

# 一种基于 Kinect 的指尖检测算法

王劲东, 武 频

(上海大学 计算机工程与科学学院, 上海 200444)

**摘 要:** 指尖检测是人机交互过程中的关键技术, 因为人手的差异, 对指尖的检测总会存在一些误判点。文中在基于曲率算法的基础上, 提出了利用凸包算法和平行向量进行指尖检测的方法。该方法首先利用 Kinect 获取人体的骨骼信息和深度信息图像, 通过人手的关节点锁定手部位置, 并利用人手肤色特征和边缘检测算法提取手部区域轮廓。然后在手部区域的轮廓上根据曲率来检测类指尖点, 结合凸包计算排除凹点和手臂点, 最后根据手指的两侧接近平行的特性排除弯曲的手指或者非手指, 最终检测出有效的指尖。实验结果表明, 该方法在复杂背景下能够对不同的类指尖点进行排除, 并且有较高的检测精度。

**关键词:** 指尖检测; 曲率; 凸包; 平行向量; Kinect

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1673-629X(2016)07-0014-05

**doi:** 10.3969/j.issn.1673-629X.2016.07.004

## An Algorithm of Fingertip Detection Based on Kinect

WANG Jin-dong, WU Pin

(School of Computer Engineering and Science, Shanghai University,  
Shanghai 200444, China)

**Abstract:** Fingertip detection is a crucial technology in the process of human-computer interaction. Because of differences in human hands, there will always be some misjudgment points in fingertip detection. Based on curvature algorithm, a method of fingertip detection is proposed using convex-concave algorithm and parallel vector. Firstly, it obtains information of human bone and in-depth image using Kinect, locking hand position by joints of human hands, and extracts the hand contour area using color characteristics of human hand and edge detection algorithm. Then on the contour of the hand region, fingertips are detected according to the curvature, and combined with the convex hull, pits and arm points are computed and excluded. At last, according to the characteristic that both sides of the finger are nearly parallel, curved fingers or non-fingers are excluded, and valid fingertips are detected. Experimental results show that this method can exclude different classes of fingertips under complex background with higher detection accuracy.

**Key words:** fingertip detection; curvature; convex-concave; parallel vector; Kinect

## 0 引 言

人机交互<sup>[1]</sup>是指人与计算机可以用某种对话语言进行信息交换, 包括计算机通过输出设备或显示设备给人提供需要的信息, 以及人通过输入设备向计算机输入相关信息。目的讨论如何使设计的计算机帮助人们更加安全可靠地、更加有效率地完成所要完成的任务。用户借助必要的设备以自然的方式与虚拟环境中的对象进行交互作用、相互影响, 从而产生亲临真实环境的感受和体验。

手势作为人机交互过程中最重要的行为特征<sup>[2]</sup>,

已经广泛用于机器人控制、视频监控、多媒体教学和虚拟现实等。其中指尖特征又是手势识别中最重要的特征, 早期的指尖检测是借助一定设备, 如带有颜色的手套<sup>[3]</sup>或者数据手套<sup>[4]</sup>。目前对于指尖的检测都是基于形状的模板匹配方法和基于轮廓的边缘特性进行分析。基于形状的模板匹配方法有归一化互相关法<sup>[5]</sup>、广义霍夫变换法<sup>[6]</sup>等。基于轮廓的边缘特性进行分析的方法有很多。文献[7]提出的边缘曲率分析, 利用肤色提取出手部区域轮廓, 再根据曲率对指尖点进行检测。文献[8]使用改进的  $K$  向量方法, 结合轮廓阈

收稿日期: 2015-11-03

修回日期: 2016-03-02

网络出版时间: 2016-06-22

基金项目: 上海市科学技术计划资助项目 (14590500500)

作者简介: 王劲东 (1988-), 男, 硕士研究生, 研究方向为软件测试、并行计算、图像处理等; 武 频, 副教授, 博士, 研究方向为 CFD 数值计算、高性能计算、图像处理等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160622.0844.040.html>

