

# 基于 Kinect 深度图像的指尖检测与手势识别<sup>①</sup>

高 晨, 张亚军

(北京化工大学 机电工程学院, 北京 100029)

**摘 要:** 针对基于普通摄像头的手势识别系统在不同光照条件和复杂环境下易受影响的问题, 提出一种基于 Kinect 深度图像进行指尖检测和手势识别的算法. 首先利用 Kinect 传感器获取深度图像, 再利用 OpenNI 手部跟踪器检测出手部的位置, 根据手部位置对手势进行深度阈值分割. 提出一种结合凸包和曲率检测指尖的算法, 检测出指尖数目和位置后, 计算出包括指尖和手掌水平方向的夹角、相邻两个指尖夹角以及指尖与掌心的距离的特征向量, 最后利用支持向量机(SVM)对预定的 9 种数字手势进行识别. 实验邀请 5 位实验者在复杂环境下每个手势做 30 次, 每次的手势角度不同, 实验结果表明该方法能够准确检测出指尖的数目和位置, 9 种数字手势平均识别率达到 97.1%, 该方法使用特征简单, 实时性好, 有较好的鲁棒性.

**关键词:** Kinect 传感器; 指尖检测; 支持向量机; 手势识别

## Fingertip Detection and Hand Gesture Recognition Based on Kinect Depth Image

GAO Chen, ZHANG Ya-Jun

(Beijing University of Chemical Technology, College of Mechanical and Electrical Engineering, Beijing 100029, China)

**Abstract:** Aiming at the problem that hand gesture recognition system based on ordinary camera is susceptible to the different lighting conditions and complex background, a fingertip detection and hand gesture recognition algorithm based on Kinect depth image is proposed. First, we get depth image by Kinect sensor. Then the hand region is extracted by putting the depth of thresholds on hand point detected by using OpenNI library. Fingertip detection based on convex hull and curvature is proposed. After the number of fingertips and the location of fingertips being detected, it calculates a feature vector including the number of fingers, the angles between fingertips and horizontal of the hand, the angles between two consecutive fingers, and the distance between fingertips and hand center point. Finally, a support vector machine(SVM) is applied to identify nine scheduled number hand gesture. Five experimenters are invited to perform 9 different hand gestures in the complex environment, which each gesture is repeated at thirty times and the angle of hand gesture is different every time. The experiment results show that this algorithm can detect the number and location of fingertips, and the recognition rate of nine hand gesture is 97.1% on average. This proposed method uses simple features and has good robustness, also it is real-time.

**Key words:** Kinect sensor; fingertip detection; support vector machine; hand gesture recognition

### 1 概述

不同于以往的鼠标、键盘等交互设备, 手势作为一种新的人机交互方式, 不需要直接接触, 这种人机交互方式更接近人与人之间的交流. 手势识别系统可以通过摄像头获取包含手势的图像信息, 根据图像信息解释手势的含义, 成熟的手势识别系统已经开始应

用于很多领域, 如体感游戏, 智能电视, 医疗设备, 手势识别和机器人控制等<sup>[1-3]</sup>.

采用普通摄像头获取图像信息进行手势识别容易受到光照和背景的影响, 针对此问题, 文献[4]通过使用用户佩带有颜色的手套来分割手势. 但是当背景有和手套颜色一样的物体时, 仍会影响手势的分割, 并

<sup>①</sup> 收稿时间:2016-07-31;收到修改稿时间:2016-09-23 [doi:10.15888/j.cnki.csa.005701]

且文献[4]使用的普通摄像头在光线不足的情况无法检测到手势. 而微软推出的 Kinect 摄像头可以同时获取彩色图像和深度图像信息, 利用深度图像进行手势识别可以避免光照和环境的影响.

近几年基于 Kinect 的指尖检测和手势识别已经有相关的研究, 文献[5]采用三维手势模型, 将手部分成 21 个不同的部分, 然后使用支持向量机(SVM)做分类器识别常用的手语. 但是其模型复杂, 运算量大, 实时性较差. 文献[6]和[7]都采用深度阈值分割手势区域, 默认手势区域是距离摄像头最近的, 但是当手掌和手臂垂直的情况下就会将手臂也分割出来, 降低手势识别的准确率. 文献[8]使用距离直方图来识别手势, 而文献[9]综合了距离特征和曲率特征使用多级支持向量机来识别手势. 文献[10]通过从深度图像中分割手部区域, 并建立表观特征, 再用决策树识别手势. 文献[8-10]使用的特征都较复杂, 需要的计算量较大.

