

Depth estimation of face images

¹Soma.chaitanya, ²Koya Sucharitha, ³V. Sridhar

¹M.Tech, Dsce, Snist, Hyderabad, India,

²M.tech, Vlsi design Hyderabad, India

³M.Tech, Dsce, Snist, Hyderabad, India,

Abstract: *In this paper, we propose an efficient algorithm to reconstruct the 3D structure of a human face from one or more of its 2D images with different poses. In our algorithm, the nonlinear least-squares model is first employed to estimate the depth values of facial feature points and the pose of the 2D face image 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 the human face into the optimization procedure, in order to alleviate the sensitivities arising from changes in pose. In addition, the regularization term, based on linear correlation, is added in the objective function to improve the estimation accuracy of the 3D structure. Further, a model-integration method is proposed to improve the depth-estimation accuracy when multiple non frontal view face images are available. Experimental results on the 2D and 3D databases demonstrate the feasibility and efficiency of the proposed methods.*

Index Terms: *3D face reconstruction, face recognition, nonlinear least-squares model.*

I. Introduction:

3D FACE models have been widely applied in face image processing, such as in face recognition, face tracking and face animation, etc., due to their superior performance over 2D models with regard to variations in pose and illumination. A comprehensive review is presented in [1] for 3D and multi-modal 3D + 2D face recognition. Moreover, the work in [2] provides a summary of 3D facial-reconstruction methods for face recognition based on a single 2D training image per person. Currently, two mainstream approaches are usually adopted to create a 3D facial model. One way is to use special equipment, like a 3D scanner, to capture the 3D shapes of human heads. However, the high cost and limited applicability of 3D sensing devices are, at present, distinct obstacles to acquiring sufficient and useful data. As an alternative, 3D face models of individuals can be reconstructed using the techniques based on 2D images, such as video sequences or multi-view photographs [2]. Reconstructing a 3D model of a human face from a monocular video sequence is an important tool that can be used in surveillance and in various multimedia applications. However, the low quality of the input video data generally greatly degrades the reconstruction performance.

When we construct a 3D face model from multi-view images or photographs, the design of an efficient reconstruction model and the utilization of the available prior information are two critical issues to be considered. Generally speaking, an efficient 3D reconstruction algorithm can greatly enhance the capabilities of existing 2D or 3D face recognition systems.

A 3D shape can be expressed in several ways, e.g., depth, surface normal, surface gradient, or surface slant. The depth value can be considered either as the relative distance from the camera to the surface points, or the relative surface height above the x-y plane. Many algorithms for 3D reconstruction have been developed so far, such as shape-from-shading (SFS) the 3D morphable model [3]–[4], structure from-motion [2] etc.

By exploring the shading information in an image, e.g., the intensity and its derivatives, SFS deals with the recovery of shape based on some reflectance models, such as the Lambertian model, specular reflectance model, and hybrid model, etc. For the SFS algorithm, although various constraints, including brightness, smoothness, and integrability, are explored in sequence, the uniqueness or existence is still an intractable problem in obtaining a correct solution [5].

A 3D morphable model is generally built from a set of 3D laser-scanned heads. As a crucial step, the scans are first registered in a dense point-by-point correspondence, using an optical-flow algorithm to reduce artifact [6]. Statistical signal-processing techniques, such as principal component analysis (PCA), is then applied on the shape and texture features of these training samples to obtain a feature subspace [3], [6]. The feature subspace, including the shape and texture feature vectors, can be regarded as a generic 3D face model. Given the model, a realistic human face can be represented as a convex combination of the shape and texture vectors. Structure-from-motion (SFM) is a popular approach to recover the 3D shape of an object when multiple frames of an image sequence are available. Given a set of observations of 2D feature points, SFM can estimate the 3D structure of the feature points. After proving the rank-3 theorem, i.e., the rank of the observation matrix is 3 under an orthographic projection, a robust factorization algorithm was proposed in [7] to factor the observation

matrix into a shape matrix and a motion matrix using the singular value decomposition (SVD) technique. By considering the observations as mixing

signals, a novel algorithm for maximizing the posterior shape was developed in [1] to estimate the shape from a perspective of blind source separation (BSS)., a 3D object is assumed to be non-rigid, and the observed shapes are represented as a linear combination of a few basis shapes. Further, a Gaussian prior is assumed for the shape coefficients, and the optimization is solved using the expectation-maximization (EM) algorithm. Instead of using an object-dependent shape basis, a kind of novel object independent shape basis-trajectory basis produced by the discrete cosine transform (DCT) - was introduced to reduce the number of unknowns and to increase the stability of the estimation.

