

文章编号:1001-9081(2014)05-1442-07

doi:10.11772/j.issn.1001-9081.2014.05.1442

基于深度图像的运动人手检测与指尖点跟踪算法

刘伟华^{1*}, 樊养余¹, 雷 涛²

(1. 西北工业大学 电子信息学院, 西安 710072; 2. 兰州交通大学 电子与信息工程学院, 兰州 730070)

(*通信作者电子邮箱 lwh86117@163.com)

摘要:针对由传统摄像头获得的二维图像难以准确检测复杂环境下运动人手目标的问题,提出了一种利用 Kinect 摄像头对复杂背景下近景人手指尖空间点的检测和跟踪算法。首先利用基于视觉的多信息融合方法对人手区域进行粗检测,并利用深度信息对得到的模糊人手区域进行精确定位及分割,从而有效抑制了复杂背景对人手区域的干扰;根据分割得到的人手区域,利用人手边界的曲率关系实现了指尖空间点的快速检测;最后利用卡尔曼滤波跟踪检测到的运动指尖点。实验结果表明,与已有指尖点检测方法相比,该算法能够准确地跟踪多种姿态下手指指尖的空间位置,具有较低的误差率。

关键词:人手分割;指尖检测;指尖跟踪;边界曲率;卡尔曼滤波

中图分类号: TP391.41 **文献标志码:**A

Human fingertip detection and tracking algorithm based on depth image

LIU Weihua^{1*}, FAN Yangyu¹, LEI Tao²

(1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China;
2. School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou Gansu 730070, China)

Abstract: To solve the problem of detecting human hand in complex background based on traditional camera, a fast, automatic method was proposed which can accurately detect and track foreground human fingertips by using Kinect camera. This method firstly used a combined vision-based information to roughly extract the hand region, then, by taking advantage of depth information, a bare hand could be successfully segmented without connecting to background. Subsequently, the fingertips of that bare hand could be extracted by using minimum circle and curvature relationship on the hand boundary. Finally, to improve the detecting accuracy, the fingertips were optimized by using Kalman filter. The experimental results show that compared with existing method the algorithm can successfully track the 3D locations of fingertips under multiple hand poses and with much lower error rate.

Key words: hand segmentation; fingertip detection; fingertip tracking; boundary curvature; Kalman filter

0 引言

人手检测技术在计算机视觉领域越来越受到人们的关注,其主要应用体现在手势识别、人机交互等方面。由于人手语义表达的多样性,其在视频流或图像中的空间位置及形状的差异使得检测过程中会出现漏检或错检的情况。因此,对视频图像中人手精确定位成为识别过程中关键的一环。在复杂背景下准确地检测出人手、手指及指尖的位置与方向,成为进一步提高手势识别准确率的前提保障。

截至目前为止,已有大量关于人手识别的技术被报道。在已有的文献中,对人手识别的方法一般分为两大类:一类是基于视觉表象的方法,如文献[1–2],即处理通过普通摄像头得到的图像来判别人手及手指的位置,达到对图像中人手识别的目的;另一类是基于模型的方法,如文献[3–4],即通过建立目标的三维几何模型,根据提取的目标特征与模型进行匹配和识别。

第一类方法中人手检测问题,一般从二维图像空间和三维空间对人手特征进行描述。二维空间描述的优点在于可利用许多高效的图形学算子对其进行检测,这些算子具有运算速度快、算法复杂度低等特点,然而其未能拓展到三维空间,没有考虑目标的三维空间特征,并且受光照、背景变化的影响较大。目前对于二维图像中的人手信息,许多文献利用了皮肤颜色信息^[5]和一些图形学算子^[6]对人手进行检测、分割^[7]。其中, Thangali 等^[8]利用方向梯度直方图(Histogram of Oriented Gradients, HOG)算子来描述非刚性的不同人手形状特征,并利用支持向量机对特征进行分类和识别。然而,该方法对人手形状的检测仅存在于二维图像空间,没有考虑相同人手形状在不同空间角度下对检测的影响。Alon 等^[9]通过利用相邻帧之间的人手运动误差和肤色信息对其进行检测,然而对于运动的背景,如运动的人体,其检测率则大大降低。针对复杂的背景干扰, Karlinsky 等^[10]利用链式模型对局部可形变目标进行检测,不仅利用了目标自身视觉特征,并且结合其

收稿日期:2013-11-08;修回日期:2013-12-18。

基金项目:国家自然科学基金资助项目(61202314);中国博士后科学基金资助项目(2012M521801)。

作者简介:刘伟华(1986-),男,陕西西安人,博士研究生,主要研究方向:计算机视觉、机器学习、虚拟现实;樊养余(1960-),男,陕西蓝田人,教授,博士,主要研究方向:图像处理、模式识别、虚拟现实;雷涛(1981-),男,陕西大荔人,副教授,博士,主要研究方向:图像处理、模式识别、计算机视觉。

周围相关部件的特征来描述检测目标,但也限于二维空间描述。对于二维空间的局限性,Elmezain 等^[11]使用双目立体摄像头检测人手,利用配准的深度信息和在 YC_bC_r颜色空间中的高斯混合模型得到的肤色概率信息检测人手,然而其深度受摄像头参数误差的影响。van den Bergh 等^[12]通过增加 ToF (Time of Flight) 摄像头获取图像深度信息并结合肤色信息对人手区域进行检测,该方法虽然对目标增加了深度信息,然而当深度阈值无法分割出单独人手时,单一的肤色概率检测会受到复杂背景中类肤色目标的干扰。Ren 等^[13]利用 Kinect 的深度和颜色图像对人手进行检测和分割,然而对人手姿态的检测仅限于单幅静止人手图像,而不是动态的人手运动序列。

