

## 四路摄像头分组重建的多触点定位方法

王德鑫 张茂军

(国防科技大学信息系统与管理学院系统工程系 长沙 410073)

**摘要:** 针对多重触控技术中的多触点遮挡难题, 该文提出一种基于4路摄像头协同的多触点定位方法。该方法利用矩形交互区4顶点处的4路摄像头同步获取交互区图像, 检测目标触点, 根据立体视觉中的极线约束和3维重建原理, 将4路摄像头分组进行标定和3维重建, 得到4条代表触点方向的直线, 最后利用双阈值拟合4条直线交点定位出触点位置。实验证明该方法能有效解决多触点遮挡问题, 可支持任意多个触点的精确定位, 且成本低廉, 便于安装和移植。

**关键词:** 目标检测; 多重触控; 人机交互; 表面计算; 多触点定位

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2010)02-0495-05

DOI: 10.3724/SP.J.1146.2009.00284

## Multi-touch Locating Method Based on Reconstruction of Four Cameras in Groups

Wang De-xin Zhang Mao-jun

(College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

**Abstract:** To solve the occlusion of multi touch, this paper presents a multi-touch locating method based on the collaboration of four cameras. It uses four cameras fixed at each corner of the frame to capture images synchronously, detects objects in each camera, builds lines directing the object according to epipolar geometry and 3D reconstruction, and finally calibrated with the four cameras to locate each object. The experiment result shows that this method can locate multi touch precisely even under occlusion and it is low cost, easy to install and transplant.

**Key words:** Target detection; Multi-touch; Human-computer interaction; Surface computing; Multi-touch locating

### 1 引言

多重触控技术借助光学和材料学的成果, 使用户能在触控平台上进行多触点操作以实现基于手势的交互, 甚至可以让多个用户同时操作实现基于协同手势的交互。这种以用户为中心的自然高效的人机交互技术将被推向众多领域, 如个人便携式信息产品、信息家电、公共信息查询、电子游戏、教育、展示等等。此外还可用于各种指挥决策中心, 不同角色的指挥员可以同时参与指挥, 提高指挥决策的效率。

多重触控技术的研究最早可追溯到 1982 年多伦多大学的 Nimish Mehta 利用毛玻璃构建的弹性交互接口<sup>[1]</sup>, 之后引起了国内外学术界和工业界的广泛关注, 尤其是微软公司 Surface 计算机<sup>[2]</sup>和苹果公司 iPhone<sup>[3]</sup>的推出, 更激发了人们的研究热情, 主要成果有电容式<sup>[3-5]</sup>, 红外式<sup>[6,7]</sup>, FTIR(Frustrated

Total Internal Reflection)式<sup>[2,8]</sup>和多摄像头协同式<sup>[9,10]</sup>。

电容式利用人体的电流感应进行工作, 容易引起误操作或无响应, 此外环境温度、湿度改变会引起漂移而定位不准。红外式的分辨率取决于红外对管数目、扫描频率及差值算法。多摄像头协同式利用上方摄像头得到的深度信息检测触摸事件并根据两路摄像头之间坐标变换得到触点坐标, 精度不高, 存在延迟, 而且占用较大的空间。这些方法的共同问题是无法解决触点间的遮挡, 从而当同一检测方向上有多个触点时检测结果将出现很多伪点, 无法确定交互区中的触点个数和具体位置。而 FTIR 式虽然具有技术简单、容易构建大尺寸交互区等特点, 但由于其采用的是背投技术, 需要短距广角投影仪、亚克力玻璃层等特殊装置和材料, 需要对现有电脑进行改装, 无法直接安装于现有显示系统, 而且要求交互环境不能太亮。

本文提出一种 4 路摄像头分组重建的多触点定位方法。该方法以红外光源为背景光, 利用安装在矩形交互区 4 顶点处且装有可见光滤波片的 4 路摄像头同步获取图像, 充分利用红外光源的稳定特性,