One difference between the above three techniques is that the information utilized is different. On the other hand, from the viewpoint of the number of data points, SFS and the 3D morphable model are used to recover the whole surface of an object, i.e. a dense 3D representation, while SFM is generally adopted to estimate the depth values of some feature points, i.e. a sparse 3D representation. Therefore, SFM has a far smaller storage requirement than SFS and the 3D morphable model, which is very helpful for real-time applications. However, when they are applied to face recognition, SFS and the 3D morphable model generally have a better recognition performance than SFM because the information about more feature points is utilized. Recently, some hybrid algorithms have been proposed to alleviate the drawbacks of any one algorithm. In particular, instead of using a 3D database, in a dense 3D generic model is built from 2D hand-labeled images by combining a novel incremental SFM approach, in which the prior 3D shape information is incorporated, with a model density method using a thin plate spline. Since no 3D database is not required, some limitations with learning a generic model in the 3D morphable model are no longer an obstacle. Applications on the CMU-PIE database demonstrate that the 3D face model obtained by the incremental SFM approach has a good generalization performance with respect to expression and identity.

Among various SFM methods, the spatial-transformation approach is one important branch. Specifically, in [2], 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 [2] 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 [2], the nonlinear least-squares (NLS) model-based methods are proposed in this paper to estimate the depth values of facial feature points, i.e., a sparse 3D face representation. In the NLS model, not only the pose parameters, but also the depth values of the facial feature points, are considered as the variables to be optimized. In addition, the symmetry information of face is utilized

further in the proposed methods in order to alleviate the sensitivity to the training samples used. Note that one frontal view face image and one non-frontal-view face image are sufficient to reconstruct a 3D face model using our proposed algorithms. And for cases when multiple non-frontal-view face images are available, a model-integration approach is proposed

in this paper to improve the depth estimation accuracy.

The CANDIDE 3D face model is a parameterized face mask specifically developed for the model-based coding of human faces. During the past several decades, CANDIDE has been a popular face model used in different face-related applications, because of its simplicity and public availability [2]. As, the CANDIDE model is also used in our algorithms as the initial 3D face structure for 3D face reconstruction. To further improve the robustness of our algorithm, Pearson's linear correlation coefficient is employed, which is a common measure of the correlation between two variables. Linear correlation has been used as an important criterion

to measure the similarity between the true signal and the estimated signal. As the CANDIDE model is a generic model, a real face should have a similar 3D face structure to this model. Based on this consideration, to keep the estimated 3D face structure similar to this generic model and to improve the accuracy and robustness of depth estimation, the linear correlation coefficient between the estimated depth values of the facial feature points and the corresponding depth values of the CANDIDE model is used as a regularization term in the objective function. Therefore, different from [2], the CANDIDE model is not only used in the initialization, but also in regularizing the estimated results in our proposed algorithms. A trimmed mean or truncated mean is a statistical measure of central tendency. This calculates the mean after discarding a given percentage of the smallest and the largest observations. The trimmed mean is a robust estimator because it is less sensitive to outliers than the mean, but still gives a reasonable estimate of central tendency or mean. To improve the depth estimation accuracy level, a model-integration method based on trimmed means is further proposed for the case of multiple non-frontal-view face images. In general, the work of this paper can be summarized as follows. First,

a NLS model is developed to estimate the depth values of some important feature points of face images by means of the similarity transform. Moreover, we show that the established 3D face model can be used to estimate the pose of the 2D face image concerned, and used for face recognition. Second, the symmetrical property of the human face is utilized in the optimization procedure in order to alleviate the sensitivities arising from changes in pose. Further, different optimization schemes are investigated with regard to the accuracy levels and the training time required. Moreover, the linear correlation between the estimated depth values and those of the feature points of the CANDIDE model is explored as a regularization term in the objective function to improve the depth-estimation accuracy for face images. Third, a model-integration method is developed to further improve the depth-estimation accuracy when multiple non-frontal-view face images are available. Experimental results on 2D and 3D databases 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 concluding remarks are presented in Section IV.

II. Methodology

The below expressions are collected for calculating depth estimation from different papers.

