

# 融合红外深度信息的视觉交互手部跟踪算法

孙 瑾<sup>1,2</sup>, 丁永晖<sup>1,2</sup>, 周 来<sup>3</sup>

<sup>1</sup>南京航空航天大学民航学院, 江苏 南京 210016;

<sup>2</sup>南京航空航天大学飞行模拟与先进培训工程技术研究中心, 江苏 南京 210016;

<sup>3</sup>上海航天技术研究院上海机电工程研究所, 上海 201109

**摘要** 在虚拟现实环境中手部跟踪是视觉交互系统的基础和核心。针对现有视觉跟踪方法在手部运动姿态、尺度变化及复杂背景条件下出现的稳健性等问题, 结合纹理和轮廓信息, 利用基于梯度方向局部二值模式特征为基础的粒子滤波跟踪算法, 建立局部和全局的特征直方图描述, 实现手部跟踪。针对粒子匮乏问题, 利用红外深度信息, 并引入基于群智能的人工蜂群算法, 将当前时刻的观测信息融合在粒子预测的采样和更新阶段, 高效完成目标的搜索和优化, 降低粒子集衰减程度, 改善状态估计精度。实验结果表明, 该方法在各种复杂背景下可以实现手部的稳健跟踪。

**关键词** 机器视觉; 视觉跟踪; 粒子滤波; 人工蜂群算法; 红外深度信息; 梯度方向二值模式特征

**中图分类号** TP391.4      **文献标识码** A

**doi:** 10.3788/AOS201737.0115002

## Visually Interactive Hand Tracking Algorithm Combined with Infrared Depth Information

Sun Jin<sup>1,2</sup>, Ding Yonghui<sup>1,2</sup>, Zhou Lai<sup>3</sup>

<sup>1</sup> College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China;

<sup>2</sup> Engineering Technology Research Center of Flight Simulation & Advanced Training, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China;

<sup>3</sup> Shanghai Electro-Mechanical Engineering Institute, Shanghai Academy of Spaceflight Technology, Shanghai 201109, China

**Abstract** Hand tracking is the basis and core problem of vision interaction in virtual reality. Due to the poor performance resulted from the existing hand tracking methods in the instances of movement, scale change, complex background, etc., a new hand tracking algorithm is presented. The proposed algorithm is under particle filtering and tracking framework, which adopts oriented gradient local binary pattern descriptor integrating with texture and contour information. Furthermore, the infrared depth information is introduced. The proposed tracking algorithm combines the observation information of current frame in the stages of particle sampling and updating by the artificial bee colony algorithm, which overcomes the degeneracy problem in particle filtering and improves hand tracking precision by optimization of the space search. Experimental results show that the proposed algorithm can achieve accurate and robust hand tracking in complex background.

**Key words** machine vision; visual tracking; particle filtering; artificial bee colony algorithm; infrared depth information; oriented gradient local binary pattern

**OCIS codes** 150.0155; 110.2960; 100.3008

## 1 引 言

虚拟现实技术具有巨大的潜力,因而受到广泛关注。虚拟现实设备,特别是头盔显示器技术的发展,促使虚拟现实技术逐步走向成熟。然而,用户在配戴头盔显示器后无法看到自己的手,导致交互性差。目前主

**收稿日期:** 2016-07-08; **收到修改稿日期:** 2016-09-19

**基金项目:** 国家自然科学基金(51305255)、上海市自然科学基金(13ZR1455900)、江苏省研究生创新计划(SJLX150130)

**作者简介:** 孙 瑾(1977—),女,博士,副教授,主要从事机器视觉、视觉跟踪与识别等方面的研究。

E-mail: sunjinly@nuaa.edu.cn

要采用数据手套实现交互,成本昂贵,并且入侵性的交互方法影响操作者的沉浸感,无法提供丰富的感知。近年来,人们对人机交互提出了更高的要求,基于视觉的非接触交互方式充分体现了自然和谐交互特征,成为国内外的关注热点。基于视觉的无标记手部交互技术从图像中检测出手,通过视觉跟踪确定图像序列中运动的手部,并识别手部姿态以驱动视景中的模型手,实现交互。手部视觉跟踪是实现自然交互的首要步骤,直接关系到后续姿态识别的准确率。

目前手部跟踪主要分为基于模型的方法和基于表观的方法。基于模型的方法利用 3D 手模型投影得到预测手型图像,通过与实时采集图像间差异比较来修正相应姿态参数。Oikonomidis 等<sup>[1]</sup>利用圆柱,Qian 等<sup>[2]</sup>设计球状体手模型,基于粒子优化算法实现姿态预测跟踪,但由于手部运动的自由度高,计算复杂度<sup>[3]</sup>高。基于表观的方法是通过预先给出的手势图像库,构建图像特征空间到姿态空间的映射,根据图像特征直接估计姿态参数,其中特征提取是关键环节。为了提高该方法的稳健性,Liu 等<sup>[4]</sup>融合多个几何特征,完成手势跟踪识别,杨波等<sup>[5]</sup>根据空间统计分布特征跟踪手部运动,Priyal 等<sup>[6]</sup>结合肤色和 Krawtchouk 矩完成手势分割识别,但多特征融合会使特征维数增大,影响算法实时性。

深度信息不易受环境干扰,稳定性好,为基于视觉手部交互提供了新的研究思路。Doliotis 等<sup>[7]</sup>比较了基于颜色和深度图像的手部检测结果,发现深度信息能明显提高检测的精度和稳定性。然而,目前研究主要集中在深度阈值分割,包括区域生长和均值聚类等<sup>[8-10]</sup>,基于深度的跟踪算法则比较少,且主要通过检测代替跟踪。例如,Park 等<sup>[11]</sup>提取颜色和纹理(RGBD)特征检测指尖实现手部跟踪,但当指尖遮挡后,跟踪失效;Chen 等<sup>[12]</sup>运用区域生长分割手部区域,采用 MeanShift 算法实现手部定位,但当手部运动过快或与其他区域连通时,会导致跟踪错误;Qi 等<sup>[13]</sup>组合颜色和深度构建特征空间,应用 K-means 聚类法分离手部区域进行跟踪,但手部快速运动或姿态改变会造成跟踪失败。

