3D Shape and Pose Estimation of Face Images Using the Nonlinear Least-Squares Model

Zhan-li Sun*† and Kin-Man Lam *
* Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University
† Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, China
E-mail: zhlsun2008@yahoo.com, enkmlam@polyu.edu.hk

Abstract—In this paper, we propose an efficient algorithm to reconstruct the 3D structure of a human face from one or a number of its 2D images with different poses. In our algorithm, the nonlinear least-squares model is employed to estimate the 3D structure of the face and the respective poses of the 2D face images concerned by means of the similarity transform. Furthermore, different optimization schemes are presented with regard to the accuracy levels and the training time required. Our algorithm also embeds the symmetrical property of human faces and the regularization term based on linear correlation into the optimization procedure so as to alleviate the sensitivities arising from changes in poses and improve the estimation accuracy of the 3D structure. Experimental results demonstrate the feasibility and efficiency of the proposed methods.

I. INTRODUCTION

3D face models have been widely applied in face image processing, such as face recognition, face tracking and face animation, etc., due to their superior performance compared to 2D models with regard to variations in pose and illumination. However, the high cost and limited applicability of 3D sensing devices are at present still distinct obstacles to acquire sufficient and useful data. As an alternative, the technique of 3D face reconstruction can be adopted to reconstruct 3D face models of individuals from their corresponding 2D images. Generally speaking, an efficient 3D reconstruction algorithm can greatly enhance the capabilities of existing 2D or 3D face recognition systems.

Many algorithms for 3D reconstruction have been proposed so far, such as shape-from-shading [4], structure-from-motion [1], 3D morphable models [9], [15], shape alignment and interpolation method correction[5], etc. Among these methods, the spatial-transformation approach is an important branch [6], [15]. Specifically, in [6], a similarity-transform-based method is proposed to derive the 3D structure of a human face from a group of face images under different poses. The proposed algorithm does not require any prior knowledge of camera calibration, and has no limitation on the possible poses or the scale of the face images. In addition, the method in [6] has been verified that it can be extended to face recognition to alleviate the effect of pose variations. Unfortunately, the genetic algorithm (GA) used to estimate the depth usually encounters a heavy computational burden. Moreover, how to design a reasonable chromosome, how to make a feasible gene operation scheme, and how to adjust the parameters remain difficult problems.

To reduce the computation of the method in [6], the nonlinear least-squares (NLS) model-based methods are proposed in this paper. Furthermore, the symmetry information of face is utilized in the proposed methods to alleviate the sensitivity to the training samples used. In addition, the regularization term, based on linear correlation [12]-[14], is added in the objective function to improve the estimation accuracy of the 3D structure. Experimental results demonstrate the feasibility and efficiency of the proposed methods.

The remainder of the paper is organized as follows. In Section II, we present our proposed algorithms. Experimental results and related discussions are given in Section III, and some concluding remarks are presented in Section IV.

II. METHODOLOGY

A. Formulation of the Nonlinear Least-Squares Model

The shape features, which are represented by the \((x, y)\) coordinates of the facial feature points, are used in our algorithm to estimate the corresponding depth values, i.e. \(z\). Assume that \(n\) feature points are marked on the face images. These feature points can be marked manually or by using any of the facial feature point detection algorithms [7]. \((M_x, M_y, M_z)\) represents the feature points of a frontal-view 3D face model \(M\), and \((q_x, q_y)\) the feature points of a non-frontal-view 2D face \(q\). The rotation matrix \(R\) for \(q\) is given as follows:

\[
R = \begin{bmatrix}
\cos \phi & \sin \phi & 0 \\
-\sin \phi & \cos \phi & 0 \\
0 & 0 & 1 \\
\end{bmatrix} \times \begin{bmatrix}
\cos \psi & 0 & -\sin \psi \\
0 & 1 & 0 \\
\sin \psi & 0 & \cos \psi \\
\end{bmatrix} \times \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \theta & \sin \theta \\
0 & -\sin \theta & \cos \theta \\
\end{bmatrix} = \begin{bmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23} \\
r_{31} & r_{32} & r_{33} \\
\end{bmatrix}, \tag{1}
\]

where the pose parameters \(\phi, \psi\) and \(\theta\) are the rotation angles around the \(x, y\) and \(z\) axes, respectively. The feature point distance between the 2D face image and the corresponding estimated face image obtained by projecting the 3D face model...
onto the 2D plane can be given by
\[
d = \|q - sR_{2 \times 3}M\|^2 \\
= \left\| q_x - s(r_{11}Mx_1 + r_{12}My_1 + r_{13}Mz_1) \right\|^2 \\
= \left\| q_x - s(r_{21}Mx_1 + r_{22}My_1 + r_{23}Mz_1) \right\|^2. \tag{2}
\]