指尖检测方面,已有文献大都利用了人手图像结构来完成对其各个部分的检测^[14~16]。Kerdvibulvech 等^[14]首先对检测出的人手图像利用 Gabor 滤波降维,并且通过神经网络算法训练特征向量,然后采用局部线性映射估计指尖的位置。Nguyen 等^[16]提出了两种指尖检测的方法:第一种方法利用指尖曲线的高度信息进行阈值分割,提取手指,这种方法检测速度快,缺点是必须对高度阈值的进行准确选择,否则会给分割带来误差;第二种方法利用人手邻域凹凸形状分解手指,但其未研究如何进一步检测指尖。Lee 等^[17]研究尺度不变性下手形轮廓曲率,达到检测指尖的目的,该方法检测速度快,然而未能对手指的特征具体研究,因此对指尖点不够精确。针对这一问题,本文对其进行了改进,对曲率检测后的指尖轮廓边缘进行拟合后再检测,从而提高了检测精度。以上算法都是基于视觉表象的人手检测^[18~19]。

对于第二类基于模型的人手姿态检测算法,通常会对人手进行三维运动学建模^[20],建立其人手外表的骨骼模型^[21]。如利用数据手套获取人手各个关键点的约束参数,应用无迹卡尔曼滤波(Unscented Kalman Filter, UKF)^[22]或者蒙特卡罗(Monte Carlo, MK)算法^[23]来完成对人手运动姿态的跟踪。虽然三维人手模型的方法可以描述范围更广的手势及姿态,然而,由于其建模对象为铰链式的多自由度形变目标,对各种人手的变形要求足够大的数据支持,并且还要考虑模糊角度下的特征提取问题,而基于视觉表象的人手检测跟踪可以仅通过在图像层面上的处理达到实时检测识别的目的。

然而,基于视觉的检测会受到外部环境的干扰,例如利用肤色概率的人手检测会受到光线改变的干扰,利用图像差分帧的检测会受到运动背景的干扰等。针对上述问题,本文首先提出了结合人手不同信息特征的检测方法,其主要结合人手肤色像素特征、人手运动特征^[24]、图像差异特征^[25]以及通过 Kinect 获取的深度特征对检测的目标加以约束,以此来排除外部环境对人手检测的影响,从而增强检测的鲁棒性。对于指尖点的检测,为了减小误差,增大检测准确率,本文改进了文献[17]中单一曲率检测指尖的算法:首先利用人手区域的最大外接椭圆^[26]分割出任意状态下的单独手指区域,利用曲率检测出指尖点集,通过拟合点集邻域手指边界点,从而找出拟合抛物线的极小值点,并通过摄像机内部参数计算其空间坐标,最终得到指尖的空间位置。同时,为了减小计算复杂度,利用 Daniel 在文献[27]中提出的摄像头校准方法,获得与颜色图像相匹配的深度图像信息。

1 多信息特征下的人手检测

目前应用最广的人手检测方法是基于肤色概率的方法,当背景中无其他类似肤色的目标时,通过单一的肤色概率检测一定程度上能够有效检测出视频中运动人手,然而当加入背景运动人体后,将无法分割出单独的人手区域。因此,本文综合考虑四种人手检测方法,提出一种基于多种图像信息融合的方法,解决了复杂背景对人手检测的问题。其中基于视觉的视频信息分别为肤色概率信息、差分帧瞬时运动信息,以及图像运动残差信息。融合这三种信息,得到目标人手最大连通区域后,通过利用深度信息对目标人手最大区域进行阈值分割,最终检测出单独人手。

1.1 肤色概率信息下的人手检测

考虑到人体肤色的一致性,通过建立统计颜色模型来计算每个像素点为皮肤颜色的概率,由此来判定所检测像素是否为人手像素;文献[14]中定义了皮肤颜色的似然性分布为: $P(r,g,b \mid skin)$,同理,非皮肤的像素概率分布为: $P(r,g,b \mid nonskin)$ 。

基于以上两种分布,图像中某点像素被定义为皮肤的概率服从贝叶斯法则:

$$P(skin \mid r,g,b) = \frac{P(r,g,b \mid skin)P(skin)}{P(r,g,b)} \quad (1)$$

其中:

$$\begin{aligned} P(r,g,b) &= P(r,g,b \mid skin) \times P(skin) + \\ &\quad P(r,g,b \mid nonskin) \times P(nonskin) \end{aligned}$$

这里,通过建立肤色直方图和非肤色直方图来分别估计 $P(r,g,b \mid skin)$ 和 $P(r,g,b \mid nonskin)$ 。

最后,通过加权候选区域中像素点的后验肤色概率并对其归一化处理,得到具有最大加权值的候选区域即为检测人手区域。对于得到的上述人手检测图像,对其二值化图像,记为 L ,对应连通区域,记为 L^k ,其检测结果如图 1(d) ~ (f) 所示。然而,由图可知,由于前景桌面被误认为肤色信息,而其深度小于目标人手,因此并不能简单地通过最小深度阈值分割出目标人手,而需要引入人手运动特征对其进行约束。

