Extraction and Application of Regionalized Implicit Function Feature of 3D Face

Jiang Zhi-chao, Mu Yong-min and Zhang Zhi-hua

Beijing Information Science and Technology University, the faculty of Computer, Beijing 100001 China

Abstract

3D face reconstruction and 3D face recognition belonged to different fields previously. In order to solve these two issues in a unified way, this paper designed a kind of regionalized implicit function feature (RIFF) algorithm frame. Firstly according to the biology characteristic, the face was segmented to five regions with different resolutions; secondly the control matrix which expressed the human face was solved for 3d face reconstruction; then dimensionality reduction was did for this control matrix; finally the low dimensional character description was computed completely for 3d face classification. The specific experiments for the RIFF method are performed and the results show that this method can acquire significant effect in the performance of reconstruction and recognition.

Keywords*: 3D face reconstruction; 3D face recognition; regionalized implicit function feature; control matrix; Low dimensional character description*

1. Introduction

With the quick development of researches on computer vision, the application of three-dimensional (3D) images has become increasingly active in various fields inclusive of safety defence, computer animation, cosmetic medicine, model-based image coding, etc. For example, in the safety defence field, 3D human face recognition, one of the hottest topics in the current image processing technology, has already become the most natural and direct means among authentication technologies and the recognition technology based on the 3D face modelling has been widely regarded as the most intuitive approach of the 3D human face recognition. Furthermore, in the field of computer animation, 3D face reconstruction technology has often been used to achieve the 3D facial structure and the expression simulation of a virtual character.

In addition, 3D face reconstruction technology has also been widely used in the cosmetic medicine, model-based image coding, and other fields.

However, it's not always the same in different fields in terms of the requirements of 3D face reconstruction technology. As for the face recognition, the most important thing for each face is to find those features that can be used to distinguish itself from other faces, *i.e.*, a training database containing several individual faces of a certain person is needed, and the distance between clusters should be large enough while the internal distance should be small enough. While for the computer animation, the fidelity of face model and the visual effects that value the most.

This paper proposes a regional implicit function feature (RIFF) algorithm framework. Basing on the combination of the 3D face recognition and reconstruction in commonality, the algorithm utilizes the reconstruction information as the high-dimensional recognition features to realise the repeat and effective use of face information, and finally achieves significant

results in face recognition and reconstruction. In fact, to achieve smooth results in face reconstruction and accurate results in face recognition simultaneously, our algorithm segments the human face into five sections, each with different resolution values, and extracts easy-discriminated 3D face data structure as a face recognition feature.

Section 2 of this paper introduces the background of the related technologies, Section 3 provides a brief introduction to our RIFF algorithm framework, Section 4 describes more implementation details of both 3D face recognition and reconstruction, Section 5 is the RIFF algorithm experimental evaluation, and finally Section 6 concludes the paper.

2. Related Work

2.1. 3D Face Reconstruction

In order to acquire the disordered 3D face recognition point cloud data utilizing sensors, reconstruction algorithm is needed first to form the structured object whose features should be easily extracted. Currently, point cloud reconstruction algorithms can be broadly divided into two categories: combinatorial algorithms and implicit function algorithms.

Combinatorial algorithms include the Ear Clipping Triangulation method, Greedy Projection Triangulation algorithm, Marching Cubes algorithm, etc. Herein, the basic idea of Marching Cubes algorithm is to divide the point cloud region to form cubic voxels using the uniform gird, and then identify all these cross faces between each cubic voxel and the iso-surface. In fact, the combination of these cross faces is just the iso-surface of the object surface.

The main idea of implicit function algorithm is based on the optimized interpolation and approximation for the point cloud data to fit curved surface, and the typical implicit function algorithms include Voronoi-based variational reconstruction of unoriented point cloud sets [1-2], wavelet streaming surface reconstruction [3], signed distance surface reconstruction [4], Poisson reconstruction [5], radial basis function reconstruction[6], etc.

Voronoi-based reconstruction algorithm can form a topology equivalent grid structure as the original surface, however, it overly dependents on the accuracy of the data set and sometimes has a harsh requirement for the sampling points, such as the need for very intensive sampling points at sharp corners.

Signed function and Poisson reconstruction methods are based on implicit function reconstruction. While the implicit function is defined by the surface contour of scalar functions. Signed distance function can be estimated by two methods, namely the tangent plane based estimation method and the voxel based estimation method. The benefits of the latter approach lie in the fact that no special registration is required to extract the surface from a range of surface data, and that the measurement limit is also considered so as to stay stable when dealing with outliers. Voxel-based signed distance function for the reconstruction of sharp angles or thin plane will bring interference among the range images, however, this technology is quite practical, therefore, it is commonly used in the pretreatment of many algorithms.

This paper utilizes the Poisson equation to solve the control matrix, which has been regarded as the classical method to solve those issues including dynamic range image, fluid mechanics, grid editing, GPU computing, etc.

From the mention above, we can find that the aforementioned point cloud reconstruction algorithms do not combine with the structure and biological characteristics of the human face, and that if the prior knowledge is introduced, the reconstruction speed and effect will be significantly improved. Regarding the solution of the implicit function as the goal, this paper introduces both the surface curvature and the organ discrimination properties to make the

reconstruction results modelling, and treats them as recognition features. Finally, our paper also compares our reconstruction method with other implicit function methods as well as combinatorial methods in terms of the performance.

2.2. 3D Face Recognition

3D face recognition methods include airspace match, local feature matching, the overall feature matching, multi-modal fusion, etc.

The RIFF method proposed in this paper is the combination of the overall feature matching and the local feature matching. The traditional algorithm that combines the overall feature matching and the local feature matching prefers to divide the whole face region into several subregions to extract features, and then makes the weighted fusion in the whole face region. However, some newly emerged methods such as MS-CFB [7] can unify these two processes, but it still dependents on the hand posture correction when dividing subregions. The RIFF method proposed in this paper can directly utilize the intrinsic geometry property of the human face for regional segmentation. In fact, the local and global features can be integrated during the reconstruction, and the process is fully automated. Actually, the direct use of the control matrix extracted in the reconstruction process directly will definitely bring the curse of dimensionality. Thus, in our paper, we translate the control matrix to the Laplace matrix, and then utilize the spectrum theory to reduce the features for the purpose of using low-dimensional feature to represent high-dimensional data category information. The existing classical spectrum algorithms contain PCA [8], LDA [9], ISOMAP [10], LLE [11], MFA [12], CDFA [13], LLD [14], etc, and newly emerged ones include M3L [15], LRE [16], GPSDA [17], etc.

