

一种新的基于 Mean Shift 的目标三自由度跟踪算法

左军毅 梁彦 赵春晖 潘泉
(西北工业大学自动化学院 西安 710072)

摘要: 标准 Mean Shift 跟踪算法仅能确定目标形心位置, 而不能确定其旋转角, 在跟踪细长形目标时鲁棒性不好。为此, 该文提出了一种三自由度 Mean Shift 跟踪算法, 新算法在计算目标特征分布直方图时, 用像素的位置转角及其到目标形心的归一化距离加权, 并将像素在局部坐标系下的特征转角作为新特征引入。这种新的目标表示模型能够方便地纳入 Mean Shift 优化框架, 通过迭代求解, 可同时精确确定目标的形心位置和方位指向。实验结果表明该算法精度高, 计算量小。

关键词: 目标跟踪; 均值漂移; 转角定位

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2008)01-0172-04

A New Mean Shift Based Algorithm for Tracking Targets with Three Degrees of Freedom

Zuo Jun-yi Liang Yan Zhao Chun-hui Pan Quan

(College of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Standard Mean Shift tracker can only successfully locate the object center, but fail to find its orientation, which make it not robust to track thin object. To remedy this, an improved mean shift tracker is proposed in this paper. The new tracker use new object representation, where pixels are weighted with both their position-angles and normalized distances from target center, furthermore, pixel's feature-angle, which can be seen as new feature, is introduced in. The new object representation can be conveniently integrated into the optimization framework of mean shift. By iterative optimization, both the location and orientation of targets can be precisely determined. Experimental results show the algorithm can get precise tracking results with low computational cost.

Key words: Object tracking; Mean shift; Orientation localization

1 引言

由于不需要逐帧检测目标, 基于匹配搜索技术的视频跟踪算法近来备受重视^[1-6]。这类算法的三要素是: 目标表示, 相似性度量和搜索策略。Comaniciu^[1]首先将 Mean Shift (MS)优化算法和非参数式的目标表示模型相结合, 通过最小化 Bhattacharyya 系数, 得到了一种高效的跟踪算法。该方法以计算量小, 对目标变形适应性强等特点而成为匹配搜索类的一种主流方法。由于这种方法只能跟踪目标形心的坐标而不能获得目标的方位指向, 本文称之为二自由度 MS 跟踪算法。当目标(尤其是细长形目标)发生图像平面内的旋转运动时, 二自由度 MS 算法的鲁棒性会因跟踪框中融入大量背景像素而急剧降低。另外, 二自由度 MS 算法所获取的目标运动信息完整性差, 有些情况下(比如在人机交互等领域), 目标的转角作为一种可被有效利用的信息有时和目标的位置信息具有相同的重要性。

针对二自由度 MS 算法的不足, Zhao^[2]提出了一种基于

颜色相关图三自由度 MS 算法。Bradski^[3]通过计算目标区域的二阶矩来实现转角跟踪, 并在人机交互中获得了应用。Jia^[4]分别在上一帧转角以及上一帧转角的基础上增/减 10° 3 个转角下执行 3 次 MS 算法, 来搜索最佳的目标转角。

作者对二自由度 MS 算法的机理进行分析后发现, 若适当修改目标表示模型, 将空间角度信息引入其中, 则转角定位问题便可采用和坐标定位类似的方法实现。据此本文提出了一种新的三自由度 MS 算法, 它可以看成是二自由度 MS 算法在转角定位方面的自然扩展。该方法的搜索效率高、目标表示模型简单, 因而比传统的一些三自由度跟踪算法计算量更小, 精度更高。

2 新的目标表示模型

定义两个坐标系: 图像坐标系 OXY 和局部坐标系 oxy 。前者以每帧图像左下角的像素为坐标原点 O , 水平向右的方向为横轴正方向 OX , 纵轴正方向 OY 垂直向上; 后者以目标形心为原点 o , 目标长轴某一方向为纵轴正方向 oy , 纵轴顺时针旋转 90° 后的短轴方向为横轴正方向 ox 。

假设第 t 帧, 目标形心的坐标为 $Z^t = [X^t, Y^t]^T$, OY 到 oy 的转角为 ϕ^t (逆时针旋转时 ϕ^t 为正), 则图像坐标系中坐

2006-11-02 收到, 2007-07-16 改回

国家自然科学基金重点项目(60634030)和国家自然科学基金(60372085)资助课题

标为 $\mathbf{Z} = [X, Y]^T$ 的像素在局部坐标系中的坐标 $\mathbf{z} = [x, y]^T$ 可表示为

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \phi^t & \sin \phi^t \\ -\sin \phi^t & \cos \phi^t \end{bmatrix} \begin{bmatrix} X - X^t \\ Y - Y^t \end{bmatrix} \quad (1)$$