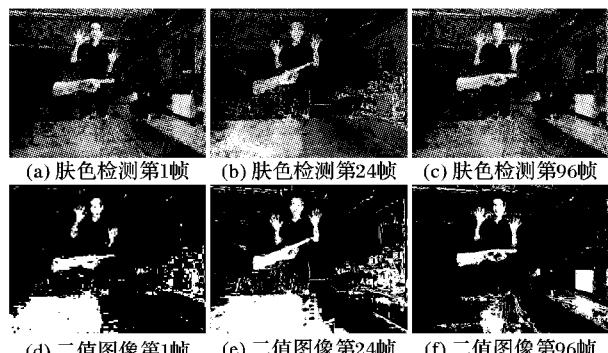


图 1 肤色概率检测与其对应二值图像

1.2 瞬时运动信息下的人手检测

由于人脸与手肤色的一致性,单独利用肤色信息进行人手检测还会受到人脸肤色的干扰。因此,需要考虑手势的运动信息,来进一步消除其他近似肤色区域的干扰。对于手势运动信息的检测,本文采用图像差分帧的方法来计算人手候选区域。这里,规定 $I(x,y,i)$ 为像素点 (x,y) 在第 i 帧的像素强度,通过比较当前帧与前 z 帧和后 z 帧之差的绝对值来计算运

动像素信息 $M(x, y, i)$:

$$\begin{cases} I_1(x, y, i) = |I(x, y, i) - I(x, y, i - z)| \\ I_2(x, y, i) = |I(x, y, i) - I(x, y, i + z)| \end{cases} \quad (2)$$

$$M(x, y, i) = \min(I_1(x, y, i), I_2(x, y, i)) \quad (3)$$

通过上述计算,人手候选区域为 M_i 图像中加权最大的后选区域。其检测结果如图 2 所示。

由图 2 可以看出,基于差分帧的算法同时检测出运动的背景人体和目标前景人手,因此,单一的图像差分检测并不能完全提取出目标人手,但其仍可作为信息融合的一个特征对目标进行约束。

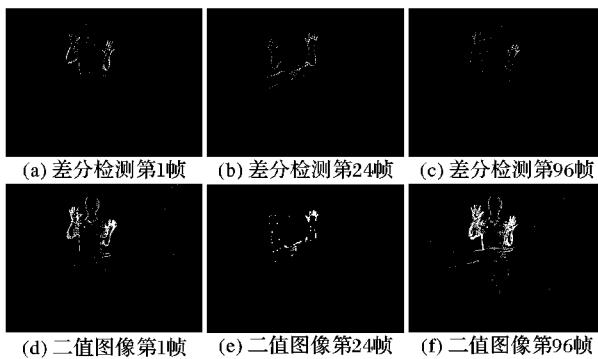


图 2 差分帧检测与其对应二值图像

1.3 运动残差信息下的人手检测

在手势视频中,人手运动的频率要远高于在背景中运动的其他对象的频率。利用这种属性,首先考虑对每两个连续的前后帧图像,进行分块处理,通常分为 8×8 或 12×12 的块,然后在后一帧中寻找与前一帧每块最为匹配的块。这里,运动残差信息定义为匹配块之间的像素强度差的加权平方和:

$$R_j(x, y) = \sum_{(x, y) \in \text{pixels}} [A_j(x, y) - B_j(x, y)]^2 \quad (4)$$

其中: A_j , B_j 表示相邻两帧中第 j 块的像素信息。这样,第 i 帧 j 第块中,每个像素点 (x, y) 的残差值定义为 $R_j(x, y)$ 。由此,分块中的每一个像素点都会被赋予残差信息。由于手势运动的显著性,因此残差值较大的块即为包含人手的区域。其检测结果如图 3 所示。



图 3 运动块残差检测

由图 3 可以看出,基于图像残差块的检测能够大致地检测出目标人手区域,然而,当残差块阈值设定过低时,有时会检测不到目标人手,设定过高则会检测出不必要的背景运动块,因此,为了保证每帧都能够检测出目标人手,实验设定阈值块数量 $N = 3$ 。

1.4 多信息融合下的人手检测

对于图像中的每个像素,通过结合上述三种特征,得到特征融合状态下的像素特征表示为:

$$F(x, y, i) = e^{-\ln(S(x, y, i) + \varepsilon) + \ln(M(x, y, i) + \varepsilon) + w_1 \times \ln(R_j(x, y, i) + \varepsilon)} \quad (5)$$

这里, $S(x, y, i)$ 为像素点 (x, y) 在第 i 帧的皮肤特征值,

同理, $M(x, y, i)$ 为其像素运动特征, $R(x, y, i)$ 为像素残差特征。 ε 为一正值常量,通过优化训练数据,权重 $w_1 = 1/2$, $\varepsilon = 0.002$ 。如图 4(a) 所示,多信息检测下的人手区域排除了前景桌面的干扰,然而,背景人体部分依然存在,因此,考虑到目标人手区域为肤色二值图像 L 中相对较大的连通区域,提出一种利用肤色连通区域分割的方法,算法流程如下:

- 1) 在肤色二值化连通区域 L^k 中,如图 4(b) 所示,遍历图像 F 的每个非零像素点 $F(x, y, i)$;
- 2) 如果 $F(x, y, i) \in L^k(x, y, i)$, 在第 k 连通域内计算 $T(k) = T(k) + 1$, 其中每一个 $T(k)$ 初值为零;
- 3) 计算所有 $T(k)$ 中最大值所对应的连通区域标记 k , $\{\max T(k)\} \rightarrow k$;
- 4) 此标记 k 即为目标人手对应的连通区域,如图 4(c) 所示;
- 5) 再次利用连通区域,填充人手区域,修复目标人手,记为 C , 如图 4(d) 所示。

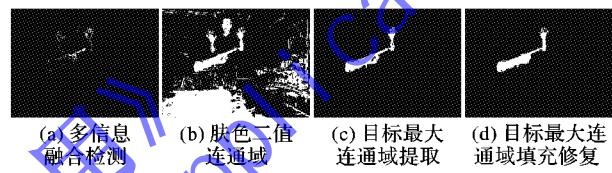


图 4 多信息融合检测(第 24 帧)