Both PCA and LDA are very popular linear graph embedding algorithms, and both of them assume that the data is Gaussian distributed and projection direction is effective. PCA focuses on the projection direction of the maximum variance, while LDA pursuits the projection direction of the most effective identification. ISOMAP is an isometric map which retains the lateral distance between the data in the data set. However, ISOMAP can only be applied with the rigid surface of flat manifolds. LLE can preserve the relationship between neighboring points when reducing the dimensionality, however, LLE is a local convergence, and has poor adaptability for the improper sampling points. In MFA, the category criterion can be obtained by calculating the quotient between the intra class density, the distance between points in the same class, and the inter class density, the distance between two points in two different classes. It's worth noting that MFA overcomes the limitation of assuming that the data must be Gaussian distributed. CDFA defines a separate manifold for each class, and establishes the connection between the popular learning method and the generated probabilistic model to deal with the nonlinear property introduced by the change of system variables. CDFA works well for recognizing those faces with changes of illumination and expression, however, considering that CDFA is modelled on the basis of a smooth manifold, higher requirements for sampling data is needed. LLD overcomes the limitation of LDA algorithm and other algorithms that they can only work in the Euclidean space, and can automatically find better discriminant subspaces. In fact, in the Euclidean space, LLD and LDA are equivalent. In recent researches, a multi-feature multi-manifold learning method, namely MMML method, is proposed to specifically deal with the situation of lacking in features in a single face image, and has achieved good results in the databases including AR, FERET, LFW, etc. Furthermore, some researchers also proposed the LRE model, which attempts to maximize the retention position information of data to ensure effective recognition features, and verified it on those databases inclusive of CMU PIE, FERET, ORL, Yale and FRGC databases. In

addition, a semi-supervised algorithm called GPSDA was also suggested to dig the hidden discrimination information of unlabeled data by introducing the measurement of probability similarity, and achieved some good results.

Different from other dimension reduction algorithms, the RIFF algorithm reduces the dimension of intermediate results, the control matrix, of the 3D modelling process. In fact, the whole process is also totally unsupervised. Furthermore, this method treats the information of 3D surface reconstruction as the structural features descriptor of 3D human face, and reconstructs the coefficient matrix, i.e., reconstructed 3D feature map. Then, the feature map will be mapped to the Laplace space to further reduce its dimensions. Finally, we can obtain the regional implicit function characteristics which can be used for the face recognition. Our proposed RIFF algorithm can map the face model in the Euclidean space to the Laplace space without changing the spatial structure to ensure that the model can be projected in the direction toward the maximum retention characteristics. The superiority of our algorithm will be partially verified in the experiment in the following.

3. RIFF Algorithm

The algorithm assumes that the input point cloud data in the face surface is the uniformly distributed sample collection CloudSet, the set CS is composed by M points as cs.p, the normal vector of each point is $cs.\vec{N}$. Assuming implicit function in the face surface is expressed as IF, S is the surface of reconstructed face model. The task of RIFF algorithm is to find the control matrix of implicit surfaces. This matrix is ordered, the face point cloud is defined as a Octree Θ, then some of the node $o \in \Theta$ will be one item of control matrix.

3.1. Basis Function Definition

The B-spline curve or surface [18~21] is widely used due to its abilities of processing local control and representing complex shape. This Paper uses bicubic B-spline function as a yuan of basis functions.

$$B_{k,3}(t) = \frac{t - t_k}{t_{k+2} - t_k} B_{k,2}(t) + \frac{t_{k+3} - t}{t_{k+3} - t_{k+1}} B_{k+1,2}(t)$$

Where $B_{k,1}(t) = \begin{cases} 1, t_k \le t \le t_{k+1} \\ 0, \text{others} \end{cases}$

According to the above definition of yuan, the basis functions of implicit function space is constructed: $B : R^3 \rightarrow R$,

$$B_k(o) = B(x, y, z) = B_{k_x, d_x}(x) B_{k_y, d_y}(y) B_{k_z, d_z}(z)$$

Where $k = (k_x, k_y, k_z)$, $d_x = d_y = d_z = 3$.

Projecting the implicit function IF onto the selected function space:

(1)
$$\widetilde{IF} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} cp_{i,j} B_{i,j}$$

On the type of N is the height or width of the matrix, $N \times N \ge M$, The part of the matrix which is greater than M are all zero vectors, The unknown control coefficient $cp_{i,j}$ is the prototype of the CP control matrix.

3.2. Decomposition of Face Region

The different areas of the face have different curvature change. The surface area whose curvature change is relatively flat and the surface area whose curvature change is relatively

sharp should adopt different resolutions, to minimize the time and space cost without damage in detail. For the resolution allocations of different regions, this paper doesn't adopt the adaptive multi-resolution structure which is similar to the multi-grid research [22] for two reasons. The first reason is that the reconstruction object of RIFF method of this paper involves only face, the structure is relatively fixed, there's no need to explore adjusting structure information in the process of reconstruction, it saves time cost undoubtedly. The second reason is that the RIFF method of this paper will be used for face recognition. Due to differences between different face structure, if we use adaptive resolution adjustment, the adopted resolution may be different in the same organ region, it will introduce non-equivalence factor in the face recognition.

This paper adopts the method of Gaussian curvature difference figure and the face is roughly divided into five face area, then the method of organ template matching is used in the precise segmentation, The realization of the specific implementation details will be described in the fourth quarter part in detail.

The different face region will be located in different depth nodes of the Octree, therefore these regions with different density use different control matrix:

$$\widetilde{IF}^{\mathrm{r}} = \sum_{i=0}^{N_r-1}\sum_{j=0}^{N_r-1} \mathrm{c}p^{\mathrm{r}}_{\mathrm{i,j}} B^r_{\mathrm{i\,j}}$$

where $B^{\mathrm{r}}_{i,j}(\mathrm{p}) = B^r(\frac{\mathrm{p}-cp_{ij}.center}{cp_{ij}.width})\frac{1}{cp_{ij}.width^3}$, $\mathrm{N_r}$ is the row width or column width of

control matrix in region r, $\sum_r \mathrm{N_r} = N$.

cp^r_{ij} is the node which is selected as control point in the Octree Θ , different depth with different bandwidth and different position with different node center. By three linear interpolation algorithm, the position coordinates of each control point is computed as the node position of Octree.