值进行检测,实现食指的指尖检测。文献[9]在曲率检测的基础上,借助物理学中“势”和“场”的概念,把势太低的点判定为手臂点去除。文献[10]则利用手指的径向对称特性来检测各个指尖。随着 Kinect 的应用,一些开始基于 Kinect 的手势识别<sup>[11]</sup>也随之出现,如文献[12]。文中也是基于 Kinect 获取深度图像和目标人手部的关节点,再根据肤色分割获取手部区域图像,然后通过边缘膨胀和腐蚀处理,最后利用文中提出的指尖检测算法获取有效的指尖。

## 1 基于 Kinect 的深度图像和手部关节点的获取

Kinect<sup>[13]</sup>是微软公司在2010年推出的一款XBOX360体感外部设备,是目前世界上唯一一款较为成熟的商业人体动作感知设备。Kinect可以玩游戏,还可以应用在医疗、教育、三维人体建模等领域。另外还有许多富有创意的应用应运而生,现在越来越多的人开始对 Kinect 进行研究。

Kinect 在实际应用中最重要的是三只眼睛,依次为红外摄影机、彩色摄像头、红外深度摄像头。文中深度图像正是由红外摄影机和红外深度摄像头获得,由深度图像进行指尖检测可以克服光照和其他物体或者人物的干扰。图1(a)为 Kinect 获取的人体深度图像。Kinect 可以追踪到目标窗口里的六个人,并且能获得两个人骨骼信息“火柴人”<sup>[14]</sup>。第一代 Kinect 获取人体20个骨骼关节点,新一代 Kinect 可以获取人体25个骨骼关节点。Kinect 将获取的人物深度信息渲染后获得人体关节点信息,如图1(b)所示。



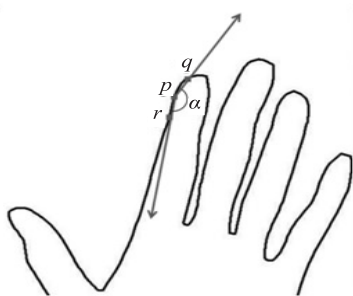
(a) 人体深度图像

(b) 人体的骨骼图像

图1 Kinect 获取的深度图像和人体骨骼图像



(a) 手部区域图像



(b) 手部轮廓曲率示意图



(c) 基于曲率获取类指尖点

图2 曲率检测指尖点示意图

文中主要用到的就是手部的关节点来追踪手部的位置,从而在复杂背景和复杂手势下进行指尖检测。

## 2 手部分割

### 2.1 目标检测人物手部的确定

Kinect 可以获取人体的骨骼信息,通过手部关节点可以获取人手的位置。但是 Kinect 可以同时检测到六个人,那么如何判定哪个才是需要检测的人手至关重要。这就要借助 Kinect 获取的深度信息,通过对不同人手部关节点的深度信息判定位于摄像头最近的手部关节点才是需要的,这样就可以排除目标检测人物后面还有其他人的干扰。

### 2.2 获取手部图像

通过深度信息可以获取目标检测人手,下一步就是进行手部分割,把手部轮廓从图像中分割出来,文中采用的是在线的自适应肤色模型<sup>[15]</sup>。首先以 Kinect 追踪到的手部关节点为圆心,以半径20 cm进行区域检测,获取手部肤色区域,提取其中的一小块作为样本。然后转换为 YCbCr 颜色空间,通过肤色聚类获取手部区域颜色空间分布进而分割出手部区域,在不同光照情况下进行检测具有很好的鲁棒性。在分割出手部轮廓边缘时会有一些噪声,手部的边缘会有一些锯齿和凸凹,需进行基于形态学的腐蚀和膨胀操作。最后利用中值滤波算法对手部边缘轮廓进行平滑处理,得到的手部区域图像如图2(a)所示。

## 3 基于曲率的二次修正指间检测算法

### 3.1 基于曲率的类指尖检测

根据肤色提取出来的手部区域轮廓点,对轮廓求曲率,因为指尖点周围的曲率都在一个固定范围内,所以可以通过阈值来筛选出类指尖点的集合。手部轮廓的曲率定义为:手部轮廓上一点  $p$ , 给出一个常数  $M$ , 取轮廓上点  $p$  的前  $M$  个点  $q$  和点  $p$  的后继第  $M$  个点  $r$ , countsize 为手轮廓像素点的总个数,那么  $p$  点的曲率可以用向量  $\vec{pq}$  和  $\vec{pr}$  的夹角  $\alpha$  的余弦值来表示,如图2(b)所示。

