# 基于 Kinect 的拉普拉斯网格形变三维人脸建模

侯守明[1]， 杜成菲[1]， 王 阳[1,2]， 张玉珍[1]

(1. 河南理工大学计算机科学与技术学院，河南 焦作 454000；
2. 广东星辉天拓互动娱乐有限公司(杭州分公司)，浙江 杭州 310000)

**摘 要：**为了快速创建真实感较强的三维人脸模型，提出了基于 Kinect 的拉普拉斯网格形变建模方法。利用 Kinect 获取彩色和深度图像信息，对深度图像进行双边滤波处理，对彩色图像进行低层级顶点定位；构建标准三维人脸模型，并为该人脸模型中的顶点建立低、中、高 3 个级别的层级结构，通过低、中层级中顶点的位置关系创建 Sibson 局部坐标约束；利用该约束构建彩色图像中间层级顶点，并结合深度信息对标准三维人脸模型进行拉普拉斯网格变形，获得真实感较强的三维人脸模型。实验结果表明，该算法在建模的真实感上得到了提高，与对比算法相比，在建模时间上得到很大的优化。

**关 键 词：**Kinect；拉普拉斯网格形变；层级模型；人脸建模
**中图分类号：**TP 391.72 **DOI：**10.11996/JG.j.2095-302X.2018050970
**文献标识码：**A **文 章 编 号：**2095-302X(2018)05-0970-06

## Laplace’s Grid Deformation 3D Face Modeling Based on Kinect

HOU Shouming[1], DU Chengfei[1], WANG Yang[1,2], ZHANG Yuzhen[1]

(1. College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo Henan 454000, China;
2. Guangdong Xinghui Tiantuo Interactive Entertainment Corporation, Hangzhou Zhejiang 310000, China)

**Abstract:** In order to create a realistic 3D face model in a fast speed, a Laplacian mesh deformation modeling method based on Kinect is proposed. The Kinect is used to obtain color and depth image information, perform bilateral filtering on the depth image, perform low-level vertex positioning on the color image, construct a standard 3D face model, and establish low, middle, and high levels for the vertices in the face model. In the hierarchical structure, Sibson local coordinate constraints are created through the positional relationship of the vertices in the low and middle levels; this constraint is used to construct the middle level vertices of the color image and the Laplacian mesh deformation of the standard 3D face model is combined with the depth information to obtain a realistic 3D face model. The experimental results show that the proposed algorithm is improved in the realism of modeling, and compared with the comparative algorithms, the length of modeling time is greatly optimized.

**Keywords:** Kinect; Laplace’s grid deformation; hierarchical model; face modeling

三维人脸建模技术一直是计算机图形学与计算机视觉领域的热门研究课题，被广泛应用于游戏娱乐、影视动画、公安安全、医疗技术等多个领域，因此如何合成具有高度真实感的三维人脸是研究者不断追求的目标[1-2]。近年来，研究者将 Kinect 深度相机应用于三维人脸建模领域，并取得了一定的研究成果。

深度相机受限于自身廉价的深度成像原理，通

**收稿日期：**2018-02-11；**定稿日期：**2018-05-20
**基金项目：**国家自然基金青年基金项目(61503124)；河南省科技攻关项目(162102310090，172102210273)；河南省教育教学改革项目(2014SJGLX478)
**第一作者：**侯守明(1972–)，男，河南焦作人，教授，博士。主要研究方向为计算机图形学与 CAD、虚拟现实与仿真。E-mail：housm@163.com