A. Formulation of the NLS Model

Nonlinear least squares is the problem of finding a set of optimal values of the parameters $\mathbf{x} = (x_1, \dots, x_k)$, which minimize the square sum of nonlinear functions $f_i(\mathbf{x})(i=1, \dots, l)$

$$\min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x})\|_2^2 = \min(f^2_1(\mathbf{x}) + f^2_2(\mathbf{x}) + \dots + f^2_l(\mathbf{x})) \quad (1)$$

Where $\mathbf{f}(\mathbf{x})$ is a vector-valued function with component i of $\mathbf{f}(\mathbf{x})$ equals to $f_i(\mathbf{x})$. 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. (M_{xi}, M_{yi}, M_{zi}) represents the i th feature point of a frontalview 3D face model \mathbf{M} , and (q_{xi}, q_{yi}) the i th feature point of a non-frontal-view 2D face \mathbf{q} . The rotation matrix R for \mathbf{q} is given as follows:

$$R = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos \psi & 0 & -\sin \psi \\ 0 & 1 & 0 \\ \sin \psi & 0 & \cos \psi \end{bmatrix} \\ \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (2)$$

Where the pose parameters ϕ, ψ and θ are the rotation angles around the x, y and z axes, respectively. The distance between the feature points, \mathbf{q} , of the 2D face image concerned and the corresponding projection points, \mathbf{M} , of the 3D face model can be given by

$$d = \|\mathbf{q} - s\mathbf{R}_{2 \times 3}\mathbf{M}\|_2^2 \\ = \left\| \begin{bmatrix} q_{x_1} - s(r_{11}M_{x_1} + r_{12}M_{y_1} + r_{13}M_{z_1}) \\ \vdots \\ q_{x_n} - s(r_{11}M_{x_n} + r_{12}M_{y_n} + r_{13}M_{z_n}) \\ q_{y_1} - s(r_{21}M_{x_1} + r_{22}M_{y_1} + r_{23}M_{z_1}) \\ \vdots \\ q_{y_n} - s(r_{21}M_{x_n} + r_{22}M_{y_n} + r_{23}M_{z_n}) \end{bmatrix} \right\|_2^2 \quad (3)$$

Where s is the scale parameter. Denoting the vector $\mathbf{x} = (\phi, \psi, \theta, s, M_{z_1}, \dots, M_{z_n})$ as the parameter vector, including both the pose parameters and the depth values of the feature

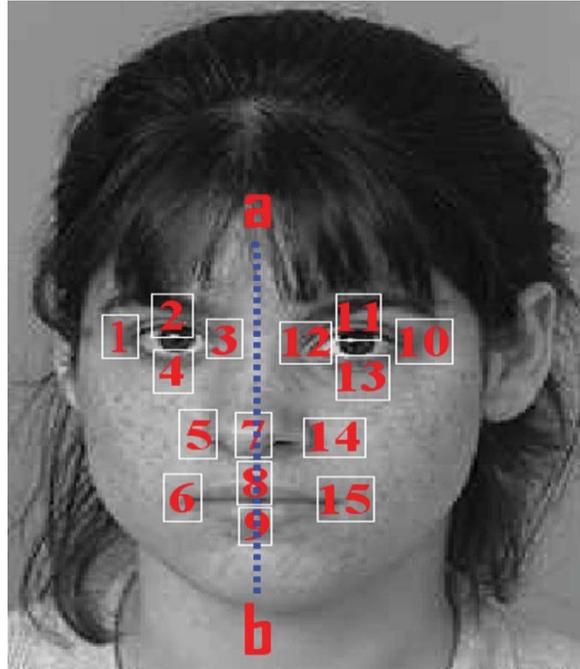


Fig. 1. Feature point positions in a frontal-view face image.

points, the similarity measurement ($\mathbf{q} - s\mathbf{R}2 \times 3\mathbf{M}$) in (3) can be rewritten as a vector function, as follows:

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x}), f_{n+1}(\mathbf{x}), \dots, f_{2n}(\mathbf{x}))^T. \quad (4)$$

The parameters φ, ψ, θ, s and the depth values Mz_i can be obtained by minimizing the distance d

$$\min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x})\|_2^2 = \min_{\mathbf{x}} \sum_{i=1}^{2n} f_i^2(\mathbf{x}). \quad (5)$$

Therefore, the pose and shape estimation problem is formulated as a NLS model. Unlike in [2], besides the pose parameters φ, ψ, θ , and s , the depth values Mz_i are also considered as the variables of the distance measurement in the NLS models. When more than one non-frontal-view face image of a subject is available, e.g., N non-frontal-view face images, the depth values Mz_i can be obtained by minimizing the sum of the similarity distances between the frontal-view face image and N non-frontal-view face images

$$\min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x})\|_2^2 = \min_{\mathbf{x}} \sum_{i=1}^{2nN} f_i^2(\mathbf{x}) \quad (6)$$

where $\mathbf{x} = (\varphi_1, \psi_1, \theta_1, s_1, \dots, \varphi_N, \psi_N, \theta_N, s_N, Mz_1, \dots, Mz_n)$.