2009-03-06 收到, 2009-07-27 改回

国家自然科学基金(60705013, 60773023)资助课题

通信作者: 王德鑫 nksky.wdx@gmail.com; nkbeeyond@163.com

简化检测背景进行触点目标检测,再根据立体视觉中的极线约束和3维重建原理,将4路摄像头分组进行3维重建,得到4条代表触点方向的直线,最后利用双阈值拟合4条直线交点定位出触点位置。

## 2 系统原理

4路摄像头分组重建多触点定位方法原理如图1(a)所示。首先在矩形交互区4个边框上安装红外光源,并在4个顶点处分别安装广角红外摄像头。利用红外光源作为背景光源有效过滤可见光,解决复杂多变背景下尤其是当交互场景中出现与手指颜色形状接近的物体时手指的精确检测难题。当交互区没有遮挡物(即触点)时,每个摄像头将拍摄到由红外光构成的白色亮带;如果有遮挡物,红外光源发出的光将被遮挡物阻挡,在亮带对应位置上将出现阴影,从而很容易检测出每路摄像头中的阴影目标,根据标定结果可确定触点方向,综合4路摄像头得到的方向数据进行拟合定位,每个真实的触点都可通过4路摄像头中的4条方向直线相交得到。

该方法中每路摄像头的光心都在所在角的角平分线上,且交互区长宽不等,保证了对角两路摄像头的光心与任意触点的连线不会重合,从而任意一个触点即便在某路摄像头中不能成像,也必定能在该摄像头的对角摄像头中成像,确保了每个触点在两个对角线方向的至少两路摄像头中成像,即能构建两条由摄像头光心到遮挡物的像的直线,再融合另两路摄像头中的直线,如果4条直线交于一点,则该点为真实的触点,否则视为伪点予以剔除。如图1(b)中,触点 $F_1$ 和 $F_4$ , $F_2$ 和 $F_3$ 在 $C_1$ 中均成同一像,但在 $C_3$ 中成像不同;而且 $F_1$ 和 $F_2$ , $F_3$ 和 $F_4$ 在 $C_2$ 中也均成同一像,但在 $C_4$ 中成像不同,所以仍然可以为每个触点构建4条直线,从而确定交互区中有 $F_1F_2F_3F_4$ 4个触点,并且直线的交点即为触点的位置。

构建代表触点方向的直线需要精确标定每个摄像头光心在交互区坐标系的坐标,而且受标定精度的影响很大,而在立体视觉中,并不需要标定每个摄像头的光心位置就可以利用两个摄像头重建3维

物体的位置,其基本原理就是极线几何关系,即空间任意一点的3维坐标可由两幅图像唯一确定<sup>[11,12]</sup>。所以从每个摄像头出发指向触点方向的光线可通过立体视觉中的外极面确定。分别利用摄像头对的外极面构建4条代表触点方向的光线,同样如果4条光线交于一点,则认为该点为真实的触点,否则视为伪点予以剔除。

## 3 系统标定

立体视觉中,得到两个摄像头的投影矩阵 $P_L$ , $P_R$ 后通过线性三角法可重构出物点的3D物理坐标,基本原理是 $\mathbf{X} = \arg \min \sum_i \|u_i - \hat{u}_i(P_i, \mathbf{X})\|^2$ <sup>[11,12]</sup>。

如果选定左摄像头为参考摄像头,则 $P_L = K_L[I | 0]$ , $P_R = K_R[R | T]$ ,其中 $K_L$ 为左摄像头的内参数矩阵, $K_R$ 为右摄像头的内参数矩阵, $R$ , $T$ 分别为右摄像头坐标系相对左摄像头坐标系的旋转和平移变换矩阵。3维重构得到的3D物理坐标是以左摄像头坐标系为参考坐标系。从而选定一个摄像头作为参考摄像头,将另外3个摄像头分别与参考摄像头进行标定,就可将每组重构结果都转为参考摄像头坐标系,再通过标定参考摄像头坐标系与交互区坐标系,实现最终的触点定位。

