Abstract

Three-dimensional human face recognition in the presence of expression is a big challenge, since the shape distortion caused by facial expression greatly weakens the rigid matching. This paper proposes a guidance-based constraint deformation (GCD) model to cope with the shape distortion by expression. The basic idea is that, the face model with non-neutral expression is deformed toward its neutral one under certain constraint so that the distortion is reduced while inter-class discriminative information is preserved. The GCD model exploits the neutral 3D face shape to guide the deformation, meanwhile applies a rigid constraint on it. Both steps are smoothly unified in the Poisson equation framework. The GCD approach only needs one neutral model for each person in the gallery. The experimental results, carried out on the large 3D face databases—FRGC v2.0, demonstrate that our method significantly outperforms ICP method for both identification and authentication mode. It shows the GCD model is promising for coping with the shape distortion in 3D face recognition.

1. Introduction

The three-dimensional human face recognition has received greatly increasing attentions in recent years[17, 3, 6, 9, 10, 22, 1]. Compared with the facial images, which essentially is the projection of the 3D human face onto a 2D planar domain, 3D facial data have more clues for recognition. Utilizing the three-dimensional information in matching face model is promising to overcome the difficulties of the image-based face recognition caused by the variations of illumination, facial posture and expression etc.

For two rigid facial surfaces, the classical matching procedure is to align the surfaces and define a similarity measure for the aligned shapes, such as Hausdorff distance [13, 18], RMS(root of mean squared distance)[5]. However, these approaches are usually not applicable to the non-rigid cases, which are mostly caused by expression variation.

A big challenge in 3D shape-based face recognition is to cope with the shape change due to facial expression, which is called distortion in this paper. The shape distortion will introduce many non-rigid components and reduce the similarity in rigidity sense between faces from the same subject. 3D face recognition should be treated as a non-rigid surface matching problem, which is still a difficult task.

There are some efforts made to attack the non-rigid surface matching problem. C.S.Chua et al [6] used Gaussian distribution to extract rigid parts of facial surface for matching. They recognize 3D face using a local descriptor, named "Point Signature", combined with "voting method". K.Chang et al [4] also extracted the rigid parts of facial surface for matching to cope with expression variation. They mainly use the nose-near region for recognition. The surface matching was achieved by ICP. A.Bronstein et al [2, 3] assumed that the facial surfaces in the presence of expression is an isometric surface which are not stretched by expressions. All 3D facial models were transformed into a canonical form for recognition, which is invariant to the isometric deformation. To reduce the effect of expression, X. Lu et al [10] introduced a person-specific deformable model into facial surface matching. By a control group, a neutral model was transferred to several models which were synthesized into a deformable model before fitted to a test scan.

This paper addresses to reducing the shape distortion in 3D face recognition, so as to reduce the effect of expression. We propose a guidance-based constraint deformation approach to perform this task. The guidance-based constraint deformation(GCD) is not to construct a general expression model but for the recognition task. We suppose there is a facial model with neutral expression in the gallery for each subject. Before each comparison between two models, the proposed GCD model will try to deform the probe model toward the gallery model with some constraints. The GCD model can make the facial models from the same individual (intra-class) more similar, while maintaining the discrimination between different individual (inter-class).
2. Framework of the Proposed Approach

The proposed framework is the deformation kind. Our motivation is to reduce the shape distortion of the input model using a kind of shape deformation method, so that the deformed probe model will enhance its discriminative characteristics for matching.

In our proposed approach, the gallery only needs one 3D face mesh with neutral expression for each individual. Given a probe mesh \( p \), the \( p \)'s ROI (region of interest) will be extracted firstly. For each comparison, the dissimilarity between \( p \) and a gallery mesh \( g \) can be computed as follows:

\[
\text{Dis}(p, g) = \text{RMS}(D^g_c(p), g)
\]

where \( D^g_c(\cdot) \) means the presented GCD deformation of \( p \) with \( g \) as the guidance and \( c \) as the constraint (detailed in the next section), \( \text{RMS}(\cdot, \cdot) \) is the root of mean squared distance. In 3D face recognition in the presence of expression, \( p \) generally is with non-neutral expression, \( g \) is with neutral expression.

The framework of our whole non-rigid matching process is shown in Fig. 1.

3. Guidance-based Constraint Deformation

We attempt to deal with the expression problem from the discriminative view, since, for the recognition task, the discriminative characteristics are the most important. On the one hand, we should make effort to reduce the intra-class distances, on the other hand, the inter-class distances should be held and even strengthened.

