

文章编号:1007-2780(2018)11-0958-07

基于人眼检测优化的立体人机交互技术的研究

薛钊鸿¹, 孙梓瀚¹, 成泽锐¹, 方 衡², 黄梓钊¹, 李佼洋², 蔡志岗², 王嘉辉^{2*}

(1. 中山大学 电子与信息工程学院, 广东 广州 510275;

2. 中山大学 物理学院 物理实验教学中心, 广东 广州 510275)

摘要:随着 3D 显示技术的逐渐成熟, 越来越多基于 3D 显示技术的产品被应用于各个领域, 例如 3D 电影、裸眼 3D 电视、体感交互式系统和影像教学等。基于 3D 显示的人机交互系统也是近年来越来越热门的研究方向, 但是却没有高效地利用深度信息作为交互的线索, 因此本文提出将 3D 显示与体感交互结合, 实现立体人机交互的虚拟键盘, 让用户只需点击屏幕前方键盘的图像即可进行对应按键信息的录入和编辑, 其帧率可达 10.6 fps。针对 3D 显示中用户视角偏离会导致按键的像在视觉上的偏移, 容易造成误输入的问题, 在系统中整合人眼检测功能, 修正因视角偏离带来按键定位误差, 从而实现了准确率 100% 的字符输入交互功能。

关键词:视角偏离修正; 立体显示; 人机交互; 人脸检测; 瞳孔位置估算

中图分类号: TN27 **文献标识码:** A **doi:** 10.3788/YJYXS20183311.0958

3D human-machine interaction based on human eye detection

XUE Zhao-hong¹, SUN Zi-han¹, CHENG Ze-rui¹, FANG Heng²,
HUANG Zi-zhao¹, LI Jiao-yang², CAI Zhi-gang², HUANG Jia-hui^{2*}

(1. School of Electronics and Information Technology, SunYat-sen University,
Guangzhou 510275, China;

2. Physics Experiment Teaching Center, School of Physics, Sun Yat-sen University,
Guangzhou 510275, China)

Abstract: With the growing of 3D display, more and more products based on 3D display are used in various fields, for example, 3D movies, auto-stereoscopic television, 3D surgery, etc. Human-computer interaction (HCI) system based on 3D display is also a popular research direction in recent years. However, most of these systems don't use depth information as a clue for interaction which limit interaction orientation into a 2D plane. To extend control on depth direction, a novel HCI system based on 3D display is presented. The system shows a keyboard style interaction interface by a 3D monitor, and uses binocular cameras to acquire the gesture and positions of the user's hand and fingers

收稿日期: 2018-04-12; **修订日期:** 2018-07-16.

基金项目: 广东省自然科学基金(No.2015A030310388)与(No.2016A030313333)

Supported by Doctoral Start-up Fund of Natural Science Foundation of Guangdong Province (No. 2015A030310388); Natural Science Foundation of Guangdong Province (No.2016A030313333).

* 通信联系人, E-mail: wangjh@mail.sysu.edu.cn

which can input information to the computer. And the system input frame rate can reach 10.6 fps. For problem that the movement of the user will cause the image of the button in his vision to be shifted correspondingly, which may cause incorrect input the human eye detection is integrated into the system. By the optimization of user's viewing angle calibration, we can correct button's positioning error, thus achieve a great accuracy of character input 100%.

Key words: visual angle deviation correction; 3D display; human-machine interaction; face detection; pupil position estimation

1 引言

近年来,随着立体(3D)显示技术取得突破进展,3D 技术和产品正革命性地影响和改变人们的沟通、工作和生活方式^[1-3]。3D 显示能提供精确的视觉深度线索,尤其是高分辨率、低串扰的 3D 显示更是深度信息的优秀载体。

体感技术也是近年来热门的新兴信息技术,在医疗卫生、金融保密、教育、游戏娱乐、零售等众多领域都有广泛的应用前景。但现在的体感技术在输入距离较长、无法充分利用深度信息进行交互等问题上有待继续提高。将 3D 显示与体感技术相结合,可在传统平面体感技术的基础上增加纵深方向的交互操作维度,不仅能覆盖较近的工作距离,而且更符合用户日常的操作习惯,有利于实现复杂、精细的操作。