本文提出的多触点定位方法的标定就是要标定每个摄像头的投影矩阵 $P$ ,以及每个摄像头相对参考摄像头的旋转变换矩阵 $R_{LR}$ 和平移变换矩阵 $T_{LR}$ 。标定过程为分别用每对摄像头同时拍摄若干幅二维棋盘格标定模板,对拍摄图像提取角点,匹配特征点,估计基础矩阵和本质矩阵,最后进行SVD分解得到每个摄像头的投影矩阵和相对参考摄像头的旋转和平移变换矩阵。流程如图2所示。

在实验中,以 $C_1$ 为参考摄像头,将 $C_2$ , $C_3$ , $C_4$ 分别与 $C_1$ 配对,但是由于交互区尺寸限制,位于交互区对角的两摄像头 $C_1C_3$ 组成的摄像头对难以同时拍全标定模板,从而取不到足够多的特征点,所以将 $C_3$ 与 $C_2$ 或 $C_4$ 配对,最后变换到 $C_1$ 坐标系。为此本文定义了如下5种坐标系如图3所示,坐标系 $C_1X_1Y_1Z_1$ , $C_2X_2Y_2Z_2$ , $C_3X_3Y_3Z_3$ , $C_4X_4Y_4Z_4$ 分别为摄像头 $C_1$ , $C_2$ , $C_3$ 和 $C_4$ 的摄像头坐标系,另外以交互区的左上角为原点定义了交互区坐标系 $IXYZ$ 。

