

# 一种分层 Mean Shift 目标跟踪算法

许海霞<sup>1,2</sup> 王耀南<sup>1</sup> 袁小芳<sup>1</sup> 周维<sup>2</sup> 朱江<sup>1</sup>

**摘要** 针对经典 Mean shift (MS) 目标跟踪算法的颜色特征鲁棒差、匹配迭代复杂的缺点, 提出一种分层 Mean shift (Hierarchical mean shift, HMS) 目标跟踪算法. 首先通过 MS 迭代将目标区域特征空间的数据点聚类于模式点, 使得以简洁的方式描述前景跟踪目标, 建立目标模型与目标候选模型的聚类模式点描述, 进行聚类块匹配. 然后, 导出聚类块模式点匹配下的相似度量函数, 进行像素点匹配, 结合邻域一致性, 计算像素平移量, 分层估计序列帧中跟踪目标质心模式点的位置, 并给出 HMS 匹配迭代跟踪算法. 实验结果表明, 与其他两种 MS 跟踪算法相比, HMS 既能提高序列帧跟踪目标表达与匹配的鲁棒性, 又无需匹配所有数据点, 算法简洁且有效可行.

**关键词** 目标跟踪, 分层 Mean shift, 聚类模式点, 匹配  
**中图分类号** TP391

## A Hierarchical Mean Shift Algorithm for Object Tracking

XU Hai-Xia<sup>1,2</sup> WANG Yao-Nan<sup>1</sup> YUAN Xiao-Fang<sup>1</sup> ZHOU Wei<sup>2</sup> ZHU Jiang<sup>1</sup>

**Abstract** We propose a hierarchical mean shift (HMS) algorithm for object tracking. Firstly, cluster modal points are obtained by mean-shift iteratively processing all the data points in the region so that they can represent foreground object in a succinct manner. The target model and the target candidate model are described by the cluster modal points, and match processes of clustered blocks are performed. Then, on the basis of cluster blocks match, similarity measure function is set up to match between target model and target candidate at pixel level. And the pixel shift vector of target is calculated with the introduction of the neighborhood consistency concept. So, the centroid of tracking object is got layer by layer in the consecutive frames, and the HMS match iteration for object tracking is presented. Experimental comparisons with other two MS algorithms demonstrate the validity and performance of the proposed algorithm.

**Key words** Object tracking, hierarchical mean shift (HMS), cluster modal point, match

运动目标跟踪是计算机视觉、人机交互等的研究热点之一. 常用的跟踪算法有 Kalman、Particle、Mean shift (MS) 及其各种变形、改进算法. MS track 最初是 Comaniciu<sup>[1]</sup> 于 2000 年将核密度估计<sup>[2]</sup> 用于跟踪而提出的一种高效的实时跟踪算法, 得到了广泛的研究与应用. 它是基于目标颜色特征的区域匹配跟踪, 通过对目标模型区与后续帧中候选区的颜色直方图进行比较, 用 Bhattacharyya 系数、Kullback Leibler 距离等方式度量其相似度, 为以后的 MS 跟踪算法奠定了基础. 进一步的改进算法从驱动 MS 迭代方式来看可以归为以下几类: 1)

建立目标模型直方图驱动的 MS 跟踪算法, 通过不断完善目标模型直方图的表达来提高跟踪性能. 如基于核的跟踪<sup>[3]</sup>, 在分析前景的同时, 加入对背景模型的考虑; 空间直方图模型<sup>[4]</sup>, 同时考虑了目标区的彩色信息及像素间的空间位置关系, 空间信息的引入使得获取目标颜色信息范围更合理, 体现了局部性的考虑; 颜色聚类的高斯分布直方图<sup>[5-6]</sup>, 引入聚类, 使得直方图量化等级有了依据, 且合理、准确. 2) 匹配 (配准) 驱动的 MS 跟踪算法, 不计算颜色直方图, 而直接度量两者的相似度, 如基于相似度量的 MS 跟踪算法<sup>[7-8]</sup>. 该类 MS 算法给定抽样点、核函数、目标的空间位置与颜色联合概率密度函数 (Probability density function, PDF), 定义目标区与候选区的 PDF 的相似度量函数, 逐点搜索比较, 直到找到最佳匹配区, 其优点是区别能力强. 3) 在概率表达及贝叶斯理论构架下建立目标模型驱动 MS 迭代, 如联合特征下目标模型的概率表达迭代<sup>[9]</sup>, 随机抽取粒子迭代<sup>[10]</sup>, MS 迭代与粒子滤波器相结合<sup>[11-12]</sup>, 用粒子群优化 (Particle swarm optimization, PSO) 算法减小 MS 迭代的样本数据点<sup>[13]</sup>, 均值平移置信传播 (Mean shift belief propagation, MSBP) 算法<sup>[14]</sup> 把 MS 放在置信传播 (Belief propagation, BP) 框架下, BP 生成样本,

收稿日期 2007-11-06 收修改稿日期 2008-06-26  
Received November 6, 2007; in revised form June 26, 2008  
国家自然科学基金重点项目 (60835004), 国家高技术研究发展计划 (863 计划) (2007AA04Z244, 2008AA04Z214), 湖南省教育厅项目 (07C073) 资助  
Supported by National Natural Science Foundation of Key Program of China (60835004), National High Technology Research and Development Program of China (863 Program) (2007AA04Z244, 2008AA04Z214), and Foundation of Hunan Educational Committee (07C073)  
1. 湖南大学电气与信息工程学院 长沙 410082 2. 湘潭大学信息工程学院 湘潭 411105  
1. School of Electrical and Information Engineering, Hunan University, Changsha 410082 2. School of Information Engineering, Xiangtan University, Xiangtan 411105  
DOI: 10.3724/SP.J.1004.2009.00401

MS 迭代加权样本. 一种新的核密度估计法<sup>[15]</sup>用于视觉跟踪, 将核密度二次计算演变成线性计算. 上述算法大多是通过减小参与迭代的样本数据量来提高计算效率.

MS 跟踪是基于目标颜色特征的区域匹配跟踪, 区域颜色的分布是随机的、不连续的点状或小区域片. 传统 MS 跟踪算法存在以下不足: 直方图驱动时, 直方图的量化是均匀的、盲目的, 造成没必要的存储和计算开销<sup>[6]</sup>. 而匹配(配准)驱动时, 1) 序列帧目标区域图像的像素颜色随时间、光照等影响会发生变化, 直接匹配鲁棒性差; 2) 区域中所有数据参与迭代运算, 需计算所有数据点的核函数与一阶矩, 运算量大; 3) 匹配过程中, 不能充分利用目标区域颜色信息的引导作用去减小搜索像素空间范围, 造成没必要的匹配比较与计算量.

