

Kinect 彩色图像在光标移动控制中的应用

裴以建, 韩甲甲, 闫 哲, 薛 端

(云南大学 信息学院, 昆明 650000)

摘 要: 传统光标控制方式不能较好满足残疾、野外工作等人员的工作需求, 而用 Kinect 识别人体不同的动作类型来实现光标控制时, 由于控制方式过于单一, 存在动作类型过多且相互干扰的情况。为此, 提出将 Kinect 提取的彩色图像和骨骼数据相结合的方法实现光标的移动控制。从彩色图像上提取可变的像素颜色, 以此减少动作类型数量, 同时避免动作之间的相互干扰。给出光标移动控制中涉及到的坐标映射相关理论的计算公式。实验结果表明, 将感应器彩色图像应用于光标移动控制中是可行的, 与单一控制方式相比, 具有更好的移动控制性能, 通过调节公式中的相关参数更加灵活地实现鼠标控制。

关键词: Kinect 感应器; 像素颜色提取; 坐标映射; 光标移动控制; 相互干扰

中文引用格式: 裴以建, 韩甲甲, 闫 哲, 等. Kinect 彩色图像在光标移动控制中的应用[J]. 计算机工程, 2017, 43(10): 241-245.

英文引用格式: PEI Yijian, HAN Jiajia, YAN Zhe, et al. Application of Kinect Color Image in Cursor Movement Control[J]. Computer Engineering, 2017, 43(10): 241-245.

Application of Kinect Color Image in Cursor Movement Control

PEI Yijian, HAN Jiajia, YAN Zhe, XUE Duan

(Information Institute, Yunnan University, Kunming 650000, China)

[Abstract] Traditional cursor control ways cannot better meet the job requirements of disabled persons, field work personnel, etc. When Kinect is used to identify different types of human actions to achieve cursor control, there are too many types of actions and interference between each other due to the single control mode. So this paper proposes the method of combining color image and skeleton data extracted by Kinect to implement the cursor movement control. The extracted pixel colors are variable, which can reduce the number of action types while avoiding the interference between actions. The calculation formulas of coordinate mapping theory related to cursor movement control are given. By adjusting the related parameters in the formula, the mouse can be controlled more flexibly. Experimental results show that it is feasible to use Kinect color image in the cursor movement control. Compared with the single control mode, the proposed method has better mobile control performance.

[Key words] Kinect sensor; pixel color extraction; coordinate mapping; cursor movement control; mutual interference

DOI: 10.3969/j.issn.1000-3428.2017.10.040

0 概述

人机交互在人们生活的各个领域扮演了重要的角色, 如机器人控制^[1]和医学^[2]等特殊环境。在人机交互中, 鼠标依然是最常见的交互工具, 用户可以使用鼠标操作界面, 但是对于残疾人或者交互环境在野外的情况下, 鼠标就很难满足需求^[3]。微软发布的 Kinect 感应器通过读取人体的深度图像数据、骨骼数据和音频数据实现人机交互^[4]。

利用景深数据实现手势识别已经发展了几十年, 拥有强大的理论支持^[5-7]。文献[8]则利用 Kinect 的深度数据读取功能实现了一套基于手势的

鼠标可控制系统, 其中, 用到了 Kinect 自带的骨骼跟踪特性, 旨在为人机交互提供一个舒适的环境。文献[9]提出了一种提高 Kinect 骨骼识别精度的方法: 关节过滤。在利用 Kinect 实现人机交互中, 往往会遇到坐标转换问题, 文献[10-11]对这一问题进行了总结和分析。为了使基于手势的鼠标控制系统有使用更少的动作类型和更灵活的操纵性, 本文在前人的基础上, 提出一种新的基于 Kinect 人机交互的思路。利用 Kinect 读取 RGB 彩色图像数据和其自带骨骼跟踪的特性, 对光标控制进行探究。在手势识别的基础上, 加入彩色数据识别, 减少要识别的动作类型, 从而减少动作之间的相互干扰, 像素颜色可选

作者简介: 裴以建(1959—), 男, 教授, 主研方向为图像处理、人工智能; 韩甲甲、闫 哲、薛 端, 硕士研究生。

收稿日期: 2016-09-26 **修回日期:** 2016-10-28 **E-mail:** hanjiajiasaisai@163.com

范围很广,提供了较大的识别域。同时,通过坐标映射^[12]将 HandLeft 骨骼点三维坐标映射到屏幕二维坐标上,并作为光标在屏幕上的坐标,此映射方法的优点在于光标移动相同的轨迹,通过调节映射公式中的相关参数,骨骼点的移动幅度是可调节的。