B. Utilization of the Facial Symmetry Information in the NLS Model

From (3)-(5), it can be seen that only one non-frontal-view face image is used, which is efficient to reconstruct the 3D face model. However, when this image has a large pose angle, the reconstruction errors at those facial-feature points with large opposite rotation angles to the non-frontal-view face image are generally greater than those of the faces with an identical rotation direction to the non-frontal-view face image. To alleviate this sensitivity, the symmetrical information on human faces is considered in the NLS models.

Take the face image shown in Fig. 1 as an example, the feature points 10, . . . , 15 can be regarded as the mirrors of the points 1, . . . , 6 with respect to the line ab . Considering this symmetry, the depth values of the points 10, . . . , 15 should have approximately equal values to the points 1, . . . , 6. In the proposed NLS model, the depth values of the symmetrical points are assumed to be equal

$$\begin{cases} M_{z_1} = M_{z_{10}}, \\ M_{z_2} = M_{z_{11}}, \\ \vdots \\ M_{z_6} = M_{z_{15}} \end{cases} \quad (7)$$

which can be regarded as a hard constraint for the symmetrical points. Under this assumption, the number of variables in the NLS model decreases from 19 to 13. For this case, the objective function (5) can be rewritten as the following form:

$$\min_{\mathbf{x}'} \|\mathbf{f}(\mathbf{x}')\|_2^2 = \min_{\mathbf{x}'} \sum_{i=1}^{2n} f_i^2(\mathbf{x}') \quad (8)$$

where $\mathbf{x}' = (\varphi, \psi, \theta, s, Mz1, \dots, Mz9)$. Besides the hard constraint as in (7), a soft constraint can also be employed, which minimizes the distance between the respective symmetrical points. The objective function with this soft constraint can be written as follows:

$$\min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x})\|_2^2 = \min_{\mathbf{x}} \sum_{i=1}^{2n} (f_i(\mathbf{x}) + \lambda(|Mz_{10} - Mz_1| + \dots + |Mz_{15} - Mz_6|))^2 \quad (9)$$

where λ is the parameter to weight the soft constraint, and $|\cdot|$ denotes the absolute value.

C. Regularization Term Based on the Correlation Coefficient

Denote $Mzc = (Mzc1, \dots, Mzcn)$ as the depth values of the feature points of the CANDIDE model that have the same position as the feature points displayed in Fig. 1, Mzc should be closely correlated to the corresponding values of the 3D face model. Thus, Pearson's linear correlation coefficient $c(Mzc, Mz)$ of Mzc and the estimated depth values Mz can be used as a regularization term in the optimization process, which is obtained by

$$c(Mzc, Mz) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{Mz_{ci} - \bar{Mzc}}{SMzc} \right) \left(\frac{Mz_i - \bar{Mz}}{SMz} \right) \quad (10)$$

where Mzc and $SMzc$ are the mean and standard deviation of Mzc , and Mz and SMz represent the mean and standard deviation of Mz . A higher absolute value of $c(Mzc, Mz)$ means that Mz is closer to Mzc . When the linear correlation is considered as a constraint in model optimization, a general approach is to subtract $c(Mzc, Mz)$ from the distance d , as given in (3). Thus the optimization of (5) is turned into

$$\min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x})\|_2^2 = \min_{\mathbf{x}} \sum_{i=1}^{2n} (f_i(\mathbf{x}) - \lambda c)^2 \quad (11)$$

where λ is the parameter to weight the regularization term c . For (11), it is not easy to find a suitable range of λ so as to achieve the best performance. Here, in our algorithm, a simpler way is used by dividing d by c directly

$$\min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x})/c\|_2^2 = \min_{\mathbf{x}} \sum_{i=1}^{2n} (f_i(\mathbf{x})/c)^2. \quad (12)$$

For (12), we can see that the solutions are not changed if we substitute c with λc . That is to say, the parameter λ is no longer needed to be considered. From (10) and (12), it can be seen that both \mathbf{x} and $c(Mzc, Mz)$ contain the variable Mz . When $c(Mzc, Mz)$ decreases, i.e., Mz is far from Mzc , the objective function (12) becomes large, and vice versa. Therefore, c can be regarded as a regularization term to keep Mz close to Mzc . Since there is generally little difference in the 3D structures of different subjects, Mz should not be far from the true depth values if Mz is not far from Mzc . It should be pointed out that other efficient functions of c can also be used as the regularization term.