## 2 人工蜂群手部粒子滤波跟踪模型

运动状态空间寻求最优解的过程,是寻找与目标描述最一致的图像区域。作为贝叶斯框架下的跟踪方法,粒子滤波(PF)通过非参数化的蒙特卡罗采样方法,实现递推贝叶斯滤波,可以解决非线性和非高斯情况下的目标跟踪问题<sup>[14]</sup>。针对手部运动姿态和尺度变化的问题,本文在粒子滤波框架下实现手部跟踪。传统粒子滤波存在粒子匮乏问题,基于群智能优化思想的 PF 则是现代粒子滤波发展的一个崭新方向<sup>[15]</sup>。Park 等<sup>[16]</sup>引入蚁群优化算法增强全局寻优能力,宋宇等<sup>[17]</sup>提出了自适应粒子群优化算法改进粒子滤波算法,田梦楚等<sup>[18]</sup>引入优胜劣汰的萤火虫算法提升精度,降低运算复杂度。针对粒子匮乏现象,本文利用深度信息融合人工蜂群算法,通过个体局部寻优行为在群体获得全局最优值以改善粒子滤波估计性能,提高跟踪精度。

### 2.1 标准粒子滤波

动态系统的数学模型通常可描述为

$$\begin{cases} \mathbf{X}_t = f(\mathbf{X}_{t-1}, \mathbf{V}_{t-1}) \\ \mathbf{Z}_t = h(\mathbf{X}_t, \mathbf{U}_t) \end{cases}, \quad (1)$$

式中  $\mathbf{X}_t$  和  $\mathbf{X}_{t-1}$  分别为  $t$  和  $t-1$  时刻的系统状态向量,  $\mathbf{Z}_t$  为  $t$  时刻的观测向量,  $\mathbf{V}_{t-1}$  为  $t-1$  时刻的系统噪声,  $\mathbf{U}_t$  为  $t$  时刻的观测噪声,  $f(\cdot)$  为系统状态转移方程,  $h(\cdot)$  为系统观测模型。

根据贝叶斯滤波原理,跟踪是在给定观测数据  $\mathbf{Z}_{1:t}$  的条件下估算状态向量  $\mathbf{X}_t$ ,即利用状态转移概率  $p(\mathbf{X}_t | \mathbf{X}_{t-1})$  和观测概率  $p(\mathbf{Z}_t | \mathbf{X}_t)$  计算后验概率  $p(\mathbf{X}_t | \mathbf{Z}_{1:t})$ ,实现目标状态估计。粒子滤波采用蒙特卡罗方法实现递归贝叶斯推理,基本思想是从概率分布  $p(\mathbf{X}_{0:t} | \mathbf{Z}_{1:t})$  中抽取  $N$  个样本  $\{\mathbf{X}_t^i, i=1, 2, \dots, N\}$ ,用样本均值来近似目标概率分布,可表示为

$$p(\mathbf{X}_t | \mathbf{Z}_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta(\mathbf{X}_t - \mathbf{X}_t^i), \quad (2)$$

式中  $\delta(\cdot)$  为狄拉克函数。

实际情况中  $p(\mathbf{X}_{0:t} | \mathbf{Z}_{1:t})$  可能是多变量且非标准的,导致抽样比较困难。运用重要性采样(IS)从一个已知且容易采样的参考分布  $q(\mathbf{X}_{0:t} | \mathbf{Z}_{1:t})$  中采样,则后验概率密度函数可以离散地加权近似为

$$p(\mathbf{X}_t | \mathbf{Z}_{1:t}) = \sum_{i=1}^N \omega_i^i \delta(\mathbf{X}_t - \mathbf{X}_t^i), \quad (3)$$

式中  $\omega_i^i$  为重要性权值,可由重要性密度函数  $q(\mathbf{X}_{0:t} | \mathbf{Z}_{1:t})$  求得:

$$\omega_i^i = \frac{p(\mathbf{Z}_{1:t} | \mathbf{X}_{0:t}^i) p(\mathbf{X}_{0:t}^i)}{q(\mathbf{X}_{0:t}^i | \mathbf{Z}_{1:t})}. \quad (4)$$

采用序列重要性抽样(SIS),将重要性密度函数表示为连乘形式:

$$q(\mathbf{X}_{0:t} | \mathbf{Z}_{1:t}) = q(\mathbf{X}_0) \prod_{j=1}^k q(\mathbf{X}_j | \mathbf{X}_{0:j-1}, \mathbf{Z}_{1:t}). \quad (5)$$

假设状态符合马尔科夫过程,且给定状态下观测条件独立,则可将重要性权值  $\omega_i^i$  表示为递推形式:

$$\omega_i^i = \omega_{i-1}^i \frac{p(\mathbf{Z}_t | \mathbf{X}_t^i) p(\mathbf{X}_t^i | \mathbf{X}_{t-1}^i)}{q(\mathbf{X}_t^i | \mathbf{X}_{0:t-1}^i, \mathbf{Z}_{1:t})}, \quad (6)$$

如果状态估计过程是最优估计,则参考分布概率密度函数只依赖于  $\mathbf{X}_{t-1}$  和  $\mathbf{Z}_t$ ,则

$$\omega_i^i = \omega_{i-1}^i \frac{p(\mathbf{Z}_t | \mathbf{X}_t^i) p(\mathbf{X}_t^i | \mathbf{X}_{t-1}^i)}{q(\mathbf{X}_t^i | \mathbf{X}_{t-1}^i, \mathbf{Z}_t)}. \quad (7)$$

粒子滤波运用序列抽样思想,从  $q(\mathbf{X}_t | \mathbf{X}_{t-1}, \mathbf{Z}_t)$  中随机抽取  $N$  个样本,计算  $p(\mathbf{X}_t | \mathbf{X}_{t-1})$  和  $p(\mathbf{Z}_t | \mathbf{X}_t)$ ,利用(7)式计算对应样本重要性权重并进行归一化处理,最后利用(3)式对  $p(\mathbf{X}_t | \mathbf{Z}_{1:t})$  进行估计。

## 2.2 人工蜂群优化

鉴于  $p(\mathbf{X}_{0:t} | \mathbf{Z}_{1:t})$  存在采样困难的问题,传统粒子滤波一般选择目标先验密度函数作为重要密度函数,可表示为

$$q(\mathbf{X}_t | \mathbf{X}_{0:t-1}, \mathbf{Z}_{1:t}) = p(\mathbf{X}_t | \mathbf{X}_{t-1}). \quad (8)$$