Kinect 不仅在鼠标控制上应用广泛,在机器人控制^[13]、三维建模^[14]、三维测量^[15]等方面同样有较多应用。本文提出的人机交互方式在骨骼跟踪的基础上添加了彩色图像识别,与以往单一的骨骼识别技术^[16]有所区别,是 Kinect 开发的新的研究方向。

1 相关技术

Kinect 感应器有 3 个镜头,中间的镜头是 RGB 彩色摄影机,用来采集彩色图像。左右两边的镜头则分别为红外线发射器和红外线 CMOS 摄影机所构成的 3D 结构光深度感应器,用来采集深度数据(场景中物体到摄像头的距离)。彩色摄像头最大支持 1 280 像素×960 像素分辨率成像,红外摄像头最大支持 640 像素×480 像素成像。Kinect 也内建阵列式麦克风,由 4 个麦克风同时收音,比对后消除杂音,并通过声音采集进行语音识别和声源定位。

1.1 Kinect V2 与 Kinect V1 的区别

相比于 Kinect V1, Kinect V2 感应器具备了更高的分辨率和色彩识别度,使识别更加精准,为开发者制作更精良的画面提供了良好支持。Kinect V2 感应器提供了更宽广的视野,尽管在有些情况下,增强视野会降低图像清晰度。Kinect V2 图像保真度大幅提高,在视野扩大的同时,不会对保真度有任何影响。同时, Kinect V2 能够实现跟踪 6 个完整骨骼和 25 个人体关节,跟踪的姿势更为精确和稳定^[17]。

1.2 骨骼空间坐标

被“骨骼跟踪”的用户位置由 X, Y, Z 坐标系表示。不同于深度图像空间坐标,该坐标是三维的,以米为单位。 Z 坐标轴则与 Kinect 感应的朝向一致。光轴与图像平面的交点,即为图像坐标系的原点,用右手坐标系表示传感器序列的原点和 Z 轴延伸方向的阵列点。 Y 轴和 X 轴延伸,如图 1 所示。

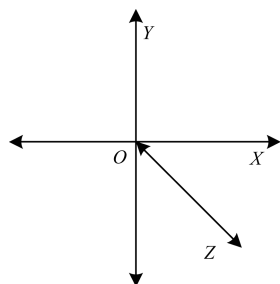


图 1 Kinect 骨骼空间坐标

1.3 骨骼跟踪

Kinect V2 感应器能够跟踪多达 6 个完整的骨骼和跟踪每人 25 个关节。与获取彩色图像的方法

类似,应用程序通过“轮训查询”或者“事件模型”获取最新的骨骼帧数据。骨骼跟踪引擎会在运行时返回一组骨骼帧数据,应用程序获得后可以进行动作识别,每一个在 Kinect 视野内运动的用户都会被分配一个唯一的“追踪 ID”。

2 算法描述

Kinect V2 感应器捕获的人体骨骼坐标和控制台窗口界面坐标不在同一个坐标系中,通过坐标转换,使得 Kinect 捕获的人体骨骼与彩色图像配准,通过转换后的 Neck 骨骼点的二维坐标提取彩色图像上的像素颜色,算法流程如图 2 所示。

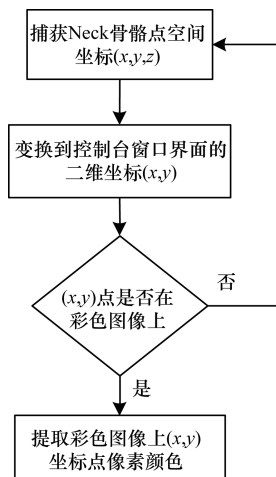


图 2 像素颜色提取算法流程

Kinect V2 感应器捕获的人体骨骼坐标和屏幕坐标不在同一个坐标系中,通过坐标映射,使得 HandLeft 骨骼点的空间坐标映射到屏幕的二维坐标中,并通过调节 l_x, l_y 值控制光标在单位时间内在水平方向和垂直方向移动距离受骨骼点移动距离的影响程度,通过改变 Neck 骨骼点在彩色图像中提取的像素颜色来控制光标移动的启动和暂停,算法流程如图 3 所示。