## 4 多触点定位

本文提出多触点定位方法流程如图4所示。

首先每路摄像头分别获取交互区图像,并利用减背景法进行阴影目标检测,利用形态学滤波和区域融合算法对结果进行优化,得到每路摄像头阴影

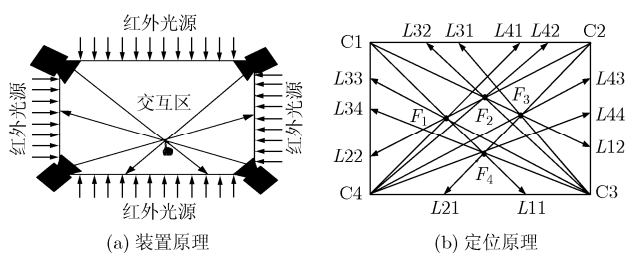


图1 原理图

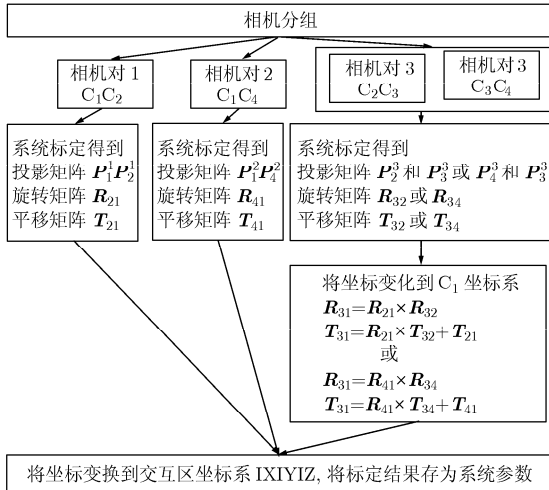


图 2 系统标定流程

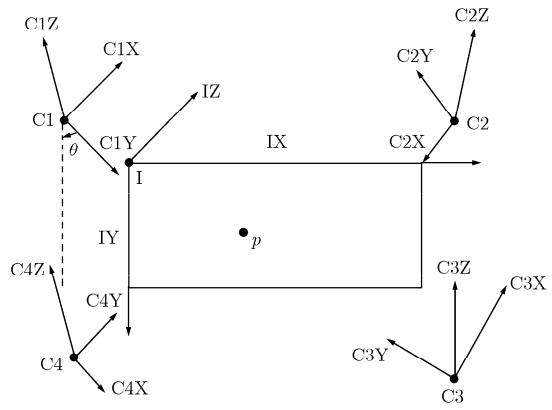


图 3 坐标系

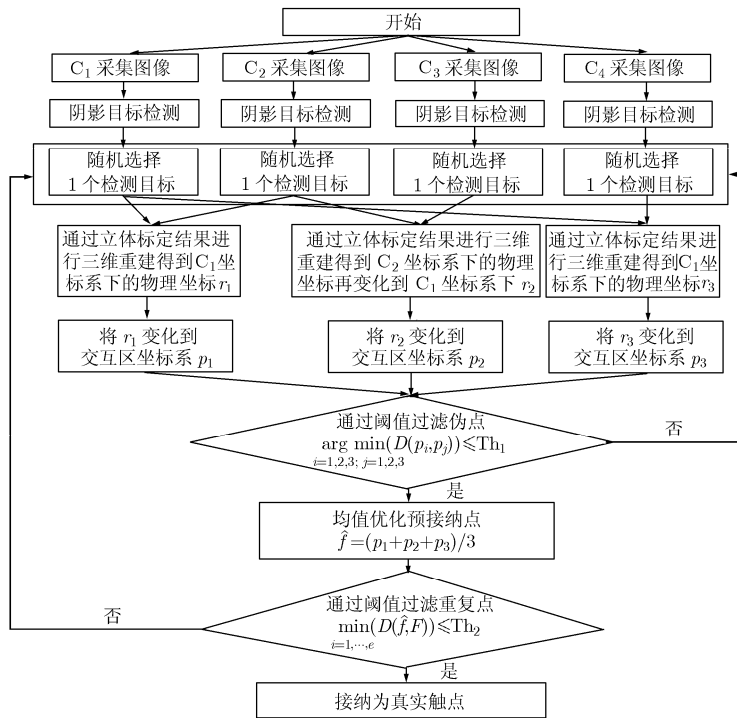


图 4 定位流程

目标检测结果  $PG = \{PG_j\}$ ,  $PG_j = \{pg_{j1}, \dots, pg_{jt_j}, pg_{jm_j}\}$ ,  $pg_{jt_j} = (x_{jt_j}, y_{jt_j})$ ,  $pg_{jt_j}$  为摄像头  $C_j$  中第  $t_j$  个触点图像坐标,  $t_j = 1, \dots, n_j$ , 检测过程如图 5 所示。

从每路摄像头检测结果中取 1 个目标, 根据摄像头配对规则分别进行 3 维重建。如第一组摄像头对  $C_1C_2$  的目标检测坐标为  $pg_{1t_1}$  与  $pg_{2t_2}$ , 标定结果为  $K_1$ ,  $K_2$ ,  $R_{21}$ ,  $T_{21}$ , 从而  $P_1 = K_1[I | 0]$ ,  $P_2 = K_2[R_{21} | T_{21}]$ , 再根据线性三角法重构出物点坐标  $r_1$ , 同理第 2 组摄像头对  $C_1C_4$  的目标检测坐标  $pg_{1t_1}$  与  $pg_{4t_4}$  的重建结果为  $r_2$ , 第 3 组摄像头对  $C_2C_3$  的

目标检测坐标  $pg_{2t_2}$  与  $pg_{3t_3}$  的重建结果为  $r_3$ , 将它们转化到交互区坐标系得到  $p_1$ ,  $p_2$  和  $p_3$ , 由于噪声的影响, 一般  $p_1$ ,  $p_2$  和  $p_3$  不会重合, 可通过阈值方法对  $p_1$ ,  $p_2$  和  $p_3$  进行融合。如果 3 点之间的距离均小于阈值, 则预接纳为真实的触点, 否则视为伪点予以剔除, 条件为  $\arg \min_{i=1,2,3; j=1,2,3} (D(p_i, p_j)) \leq Th_1$ , 其中

$D(p, p')$  表示两点之间的距离,  $Th_1$  取触点在摄像头中成像的平均宽度, 如在中星微芯片 (ZC0301PL, 30 万像素) 的普通 USB 摄像头下取 5 即可。符合 3 点重合条件后, 得到预接纳点  $\hat{f} = (p_1 + p_2 + p_3) / 3$ ,

