PCA-based Reconstruction of 3D Face shapes using Tikhonov Regularization

Ashraf Y. A. Maghari, Ibrahim Venkat, Iman Yi Liao and Bahari Belaton

School of Computer Sciences, Universiti Sains Malaysia, Pinang, Malaysia
e-mail: myashraf2@gmail.com

School of Computer Sciences, Universiti Sains Malaysia, Pinang, Malaysia
e-mail: ibrahim@cs.usm.my

School of Computer Science, University of Nottingham Malaysia Campus
e-mail: Iman.Liao@nottingham.edu.my

School of Computer Sciences, Universiti Sains Malaysia, Pinang, Malaysia
e-mail: bahari@cs.usm.my

Abstract

Reconstructing a 3D face shape from a limited number of 2D facial feature points is considered as an ill-posed problem which can be solved using regularization. Tikhonov regularization is a popular method that incorporates prior information towards providing the existence of closed-form solutions which we obtain as a result of applying PCA, in order to solve the ill-posed problem. The common factors that generally affect Tikhonov regularization are the regularization matrix, the number of feature points, regularization parameters and noise. In this study we report our findings on how various factors influence the reconstruction accuracy based on a case study performed on the USF Human ID 3D database. Further, a statistical comparison between two Tikhonov regularization matrices viz., the identity matrix and the diagonal matrix comprising of the eigenvalues (eigenvalue matrix), has been performed. We found that, the reconstruction error can be reduced significantly by using the later one. Finally our study aids to determine the most feasible interval in conjunction with optimal regularization parameters which would lead towards achieving accurate and plausible solutions.

Keywords: Tikhonov Regularization, 3D face reconstruction, PCA, Tikhonov Regularization.
1 Introduction

3D shape provides properties that are invariant to changes such as view point, illumination, background clutter and occlusion. These properties can help in applications like recognition, animation, medical application and allow prediction of appearance under new viewing conditions [1]. A major problem in computer vision is to extract such properties given a single 2D image.

There are many approaches for the reconstruction of 3D faces from single images. One of such early techniques being utilized is Shape-from-Shading (SFS) [2-4], which capitalizes the idea that the depth information is related to the intensity of a face image acquired through a given/chosen reflectance model. SFS estimates the illumination direction in the 2D image to infer the 3D shape of the surface. It has been shown that SFS suffers from poor global shape control. Recently, Kemelmacher-Shlizerman and Basri proposed an approach that combines shading information with generic shape information derived from a single reference model by utilizing the global similarity of faces [1]. This method uses only a single reference model of a different person's face to reconstruct the 3D face shape. It does not need a learning stage to build a model for representing input faces. Consequently, not counting on a 3D reference model which keeps shape similarities with the input image may result in inaccurate 3D shape estimation.

There are also conventional learning-based methods, such as neural networks [5, 6] and typical statistical learning-based methods such as Hidden Markov Model (HMM) [7], Markov Random Field (MRF) [8] and analysis by synthesis using 3D Morphable Model (3DMM) [9]. Interestingly computational intelligence based approaches such as neural networks have also been applied in related computer vision areas such as medical image segmentation [10, 11].

In the last few decades much interest has been shown in the area of extracting 3D surfaces from observed 2D images by using statistical models which allows more realistic face reconstruction than other methods (e. g. [12] & [13]). These models can be used as prior information which can be incorporated with a fitting algorithm to estimate the complete 3D face shape from the given information such as a set of facial feature points. One of the reconstruction methods that uses prior knowledge to estimate the shape coefficients from a set of facial points is regularization [14]. In [14], Jiang et al. use a regularization equation that estimate the geometry coefficient in an iterative procedure. Alternatively regularization method has also been presented in [15].

Reconstructing the 3D face shape from a small set of 2D feature coordinates is considered as an ill-posed problem which can be solved by using regularization. One of the most appropriate methods that can be employed is the Tikhonov regularization [16]. In the last few decades, Tikhonov regularization has been well addressed in the field of inverse problems [17]. It is a popular and effective
method that can be easily incorporated with prior information embedded in a closed-form solution which can be readily obtained by applying PCA on the 3D face shapes. Regularization is necessary for robust, stable, and plausible reconstruction. It finds a tradeoff between fitting the 3D shape to the available facial feature points and producing plausible results in terms of prior information about the 3D face object [18]. Importantly reconstruction by means of a Tikhonov regularization method can be computed in one step (non-iterative way), thereby enabling faster 3D reconstructions. Hence, Tikhonov method is an efficient choice for several 3D oriented interactive tools.