鉴于此,本文提出了一种利用 3D 显示为载体的新型人机交互技术——虚拟键盘^[4],将需要交互的键盘的图像浮在屏幕前方显示。用户不用接触任何辅助器件,只需手指凌空点击键盘的像,即可输入和编辑字符。同时针对因 3D 显示引发的视角偏离造成手指定位误差问题,加入人眼检测定位进行修正,从而实现了具有纵深感、接近自然运动习惯的信息交互。

2 实验原理

2.1 双目视差产生立体视觉

3D 屏幕是本文立体人机交互系统的显示载体。生理学研究表明,人眼对于客观物体的深度感知最主要来源于视差^[5]。3D 显示就是利用两幅图像间视差,模拟出物体浮出屏幕的立体感。假设在屏幕上显示了目标的左右视角两幅子图像。这两幅子图像可被眼镜分光后传输到左右

眼。由于大脑的融合作用,即可产生目标出现在屏幕前方的立体感。而目标出屏距离 z 遵从以下表示:

$$z = \frac{SL}{S + D}, \quad (1)$$

式中,观众到观察屏距离为 L ,左右两幅视差图像间距是 S ,瞳距为 D 。因此,在实验中,我们利用了快门眼镜式 3D 屏幕显示交互场景的两幅子图像,将需要互动的目标物浮出屏幕前方,供用户进行操作。

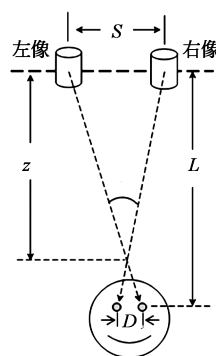


图 1 视差关系演示图

Fig.1 Parallax relationship diagram

2.2 双目摄像还原世界坐标

双目视觉则是一种实现目标精确定位重要技术。它模仿人的双眼,通过两枚摄像头从不同视角拍摄同一目标,并根据目标像点在左右两个摄像头中的图像坐标还原目标在真实空间中的世界坐标,可视为立体显示的逆过程^[6]。实验中,我们将两个摄像头沿水平线的 45° 放置,且摄像头所处的平面与交互画面所在的像平面重合,其结构如图 2 所示。系统利用此结构实时检测指尖并计算其世界坐标,再将指尖的世界坐标和各个按钮的像的世界坐标范围进行匹配,寻找出点击的按键,即可完成信息的输入。

以两摄像机连线的中点为直角坐标系 XOY

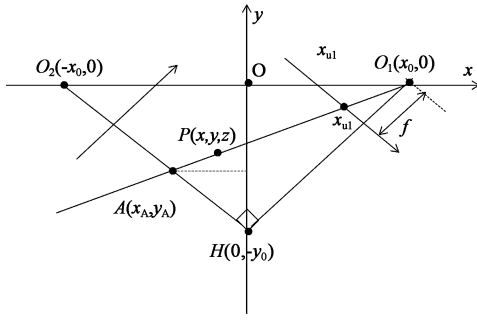


图 2 系统使用的双 45°双目摄像结构原理图

Fig.2 Double 45° binocular camera structure diagram used in the system

坐标原点,如图 2 所示,摄像机的像平面与水平面成 45°角,摄像机 1 和 2 的坐标分别为 $O_1(x_0, 0)$ 和 $O_2(-x_0, 0)$ 。摄像机像平面的 y_u 坐标体现了指尖在 z 轴方向的深度,用于确定图像中的指尖位置; x_u 坐标可用于求出指尖在 XOY 平面的坐标 $P(x, y)$ 。