针对以上问题, 本文提出一种分层 MS 跟踪算法, 将目标模型区域与目标候选区域的整个匹配跟踪过程分为两层进行. 第一层, MS 迭代算法对目标区域特征量聚类, 聚类模式点估计, 通过聚类分割获得稳定的块, 进行聚类块匹配. 这样, 聚类块可以体现区域的整体特征, 能够克服像素受干扰而变化所带来的影响, 而且聚类模式点以简洁的方式表达目标的分布特征, 提高了匹配的鲁棒性与有效性. 第二层, 在匹配块内计算像素偏移量, 进行像素匹配, 取得跟踪目标质心模式点的估计. 区域聚类于块状、块模式点聚于跟踪质心, 即先块再像素点分层匹配迭代.

## 1 分层 MS 跟踪算法 (HMS)

### 1.1 目标区域颜色与空间联合聚类分析

在基于颜色特征的区域匹配目标跟踪算法中, 区域内像素颜色随时间受光照、噪声等影响会发生细微的变化, 采用聚类可以提高像素颜色的抗干扰性, 而且聚类模式点能以简洁的方式表达目标的分布特征. 通常情况下, 非参数 MS 聚类迭代算法及应用<sup>[16-17]</sup>, 都是对一幅给定的静态图像进行聚类、区域分割, 而在动态图像序列跟踪帧中对目标区域聚类进行分析, 目标在运动中除了平移, 还会有尺度变化. 因此, 聚类带宽须能够随目标大小而变化、更新, 即对序列帧需进行可变带宽聚类. 带宽的更新有两种方式<sup>[18]</sup>: 窗口内所有数据点采用相同尺度带宽核与不同尺度带宽核更新. 由于目标区域内前景目标的顏色信息不会发生尺度变化, 而位置信息会发生尺度变化, 位置数据点是相邻的, 则发生近似相同尺度变化, 所以, 所有数据点采用相同尺度带宽核的更新方式. 在文献 [17] 的基础上, 本文给出了适合于动态图像序列帧的 MS 变带宽聚类如式 (1), 式中

对目标区域数据点集  $\{\mathbf{x}_n\}_{n=1}^N \in \mathbf{R}^5$  进行自动聚类, 收敛于模式点  $\mathbf{x}$ .

$$f_{\sigma,s}(\mathbf{x}) = \frac{\sum_{j=1}^m \sum_{\mathbf{x}_n \in \text{set}_j} \omega \left( \left\| \frac{\mathbf{x}_n - \mathbf{x}}{\sigma, s} \right\|^2 \right) \mathbf{x}_n}{\sum_{j=1}^m \sum_{\mathbf{x}_n \in \text{set}_j} \omega \left( \left\| \frac{\mathbf{x}_n - \mathbf{x}}{\sigma, s} \right\|^2 \right)} \quad (1)$$

式 (1) 中,  $\mathbf{x}_n = \{i, j, r, g, b\}$  是颜色分量  $\{r, g, b\}$  与空间位置分量  $\{i, j\}$  的 5D 信息数据点;  $N$  个数据点预划分在  $m$  个子集  $\text{set}_j$  ( $j = 1, \dots, m$ ) 内,  $m$  在聚类过程中自动更新;  $\omega(\cdot)$  为带宽为  $\sigma$  的高斯核函数, 聚类带宽  $\sigma$  控制聚类程度和聚类数;  $s$  是更新聚类带宽的尺度因子, 由目标尺度变化大小来确定.

目标模型区域的颜色与空间 5 维联合信息按式 (1) 进行聚类分析, 形成  $d$  个聚类块, 每一聚类块的模式点  $\mathbf{C}_k = \{\mathbf{C}\mathbf{x}_k, \mathbf{C}\mathbf{u}_k\}$ ,  $k = 1, \dots, d$ , 其中  $\mathbf{C}\mathbf{x}_k$  为聚类空间位置,  $\mathbf{C}\mathbf{u}_k$  为聚类颜色, 每类的像素数为  $n_k$ . 颜色与空间位置的联合聚类分析, 使得目标区域特征表达鲁棒性强. 图 1 显示了两种不同的 RGB 彩色目标模型区图像: 颜色分布变化大的球员模型区和颜色分布变化细腻的人脸模型区, 及经过 5 维空间、颜色联合特征聚类后的自适应聚类分割块. 根据目标颜色分布变化情况, 选择不同的聚类带宽, 实现不同程度的聚类. 球员各部分颜色分布变化大, 设聚类带宽  $\sigma$  为 38, 自适应聚为 10 类. 而相对于球员整体目标来讲, 人脸是局部区域目标, 其变化细腻, 聚类带宽设为 14, 聚为 84 类, 视觉上与原始图

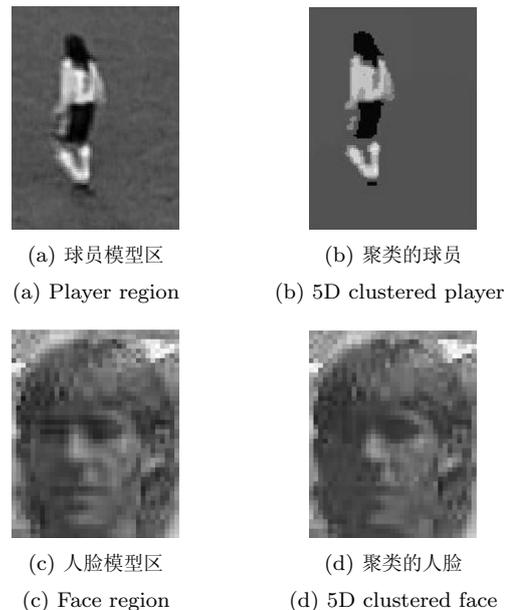


图 1 两种不同的颜色分布目标模型区及 5D 聚类分析  
Fig. 1 Two object model regions with distinct color distributions and their 5D clustering analysis

像差别不大.

## 1.2 目标模型聚类模式点描述

MS 跟踪迭代算法是以目标区前景 ( $fg$ ) 颜色信息引导及驱动得到跟踪平移量, 所以目标模型的建立需去除掉背景. 背景 ( $bg$ ) 像素的混入对跟踪效果影响非常大, 前景与背景能否准确分离是影响目标模型及 MS 跟踪性能好坏的关键要素之一. 给定的目标区域经过颜色与空间的联合特征聚类分析, 形成聚类分割块, 并以聚类模式点  $\mathbf{C}_k$  ( $k = 1, \dots, d$ ) 描述. 下面首先判别每一聚类块属于前景还是背景, 然后再提取出前景类, 建立目标模型描述.

图像区域矩特征具有不变性, 目标质心是一个有效的矩特征, 计算目标质心可进行前景与背景的区别判别, 目标质心的计算过程如下:

1) 初始化质心为目标区域中心  $\mathbf{x}_c = \mathbf{x}^*$  ( $\mathbf{x}^*$  为目标模型区中心);

2) 目标区域前景聚类块的组成部分是有序连接的, 前景类、背景类与目标区域质心的距离远近不同, 利用目标区域内各聚类的模式点  $\mathbf{C}_k$  与质心的距离, 区别判别  $fg$  和  $bg$ , 即

$$\begin{cases} \mathbf{C}_k \in fg, & \text{若 } |\mathbf{C}_k - \mathbf{x}_c| < T_d \\ \mathbf{C}_k \in bg, & \text{否则} \end{cases} \quad (2)$$

聚类模式点位置距离质心  $\mathbf{x}_c$  较远者标记为  $bg$ ,  $T_d$  为最大距离阈值.