The evaluation of r in the last-written is 1~5, r respectively represents five different regions of eyes, mouth, nose, Smooth cheeks part, Cheek boundary.

Therefore, the implicit function of face surface can be represented as:

$$IF = \sum_r \lambda^r IF^{\mathrm{r}}$$

Where λ^{r} is the weight of each regions:

(2)
$$\lambda^{\mathrm{r}}(\mathrm{p}) = \sum_{o\in U_8(e),e\in CS^r} w_{o,s} B^{\mathrm{r}}_{o.i,o.j}(p)$$

Where $U_8(e)$ represents the eight neighbour nodes of element e, $w_{o,s}$ is three linear weighted value, CS^r is the original points of r region.

3.3. Face region integration

In order to generate the control matrix of the whole face, the first problem is how to represents the region with different resolutions which is said different control point density in one matrix; the second problem is how to solve the boundary cohesion problem between regions.

The max depth of these five region in Octree is D, $D_{\mathrm{r}} \leq D$, $\mathrm{r} = 1,\dots 5$, so the size of control matrix is $N \times N$, where $N = 2^D$. There are large of zero vectors in the control matrix. These zero vectors are composed of two parts primarily, one part is composed of zero elements which inside the control matrix and outside the face region; the other part is the zero vectors resulted by sparse distribution of region control points whose depth is less than D.

The figure 1 reflects the distribution of control coefficient of each region from arrangement position of each region of matrix A. The space is limited, so the matrix A don't show the detail of each region completely.
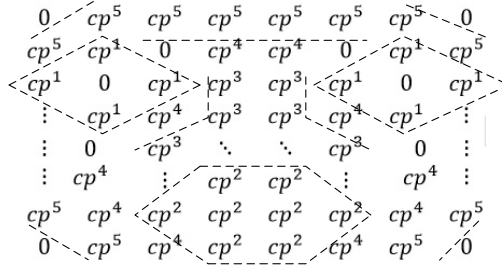
$$
\begin{matrix}
0 & cp^5 & cp^5 & cp^5 & cp^5 & cp^5 & cp^5 & 0 \\
cp^5 & cp^1 & 0 & cp^4 & cp^4 & 0 & cp^1 & cp^5 \\
cp^1 & 0 & cp^1 & cp^3 & cp^3 & cp^1 & 0 & cp^1 \\
\vdots & cp^1 & cp^4 & cp^3 & cp^3 & cp^4 & cp^1 & \vdots \\
\vdots & 0 & cp^3 & \ddots & \ddots & cp^3 & 0 & \vdots \\
\vdots & cp^4 & \vdots & cp^2 & cp^2 & \vdots & cp^4 & \vdots \\
cp^5 & cp^4 & cp^2 & cp^2 & cp^2 & cp^2 & cp^4 & cp^5 \\
0 & cp^5 & cp^4 & cp^2 & cp^2 & cp^4 & cp^5 & 0
\end{matrix}
$$

Figure 1. The Distribution of Face Control Matrix A

In the divided five regions, the edge jump is existent between eyes, nose, mouth, cheek border area with cheek smooth area, so the resolution of cheek smooth area inevitably belows the other parts. So the method to reduce control points density is external expansion. Taking the example of mouth, assuming matrix A is a matrix with the size of D*D, the mouth depth is $D_3 = D - 1$, the mouth and boundary region of matrix A which is shown in the left of figure 2 is amplified, the effect that the control point matrix expanded outward and reduced by once is described in the right of Figure 2.
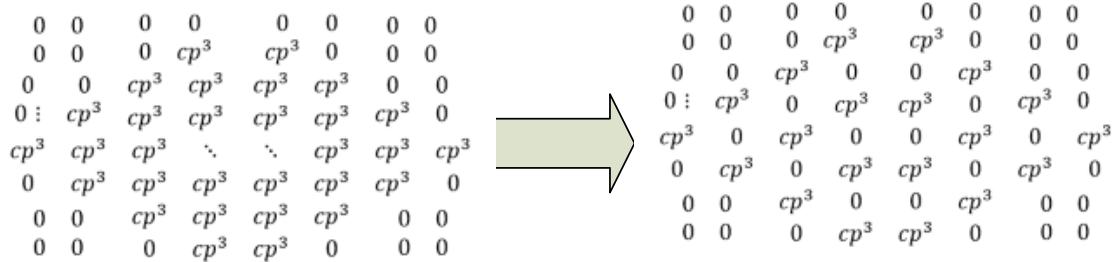

Figure 2. The Reduction Process of Control Matrix

Next, firstly we compute the distance d from A_{ij} to the center of A.c, $d = \sqrt{(i - c.i)^2 + (j - c.j)^2}$, then we compute the angle $\theta = \arctan((i - c.i - 1)/(j - c.j - 1))$, then we rotate the matrix with 45 degrees around the center of matrix and get the new matrix B, $B_{mn} = A_{ij}$, where $m = c.i + 1 + \lfloor d \cdot \cos(\theta + 45°) \rfloor$, $n = c.j + 1 + \lfloor d \cdot \sin(\theta + 45°) \rfloor$. Keeping the displacement vector after each region's center rotated $v_r = (r.d, r.\theta + 45)$, these vectors will be the weighted vectors of the composition of each region.

Figure 3 show the face structure relationship included in the control matrix after rotating 45 degrees from the apparent.
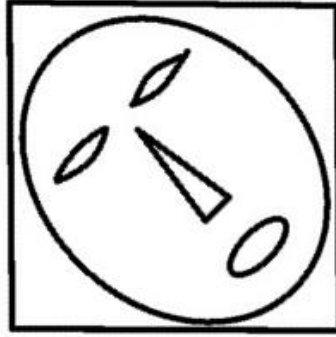
Figure 3. The Face Structure of Control Matrix

From Figure 1 and Figure 2, we can see that the control matrix is sparse symmetric matrices. This characteristic is the basis of the transition from control matrix to Laplacian matrix and the premise of dimensionality reduction.