针对上述问题本文提出一种把指尖数目, 手部方向, 掌心和指尖距离以及指尖相对于手部水平方向夹角作为特征向量, 再通过支持向量机的方法进行手势识别的算法, 如图 1 所示.

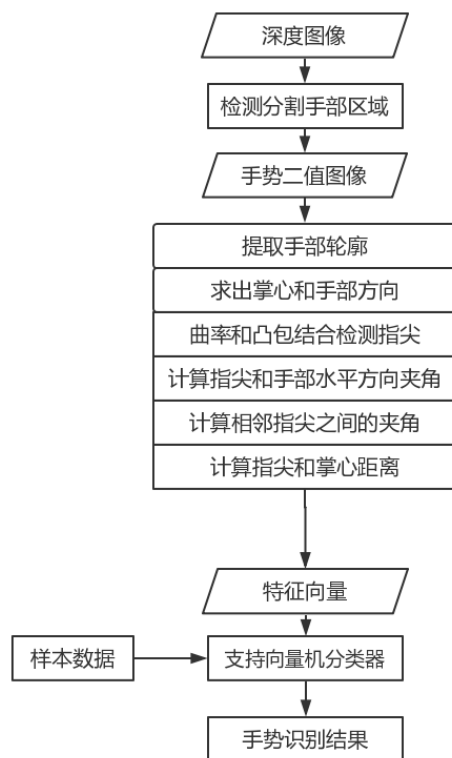


图 1 手势识别流程图

主要的工作步骤如下:

Step 1. 检测到手部区域, 将手部区域从复杂背景中分割出来;

Step 2. 根据分割出来的手势区域, 计算出手掌心、手部水平方向, 再结合凸包和曲率检测指尖, 最后计算出指尖相对于手部水平方向的夹角, 相邻指尖之间的夹角和指尖与掌心的距离;

Step3. 最后采集 9 种预设数字手势样本, 用支持向量机方法训练识别手势.

## 2 手势检测和分割

Kinect 摄像头可以同时获取深度图像和 RGB 彩色图像, 并且支持骨骼跟踪和动作识别. Kinect 一共有三个摄像头, 左边镜头为红外线发射器, 中间镜头是一般常见的 RGB 彩色摄像头, 右边镜头是由红外线 CMOS 摄像头所构成的 3D 深度传感器. 根据深度传感器可以获取摄像头一定范围内物体与摄像头的距离.

### 2.1 基于 OpenNI 的手势检测

通过开源库 OpenNI 可以检测到挥手、前推和举起三种预设手势动作并获得手部掌心的坐标. 本文采用识别效果较好的前推动作和挥动手势, 手部对着 Kinect 摄像头前推或者挥手时, 即可检测到手部区域, 并且给出掌心的粗略坐标, 识别效果如图 2 所示. 距离摄像头越远的点灰度值越大, 离摄像头远的点灰度值越小, 红色的点为识别出来的掌心位置, 从图 2 可以发现 OpenNI 给出的掌心位置不准确, 本文将根据手部轮廓的最大内切圆重新计算掌心坐标.



图 2 OpenNI 检测出掌心位置

### 2.2 基于深度阈值的手势分割

Kinect 获取到的深度数据为 16 位, 若使用不带用户 ID 的方式存储深度值, 则 16 位中的低 12 位为深度

值, 即最大值为 4096, 其代表 4.096m. 通过 OpenNI 获得的粗略掌心坐标包括二维坐标和深度值, 本文采用双深度阈值方法分割手势区域, 对掌心深度  $d$  在加减阈值  $T$  范围内的像素按照式(1)进行二值化, 公式如下所示:

$$H(x, y) = \begin{cases} 255, & d - T \leq D(x, y) \leq d + T \\ 0 & \end{cases} \quad (1)$$

式(1)中  $d$  为掌心的深度值,  $T$  为手掌厚度的深度阈值,  $H(x, y)$  为分割后手部深度图像中点  $(x, y)$  所对应的像素值,  $D(x, y)$  为深度图像中点  $(x, y)$  所对应的深度值.