这种方法易于实现,但不考虑后验信息会导致估计精度有限,特别是在手部跟踪过程中,手部图像尺度随着手部姿态的改变而发生变化,如果不考虑当前观察信息,就不能准确提取目标特征,会引起目标特征漂移,导致跟踪精度下降甚至丢失目标,最后出现粒子匮乏现象。重采样牺牲计算量来消除粒子匮乏,并且重采样后出现大量重复的粒子,缺少多样性<sup>[19]</sup>。

针对手部姿态和轮廓在跟踪期间随机变化性大的问题,胡士强等<sup>[19]</sup>选用了随机扰动模型,但误差较大;贾静平等<sup>[20]</sup>利用 Lindeberg 尺度空间理论,在不同尺度不同位置处计算 Laplacian 滤波函数极值确定目标大小,但计算量较大。本文引入当前帧的红外深度观测信息,采用人工蜂群(ABC)算法对粒子分布进行优化。

ABC算法是2005年由Karaboga<sup>[21]</sup>提出的一种基于蜜蜂觅食的群体智能算法。蜂群在觅食过程中有三个重要要素:花蜜源、被雇佣蜜蜂和未被雇佣蜜蜂。被雇佣蜜蜂又称为引领蜂,未被雇佣蜜蜂分为跟随蜂和侦查蜂。侦查蜂寻找蜜源,引领蜂在该蜜源邻域内搜索新的蜜源。如果新蜜源所含花粉数量更多(适应度更优),则标记新蜜源,飞回巢穴通过摇摆舞向跟随蜂分享蜜源信息。跟随蜂通过一定的概率选择蜜源,并在其邻域内继续搜索新的蜜源,择优选择,最终找到最优蜜源。

本文将粒子作为蜜源,每个引领蜂和跟随蜂的位置代表每帧跟踪的手部位置,蜜源的丰富程度对应该位置与目标模板的相似度,跟踪的目标就是寻找适应度最大的引领蜂或跟随蜂的位置。粒子滤波算法主要包括预测采样、更新、重采样和状态估计输出。预测采样阶段建立手部观测模型,利用当前帧观察像素深度值,由引领蜂和跟随蜂的局部寻优解决目标尺度自适应变化跟踪。随着手部姿态改变,若粒子(蜜源)权重无法提高,则释放侦查蜂寻找新的蜜源(粒子);更新阶段针对手部运动启动目标模板更新以提高跟踪准确度。算法跟踪流程如图1所示。

相邻帧之间手部运动机动性不大,可采用常速模型描述,因此采用一阶自回归(ARP)方程建立手部运动模型。设目标状态为  $\mathbf{X} = [x, \Delta T, y, g]$ ,其中,目标中心位置为  $(x, y)$ ,  $\Delta T$  为采样周期,  $g$  为对应目标尺度收缩系数。目标状态随时间变化的过程可以描述为

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\mathbf{w}_{t-1} = \begin{bmatrix} 1 & v_x & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & v_y & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{X}_{t-1} + \mathbf{B}\mathbf{w}_{t-1}, \quad (9)$$

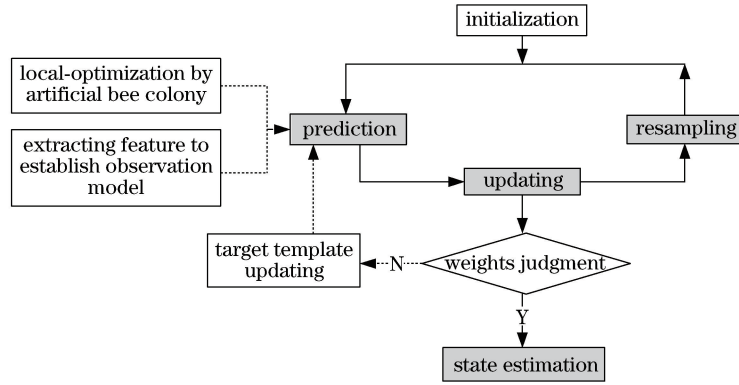


图 1 基于人工蜂群算法的粒子滤波跟踪流程

Fig. 1 Particle filtering and tracking process based on ABC algorithm

式中  $\mathbf{A}$  为系统状态转移矩阵,  $B$  为粒子传播半径,  $\mathbf{w}_{t-1} = [\omega_x, 1, \omega_y, 1]^T$  为多元高斯随机噪声,  $\omega_x$  和  $\omega_y$  为高斯白噪声, 满足  $\omega_x \sim N(0, \sigma_x^2)$ ,  $\omega_y \sim N(0, \sigma_y^2)$ ,  $v_x$  和  $v_y$  分别为目标中心在图像中  $x$  和  $y$  方向的速度。

在预测采样阶段, 根据(9)式获得蜜源(粒子)位置。为优化蜜源位置, ABC 算法中引领蜂在每个蜜源位置进行邻域搜索, 产生新的蜜源为

$$\mathbf{X}_{it}^{\text{new}} = \mathbf{X}_{it} + \varphi(\mathbf{X}_{it} - \mathbf{X}_{jt}), \quad (10)$$

式中  $\mathbf{X}_{it}$  为  $t$  时刻当前粒子(蜜源) $i$  的位置,  $\mathbf{X}_{jt}$  为  $t$  时刻随机选择的邻域新蜜源  $j$  的位置,  $\varphi$  为  $[-1, 1]$  区间内的随机数。

新旧蜜源根据贪婪原则对比完成择优替换, ABC 算法中新蜜源  $\mathbf{X}_{it}^{\text{new}}$  随机产生, 效率低, 且无法对应手部尺度和姿态变化。根据当前蜜源的适应度  $F_i$ , 选取新蜜源对应目标尺度收缩系数  $g_i = 1/F_i$ 。根据手部图像深度平滑连续性, 由 Kinect 获取深度信息, 在新旧蜜源位置选取深度最大聚集区域作为目标候选区域, 计算特征值相似度。如果新蜜源优于旧蜜源则进行替换, 并记录当前蜜源(粒子)深度分布值  $d_i$ 。该方法避免了随机产生的无效蜜源(粒子)的计算, 并通过深度信息引入当前观测信息。

蜜源的适应度  $F_i$  定义为粒子  $i$  与目标模板的相似度, 计算公式为

$$F_i = \rho_i^i[\mathbf{p}, \mathbf{q}], \quad (11)$$

式中  $\rho_i^i$  为  $t$  时刻粒子  $i$  与目标模板的相似系数, 通常采用 Bhattacharyya 系数<sup>[18]</sup> 计算:

$$\rho_i^i[\mathbf{p}, \mathbf{q}] = \sum_{u=1}^m \sqrt{p(u)q_i^i(u)}, \quad (12)$$