3.4. Obtaining the Control Matrix

According to the definition of implicit function of each area, it can be acquired that:

$$\nabla(IF^{r})\,(p) \equiv \vec{N}^{r}(p)$$

where $\vec{N}^{r}(p)$ is the normal of each control point. Because the position coordinate of control point is acquired by three trilinear interpolation, and each region's depth is inversely proportional to this region's density of control point which is said the contribution of this region to the total control matrix, so:

$$\vec{N}^{r}(p) = \sum_{e \in cs^{r}, o \in U_{8}(e)} w_{e,s} B^{r}_{o.i,o.j}(p)e.\vec{N} \,/\, \lambda^{r}(p)$$

In order to solve the control coefficient of implicit function space, we take the divergence of both sides of the last formula, then we can solve the following formula using the classical Poisson equation:

(3)
$$\Delta IF^{r} = \nabla \cdot \vec{N}^{r}$$

We can transform this question to a minimum problem:

$$E(IF^{r}) = \int \left\| \Delta IF^{r}(p) - \nabla \cdot \vec{N}^{r}(p) \right\|^{2} dp$$

We describes the solving process by the form of matrix, that is to solve:

$$\min \left\| C^{r} y^{r} - b^{r} \right\|^{2}$$

Where $b^{r} = \nabla \cdot \vec{N}^{r}$, $C^{r}_{ij} = \dfrac{\partial^{2} B_{r}(o)}{\partial x^{2}} + \dfrac{\partial^{2} B_{r}(o)}{\partial y^{2}} + \dfrac{\partial^{2} B_{r}(o)}{\partial z^{2}}$.

we obtain y^{r} of each region, and map it to the control matrix of each region, namely $CP^{r}_{ij} = y^{r}_{(i-1) \times N_{r} + j}$.

According to the displacement information, we compose the control matrices of all regions and get $CP = \sum_{r} CP^{r} \cdot v_{r}$.

4. Implementation

4.1. Implementation Diagram

The RIFF algorithm can embed both the face reconstruction and the face recognition in the same frame which is demonstrated in Figure 4.
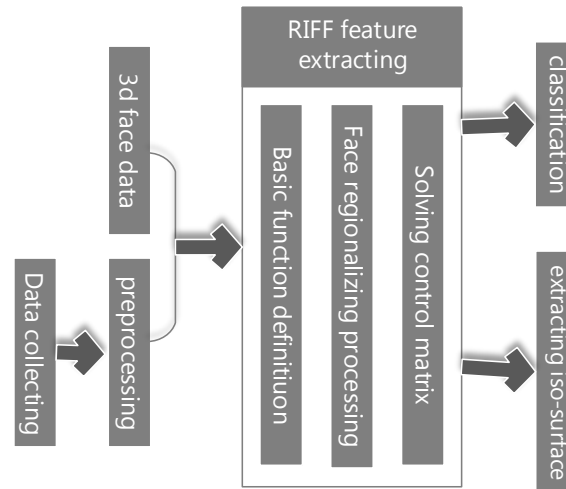

Figure 4. RIFF Implement System Block Diagram

In Figure 4, there are two inputs, kinect and cyberware, in the RIFF implementation framework. For the kinect, in order to acquire a pure human face, the preprocessing is needed since the image is collected in the natural environment, and will definitely contain complicated background noise.

To acquire a pure human face, the preprocessing module in Figure 4 contains a smoothing process for the point cloud data, including extracting the point cloud cluster from the complicated background and making an optimization or supplement based on the existing methods. After the preprocessing, the post-processing including the reconstruction and the recognition can be carried out.

The next module is the RIFF algorithm processing. After selecting a basis function space and defining the vector area of sampling points, the implicit function can be converted to be the Poisson equation. Then, we can obtain the coordinates coefficients of the implicit function in the selected basis function space through the Poisson equation to resolve the implicit function. In fact, the RIFF module also contains the facial region segmentation and fusion. The facial region can be divided into five subregions which will be reconstructed with different resolution values in order to optimize the use of memory space and reduce the redundant time overhead. Furthermore, this strategy of division distinguishes the influence to the class feature descriptor of the human face to improve the face recognition results. While the control matrix of the whole face obtained by the facial region fusion will get ready for the extraction of iso-surface and the calculation of category discriminant.

There are two outputs in the framework demonstrated in Figure 4. One utilizes the implicit function to extract iso-surface, and then gets the surface reconstruction. The other treats the control matrix as the source matrix whose Laplace transform matrix can be regarded as feature descriptor, and finally obtains the category discriminant, the quantized values that can be used to distinguish different faces, by projecting the feature descriptor into PCA space to reduce dimensions and performing the Kernel operation to transform the non-linear

classification to linear classification. Figure 4 also has demonstrated the dual nature of the RIFF algorithm since both those two outputs have repeatedly used the control matrix.

4.2. Preprocessing

Our paper uses the kinect equipment to collect the original point cloud to carry on the preprocessing. To acquire the 3D depth data, the kinect needs to emit the infrared laser light, and then receives the speckle information in the measuring volume through infrared cameras. Although the cost of this technology is low, the data acquired contains a lot of noise, and the situation of lacking data can easily be exacerbated because of the change of light.

Since the data collected by the kinect contains the complex character, background and noise information, firstly, the Euclidean cluster method is applied to extract point cloud clusters. In fact, the face is always positioned closest to the camera (otherwise the face will be blocked, which is beyond the scope of the paper), while those areas between the face and the back of the head cannot be blocked, therefore, we can filter out those point cloud data within the depth of 10-15cm from the nearest point. It's worth nothing that these data still contains noise information including holes, sharp points, etc, which need to be further smoothly dealt with. In this paper, we use the bilateral filtering and sampling interpolation methods to process the data. Moreover, the results after smooth processing may still contain the face and the body, and need to be further identified.

Since the geometry characteristics of point cloud can be used to identify the object, in this paper, we adopt the VFH feature descriptor which regards the point pair between the central point and other points on the surface as the evaluation objects, and achieves the target of object recognition by calculating the angle between the viewpoint direction and the estimated normal direction and the angle between viewpoint to central point direction and surface orientation. In fact, this feature can also detect the orientations of those objects whose feature contains a clear change of direction, which plays the fundamental role in the human face posture correction. Finally, we can extract the pure face in the scene point cloud clusters in real time through training the acquired VFH features of different faces to distinguish other point cloud clusters. The method proposed by this paper can also be applied to other human face extraction scenes.