3) 由于背景像素分布是随机的, 可能会出现背景类模式点位置距离目标质心较近的情况, 因此仅用距离信息可能会存在误判. 这里, 在步骤 2) 的基础上, 进一步对被判为  $fg$  的聚类块及其模式点  $\mathbf{C}_k$  进行分析, 对目标区域聚类块计算对数似然函数比  $L(k)$ <sup>[19]</sup>, 通过比值  $L$  的正负判别  $fg$  和  $bg$ , 即

$$\begin{cases} \mathbf{C}_k \in fg, & \text{若 } L(k) > 0 \\ \mathbf{C}_k \in bg, & \text{若 } L(k) \leq 0 \end{cases} \quad (3)$$

4) 去掉背景类, 再采用聚类模式点  $\mathbf{C}_k$  定义质心

$$\mathbf{x}_c = \frac{\sum_{\mathbf{C}_k \in fg} \mathbf{C}_k \cdot H(\mathbf{C}_k)}{\sum_{\mathbf{C}_k \in fg} H(\mathbf{C}_k)} \quad (4)$$

其中  $H(\mathbf{C}_k)$  是该类在 HSV 色彩空间的  $H$  分量值, 作为计算质心的图像属性.

5) 重复步骤 2) ~ 4) 求  $\mathbf{x}_c$ , 直到质心的位置趋于不变. 同时, 也得到越来越准确的前景, 目标区前景与质心相互制约、相互影响.

因此, 根据图像矩特征去分离前景与背景, 可以将目标区域  $d$  类分解为属于前景的有  $df$  类、属于背

景的有  $db$  类 ( $d = df + db$ ). 这样, 目标模型的聚类模式点描述为:  $\mathbf{S}_k = \{\mathbf{C}_k, \mathbf{C}_k\}$ ,  $k = 1, \dots, df$ . 图 2 显示了图 1 目标区在联合空间多维信息聚类后前景与背景分离, 提取出前景目标建立的聚类模式点描述的目标模型. 球员前景目标呈区域片状, 而人脸聚类程度低, 几乎还是呈点状.

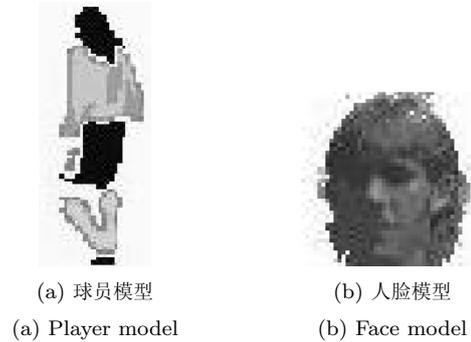


图 2 聚类模式点描述的目标模型  
Fig. 2 Object model in the description of cluster modal points

## 1.3 聚类块与像素点分层匹配的相似度量函数

给定目标模型区域抽样点集合  $\mathbf{S}_x$  和核函数, 利用核密度估计法对目标区的空间位置与颜色联合概率密度函数 (PDF) 估计, 在前景颜色信息的引导下进行目标跟踪. 假定  $\mathbf{S}_x = \{\mathbf{x}_i, \mathbf{u}_i\}_{i=1}^N$  是源于 PDF 为  $P(\mathbf{x}, \mathbf{u})$  分布的抽样点集合,  $\mathbf{x}_i$  为像素二维坐标点,  $\mathbf{u}_i = u(\mathbf{x}_i)$  为在位置  $\mathbf{x}_i$  处  $\{r, g, b\}$  三维彩色信息, 则  $P(\mathbf{x}, \mathbf{u})$  用多变量联合核密度函数估计<sup>[20]</sup> 为

$$\hat{P}(\mathbf{x}, \mathbf{u}) = \frac{1}{N} \sum_{i=1}^N \omega \left( \left\| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right\|^2 \right) k \left( \left\| \frac{\mathbf{u} - \mathbf{u}_i}{h} \right\|^2 \right) \quad (5)$$

其中  $\omega(\cdot)$ ,  $k(\cdot)$  分别是带宽为  $\sigma$ ,  $h$  的高斯核函数, 记作  $\omega_\sigma(\cdot)$ ,  $k_h(\cdot)$ .

经过以上聚类分割分析, 确定了目标模型区域的前景类和背景类, 建立了目标模型的聚类模式点描述, 则目标模型区域前景的联合概率密度函数估计可表示为

$$\hat{P}(\mathbf{x}, \mathbf{u}) = \frac{1}{N_f} \sum_{k=1}^{df} \sum_{\mathbf{x}_i \in \mathbf{C}_k} \omega \left( \left\| \frac{\mathbf{x} - \mathbf{x}_i}{\sigma} \right\|^2 \right) k \left( \left\| \frac{\mathbf{u} - \mathbf{u}_i}{h} \right\|^2 \right) \quad (6)$$

其中  $N_f = \sum_{k=1}^{df} n_k$  是前景类的总像素数.

相应的后续跟踪帧中, 候选目标区域的抽样点集合  $\mathbf{S}_y = \{\mathbf{y}_j, \mathbf{v}_j\}_{j=1}^M$ , 聚类为  $d'$  类, 聚类中心  $\mathbf{C}_l = \{\mathbf{C}_l, \mathbf{C}_l\}$ ,  $l = 1, \dots, d'$ , 前景有  $df'$  类, 背景有  $db'$  类, 每类的像素数为  $n'_l$ . 目标候选模型的聚类模式

点描述为:  $Sy_l = \{Cy_l, Cv_l\}, l = 1, \dots, df'$ .

通过匹配法在后续帧的候选区域找出与目标模型区域前景相似的前景, 本文以两者前景颜色为引导信息, 首先在两者间颜色相近的聚类块之间进行搜索匹配, 即对目标模型区域  $df$  个前景聚类块与目标候选模型区域  $df'$  个前景聚类块进行匹配.

采取在 HSV 空间下, 通过两者颜色的差别来衡量其是否相匹配, 按以下两步为目标模型描述的每一类  $C_k$  找到与候选目标模型描述类  $C_l$  中最接近的相匹配的类.

1) 将两者的聚类中心以颜色分量  $Cu_k, Cv_l$  的  $H$  分量值从大到小进行排序;

2) 对于每一个值  $k$ , 计算  $Cu_k$  与  $Cv_l$  的  $H$  分量的差别, 取  $\arg \min_l |Cv_l - Cu_k|$  且满足条件  $|Cv_l - Cu_k| < T_c$ , 得到第  $l$  类. 若  $l$  存在, 即为与  $C_k$  最相匹配的是  $C_l$ ; 若  $l$  不存在, 即候选区中没有与其相匹配的类. 其中  $1 < k < df, 1 < l < df', T_c$  是颜色相匹配所容许的误差范围阈值.

若每类模式点  $C_k$  都找到对应的类模式点  $C_l$ , 设  $l = k = 1, \dots, d$ , 候选模型区与目标模型区的聚类块匹配对如图 3(a). 若找不到对应类或对应的数量太少, 则重新选择目标. 相比而言, 原 MS 匹配法需逐点搜索 (如图 3(b)) 所有像素点对进行匹配.