当指尖处于 P 点时,由小孔摄像机模型下的理想透视投影变换可知,摄像机 1 的位置 $O_1(x_0, 0)$ 与 P 点在摄像机 1 的像点 (x_{u1}, y_{u1}) 以及交互画面像平面内的 $P(x, y)$ 三点共线,由几何关系得到:

$$\frac{\overline{O_1H}}{f} = \frac{\overline{AH}}{|x_{u1}|}, \quad (2)$$

式中: x_{u1} 为 P 点在摄像机 1 的像点的横坐标(坐标原点为像平面中心); f 为摄像机的焦距;点 A 是直线 O_1P 与摄像机 2 的法线的交点,点 H 为两摄像机法线的交点,有几何关系可知 $\angle O_1H O_2$ 为直角。再由摄像机与水平面成 45°角,可求得:

$$x_A = \frac{x_0 x_{u1}}{f}, \quad (3)$$

其中: A 点坐标应满足 O_2H 的直线方程,故可得到:

$$y_A = -\frac{x_0 x_{u1}}{f} - x_0, \quad (4)$$

则由两已知点 $A(x_A, y_A)$, $O_1(x_0, y_0)$ 可得到直线 O_1P 的方程如下:

$$\frac{y - 0}{0 + \frac{x_0 x_{u1}}{f} + x_0} = \frac{x - x_0}{x_0 - \frac{x_0 x_{u1}}{f}}, \quad (5)$$

同理,可求出直线 O_2P 的方程如下

$$\frac{y - 0}{0 - \frac{x_0 x_{u2}}{f} + x_0} = \frac{x + x_0}{-x_0 - \frac{x_0 x_{u2}}{f}}. \quad (6)$$

综上,由直线 O_1P 与直线 O_2P 均过点 P ,将式(5)与(6)两式联立,则可以根据 P 点在摄像机 1、2 的像点横坐标 x_{u1}, x_{u2} , 求出 P 点在虚拟触摸屏内的坐标 $P(x, y)$ 。

2.3 视角修正

根据 3D 成像原理,我们在 3D 屏幕上显示分属左右眼的两幅视差子图像。由图 3 可知,用户在不同的视角下,所看到的目标像点 P 具有不同的位置,所以在实际工作中,需要对用户视角进行实时检测,并修正各按钮的像的边沿世界坐标范围。

用户双眼平视屏幕中心时,双眼连线中点在屏幕上的投影为 XYZ 坐标系的坐标原点 O ,其中 XOY 为屏幕平面, Z 轴垂直于屏幕指向用户。设像点坐标为 $P(X_P, Y_P, Z_P)$ 。对应屏幕上的两点 $P_L(X_L, Y_L, Z_L)$ 和 $P_R(X_R, Y_R, Z_R)$ 。设左右眼所看到的图像在屏幕上的位差 $X_L - X_R = S$, 用户瞳距为 D ,与屏幕距离为 L ,成像示意图如图 3 所示。

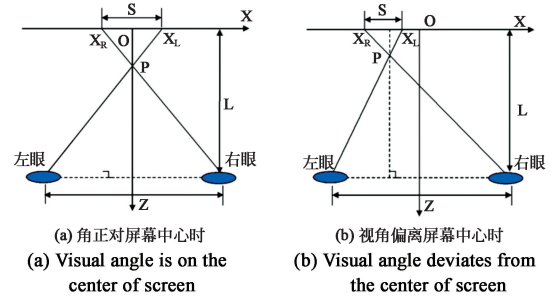


图 3 不同视角下目标像点的偏移情况

Fig.3 Target image point in the case of different visual angles offset

当用户平视屏幕中心时,由图 $x(a)$ 的相似三角形关系,得到图像出屏距离 Z_R 与位差 S 、瞳距 D 以及观察距离 L 有关,具体表达式为:

$$Z_P = \frac{S \cdot L}{S + D}, \quad (7)$$

而得到像点的横坐标 X_P 和纵坐标 Y_P 由屏幕对应点的横坐标 X_R 、瞳距 D 以及位差 S 决定,与观察距离 L 无关,具体表达式为:

$$X_P = \frac{S \cdot D}{2S + 2D} + \frac{D}{S + D} \cdot X_R, \quad (8)$$

$$Y_P = \frac{D}{S + D} \cdot Y_R. \quad (9)$$

当用户往水平(即 x)方向偏离屏幕中心,不再直视时,光路如图 3 (b)所示,根据几何三角关系,可以求得屏距离的公式如下:

$$Z_P = \frac{S \cdot L}{S + D}, \quad (10)$$

可见, Z_P 与直视屏幕中心时一致。由于用户只是侧视,所以纵坐标 Y_P 的对应关系也与直视屏幕中心时一致。像点横坐标 X_P 受到视角偏移量(即视差子图像中心线与瞳距中心线的间距)、瞳距 D 以及位差 S 有关,与观察距离 L 无关。最后进行坐标变换,设定世界坐标系 $X'Y'Z'$ 的原点为屏幕左下角。若双眼平视时,双眼连线中点在屏幕上的投影的计算机坐标为 (X_O, Y_O) , 则视角偏离后的像点横坐标修正公式为:

$$X_P' = \frac{S \cdot D}{2S + 2D} + \frac{D}{S + D} \cdot (X_R' - X_O) + X_O. \quad (11)$$