在局部坐标系中, 对目标区域的每个像素 P_i (局部坐标为 (x, y)) 定义两个角度: 位置转角和特征转角。 P_i 的位置转角 θ_i 表示 P_i 相对于 oy 正或负半轴的转角, 定义为

$$\theta_i = \begin{cases} 0, & x = 0 \text{ 且 } y = 0 \\ \pi/2, & x < 0 \text{ 且 } y = 0 \\ -\pi/2, & x > 0 \text{ 且 } y = 0 \\ -\arctg(x/y), & \text{其它} \end{cases} \quad (2)$$

显然, $\theta \in [-\pi/2, \pi/2]$ 。像素 P_i 的特征转角 φ_i 定义为横轴正方向 ox 沿逆时针转至 oP_i 时经历的转角, $\varphi_i \in [0, 2\pi)$, 规定原点处像素的特征转角为 0。

假设初始时刻, 目标相对于图像坐标系的位置和转角分别为 $\mathbf{Z}^0 = [X^0, Y^0]^T$ 和 ϕ^0 , 并令 $\mathbf{S}^0 = [(\mathbf{Z}^0)^T, \phi^0]^T$, 任给目标区域的像素点 \mathbf{Z}_i , 根据式(1), 式(2)可求得其对应的局部坐标系下的坐标和位置转角 $\mathbf{s}_i = [\mathbf{z}_i^T, \theta_i]^T$, 其中 $\mathbf{z}_i = [x_i, y_i]^T$ 。将特征转角的定义域等分成 m_φ 个量化级, 将颜色空间量化成 m_p 个量化级, 则目标模型可用加权直方图表示为 $\mathbf{q} = \{q_{uv}\}_{u=1 \dots m_p, v=1 \dots m_\varphi}$, 其中

$$q_{uv} = C \sum_{i=1}^N K(\mathbf{s}_i) \delta[b_p(\mathbf{z}_i) - u] \delta[b_\varphi(\mathbf{z}_i) - v] \quad (3)$$

函数 $b_p(\mathbf{z}_i): R^2 \rightarrow \{1 \dots m_p\}$ 将跟踪窗中坐标为 \mathbf{z}_i 的像素按照其颜色特征映射到颜色量化空间中; $b_\varphi(\mathbf{z}_i): R^2 \rightarrow \{1 \dots m_\varphi\}$ 将像素 \mathbf{z}_i 按照其特征转角映射到角度量化空间中; $\delta[\cdot]$ 是 Kronecker 函数; u 和 v 分别为颜色空间和角度空间中量化的索引; N 为目标模板中像素的个数; $C = 1/\sum_{i=1}^N K(\mathbf{s}_i)$ 为

归一化因子, 使 $\sum_{u=1}^{m_p} \sum_{v=1}^{m_\varphi} q_{uv} = 1$; $K(\mathbf{s}_i)$ 为选定的核函数, 定义其轮廓函数为 $k(\mathbf{s}_i)$, 满足

$$K(\mathbf{s}_i) = c_k k(\|\mathbf{s}_i\|_{\mathbf{H}^{-1}}^2) \quad (4)$$

$\|\mathbf{s}_i\|_{\mathbf{H}^{-1}}$ 表示向量 \mathbf{s}_i 关于正定对角阵 \mathbf{H}^{-1} 的椭圆范数, 这里 $\mathbf{H}^{-1} = \text{diag}[1/h_x^2, 1/h_y^2, 1/h_\theta^2]$, 其中 h_x 和 h_y 为位置空间核的带宽, h_θ 为转角空间核的带宽, 并称 $\mathbf{H}^{1/2}$ 为带宽矩阵。于是式(3)可改写为

$$q_{uv} = C \sum_{i=1}^N k(\mathbf{s}_i^T \mathbf{H}^{-1} \mathbf{s}_i) \delta[b_p(\mathbf{z}_i) - u] \delta[b_\varphi(\mathbf{z}_i) - v] \quad (5)$$

文献[1]仅用像素到目标形心的马氏距离加权, 距离越小权值越大; 而式(5)表示像素的权值不仅取决于它到目标形心的距离, 还取决于它的位置转角, 位置转角越大权值越小。Comanicu^[7]指出当核轮廓函数 $k(x)$ 是凸的单调递减函数时, MS 算法收敛, 因此, 之所以对位置转角采用这种加权方式是出于对 MS 算法收敛性的考虑。