The common factors that generally affect Tikhonov regularization are the Tikhonov regularization matrix (stabilizing item), the number of feature points, regularization parameters and noise. The objective of this contribution is to study the effects of these factors on the reconstruction accuracy and plausibility.

Blanz and Vetter [15] have studied the effect of the regularization parameters and the noise on the accuracy with the aid of a Bayesian approach based regularization technique. The effect of feature points on 3D face shape reconstruction has been studied by Maghari et al. [19] using the Tikhonov Method that uses the eigenvalues as the regularization matrix. In this study we aim to estimate the effect of the other factors without varying the number of feature points.

When the number of feature points available are large with respect to the degrees of freedom of the model, approximate solution tends to the true function when the number of feature points tends to infinity [20]. However, in the real application, in addition to the noise which depends on the acquisition system, the available feature points are limited. Hence, this study uses a small set of feature points (25 2D points). By small number of feature points, overfitting can easily occur. To avoid overfitting, regularization that enforces the results to be plausible according to the prior information is needed. The regularization mechanism needs to find a tradeoff between fitting 3D shape to the given 2D facial landmarks and producing plausible solution in terms of prior knowledge [18]. Initially a PCA-based 3D face shape model is incorporated to gain prior information about the 3D face shape. Then the Tikhonov method is deployed to show how a complete face shape can be estimated from a limited number of feature points. The PCA-based 3D shape learning model relies on examples of 3D scans from the USF Human ID 3D Face database [9].

In this paper we will study the effect of regularization parameters and the noise on the Tikhonov regularization with two different Tikhonov matrices. We will refer to the Tikhonov method that uses the identity matrix as STR (Standard Tikhonov Regularization) and the one that uses the diagonal eigenvalue matrix as ETR (Eigenvalue Tikhonov Regularization). We will also refer to the matrix that has the eigenvalues in its diagonal as eigenvalue matrix. ETR has been compared with STR in terms of accuracy, plausibility and robustness to noise. As part of this
contribution, we demonstrate how to determine the best interval for the regularization parameter that produces the best solution.

This paper is organized as follows: Section 2 provides details on the PCA-based shape face representation. Reconstruction approaches based on regularization algorithms are briefed in Section 3. Section 4 describes distance-based evaluation procedures. In section 5, the regularization parameter interval is demonstrated. Experiments and statistical analysis are reported in section 6 and finally section 7 concludes the outcome of our paper.

2 PCA-based 3D face shape representation

In order to build a PCA based face shape model, PCA decomposition has been applied to the 3D face shapes offered by the USF Human ID 3D database [10]. Each training face shape is represented using the 3D coordinates of all its vertices in terms of a triangulated mesh. Then these 3D shapes are aligned with each other in such a way that a 3D-3D correspondence for all the vertices are obtained [9]. Each face shape has \( p \) corresponding vertices which can be vectorized as:

\[
\mathbf{s}_i = (x_{i1}, y_{i1}, z_{i1}, \ldots, x_{ip}, y_{ip}, z_{ip})^T,
\]

where \( s_i \) has the dimension \( n = 3 \times p \), \( n \) is the number of vertices and \( i = 1, \ldots, m \), where \( m \) represent the number of training shapes. Based on this vectorization, the mean face shape among the training face shapes is computed and the deviation of each training shape from the mean is calculated. Singular Value Decomposition (SVD) [21] is then used to perform PCA on the covariance matrix resulting in a morphable model.