式中  $\mathbf{p} = \{p(u)\}$  和  $\mathbf{q} = \{q_i^i(u)\}$  分别表示目标模板和  $t$  时刻各个粒子的特征分布,  $m$  为特征分布维数。

跟随蜂根据引领蜂传递的蜜源(粒子)信息, 按照一定的概率选择蜜源(粒子):

$$P_i = F_i / \sum_{i=1}^N F_i, \quad (13)$$

式中  $P_i$  为蜜源(粒子) $i$  的归一化适应度。

对  $P_i$  从大到小排序, 只选取前  $N_{\text{eff}}$  个蜜源(粒子)进行邻域搜索,  $N_{\text{eff}}$  为有效粒子数<sup>[19]</sup>,

$$N_{\text{eff}} = 1 / \sum_{i=1}^N (\omega_i^i)^2. \quad (14)$$

跟踪蜂同样采用深度信息, 搜索有效蜜源(粒子) $i$  的位置, 对应深度值  $d_i$  的最大连通区域作为新蜜源, 计算新旧蜜源特征值相似度进行选择替换。

经过多帧搜索后, 如果位置  $k$  处蜜源(粒子)连续  $n$  帧都没有成为有效蜜源(粒子), 则与该蜜源对应的引领蜂变为侦查蜂, 产生新蜜源(粒子)为

$$\mathbf{X}_{kt}^{\text{new}} = \mathbf{X}_{t-1} + \text{rand}(0, 1) \times (\mathbf{X}_{t-1} - \mathbf{X}_{t-2}), \quad (15)$$

式中  $\mathbf{X}_{t-1}$  和  $\mathbf{X}_{t-2}$  分别对应  $t-1$  和  $t-2$  时刻的目标输出位置。

### 2.3 目标模板更新

目标模板有效性是实现稳健视觉跟踪的前提和基础<sup>[22]</sup>。运动过程中手部尺度变化和姿态改变造成与

初始目标模板的相似度降低,此时需更新目标模板。目标表观特征的变化可能是由目标自身的变化或遮挡引起的。通过计算输出粒子  $i$  与当前目标模板的相似程度进行判断,考虑发生遮挡时相似度较低,根据实验设定相似性系数双阈值为  $\rho_{\min}=0.3, \rho_{\max}=0.5$ 。 $t$  时刻跟踪目标相似度为  $\rho_t, \rho_t < \rho_{\min}$  时判定发生遮挡不更新模板;  $\rho_t < \rho_{\max}$  时判定目标姿态发生变化,启动模板更新。以当前帧的深度阈值为依据,在此深度范围内融合肤色特征实现模板更新。

### 3 系统的观测模型

为易于抽样,选取重要性密度函数为状态转移模型,根据(7)式计算每个粒子的权值为

$$\omega_i = \omega_{i-1} p(\mathbf{Z}_i | \mathbf{X}_i^i). \quad (16)$$

可以看出,每个粒子的权值与观测似然函数相关,观测似然函数用来表征系统状态的相似性,与提取的目标特征表示相关,直接影响跟踪的稳健性<sup>[23]</sup>。

肤色特征是手部检测跟踪中常用的特征<sup>[24-25]</sup>,然而当图像中出现人脸后将无法准确定位手部。为此,将肤色与 haar 特征融合,利用 haar 特征排除人脸进行手部定位跟踪<sup>[26-28]</sup>。鉴于对环境光照的敏感性,且要求被识别对象的形状一致性较好,这种方法仅适用于固定手势特征的提取。尺度不变特征变换(SIFT)<sup>[29-30]</sup>具有尺度变换和仿射变换不变性,但实验发现手部区域 SIFT 特征点数量少,可区分度低;形状(或轮廓)特征<sup>[31]</sup>对光照变化不敏感。Chang 等<sup>[32]</sup>提取 Zernike 矩特征,Liu 等<sup>[33]</sup>提取手部 Hu 特征,实验表明不变量不足以准确描述复杂形状,只能对固定手势进行跟踪定位<sup>[34]</sup>。梯度方向直方图(HOG)最初用于行人检测<sup>[35-36]</sup>,周来<sup>[37]</sup>融合 HOG 与其他特征进行手部识别,但 HOG 特征向量维数过高会增加特征提取及跟踪过程中的相似性匹配计算量,影响系统运行速度。

为排除脸部对手部跟踪的干扰,考虑手部运动瞬时变化量小,采用梯度方向局部二值模式(OG\_LBP)特征<sup>[38]</sup>,根据邻域内八方向位置的对应灰度差粗略确定边缘梯度方向<sup>[39]</sup>,通过双阈值构造梯度方向 LBP 特征,并将纹理与轮廓信息结合,实现手部的快速、稳定跟踪。

OG\_LBP 特征通过比较邻域对称点之间的差异捕捉梯度信息,进行二值描述为

$$B_{OL,P,R} = \sum_{i=0}^{P/2-1} v(g_i - g_{i+P/2}) \cdot 2^i, \quad (17)$$

式中  $g_i$  和  $g_{i+P/2}$  分别代表邻域内以中心点为对称点的一组像素点的灰度值; $P$  是半径为  $R$  的环形邻域内的像素个数,坐标不是整数位置的采样点,可通过双线性插值得到; $v(\cdot)$  为阶跃响应函数,

$$v(x) = \begin{cases} 1, & |x| \geq T \\ 0, & |x| < T \end{cases}, \quad (18)$$

式中  $T$  为设置的阈值。

为消除图像旋转影响,参考扩展 LBP 特征<sup>[40]</sup>定义旋转不变 OG\_LBP 特征为

$$R_{OL,P,R}^i = \min\{\text{ROR}(B_{OL,P,R}, k) \mid k = 0, 1, \dots, P/2 - 1\}, \quad (19)$$

