a satisfactory 3D face model, see Figure 3. This suggests that we may need a large amount of 3D faces to build a generic 3D faces model and this is consistent with the comments made in (Kemelmacher-Shlizerman & Basri 2011), (Jose et al. 2010).

Note that there may be other methods to compensate for this need if large data is not available. However, this issue is beyond the topic of this study.

4.2 The dataset variations & PCA-based Models
The results show that as far as the small sample size training set is concerned, in many cases the representational power may vary if we train the model with a different sample. It was inferred that the representational power of the model may be different if the training data set is changed. Table 2 shows the results of t-test conducted on four cases. Two cases (PCA40, PCA40R) and (PCA70, PCA70R) show significant differences in the RP, whereas the p-value of t-test is less than $\alpha = 0.05$ in the two cases.

4.3 Feature points selection for reconstruction

For the selection of feature points, we found that, regardless of the value of the weighting factor $\lambda$, the accuracy of reconstruction by a large number of randomly selected points (greater than 200 points) is relatively the same in all cases even with different locations of points on the face. However, if the number of feature points is equal to the number of training faces, the produced face will have the most inaccurate shape particularly when $\lambda \leq 1$. This is because the algorithm became unstable when the number of training faces is almost the same as the number of feature points. When $\lambda > 20$ and the number of selected points is greater than the number of training faces, the accuracy of reconstruction is relatively the same in all cases and is further associated with slight improvement related to the larger number of selected points as shown in Figure 4.

Furthermore, we found that different locations of a consistent number of feature points have no significant quantitative effect on the reconstruction results. Whether the feature points are dependent or independent of each other may be an issue of interest that needs to be further experimented, the matter which is beyond the scope of this study.

5 SUMMARY AND FUTURE WORK
The representational power of the 3D shape model which is based on PCA was evaluated by analyzing the 3D face database we obtained from University South Florida (USF) with a series of experiments and statistical analysis. The current USF Human ID 3D database
has 100 faces. All 100 faces were used for training and testing purposes. The functional relationship between the number of independent faces in the training data set and the representational power of the model were estimated. Based on this relationship, a minimum number of 250 shape faces is suggested to build a satisfactory 3D face model, especially if the faces are neither too identical nor too different in their shapes. The findings of this relationship show that we need to have a way to increase the representational power of the model by involving more face or predicting the increasing behavior of the model based on increasing the number of faces.

On the other hand, one type of regularization-based 3D face reconstruction algorithm was analyzed to find the relationship between the number of the feature points and the accuracy of the reconstructed 3D face shape. The extensive experimental results showed that if the test face is from the training set, then any set of any number greater than or equal to the number of training faces -1 can reconstruct exact 3D face. If the test face does not belong to the training set, it will hardly reconstruct the exact 3D face using 3D PAC-based models. However, it could reconstruct an approximate face depending on the number of feature points and the weighting factor. This is consistent with the finding on the representational power of the current database. This indicates that we may choose any set of feature points from the most reliable part of the face as long as they are independent of each other, rather than the ones that are from the parts undergoing the facial expressions. Further studies are suggested to test this idea.

ACKNOWLEDGMENT

The work presented in this paper is sponsored by RU grant 1001/PKCOMP/817055, Universiti Sains Malaysia.

APPENDIX

![Figure 5](image.png)

Figure 5. (a) test face. (b) and (c) are reconstructions from 55 randomly selected points. (e) and (f) are reconstructions from 250 points. (d) PCA representation of (a).

REFERENCES