A new shape vector \( \mathbf{s}_{\text{rec}} \in \mathbb{R}^n \) can be expressed as

\[
\mathbf{s}_{\text{rec}} = \mathbf{s}_0 + E\alpha = \mathbf{s}_0 + \sum_{i=1}^m \alpha_i \mathbf{e}_i,
\]

where \( \mathbf{s}_0 \) is the mean 3D shape, \( E = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m] \) is a basis vector matrix and \( \alpha_i \) is the coefficient of the scaled basis vector \( \mathbf{e}_i \). Assuming \( \mathbf{x} = \mathbf{s} - \mathbf{s}_0 \in \mathbb{R}^n \), Eq. (2) can be minimized to

\[
\mathbf{x}_{\text{rec}} = E\alpha.
\]

Since \( E \) is an orthogonal matrix, the PCA-coefficients \( \alpha \) of a vector \( \mathbf{x} \) can be derived from Eq. (3) as
\[ \alpha = E^T x. \quad (4) \]

A new 3D face shape can be approximately represented by making use of Eq. (4) and Eq. (2). Taking into account the data uncertainties and the few number of available feature coordinates \( f << p \) on the real 2D face images, Eq. (3) becomes

\[ x_f = A\alpha + \varepsilon. \quad (5) \]

Here, \( A \in \mathbb{R}^{lm} \) is the corresponding subset of the matrix of scaled basis vectors \( E \in \mathbb{R}^{nm} \) and \( \varepsilon \in \mathbb{R}^l \) can be considered as measurement errors with unknown properties. \( x_f = s_f - s_{0f}, s_f \in \mathbb{R}^l(\ell = 2f) \) and \( s_{0f} \) are corresponding points on \( s_0 \) (the average 3D face shape).

The primary least square (LS) method can be used to solve for \( \alpha \)

\[ \alpha = (A^T A)^{-1} A^T x_f. \quad (6) \]

However, in addition to the measuring errors, the measured feature points \( x_f \) captures only a small portion of the original image \( x \), which introduces errors in the recovered model. This problem is considered as an ill-posed problem. Therefore, we see that, regularization is necessary for a robust and plausible solution, which finds a tradeoff between plausible results (in terms of the prior information) and fitting the 3D shape to the given facial points.

### 3 Regularization based Reconstruction

The standard Tikhonov regularized solution of Eq. (5) is to minimize the following function:

\[ \alpha_{reg} = \arg \min_{\alpha} \{ \| A\alpha - x \|^2 + \lambda \| \alpha \|^2 \}. \quad (7) \]

Here, \( \| \alpha \|^2 \) is the stabilizing term and \( \lambda > 0 \) is called the regularization parameter which is chosen to balance the data fitting error so as to get a plausible solution. A standard regularized approximation for \( \alpha \) is given by

\[ \alpha_{reg} = (A^T A + \lambda I)^{-1} A^T (x_f), \quad (8) \]

where \( I \) is the identity matrix.
Different regularization matrices have been employed in order to improve the standard Tikhonov regularization. Obviously, different reconstruction accuracies will be obtained by using different regularization matrices. The objective function in Eq. (7) can be generalized as:

$$\alpha_{reg} = \arg \min_\alpha \{ \| A\alpha - x \|^2 + \lambda \| \alpha - \alpha_0 \|^2 \}. \quad (9)$$

According to the Bayesian posterior principle, $\alpha_0$ is the expected value of $\alpha$ which can be computed as the average vector comprising of all the face shape parameter vectors. Eq. (9) has an explicit optimal solution for $\alpha_{reg}$ and is given by

$$\alpha_{reg} = \alpha_0 + (A^T A + \lambda I) \alpha_0 (x_f - A\alpha_0). \quad (10)$$

Owing to the Bayes’ theorem and under the assumption that the data matrix $X = S - S_0$ is a multivariate normal distribution and the errors in $x_f$ are independent with zero mean, the Tikhonov-regularized solution yields the most probable solution [22]. Let $C$ be the inverse of the covariance matrix of $\alpha$, then according to the maximum posterior probability, the Tikhonov matrix can be given as $C = L^T L$. The covariance matrix is the diagonalised matrix with eigenvalues. So, the stabilizing item is chosen to be the inverse of the diagonal eigenvalue matrix $W$. Assume that $\alpha_0 = 0$, then model parameter $\alpha$ can be estimated as

$$\hat{\alpha} = (A^T A + \lambda W^{-1})^{-1} A^T x_f. \quad (11)$$

Then, a new face shape $s_{rec}$ can be obtained by applying $\hat{\alpha}$ to Eq. (2). In the next section we will discuss how to evaluate the Tikhonov method that uses the eigenvalue matrix as a regularization matrix (ETR: Eq. (11)) compared with the standard Tikhonov (STR: Eq. (8)).