为了避免手臂和手腕的干扰, 取距离掌心长宽一定的矩形作为包含手势的感兴趣区域, 分割效果如图 3 所示, 从图 3(b)中可以看到经过式(1)分割出的手部区域, 其边缘存在锯齿, 会对之后的指尖检测造成干扰, 针对此问题, 本文采用中值滤波和形态学开闭运算对其进行去除噪声处理, 使手部边缘平滑, 效果如图 3(c)所示.



(a)深度图像 (b)手部分割 (c)去除噪声

图 3 基于深度阈值的手势分割

### 3 手部特征提取

#### 3.1 手部边缘检测

为了获取手部轮廓的特征, 本文采用 Canny 边缘检测算法, 边缘检测即检测出图像亮度变化大的区域, Canny 边缘检测算法是对信噪比和定位之乘积的最优化逼近算子, 用来计算图像边缘的梯度幅度和相位值, 选取合理幅度阈值来确定边缘点<sup>[11]</sup>. 从抑制噪声方面来讲, Canny 边缘检测子不容易受到噪声的干扰, 而相同尺度下 LOG 算子却容易受到噪声的干扰, 抑制噪声的能力要弱一些. 此外, 由于 Canny 边缘检测算子采用两种不同的阈值分别检测强边缘和弱边缘, 并且当弱边缘和强边缘相连时, 只将弱边缘包含在输出图像中, 故而 Canny 算子更容易检测出真正的弱边缘, 因此本文采用 Canny 边缘检测来提取手部轮廓. 检测效果如图 4 所示.



图 4 Canny 边缘检测

#### 3.2 计算手掌掌心位置

通过 OpenNI 获得的手掌掌心坐标不够准确, 本文通过计算手部轮廓的最大内切圆来确定手掌掌心坐标和掌心圆半径. 首先计算手部轮廓内每个点到轮廓边界距离的最小值, 找出其中最大距离值所对应的点即为掌心点, 距离即掌心圆半径. 这个过程计算量会很大, 影响手势识别的实时性, 因此本文不计算感兴趣区间内的所有点, 而是每隔  $N$  个点计算一个点, 经过实验验证当  $N=4$  时可以保证速度和准确性. 求出的掌心圆如图 5 所示.



图 5 最大内切圆计算掌心圆

#### 3.3 计算手部水平方向

为了计算指尖的相对位置, 要先计算手部水平方向, 如图 6 所示,  $xh$  为手部水平方向,  $\alpha$  为手部水平方向和  $x$  轴方向的夹角,  $\theta$  为手部主轴和  $x$  轴方向的夹角, 这两个夹角可以根据文献[12]中提出的方法, 利用图像中心矩计算, 公式如下所示:

$$\theta = \frac{1}{2} \left( \arctan \left( \frac{2\mu(1,1)}{\mu(2,0) - \mu(0,2)} \right) \right) \quad (2)$$

$$\alpha = 90 - \theta \quad (3)$$

式(2)中  $\theta$  为手部主轴和水平方向的夹角,  $\alpha$  为手部水平方向和  $x$  轴水平方向的夹角,  $\mu(2,0)$  和  $\mu(0,2)$  为图像的二阶中心矩.

#### 3.4 指尖检测

##### 3.4.1 凸包结合曲率检测指尖算法

本文采用凸包算法与曲率检测算法相结合的算法, 可以精确的获得指尖的位置, 首先求出包围手部

轮廓的最小外接圆, 然后利用 OpenCV 计算机视觉开源库中函数 `convexHull` 和 `convexityDefects` 求出包含手部轮廓的凸包和凸缺陷, 手部轮廓的凸包和凸缺陷模型如图 7 所示, A、B、C 等为手部轮廓的凸缺陷, 凸缺陷包括起始点  $P_s$ 、终止点  $P_e$ 、缺陷上距离轮廓凸包最远的点  $p_d$ , 凸包缺陷深度  $I_d$ . 根据指尖的几何特征, 指尖点应该满足的条件如下所示:

- 1) 凸缺陷深度应该大于掌心圆半径, 但小于手部轮廓最小外接圆半径, 即  $r_a < I_d < r_b$ .
- 2) 凸缺陷起始点和终止点的夹角应小于  $90^\circ$ , 即  $\theta_a < 90^\circ$ .
- 3) 指尖点附近最小曲率值应小于  $60^\circ$ , 即  $\theta_k < 60^\circ$ .