式中  $\text{ROR}(x, k)$  为二进制数  $x$  循环右移  $k$  位操作。

实验发现不同  $T$  值下 OG\_LBP 描述特征不同(图 2), $T$  值较小时,注重细节描述,随着  $T$  值不断增大,图像纹理信息逐渐减弱,轮廓信息得到加强。本文采用双阈值,选用具有平移、旋转、缩放不变性等特点,且计算代价小的直方图进行特征描述。利用  $T_1$  和  $T_2$  分别提取手部和脸部的局部特征 OG\_LBP 直方图(图 3),其中  $T_1$  取灰度图像像素级的 1%,即  $T_1 = 256 \times 1\% = 2.56 \approx 3$ ;  $T_2$  采用 Otsu 阈值确定方法,由类间方差最大准则确定,选取阈值  $T_2 = 20$ 。从图 3 可以看出, $T_1$  阈值下提取纹理特征,脸部和手部直方图相似,以此排除非肤色区域干扰。 $T_2$  提取形状特征,脸部和手部直方图存在差异,以此排除其他肤色区域干扰。最后将不同阈值下的归一化直方图连接,获得图像 OG\_LBP 直方图。以  $3 \times 3$  邻域为例,传统 LBP 特征维数为  $2^8 = 256$ ; OG\_LBP 特征直方图虽然采用双阈值,但计算维度为  $2^4 \times 2 = 32$ ,有效降低了计算复杂度。

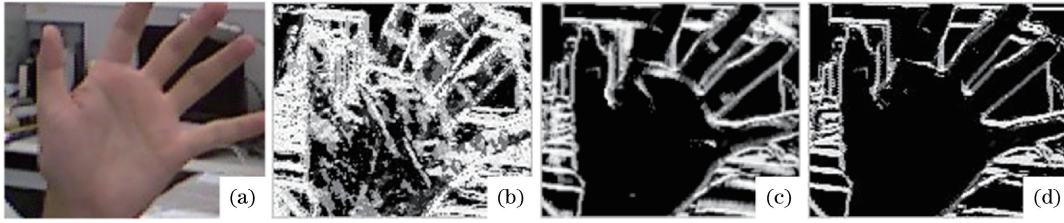


图 2 不同  $T$  值检测效果比较。(a) 原图; (b)  $T=3$ ; (c)  $T=10$ ; (d)  $T=20$

Fig. 2 Detection results of different  $T$  values. (a) Original image; (b)  $T=3$ ; (c)  $T=10$ ; (d)  $T=20$

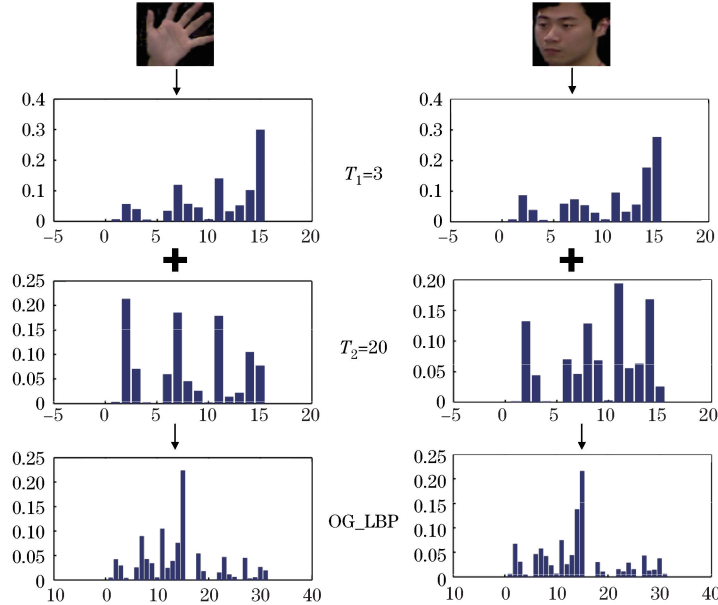


图 3 OG\_LBP 算子直方图描述

Fig. 3 Histogram description for OG\_LBP

#### 4 手部粒子滤波跟踪算法描述

针对手部姿态和尺度变化,结合红外深度信息,引入人工蜂群算法,利用当前观测信息,采用双阈值融合纹理和形状特征提取 OG\_LBP 特征,计算粒子观测似然函数,修改粒子权重  $\omega_i^t$ ,判断粒子是否需要重采样之后进入下一帧循环,不断循环实现手部运动自适应跟踪。每个粒子观测值的似然函数为

$$p(\mathbf{Z}_t | \mathbf{X}_t^i) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{D_i^2}{2\sigma^2}\right), \quad (20)$$

式中  $D_i$  为巴氏距离。

重采样以牺牲计算量来消除粒子匮乏现象,且重采样后有大量重复粒子,因此并不是每一时刻都需要进行重采样,当  $N_{\text{eff}}$  小于一定阈值时启动重采样算法<sup>[19]</sup>。

手部跟踪算法可以描述如下。

1) 初始化:融合深度和肤色特征获取目标模板,计算目标 OG\_LBP 特征;设定初始粒子集  $\{\mathbf{X}_0^i | i=1, 2, \dots, N\}$ ,所有粒子权重为  $1/N$ 。

2) 粒子传播及预测:对  $t$  时刻粒子按照(9)式系统模型,根据  $t-1$  帧计算的粒子半径进行传播;引入的人工蜂群算法步骤为:① 引领蜂根据每个蜜源的适应度  $F_i$ ,选取新蜜源对应目标尺度收缩系数  $g_i = 1/F_i$ ,结合深度信息确定新蜜源位置,计算新旧蜜源特征值相似度进行选择替换,并记录当前蜜源深度分布值  $d_i$ ; ② 跟随蜂对蜜源归一化适应度按照从大到小的顺序选取前  $N_{\text{eff}}$  个有效蜜源,以对应深度值  $d_i$  的最大连通区域作为新蜜源,计算新旧蜜源特征值相似度进行选择替换,其中  $N_{\text{eff}}$  计算参照(14)式; ③ 若位置  $i$  蜜源连续  $n$  帧都没有成为有效蜜源,则与该蜜源对应的引领蜂变为侦查蜂,按照(15)式产生新蜜源,其中适应度通过

计算对应的 OG\_LBP 特征直方图得到,计算每个粒子与目标的相似距离  $D_i^t$  构建似然函数  $p(\mathbf{Z}_t | \mathbf{X}_t)$ ,如(11)、(12)和(20)式所示。

3) 更新:邻域搜索后,根据(16)式更新粒子权重并进行归一化,得到  $t$  时刻粒子  $\{\mathbf{X}_t^i, \omega_t^i | i=1, 2, \dots, N\}$ ;根据(12)式计算  $t$  时刻相似系数  $\rho_t$ ,判断是否需要目标模板更新。