4 Distance-based Evaluation

The evaluation is based on computing the average distance between the points of reconstructed face shape vector and the ground-truth vector

$$E_s = \frac{1}{n} \sum_{i=1}^{n} \| s_{i,orig} - s_{i,recon} \|. \quad (12)$$

To compare between the STR (Eq. (8)) and ETR (Eq. (11)), new test face shapes are reconstructed from a set of feature points using both the methods. Then, according to Eq. (12), for every test face shape, the minimum $E_s$ is computed by
the optimal regularization parameter $\lambda$. The 3D faces in the database are already aligned with each other, which mean that the points can be easily selected by their indices. A set of $f = 25$ salient feature points such as eye corners, eyebrows, mouth corners, nose and face contours have been selected from the face model. These points can be directly used to compute the 3D shape coefficient $\alpha$ using Eq. (8) or Eq. (11). Then, the coefficients are used to reconstruct the 3D face shape using Eq. (2). The resulted face shape is compared with the original shape by calculating $E_r$.

ETR is evaluated quantitatively in order to compare it to the STR. By quantitative evaluation, $E_r$ of the reconstructed 3D face shape has been computed for all the test face shapes. Then we have applied the t-Test to compare the effect of the two matrices on the reconstruction error. The t-Test is a statistical test which allows the comparison of any two methods on objective terms with knowledge of risks associated with reaching the wrong conclusion. On the other hand, examples of reconstructed 3D face shapes have been visualized to clearly justify the qualitative features of the reconstructed face shapes. This is to check how well the qualitative evaluation complies to the results which have been obtained using quantitative techniques. Further, we compute the Euclidean distance between the reconstructed face shape and the mean face to determine how close is the solution to the mean face using Eq.13.

$$D_m = \frac{1}{n} \sum_{i=1}^{n} ||s_{0i} - s_{i,\text{reconst}}||.$$  \hspace{1cm} (13)

5 Regularization Parameter Interval

In this section, we examine how to determine the interval for $\lambda$ that reduces the $E_r$ between the reconstructed face shape and its ground truth while producing a plausible solution in terms of prior knowledge. Further, to avoid the excessive smoothness of the solution (the highly closeness to the mean face), $D_m$ between the reconstructed face and the mean face is computed. By this way, we can find out the best interval for $\lambda$ that could produce a plausible but not necessarily a very smooth solution. The quality of reconstructed faces can be controlled by $\lambda$ as shown in Fig. 1. Low values of $\lambda$ can lead to overfitting and high values leads to the mean face. For example, by applying STR (Ps. see upper row of Fig. 2), when $\lambda > 0.01$, $D_m$ converges to zero and the results becomes closer to the mean face; However we see that $E_r$ grows and overfitting occurs when $\lambda$ converges to 0. According to the results shown in Fig. 1 (right) where the features are noise free, we can conclude that a good interval for an optimal $\lambda$, in terms of the minimum average of $E_r$ (for all 20 test faces) and plausible solution, is located in the interval $(0.000001, 0.001)$. However we can see that, for noisy feature points the
best interval can be (0.0001, 0.001) according to the average reconstruction errors $E_r$ shown in Fig. 1 (left).

**Fig. 1.** The effect of $\lambda$ on the average reconstruction errors $E_r$ and the distance from the mean face for 20 test faces, for a given set of 25 feature points using the Standard Tikhonov Regularization method (STR).

For ETR, the association to $\lambda$ has the same manner as experienced by the STR but with a different set of values. Fig. 2 (lower row) shows reconstructed face shapes for the same probe face shapes with different values of $\lambda$. For smaller values of $\lambda$, the face shape becomes far from the original shape, $E_r$ gets larger and also the face is distorted when $\lambda$ converges to 0. In contrast, the shape becomes more smooth when $E_r$ decreases, and reach its minimum at $\lambda = 10$. For all 20 test faces, the values of $\lambda$ in the interval (10, 1000) produce the best average of $E_r$ values with slight convergence to the average face as $\lambda$ increases (see Fig. 3 (left)). If $\lambda >$
1000, $E_r$ increases and $D_m$ remarkably decreases until $\lambda = 100000$; further the reconstructed face becomes the mean face for very large $\lambda$ values whereas $D_m$ converges to 0.