In this section, we will propose the guidance-based constraint deformation (GCD) for non-rigid 3D face recognition. The GCD model includes the two key parts:

(1) **Guidance-based deformation**: try to deform the non-neutral probe mesh toward a neutral gallery mesh so that the non-rigid distortion can be reduced as much as possible, which possibly reduces the intra-class distances.  

(2) **Rigid constraint**: apply the constraint on the deformation so that the comparatively rigid components keeps unchanged, which is for not reducing the inter-class distances.

3.1. Theoretical Background

We represent a facial surface as a 2-manifold mesh \( M = (P, K) \), where \( P = \{p_i \in \mathbb{R}^3 \mid 1 \leq i \leq N \} \) is the set of vertices and \( K \) encodes the connectivity of the simplicial complex including the vertices, edges and triangles. For a 1-ring mesh is shown in Fig. 2. Some notions are introduced here at first [11, 19, 15, 16].

**Discrete Scalar Fields and Discrete Vector Fields**: A scalar field \( u \) on the triangular mesh \( M \) is expressed as:

\[
u(v) = \sum \phi_i(v)u_i
\]

where \( \phi_i \) is the piecewise-linear basis function valued 1 at vertex \( v_i \), and 0 at all other vertices of \( M \). The scalar function \( u \) has value \( u_i \) at \( v_i \). Obviously, The three coordinates, \( x, y, z \), can be seen as three scalar fields defined on \( M \).
Similar to discrete scalar fields, we call $\xi$ a vector field defined on $M$ if:

$$\xi(v) = \sum_{k} \psi_k(v) \xi_k$$  \(3\)

with $\psi_k$ being the piecewise-constant basis function valued 1 within the triangle $T_k$ and 0 outside of $T_k$. The gradient of the scalar field is a vector field.

**Discrete Gradient Operator:** The gradient of scalar field $u$ on a triangular mesh can be expressed as:

$$\nabla u(v) = \sum_{i \in T_v} u_i \nabla \phi_i(v)$$  \(4\)

where $T_v$ is the triangle $v$ attached and its three vertices are $v_0, v_1, v_2$, and

$$\nabla \phi_i = \frac{1}{2A_T} R^{90}((v_{(i+2)} \% 3) - v_{(i+1)} \% 3)$$  \(5\)

$A_T$ is the area of $T_v$ and $\%$ is module operator. $R^{90}(\cdot)$ denotes anticlockwise rotating the vector of 90 degree along the normal of $T_v$. The gradient field is constant inside $T_v$.

**Discrete Divergence:** For a vector field $\xi$ over the $M$, the discrete divergence at a vertex $v_i$ is defined as [19]:

$$\text{Div}(\xi)(v_i) = \sum_{T_k \in N_T(i)} \nabla \phi_i \cdot \xi_k A_k$$  \(6\)

where $N_T(i)$ is the neighbor triangles of vertex $v_i$.

**Discrete Laplace-Beltrami Operator:** The discrete Laplace-Beltrami operator in 1-ring neighborhood case (shown in Fig. 2) could be written as:

$$\Delta u(v_i) = \sum_{j \in N(i)} \frac{1}{2} (\cot \alpha_j + \cot \beta_j)(u_i - u_j)$$  \(7\)

### 3.2. Guidance-based Deformation of Facial Mesh

The problem of the guidance-based deformation is: when given a facial mesh $M_p$ with neutral expression and a distorted facial mesh $M_g$, how can we efficiently deform the distorted version $M_g$ back to its neutral one exploiting $M_p$ as the guidance (also the target). The facial shape distortion due to expression should be mostly changed back to its neutral expression version in a natural way which means the overall shape follows the specified guidance while important geometric details are locally deformed in a way keeping their appearance characteristics. In this paper, we incorporate the Poisson-based gradient field manipulation technique [21] into the target-guided deformation of facial mesh.

#### 3.2.1 Poisson-based Mesh Deformation

For a facial triangular mesh $M$, each coordinate-component of coordinate 3-tuple of its vertices could be treated as a scalar field on $M$. Thus, $M$ could be represented by three scalar fields. In the Poisson-based mesh editing framework, the manipulation on the gradient fields of a mesh can be transformed back to the scalar fields of the mesh, i.e. the three coordinate components can be reconstructed from the modified gradient fields [21].