4) 重采样:考虑有效粒子数,若  $N_{\text{eff}} < N_{\text{threshold}}$ ,则将原来的带权样本  $\{\mathbf{X}_t^i, \omega_t^i | i=1, 2, \dots, N\}$  映射为等权样本  $\{\mathbf{X}_t^i, N^{-1} | i=1, 2, \dots, N\}$  进行重采样。

5) 输出:基于最大后验准则(MAP)确定  $t$  时刻跟踪位置。

## 5 实 验

实验利用 Kinect 摄像头连接到 Intel 主频 2.50 GHz,内存 8.00 GB 的 PC 机上,采用 VS2010 软件平台运行。实验中采用的粒子数为 50。

### 5.1 抗干扰实验

为验证算法的抗干扰性,设置脸部和手臂肤色区域进行干扰和遮挡情况下的跟踪实验,脸部和手臂的干扰跟踪结果分别如图 4 和图 5 所示,遮挡情况下的跟踪结果如图 6 所示。可以看出,存在人体肤色区域干扰时本文算法依然能实现稳定跟踪,遮挡部分手掌时手部跟踪仍稳定有效。



图 4 脸部干扰下跟踪结果

Fig. 4 Tracking results with the influence of face



图 5 手臂干扰跟踪结果

Fig. 5 Tracking results with the influence of arm



图 6 遮挡情况下跟踪结果

Fig. 6 Tracking results with occlusion

### 5.2 准确性实验

考虑手部运动的时变特性,录制 200 帧视频进行跟踪实验,以 5 帧为间隔,截取 20 帧(第 20 帧至第 115 帧)图像(图 7),其中 20~40 帧为手部相对摄像机所做的前后移动,即不同深度情况引起手部尺度变化的跟踪结果;45~60 帧为手部在相对摄像机平面,即同一深度范围内移动的跟踪结果,65~90 帧是手部姿态变化的跟踪结果,95~115 帧为手部进行变速运动的跟踪结果。可以看出,实验实现了上述各种情况下的准确跟踪。

比较 OG\_LBP 特征与传统 LBP 特征和 HOG 特征的跟踪精度并进行量化分析(图 8),其中 HOG 特征角度量化级别(bin)分别设为 36、18 和 9,即  $[0^\circ, 360^\circ]$  区间内每隔  $10^\circ$ 、 $20^\circ$  和  $40^\circ$  设置区间。实验中以手腕以上部分作为需要检测的手部区域,跟踪框随着手部姿态大小变化,准确度采用  $S_-/S_1$  和  $S_+/S_2$  进行判别,其中  $S_-$  为漏检的手部面积,  $S_1$  为跟踪框中实际手部面积,  $S_+$  为跟踪框中非手部区域面积,  $S_2$  为跟踪框面积,



图 7 手部尺度和姿态变化跟踪结果

Fig. 7 Tracking results in scale and posture changes

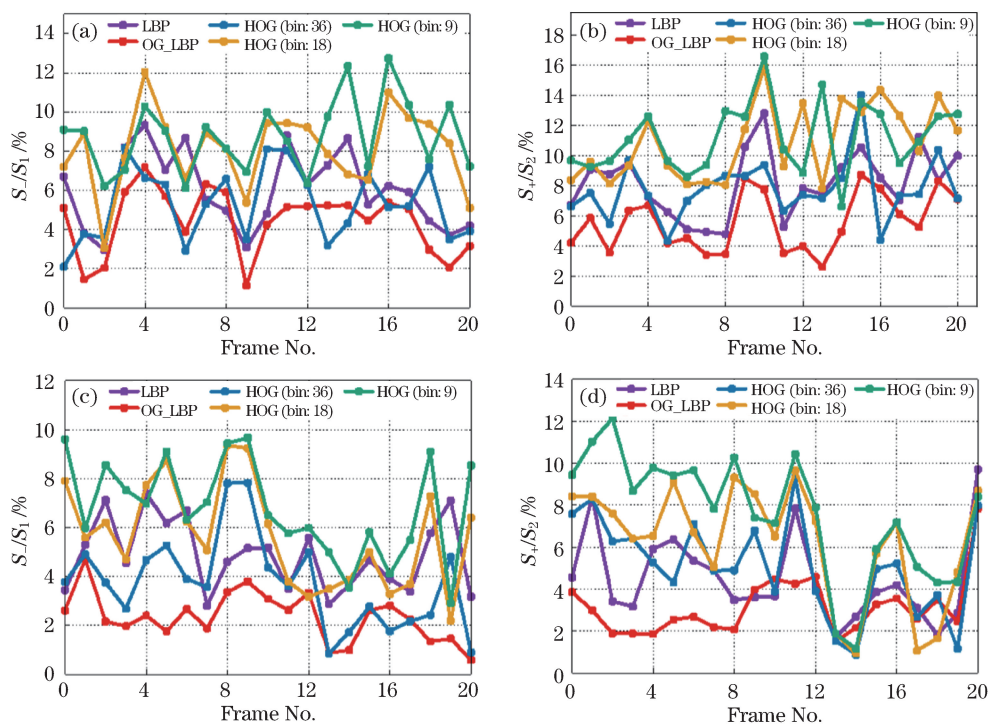


图 8 OG\_LBP、传统 LBP 和 HOG 特征跟踪误差比较。(a) 相同深度范围内运动漏检误差；  
(b) 相同深度范围内运动多检误差；(c) 不同深度范围内运动漏检误差；(d) 不同深度范围内运动多检误差

Fig. 8 Comparison of tracking results between OG\_LBP, traditional LBP and HOG features.

(a) Detection missing error at the same depth; (b) over-detection error at the same depth;

(c) detection missing error at different depths; (d) over-detection error at different depths

此处面积均指区域外接矩形面积,以像素计。从比较结果可以看出,与传统 LBP 和 HOG 特征相比,OG\_LBP 特征跟踪误差较低,具有较高精度及稳健性。

为验证 OG\_LBP 特征的稳健性,进行两方面实验:1) 在跟踪过程中调整手部平移速度,使手部跟踪产



生较大偏差乃至丢失;2) 前臂在同一深度范围内进行干扰。利用  $S_-/S_1$  和  $S_+/S_2$  进行跟踪精度比较的结果如图 9 所示,可以看出,OG\_LBP 特征在跟踪丢失后,能更快地实现目标跟踪;前臂和手部在同一深度下跟踪结果出现波动,误差增大,但 OG\_LBP 特征波动明显低于其他两种方法。