常分辨率较低，且误差较大，如何有效地利用其获得的深度信息是使用深度相机进行建模最大的难题。HERNANDEZ 等[3]提出了一种融合多帧深度图像产生人脸模型的方法，用 Kinect 采集人脸深度图像，并通过人脸检测算法从深度图像中截取人脸部分，将分割出的人脸深度图像注册并集成到 3D 模型中完成建模工作。Kinect Fusion[4-5]是微软官方提供的一种利用 Kinect 进行三维重建的方案，是基于 RGB-D 相机三维重建的开山之作，其使用 TSDF 模型，通过迭代最近点算法，实时的融合深度图像重建三维模型，为三维建模提供了强有力的工具。MEYER 和 DO[6]在 Kinect Fusion 的基础上提出改进方法，通过快速的整合多帧图像完成三维人脸的重建。但上述方法需要比较强大的 CUDA 显卡的支持，对计算量的要求过大。文献[7]使用单个 Kinect 对正面人脸进行采集，并通过脸部特征点的位置将标准的人脸模型与相应的深度图像对齐，然后采用非刚性配准和改进的高斯牛顿算法对标准人脸进行优化变形，使其与深度图像最佳匹配以完成人脸建模工作。由于点云数据量较大，优化变形过程计算复杂，在建模时间效率上无法做到高效性。文献[8]将其方法整合了三维形变模型的思想，使用非刚性配准和基于 GPU 的快速非线性拟牛顿算法构建更完整、更细节的三维人脸模型。但该方法过于依赖 GPU，难以向移动平台扩展。

为了改进上述方法的不足，本文提出了基于 Kinect 的拉普拉斯网格形变人脸建模方法，利用 Kinect 获取彩色和深度图像信息，对彩色图像进行低层级特征点定位，创建标准人脸模型并对其进行层级化处理；并创建 Sibson 局部坐标关系，根据该关系创建人脸中间层级顶点，对标准人脸模型进行拉普拉斯网格形变，变形得到的人脸网格模型能够具备相应的人脸形状特征信息。本文提出的三维人脸建模方法能够有效构建特征化的三维人脸模型，具有较好的真实感，且所建立的人脸模型具有良好的网格顶点布线方式。同时，利用拉普拉斯网格形变算法建模的过程快速高效，无需 GPU 的参与，对移动端也具有通用性。

# 1　深度图像获取及处理

本文以 Kinect V2 获取彩色和深度图像信息。由于 Kinect 采集到的深度图像存在跳变现象，如果只对一幅深度图进行滤波处理，仍会出现跳变现象，本文对多帧(7 帧)深度数据做加权平均处理，以减小跳变现象。因此采用一种多帧加权平均的双边滤波算法[9]，以减小跳变现象对 Kinect 获得的深度数据的影响。

# 2　人脸低层级顶点定位

采用基于改进主动形状模型(active shape model, ASM)算法的堆叠修剪活动形状模型(stacked trimmed actived shape model, STASM)[10]算法对 Kinect 采集到的正面人脸彩色图像进行特征点定位构建局部特征模型，所确定的二维人脸特征点为低层级顶点，定位结果如图 1 所示。

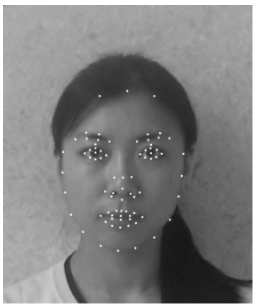


**图 1**　STASM 人脸特征点定位结果

# 3　标准人脸模型的构建与层级处理

## 3.1　标准人脸模型的构建

标准人脸模型通常以经典的 Candide-3 参数化人脸模型为基础，通过丰富细节来完成标准模型的建立。本文结合 Candide-3 的布线特点，以三角形网格的形式，建立由 1 347 个顶点和 2 630 个三角面组成的标准三维人脸模型，如图 2 所示。

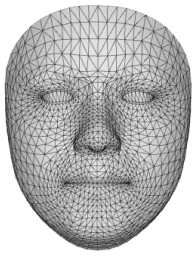


**图 2**　标准人脸模型

## 3.2　标准人脸模型的层级处理

借鉴多分辨率网格的思想，为标准人脸模型中的顶点手动建立低层、中间层和高层 3 个级别的层级结构。标准人脸模型中顶点的划分效果如图 3 所示，其中绿色顶点表示低层级顶点，红色顶点表示中间层级顶点，高层级包含所有的顶点。