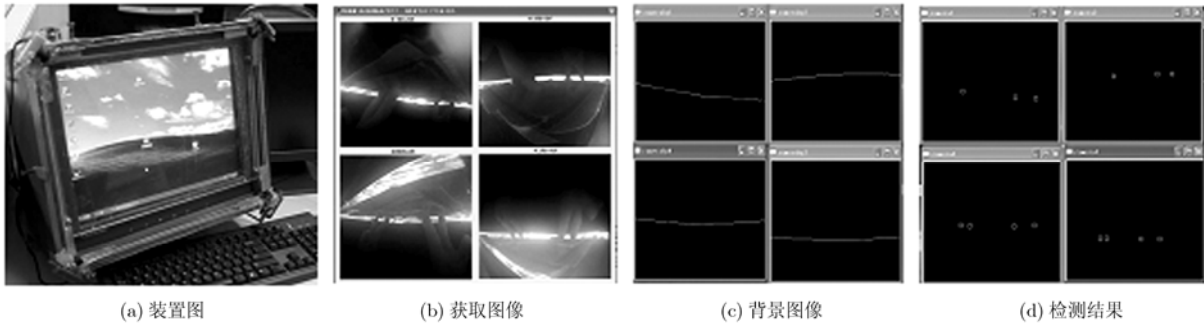


图 5 检测过程

然后还需对  $C_{n_1}^1 \cdot C_{n_2}^1 \cdot C_{n_3}^1 \cdot C_{n_4}^1$  组重建结果去除重复点, 条件是  $\min_{i=1 \dots e} (D(\hat{f}, F)) \leq Th_2$ , 其中  $F = \{f_e\}$  为已定位的触点集合,  $e$  为已定位的触点个数,  $Th_2$  为判断点是否重合的阈值, 一般取 1 即可, 如果符合则接受为真实的触点, 即  $f_{e+1} = \hat{f}$ 。

## 5 精度分析

影响本文方法精度的主要因素有摄像头分辨率, 标定算法和交互区尺寸。假设摄像头采集图像序列大小为  $iw \cdot ih$ ,  $iw$  为图像宽度,  $ih$  为图像高度; 交互场景宽为  $pw$ , 高为  $ph$ , 由于每个摄像头采集到的图像为一条直线, 对应于对角的边长度为  $pl = pw + ph$ , 从而本文方法的像素分辨率为  $PD = iw / pl$ 。

假设每个摄像头标定结果的重投影误差分别为  $[T_X^1 \ T_Y^1]$ ,  $[T_X^2 \ T_Y^2]$ ,  $[T_X^3 \ T_Y^3]$ ,  $[T_X^4 \ T_Y^4]$ ,  $T_X^i$  表示水平投影误差,  $T_Y^i$  为竖直投影误差。4 路摄像头分组重构的结果将在一个以真实点为中心, 宽  $2 \cdot TH_X$ , 高  $2 \cdot TH_Y$  的矩形内, 其中  $TH_X = \max(T_X^1, T_X^2, T_X^3, T_X^4)$ ,  $TH_Y = \max(T_Y^1, T_Y^2, T_Y^3, T_Y^4)$ 。

从而本文提出方法的物理分辨率为, 水平误差在  $PD \cdot TH_X$  内, 竖直误差在  $PD \cdot TH_Y$  内。

## 6 实验结果

本文设计并实现了一个原型系统, 交互区的尺寸与 43 cm 显示器的显示区域相同 ( $350 \text{ mm} \times 400 \text{ mm}$ ), 图像采集设备为中星微芯片 (ZC0301PL, 30 万像素) 的普通 USB 摄像头, 以  $320 \times 240$  分辨率采集交互区图像, 标定过程中采用标准网格模板, 用文献 [13] 标定内参, 并用文献 [12] 进行基础矩阵的估计, 标定精度水平最大误差为 0.5 像素, 竖直最大误差为 0.8 像素, 标定结果见表 1。

本文实验所使用计算机的主频为 2.0 GHz, 系统运行时, 利用 directshow 技术近似同步地采集分辨率为  $320 \times 240$  像素的图像序列, 系统的视觉处理部分速度可以达到 15 帧/秒。实验中统计贴在屏幕