曲率的值如式(1)所示。

$$\cos\alpha = \frac{\vec{pq} \cdot \vec{pr}}{|\vec{pq}| |\vec{pr}|} \quad (1)$$

实验中取常数  $M$  为 5, 从公式可以看出曲率是在区间  $[-1, 1]$  之间, 所以采取一个阈值  $L$  来判断类指尖点将曲率大于等于  $L$  的像素点筛选出来。经过曲率检测所有的指尖点都能被检测出来, 但是也产生了很多误判点, 例如手指间的凹陷点和手部与手臂的一些切割点, 如图 2(c) 所示。

### 3.2 基于凸包算法的凹槽点和手臂点过滤

对于手部区域轮廓有很多的凸点和凹点, 通过手部轮廓的这种凸凹性可以很好地排除由曲率检测出来的类指尖点。首先介绍什么是凸包, 可以想象用一根绳子围住一片树木, 拉紧绳子就会形成一个多边形。可以把树木理解成平面上的点集, 那么绳子围成的多边形就是这些平面点集的最小凸包多边形。文中采用文献[16]提出的凸包算法, 求取手部轮廓的凸包点, 只有在凸包上的点和由 3.1 节中曲率检测后的类指尖点的交点才能第二次判断为类指尖点, 如图 3(a) 所示, 但是也把关节处切割点误判为指尖点。这里采用一种基于距离的判定方法来排除关节处的切割点, 即如果凸点到手部区关节点的距离大于该凸点到手部重心的距离, 即判定为指尖点。可以利用图像的几何特征矩来确定。图像的几何矩能够有效地描述对象的形态特征, 一个图像的  $(p, q)$  矩可以表示为:

$$M_{p,q} = \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy, p, q = 0, 1, 2, \dots \quad (2)$$

其中,  $x^p$  和  $y^q$  为变换核;  $f(x, y)$  为二维图像函数。

零阶矩代表图形轮廓的质量, 一阶矩代表轮廓的重心, 这里把手部轮廓像素信息描述为二维函数  $g(x, y)$ , 其重心计算过程如下:

$$M_{0,0} = \sum_x \sum_y g(x, y) \quad (3)$$

$$M_{1,0} = \sum_x \sum_y x * g(x, y) \quad (4)$$

$$M_{0,1} = \sum_x \sum_y y * g(x, y) \quad (5)$$

$$x_0 = M_{1,0} / M_{0,0} \quad (6)$$

$$y_0 = M_{0,1} / M_{0,0}$$

最终求得点  $(x_0, y_0)$  为手部的重心点, 如图 3(b) 中的手掌心的点。然后分别求每个凸点  $p$  到手部关节点的距离  $D_{\text{joint}}$  以及到手部重心点的距离  $D_{\text{center}}$ , 其中:

$$D_{\text{joint}} = \sqrt{(p_x - x_{\text{joint}})^2 + (p_y - y_{\text{joint}})^2} \quad (7)$$