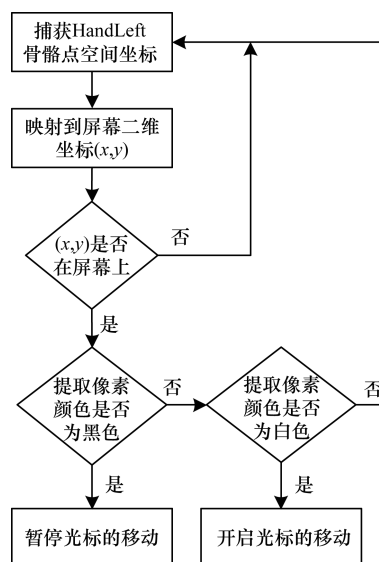


图 3 光标移动控制算法流程

2.1 像素颜色提取

用 OpenCV 获取 Kinect V2 捕获的彩色图像数据,并显示彩色图像,由于控制台窗口界面以左上角为原点,水平方向为 X 轴,垂直方向为 Y 轴,区别于传统二维坐标系,控制台界面 (x,y) 坐标点的像素颜色通过存储彩色图像数据的矩阵的 (y,x) 下标获得。

提取的彩色图像拷贝到 OpenCV 的 Mat 矩阵中,通过访问 Mat 矩阵中存储的数据提取像素颜色,转换后的骨骼点二维坐标为提取像素颜色的位置,像素颜色 4 个通道的计算公式如式(1)~式(4)所示。

$$b = I.at<Vec4b>(y,x).val[0] \quad (1)$$

$$g = I.at<Vec4b>(y,x).val[1] \quad (2)$$

$$r = I.at<Vec4b>(y,x).val[2] \quad (3)$$

$$a = I.at<Vec4b>(y,x).val[3] \quad (4)$$

其中, I 为存储彩色数据的矩阵; b, g, r, a 为像素颜色的 4 个通道; (y,x) 为提取像素颜色点的位置。

2.2 坐标映射

如图 4 所示,人体左手骨骼点是以 Kinect V2 感应器为坐标原点 O_1 的三维坐标,屏幕坐标是以屏幕左上角为坐标原点 O_2 的二维坐标,水平方向为 X 轴,垂直方向为 Y 轴,以 O_1 为原点的 HandLeft 骨骼点坐标 (x,y,z) ($-l_x \leq x \leq l_x, -l_y \leq y \leq l_y$)。其中, l_x, l_y 值控制光标在单位时间内在水平方向和垂直方向移动距离受骨骼点移动距离的影响程度。

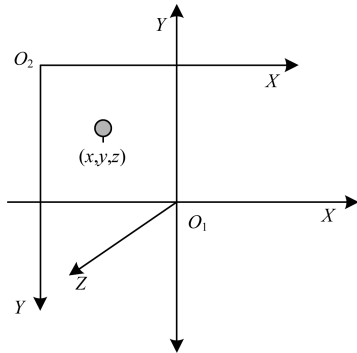


图4 骨骼空间坐标和屏幕坐标映射关系

人体左手骨骼三维坐标点和屏幕二维坐标点映射关系如式(5)~式(8)所示。

$$\begin{cases} x = \frac{(X + l_x)}{2 \times l_x} \times screenWidth \\ y = \frac{(|l_y - y|)}{2 \times l_y} \times screenWidth \\ z = 0 \end{cases} \quad (5)$$

其中, $X \geq 0, Y \geq 0$ 。

$$\begin{cases} x = \frac{(|l_x - |X||)}{2 \times l_x} \times screenWidth \\ y = \frac{(|l_y - Y|)}{2 \times l_y} \times screenWidth \\ z = 0 \end{cases} \quad (6)$$

其中, $X < 0, Y > 0$ 。

$$\begin{cases} x = \frac{(|l_x - |X||)}{2 \times l_x} \times screenWidth \\ y = \frac{(|l_y + |Y||)}{2 \times l_y} \times screenWidth \\ z = 0 \end{cases} \quad (7)$$

其中, $X \leq 0, Y \leq 0$ 。

$$\begin{cases} x = \frac{(l_x + X)}{2 \times l_x} \times screenWidth \\ y = \frac{(|l_y + |Y||)}{2 \times l_y} \times screenWidth \\ z = 0 \end{cases} \quad (8)$$