同理,当用户往竖直(即 y)方向偏离屏幕中心,不再直视时,其视角偏离后的像点纵坐标修正公式为:

$$Y_P' = \frac{D}{S + D} \cdot (Y_R' - Y_O) + Y_O. \quad (12)$$

由于像点出屏的深度坐标不受横纵坐标的影响,所以他它无需修正,表达式为:

$$Z_P' = \frac{S \cdot L}{S + D}. \quad (13)$$

根据上述的普适性修正公式,我们即可对不同视角下的像点偏离进行校正,从而减少人机互动位置的误差。

2.4 人脸检测与瞳孔估算

前述内容提到用户视角的偏离会导致目标像点随之偏移,所以在交互过程中先需要检测用户人脸位置,然后求得双目瞳孔的世界坐标,进而计算出观看距离 L , 结合上述两者才能完成目标像点在不同视角下的修正。

本系统基于开发时间和运算时间的综合考虑,采用了 Haar 特征作为目标特征值,利用 Ada-boost 以及 Cascade 算法进行分类器训练的方式实现人脸检测的功能^[7]。训练人脸识别分类器的图像往往是毫无遮挡、曝光良好、轮廓清晰的标准正面人脸,经由这些样本所训练得到的分类器同样只能用来识别不存在面部遮挡的人脸,这与系

统中由于佩戴快门式 3D 眼镜造成的双目遮挡相矛盾。所以,在实验中,我们在人脸检测的基础上,采用三庭五眼的先验法则实现瞳孔位置的推算。所谓三庭五眼^[8-10]指的是根据统计学得出的人五官在人脸面上的统计分布规律,如图 4 所示。根据三庭五眼的规律能简单高效地确定出双目区域在检测得到的人脸区域中的大致位置。

由于人眼瞳距基本在 60~65 mm 之前,不同人的双眼瞳距差距较小,实际工作中取平均值 62 mm 为瞳距 D , 这样根据前置摄像头所拍摄到的图像及人眼推算,就可以进行观看距离 Z 的粗定位,并且还原出用户双眼中心的世界坐标。

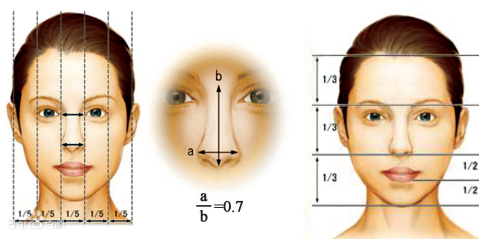


图 4 三庭五眼原理示意图

Fig.4 Schematic diagram of the five eyes of the three tribunals

3 系统搭建

本文搭建的基于人眼跟踪优化的立体人机交互系统——虚拟键盘,实物图如图 5 所示,由 2 个工业摄像机、1 个带 3D 显示功能的手提电脑及其 3D 眼镜组成。

用户戴上 3D 眼镜,可看到浮出显示器的人机交互画面——1 个 12 宫格键盘(键值为数字 0~9、退格和回车)。当用户用手指点击交互画面中的按键时,可在 2 个摄像机采集的图像中分别检测到指尖并得到指尖的图像坐标,计算出指尖在虚拟触摸屏平面内的坐标。

同时,手提电脑前置摄像头会实时采集用户的位置,检测出用户的人脸的区域,结合三庭五眼的先验知识,估算出瞳孔的世界坐标。根据两个瞳孔的世界坐标,系统可以给出每个按键的世界坐标范围修正值。利用按键修正后的世界坐标范围与指尖的世界坐标匹配,即可确认指尖点击的按键值。最后通过手提电脑的屏幕刷新显示,就可在输入框中显示出点击的按键值。