$$D_{\text{center}} = \sqrt{(p_x - x_0)^2 + (p_y - y_0)^2} \quad (8)$$

若  $D_{\text{joint}} > D_{\text{center}}$ , 则此凸点为保留; 反之, 把此凸点排除。

从图 3(b) 中可以看到, 利用该距离判定法后, 能

有效地排除手臂关节处的切割点。

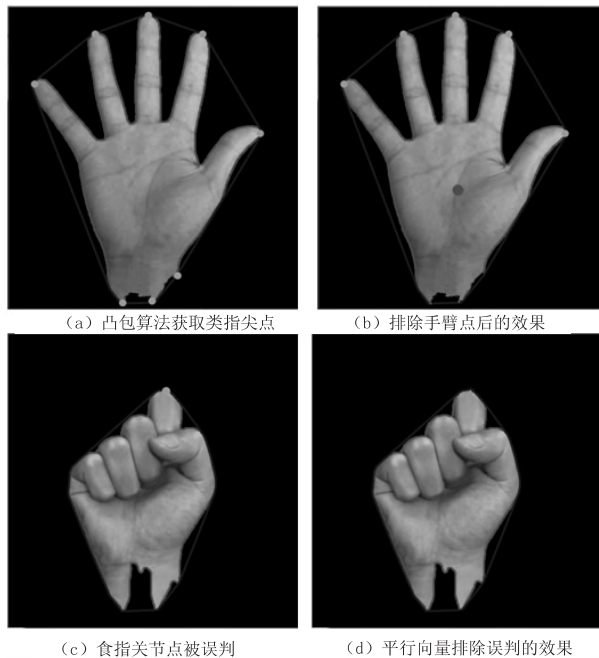


图 3 凸包检测算法和平行向量方法排除误判点

### 3.3 基于平行向量的手指确认

通过曲率检测和凸包计算可以找出所有的指尖点, 但是对于一些手部的特殊姿势仍然会造成一定的误判, 其中就包括手指弯曲的情况和手指的关节点。如图 3(c) 中的食指弯曲时的关节点, 该点经过曲率检测和凸包计算后, 曲率在阈值内, 也属于凸点, 所以仍被认为是指尖点。对于这种情况可以利用手指两侧边缘接近平行的特性去判断是不是有效的指尖。手指两侧的曲线可以看作是一组平行线, 当然手指两侧边缘不可能完全平行, 根据手指的差异性, 文中经过反复实验测试, 设定一个阈值  $20^\circ$ , 如果取得手指两侧的平行向量的夹角小于  $20^\circ$ , 则可以认为是有效手指。如何获取平行向量如下所述。3.1 节在曲率检测过程中在指尖位置两侧有两个阈值点, 从左侧阈值点向前取第  $N$  个像素点  $p_1$  和第  $N+10$  个像素点  $p_2$ , 同理从右侧的阈值点向后取第  $N$  个像素点  $q_1$  和第  $N+10$  个像素点  $q_2$ 。求取向量  $\vec{p_1 p_2}$  和向量  $\vec{q_1 q_2}$  之间的夹角, 若小于  $20^\circ$  则为平行向量, 那么该类指尖点为有效指尖点; 如果夹角大于  $20^\circ$ , 那么该类指尖点为无效指尖点, 如图 3(d) 所示。

经过实验测试, 文中  $N$  取值为 20, 如果  $N$  取值过大时对手指较短的情况发生误判, 如果  $N$  取值较小会对弯曲的手指误判为有效手指。

## 4 实验分析

为了验证算法的性能, 文中采用的硬件环境为二代 Kinect, CPU 为 Intel(R) Core(TM) i5-4590 CPU @



3.30 GHz、内存为 8 G 的笔记本,软件环境为 VS2015、OpenCV3.0。主要对人手的 9 种手势进行检测,如图 4 所示。

在复杂背景下,实验通过对不同的手势进行识别,并且同一个手势在不同位置方向都进行了测试。文中对图 4 中的 9 种手势进行了采样,利用 Kinect 获取图像信息,每个手势采取 200 幅图像,且这 200 幅图像中的手势一样,但是指尖的指向和手部关节点位置不是固定的。