4.3. Face Regionalization

In order to divide the pure face extracted according to the upper section into five parts more precisely, firstly, we need to calculate the Gaussian curvature difference map and roughly divide the organ region using the seed region growing algorithm. Then, the accurate boundary can be fitted out through the template matching method. Considering the requirement of precise division for those five face regions, this paper adds the nose template based on the literature GPSDA [22].

The nose template contains three parts, namely the left wing nose, the right wing nose and nose bridge.

$$z = a(\theta - b)^{2n} + c$$
$$z = a(\theta - b - L)^{2n} + c$$

Where $\theta \in [b, b + L]$, $z \in [c, \dfrac{z_{left\_eye} + z_{right\_eye}}{2}]$, and L denotes the length of nose.

After obtaining the fitting curves of each organ, we just need another two rounds of mapping to acquire the facial region.

So far, the division of eyes, mouth and nose regions have been done. The following will discuss the extraction of the cheek. Firstly, we need to extract the face to the lower intrinsic alignment coordinate system, and the boundaries according to the depth image calculated in the current coordinate system. Secondly, we gradually expand the neighborhood of boundaries inwards until the Gaussian curvature tends to be smooth. Then, the neighborhood will be just the cheek area. Furthermore, if we remove the eyes, mouth, nose and cheek regions, the rest is just the smooth cheek part.

According to the Gaussian curvature differential maps and the organ templates, we can extract five point cloud blocks including the eyes region, mouth region, nose region, smooth part of cheek region and cheek boundaries. Since the ears are often blocked, therefore, this module is ignored in our paper.

Actually, these five regions are reconstructed based on different resolution values. If the resolution value is too high, the time complexity is too large, while if the resolution value is too low, the effect is not that good. Therefore, in our paper, the optimal resolution value for each region will be selected as the intermediate value during the experiment.

Figure 5 demonstrates the resolution distribution of each organ in the human face depth map.
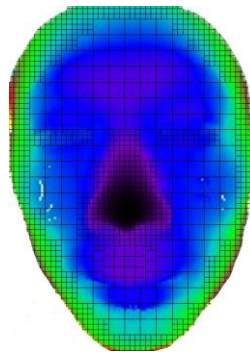

Figure 5. The Distribution of Face Resolution

Figure 5 roughly shows the resolution distribution of the eyes, mouth, nose, smooth part of the cheek and the cheek boundary. From Figure 5, The larger the Gaussian curvature changes, the larger density of resolution values there will be. The idea of adjusting resolution values according to the face expression can achieve the effective use of resources while ensuring the reconstruction details. Furthermore, the subregion face recognition results can still maintain robustness even when the face surface changes non-grid caused by expression.

Indeed, for these five regions, the generally range of resolution depends on the distribution property of the human face. Within the narrow boundaries of the cheek area, since the Gaussian curvature has a larger turning point, therefore, data-intensive reconstruction work is needed to ensure that the sharp edge features will not disappear. In fact, the edge of face plays an important role in the face recognition because it can well define the shape of the human face. Considering that the eyes region differs greatly from each other and the Gaussian curvature occurs within a narrower region than the cheek area, intensive data is also needed. As for the nose and mouth, because of the large area of the Gaussian curvature change and the non-rigid characteristics, high resolution is necessary in these areas to capture expression and represent the feature of the human face. While in the rest area, the Gaussian curvature changes slowly and less details about the reconstruction is needed, therefore, low resolution is

preferred in this area. Of course, the cheekbone bumps of some special human face on the cheek are quite obvious, so the resolution is not too low in this area.

For the resolution of these five regions, more precise selection depends on the experimental data, *i.e.*, it should consider the total time overhead, total space overhead and the face recognition results.

In the region dividing process, there is no need to make a too precise division for the boundary area, since the error of resolution at the boundary area will just lead to a constant level increase in terms of the time and space complexity. Furthermore, the slight changes in the posture and the facial expression just lead to the difference at the same region of the same person, and the difference is a very small proportion, whose influence to the global feature of the human face can be omitted.

According to the RIFF algorithm mentioned in Section III, the paper first defines the basic function space and maps the implicit function to the function space, and then extracts the face to calculate the control matrix at each region and finally combines them into the human face control matrix.

4.4. Face Reconstruction

Based on the IF^r calculated in Section III, the extraction of iso-surface S can be written as

$$S \equiv \{p \in R^3 \mid I\widetilde{F}(p) = \gamma\}, \gamma = \sum_r \frac{1}{\sum 1/\lambda^r(p)} \sum_{s \in CS^r} \frac{1}{\lambda^r(p)} I\widetilde{F}^r(s.p) \text{ Where}$$

$\lambda^r(p)$ denotes the impact factor of p in the region r.

4.5. Face Recognition

The essential of the indicator function IF that formulated in the reconstruction process is a mapping from three-dimensional space to four-dimensional space, and the iso-surface r, in fact, is just the flat four-dimensional space. Considering that the indicator function IF represents the construction information of the 3D surface and also bears the relative information between various parts of the face, this paper selects the indicator function IF as the feature descriptor of 3D face. Let's construct the intra-class Laplace matrix L_{intra} and the inter-class Laplace matrix L_{inter} in the non-Euclidean space, and expand the generic graph embedding framework in the literature [11]. Then,

$$L_{\mathrm{intra}} = D_{\mathrm{intra}} - A_{\mathrm{intra}}, \quad L_{\mathrm{inter}} = D_{\mathrm{inter}} - A_{\mathrm{inter}}$$

$$A^{ij}_{\mathrm{intra}} = \exp\{\|CP^{m,n}_{1_x} - CP^{k,t}_{1_x}\|^2 / s\}, \quad A^{ij}_{\mathrm{inter}} = \exp\{\|CP^{m,n}_{1_x} - CP^{k,t}_{1_y}\|^2 / s\}$$

Where both D_{intra}, D_{inter} are diagonal matrix, $D_{\mathrm{intra}}^{ii} = \sum_j A_{\mathrm{intra}}^{ij}$, $D_{\mathrm{inter}}^{ii} = \sum_j A_{\mathrm{inter}}^{ij}$.