表 1 标定结果

	内参 $K$	$R$	$T$
$C_1$	$\begin{bmatrix} 231.83 & 0 & 155.86; \\ 0 & 232.31 & 96.21; \\ 0 & 0 & 1 \end{bmatrix}$	$I$	$0$
$C_2$	$\begin{bmatrix} 222.41 & 0 & 162.26; \\ 0 & 221.49 & 118.42; \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.0644 & 0.0633 & -0.9959; \\ -0.1268 & 0.9894 & 0.0711; \\ 0.9898 & 0.1309 & -0.0557 \end{bmatrix}$	$\begin{bmatrix} 251.49688 \\ -16.82527 \\ 272.53985 \end{bmatrix}$
$C_3$	$\begin{bmatrix} 208.24 & 0 & 153.97; \\ 0 & 207.22 & 120.87; \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.9908 & -0.0142 & 0.1343; \\ -0.0199 & 0.9988 & -0.0432; \\ -0.1336 & -0.0455 & -0.99 \end{bmatrix}$	$\begin{bmatrix} 38.14 \\ -33.831 \\ 487.99 \end{bmatrix}$
$C_4$	$\begin{bmatrix} 229.59 & 0 & 181.42; \\ 0 & 230.85 & 99.44; \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.0476 & -0.0285 & 0.9985; \\ 0.0611 & 0.9976 & 0.0313; \\ -0.9970 & 0.0625 & -0.0457 \end{bmatrix}$	$\begin{bmatrix} -221.3245 \\ -13.85174 \\ 213.08921 \end{bmatrix}$
$C_{XYZ}$		$\begin{bmatrix} -0.1204 & -0.1765 & -0.9769; \\ -0.0783 & 0.9827 & -0.1679; \\ 0.9896 & 0.0563 & -0.1322 \end{bmatrix}$	$\begin{bmatrix} 225.68294 \\ -6.10775 \\ 201.98476 \end{bmatrix}$

上的网格纸每个节点出的定位误差, 得出水平定位精度约为 0.5 mm (接近于  $0.5 \times 320 / (350 + 400) = 0.213 \text{ mm}$  的理论值), 竖直定位精度为 0.7 mm (接近于  $0.8 \times 320 / (350 + 400) = 0.34 \text{ mm}$  的理论值), 效果如图 6 所示。

现有方法<sup>[2-5,8,9]</sup>都只有两个信号检测方向, 即便在 2 指时都将检测出很多伪点。如图 6(a) 中, 触点  $F_1$  和  $F_3$ ,  $F_2$  和  $F_3$  在  $Y$  方向均成一像; 而且  $F_1$  和  $F_2$ ,  $F_3$  和  $F_4$  在  $X$  方向也均成一像, 所以场景  $F_1F_3, F_2F_4, F_1F_3F_2, F_1F_3F_4, F_2F_4F_1, F_2F_4F_3, F_1F_2F_3F_4$  的检测结果都一致, 无法确定交互区中的触点个数和具体位置。利用本文方法则能有效解决该问题, 两指的定位效果如图 6(b) 所示; 此外从图 6(c) 中也可看出, 虽然第 1, 2, 3 摄像头中分别都存在触点遮挡问题 (即两个触点在同一光线上), 但利用本文方法仍能精确地定位出每个触点的位置。