不同手势的检测效果图见图 5。  
从图中可见,对于复杂背景和同一手势在不同位置方向的检测都达到了很好的效果,指尖的检测结果如表 1 所示。

由表 1 可以看出,该算法检测结果是比較高效的,对手势 0 全部检测出来了。还可以看出对手势 5 的检测误检率最高,这是因为在伸开五指时容易造成手指

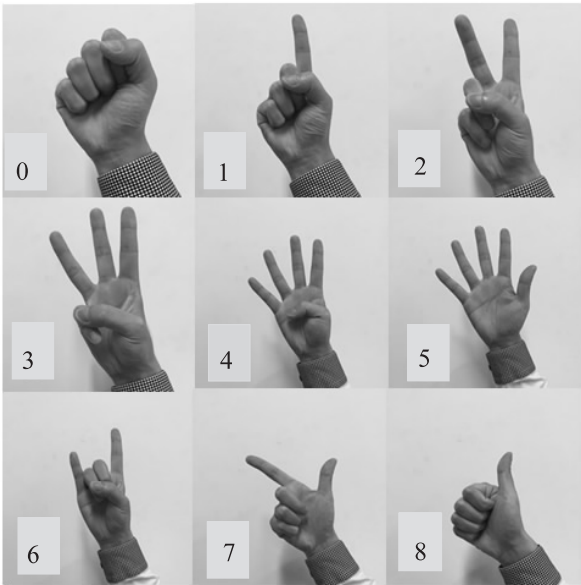


图 4 9 种手势

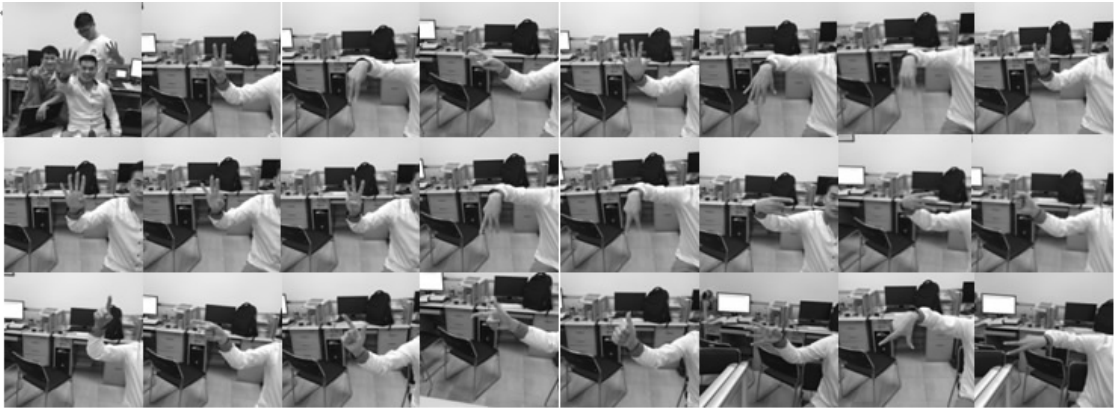


图 5 指尖的检测效果图

表 1 指尖的检测结果

手势	选取的帧数	误检帧数
0	200	0
1	200	1
2	200	3
3	200	2
4	200	3
5	200	6
6	200	2
7	200	3
8	200	5

之间的重叠和遮挡,其次就是手势 8 的检测,即对大拇指的检测。由于大拇指较短,在指尖指向不同位置时,造成了手指过短而形成误判。为了验证该算法的高效性,在同一环境下,将文中指尖检测方法的实验结果与基于曲率的检测方法<sup>[9]</sup>和基于人体骨骼和深度图像信息的指尖检测方法<sup>[12]</sup>进行比较。结果文献[9]中算法的检测准确率为 97.48%,文献[12]中算法的检测准

确率为 97.95%,文中算法的准确率达到了 98.63%。所以文中算法在检测的准确性上具有优越性。

5 结束语

文中提出了一种基于 Kinect 的指尖检测算法,通过 Kinect 获取的深度图像和关节点信息,可以很好地在手部区域分割时避免复杂背景环境的影响。文中利用手部关节点位置结合肤色特征进行手部区域分割,获得了很好的效果,并且对轮廓边缘进行膨胀和腐蚀处理获取相对平滑的轮廓边缘。在基于曲率初步检测的基础上获得初始类指尖点,通过基于凸包的距离计算排除手臂点和凹点,最后利用平行向量的思想判断类指尖点是否是弯曲的手指关节点或者其他干扰点。实验结果表明,该算法检测的精度较高,能在复杂背景和不同方向对指尖进行检测。