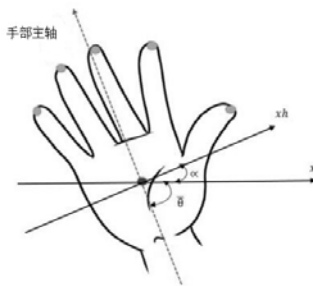


图 6 手部水平方向

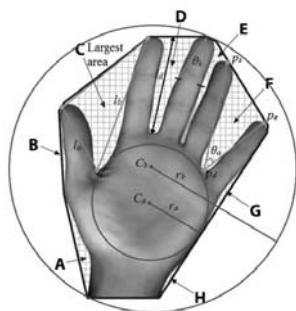


图 7 凸包和曲率检测指尖

检测指尖时首先将满足上述条件(1)和(2)的凸缺陷的起始点和终止点保存到点集  $A_p$  中, 并且删除重复和距离过近的点, 然后以点集  $A_p$  中的点  $P_i$  为原点, 在手部轮廓中各取  $P_i$  之前和之后的 10 个点, 计算这 20 个点的 K 曲率, 即向量  $(A_i(j), A_i(j-k))$  与  $(A_i(j), A_i(j+k))$  之间的夹角, 若点  $P_i$  的 K 曲率值小于  $60^\circ$  并且是 20 个点中 K 曲率值最小的点, 则其为指尖点. 本文算法相比较于传统的 K 曲率检测指尖运算量小, 只需要计算凸缺陷起始点和终止点极其附近的 20 个点的 K 曲率, 同时有效地排除了两个手指指尖的凹点, 通过实验验证当 K 取值 30 时效果最好.

上述三个条件虽然能较好的满足基本的指尖检测, 但是也存在一些不足, 如只有一个指尖时, 是没有凸缺陷满足条件 1) 和 2), 此时我们把凸缺陷起始点和终止点夹角 并且凸缺陷深度  $I_d$  大于阈值  $T_d$  的凸缺陷起始点和终止点存储到点集  $A_p$  中, 然后通过上述方法计算点集  $A_p$  中点的 K 曲率通过条件 3) 来检测指尖.

### 3.4.2 计算指尖点与手部水平方向夹角

在获得了掌心圆、手部方向, 指尖坐标等特征之后, 接下来计算指尖点和手部水平方向的夹角, 以及相邻两个指尖与掌心的夹角. 如图 8 所示, 分别计算每个指尖与手部水平方向的夹角  $\alpha[i]$  ( $i=0, 1, 2, 3, 4$ ), 在计算相邻两个指尖和掌心形成的夹角  $\beta[i]$  ( $i=0, 1, 2, 3$ ), 最后计算每个指尖和掌心指尖的距离  $D[i]$  ( $i=0, 1, 2, 3, 4$ ), 最后把计算得到的结果保存到特征向量数组中, 如式(4)所示:

$$V_f = \{N, \alpha[i], \beta[i], D[i]\} \quad (4)$$

式(4)中  $V_f$  储存特征向量的数组,  $N$  为指尖数目,  $\alpha[i]$  为每个指尖和手部水平方向的夹角,  $\beta[i]$  为相邻两个指尖的夹角,  $D[i]$  为每个指尖与掌心的距离.

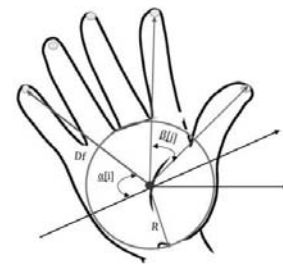


图 8 指尖夹角特征

## 4 手势训练和识别

在得到手部特征向量之后, 要设计分类器根据每种手势的特征对手势进行分类. 本文采用支持向量机作为分类器, 支持向量机具有能有效发现目标函数的全局最优解, 最大化决策边界等特点<sup>[13]</sup>. 在解决线性问题中, SVM 能直接在该空间中划分出分割平面, 但在实际问题中更多的是非线性的问题, 支持向量机首先在低维空间中完成计算, 然后通过核函数将输入空间映射到高维特征空间, 最终在高维特征空间中构造出最优分离超平面, 从而把平面上本身不好分的非线性数据分开, 常用的核函数包括多项式核函数、径向基(RBF)核函数和 Sigmoid 核函数. 本文采用径向基函数, 因为其模型简单, 参数相对较少, 可以有效的