The Poisson equation with Dirichlet boundary condition is formulated as

$$\nabla^2 f = \text{Div}(\xi) \quad f|\partial \Omega = f^*|\partial \Omega$$  \(8\)

where $f$ is an unknown scalar function, $\xi$ is a guidance vector field, $f^*$ provides the desirable values on the boundary $\partial \Omega$. When considering piecewise linear functions over triangular mesh, the discrete Poisson equation can be written as

$$\Delta(u) \equiv \text{Div}((\nabla u) = \text{Div}(\xi)$$  \(9\)

which leads to a sparse linear system when yielding Eq. 4, Eq. 6 and Eq. 7:

$$AU = b$$  \(10\)

where $U$ is the coordinates of the unknown deformed mesh, $b$ is the divergence of the modified gradient fields, and $A$ is a sparse matrix as follows:

$$A_{ij} = \begin{cases} -\frac{1}{2}(\cot(\alpha_{ij}) + \cot(\beta_{ij})) & \text{if } j \in N(i) \\ -\sum_{k \in N(i)} A_{ik} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$  \(11\)

Using Eq. 9, we can solve an unknown mesh with known topology (i.e. vertex connectivity) but unknown geometry (i.e. vertex coordinates) via the discrete guidance vector fields. For instance, we can deform the face mesh by firstly editing some vertices’ positions, then computing the three modified gradient fields, and finally solving Eq. 9.

#### 3.2.2 Gradient Fields Computing

In 3D face recognition in the presence of expression, our purpose is to remove the distortion introduced by expression. Thus, the key problem is how to modify the gradient fields, which can be described as: Given a probe mesh $M_p$ and a gallery mesh $M_g$ with neutral expression, modify $M_g$’s gradient fields so that the reconstructed mesh has less shape distortion by expression. We use $M_g$ as guidance, which makes $M_p$ have chance to become the neutral one (when $M_g$ is exactly the neutral model of the same subject of $M_p$).

Firstly, we put $M_p$ and $M_g$ in the canonical coordinate system which can be done by a robust facial feature points detection method which will be presented in Sec.4.2.

Secondly, we try to transform the gradient field of $M_p$ toward that of $M_g$. In order to be consistent between three coordinates, three gradient fields(corresponding to three coordinates) are applied an uniform transformation. It is done...
by introducing a local transformation matrix $H_i$ for each triangle $T_i^g$ of $M_g$, since the gradient within $T_i^p$ is constant. We search a closest triangle $T_i^p$ in $M_p$ and transform $T_i^p$ to $T_i^g$ so that $T_i^p$ is parallel with $T_i^g$. Suppose the coordinate of vertex $v_j$ of $T_i^p$ is $X_{ij}$ and that of $T_i^p$ is $X'_{ij}$, then:

$$X_{ij} = H_i \cdot X_{ij}, j = 0, 1, 2$$ (12)

For $T_i^g$ and $T_i^p$, their nearest vertices pairs can be corresponded and combined with the normals of triangles, two local orthogonal frames $F^g_i, F^p_i$ can be set up. Since the gradient vector is translation invariant, we only consider the rotation. The $H_i$ can be obtained by:

$$H_i = F^g_i \cdot (F^g_i)^{-1}$$ (13)

The transformation is around the barycenter of the triangles. The coordinates of three vertices of $T_i^p$ are applied to Eq.4 to compute the modified gradient field. Since the $H_i$ on every triangle of $M_p$ is different, the adjacent triangles in $M_p$ will be broken. When we use the divergence of the modified gradient filed as right side of Eq.9, the solution of Poisson-equation paste the broken triangles and the integrated mesh is reconstructed by Poisson-equation.

### 3.3. Rigid Constraint on Deformation

![Figure 3](image3.png)

**Figure 3.** Constraint vs Without-constraint. (a) Intra-class deformation without constraint. (b) Intra-class deformation with rigid constraint. (c) Inter-class deformation without constraint. (d) Inter-class deformation with rigid constraint.

If the guidance and the probe are with the same subject, the deformed probe will be very close to the guidance with most distortion removed, as we expected, shown in Fig. 3(a). Otherwise, when the guidance and the probe are with two different subjects, the domination of the guidance during the deformation will lead to losing much discriminative information of the probe, which causes the deformed probe will also be similar to the guidance, shown in Fig. 3(c). This is attributed to lack of the probe’s control in the guidance-based deformation. A possible way is to impose some constraints from the probe on the deformation to weaken the governing of the guidance.

One important observation is that the within-class distortion is more stable than the between-class distortion under the expression variation. Figure 4 shows some results of two kinds of distortions, from which it can be seen that the within-class distortion concentrates on the regions of eyes, the mouth and chin while the between-class distortion is almost distributed over the whole face surface. Those relatively stable regions for the within-class distortion can be called nearly-rigid components.