Here $1_x, 1_y \in \{C_1 .. C_M\}, y \neq x$ denote the category labels, and $i = m \times N + n$, $j = k \times N + t$.

Both L_{intra} and L_{inter} are $N^2 \times N^2$ matrix, and N^2 denotes the node number of the octree. Here, N^2 is proportional to both the number of sampling points of the human face and the density distribution of the depth of each part. In fact, the control coefficients of L_{intra} and L_{inter} are related to the locations where the sampling points are obtained, and we need to

reduce the relativity between L_{intra} (L_{inter}) and the sample data so as to increase the relativity between L_{intra} (L_{inter}) and the intrinsic characteristic of the human face. Therefore, L_{intra} and L_{inter} are projected to the two-dimensional PCA subspace, i.e., $L_{\text{intra}}{}^{PCA} = W_{PCA}L_{\text{intra}}$ and $L_{\text{inter}}{}^{PCA} = W_{PCA}L_{\text{inter}}$. It's worth noting that in this situation, W_{PCA} compresses the reconstruction feature of the human face without changing its main structure.

Based on all those analysis mentioned above, in order to find the proper projection direction, our paper defines the optimization criterion of the projection matrix, that is

(4)
$$\alpha' = \arg \min_{\alpha} \sum_{i,j} \left[\left[\alpha^T X L_{\text{intra}}{}^{PCA} X^T \alpha \right] \times \left[\alpha^T X L_{\text{inter}}{}^{PCA} X^T \alpha \right]^{-1} \right]_{ij}^2$$

Where X represents the training control matrix.

Since a proper value of α should minimize the intra-class density while maximize the inter-class density, therefore, α needs to be iteratively optimized. After α converges to a stable state, it can be utilized to calculate the projection feature map Y:

$$Y = X^T \alpha$$

Finally, the face recognition process can be carried on through calculating the distance between different feature maps.

5. Experimental Results

This paper respectively uses the 3d face data collected by Kinect in time and the BJUT-3D database collected by Cyberware3030RGBPPS, and compare the differences of the two from the aspects of reconstruction performance, reconstruction tine and space complexity. Then we use FRGC V2.0 database to be the training set and test set and implement a face recognition system, the result indicates that different resolution or precision can't effect the recognition performance. So it is turned out that the implicit function feature extracted from RIFF method sufficiently characterizes the face structure information and ignores the information difference brought by measuring equipment and measuring environment.

5.1. Reconstruction Performance

Table 1 respectively gives out the time cost and memory cost of the 3d face data of kinect and BJUT-3D, the experiment platform is DELL M4600 PC. It's 8 core CPU, 12G memory, the RIFF reconstruction algorithm will use the 8 core CPU sufficiently.

Table 1. Cost

Collect Equipment	Point cloud number	Memory Peak/MB	Time cost/ms
kinect	30496	118	589
Cyberware	73303	324	1373

Where the time cost is the construction time of the afferent ordered point cloud, not include the acquisition process of the ordered point (that is the preprocessing and normal acquisition). It can be seen that the time cost and memory cost depend linearly on the number of point cloud.

Figure 6 respectively gives out the reconstruction effect of two kinds of collect equipment, (a) is the data collected from kinect, (b) is the data collected from Cyberware. The data from kinect is collected in the nature environment, including large amounts of holes, noisy point

and so on. The experiment smooth it first and retain the head data only but completely; and use the face data of BJUT-3D to reconstruct directly without any preprocessing.
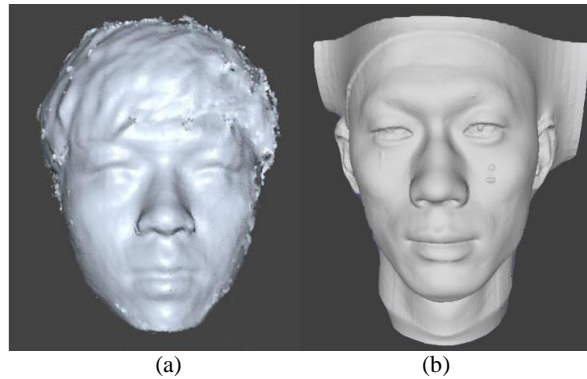

(a)   (b)

Figure 6. The Comparison of Reconstruction

From the Figure 6, we can see that the information collected by Cyberware is more accurate than kinect. Especially the eye area include more details, the edge of lip area is more obvious.

Table 2 respectively give out the comparison of RIFF method and the existing other methods, the test data use 73303 points. This paper uses the composition of resolution of (11, 10, 9, 7, 10). The eye, mouth, nose, smooth cheek part, cheek border area respectively choose the depth of 11,10,9,7,10.

Table 2. Comparison of Reconstruction Methods

Methods	Time cost(ms)	Memory Peak(MB)
Voronoi	4513	266
MarchingCubes	12947	471
RBF	16853	156
RIFF	1373	324

From the Table 2, we can see that the time cost of RIFF method is the lowest, and the RBF method has the lowest memory peak.

Figure 7 show the effect comparison of the algorithms in Table 2. Where (a) is the Voronoi method, (b) is the MarchingCubes method, (c) is the RBF method, (d) is the RIFF method of our paper.
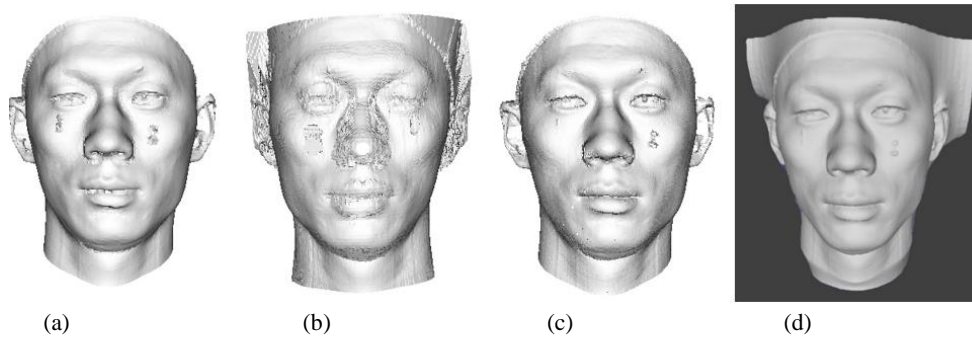

(a)   (b)   (c)   (d)

Figure 7. The Comparison of Reconstruction Effect

From figure 7 we can see that, the reconstruction result of Voronoi method is close to our RIFF method, but the interpolation effect of unknown database is slightly worse than RIFF method; Considering the time factor, because the RIFF method need to compute normal vectors in advance, this will spend about 2~3s, so the time cost of these two methods with 70 thousand points are similar; the reconstuction time of MarchingCubes is the longest and the reconstruction result contain much noise; the data smoothing effect of RBF method is slightly worse than the method of our paper; To sum up, comparing the RIFF method and the above-mentioned classical algorithm, our RIFF method acquire remarkable improvement in time cost and effect.