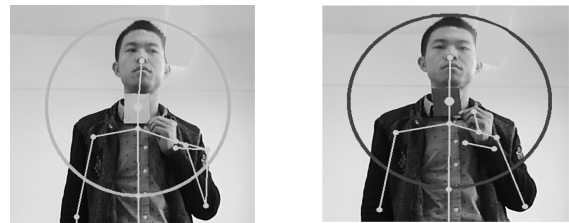
其中, $X > 0, Y < 0$; $screenWidth$ 为屏幕的宽度; $screenHeight$ 为屏幕的高度; (X, Y, Z) 为左手骨骼点以 Kinect V2 感应器为原点的三维坐标; (x, y, z) 为映射后在屏幕上的坐标。

3 实验与结果分析

3.1 像素颜色提取实验

在像素颜色提取实验中,选取 Kinect V2 感应器提供的 25 个骨骼点中的 Neck 骨骼点。假设 Neck 骨骼点转换到彩色图像中的坐标为 (x,y) ,提取彩色图像中 (x,y) 坐标点像素颜色,并以该像素颜色绘制 (x,y) 为原点,半径 $R = 15$ 的圆。

实验结果如图 5 所示,改变转换后在彩色图像上的坐标为 (x,y) 的 Neck 骨骼点的像素颜色为白色和红色。



(a)提取颜色为白色

(b)提取颜色为红色

图5 像素颜色提取实验

3.2 光标移动控制实验

在光标移动实验中,选取 Kinect V2 感应器中能够捕获的 25 个骨骼点中的 HandLeft 骨骼点,光标会随着 HandLeft 骨骼点的运动而运动,该实验中主要对坐标映射公式的正确性进行验证,并探究了不同的 l_x, l_y 值对光标移动的影响。

移动左手掌位置,每隔 $t = 1$ s 间隔时间获取在 $(l_x = 0.5, l_y = 0.5)$ 条件下屏幕上光标的坐标为 (x_1, y_1) 和 $(l_x = 1, l_y = 1)$ 条件下屏幕上光标的坐标为 (x_2, y_2) ,如表 1 所示。图 6 所示为在相同 HandLeft 骨骼点运动轨迹下,分别在 $(l_x = 0.5, l_y = 0.5)$ 和 $(l_x = 1, l_y = 1)$ 条件下光标的运动图像,从图 6 可以看出,相同 HandLeft 运动轨迹条件下,光标的运动轨迹

基本相同且 l_x, l_y 值越大, 光标在单位时间内水平方向和垂直方向上运动的距离越小。

表 1 单位时间内不同 l_x, l_y 值条件下光标在屏幕上的坐标值

t/s	x_1	y_1	x_2	y_2
1	388	397	551	386
2	420	389	590	371
3	437	359	610	340
4	537	397	614	344
5	546	305	675	390
6	668	397	683	410
7	684	437	737	430
8	791	476	797	344
9	912	304	815	340
10	947	297	821	333
11	960	287	863	429
12	1 044	474	870	444

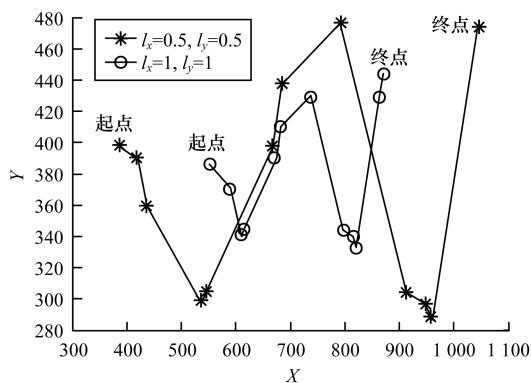


图 6 不同 l_x, l_y 下光标的运动轨迹

实验分析了鼠标移动控制中未结合和结合彩色图像识别时, 分别涉及到的动作类型、动作数量、骨骼点的类型和骨骼点数量, 如表 2 所示。其中, $l_x = 0.5, l_y = 0.5$ 。从表 2 得出结论: 结合彩色图像后, 明显减少了动作类型和骨骼点的数量, 由此也避免了多个动作之间的相互影响。

表 2 不同状态条件下图像识别性能对比

类别	未结合彩色图像识别	结合彩色图像识别
动作类型	左手后伸(暂停运动)	红色(暂停运动)
	左手前伸(开启运动)	白色(开启运动)
	左手下垂(重置)	其他颜色(重置)
	右手掌移动(光标移动)	右手掌移动(光标移动)
动作类型数量	4	1
骨骼点类型	RightHand ShoulderLeft LeftHand	RightHand
骨骼点数量	3	1