由图 4(d) 所示,提取的包含目标人手的区域能够剔除掉部分背景干扰,然而并不完全。为了能够完全分割出单独人手形状,这里引入图像的深度信息,通过利用最小深度阈值,分割出单独的前景人手。

1.5 人手深度图像预处理

为了使得深度阈值自适应于每一帧图像的分割,需要对深度图像进行归一化处理:

$$I_{\text{normal}}(x, y) = \frac{C_{\text{depth}}(x, y)}{L_{\text{depth}}^{\max} - C_{\text{depth}}^{\min}} \quad (6)$$

其中: C_{depth}^{\min} 为图像 C 中非零值像素对应深度最小值; L_{depth}^{\max} 为图像 L 非零值像素对应深度最大值。

考虑到目标人手为图像 C 中前景图像,因此通过设定比例阈值,分割出目标人手的二值图像,其算法流程如下。

- 1) 利用文献[19]中的配准方法,对 Kinect 获取的颜色图像和深度图像进行配准,匹配后图像如图 5(a) 所示。
- 2) 利用深度二值掩模 D^κ 对图像 C 进行或运算,得到对目标前景人手的精确分割。

其中,二值掩模为:

$$D^\kappa(x, y) = \begin{cases} 1, & I_{\text{normal}}(x, y) < \kappa \\ 0, & \text{其他} \end{cases} \quad (7)$$

由实验数据得深度阈值 $\kappa = 0.13$ 。

或运算后分割出单独人手 O_{object} :

$$O_{\text{object}} = D^\kappa \cap C \quad (8)$$

掩模后目标区域如图 5(d) 所示;这里,进行掩模检测不但能够独立分割出目标人手,另一个主要原因是由单独的深度图像会对目标产生模糊的边界,对指尖点的检测造成干扰,从而对人手分割带来误差;而多信息检测下的目标区域可能会连接背景区域,但能够检测出较为准确的目标边界,因此,结合深度和颜色图像,对目标人手连通域进行掩模分割,达到精确检测人手轮廓、剔除复杂背景的目的。

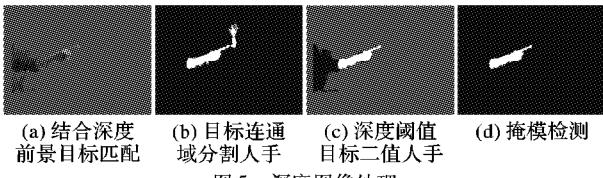


图5 深度图像处理

2 人手部件分割

对于检测到的人手,如图5(d)所示,通常情况下与小臂相连。为了后续对指尖的检测,需要把手从整个人手区域中分割出来。其算法如下。

1)Canny算子对图5(d)中人手区域进行边缘检测,得到上肢边界点;为了使得边缘轮廓闭合,利用形态学膨胀算子连通目标区域。这样,得到新的上肢边界点为:

$$\mathbf{s} = (x_1, x_2, \dots, x_m) \quad (9)$$

2)计算边界点 \mathbf{s} 的最小外接椭圆,如图6(a),得到沿手臂方向椭圆的长轴,及以长轴为横坐标的上肢宽度序列;

这里,椭圆方程可表述为:

$$\zeta = \{x' \in \mathbb{R}^2 \mid (x' - c)^T E (x' - c) \leq 1\} \quad (10)$$

其中: c 为椭圆 ζ 的中心, E 为一个 2×2 的正定对称矩阵。通过求解最优问题可得到最小闭合外接椭圆:

$$\min_{E, c} \det(E^{-1})$$

使得

$$(x' - c)^T E (x' - c) \leq 1; \quad i = 1, \dots, k; \quad E > 0 \quad (11)$$

由上式得到长轴上的像素点 $\mathbf{P} = (p_1, p_2, \dots, p_m)$ 。

3)水平变换人手长轴与横坐标平行,其夹角为 θ_{hand} 。对于每个 p_i ,计算其长轴垂线到人手边界点 x_i 的最大距离,因为手掌与手指的分割点为距离大小的突变点,即局部极值点,通过计算梯度值寻找局部极值,即为截断点。首先,找到全局极大值点,其始终存在于人手区域内;然后,沿着手指的方向移动此点直到找出局部极小值即为分割点,如图6(b),图7所示。

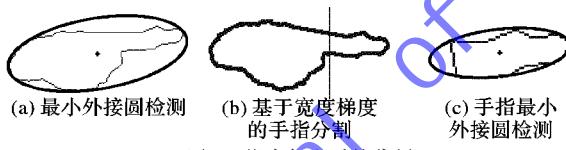


图6 指尖与人手的分割

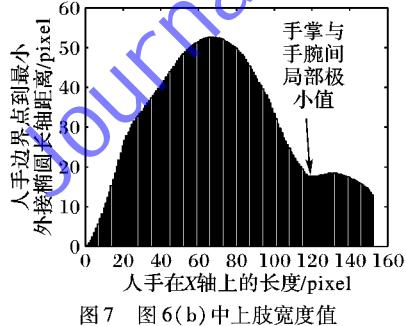


图7 图6(b)中上肢宽度值

3 指尖关键点检测

指尖关键点为指尖的中心顶点,为了查找出关键点,对于水平变换后的人手,计算闭合人手轮廓 s 的最大内接圆,利用圆心与轮廓点的曲率夹角计算指尖点集,利用指尖点集查找指尖边界点集,并拟合出指尖曲线模型,最后利用模型精确估计出指尖所在位置。