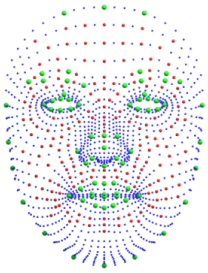


**图 3** 标准人脸模型的层级划分

(1) 低层级。其由 STASM 算法中定义的除去两个眼睛中心点剩余的 75 个特征点对应的顶点组成。定义眼睛两侧的两个顶点为 $M_1$、$M_2$，该两点作为约束人脸模型整体大小的参考点；定义鼻尖处顶点为 $M_3$，该点作为深度信息与人脸模型坐标对齐的参考点。低层级顶点分布如图 4 所示。

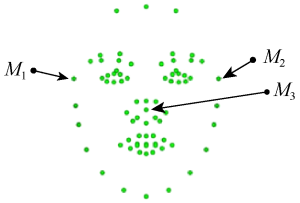



**图 4** 低层级顶点分布图

(2) 中间层级。Kinect 受限于自身的低分辨率和误差性，深度图像经过相应的平滑处理后，对于一些细节部位(鼻尖、眼睛等)，深度图仍难以有效反映出形状的细节信息。然而据观察，在深度变化不明显的“大表面”区域(额头、眼睛外轮廓等)，表面较为平滑、细节较少，深度信息较为平滑且稳定，可以获得较为可信的深度数据。同时这些部位的顶点能够勾勒出人脸的外轮廓信息，可以很好地反映人脸的形状特征。因此对于中间层级中顶点的选取，主要集中在深度信息可信的“大表面”区域。图 5 为中间层级中 163 个顶点的分布图。

(a) 正面图 (b) 侧面图

**图 5** 中间层级顶点分布图

(3) 高层级。其包含了人脸模型的所有顶点，以中间层级中顶点作为约束顶点对标准人脸模型进行拉普拉斯网格变形所得的结果，即是最终特征化人脸的建模结果。

## 4 Sibson 局部坐标系

标准人脸模型中低层级顶点与中间层级顶点具备一种二维位置约束关系。以低层级中顶点作为控制点在 Voronoi 图(V 图)的基础上构建二维的 Sibson 局部坐标系[11]表示两个层级间的二维位置对应关系。

## 5 人脸中间层级顶点

由于顶点的 Sibson 局部坐标考察的是顶点坐标间的二维位置比例关系，因此，在彩色图像坐标系下同样适用。根据标准人脸模型中低层级顶点与中间层级顶点间的 Sibson 关系(图 6)，可求得图 6(c)中间层级顶点对应的彩色二维图像坐标。

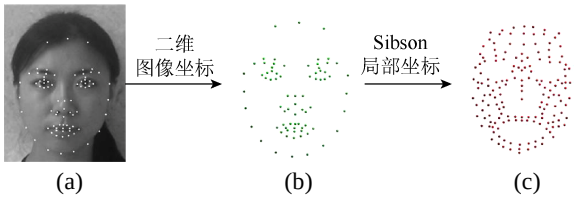



**图 6** 中间层级顶点的二维位置信息((a)STASM 算法提取的低层级顶点；(b)除去两眼中心点的低层级顶点；(c)中间层级顶点)

得到的中间层级顶点，将作为拉普拉斯变形中的控制顶点，同时在这些点处通过滤波处理的深度图获得较为准确的三维位置信息。

## 6 拉普拉斯网格形变算法

本文使用拉普拉斯网格变形算法，通过对一个标准的人脸模型添加测量者脸部的特征顶点信息作为约束条件，最终将标准人脸模型变形成为相匹配的特征化模型。

将三角网格模型表示为 $M=(E,V)$，其中 $E$ 为边的集合；$V$ 为顶点的集合。$V=(v_1,\cdots,v_n)$ 为网格模型中所有顶点的三维位置信息，$n$ 为模型中顶点的数量。点 $v_i$ 的拉普拉斯坐标 $\delta_i$ 可表示为

$$\delta_i=(\delta_i^x,\delta_i^y,\delta_i^z)=v-\frac{1}{d_i}\sum_{j\in N(i)}v_j \tag{1}$$

其中，$N(i)=\{j|(i,j)\in E\}$ 为 $v_i$ 的所有邻接顶点的集合；$d_i=|N(i)|$ 为顶点 $v_i$ 的度，即邻接顶点的个数。