Denoting the vector \( \mathbf{x} = (\phi, \psi, \theta, s, M_{z_1}, \ldots, M_{z_{10}}) \) as the parameter vector, including both the pose parameters and the depth values of the feature points, the similarity measurement \( (\mathbf{q} - sR_{2 \times 3}M) \) in (2) can be rewritten as a vector function as follows:
\[
\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_n(\mathbf{x}), f_{n+1}(\mathbf{x}), \ldots, f_{2n}(\mathbf{x}))^T. \tag{3}
\]
The pose parameters and the depth values can be obtained by minimizing the distance \( d \), i.e.,
\[
\min_{\mathbf{x}} \| \mathbf{f}(\mathbf{x}) \|^2 = \min_{\mathbf{x}} \sum_{i=1}^{2n} f_i^2(\mathbf{x}). \tag{4}
\]

Therefore, the pose and shape estimation problem is formulated as a NLS model. Unlike in [6], besides the pose parameters \( \phi, \psi, \theta, \) and \( s \), the depth values \( M_{z_i} \) are also considered as the variables of the distance measurement in the NLS models. Note that only one non-frontal-view face image is used in the training. When this image has a large pose angle as compared to the frontal-view face, the reconstruction errors of the faces with the large opposite rotation angles to the training sample are generally larger than those of the faces with an identical rotation direction to the training sample. To alleviate this sensitivity, the symmetrical information on human faces is considered in the NLS models.

Fig. 1. The feature point positions in a frontal-view face image.

As shown in Fig. 1, the feature points 10, \ldots, 15 can be regarded as the mirrors of the points 1, \ldots, 6 with respect to the line \( ab \). Considering this symmetry, the depth values of points 10, \ldots, 15 should have approximately equal values to the points 1, \ldots, 6. In the NLS model, the depth values of the symmetrical points are assumed to be equal, i.e., \( M_{z_1} = M_{z_{15}}, \ldots, M_{z_6} = M_{z_{11}} \). Under this assumption, the number of variables in the NLS model decreases from 19 to 13.

To keep the estimated 3D face structure not far away from the real case and to improve the depth estimation accuracy, the linear correlation coefficient \( (c) \) between the estimated depth values of the facial feature points and the depth values of the CANDIDE model is used as a regularization term in the objective function, i.e.,
\[
\min_{\mathbf{x}} \| \mathbf{f}(\mathbf{x})/c \|^2 = \min_{\mathbf{x}} \sum_{i=1}^{2n} (f_i(\mathbf{x})/c)^2. \tag{5}
\]

B. Nonlinear Least-Squares Model Optimization

There are two schemes to estimate the optimal pose parameters and the depth values. When all the parameters are estimated simultaneously, both the pose parameters and the depth values are regarded as the variables to be optimized in the model depicted in (4), i.e., \( \mathbf{x} = (\phi, \psi, \theta, s, M_{z_1}, \ldots, M_{z_{10}}) \).

The optimization can be carried out using the trust-region-reflective (TRR) algorithm [3].

To improve the accuracy, the pose parameters and the depth values can also be estimated in two steps. Assume that the depth values are known, the pose parameters can be estimated by
\[
\min_{\phi, \psi, \theta, s} \| \mathbf{f}(\phi, \psi, \theta, s)/c \|^2 = \min_{\phi, \psi, \theta, s} \sum_{i=1}^{2n} (f_i(\phi, \psi, \theta, s)/c)^2. \tag{6}
\]

Similarly, given the estimated pose parameters, the depth values \( M_{z_i} \) can be estimated by
\[
\min_{M_{z_i}} \| \mathbf{f}(M_{z_1}, \ldots, M_{z_{10}})/c \|^2 = \min_{M_{z_i}} \sum_{i=1}^{2n} (f_i(M_{z_1}, \ldots, M_{z_{10}})/c)^2. \tag{7}
\]

The procedure is then repeated until some criteria are reached. The depth values can also be initially estimated. Usually, there is not much difference between these two cases. Two methods can be used to determine when the loop should be terminated. One is to set the maximum number of iterations, and the other is to check if the objective value difference between two successive iterations is small enough. Given a small positive real number \( \varepsilon \), if the condition
\[
d_{k+1} - d_k < \varepsilon \tag{8}
\]
is satisfied, then the iteration can be terminated. The specific steps of the latter method are given as follows.

\textbf{Step 1:} Initialize the depth values \( M_{z_1}, \ldots, M_{z_{10}} \), set the constant \( \varepsilon \) and the maximum number of iteration \( N_{\text{iter}} \).