3.1 目标人手及指尖水平变换

由于研究目标为人手手指,为了使其计算方便,需要对其进行水平变换。对于手指相对于水平横坐标的角度,利用第2章所述方法,再次对手指边界点计算其最小外接圆,此时,得到其长轴方向向量与图像横坐标的夹角 θ_{finger} ,即为目标手指水平变换夹角: $\theta = \theta_{finger}$;对手指坐标点进行水平位移变换,表达公式为:

$$\mathbf{z}_t(i) = \mathbf{T}_\theta(\mathbf{z}(i)) \quad (12)$$

其中 $\mathbf{z}(i), \mathbf{z}_t(i)$ 分别为角度旋转变换前后的手指边界坐标点。

3.2 指尖关键点初始检测

本文定义闭合轮廓的最大内接圆心为距离目标人手边界最远的像素点,因此考虑对于轮廓内每个像素,计算其到轮廓边界的最小距离集合,找出最小距离中的最大值,认定其为最大内接圆圆心:

$$D(r', c') = \max \left\{ \min_{(i, j) \in \Omega_{object}} (d(s(x, y), r(i, j))) \right\} \quad (13)$$

其中: $s(x, y)$ 为目标人手边界点; $r(i, j)$ 为边界内部像素点; Ω_{object} 为目标人手区域,其最小最大距离为与边界 r' 距离最大圆心的 c' ,如图8(a)、(b)所示。

考虑到指尖点为轮廓曲线中曲率较大的边界点,因此,通过计算轮廓边界点的曲率,可以提取出指尖关键点,其算法表述如下:对于边界点 $s(i)$,计算其 k 相邻的边界点 $s(i+k), s(i-k)$ 与最大内接圆圆心 c' 所构成的向量 $V(i+k), V(i-k)$,并且计算两个向量所形成的小角度 α ;如果 α 小于设定的阈值,认定其为候选关键点;然而,对于指缝所形成的小于阈值的角度会对指尖造成误检。因此,通过判定检测关键点与掌心的距离来剔除错检指尖点,即当两点距离小于设定阈值时,去除此检测点。

3.3 指尖关键点的优化

由于检测步长 k 的不同,所检测出的候选指尖点并不都位于手指指尖中心。因此,为了更好地得到指尖中心,考虑指尖形状模型可以通过抛物线来表达,这里,对检测到的候选指尖点以及其邻域指尖边缘进行二次曲线拟合,计算抛物线的极值点,通过实验可得,指尖中心点为极值点法线方向与指尖边缘的交点,如图8(d),由此,即认为此交点为指尖中心点。最后通过反变换得到真实指尖点在像素坐标中位置: $\mathbf{P}(i) = \mathbf{T}_\theta^{-1}(\mathbf{P}_t(i))$ 。

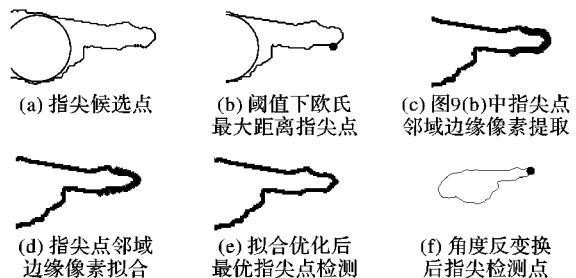


图8 指尖点初始检测

3.4 指尖关键点跟踪

针对检测出的指尖关键点,由于复杂背景及图像噪声的

影响会对检测指尖点产生误差。为了更加精确地描述手指指尖在空间中的位移变化,采用卡尔曼滤波对颜色深度匹配的视频序列中指尖点进行预测定位,达到精确检测的目的。

利用卡尔曼滤波跟踪指尖点,首先将人手运动看作一个动态系统,其运动状态为近似的匀速运动,由于存在图像深度信息,因此在预测过程中系统的状态矢量可表示为 $\mathbf{x}(k) = (u(k), v(k), d(k), u'(k), v'(k), d'(k))^T$ 。其中: $u(k), v(k)$ 和 $d(k)$ 分别表示指尖点在 Kinect 颜色摄像头坐标系中的空间坐标,而 $u'(k), v'(k)$ 和 $d'(k)$ 分别表示指尖点在三个坐标方向上的速度分量。

由此建立跟踪指尖点的系统动态方程:

$$\mathbf{x}(k) = \mathbf{F} \mathbf{x}(k-1) + \mathbf{B} \mathbf{u} + \mathbf{w}(k) \quad (14)$$

由检测得到的指尖点观测值与理想值之间建立系统观测方程为:

$$\mathbf{z}(k) = \mathbf{H} \mathbf{x}(k) + \mathbf{v}(k) \quad (15)$$

其中 $\mathbf{w}(k), \mathbf{v}(k)$ 为服从均值为零的白噪声。

4 实验及分析

本实验视频数据通过利用 OpenNI 开发环境,微软 Kinect 摄像头捕获到每秒 30 帧,大小为 480×640 的颜色和深度图像帧,如图 9,10 中分别选取原始未经配准的颜色图像帧和深度图像帧。实验结果主要分为两部分,即目标人手的识别和指尖点检测。下面对这两部分分别进行分析。