求得网格 $M$ 中所有顶点的拉普拉斯坐标。设有矩阵 $\boldsymbol{L}$，使得 $\boldsymbol{LV}=\boldsymbol{\Delta}$，其中 $\boldsymbol{\Delta}$ 是顶点的拉普拉斯坐标组合的矩阵。

用网格的邻接矩阵 $\boldsymbol{A}$ 可表示网格中各顶点的连接关系，即

$$A_{ij}=\begin{cases}1, & (i,j)\in E\\ 0, & \text{otherwise}\end{cases} \tag{2}$$

当顶点 $v_i$ 与 $v_j$ 之间有边将其直接相连时，$A_{ij}=1$；当顶点 $v_i$ 与 $v_j$ 不直接相连时，$A_{ij}=0$。

用对角矩阵 $\boldsymbol{D}=diag(d_1,\cdots,d_n)$ 表示网格的度矩阵，其中 $d_i$ 为 $\boldsymbol{D}$ 对角线上的元素 $\boldsymbol{D}_{ii}$，所有非对角线上的元素 $D_{ij}=0(i\neq j)$。可以求得 $\boldsymbol{L}=\boldsymbol{I}-D^{-1}\boldsymbol{A}$ 满足 $\boldsymbol{L}V=\boldsymbol{\Delta}$，其中 $\boldsymbol{I}$ 为单位矩阵。$\boldsymbol{L}$ 即是网格 $M$ 的拉普拉斯转换矩阵，且满足

$$L_{ij}=\begin{cases}1, & i=j\\ -\dfrac{1}{d_i}, & (i,j)\in E\\ 0, & \text{others}\end{cases} \tag{3}$$

当网格中某些顶点的位置作为约束条件发生变化时，对于网格变形的过程，即是通过顶点已经建立好的拉普拉斯坐标反求绝对坐标的过程。由于 $\boldsymbol{L}$ 为奇异矩阵，因此不能直接通过 $V'=L^{-1}\boldsymbol{\Delta}$ 方式求得绝对坐标。

记 $F=\{k_1,k_2,\cdots,k_s\}$ 表示已知的 $s$ 个约束点对应的索引值的集合，此时可以根据已知的约束条件的顶点信息对 $\boldsymbol{L}V'=\boldsymbol{\Delta}$ 方程组进行扩充，使其有解

$$\boldsymbol{L}_C V'=\begin{pmatrix}\boldsymbol{L}\\ \boldsymbol{C}\end{pmatrix}V'=\boldsymbol{\Delta}_C=\begin{pmatrix}\boldsymbol{\Delta}\\ \boldsymbol{\delta}^C\end{pmatrix} \tag{4}$$

其中，$\boldsymbol{L}_C$ 为扩充后带约束的拉普拉斯变换矩阵；$\boldsymbol{C}$ 是一个 $s\times n$ 的矩阵，且满足

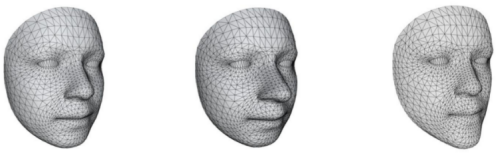
$$C_{ij}=\begin{cases}1, & j=k_i\in F\\ 0, & \text{others}\end{cases} \tag{5}$$

其中，$\boldsymbol{\Delta}_C$ 为扩充后的拉普拉斯坐标矩阵；$\boldsymbol{\delta}^C$ 是 $s$ 个约束点对应的拉普拉斯坐标组成的 $s\times 1$ 的列向量。

运用最小二乘法将等式 $\boldsymbol{L}_C V'=\boldsymbol{\Delta}_C$ 左右两边同时乘以 $\boldsymbol{L}_C$ 的转置矩阵 $\boldsymbol{L}_C^{\mathrm{T}}$，得到

$$(\boldsymbol{L}_C^{\mathrm{T}}\boldsymbol{L}_C)V'=\boldsymbol{L}_C^{\mathrm{T}}\boldsymbol{\Delta}_C \tag{6}$$