5.2. Recognition Performance

This paper use the 3d face data of FRGC-v2.0 to verify the classification effect of implicit function feature. The FRGC-v2.0 database contain 4007 3d face model of 466 people, include different ages, expression and races, the collection equipment is Minolta Vivid 900/910 series scanner. The paper divides the FRGC-v2.0 database into 3 groups, the first group(G1) take the 466 nature face to be the training library and the left 3541 faces to be the test set; the second group (G2) take the three faces of each people of 352 people to be the training library and the left 2747 faces to be the test set; the third group(G3) take the 6 faces of each people of 257 people to be the training library and the left 1794 faces to be the test set. We respectively extract the control matrix of each face in the training set to be one item of face data for system storage, then we compute the control matrix of each face of the test set in the recognition process, and acquire respectively the Laplace matrices of them and every items in the test set, finally, we compute the value of category discriminant to be the current category.

In order to choose an appropriate resolution composition, the experiment firstly chooses the data of the third group and gives out a lowest resolution composition(6,6,6,5,6) to be the initial value (The resolution compositions which are lower than this initial value are meaningless). We increase the resolution of various regions gradually to compare their recognition rate. In addition we choose the training set of 1000 dimensions, the recognition performance will be compared among different dimensions. The figure 8 is the recognition results of six representative resolution composition.
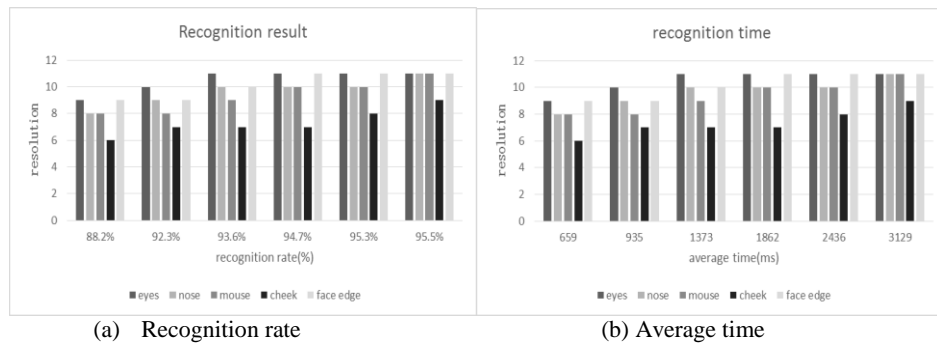

(a)   Recognition rate   (b) Average time

Figure 8. The Recognition Performance

The vertical ordinate of Figure 8(a) represents that different human organ selects different resolution to reconstruct, and the horizontal ordinate represents the recognition of various combinations. From the picture we can see that the recognition rate improves gradually as the resolution increases. The first resolution combination has significant changes compared with other combinations because its value is too low. Due to the variation of expressing in the

FRGC v2.0 database, the recognition of our method is not the best. How to recover the non-rigid transforming brought by the variation of expressing from the reconstruction matrix is our next study context.

The time costs of different resolutions are show as Figure 8 (b).

The horizontal ordinate of figure 8(b) represent the time costs of different resolution combinations, its combination mode is a one-to-one correspondence with figure 8(a). When the resolution combination arrives at the highest point, the average recognition time up to 3s and it will not satisfy the real-time identification requirements.

The figure 8(a) and figure 8(b) select six representative resolution combination. From the result, we can see that the recognition rate basically remains 95% after the fifth group. The recognition time is about 2s and it satisfies the real-time acquisition of face recognition system mainly.

We select the fifth data to be the representation of RIFF method and compare it with other algorithms. The Figure 9 compares the Rank-1 recognition result of different feature dimensionalities.
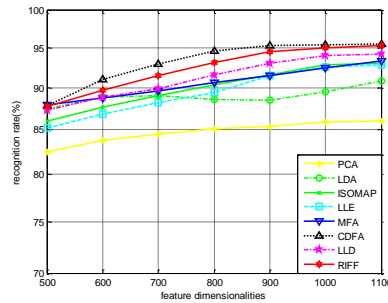

Figure 9. The Feature Dimensionalities

From the Figure 9, we can find that RIFF method is slightly less than the CDFA method. The recognition rate of RIFF method smooth-out after the 900 dimensions. We select the feature dimensionalities 1000 and compared the fifth group with other graphic embedded frame algorithms, the result is shown in table 3.

Table 3. Comparison of Reconstruction Methods

Algorithms	G3	Algorithms	G3
PCA	85.9	MFA	92.4
LDA	89.5	CDFA	95.5
ISOMAP	92.8	LLD	94
LLE	92.5	RIFF	95.3

Table 3 compared the RIFF method and other graphic embedded algorithms, the recognition of our method is only second to CDFA. The reconstruction matrix includes the elements that the 3d reconstruction matrix can resume to the matrix of nature expression when the expression varies. It is robust to the environment. Although the recognition rate is less than CDFA, it can improve significantly if we implanted the characteristic relative changes of facial areas in the matrix. Certainly it will introduce side effect and that will be solved in the subsequent study.