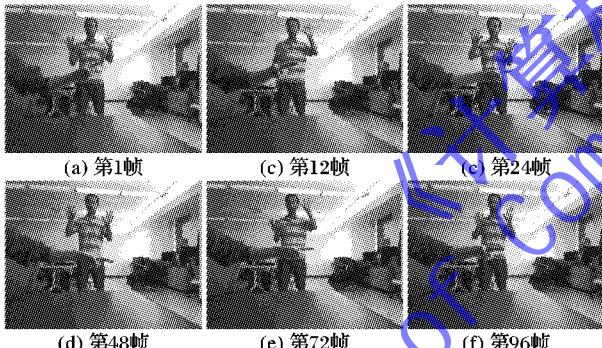


图 9 手指运动颜色图像帧序列

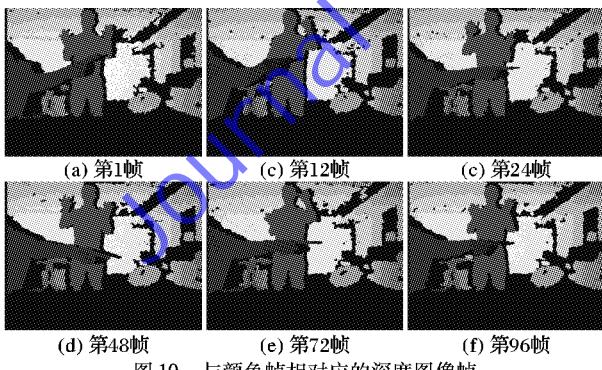


图 10 与颜色帧相对应的深度图像帧

4.1 人手检测分析

为了精确地识别出人手区域,算法融合各种信息特征,结合匹配深度图像最终识别的目标人手如图 5(d)所示;图 1~3 分别为基于肤色、差分帧和运动块残差的检测,然而,复杂背景会对单独人手的检测带来干扰,图 1(d)、(e)、(f)中,由于前景桌子和一些背景的颜色近似于肤色,因此,检测后并不能单独提取出目标人手。图 2(e)中,由于背景人的运动,检

测后分别得到目标人手与背景人运动区域。图 3 基于块残差的检测仍然不能完全检测出目标。因此结合以上信息,得到如图 4(a)中的检测结果,利用文中提出的最大连通域算法得到目标人手存在的最小区域,如图 4(b)、(c)所示。这里,背景的干扰降到最小,然而并不能完全消除,此时,引入校准后的深度信息,图 11 为对图像的深度预处理结果。为了得到目标前景人手所在区域,对深度图像进行归一化后,结合阈值进行前景目标分割,图 11 分别为对第 1 帧和 24 帧图像进行深度预处理,通过实验分析,当阈值 $\alpha = 0.13$ 时,能够大致单独地分割出目标人手,如图 11(b)、(e)所示;图 11(c)、(f)为分割后二值化结果。图 5(a)、(b)为第 24 帧提取后的前景深度图像与多信息检测后图像的叠加结果,其中,目标区域外的灰色为背景颜色。掩模后得到理想的目标人手区域如图 5(c)所示。

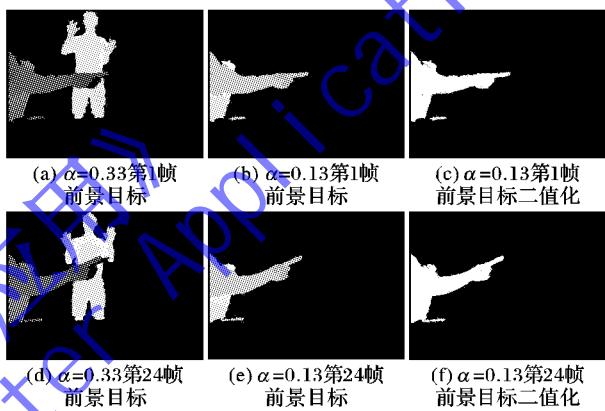


图 11 归一化深度图像前景目标人手提取

4.2 指尖点检测分析

对检测后目标人手水平变换如图 6 所示。其中:(a)为对目标的最小外接圆检测;(b)为初次水平变换后利用垂直方向边缘像素距离导数对手指的分离;(c)为再次利用最小外接圆,计算手指的水平夹角。图 8(a)首先通过计算最大内接圆计算得到指尖候选点集,选取距离圆心最大的点为初始指尖估计点,如图 8(b)所示;再对初始估计点邻域边界点,图 8(c),进行拟合得到拟合曲线为图 8(d);通过计算极值点两侧相邻点所构成的法线,其与手轮廓最近的交点即为精确估计得到的指尖中心点,如图 8(e)所示,图 8(f)为指尖点在颜色图像中的所在位置。

图 12 分别为目标人手在不同空间角度下(俯仰、偏航、翻转),检测指尖点与实际指尖点(Ground Truth, GT)在深度摄像头坐标下的误差(本文规定深度摄像头坐标系等于世界坐标系);由图 12(a)、(b)、(c)中,菱形标记为经过拟合优化后检测到的指尖点与真实点之间的误差,圆圈标记为未经过拟合优化后(距离最大阈值方法)的指尖点与真实值指尖的误差,由图可以看出,本文算法的误差均小于距离最大阈值计算下的指尖点(圆圈标记),并且保持较低的误差值,表明了本文算法的有效性。

图 13 比较了在本文算法基础上加入卡尔曼滤波得到的指尖点和利用视觉信息(未加入深度信息)检测到的指尖点分别与人工标注真实点之间的误差。图 13(a)中圆圈标记为本文算法,其误差大大低于仅通过视觉检测得到的指尖点误差(方块标记),由于单一视觉检测未能加入图像深度信息,

因此,在某些帧上由于背景的干扰使得其检测到的指尖点误差很大,如(20~60帧)。图13(b)为(a)中卡尔曼跟踪指尖点误差的放大图。

为了进一步验证本文检测指尖方法的有效性,对手写数字视频数据库中的指尖轨迹进行跟踪。在手写数字数据库中,阿拉伯数字(0~9)通过手指运动被描绘出来,如图14

所示。数字轨迹所在平面大致与深度摄像头 y - z 平面平行,这里,规定深度摄像头坐标系为世界坐标系。对指尖描绘的每个数字轨迹进行跟踪,当检测指尖点与真实值误差 $e \leq 3$ mm时,即认为成功检测此点,其检测率如表1所示。由表1可看出,在数字(0~9)中,本文算法得到的指尖点检测率高于传统检测算法(最大阈值法),从而验证了本文算法的有效性。