为了验证将拉普拉斯形变算法应用于人脸建模的可行性，本文选择不同位置区域的控制点集对人脸模型进行拉普拉斯变形结果如图 7 所示。其中，图 7(b)选择鼻尖附近的 4 个点作为控制点向外拉伸的变形效果；图 7(c)选择下巴处 6 个顶点作为控制点向下拉伸的变形效果。



(a) 原始模型　　(b) 变形一　　(c) 变形二

**图 7**　对人脸模型进行拉普拉斯网格变形效果图

## 7　基于 Kinect 的三维人脸建模流程

本文提出的三维人脸建模方法仅需单个 Kinect 对正面人脸进行信息采集，并根据 Kinect 深度图像特点，以脸部稀疏的顶点位置信息作为约束，对标准人脸模型进行拉普拉斯网格形变完成人脸建模。

图 8 展示了本文三维人脸建模的详细流程，具体如下：

(1) 使用 Kinect 获取正面人脸彩色和深度数据信息。

(2) 应用 STASM 特征点定位算法对正面人脸彩色图像进行特征点定位，确定低层级顶点对应的图像坐标信息；建立彩色图像坐标系与标准人脸模型坐标系之间的对应关系，并根据低层级与中间层级间的 Sibson 局部坐标对应关系求得中间层级中顶点对应的二维图像坐标。

(3) 通过深度图像信息及坐标系转换关系获得中间层级顶点在模型坐标系下的三维位置信息，并将其作为拉普拉斯网格变形的控制顶点对标准人脸进行变形，生成特征化的三维人脸模型。

(4) 通过纹理映射，完成最终真实感的三维人脸模型的建立。

## 8　实验及结果分析

本文使用基于 Kinect V2 深度相机与拉普拉斯网格形变算法相结合的建模方法，以 Unity3D 为三维平台，整合 Kinect For Windows 2.0、Visual Studio 2015 以及 OpenCV，组成以 C#为编程语言的混合编程环境。Kinect For Windows SDK 中整合了常用的源数据流获取、图像对齐等 Kinect 相关操作；OpenCV 被应用于深度图像处理以及人脸特征点定位工作，以 P/Invoke 平台调用的方式整合于 Unity3D，使得可以通过 C#语言进行相关功能的调用。

### 8.1　建模结果展示

图 9 展示了对 4 组不同的测量者进行人脸建模的效果。

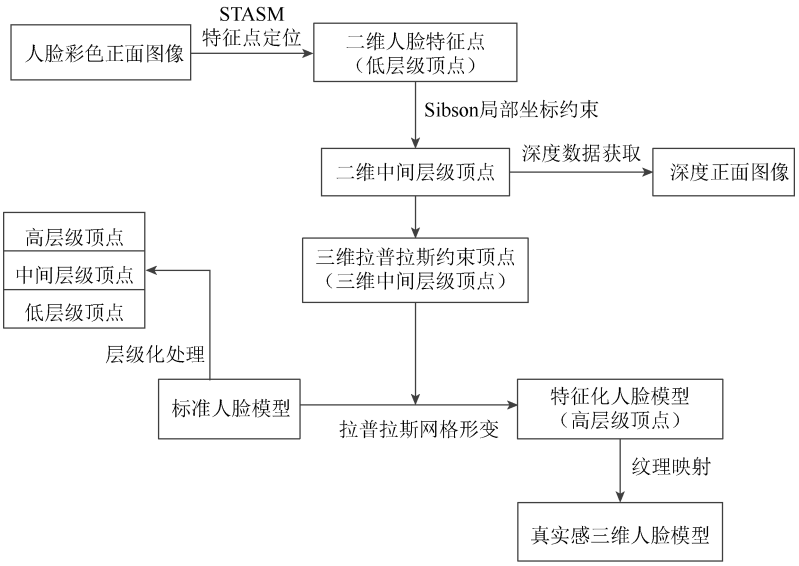
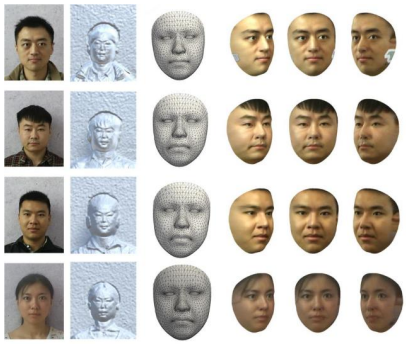