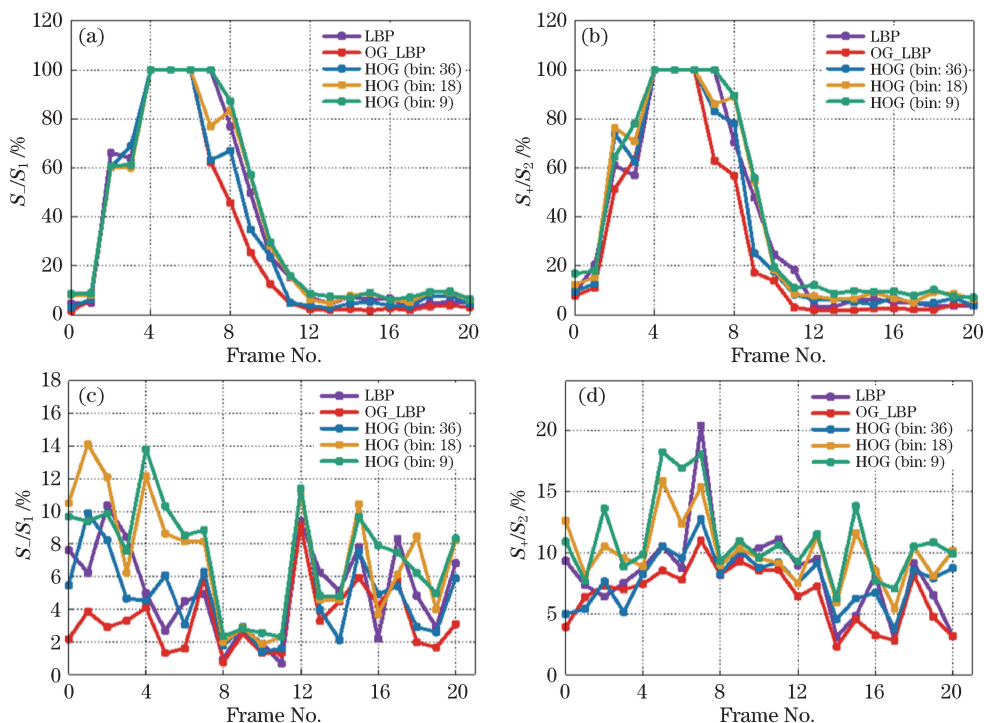


图 9 OG\_LBP 特征、传统 LBP 特征与 HOG 特征跟踪稳健性比较。(a) 变速运动漏检误差; (b) 变速运动多检误差;  
(c) 手臂干扰情况下漏检误差; (d) 手臂干扰情况下多检误差

Fig. 9 Comparison of tracing robustness between OG\_LBP, LBP and HOG features.

(a) Detection missing error in variable motion; (b) over-detection error in variable motion;  
(c) detection missing error with arm occlusion; (d) over-detection error with arm occlusion

## 6 结 论

手部跟踪在虚拟现实和自然人机交互领域具有重要的意义。提出 OG\_LBP 特征,建立准确的手部观测模型,降低特征维数,同时建立局部和全局的特征直方图描述,提高识别效率。在粒子滤波基本框架下,利用红外深度信息,通过人工蜂群算法将当前时刻的观测信息融合在滤波过程中,减轻了粒子退化现象。实验结果表明本文算法实现了复杂观测环境下手部尺度和姿态变化的自适应稳健跟踪,但在算法实时性上仍有提升空间。下一步将通过手部姿态的准确识别,提高手部交互在远程医疗和机器人深海控制等方面的应用价值。

## 参 考 文 献

- [1] Oikonomidis I, Kyriazis N, Argyros A. Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints[C]. IEEE International Conference on Computer Vision, 2011: 2088-2095.
- [2] Qian C, Sun X, Wei Y, *et al.* Realtime and robust hand tracking from depth[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1106-1113.
- [3] Sridhar S, Mueller F, Oulasvirta A, *et al.* Fast and robust hand tracking using detection-guided optimization[C]. Computer Vision & Pattern Recognition, 2015: 3213-3221.
- [4] Liu Y, Zhang L F, Zhang S J. A hand gesture recognition method based on multi-feature fusion and template matching [J]. Procedia Engineering, 2012, 29(4): 1678-1684.
- [5] Yang Bo, Song Xiaona, Feng Zhiqian, *et al.* Gesture recognition in complex background based on distribution features of hand[J]. Journal of Computer-Aided Design & Graphics, 2010, 22(10): 1841-1848.