## 7 结论

本文提出的 4 路摄像头分组重建的多触点定位

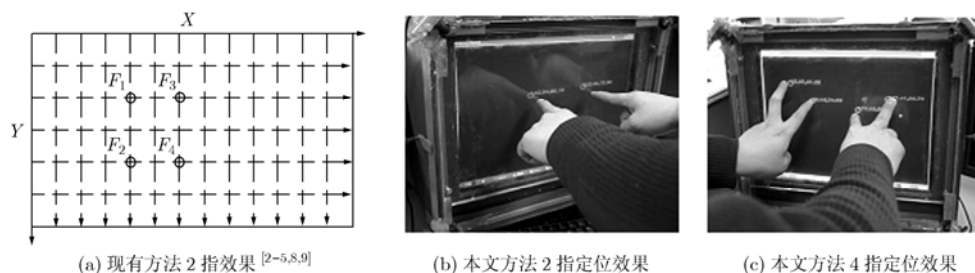


图 6 多触点定位

方法舍弃了 3 维重建结果中误差最大的 Z 轴坐标, 并且通过 3 组摄像头分组标定确定摄像头之间的外参数约束, 再根据双阈值对 3 组 3 维重建结果进行拟合, 有效保证了重建的精度。实验证明该方法能有效解决多触点遮挡问题, 可支持任意多个触点的精确定位, 而且能直接安装于现有显示平台前, 能构建任意尺寸的交互平台, 此外对交互环境的亮度、湿度、温度等都没有特殊要求, 更利于推广应用。

利用本文提出的多触点定位方法得到触点位置信息后, 根据触点位置坐标的变化情况确定各个触点的移动轨迹, 据此定义不同的交互手势, 比如两个触点运动方向相反表示进行放大或缩小操作; 一个触点不动, 另一个触点作弧线运动, 表示进行旋转操作等等, 完成单点触控系统所无法完成的功能, 实现更强大、更人性化的人机交互。

### 参 考 文 献

- [1] Bill B. Multi-touch systems that I have known and loved [OL]. (2008-11-11)[2008-12-20]. <http://www.billbuxton.com/multitouchOverview.html>.
- [2] Bill B. Surface and tangible computing, and the “Small” matter of people and design[C]. Proceedings of Solid-State Circuits Conference, San Francisco, 2008: 24-29.
- [3] Apple Inc. iPhone [OL]. (2008-11-11)[2008-12-20]. <http://www.apple.com/iphone>.
- [4] Rekimoto J. SmartSkin: An infrastructure for freehand manipulation on interactive surfaces[C]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves, Minneapolis, 2002: 113-120.
- [5] Dietz P and Leigh D. DiamondTouch: A multi-user touch technology[C]. Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, Orlando, 2001: 219-226.
- [6] Izadi S, Hodges S, Butler A, Rrustemi A, and Buxton B. ThinSight: integrated optical multi-touch sensing through thin form-factor displays. (San Diego, California, August 04-04, 2007). EDT '07, Vol. 252. ACM, New York, NY, 6, 2007.
- [7] 卢如西, 周春景, 李军明. 一种红外线触摸屏及其多点触摸定位方法: 中国专利, 200710031082.6[P], 2008-03-26. Lu Ru-xi, Zhou Chun-Jing, and Li Jun-ming. An infrared touch screen and its multi-touch detecting method, China, patent application NO. 200710031082.6[P], March.26, 2008.
- [8] Jefferson H. Low-cost multi-touch sensing through frustrated total internal reflection[C]. Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, Seattle, 2005: 315-319.
- [9] Itai K, Kevin G, and Hamid A. A multi-touch surface using multiple cameras[C]. Proceedings of Advanced Concepts for Intelligent Vision Systems, Delft, 2007: 97-108.
- [10] Ankur A, Shahram I, and Manmohan C, et al. High precision multi-touch sensing on surfaces using overhead cameras[C]. Proceedings of The second IEEE International Workshop on Horizontal Interactive Human-Computer Systems, Rhode Island, 2007: 197-200.
- [11] 马颂德, 张正友. 计算机视觉: 计算理论与算法基础[M]. 北京: 科学出版社, 1998: 60-63. Ma Song-de and Zhang Zheng-you. Computer Vision [M]. Beijing: Science Press, 1998: 60-63.
- [12] Torr P H S. A structure and motion toolkit in matlab. Technical Report MSR-TR-2002-56, Microsoft Research, 7 JJ Thomson Avenue, Cambridge, CB3 0FB, UK, <http://research.microsoft.com/~philtorr/>, 2002.
- [13] Camera Calibration Toolbox for Matlab, [http://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/index.html)

王德鑫: 男, 1983 年生, 博士生, 研究方向为人机交互、虚拟现实。

张茂军: 男, 1971 年生, 教授, 博士生导师, 研究方向为虚拟现实与系统仿真技术。