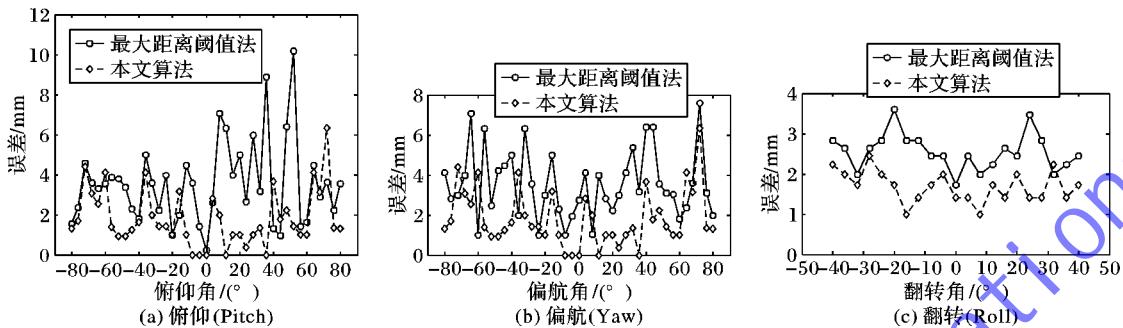


图12 指尖检测点不同姿态下误差比较

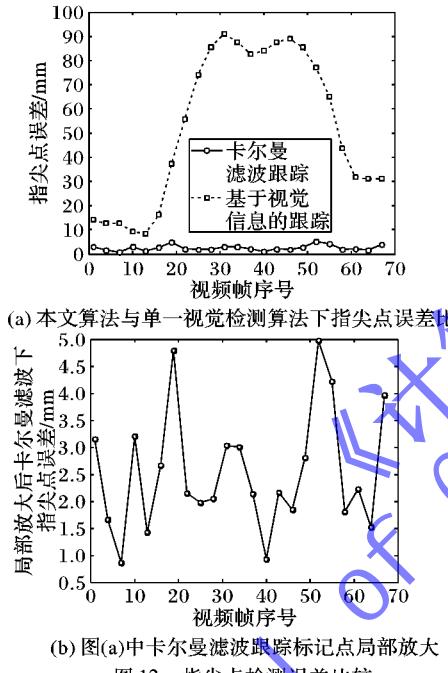


图13 指尖点检测误差比较

5 结语

本文提出了一种复杂背景下对目标人手指尖点检测的方法,将指尖检测问题分为两部分,首先对视频中的目标人手进行基于多信息融合下的检测,结合校准后的深度信息,进一步提取单独的目标人手;其次,对目标手指分割、水平变换后,利用人手边界像素的曲率关系检测初始指尖点,拟合优化,空间变换后,得到检测点;最后利用卡尔曼滤波,获得实时的精确指尖点空间位置。实验表明,与传统方法相比,本文算法所检测到的指尖与真实指尖数据有较小的误差,具有较高的鲁棒性。下一步将通过研究指尖点和摄像机空间变换关系,进一步计算指尖在空间中指示方向。

表1 手写数字数据库上指尖点跟踪检测率对比 %

手写数字	本文方法	最大阈值法	手写数字	本文方法	最大阈值法
0	94.0	85.3	5	87.3	85.5
1	96.0	89.5	6	91.5	79.4
2	94.7	86.4	7	90.3	82.3
3	93.5	89.5	8	94.7	90.0
4	95.4	90.4	9	92.5	84.0

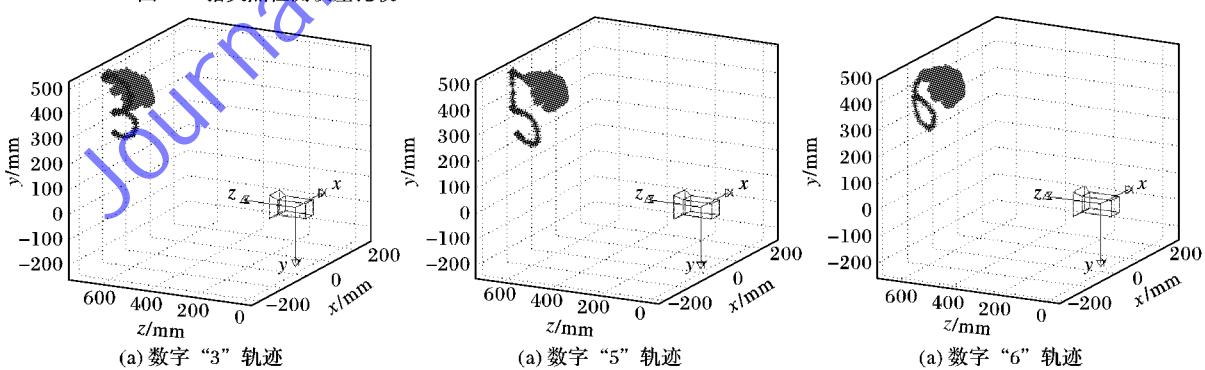


图14 三维视角数字手指运动轨迹在深度摄像头坐标系中可视化显示

参考文献:

- [1] ROUSSOS A, THEODORAKIS S, PITSIKALIS V, et al. Hand tracking and affine shape-appearance hand shape sub-units in continuous sign language recognition [C]// ECCV '10: Proceedings of the 11th European Conference on Trends and Topics in Computer Vision—Volume Part I. Berlin: Springer, 2012: 258–272.
- [2] WACHS J P, KÖLSCH M, STERN H, et al. Vision-based hand-gesture application [J]. Communications of the ACM, 2011, 54(2): 60–71.
- [3] BROWN J A, CAPSON D W. A framework for 3D model-based

- visual tracking using a GPU-accelerated particle filter [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(1): 68–80.
- [4] OIKONOMIDIS I, KYRIAZIS N, ARGYROS A A. Efficient model-based 3D tracking of hand articulations using Kinect [C]// Proceedings of the 22nd British Machine Vision Conference. Norwich: BMVA Press, 2011: 1–11.
- [5] LEE D, PARK Y. Vision-based remote control system by motion detection and open finger counting [J]. *IEEE Transactions on Consumer Electronics*, 2009, 55(4): 2308–2313.
- [6] KANG S K, NAM M Y, RHEE P K. Color based hand and finger detection technology for user interaction [C]// Proceedings of the 2008 International Conference on Convergence and Hybrid Information Technology. Washington, DC: IEEE Computer Society, 2008: 229–236.
- [7] NGUYEN D D, PHAM T C, PHAM X D, et al. Finger extraction from scene with grayscale morphology and blob analysis [C]// Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics. Piscataway: IEEE Press, 2009: 324–329.
- [8] THANGALI A, SCLAROFF S. An alignment based similarity measure for hand detection in cluttered sign language video [C]// Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 89–96.
- [9] ALON J, ATHITSOS V, YUAN Q, et al. A unified framework for gesture recognition and spatiotemporal gesture segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(9): 1685–1699.
- [10] KARLINSKY L, DINERSTEIN M, HARARI D, et al. The chains model for detecting parts by their context [C]// Proceedings of the 2010 Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2010: 25–32.
- [11] ELMEZAIN M, AL-HAMADI A, MICHAELIS B. Real-time capable system for hand gesture recognition using hidden Markov models in stereo color image sequences [J]. *Journal of Winter School of Computer Graphics*, 2008, 16(1): 65–72.
- [12] van den BERGH M, van GOOL L. Combining RGB and ToF cameras for real-time 3D hand gesture interaction [C]// Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision. Piscataway: IEEE Press, 2011: 66–72.
- [13] REN Z, MENG J, YUAN J, et al. Robust hand gesture recognition with Kinect sensor [C]// Proceedings of the 19th ACM International Conference on Multimedia. New York: ACM, 2011: 759–760.
- [14] KERDVIBULVECH C, SAITO H. Vision-based detection of guitar players' fingertips without markers [C]// Proceedings of the 2007 Computer Graphics, Imaging and Visualization. Piscataway: IEEE Press, 2007: 419–428.
- [15] VERMA R, DEV A. Vision based hand gesture recognition using finite state machines and fuzzy logic [C]// Proceedings of the 2009 International Conference on Ultra Modern Telecommunications and Workshop. Piscataway: IEEE Press, 2009: 1–6.
- [16] NGUYEN D D, PHAM T C, JEON J W. Fingertip detection with morphology and geometric calculation [C]// Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2009: 1460–1465.
- [17] LEE D, LEE S. Vision-based finger action recognition by angle detection and contour analysis [J]. *ETRI Journal*, 2011, 33(3): 415–422.
- [18] SPONG M W, HUTCHINSON S, VIDYASAGAR M. Robot modeling and control [M]. Hoboken: John Wiley & Sons, 2005: 163–182.
- [19] STENGER B, MENDONCA P R S, CIOPOLLA R. Model-based 3D tracking of an articulated hand [C]// Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2001, 2: 310–315.
- [20] QIN S, YANG Y, JIANG Y. Model-based 3D tracking of an articulated hand from single depth images [J]. *Pattern Recognition Letters*, 2013, 34(12): 1437–1445.
- [21] ALPRECHT I, HABER J, SEIDEL H P. Construction and animation of anatomically based human hand models [C]// Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Aire-la-Ville: Eurographics Association Press, 2003: 98–109.
- [22] JULIER S J, UHLMANN J K, DURRANT-WHYTE H F. A new approach for filtering nonlinear system [C]// Proceedings of the 1995 American Control Conference. Piscataway: IEEE Press, 1995, 3: 1628–1632.
- [23] RUBINSTEIN R Y, KROESE D P. Simulation and the Monte Carlo method [M]. 2nd ed. Hoboken: John Wiley & Sons, 2007: 25–37.
- [24] REN H, ZHU Y, XU G, et al. Hand gesture segmentation and recognition with complex background [J]. *Acta Automatica Sinica*, 2002, 28(2): 256–261. (任海兵, 祝远新, 徐光祐, 等. 复杂背景下的手势分割与识别[J]. 自动化学报, 2002, 28(2): 256–261.)
- [25] YUAN Q, SCLAROFF S, ATHITSOS V. Automatic 2D hand tracking in video sequences [C]// WACV-MOTION '05: Proceedings of the 7th IEEE Workshops on Application of Computer Vision. Washington, DC: IEEE Computer Society, 2005, 1: 250–256.
- [26] LIU B, YAN L, ZHOU D. Robust template matching method based on minimum maximum circular similarity measurement [J]. *Journal of Infrared and Laser Engineering*, 2006, 35(5): 618–623. (刘宝生, 闫莉萍, 周东华. 基于最小最大圆度量的鲁棒模板匹配方法[J]. 红外与激光工程, 2006, 35(5): 618–623.)
- [27] DANIEL H C, JUHO K, JANNE H. Joint depth and color camera calibration with distortion correction [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(10): 2058–2064.