文中算法也有一定的缺陷,在手指并在一起时检测效果受到一定的影响。下一步的研究目标是对手势动作特征进行定义,通过实时追踪指尖点的位置信息

来识别动作,目前已经取得了一些研究进展和成果。

# 参考文献:

- [1] 宋鸣侨. 浅析人机交互技术的发展趋势[J]. 现代装饰:理论,2012(2):148-148.
  - [2] Carroll J M. Human-computer interaction: psychology as a science of design[J]. Annual Review of Psychology,1997,48(1):61-83.
  - [3] Iwai Y, Watanabe K, Yagi Y, et al. Gesture recognition using colored glove[C]//Proceedings of the 13th international conference on pattern recognition. [s. l.]: [s. n.], 1996:662-666.
  - [4] Weissmann J, Salomon R. Gesture recognition for virtual reality applications using data gloves and neural networks[C]//Proceedings of international joint conference on neural networks. [s. l.]: IEEE,1999:2043-2046.
  - [5] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of accepted conference on computer vision and pattern recognition. [s. l.]: [s. n.], 2001:511-518.
  - [6] Barrho J, Adam M, Kiencke U. Finger localization and classification in images based on generalized Hough transform and probabilistic models [C]//Proceedings of 9th international conference on control, automation, robotics and vision. [s. l.]: [s. n.], 2007:1-6.
  - [7] Lee D, Lee S G. Vision-based finger action recognition by angle detection and contour analysis[J]. ETRI Journal, 2011, 33(3):415-422.
  - [8] Guo K, Zhang M, Sun C, et al. 3D fingertip tracking algorithm based on computer vision[J]. Journal of Computer Research and Development, 2010, 47(6):1013-1019.
  - [9] 李博男, 林 凡. 基于曲率的指尖检测方法[J]. 南京航空航天大学学报, 2012, 44(4):587-591.
  - [10] 梅萍华, 李 斌, 朱中的, 等. 基于径向对称变换的实时指尖检测算法[J]. 中国科学技术大学学报, 2011, 41(2):101-107.
  - [11] 刘 佳, 郑 勇, 张小瑞, 等. 基于 Kinect 的手势跟踪概述[J]. 计算机应用研究, 2015, 32(7):1921-1925.
  - [12] 宋海声, 刘平和, 王全州, 等. 基于人体骨骼和深度图像信息的指尖检测方法[J]. 计算机工程与科学, 2014, 36(9):1788-1794.
  - [13] Wikipedia. Kinect[EB/OL]. 2011-01-13. <http://en.wikipedia.org/wiki/Kinect>.
  - [14] Clark R A, Pua Yong-Hao, Karine F, et al. Validity of the Microsoft Kinect for assessment of postural control[J]. Gait & Posture, 2012, 36(3):372-377.
  - [15] Dawod A Y, Abdullah J, Alam M J. Adaptive skin color model for hand segmentation[C]//Proceedings of international conference on computer applications and industrial electronics. [s. l.]: [s. n.], 2010:486-489.
  - [16] Homma K, Takenaka E. An image processing method for feature extraction of space-occupying lesions[J]. J Nucl Med, 1985, 26(12):1472-1477.
- +++++
- (上接第 13)
- IEEE, 2002:255-260.
  - [3] Ng J, Gong S. Composite support vector machines for detection of faces across views and pose estimation[J]. Image & Vision Computing, 2002, 20(5):359-368.
  - [4] Baggio D L, Emami S, Escriba D M, et al. Mastering OpenCV with practical computer vision projects[M]. Birmingham: Packt Publishing Ltd, 2012:208-254.
  - [5] 施 华. 头部姿态估计与跟踪系统的研究与实现[D]. 上海:华东师范大学, 2015.
  - [6] Darrell T, Moghaddam B, Pentland A P. Active face tracking and pose estimation in an interactive room[C]//Proc of 2013 IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE Computer Society, 1996:67-67.
  - [7] Hogg T, Rees D, Talhami H. Three-dimensional pose from two-dimensional images: a novel approach using synergetic networks[C]//Proceedings of IEEE international conference on neural networks. [s. l.]: IEEE, 1995:1140-1144.
  - [8] 余 凯, 贾 磊, 陈雨强, 等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展, 2013, 50(9):1799-1804.
  - [9] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[M]//Advances in neural information processing systems. [s. l.]: [s. n.], 2012:1097-1105.
  - [10] Luo P. Hierarchical face parsing via deep learning[C]//Proc of IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2012:2480-2487.
  - [11] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[C]//Proceedings of IEEE computer society conference on computer vision and pattern recognition. [s. l.]: IEEE, 2013:3476-3483.
  - [12] Zhu Z, Luo P, Wang X, et al. Deep learning identity-preserving face space[C]//Proceedings of 2013 IEEE international conference on computer vision. [s. l.]: IEEE Computer Society, 2013:113-120.
  - [13] 王 辉. 主成分分析及支持向量机在人脸识别中的应用[J]. 计算机技术与发展, 2006, 16(8):24-26.
  - [14] Huang G B. Learning hierarchical representations for face verification with convolutional deep belief networks[C]//Proc of IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2012:2518-2525.