在光标移动控制实验中, 选取 Kinect 体感机提供的 25 个骨骼点中的 HandLeft 骨骼。移动左手掌位置, 每隔 $t = 1$ s 时间间隔获取屏幕上光标的坐标为 (x, y) , 如表 3 所示。其中, $l_x = 0.5; l_y = 0.5$ 。在

屏幕二维坐标基础上, 加上表示取样时间的 Z 轴, 绘制三维坐标轴, 如图 7 所示。其中, e, d, c, b, a 为提取像素颜色为黑色的点, 即光标没有随着 HandLeft 骨骼点运动而运动, 其他点提取像素颜色为白色, 光标随着 HandLeft 骨骼点的运动而运动。从实验中可以看出, 通过改变在彩色图像中提取的像素颜色值, 能够较好地实现鼠标的移动控制。

表 3 单位时间内不同像素颜色条件下光标的屏幕坐标

t/s	x	y
1	326	281
2	411	240
3	407	258
4	511	325
5	661	222
6	901	366
7	901	366
8	901	366
9	901	366
10	901	366
11	1 059	278
12	1 077	241
13	1 093	234

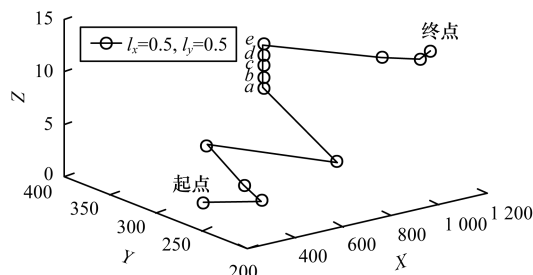


图 7 光标在不同像素颜色下的运动轨迹

4 结束语

在基于感应器开发的应用中主要使用 Kinect 捕获的深度数据和骨骼数据, 而较少使用彩色数据。本文研究通过变换后的骨骼点坐标提取彩色图像像素点的方式, 将感应器捕获的彩色图像应用到光标的移动控制中, 同时也探究了鼠标控制中坐标转换的相关内容, 并给出了计算公式, 通过调节公式中的相关参数更加灵活地实现鼠标控制。

实验结果表明, 本文提出的方法能够精确地提取转换后骨骼点坐标处彩色图像的像素颜色, 结合彩色图像实现光标控制是可行的。通过改变提取到的像素颜色值能够明显减少动作类型, 避免由于动作类型过多而产生的动作之间的相互干扰。针对实验中光标的移动控制, 采用改变彩色图像上转换后骨骼点坐标的像素颜色, 如白色转变为黑色, 或者黑色转变为白色的方式, 能够精确地控制鼠标移动的暂停和启动。

作为一种价廉且使用简便的传感设备,感应器随着测量精度的不断提高,会体现明显的优越性而逐渐在各个领域得到更广泛的应用。下一步将研究感应器捕获的彩色图像数据和骨骼数据、深度数据的结合使用,为 Kinect 相关开发提供更多思路。