在目标模型与目标候选模型的聚类模式点匹配后, 即在前景聚类块匹配基础上, 进一步再在匹配块内进行像素匹配, 计算目标的运动平移量, 定义目标模型区与候选模型区的相似度量函数为式 (7).

式 (7) 中,  $\delta(\cdot)$  是 delta 函数, 当候选目标模型聚类模式点  $C_l$  与目标模型聚类模式点  $C_k$  相匹配时为 1, 否则为 0;  $N_p$  为相匹配的总像素数;  $n'_l, n_k$  分别为第  $l', k$  类的像素数. 即, 在对应类块内进行像素点匹配搜索, 而在非对应类区不进行任何操作,

只有  $d$  对对应类作累加.

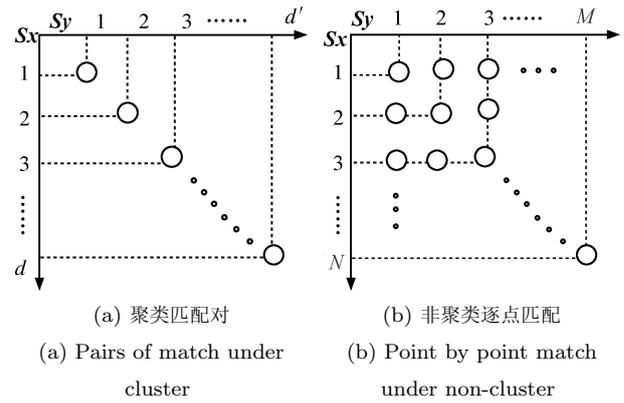


图 3 聚类匹配与非聚类匹配

Fig. 3 Cluster and non-cluster matches

#### 1.4 分层 MS 迭代跟踪算法

在目标模型描述的前景颜色信息的引导下, 先聚类块再像素点逐级分层确定两者相接近的待搜索像素空间. 设目标模型区域中心为  $x^*$ , 跟踪帧候选区中心为  $y$ , 若只考虑其平移运动, 则式 (7) 可表示为式 (8) 的形式.

通过最大化目标模型与候选模型的相似度量函数  $J$ , 确定跟踪目标在候选区的位置. 采用梯度下降 MS 迭代算法寻找式 (8) 的极值. 关于候选区位置中心向量  $y$  的梯度定义为

$$\nabla L(\mathbf{y}) = -\frac{\nabla J(\mathbf{y})}{J(\mathbf{y})} \quad (9)$$

其中  $\nabla J(\mathbf{y})$  的定义见式 (10). 式 (10) 中,  $\Delta \mathbf{y}_j = (\mathbf{y}_j - \mathbf{y}), \Delta \mathbf{x}_i = (\mathbf{x}_i - \mathbf{x}^*), \omega'(\cdot)$  表示求导. 则进一步得到式 (11).

在式 (11) 迭代确定目标在候选区位置  $y$  的计算

$$J(Sx_k, Sy_l) = \frac{1}{N_p} \sum_{l=1}^{df'} \sum_{k=1}^{df} \left[ \delta(C_l, C_k) \left( \sum_{\mathbf{y}_j \in C_l} \sum_{\mathbf{x}_i \in C_k} \omega_\sigma(\mathbf{y}_j - \mathbf{x}_i) k_h(\mathbf{v}_j - \mathbf{u}_i) \right) \right] \quad (7)$$

$$J(Sx_k, Sy_l) = \frac{1}{N_p} \sum_{l=1}^{df'} \sum_{k=1}^{df} \left[ \delta(C_l, C_k) \left( \sum_{\mathbf{y}_j \in C_l} \sum_{\mathbf{x}_i \in C_k} \omega_\sigma((\mathbf{y}_j - \mathbf{y}) - (\mathbf{x}_i - \mathbf{x}^*)) k_h(\mathbf{v}_j - \mathbf{u}_i) \right) \right] \quad (8)$$

$$\nabla J(\mathbf{y}) = \sum_{l=1}^{df'} \sum_{k=1}^{df} \left[ \delta(C_l, C_k) \left( \sum_{\mathbf{y}_j \in C_l} \sum_{\mathbf{x}_i \in C_k} (\Delta \mathbf{x}_i - \Delta \mathbf{y}_j) \omega'_\sigma(\Delta \mathbf{y}_j - \Delta \mathbf{x}_i) k_h(\mathbf{v}_j - \mathbf{u}_i) \right) \right] \quad (10)$$

$$ms(\mathbf{y}) = \frac{\sum_{l=1}^{df'} \sum_{k=1}^{df} \left[ \delta(C_l, C_k) \left( \sum_{\mathbf{y}_j \in C_l} \sum_{\mathbf{x}_i \in C_k} (\mathbf{y}_j - \mathbf{x}_i) \omega_\sigma(\Delta \mathbf{y}_j - \Delta \mathbf{x}_i) k_h(\mathbf{v}_j - \mathbf{u}_i) \right) \right]}{\sum_{l=1}^{df'} \sum_{k=1}^{df} \left[ \delta(C_l, C_k) \left( \sum_{\mathbf{y}_j \in C_l} \sum_{\mathbf{x}_i \in C_k} \omega_\sigma(\Delta \mathbf{y}_j - \Delta \mathbf{x}_i) k_h(\mathbf{v}_j - \mathbf{u}_i) \right) \right]} + \mathbf{x}^* \quad (11)$$

过程中, 相匹配的前景类块内, 采用 FGF 算法<sup>[7]</sup> 计算高斯核, 而本文给出结合邻域一致性<sup>[17]</sup> 分层近似计算, 效率更高, 过程如下:

首先, 第一层匹配中, MS 聚类过程有  $\mathbf{C}\mathbf{x}_k = (1/n_k) \sum_{\mathbf{x}_i \in \mathbf{C}_k} \mathbf{x}_i$ ,  $\mathbf{C}\mathbf{y}_l = (1/n'_l) \sum_{\mathbf{y}_j \in \mathbf{C}_l} \mathbf{y}_j$ . 因此, 由邻域一致性, 每对相匹配的聚类块有

$$\begin{aligned} & \sum_{\mathbf{y}_j \in \mathbf{C}_l} \sum_{\mathbf{x}_i \in \mathbf{C}_k} \mathbf{y}_j \omega_\sigma((\mathbf{y}_j - \mathbf{y}) - (\mathbf{x}_i - \mathbf{x}^*)) k_h(\mathbf{v}_j - \mathbf{u}_i) \approx \\ & \sum_{\mathbf{x}_i \in \mathbf{C}_k} n'_l \cdot \mathbf{C}\mathbf{y}_l \cdot \omega_\sigma((\mathbf{C}\mathbf{y}_l - \mathbf{y}) - (\mathbf{x}_i - \mathbf{x}^*)) k_h(\mathbf{v}_{\mathbf{C}\mathbf{y}_l} - \\ & \mathbf{u}_i) \approx n_k \cdot n'_l \cdot \mathbf{C}\mathbf{y}_l \cdot \omega_\sigma((\mathbf{C}\mathbf{y}_l - \mathbf{y}) - (\mathbf{C}\mathbf{x}_k - \mathbf{x}^*)) \times \\ & k_h(\mathbf{v}_{\mathbf{C}\mathbf{y}_l} - \mathbf{u}_{\mathbf{C}\mathbf{x}_k}) \approx n_k \cdot n'_l \cdot \mathbf{C}\mathbf{y}_l \cdot \omega_\sigma(\cdot) k_h(\cdot) \quad (12) \end{aligned}$$