**图 8** 基于 Kinect 的三维人脸建模流程图



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

**图 9** 建模效果展示图((a)为正面人脸彩色图像；(b)为 Kinect 深度相机采集到的深度信息；(c)为所建立的特征化的人脸模型；(d)为纹理贴图后的结果)

## 8.2 建模性能分析

三维人脸建模的时间效率是衡量建模方法是否能推广到应用领域的一个重要指标。本文使用的 PC 端硬件环境为 Intel Core i5-6500 CPU 以及 NVIDIA Geforce GTX960 GPU，对于图 9 中展示的 4 组样本，每组进行 10 次重复建模实验测试，经统计建模各阶段平均时间消耗见表 1。

**表 1** PC 端各建模阶段时间效率统计

| 建模阶段 | 平均耗时(ms) | | | | |
|---|---|---|---|---|---|
| | 组 1 | 组 2 | 组 3 | 组 4 | 平均值 |
| 数据获取 | 312.6 | 319.2 | 322.3 | 318.7 | 318.2 (连续 7 帧数据) |
| 人脸特征点定位 | 41.6 | 44.1 | 46.7 | 40.0 | 43.1 |
| 拉普拉斯网格形变 | 110.8 | 98.1 | 102.5 | 101.0 | 103.7 |
| 总计 | 449.4 | 445.3 | 455.5 | 444.2 | 469.2 |

由表 1 数据可看出，由于要获取总计 7 帧深度数据，该建模方法在数据的获取环节消耗了大部分的时间；而由于使用了高效的拉普拉斯网格形变算法，同时又选用了顶点数较少的标准人脸模型，在真正的建模环节，仅使用了 0.15 s 完成。可见本文的人脸建模方法在时间效率上的高效性。

为了验证该算法对于移动平台的适用性，本文在 PC 端使用 Kinect 完成数据获取工作，并将数据记录，通过 Unity3D 发布于 Android 移动平台完成建模过程。使用的移动平台硬件环境为海思 Kirin950CPU、Mali-T880GPU，选择与表 1 相同的数据样本进行 10 次重复建模实验验证，表 2 为时间效率统计数据。

**表 2** Android 移动平台各阶段时间效率统计

| 建模阶段 | 平均耗时(ms) |
|---|---|
| 人脸特征点定位 | 313.5 |
| 拉普拉斯网格形变 | 834.6 |
| 总计 | 1 148.1 |

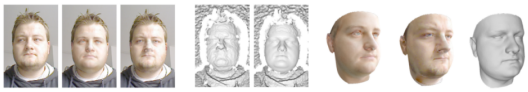
通过表 2 数据可见，在有效获取深度数据的前提下，本文方法在中高端移动设备上，可以在 1.2 s 内完成建模，在实时性和用户体验上具有很大的优势，并有很好的实用前景。

## 8.3 与同类建模方法之间的比较

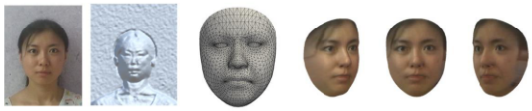
文献[7]、[8]的方法使用单个 Kinect 对正面人脸进行信息采集完成三维人脸建模，综合建模的真实感和时间效率两个方面将本文方法与其做比较。图 10 是 3 种不同的人脸建模方法的建模效果图。

表 3 为 3 种不同的人脸建模方法的建模情况对比。

(a) 文献[7]方法的建模效果



(b) 文献[8]方法的建模效果



(c) 本文方法的建模效果

**图 10**　不同方法建模效果对比图

**表 3　不同方法建模时间效率对比**

| 建模方法 | 硬件环境 | 建模时间(s) | GPU 参与 |
|---|---|---|---|
| 文献[7] | Intel Core i7-2700K CPU NVIDIA Geforce GTX690 GPU | 13 | 否 |
| 文献[8] | Intel Core i7-3770 CPU NVIDIA Geforce GTX780 GPU | 3~5 | 是 |
| 本文 | Intel Core i5-6500 CPU NVIDIA Geforce GTX960 GPU | 0.5 | 否 |