Motivated by this observation, we apply a rigid constraint to the deformation so that some nearly-rigid components of the probe is kept unchanged during the deformation. From Poisson equation in Eq.8, the Dirichlet boundary condition is to specify the values of some variables as:

$$u(v_i) = u^*(v_i)$$ (14)

This can be easily achieved by fixing some points in Eq.10. Therefore, the problem becomes how to extract the nearly-rigid components of facial surface under different expressions.

A rigid template can be built to denote the nearly-rigid components from the training data. We select a set $S_t$, in-
including 60 subjects with each has 4 meshes, where one is the neutral model and the other three with various expressions in different degree (randomly selected).

To set up the correspondence between two meshes, we parameterize each model to a disk to form a parametrization domain[7]. The statistical distribution of the distortion is based on parametrization domain. As shown in Fig. 5, the parametrization domain is discretized into many cells by evenly segmenting the radius and rotational angle. Each vertex on the mesh will fall into these cells after parametrization.

Using $S_t$ to calculate the within-class distortion will result in 180 distances for each vertex. For each cell, we calculate the average distortion over those vertices falling into this cell. Suppose $V_i$ is the vertices set whose element is falling in the cell $i$, and a vertex in $V_i$ has Euclid distance $d_{ij}$ from the vertex to its nearest vertex after matching, the mean distortion $DT_i$ of cell $i$ can be computed as:

$$DT_i = \frac{1}{|V_i|} \sum_{1 \leq j \leq |V_i|} d_{ij}$$

(15)

After calculating all cells on parametrization domain and normalizing the distortion, an average distortion map can be generated. Figure 5(e) is the sampled image of the distortion map. The darker indicates larger distortion. This map indicates the with-class distortion of shapes introduced by expression. According to this map, setting a certain percentage of the domain to be the nearly-rigid components for the mesh deformation can obtain a binary mask, which is used as constraints in our experiments. This percentage is called constraint ratio (CR) in this paper. Figure 5(f) is a constraint mask with 50% CR. Figure 3(b)(d) are the deformation results using the 50% binary mask. It can be seen that after adding the rigid constraints, the models from the same subject still can be deformed to be similar, while the deformation between different subjects holds the difference.

4. Experiments

4.1. Data

Figure 6. The model examples from FRGC database.

Our experiments use the 3D facial data from FRGC database version 2.0[14], totally 557 subjects and 4950 3D models. Each face model has the expression label and capturing timestamp. All the non-neutral models in FRGC database are adopted as our probe dataset, totally 1538 models for 353 subjects, who have the neutral expression model captured before the non-neutral scan. We use the earliest neutral models of the 353 subjects as the gallery, one model for each subject. The probe data set is divided into nine probes in a natural manner of capturing date, shown as Table 1. Figure 6 shows some non-neutral examples of FRGC database.

<table>
<thead>
<tr>
<th>Capturing date</th>
<th>Expression</th>
<th>Models#</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/07/2003-10/09/2003</td>
<td>smiling</td>
<td>224</td>
<td>1</td>
</tr>
<tr>
<td>10/14/2003-10/16/2003</td>
<td>frowning</td>
<td>161</td>
<td>2</td>
</tr>
<tr>
<td>10/28/2003-10/30/2003</td>
<td>surprise</td>
<td>171</td>
<td>3</td>
</tr>
<tr>
<td>11/04/2003-11/06/2003</td>
<td>disgust</td>
<td>188</td>
<td>4</td>
</tr>
<tr>
<td>02/10/2004-02/12/2004</td>
<td>smiling</td>
<td>115</td>
<td>6</td>
</tr>
<tr>
<td>02/17/2004-02/19/2004</td>
<td>surprise</td>
<td>156</td>
<td>7</td>
</tr>
<tr>
<td>02/24/2004-02/26/2004</td>
<td>puffy cheeks</td>
<td>190</td>
<td>8</td>
</tr>
<tr>
<td>03/02/2004-03/04/2004</td>
<td>frowning</td>
<td>168</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 1. The probe partition of FRGC v2.0

4.2. Preprocessing

Commonly, there exist noises or outliers in the captured 3D face model. Several preprocessing steps are carried out before the deformation. These steps are briefly presented as follows.

(1) If the facial data have only the points information, the surface is reconstructed by triangulation. Then, ROI (region of interest) of the face model is roughly cropped by the curvature analysis of surface. For details, see[20].

(2) After that, the mesh simplification is applied on the face mesh to reduce the number of vertices to meet the requirement of efficient computation[8]. About 8000 vertices and 160000 triangles are preserved in every simplified facial mesh in our experiments.