$$\begin{aligned} & \sum_{\mathbf{y}_j \in \mathbf{C}_l} \sum_{\mathbf{x}_i \in \mathbf{C}_k} \mathbf{x}_i \omega_\sigma((\mathbf{y}_j - \mathbf{y}) - (\mathbf{x}_i - \mathbf{x}^*)) k_h(\mathbf{v}_j - \mathbf{u}_i) \approx \\ & n'_l \cdot n_k \cdot \mathbf{C}\mathbf{x}_k \cdot \omega_\sigma(\cdot) k_h(\cdot) \quad (13) \end{aligned}$$

所以, 每一匹配对求和  $\approx n_k \cdot n'_l \cdot \omega_\sigma(\cdot) k_h(\cdot) (\mathbf{C}\mathbf{y}_l - \mathbf{C}\mathbf{x}_k)$ , 则式 (11) 最终简化为 HMS 跟踪迭代式 (14).

在图像序列帧跟踪过程中, 如果目标伴随有尺度变化, 则根据文献 [7, 21] 重新定义相似度量函数为

$$\begin{aligned} J(\mathbf{S}\mathbf{x}_k, \mathbf{S}\mathbf{y}_l) &= \frac{1}{N_p} \sum_{l=1}^{df'} \sum_{k=1}^{df} \left[ \delta(\mathbf{C}_l, \mathbf{C}_k) \times \right. \\ & \left. \left( \sum_{\mathbf{y}_j \in \mathbf{C}_l} \sum_{\mathbf{x}_i \in \mathbf{C}_k} \omega \left( \left\| \frac{\Delta \mathbf{y}_j}{\sqrt{s} \sigma} - \left( \frac{\sqrt{s} \Delta \mathbf{x}_i}{\sigma} \right) \right\|^2 \right) k_h(\mathbf{v}_j - \mathbf{u}_i) \right) \right] \quad (15) \end{aligned}$$

其中,  $s$  是图像序列帧之间目标尺度变化因子.

相应地, 通过函数  $J$  对目标位置  $\mathbf{y}$  及尺度  $s$  求归一化梯度, 可得  $\mathbf{y}$  与  $s$  更新迭代式 (16) 和 (17), 其中,  $\omega_{\sigma,s}(\cdot) = \omega \left( \left\| \frac{\mathbf{C}\mathbf{x}_k - \mathbf{x}^*}{\sigma} - \frac{\mathbf{C}\mathbf{y}_l - \mathbf{y}}{s} \right\|^2 \right)$ .

HMS 跟踪算法的整个原理过程如图 4 (见下页) 所示. 首先手工选定待跟踪的目标模型区, 对目标区的颜色与位置 5D 数据联合聚类分析及去背景, 建立目标模型  $\mathbf{S}\mathbf{x}_k$ ; 然后在后续帧中分层匹配搜索目标, 建立目标候选模型  $\mathbf{S}\mathbf{y}_l$ ; 目标模型与候选模型聚类块模式点匹配后, 再进行像素点匹配. 候选区初始位置中心为  $\mathbf{y}_0$ , 尺度为  $s_0$ , 执行式 (16) 和 (17) 迭代更新当前帧目标候选区中心  $\mathbf{y}$ , 尺度  $s$ , 直至其收敛.

调节聚类带宽  $\sigma$ , 可实现不同程度的聚类, 当带宽非常小, 以至于每类的大小  $n_k = 1$  时, 则转化为 Yang<sup>[7]</sup> 的 MS 迭代跟踪.

## 2 算法的复杂性分析

匹配驱动的 MS 跟踪, 不计算直方图, 直接度量目标模型区与候选区的相似性, 如文献 [7] 没有利用颜色的引导性, 在搜索过程中逐点比较目标模型区  $\mathbf{x}_i$  与候选区  $\mathbf{y}_j$ , 匹配计算量为  $(M \times N)^2$ , 并且每搜索一步还需计算所有数据点的核函数及计算一阶矩. 与之相比, 本文分层计算, 先聚类块匹配, 再像素匹配, 结合邻域一致性, 最后只需计算  $d$  个聚类模式点的核函数及数乘运算即可计算出目标运动的运动量. 本文算法的计算代价: 第一层, 设聚类块匹配的计算量为  $N_1$ , 其中包括 5 维信息聚类过程, 设其计算量为  $N_c$ ,  $N_c$  的大小主要取决于聚类带宽  $\sigma$ <sup>[16-17]</sup>, 其次与子集数  $m$  及数据点数量  $N$  有关.  $\sigma$  越小, 计算

$$ms(\mathbf{y}) = \frac{\sum_{l=1}^{df'} \sum_{k=1}^{df} \delta(\mathbf{C}_l, \mathbf{C}_k) n_k \cdot n'_l \cdot \omega_\sigma(\cdot) k_h(\cdot) (\mathbf{C}\mathbf{y}_l - \mathbf{C}\mathbf{x}_k)}{\sum_{l=1}^{df'} \sum_{k=1}^{df} \delta(\mathbf{C}_l, \mathbf{C}_k) n_k \cdot n'_l \cdot \omega_\sigma(\cdot) k_h(\cdot)} + \mathbf{x}^* \quad (14)$$

$$ms(\mathbf{y}) = \frac{\sum_{l=1}^{df'} \sum_{k=1}^{df} \delta(\mathbf{C}_l, \mathbf{C}_k) n_k \cdot n'_l \cdot \omega_{\sigma,s}(\cdot) k_h(\cdot) (\mathbf{C}\mathbf{y}_l - s\mathbf{C}\mathbf{x}_k)}{\sum_{l=1}^{df'} \sum_{k=1}^{df} \delta(\mathbf{C}_l, \mathbf{C}_k) n_k \cdot n'_l \cdot \omega_{\sigma,s}(\cdot) k_h(\cdot)} + s\mathbf{x}^* \quad (16)$$

$$ms(s) = \frac{\sum_{l=1}^{df'} \sum_{k=1}^{df} \delta(\mathbf{C}_l, \mathbf{C}_k) n_k \cdot n'_l \cdot \omega_{\sigma,s}(\cdot) k_h(\cdot) \left( 1 + \left( \frac{\mathbf{C}\mathbf{y}_l}{s\sigma} \right)^2 - \left( \frac{\mathbf{C}\mathbf{x}_k}{\sigma} \right)^2 \right)}{\sum_{l=1}^{df'} \sum_{k=1}^{df} \delta(\mathbf{C}_l, \mathbf{C}_k) n_k \cdot n'_l \cdot \omega_{\sigma,s}(\cdot) k_h(\cdot)} s \quad (17)$$