D. Optimization Schemes of the NLS Model

There are two schemes used 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 (5), i.e. $\mathbf{x} = (\varphi, \psi, \theta, s, Mz1, \dots, Mzn)$. The optimization can be carried out using the trust region reflective (TRR) algorithm.

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

$$\min_{\phi, \psi, \theta, s} \|\mathbf{f}(\phi, \psi, \theta, s)\|_2^2 = \min_{\phi, \psi, \theta, s} \sum_{i=1}^{2n} f_i^2(\phi, \psi, \theta, s). \quad (13)$$

Similarly, given the estimated pose parameters, the depth values Mz_i can be estimated by

$$\min_{Mz_i, i=1, \dots, n} \|f(Mz_1, \dots, Mz_n)\|_2^2 = \min_{Mz_i, i=1, \dots, n} \sum_{i=1}^{2n} f_i^2(Mz_1, \dots, Mz_n). \quad (14)$$

When the symmetrical information is embedded into the optimization procedure, (14) is changed into the following form for the example shown in Fig. 1:

$$\min_{Mz_i, i=1, \dots, 9} \|f(Mz_1, \dots, Mz_9)\|_2^2 = \min_{Mz_i, i=1, \dots, 9} \sum_{i=1}^{2n} f_i^2(Mz_1, \dots, Mz_9). \quad (15)$$

Correspondingly, the similarity measure in (3) is changed into The iteration procedure is hen 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 iteration 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.

$$d = \|q - sR_{2 \times 3}M\|_2^2 = \left\| \begin{array}{c} q_{x_1} - s(r_{11}Mx_1 + r_{12}My_1 + r_{13}Mz_1) \\ \vdots \\ q_{x_9} - s(r_{11}Mx_9 + r_{12}My_9 + r_{13}Mz_9) \\ q_{x_{10}} - s(r_{11}Mx_{10} + r_{12}My_{10} + r_{13}Mz_{10}) \\ \vdots \\ q_{x_{15}} - s(r_{11}Mx_{15} + r_{12}My_{15} + r_{13}Mz_{15}) \\ q_{y_1} - s(r_{11}Mx_1 + r_{12}My_1 + r_{13}Mz_1) \\ \vdots \\ q_{y_9} - s(r_{11}Mx_9 + r_{12}My_9 + r_{13}Mz_9) \\ q_{y_{10}} - s(r_{11}Mx_{10} + r_{12}My_{10} + r_{13}Mz_{10}) \\ \vdots \\ q_{y_{15}} - s(r_{11}Mx_{15} + r_{12}My_{15} + r_{13}Mz_{15}) \end{array} \right\|_2. \quad (16)$$

Given a small positive real number ε , if the condition

$$dk+1 - dk < \varepsilon \quad (17)$$

is satisfied, then the iteration is terminated. The specific steps of the latter method are given as follows.

Step 1: Initialize the depth values Mz_1, \dots, Mz_n according to the CANDIDE model, and set the constant ε or the maximum number of iterations ($Niter$).

Step 2: With the initial depth values, estimate the pose parameters φ, ψ, θ, s by minimizing the objective function in (13) via the TRR algorithm.

Step 3: With the estimated pose parameters, estimate the depth values Mz_1, \dots, Mz_n by minimizing the objective function in (14) 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 (17) 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:

NLS1: one-step optimization without considering the symmetry information;

NLS1_S: one-step optimization considering the symmetry information;

NLS2: two-step optimization without considering the symmetry information;

NLS2_S: two-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 $NLS1_R, NLS1_SR, NLS2_R$ and $NLS2_SR$, respectively.