表 1 系统关键部件列表
Tab.1 List of key system

名称	型号	参 数
手提电脑	华硕 G51J 3D	i7 720QM,4G 内存,Nvidia 时分式 3D 显示技术,前置摄像头 200 万像素
工业摄像头	大恒 HV1303UC	15fps@130 万像素,USB2.0
3D 眼镜	Nvidia 3D Vision	液晶镜片,单眼刷新率 60 Hz
交互画面	自行绘制键盘	内容:数字 0 至 9,退格和回车;出屏距离:8 cm,按钮尺寸:4 cm×3 cm

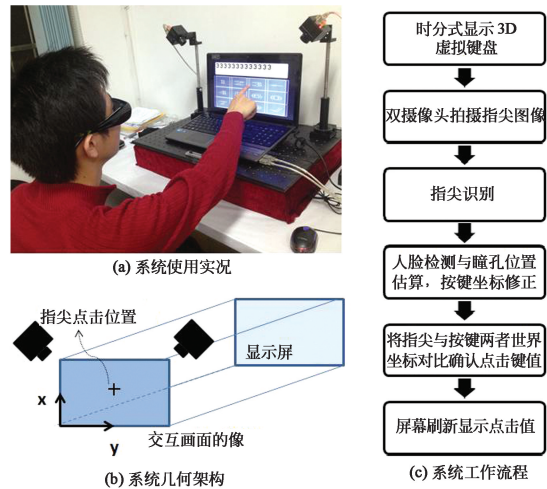


图 5 基于人眼跟踪优化的立体人机交互系统与工作原理
Fig.5 3D human-machine interaction system and working principle based on human eye tracking

4 实验结果

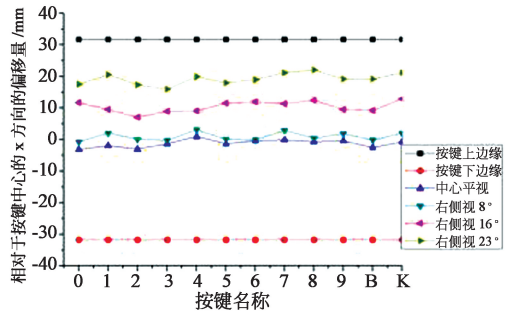
4.1 人机交互实验结果

本项目所研发的立体人机交互系统将虚拟键盘的影像在屏幕前方进行立体显示,能让用户凭借手指凌空点击虚拟键盘的立体影像即可准确、实时地完成数字 0 至 9 的输入和屏幕刷新显示,也可以通过对退格键和回车键的点击,实现输入字符的删去和回车换行。其字符输入速率可达到 10.6 fps。

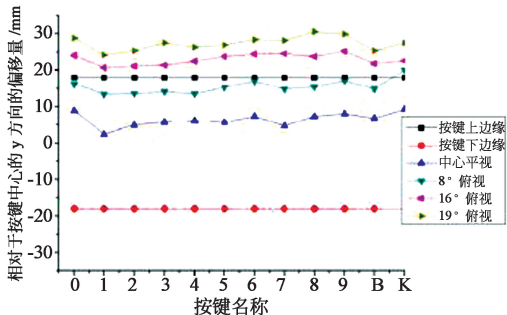
为了验证用户的点击准确性,我们让 5 名用户在 8 个视角(观看距离 50 cm,视角选择为右侧视:0°、8°、16°和 23°;俯视 0°、8°、16°和 19°)下,对每个按键点击 10 次。每次点击后计算指尖点击位置世界坐标和交互画面中对应按键的中心位置世界坐标的偏移量,综合 5 名用户平均值得到每

个按钮的点击偏移量。偏移量小于按键的边缘范围时,表示用户仍点在按键内,完成正确输入;反之,则认为点击了其他按键,引起了误输入。

在未加入视差修正算法前,用户的交互准确性结果,即点击位置与按键中心位置的误差关系如图 6 所示。当用户正对屏幕(即视角在水平和俯仰方向均为 0°)时,手指点击位置与按键中心理论位置两者世界坐标在 x 和 y 方向的差距小,能实现 100%正确的输入;但是用户观看位置逐渐偏移时,随着人眼偏移角度的增大,指尖点击位