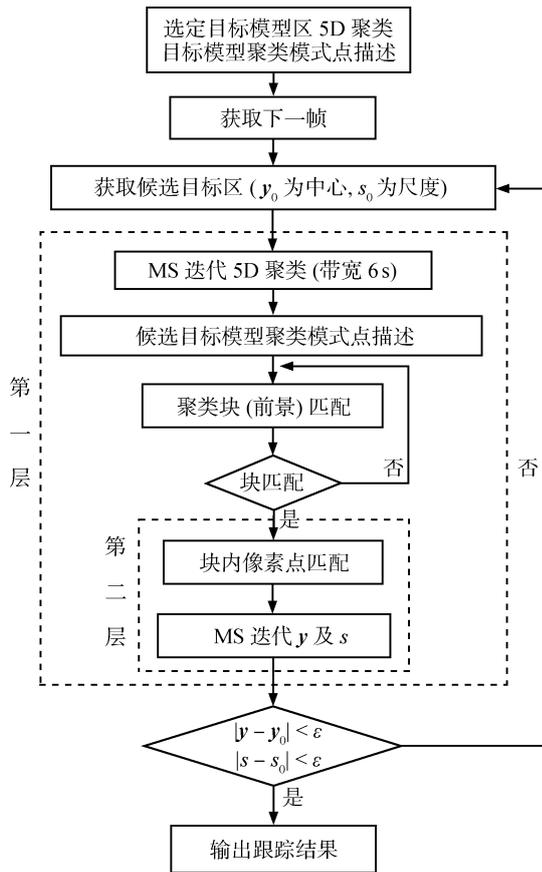


图 4 HMS 跟踪原理框图

Fig. 4 Block diagram of HMS tracking process

量越大, 得到的聚类数  $d$  越大.  $\sigma$  太大, 会使聚类分割丢掉太多细节;  $\sigma$  太小, 不能起到增强颜色特征的鲁棒性的作用 (本文初始  $\sigma$  参数是经过反复实验的经验值). 设聚类后求质心、前景/背景分离、匹配模式点的计算量为  $N_m$ , 由于聚类个数  $d$  很小, 因而产

生的计算量  $N_m$  也很小. 那么,  $N_1 = N_c + N_m$ . 第二层, 像素匹配的计算量为  $N_2$ , 若聚类块匹配后直接运算,  $d$  对相匹配的聚类逐点比较产生的计算量  $N_2 = d \times (n_k \times n'_k)^2$ . 而本文给出结合邻域一致性分层匹配, 只需计算聚类模式点近似比较  $N_2 = d \times 1$ . 令  $N_s$  为适应目标尺度  $s$  的变化而迭代搜索的尺度数目的计算量, 那么执行一次分层 MS 匹配迭代算法总的计算代价为  $C = N_s(N_1 + N_2) = N_s(N_c + N_m + d \times 1)$ . 显然, 本文的匹配计算量  $C$ , 除聚类计算量  $N_c$  较大之外, 其他都是几个模式点的匹配及数乘运算.

### 3 实验

实验平台环境: 硬件为主频 1.5 GHz, 内存 512M 的 R50e 的 IBM 笔记本电脑, 软件为 Matlab 7.1. 通过大量真实实验调试, 与其他两种算法<sup>[3, 7]</sup>相比, 本文算法是有效可行的. 在跟踪准确性方面, 抗光照、遮挡、阴影噪声干扰等鲁棒性方面强于经典方法, 搜索迭代次数随所选目标不同而有所不同, 略高于经典方法. 下面给出一些本文算法跟踪结果及性能分析, 第一帧手动初始化目标模型区, 匹配允许误差范围阈值  $T_c = 1$ .

**实验 1.** 户外足球比赛图像序列, 图像大小为  $720 \times 560$  像素, 共 285 帧. 球员之间相互遮挡, 存在较多的颜色特征相似的目标, 对快速运动的球员目标进行跟踪. 第一层初始聚类带宽及前景/背景分离距离参数为  $[\sigma, T_h] = [35, 18]$ , 第二层核函数带宽  $[\sigma, h] = [18, 12]$ . 设定的聚类带宽较大, 聚类程度较高, 自适应聚类为 15 类. 中途 101~120 帧、142~150 帧出现运动目标遮挡及碰撞. 序列帧 90, 105, 114, 162, 188, 233 跟踪结果如图 5 所示, 每帧迭代

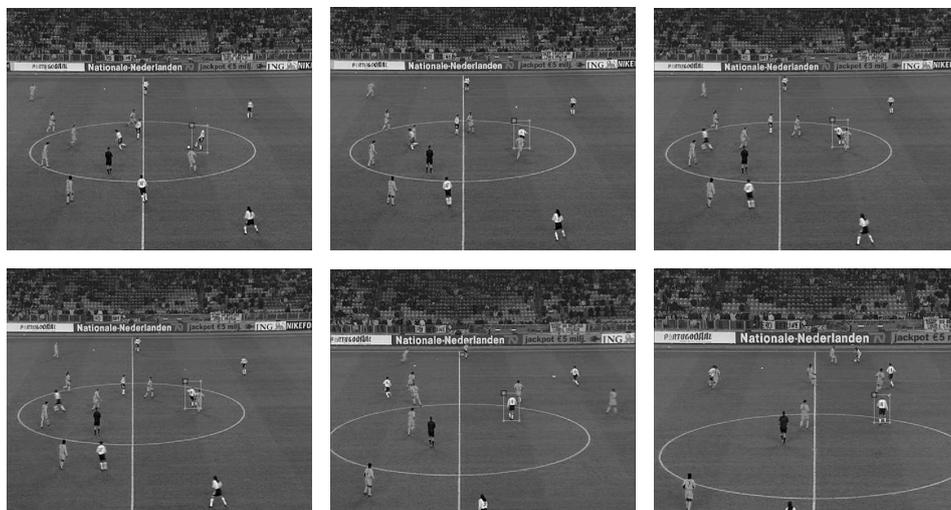


图 5 足球赛运动球员 HMS 算法跟踪结果

Fig. 5 Tracking results of football player using HMS

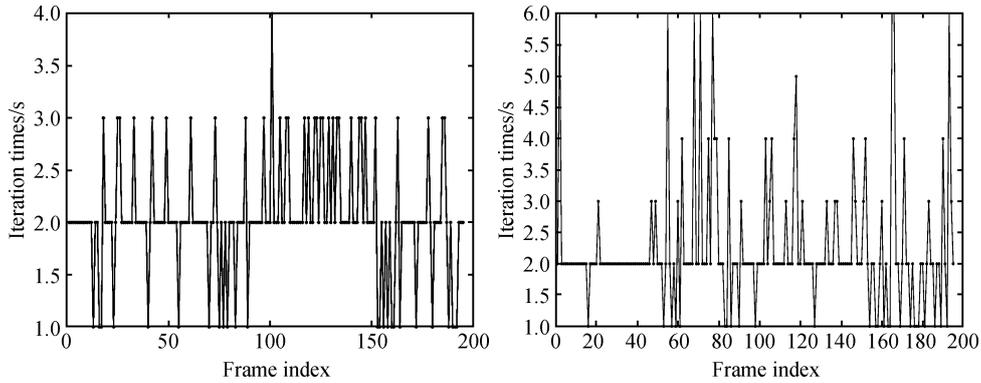


图 6 文献 [3] (左) 与本文 (右) 每帧迭代次数的对比

Fig. 6 The numbers of iterations with regard to the frame index using reference [3] (left) and HMS (right)

次数如图 6 所示. 与文献 [3] 算法相比, 本文算法平均每帧迭代 2.2512 次, 文献 [3] 算法为 2.0461 次.