(3) Then, the model is placed to a canonical coordinate system by feature points detection. Using the method proposed by Pan[12], the nose tip $p_{nt}$, nose base $p_{nb}$, and direction of symmetrical plane $d_s$ of face are calculated to fix the six degrees of freedom of facial surface.

(4) Finally, a geodesic-based cropping is performed to refine the ROI

4.3. Evaluation of Constraints Ratio Effect

Constraint ratio (CR) mentioned in Section 3.3 plays an important role in the constraint deformation, since it controls the percent of the nearly-rigid component for constraint.

To evaluate the effect of different $CR$, nine $CR$ values from 10% to 90% in an interval of 10% is tested in the identification. We choose 100 subjects from 353 neutral models

of as gallery, and randomly select 2 non-neutral models for each subject from 1538 scans. A total of 200 non-neutral scans for the probe. The recognition result is shown in Fig. 7. The best rank-1 rate is achieved with CR around 50% or 60%. It can be seen that the recognition performance is low when CR is 90%, as we expected, since most part of the face model is forced to keep unchanged during deformation in that case. In the following experiments, we set the CR to 50%.

4.4. ICP vs GCD

In order to verify the advantages of our approach, we make the comprehensive comparison experiments. In the experiments, we use ICP method as the baseline algorithm. ICP method is one of the most popular 3D face recognition method and has the relatively good performance. We carry out both face identification and face authentication experiments using the gallery of 353 models and nine probes in Table.1.

3D Face Identification: The result is shown in Fig. 8. For all the nine probes, our method outperforms ICP approach. The biggest improvement happens on the probe1 and probe6, which is smiling expression. More than 25% of increasing is achieved for both probes. An average improvement of 11.8% are obtained over 9 probes.

3D Face Authentication: Table.2 shows the results of the equal error rate(EER). Obviously, our method gets better performance than ICP method. Among the nine probes, ICP method has an average EER 12.5% and the worst EER 18.6%. Our method achieves an average of 6.2% and the worst EER of 8.51%. The average improvement of our method is more than 6%.

From the above experimental results, it is obvious that GCD model prominently improves the ICP matching method. This mainly contributes to its removing the majority of distortion induced by different expressions.

4.5. Computational Performance Analysis

The main steps included in GCD-based method is canonical coordinate transformation(CT) for the models, parametrization(PARA), the Poisson Equation solution (PES) and RMS calculation.

The parametrization and Poisson Equation can be both formalized as solving a sparse linear system. By taking a Cholesky decomposition of matrix in the parametrization and Poisson equation, the solution could be obtained from back substitution. The decomposition is taken once, and the cost will be amortized among the whole matching procedure. Either, the canonical coordinate transformation will be done only once. Its cost also amortized by all probe models.

The RMS calculation requires finding the corresponding nearest points pair between two facial meshes. The k-d tree technique is used to compute the pairs.

The experiments are carried out on our system of CPU P4 2.4GHz with 1024MB DDR333 RAM. Table 3 shows the performance of our method.

5. Conclusion

This paper has proposed a framework for robust 3D face shape matching in the presence of non-rigid distortion in-
duced by the expression variation. The key component is GCD model, which is to reduce the shape distortion via a guidance-based constraint deformation. On the one hand, GCD model employs a neutral face shape as the target to guide deformation. On the other hand, it exploits rigid component of the original non-neutral face shape to suppress deformation. Essentially, GCD model makes the trade-off between inter-class and intra-class distances from the discriminative capability view.

The evaluation of the proposed scheme is conducted using the large 3D face databases, FRGC v2.0 with 1538 non-neutral face models. The ICP method acts as the baseline for comparison. Both face identification and face authentication experiments are performed. The experimental results on these 3D models with wide expression variation indicate that the GCD approach is prominently superior to ICP in the presence of expressions, especially in the case of large distortion. Since the guidance-based constraint deformation is achieved by solving a sparse linear system, the computational cost is reasonably acceptable.

The proposed scheme could be easily integrated with most other 3D face recognition methods. Actually it could be considered as a kind of analysis-by-synthesis approaches. It generates a virtual 3D model for the input, replacing the original probe one. Therefore, the GCD could be regarded as a pre-processing for other approaches.

6. Acknowledgements

This work is partly supported by NSFC (60503019, 60525202), Key Program of NSFC(60533040) and Program for New Century Excellent Talents in University (NCET-04-0545). The authors would also thank Dr. Kun Zhou for valuable discussions on Poisson Equation.

References