(a) x 方向点击误差
(a) X-direction click error



(b) y 方向点击误差
(b) Y-direction click error

图 6 用户点击位置误差与观看角度的关系
Fig.6 Relationship between user click position error and visual angle

置与系统算出的按键中心的偏差也在增大。特别是 y 坐标的偏移,可以看到当俯视角大于 16° 时,指尖点击位置已经超出了对应按键范围,造成了很高的错误点击率。因此,证明需要在系统中加入人眼的定位对按键在不同视角下的位置范围加以校正。

4.2 按键视差修正效果

按照前述的原理,我们利用三庭五眼的先验知识,对人眼的瞳孔加以定位,其效果如图 7 所示。根据两个瞳孔的图像坐标和瞳距的均值,可以计算出用户所在视角下每个按钮的边缘世界坐标范围。由于每个按钮的尺寸较大,所以尽管瞳孔世界坐标推算有一定误差,导致每个按钮的边缘世界坐标范围受到影响,但是仍在可以接受的范围内。图 8 所示为经过视差修正后,用户在不

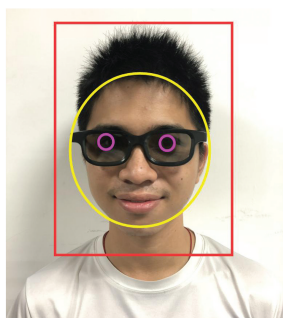
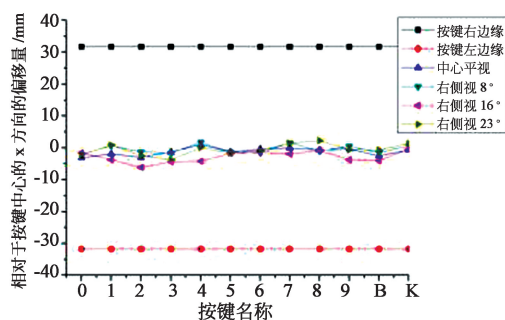


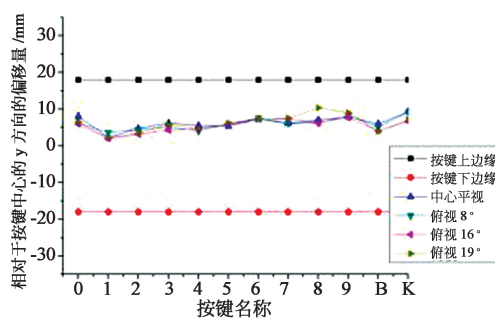
图 7 基于三庭五眼先验法则双目定位结果图

Fig.7 Binocular positioning result graph based on three courts and five eyes



(a) X 方向点击误差

(a) X-direction click error



(b) Y 方向点击误差

(b) Y-direction click error

图 8 视差修正后点击准确性的优化结果

Fig.8 Optimum results of click accuracy after parallax correction

同视角下点击位置与按键中心位置的误差关系。可以发现,不管在 x 或者 y 方向,点击位置与按键中心位置的差距明显减小,并都局限在修正后的边缘范围内,证明在侧视角 $0 \sim 23^\circ$ /俯仰角在 $0 \sim 19^\circ$ 以内,因用户视角带来的视差均能得到良好修正,实现 100% 的点击成功率。

5 结 论

本文提出一种利用 3D 显示作为载体的立体体感技术——虚拟键盘,让用户可在屏幕前方的虚拟交互画面中,通过手指的凌空点击实现简单字符的输入和编辑,其帧率达到 10.6 fps,满足实时工作的速度要求。由于使用过程中,用户视角改变将带来视差偏移,并导致误输入问题,所以我们基于人脸检测和瞳孔估算的方法的修正方法,利用三庭五眼先验法则推算因为 3D 眼镜遮蔽而无法直观检测的用户双眼瞳孔所在的世界坐标,并以此为据,对各个按键边缘的世界坐标范围进行修正。实验结果表明,经过修正后,用户在较大的视角范围(侧视角 $0 \sim 23^\circ$ /俯仰角在 $0 \sim 19^\circ$)内的点击位置与与按键中心位置的误差的偏离度和离散度均明显减小,消除了误输入的可能性,可实现 100% 的点击成功率。