**实验 2.** 室内人脸图像序列, 图像大小为  $128 \times 96$  像素, 共 100 帧. 人脸从正面转向反面, 再转过来, 遮挡严重. 相对于实验 1 而言, 实验 2 跟踪的目标颜色分布较细腻、变化较小. 第一层初始聚类带宽及前景/背景分离距离参数为  $[\sigma, T_h] = [26, 20]$ , 第二层核函数带宽  $[\sigma, h] = [12, 10]$ . 设定的聚类带宽较小, 聚类程度较低. 图 7 显示了本文算法与文献 [3] 算法对序列帧 18, 24, 90 的跟踪结果. 与其他两种 MS 算法 (文献 [3] 中的 MS 算法 1, 文献 [7] 中的 MS 算法 2) 相比, 目标质心的跟踪误差如图 8 所示, 本文算法的准确性更高.



图 7 本文算法 (上) 与文献 [3] (下) 的人脸跟踪结果

Fig. 7 Tracking results of face using HMS (up row) and reference [3] (down row)

**实验 3.** 户外强光照下穿梭于草地、沙漠、公路上的车辆图像序列. 车辆较长时间行驶在强光照下, 阴影影响较严重, 目标的尺度变化大, 背景环境变化大. 图像序列来源于 PETS 2005, 图像大小为  $160 \times 107$  像素, 共 1810 帧, 并提供有真实数据. 设定的聚类带宽及前景/背景分离距离参数为  $[\sigma, T_h] = [30, 23]$ , 核函数带宽  $[\sigma, h] = [18, 12]$ , 聚为 9 类. 序列帧 60, 210, 580, 1250, 1510, 1800 的跟踪结果如图 9 (见下页) 所示. 与其他两种 MS 算法 [3, 7] 相

比, 目标质心的跟踪误差如图 10 (见下页) 所示, 显然, 本文算法的跟踪准确性更高.

从以上几个实验的跟踪效果图及性能图比较分析来看, 与其他两种 MS 算法 [3, 7] 相比, 本文算法中, 聚类块模式点增强了颜色特征表征的鲁棒性, 在一定程度上克服了光照、噪声等的影响, 可获取稳定的

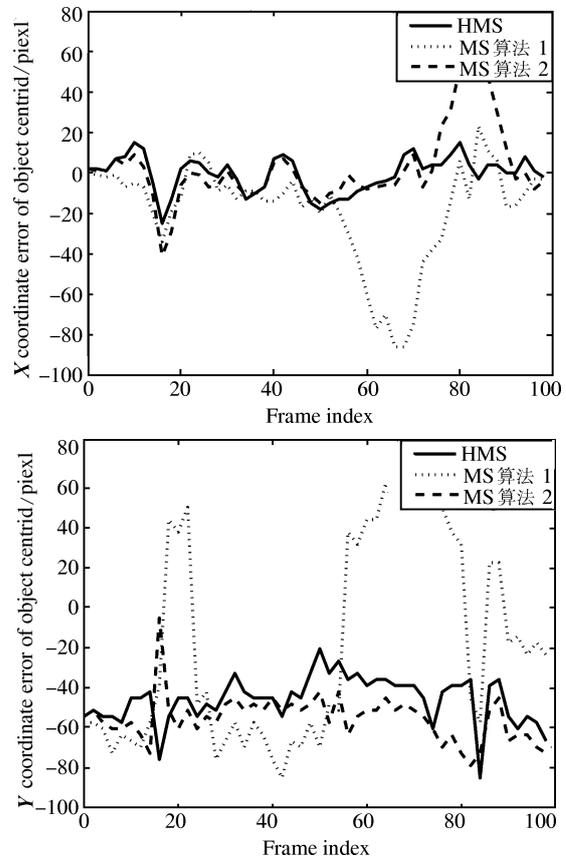


图 8 室内人脸序列的文献 [3, 7] 算法与本文算法的跟踪误差比较

Fig. 8 Comparison of tracking errors of face sequence indoor for three different algorithms, references [3, 7], and HMS



图9 穿梭于草地、沙漠公路上的车辆 HMS 算法跟踪结果  
Fig. 9 Tracking results of running car in the grass, desert and road using HMS

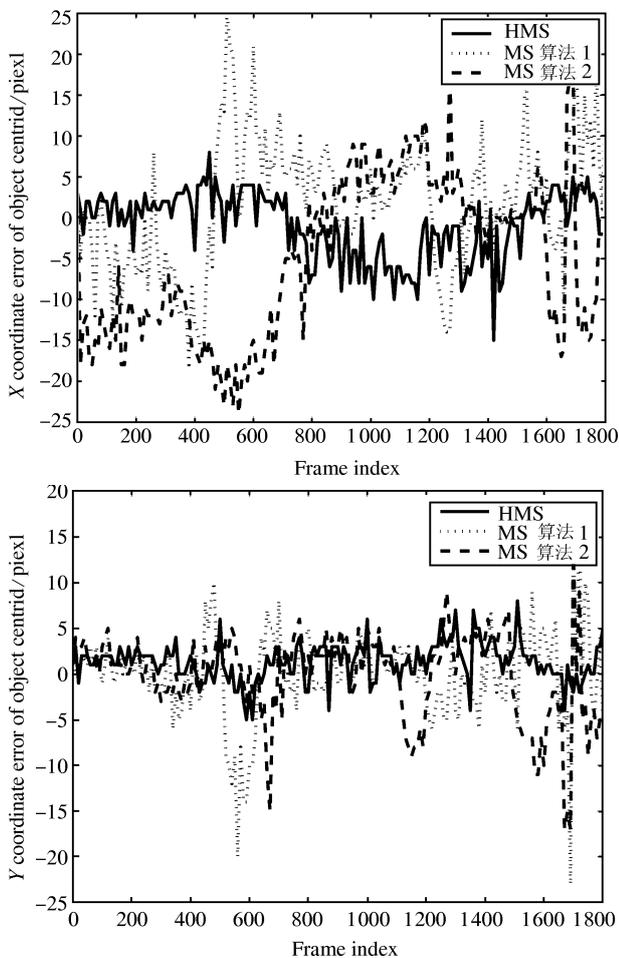


图10 车辆序列中文献 [3, 7] 算法与本文算法跟踪误差比较  
Fig. 10 Comparison of tracking errors in car sequence for three algorithms, references [3, 7] and HMS