\textbf{Step 2:} With the initial depth values, estimate the pose parameters \( \phi, \psi, \theta, s \) by minimizing the objective function in (6) via the TRR algorithm.
Step 3: With the estimated pose parameters, estimate the depth values $M_{\alpha_1}, \cdots, M_{\alpha_n}$ by minimizing the objective function in (7) via the TRR algorithm. Then the estimated depth values are used as the initial depth values of Step 2.

Step 4: Repeat Steps 2 and 3 until either the condition of (8) is satisfied or the iteration number exceeds the maximum iteration number.

Considering the symmetrical information and the optimization scheme, four NLS-based methods are given as follows. To make the expression more convenient, their abbreviations are also given:

- $NLS_I$: two-step optimization without considering the symmetry information;
- $NLS_{IS}$: two-step optimization considering the symmetry information;
- $NLS_S$: one-step optimization without considering the symmetry information;
- $NLS_{SS}$: one-step optimization considering the symmetry information.

If the regularization term is added in the objective function, the abbreviations of the above four methods are denoted as $NLS_IP$, $NLS_{ISP}$, $NLS_SP$ and $NLS_{SSP}$, respectively.

III. EXPERIMENTAL RESULTS

We evaluate the performance of the proposed methods based on two well known databases: the FERET database [10] and the Bosphorus database [11]. All simulations were conducted in MATLAB environment running on an ordinary personal computer.

A. Reconstruction Error Analysis on FERET Database

Thirty subjects from the FERET database [10] are used in the experiments. Taking the subject shown in Fig. 2 as an example, the first experiment is carried out on this subject to serve as an illustration of the proposed methods. In the experiment, the pose parameters and the depth values are all initially set to be zeros. When the face image $b_{\gamma}$ is used to construct the face model, the reconstruction errors of the remaining face images based on each of the eight approaches are shown in Fig. 3. As a comparison, the reconstruction error using the GA method [6] is also given in Fig. 3. It can be seen that the GA method and the various NLS models have very similar reconstruction errors. From Fig. 3, we can see that the reconstruction errors of the faces with big opposite rotation angles to $b_{\gamma}$ are obviously larger than those of the faces with the identical rotation direction to $b_{\gamma}$ for the methods $NLS_I$ and $NLS_S$. Therefore, these two methods are sensitive to the selection of training samples. However, the NLS-based methods incorporating the symmetry information and the regularization term generally have smaller reconstruction errors than the methods that do not utilize them. In addition, the two-step NLS-based methods generally have fewer reconstruction errors than the one-step NLS-based methods when the symmetry information and the regularization term are not used in the optimization. As an example, the pose estimation results obtained by $NLS_{SSP}$ are shown in Fig. 2. These results show that the poses are satisfactorily estimated by using $NLS_{SSP}$. Similar conclusions can also be drawn for different training images and other subjects.

| TABLE I |
|-----------------|----------------|----------------|----------------|----------------|
|                | $GA$           | $NLS_I$        | $NLS_{IP}$     | $NLS_{ISP}$    |
| $NLS_S$        | 0.08           | 0.19           | 0.03           | 0.10           |
| $NLS_{SP}$     |                |                |                |                |
| $NLS_{SS}$     |                |                |                |                |
| $NLS_{SSP}$    |                |                |                |                |

The training times of GA and various NLS methods are tabulated in Table I. The results show that the training time of the GA-based method is far longer than that of the NLS-based methods.

B. Depth Value Estimation Results on Bosphorus Database

Take subject 1 ("bs000") for example, the mean and standard deviation ($\mu \pm std$) of the correlation coefficients of the true depth values and the estimated depth values are given in Table II. From Table II, it can be seen that the correlation coefficients obtained by the NLS-based methods with the regularization term are obviously larger than those obtained by the NLS-based methods without the regularization term. Therefore, the performance is improved significantly when the regularization term is added in the objective function. In addition, we can see that mean values of the correlation
coefficients obtained by $\text{NIS}_{\text{ISP}}$ and $\text{NIS}_{\text{SSP}}$ are higher than those obtained by $\text{NIS}_{\text{IP}}$ and $\text{NIS}_{\text{SP}}$. Moreover, the standard deviations of $\text{NIS}_{\text{ISP}}$ and $\text{NIS}_{\text{SSP}}$ are only approximately a quarter of those of $\text{NIS}_{\text{IP}}$ and $\text{NIS}_{\text{SP}}$. Therefore, the performance and the robustness of the various NLS-based methods can be improved further when the symmetrical information is embedded into the optimization procedure. Similar conclusions can also be drawn from other subjects.

IV. CONCLUSION

In this paper, NLS-based methods are proposed to estimate the 3D shape and pose of face images. The usefulness of iteration, symmetry information and regularization term is investigated and analyzed. Compared to the GA-based method [6], the proposed methods have a comparable performance, while the training times decreases significantly. Experimental results have demonstrated the feasibility and efficiency of the proposed methods.

REFERENCES