本文使用拉普拉斯网格形变完成建模，通过控制顶点的约束产生平滑的变形效果，可以较好地保证人脸各器官的尺寸形状特点，通过纹理贴图获得真实感较好的三维人脸模型。由于本文使用的拉普拉斯网格形变算法建模的过程快速高效，因此本文方法在时间效率上具有极大的优势，建模时间远远小于其他两种方法，且建模过程无需 GPU 的辅助，对于移动平台具有良好的适用性和应用前景。然而相比于其他两种建模方法，缺点是对于人脸器官的细节特点无法充分的体现。

## 9　结 束 语

本文基于 Kinect 深度相机对三维人脸建模展开研究，实现了一种高效的三维人脸建模方法，该方法通过结合深度相机采集到的彩色和深度图像获取脸部离散的顶点位置信息，并以此作为约束对一个标准的三维人脸模型进行拉普拉斯网格变形，变形得到的人脸网格模型通过纹理贴图可获得较好真实感的三维人脸模型。实验结果表明在保证真实感的同时，有效地提高了建模时间效率。但是由于获取的深度图像分辨率低，该方法只选取人脸平滑区域的控制点，致使眼睛、鼻等细节特征易依赖标准人脸模型，今后可将捕捉脸部细节特征作为进一步的研究方向，实现真实感更强的三维人脸建模。另外，进一步的研究可以考虑将纹理贴图网络化后同人脸特征点对齐，实现三维人脸模型和纹理的一体化形变。

## 参 考 文 献


[1] 崔微, 朱民耀, 颜柯, 等. 基于 Kinect 的三维重建技术在三维显示中的应用[J]. 电子测量技术, 2017, 40(3): 113-116.

[2] 董洪伟. 基于网格变形的从图像重建三维人脸[J]. 计算机辅助设计与图形学学报, 2012, 24(7): 932-940.

[3] HERNANDEZ M, CHOI J, MEDIONI G. Laser scan quality 3D face modeling using a low-cost depth camera [C]//Proceedings of the 20th European Signal Processing Conference. New York: IEEE Press, 2012: 1995-1999.

[4] IZADI S, KIM D, HILIGES O, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera [C]//Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. New York: ACM Press, 2011: 559-568.

[5] NEWCOMBE R A, IZADI S, HILLIGES O, et al. KinectFusion: real-time dense surface Software mapping and tracking [C]//Proceedings of 10th IEEE International Symposium on Mixed and Audmented Reality. New York: IEEE Press, 2012: 127-136.

[6] MEYER G P, DO M N. Real-time 3D face modeling with a commodity depth camera [C]//IEEE INternationnal Conference on Multimedia and Expo Workshops (ICMEW). New York: IEEE Press, 2013: 1-4.

[7] ZOLLHÖFER M, MARTINEK M, STAMMINGER M, et al. Automatic reconstruction of personalized avatars from 3D face scans [J]. Computer Animation and Virtual Worlds, 2012, 22(2-3): 195-202.

[8] ZOLLHÖFER M, THIES J, COLAIANNI M, et al. Interactive model‐based reconstruction of the human head using an RGB‐D sensor [J]. Computer Animation and Virtual Worlds, 2015, 25(3-4): 213-222.

[9] TOMASI C, MANDUCHI R. Bilateral filtering for gray and color images [C]//Sixth International Conference on Computer Vision. New York: IEEE Press, 1998: 839-846.

[10] MILBORROW S, NICOLLS F. Locating facial features with an extended active shape model [C]//10th European Conference on Computer Vison (ECCV 2008). Berlin: Springer, 2008: 504-513.

[11] MOCCOZET L, THALMANN N M. Dirichlet free-form deformations and their application to hand simulation [C]//Proceedings Computer Animation'97. New York: IEEE Press, 1997: 93-102.