参 考 文 献:

- [1] 王琼华,邓欢.集成成像 3D 拍摄与显示方法[J].液晶与显示,2014,29(2):153-158.
WANG Q H, DENG H. 3D pickup and display method of integral imaging [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(2): 153-158. (in Chinese)

- [2] LUEDER E. 3D 显示技术[M]. 余宏生, 王彤彤, 梅军进, 译. 北京: 电子工业出版社, 2016.
LUEDER E. *3D display* [M]. YU H S, WANG T T, MEI J J, trans. Beijing: Publishing House of Electronics Industry, 2016.
- [3] 孙博. 裸眼 3D 显示技术的现状与发展[J]. 科技传播, 2017, 9(21): 115-116.
SUN B. Present situation and development of naked eye 3D display technology [J]. *Public Communication of Science & Technology*, 2017, 9(21): 115-116.
- [4] 林孜锐, 戴思璋, 梁浩文, 等. 基于混合颜色空间与码本背景减除的虚拟触控优化技术[J]. 液晶与显示, 2015, 30(4): 722-728.
LIN Z K, DAI S Z, LIANG H W, *et al.* Touch technology based on mixed color space and codebook background difference [J]. *Chinese Journal of Liquid Crystals and Displays*, 2015, 30(4): 722-728. (in Chinese)
- [5] 张超, 王琼华, 李大海, 等. 从平面图生成多视点自由立体显示视差图的研究[J]. 液晶与显示, 2009, 24(2): 258-261.
ZHANG C, WANG Q H, LI D H, *et al.* Capture of parallax image from 2D images for multi-view auto-stereoscopic display [J]. *Chinese Journal of Liquid Crystals and Displays*, 2009, 24(2): 258-261. (in Chinese)
- [6] 刘江华, 陈佳品, 程君实. 双目视觉平台的研究[J]. 机器人技术与应用, 2002(1): 36-40.
LIU J H, CHEN J P, CHENG J S. Research on binocular vision platform [J]. *Robot Technique and Application*, 2002(1): 36-40. (in Chinese)
- [7] 林岱昆, 梁伟棠, 梁浩文, 等. 基于人脸跟踪的时空无缝裸眼 3D 显示技术研究[J]. 电视技术, 2015, 39(5): 70-73, 77.
LIN D K, LIANG W T, LIANG H W, *et al.* Spatially and temporally seamless autostereoscopic display technology research based on face-tracking [J]. *Video Engineering*, 2015, 39(5): 70-73, 77. (in Chinese)
- [8] 龙伶敏. 基于 Adaboost 的人脸检测方法 & 眼睛定位算法研究[D]. 成都: 电子科技大学, 2008.
LONG L M. Face detection method and eye location algorithm based on Adaboost [D]. Chengdu: University of Electronic Science and Technology of China, 2008. (in Chinese)
- [9] 胡涛, 冯玉娜, 王亚丹, 等. 一种人脸三庭五眼自动定位及比例确定方法: 中国, 104021550B[P]. 2017-01-18.
HU T, FENG Y N, WANG Y D, *et al.* A method of automatic location and proportion determination of three faces and five eyes in human faces; CN, 104021550B [P]. 2017-01-18. (in Chinese)
- [10] 张华坤, 王敏, 罗开, 等. 三庭五眼的最佳定点探索[J]. 医学与哲学: B, 2017, 38(5): 64-69.
ZHANG H K, WANG M, LUO K, *et al.* A study of the best fixed-point of facial proportions [J]. *Medicine & Philosophy: B*, 2017, 38(5): 64-69. (in Chinese)

作者简介: 薛钊鸿(1993—), 男, 广东江门人, 硕士研究生, 主要从事图像处理技术及应用的研究。E-mail: xue_zhaohong@qq.com

王嘉辉(1979—), 男, 广东广州人, 博士, 实验师, 2009 年于中山大学获得硕士学位, 2014 年于中山大学获得博士学位, 主要从事 3D 显示、图像处理方面研究。E-mail: wangjh@mail.sysu.edu.cn