把非线性问题转换为线性问题，并经过实验验证径向基核函数的效果最好。

### 5 实验结果和分析

为了验证本文提出方法的可靠性和准确性，本文使用内存 8G, CPU 为 Intel(R)Core™i5-4200H CPU @2.80GHz 的笔记本作为处理平台，使用微软 Kinect 一代摄像头采集图像，通过使用开源库 OpenNI、Opencv 进行编程，编程环境为 VS2015，实验包括指尖检测和手势识别两个部分，实验过程中光照环境不稳定，实验者与 Kinect 摄像头距离约为 1m。

#### 5.1 指尖检测实验分析

本文分别对指尖数目为 0 到 5 的六种情况进行测试，测试过程中，指尖的方向不断变化，测试结果如图 9 所示。然后将每种指尖数目的视频序列保存 1000 帧，每隔 5 帧提取 1 帧，共提取 200 帧，统计出检测结果如表 1 所示。

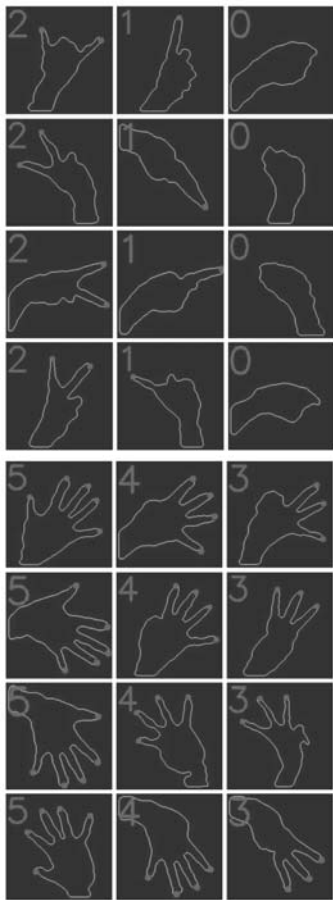


图 9 不同数目、方向指尖检测结果

表 1 指尖检测准确率

指尖数目	提取帧数	误检帧数	检测准确率(%)
0	200	1	99.5
1	200	6	97
2	200	4	98
3	200	5	97.5
4	200	3	98.5
5	200	2	99

通过图 9 和表 1 可以看出，本文提出的结合凸包和曲率检测指尖的算法平均准确率可以达到 97.9%，在手部运动的过程中能准确的检测出指尖的位置和数目，可以有效排除手腕处的干扰点，针对不同方向的指尖都可以准确检测出来，具有较好的鲁棒性。同时可以看出 1 个手指和三个手指的误判比其他的多，因为三个手指的情况较多，在手指没伸直的情况下，容易造成误判，而一个手指则因为无法使用凸包和曲率结合算法，容易受手腕干扰，不同光照下的检测结果如图 10 所示。



图 10 不同光照下指尖检测结果

#### 5.1 数字手势识别实验

为了验证本文提出的手势识别算法，本实验针对如图 11 所示的 9 种数字手势进行识别，实验前先构建训练样本数据库，邀请 5 位实验者针对每个手势做 10 次，要求每个手势的角度不同，总共采集 450 张样本图像，然后对样本图像进行训练，计算每张样本图像对应的特征向量并保存结果。训练完成之后，为了验证算法的鲁棒性，再邀请另外的 5 个人实时检测手势，每个手势做 30 次，即测试样本图像共 1350 张，每种手势的识别率如表 2 所示。

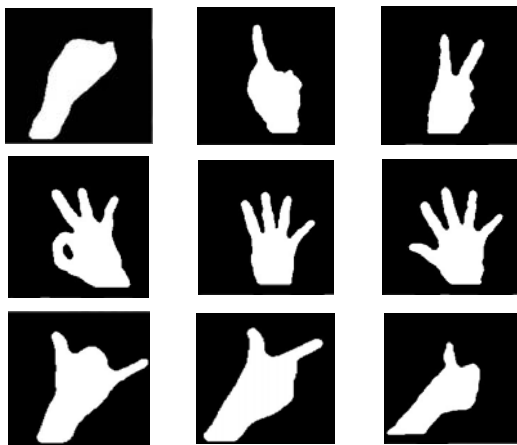


图 11 9种数字手势

表 2 9种数字手势检测结果

手势	测试帧数	误检帧数	手势识别率(%)
0	150	1	99.3
1	150	3	98
2	150	4	97.3
3	150	9	94
4	150	3	98
5	150	0	100
6	150	8	94.7
7	150	4	97.3
8	150	7	95.3