In terms of plausibility and closeness to the ground truth, the best interval has been suggested to be within the interval (10, 1000), whereas the average $E_r$ for all test faces has the minimum value. For noisy feature points, the best interval can be (100, 10000) whereas the average reconstruction errors $E_r$ has the minimum values and $D_m$ is large enough for non successive smooth solution (see Fig. 3, right).

Fig. 3. Using ETR, the effect of $\lambda$ on the average reconstruction errors $E_r$ and the distance from the mean face for 20 test faces, given 25 feature points.

6 Experiments and Statistical Analysis

Let us recall that, this work aims to explore the influence of the regularization matrices with respect to the accuracy of 3D face shape reconstruction by using Tikhonov regularization method. We have evaluated the accuracy of reconstruction for both of the standard identity matrix and the Tikonov’s eigenvalue matrix and compared them. The USF Human ID 3D Face database [9] which contains 100 3D faces has been used for our experimental validations. The PCA model has been trained with the 80 3D face shapes. The remaining 20 3D face shapes have been used for the purpose of testing.

6.1 Reconstruction from noise free feature points

In this section, we analyze how the two regularization matrices perform on the test 3D face shapes that have been chosen from outside the training dataset. The xy coordinates of the selected feature vertices have been used for reconstruction. The results in Fig. 4 shows that the reconstruction errors yielded by reconstruction of 20 test faces using the ETR is minimal when compared to the STR. Moreover, the
average results of the 20 test faces shown in Table 1 indicates that ETR outperforms STR with a 95% confidence level, whereas the P-value of the t-Test corresponding to the two methods is less than 0.05 (level of significance). This indicates that there is a statistically significant difference between the reconstruction errors of the two methods. This also justifies that using ETR yields better accuracy than that of STR. Fig. 5 visualizes the effect of reconstructed 3D face shapes for the same original face (left) by using ETR and STR.

![Graph showing reconstruction error from noise free feature coordinates](image)

**Fig. 4.** Comparison between the ETR and STR in terms of "reconstruction from 25 noise free feature points". \( E_r \) is computed by the optimal regularization parameter \( \lambda \).

**Table 1.** Comparison of results between the reconstruction errors yielded by ETR and STR.

<table>
<thead>
<tr>
<th>( f = 25 ) values( \times 100 )</th>
<th>ETR</th>
<th></th>
<th>STR</th>
<th></th>
<th>t-Test against significant eigenvalues (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
<td>Std.</td>
<td></td>
</tr>
<tr>
<td>Noise free feature points</td>
<td>0.61</td>
<td>0.151</td>
<td>0.92</td>
<td>0.184</td>
<td>1.39E-07</td>
</tr>
<tr>
<td>Noisy feature points</td>
<td>0.75</td>
<td>0.19</td>
<td>1</td>
<td>0.18</td>
<td>1.49E-05</td>
</tr>
</tbody>
</table>
PCA-based Reconstruction of 3D Face

Fig. 5. Reconstruction of 3D face shapes from two different sets of 25 feature points viz., noisy and noise free feature points using the different regularization matrices. The results depend on the optimal $\lambda$ which produces the minimum $E_r$.

6.2 Reconstruction from Noisy feature points

In the previous section, we have demonstrated how the reconstruction error can be reduced by using different regularization matrices. However, if we add random noise to all feature points coordinates, $\alpha$ becomes more difficult to estimate, and overfitting can easily occur. Fig. 6 shows the reconstruction errors found by reconstruction of 20 test faces from 25 feature points using both the matrices. This further justifies that the eigenvalue Tikhonov (ETR) outperforms the standard Tikhonov (STR) in terms of reconstruction accuracy for all test faces. Moreover, the average errors of the 20 test faces shown in Table 1 indicates that the reconstruction errors of ETR are significantly lower than those for STR at 95% confidence level. Fig. 5 visualizes the reconstructed 3D face shapes for the same original face (left) by optimal $\lambda$ using STR and ETR. It can be seen that the quality of reconstructed faces using ETR looks more precise and intuitive than those reconstructed by the STR for noisy and noise free feature points.