E. Model Integration for Multiple Nonfrontal-View Face Images

When more than one non-frontal-view face image of a subject is available, the depth values can be obtained via the method of (6) (denoted as $NLS1_R_MS$). Assume that there are N non-frontal-view face images to be used in the optimization, it can be seen from (3)–(6) that the dimension of the function vector and the parameter number increase linearly with N . As a result, the 3D structure estimation accuracy level will decrease greatly due to the increase in the number of variables in the NLS model. Therefore, a model integration method (denoted as $NLS1_R_MI$) based on the trimmed mean is proposed in this paper for the case of multiple nonfrontal-view face images. Assume that P percentage of the observations are trimmed, the number of the smallest or the largest observations to be discarded is

$$K = [N(P/100)/2], \quad (18)$$

where $[\cdot]$ denotes a rounding operation. Further, assume that the j th entry of Miz are ordered so that $M1z.j < M2z.j < \dots < MN.j$, the trimmed mean of the j th entry $Mtzj$ can be given by

$$M_{zj}^t = \frac{1}{N - 2K} \sum_{i=K+1}^{N-K} M_{zj}^i, \quad j = 1, \dots, n. \quad (19)$$

The method $NLS1_R$ is used as an example here to explain how to perform model integration due to its simplicity. As shown in Fig. 2, given the initial parameters and M_c , a depth-estimation Miz ($i = 1, \dots, N$) is obtained for each nonfrontal-view face image q_i by first using $NLS1_R$. For each of the models, the mean reconstruction error dk of the k th iteration is calculated for all testing face images. If $dk+1 \geq dk$, the iteration is terminated and the trimmed mean Mtz of Miz in the k th iteration is regarded as the final depth estimation value. Otherwise, Mtz is used as the initial value of $NLS1_R$ to substitute M_c , and the iteration is repeated till $dk+1 \geq dk$. Those estimated depth values with high errors are removed in the calculation of the trimmed mean. As a result, the trimmed mean can be regarded as a more correct estimation of the depth than that provided by the CANDIDE model. Therefore, the depth estimation accuracy level will be improved via the model integration method.

F. Face Recognition

If the depths Miz ($i = 1, \dots, k$) are estimated as described above for each subject, then k models are constructed for k subjects. For a query face, the feature point distances between it and the k subjects can be computed using (3). In general, the distances between this face model and a face image of this particular subject should be smaller than that of other subjects. Therefore, the constructed 3D model can be used for face recognition. This can also be used as a performance index to show the accuracy of the estimated 3D face models.

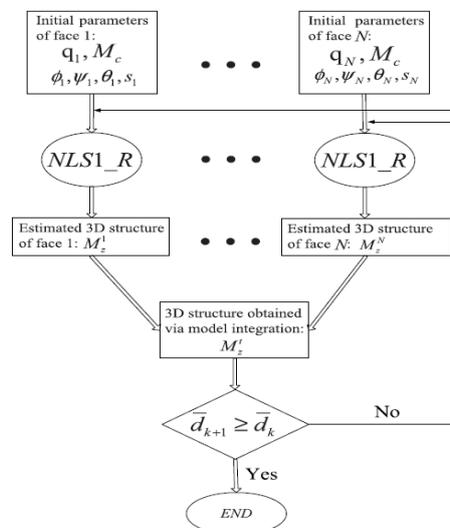


Fig. 2. Flowchart of the model integration method.

III. Conclusion

In this paper, NLS-based methods are proposed to estimate the 3D shape and pose of face images. The two-step optimization scheme, the symmetry information and the correlation coefficient- based regularization term are proposed, and it has been verified by experiments that they can improve the 3D face-reconstruction accuracy and the recognition performance. Compared to the GA-based method, the proposed methods have a comparable reconstruction and recognition performance, while the training times required decrease significantly. Experimental results have demonstrated the feasibility and efficiency of the proposed methods. In future work, we will further improve the efficiency of the model-integration method and investigate other regularization terms so as to improve the accuracy of both the depth values estimated and the recognition performance.

References

- [1] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Comput. Vis. Image Understand.*, vol. 101, no. 1, pp. 1–15, Jan. 2006.
- [2] H. S. Koo and K. M. Lam, "Recovering the 3D shape and poses of face images based on the similarity transform," *Pattern Recognit. Lett.*, vol. 29, pp. 712–723, Apr. 2008.
- [3] D. L. Jiang, Y. X. Hu, S. H. Yan, L. Zhang, H. J. Zhang, and W. Gao, "Efficient 3D reconstruction for face recognition," *Pattern Recognit.*, vol. 38, no. 6, pp. 787–798, Jun. 2005.
- [4] C. Z. Zhang and F. S. Cohen, "3D face structure extraction and recognition from images using 3D morphing and distance mapping," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1249–1259, Nov. 2002.
- [5] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 690–706, Aug. 1999.
- [6] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.