参考文献

- [1] 余振中,邓为涛.基于 Kinect 的移动机器人实时局部路径规划[J].计算机工程,2013,39(4):1-2.
- [2] 李小龙.基于 Kinect 的手势识别在虚拟人体解剖教学系统的设计与实现[D].北京:北京工业大学,2014:1-3.
- [3] SZILVIA S, ZICHAR M, ATTILA F. Gesture-based Computer Mouse Using Kinect Sensor[C]//Proceeding of the 5th IEEE International Conference on Cognitive Infocommunications. Washington D. C., USA: IEEE Press, 2014:419-424.
- [4] 马风力.基于 Kinect 自然人机交互系统的设计与实现[D].杭州:浙江大学,2016:10-11.
- [5] REN Zhou, YUAN Junsong, ZHANG Zhengyou. Robust Hand Gesture Recognition Based on Finger-earth Mover's Distance with a Commodity Depth Camera [C]//Proceedings of the 19th ACM International Conference on Multimedia. New York, USA: ACM Press, 2011: 1093-1096.
- [6] KLOMPMAKER F, NEBE K, FAST A. dSensingNI: A Framework for Advanced Tangible Interaction Using a Depth Camera[C]//Proceedings of the 6th International Conference on Tangible, Embedded and Embodied Interaction. New York, USA: ACM Press, 2012: 217-224.
- [7] YEO Hui-shyong, LEE Byung-gook, LIM Hyotaek. Hand Tracking and Gesture Recognition System for Human-computer Interaction Using Low-cost Hardware[J]. Multimedia Tools and Applications, 2015, 74(8): 2687-2715.
- [8] QIN Shuxin, ZHU Xiaoyang, YANG Yiping, et al. Real-time Hand Gesture Recognition from Depth Images Using Convex Shape Decomposition Method[J]. Journal of Signal Processing Systems, 2014, 74(1): 47-58.
- [9] MICHAEL A, YANG Yi, LI Lian. Joint Filtering: Enhancing Gesture and Mouse Movement in Microsoft Kinect Application[C]//Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery. Washington D. C., USA: IEEE Press, 2015: 2528-2532.
- [10] 邵浩,侯文立. Kinect 用于其他应用时坐标系转换问题研究[J].现代计算机,2014(4):42-45.
- [11] 陈建军.基于 Kinect 手势识别的网页软件控制设计[D].太原:太原理工大学,2015.
- [12] 李恒.基于 Kinect 骨骼跟踪功能的骨骼识别系统研究[D].西安:西安电子科技大学,2013:19-21.
- [13] WANG Yali, SONG Guangming, QIAO Guifang, et al. Wheeled Robot Control Based on Gesture Recognition Using the Kinect Sensor [C]//Proceedings of IEEE International Conference on Robotics and Biomimetics. Washington D. C., USA: IEEE Press, 2013: 378-383.
- [14] 宋诗超,禹素萍,许武军.基于 Kinect 的三维人体扫描、重建及测量技术的研究[J].天津工业大学学报, 2014, 31(5): 4-5.
- [15] DUTTA T. Evaluation of the KinectTM Sensor for 3-D Kinematic Measurement in the Workplace[J]. Applied Ergonomics, 2012, 43(4): 645-649.
- [16] 李红波,李双生.基于 Kinect 骨骼数据的人体动作姿势识别方法[J].计算机工程与设计,2016(4): 1-3.
- [17] Microsoft. Kinect for Windows 2.0 主要性能改进[EB/OL]. (2014-09-16). <http://www.cnkinect.com/article-6772-1.html>.
- [12] GAUTIER J, LE MEUR O, GUILLEMOT C. Depth-based Image Completion for View Synthesis[C]//Proceedings of 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video. Washington D. C., USA: IEEE Press, 2011: 1-4.
- [13] ZENZO S D. A Note on the Gradient of a Multi-image[J]. Computer Vision, Graphics, and Image Processing, 1986, 33(1): 116-125.
- [14] LIM H, KIM Y S, LEE S, et al. Bi-layer Inpainting for Novel View Synthesis [C]//Proceedings of the 18th IEEE International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2011: 1089-1092.
- [15] AHN I, KIM C. A Novel Depth-based Virtual View Synthesis Method for Free Viewpoint Video[J]. IEEE Transactions on Broadcasting, 2013, 59(4): 614-626.
- [16] 骆凯,李东晓,冯雅美,等.基于 DIBR 和图像修复的任意视点绘制[J].中国图象图形学报,2010,15(3): 443-449.
- [17] 陈悦,姚剑敏,郭太良.基于背景提取和分区修复的 DIBR 空洞填补方法[J].电视技术,2014,38(23): 34-37.
- [18] CHEUNG C H, SHENG L, NGAN K N. A Disocclusion Filling Method Using Multiple Sprites with Depth for Virtual View Synthesis [C]//Proceedings of IEEE International Conference on Multimedia & Expo Workshops. Washington D. C., USA: IEEE Press, 2015: 1-6.
- [19] MCMILLAN J L. An Image-based Approach to Three-dimensional Computer Graphics[D]. Chapel Hill, USA: University of North Carolina at Chapel Hill, 1997.
- [20] ZINGER S, DO L, de WIT P H N. Free-viewpoint Depth Image Based Rendering [J]. Journal of Visual Communication and Image Representation, 2010, 21(5): 533-541.
- [21] OTSU N. A Threshold Selection Method from Gray-level Histograms [J]. IEEE Transactions on Systems, Man and Cybernetics, 1979, 9(1): 62-66.
- [22] LAWRENCE C, SING Z, KANG B, et al. High-quality Video View Interpolation Using a Layered Representation[J]. ACM Transactions on Graphics, 2008, 23(3): 600-608.

编辑 顾逸斐

编辑 刘冰

(上接第240页)