从图 11 和表 2 可以看出,用指尖与掌心的距离、指尖与手部水平方向的夹角以及相邻指尖之间的夹角作为特征向量可以明确的表示出每种手势的特征,平均手势识别率可以达到 97.1%,其中手势 3 和手势 6 误检率最高,因为手势 3 不是正对着摄像头时容易因为遮挡而识别错误,手势 6 因为在运动过程中小拇指会因为噪声干扰而误判为手势 8。每帧手势识别处理时间平均为 0.015 秒,实时性比较好。本文方法与常用的 HOG+SVM、几何特征、模板匹配方法 14 相比使用的特征简单,计算量小,准确率也相差不多。

## 6 结语

本文通过使用 Kinect 深度图像信息进行指尖检测和手势识别,有效地解决了在不同光照条件和复杂背景下指尖检测和手势识别的难题,提出一种结合凸包和曲率检测指尖算法,避免传统曲率检测误判点多和凸包检测位置不准确的缺点。检测出指尖后,提出一种需要样本少,方法简单的手势识别算法,该方法利用指尖与手部水平方向夹角,相邻指尖夹角以及指尖

和掌心距离作为特征向量,使用支持向量机做分类器。通过实验证明该算法手势识别精度较高,指尖检测较准确,能够在不同光照条件和复杂背景下进行检测,鲁棒性较好。

但是本文提出的方法在手指并在一起等一些特殊情况无法准确检测出指尖,OpenNI 的手部跟踪器效果不是很好,在有遮挡的情况会跟丢手部区域,在今后的工作中研究更精确,鲁棒性更好的跟踪算法。

## 参考文献

- 1 孙丽娟,张立材,郭彩龙.基于视觉的手势识别技术.计算机技术与发展,2008,18(10):214-216.
- 2 易靖国,程江华,库锡树.视觉手势识别综述.计算机科学,2016,43(6A):103-108.
- 3 李勇,高文,姚鸿勋.基于颜色手套的中国手指语字母的动静态识别.计算机工程与应用,2002,17:55-58.
- 4 Lamberti L, Camastra F. Real-time hand gesture recognition using a color glove. Springer Proc. of the 16th International Conference on Image Analysis and Processing, 2011, 6978(1): 365-373.
- 5 Keskin C, Kirac F, et al. Real time hand pose estimation using depth sensors. IEEE International Conference on Computer Vision Workshops, 2011, 28(5): 1228-1234.
- 6 王艳,张奇志.基于 Kinect 深度信息的手势识别.北京信息科技大学学报,2013,28(1): 22-26.
- 7 邓瑞,周玲玲,应忍冬.基于 Kinect 深度信息的手势提取与识别研究.计算机应用研究,2013,30(4):1263-1265.
- 8 Ren Z, Yuan J, Zhang Z. Robust hand gesture recognition based on finger earth mover's distance with a commodity depth camera. Proc. of ACM Conference on Multimedia. 2011. 1093-1096.
- 9 Domini F, Donadeo M, Marin G. Hand gesture recognition with depth data. ARTEMIS '13 Proc. of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream. 2013. 9-16
- 10 曹维清,李瑞峰,赵立军.基于深度图像技术的手势识别方法.计算机工程,2012,38(8):16-18.
- 11 薛丽霞,李涛,王佐成.一种自适应的 Canny 边缘检测算法.计算机应用研究,2012,27(9):3588-3590.
- 12 Liao SX, Pawlak M. Image Analysis by Moments. The University of Manitoba Winnipeg, 1993.
- 13 祁亨年.支持向量机及其应用研究综述.计算机工程,2004, 30(10):6-9.
- 14 刘淑萍,刘羽,淤俊,汪增福.结合手指检测的 HOG 特征的分层静态手势识别.中国图像图形学报,2015,20(6):781-788.