We respectively adjust G1, G2, G3 and get P1, P2, P3, where P1, P2, P3 test set respectively has 3541, 2951, 2465 faces. Figure 10 compares the relationship between the verification rate and false recept rate.
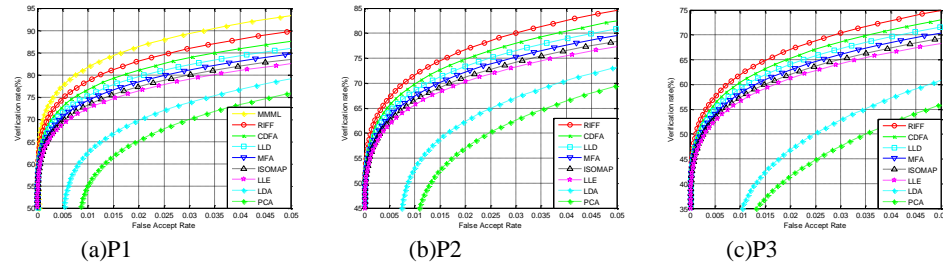
(a)P1   (b)P2   (c)P3

Figure 10. The ROC Curve

From Figure 10, we can see that the effect of MMML is the best in the P1 test group. That method is essentially used to single face data and the RIFF method is only second to MMML. In the P2, P3 test group, the RIFF method is the best, these two group can't use MLLL method to verify. And the verification rate in the group P1 is higher than the other two groups, so we can see that the more training data people use, the easier it can cause interference to the verification results of different consumer.

6. Conclusion and Future Work

The RIFF approach proposed in our paper accomplishes the face recognition process through representing the 3D face surface with the implicit function, resolving the implicit function with Poisson equation, and then treating implicit function as the feature descriptor of the 3D face. Furthermore, our approach also divides the whole face region into several subregions with different resolution values, which can greatly reduce the time required for reconstruction without decreasing the accuracy, and reflect the influence of different parts to the whole face recognition process, simultaneously.

Our experiments utilize the real-time 3D face data collected by the kinect and the classical 3D face data in BJUT-3D database to carry on the 3D reconstruction, and further analyze their performances of reconstruction. Then, we use the data in FRGC-v2.0 database as a training set and test set to observe the validity of RIFF algorithm in 3D face recognition. Experiment results demonstrate that our RIFF algorithm is better than most of the classical algorithms in terms of the recognition rate and authentication feature, however, it still has much room for improvement.

There are still some optimization works needed to be done:
- Since the lower half of the cheek can make great changes along with the expression, therefore, this region should be extracted separately.
- The biological properties of the organ accurately extracted have not been fully utilized.
- A reversible factor needs to be embeded into the reconstruction matrix that obtained after the same individual's expression changes to improve the robustness of the recognition effects.

References

[1] P. Alliez, D. Cohen-Steiner, Y. Tong and M Desbrun, "Voronoi-based variational reconstruction of unoriented point sets", In Proceedings of the Symposium on Geometry Processing, (**2007**), pp. 39–48.

[2] W. Y. Chang, "Surface Reconstruction from Points", Department of Computer Science and Engineering University of California, San Diego.

[3] J. Manson, G. Petrova and S. Schaefer, "Streaming surface reconstruction using wavelets", In Proceedings of the Symposium on Geometry Processing, (**2008**), pp. 1411–1420.

[4]   F. Calakli and G. Taubin, "SSD: Smooth signed distance surface reconstruction", Comput. Graph. Forum, vol. 30, no. 7, (**2011**).

[5]   M. Kazhdan, M. Bolitho and H. Hoppe, "Poisson surface reconstruction", In Proceedings of the Symposium on Geometry Processing, (**2006**), pp. 61–70.

[6]   G. Turk and J. O'Brien, "Modelling with implicit surfaces that interpolate", In TOG, (**2002**), pp. 855–873.

[7]   Y. Yan, H. Wang and D. Suter, "Multi-subregion based correlation filter bank for robust face recognition", Pattern Recognition, vol. 47, (**2014**), pp. 3487–3501.

[8]   Joliffe, "Principal Component Analysis", Springer-Verlag, New York, (**1986).**

[9]   H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data-with application to face recognition:, Pattern Recognition, vol. 34, (**2001**), pp. 2067-2070.

[10] J. Tenenbaum, V. Silva, and J. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction", Science, vol. 290, (**2000**) December 22.

[11] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", Science, vol. 290, (**2000**), December 22.

[12] S. Yan, D. Xu, B. Zhang and H.-J. Zhang, "Graph Embedding: A General Framework for Dimensionality Reduction, (CVPR'05).

[13] B. Tunc, V. Dagli and M. Gokmen, "Class dependent factor analysis and its application to face recognition", Pattern Recognition, vol. 45, (**2012**), pp. 4092–4102.

[14] D. Zhao, Z. Lin, R. Xiao, X. Tang, "Linear Laplacian Discrimination for Feature Extraction", Proceedings of Computer Vision and Pattern Recognition (CVPR07).

[15] H. Yan, J. Lu, X. Zhou and Y. Shang, "Multi-feature multi-manifold learning for single-sample face recognition", Neurocomputing (**2014**).

[16] Y. HanPang, A. B. J. Teoh and F. S. Hiew, "Locality Regularization Embedding for face verification", Pattern Recognition (**2014**).

[17] W. Li, Q. Ruan and J. Wan, "Dimensionality reduction using graph-embedded probability-based semi-supervised discriminant analysis", Neuro computing, vol. 138, (**2014**), pp. 283–296.

[18] Y. Kineri, M. Wang, H. Lin and T. Maekawa, "B-spline surface fitting by iterative geometric interpolation/approximation algorithms", Computer-Aided Design, vol. 44, (**2012**), pp. 697–708.

[19] A. Gálvez and A. Iglesias, "Particle swarm optimization for non-uniform rational B-spline surface reconstruction from clouds of 3D data points", Information Sciences, vol. 192, (**2012**), pp. 174–192.

[20] D.-J. Yoo, "Three-dimensional surface reconstruction of human bone using a B-spline based interpolation approach", Computer-Aided Design, vol. 43, (**2011**), pp. 934–947.

[21] H. Yoshihara, T. Yoshii, T. Shibutani and T. Maekawa, "Topologically robust B-spline surface reconstruction from point clouds using level set methods and iterative geometric fitting algorithms", Computer Aided Geometric Design, vol. 29, (**2012**), pp. 422–434.

[22] E. Grinspun, P. Krysl and P. Schröder, "Charms: a simple framework for adaptive simulation", In SIGGRAPH, (**2002**), pp. 281–290.

[23] X. Liu, "Face recognition technology research based on the three-dimensional model", Northwest University, (**2006**).