运动块, 先聚类块匹配、再像素匹配提高了目标匹配的有效性, 跟踪效果更好, 分层匹配算法有效可行。

## 4 结论

本文提出了一种分层 Mean shift 目标跟踪算法, 通过多组实验, 分别从直观跟踪效果、迭代运算次数和跟踪误差方面, 与其他两种 MS 算法进行比较, 分析了本文给出算法的性能及有效性. 本文的贡献在于: 1) 将区域匹配的 MS 目标跟踪算法分层执行, 提出分层搜索的思想, 即先块匹配, 再像素匹配; 2) 以聚类块模式点描述目标模型, 利用邻域一致性计算目标运动平移量, 给出分层的 MS 匹配迭代跟踪算法. 将  $N$  个数据点收敛于  $d$  个模式点, 再收敛于目标质心模式点. 与经典方法相比, 分层处理提高了目标模型表达与匹配的鲁棒性, 无需逐点计算高斯核, 大大节省了计算量, 但是增加了聚类的运算。

本文算法存在的主要问题是: 在跟踪过程中, 目标 (参考) 模型描述没有在线更新; 环境的变化可能导致目标的前景颜色发生较大变化, 而本文的匹配误差允许阈值不能设定的太大, 这样可能使目标丢失而跟踪失败. 因此将来的工作是着重解决如何在线更新目标 (参考) 模型描述的问题, 以获得更为鲁棒的跟踪, 以及研究效率更高的聚类算法。

## References

- Comaniciu D, Ramesh V, Meer P. Real-time tracking of nonrigid objects using mean shift. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Hilton Head Island, USA: IEEE, 2000. 142–149
- Fukanaga K, Hostetler L D. The estimation of the gradient of a density function with applications in pattern recognition. *IEEE Transactions on Information Theory*, 1975, **21**(1): 32–40
- Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5): 564–575
- Bircheld S T, Rangarajan S. Spatiograms versus histograms for region-based tracking. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 1158–1163
- Li P H. A clustering based color model and fast algorithm for object tracking. In: Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China: IEEE, 2006. 671–674
- Li Pei-Hua. An improved mean shift algorithm for object tracking. *Acta Automatica Sinica*, 2007, **33**(4): 347–354 (李培华. 一种改进的 Mean Shift 跟踪算法. 自动化学报, 2007, **33**(4): 347–354)
- Yang C J, Duraiswami R, Davis L S. Efficient Mean-Shift tracking via a new similarity measure. In: Proceedings of IEEE Computer Society Conference on Computer Vision

- and Pattern Recognition. San Diego, USA: IEEE, 2005. 176–183
- 8 Yang C J, Duraiswami R, Elgammal A, Davis L S. Real-time Kernel-based Tracking in Joint Feature-spatial Spaces, Technical Report CS-TR-4567, Computer Science Department, University of Maryland, USA, 2004
  - 9 Elgammal A, Duraiswami R, Davis L S. Probabilistic tracking in joint feature-spatial spaces. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, USA: IEEE, 2003. 781–788
  - 10 Leung A P, Gong S G. Mean-Shift tracking with random sampling. In: Proceedings of the 17th British Machine Vision Conference. Edinburgh, British: University Trier Press, 2006. 729–738
  - 11 Naeem A, Mills S, Pridmore T. Structured combination of particle filter and kernel Mean-Shift tracking. In: Proceedings of the 21st International Conference on Image and Vision Computing. Great Barrier Island, New Zealand: IEEE, 2006. 1–6
  - 12 Han B, Joo S W, Davis L S. Probabilistic fusion tracking using mixture kernel-based Bayesian filtering. In: Proceedings of the 11th International Conference on Computer Vision. Janeiro, Brazil: IEEE, 2007. 1–8
  - 13 Thomas M, Kambhampettu C. An approximation to Mean-Shift via swarm intelligence. In: Proceedings of the 18th International Conference on Tools with Artificial Intelligence. Washington D. C., USA: IEEE, 2006. 583–590
  - 14 Park M, Liu Y X, Collins R T. Efficient mean shift belief propagation for vision tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Alaska, USA: IEEE, 2008. 1–8
  - 15 Han B, Comaniciu D, Zhu Y, Davis L S. Sequential kernel density approximation and its application to real-time visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(7): 1186–1197
  - 16 Miguel Á, Carreira-Perpiñán. Fast nonparametric clustering with Gaussian blurring Mean-Shift. In: Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh, USA: IEEE, 2006. 153–160
  - 17 Zhang K, Tang M, Kwok J T. Applying neighborhood consistency for fast clustering and kernel density estimation. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 1001–1007
  - 18 Comaniciu D, Ramesh V, Meer P. The variable bandwidth mean shift and data-driven scale selection. In: Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 438–445
  - 19 Collins R T, Liu Y X, Leordeanu M. On-line selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1631–1643
  - 20 Parzen E. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 1962, **33**(3): 1065–1076
  - 21 Rangarajan A, Chui H, Bookstein F L. The softassign procrustes matching algorithm. In: Proceedings of the 15th International Conference on Information Processing in Medical Imaging. Poultney, USA: IEEE, 1997. 29–42



许海霞 湖南大学电气与信息工程学院自动化重点实验室博士研究生。主要研究方向为计算机视觉, 目标跟踪和摄像机标定。本文通信作者。

E-mail: xhxia2002@126.com

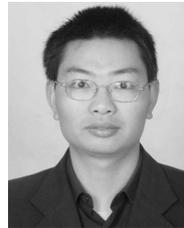
(XU Hai-Xia Ph.D. candidate at the Key Laboratory of Automation, Institute of Electrical and Information Engineering, Hunan University. Her research interest covers computer vision, object tracking, and camera calibration. Corresponding author of this paper.)



王耀南 湖南大学电气与信息工程学院教授。主要研究方向为机器视觉, 智能图像处理 and 智能机器人。

E-mail: yaonan@hnu.cn

(WANG Yao-Nan Professor at the Institute of Electrical and Information Engineering, Hunan University. His research interest covers computer vision, intelligent image processing, and intelligent robot.)



袁小芳 湖南大学电气与信息工程学院自动化重点实验室博士研究生。主要研究方向为智能控制, 神经网络。

E-mail: yuanxiaof@21cn.com

(YUAN Xiao-Fang Ph.D. candidate at the Key Laboratory of Automation, Institute of Electrical and Information Engineering, Hunan University. His research interest covers intelligent control and neural networks.)



周维 湘潭大学信息工程学院硕士研究生。主要研究方向为网络与信息安全, 图像处理。

E-mail: bigdaddy2004@126.com

(ZHOU Wei Master student at the Institute of Information Engineering, Xiangtan University. His research interest covers network and information security, and image processing.)



朱江 湖南大学电气与信息工程学院自动化重点实验室博士研究生。主要研究方向为计算机视觉, 机器人导航控制。

E-mail: jiang126@126.com

(ZHU Jiang Ph.D. candidate at the Key Laboratory of Automation, Institute of Electrical and Information Engineering, Hunan University. His research interest covers computer vision and robot navigation control.)