- 杨 波, 宋晓娜, 冯志全, 等. 复杂背景下基于空间分布特征的手势识别算法[J]. 计算机辅助设计与图形学学报, 2010, 22(10): 1841-1848.
- [6] Priyal S P, Bora P K. A robust static hand gesture recognition system using geometry based normalizations and Krawtchouk moments[J]. Pattern Recognition, 2013, 46(8): 2202-2219.
- [7] Doliotis P, Stefan A, McMurrough C, *et al.* Comparing gesture recognition accuracy using color and depth information [C]. Proceedings of the 4<sup>th</sup> International Conference on PEervasive Technologies Related to Assistive Environments, 2011: 20.
- [8] Sridhar S, Oulasvirta A, Theobalt C. Interactive markerless articulated hand motion tracking using RGB and depth data [C]. IEEE International Conference on Computer Vision, 2013: 2456-2463.
- [9] Tzionas D, Srikantha A, Aponte P, *et al.* Capturing hand motion with an RGB-D sensor, fusing a generative model with salient points[C]. German Conference on Pattern Recognition, 2014: 277-289.
- [10] Xu C, Cheng L. Efficient hand pose estimation from a single depth image [C]. IEEE International Conference on Computer Vision, 2013: 3456-3462.
- [11] Park S, Yu S, Kim J, *et al.* 3D hand tracking using Kalman filter in depth space[J]. Journal on Advances in Signal Processing, 2012(1): 1-18.
- [12] Chen C P, Chen Y T, Lee P H, *et al.* Real-time hand tracking on depth images[C]. Visual Communications and Image Processing, 2011: 1-4.
- [13] Qi Y Q, Kazumasa S, Wu H Y, *et al.* EK-means tracker: A pixel-wise tracking algorithm using kinect[C]. Third Chinese Conference on Intelligent Visual Surveillance, 2011: 77-80.
- [14] Zhang Yueqiang, Su Ang, Liu Haibo, *et al.* Three dimensional rigid objects pose tracking and optimization based on multilevel line representation and M-estimation[J]. Acta Optica Sinica, 2015, 35(1): 0115003.  
张跃强, 苏 昂, 刘海波, 等. 基于多级直线表述和 M-估计的三维目标姿态跟踪优化算法[J]. 光学学报, 2015, 35(1): 0115003.
- [15] Yu Y, Zheng X. Particle filter with ant colony optimization for frequency offset estimation in OFDM systems with unknown noise distribution[J]. Signal Processing, 2011, 91(5): 1339-1342.
- [16] Park S, Hwang J P, Kim E, *et al.* A new evolutionary particle filter for the prevention of sample impoverishment[J]. IEEE Transactions on Evolutionary Computation, 2009, 13(4): 801-809.
- [17] Song Yu, Li Qingling, Kang Yifei, *et al.* SLAM with square-root cubature Rao-Blackwillised particle filter[J]. Acta Automatica Sinica. 2014, 40(2): 357-367.  
宋 宇, 李庆玲, 康轶非, 等. 平方根容积 Rao-Blackwillised 粒子滤波 SLAM 算法[J]. 自动化学报, 2014, 40(2): 357-367.
- [18] Tian Mengchu, Bo Yuming, Chen Zhimin, *et al.* Firefly algorithm intelligence optimized particle filter[J]. Acta Automatica Sinica, 2016, 42(1): 89-97.  
田梦楚, 薄煜明, 陈志敏, 等. 萤火虫算法智能优化粒子滤波[J]. 自动化学报, 2016, 42(1): 89-97.
- [19] Hu Shiqiang, Jing Zhongliang. The theory and application of particle filter[M]. Beijing: Science Press, 2010: 21-27.  
胡士强, 敬忠良. 粒子滤波原理及其应用[M]. 北京: 科学出版社, 2010: 21-27.
- [20] Jia Jingping, Xia Hong, Xie Ping. Visual tracking by scale space particle filter with multiple reference color histograms [J]. Systems Engineering and Electronics, 2015, 37(5): 1206-1212.  
贾静平, 夏 宏, 谢 萍. 基于尺度空间粒子滤波器的多参考直方图目标跟踪算法[J]. 系统工程与电子技术, 2015, 37(5): 1206-1212.
- [21] Karaboga D. An idea based on honey bee swarm for numerical optimization[R]. Technical Report TR06, Erciyes University, 2005.
- [22] Zhang Huanlong, Hu Shiqiang, Yang Guosheng. Video object tracking based on appearance models learning [J]. Journal of Computer Research and Development, 2015, 52(1): 177-190.  
张焕龙, 胡士强, 杨国胜. 基于外观模型学习的视频目标跟踪方法综述[J]. 计算研究与发展, 2015, 52(1): 177-190.
- [23] Li Peihua. A novel color based particle filter algorithm for object tracking[J]. Chinese Journal of Computers, 2009, 32(12): 2455-2463.  
李培华. 一种新颖的基于颜色信息的粒子滤波器跟踪算法[J]. 计算机学报, 2009, 32(12): 2454-2463.
- [24] Mittal A, Zisserman A, Torr P. Hand detection using multiple proposals[C]. British Machine Vision Conference, 2011, 75.1-75.11.
- [25] Suk H I, Sin B K, Lee S W. Hand gesture recognition based on dynamic Bayesian network framework [J]. Pattern

- Recognition, 2010, 43(9): 3059-3072.
- [26] Suksil T, Chalidabhongse T H. Hand detection and feature extraction for static Thai Sign Language recognition[C]. Proceedings of the 7<sup>th</sup> International Conference on Ubiquitous Information Management and Communication, 2013: 23.
- [27] Ong E J, Bowden R. A boosted classifier tree for hand shape detection[C]. 6<sup>th</sup> IEEE International Conference on Automatic Face and Gesture Recognition, 2004: 889-894.
- [28] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]. Computer Vision and Pattern Recognition, 2001: 511-518.
- [29] Niu Changfeng, Chen Dengfeng, Liu Yushu. Tacking object based on SIFT features and particle filter[J]. Robot, 2010, 32(2): 241-247.  
牛长锋, 陈登峰, 刘玉树. 基于 SIFT 特征和粒子滤波的目标跟踪方法[J]. 机器人, 2010, 32(2): 241-247.
- [30] Zheng Yongbin, Huang Xinsheng, Feng Songjiang. An image matching algorithm based on combination of SIFT and the rotation invariant LBP[J]. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(2): 286-292.  
郑永斌, 黄新生, 丰松江. SIFT 和旋转不变 LBP 相结合的图像匹配算法[J]. 计算机辅助设计与图形学学报, 2010, 22(2): 286-292.
- [31] Rathi Y, Vaswani N, Tannenbaum A, *et al.* Tracking deforming objects using particle filtering for geometric active contours[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(8): 1470-1475.
- [32] Chang C C, Chen J J, Tai W K, *et al.* New approach for static gesture recognition[J]. Journal of Information Science & Engineering, 2006, 22(5): 1047-1057.
- [33] Liu Y, Gan Z, Sun Y. Static hand gesture recognition and its application based on support vector machines[C]. ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008: 517-521.
- [34] Zhang D S. Image retrieval based on shape[D]. Australia: Monash University, 2002.
- [35] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. IEEEConference on Computer Vision & Pattern Recognition, 2005: 886-893.
- [36] Papageorgiou C, Poggio T. A trainable system for object detection[J]. International Journal of Computer Vision, 2000, 38(1): 15-33.
- [37] Zhou Lai. Research of visual interaction for virtual reality flight training [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012.  
周 来. 面向虚拟现实飞行模拟训练的视觉手交互技术研究[D]. 南京: 南京航空航天大学, 2012.
- [38] Sun Jin, Gu Hongbin, Hou Jianbo. Airplane wingtip tracking algorithm based on oriented gradient local binary pattern [J]. Acta Aeronautica et Astronautica Sinica, 2011, 32(11): 2062-2072.  
孙 瑾, 顾宏斌, 侯建波. 基于梯度方向二值模式特征的飞机翼尖跟踪技术[J]. 航空学报, 2011, 32(11): 2062-2072.
- [39] Sun Jin, Gu Hongbin, Zheng Jiping. A gradient direction based moving object detection[J]. Journal of Image and Graphics, 2008, 13(3): 571-579.  
孙 瑾, 顾宏斌, 郑吉平. 一种基于梯度方向信息的运动目标检测算法[J]. 中国图象图形学报, 2008, 13(3): 571-579.
- [40] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7): 971-987.