Fig. 6. Reconstruction errors generated with the aid of choosing 25 noisy feature points. $E_r$ is computed by the optimal regularization parameter $\lambda$. 
6.3 Reconstruction of 3D Face Shapes from Images

The CMU-PIE database [23] has been used for testing the visual effects of the proposed model. We intend to reconstruct 3D models for the 2D images of CMU-PIE database. From small number of 2D facial landmarks, the proposed algorithm can recover the 3D shape of the given 2D face image. In this work, the input 2D images are in near frontal pose with most of their expression being neutral. The feature points which have been manually selected, have been aligned with the reference 3D model using Procrustes Analysis, which is the usual preliminary step before the reconstruction stage. The aligned feature points have been used to compute the model parameters $\alpha$ using ETR Eq. (11). Then, $\alpha$ has been used to reconstruct the 3D shape using Eq.(2). Fig. 7 shows two typical 2D images (left) and their corresponding reconstructed 3D faces using ETR (right).

For each of these 2D images, 6 reconstructed faces with different distances $D_m$ have been shown. It can be seen that the quality of reconstructed faces gets degraded if $D_m$ increases. In other words, we can infer that the reconstruction of faces gets gradually smoother when $D_m$ decreases. Hence, we can employ the distance from the mean face ($D_m$) to evaluate the 3D faces reconstructed from real 2D face images. According to the results seen in Fig. 7, we can see that an optimal $D_m$ may be the value around 0.001 (the average distances $D_m$ of all 3D faces in the training data set). Assuming the feature points are available, our MATLAB implementation of the ETR algorithm for reconstructing the complete face shape vector takes about one second on a workstation with Intel(R) Xeon(R) CPU E5620 @ 2.40GH which justifies that the ETR based 3D face reconstruction is quite fast and hence feasible to be deployed for real time applications.
7 Conclusion

A major problem in computer vision is to extract 3D shape properties from single 2D images. Example-based models allow more realistic face reconstruction than other methods. In this paper, the standard PCA-model has been used as a holistic model to represent the object class of 3D face shapes, which aids in solving the ill-posed problem of reconstruction of complete 3D face shapes from their 2D images. This holistic model has been incorporated with the Tikhonov method to show how to estimate the 3D shapes of novel faces using a limited number of feature points. The common factors that generally affect Tikhonov regularization are the regularization matrix, the number of feature points, value of regularization parameters $\lambda$, and noise. This work is basically an effective case study to show the effect of the regularization matrices on the quality of reconstructed faces with respect to a small set of feature points (25 2D feature points). The effects of the regularization parameter and noise have been also demonstrated.

Our extensive experimental results justifies that by using ETR, the reconstruction errors could be significantly reduced for all of the test faces when compared to the STR. Further we have shown that ETR outperforms the STR with 95% confident level in terms of reconstruction errors. In terms of plausibility and closeness to the ground truth, our results demonstrate that the quality of reconstructed faces using ETR is better than those reconstructed using STR even in the presence of noise. We have also shown a mechanism of determining the best interval of $\lambda$ in order to find the optimal solution. Our experiments on real 2D images demonstrate that the ETR method was able to reconstruct plausible 3D face shapes from limited number of feature points in an effective way. Importantly as the ETR method incurs less computation time, it can be readily deployed in many real world interactive applications. For our future work, we intend to evaluate the more challenging task of reconstruction of images with varying degrees of poses and expressions.

ACKNOWLEDGMENT

The first author would like to thank the Ministry of Higher Education (MOHE), Malaysia for supporting him with the MIS scholarship. This research is also partly supported by an RU Grant (1001/PKCOMP/817055) offered by Universiti Sains Malaysia and an ERGS grant (203/PKCOMP/6730075) offered by MOHE.
References


PCA-based Reconstruction of 3D Face