假设候选目标状态在局部坐标系下可表示为 \mathbf{r} , \mathbf{r} 是三维向量, 分别表示目标的形心坐标和转角(这里对 \mathbf{r} 需稍作解释: 要从第 t 帧目标状态开始获取第 $t+1$ 帧的目标状态需经过若干次迭代, 第 k 次迭代后由于局部坐标系的平移旋转, 局部坐标系下的目标状态向量必为零向量, 这里的 \mathbf{r} 是对下一次迭代结果在当前局部坐标系下状态向量的一个猜测, 因此 \mathbf{r} 是相对于三维零向量的一个偏移量), 于是候选目标可用加权直方图表示为 $\mathbf{p}(\mathbf{r}) = \{p_{uv}(\mathbf{r})\}_{u=1 \dots m_p, v=1 \dots m_\varphi}$, 其中

$$p_{uv}(\mathbf{r}) = C_r \sum_{i=1}^{N_s} k((\mathbf{r} - \mathbf{s}_i)^T \mathbf{H}^{-1} (\mathbf{r} - \mathbf{s}_i)) \delta[b_p(\mathbf{z}_i) - u] \delta[b_\varphi(\mathbf{z}_i) - v] \quad (6)$$

相应的归一化系数为 $C_r = 1/\sum_{i=1}^{N_s} k((\mathbf{r} - \mathbf{s}_i)^T \mathbf{H}^{-1} (\mathbf{r} - \mathbf{s}_i))$,

N_s 表示跟踪窗中的像素个数, 其他相关参数的定义与式(3)同。式(5), 式(6)中, 若 \mathbf{H}^{-1} 中的 h_θ 取无穷大, 则目标表示模型退化成文献[1]的只采用距离加权的形式。

3 目标三自由度跟踪算法

本文采用 Bhattacharyya 系数

$$\rho(\mathbf{r}) = \rho[\mathbf{p}(\mathbf{r}), \mathbf{q}] = \sum_{u=1}^{m_p} \sum_{v=1}^{m_\varphi} \sqrt{p_{uv}(\mathbf{r}) q_{uv}} \quad (7)$$

来描述目标模型和候选目标的相似程度, $\rho(\mathbf{r})$ 越大, 目标和候选目标越相似。于是跟踪问题转化为优化问题:

$$\mathbf{r} = \max_{\mathbf{r}} \rho(\mathbf{r}) \quad (8)$$

求解式(8)的迭代公式可按与文献[1]类似的推导过程获得, 推导结果为

$$\mathbf{r}_{k+1} = \frac{\sum_{i=1}^{N_s} w_i \mathbf{s}_i g((\mathbf{r}_k - \mathbf{s}_i)^T \mathbf{H}^{-1} (\mathbf{r}_k - \mathbf{s}_i))}{\sum_{i=1}^{N_s} w_i g((\mathbf{r}_k - \mathbf{s}_i)^T \mathbf{H}^{-1} (\mathbf{r}_k - \mathbf{s}_i))} \quad (9)$$

其中

$$w_i = \sum_{u=1}^{m_p} \sum_{v=1}^{m_\varphi} \left\{ \delta[b_p(\mathbf{z}_i) - u] \delta[b_\varphi(\mathbf{z}_i) - v] \sqrt{\frac{q_{uv}}{p_{uv}(\mathbf{r}_k)}} \right\} \quad (10)$$

函数 $g(x)$ 满足: $g(x) = -k'(x)$ 。当 $k(x)$ 取为 Epanechnikov 核时, $g(x)$ 为常数, 式(9)可得到简化。 \mathbf{r}_k 中的下标 k 表示迭代次数。

由于目标表示方法以及转角等都是在局部坐标系下定义的, 跟踪算法的核心部分也是在局部坐标系下执行的, 每步迭代后, 局部坐标系都要平移、旋转到新的位置和方向, 以去除坐标分量和转角分量的相关性, 因此在局部坐标系中, 新一轮迭代都是从坐标原点和零转角开始的。任何时刻, 目标在图像坐标系中的位置、转角信息都被包含在两种坐标系的相对关系中。若在第 t 帧, 经过 $k+1$ 次迭代后的结果满足收敛判别条件, 则第 t 帧输出的目标形心坐标和转角值为

$$\mathbf{S}_{k+1}^t = \mathbf{S}_k^t + \mathbf{r}_{k+1} \quad (11)$$

$\mathbf{S}_k^t = [(\mathbf{Z}_k^t)^T, \phi_k^t]^T$ 表示第 t 帧经过 k 次迭代后局部坐标系的位置和转角。

和二自由度 MS 算法一样, 带宽矩阵 $\mathbf{H}^{1/2}$ 影响着参与计算的像素的个数, 从而影响算法的性能。若直接借鉴文献[1], 并根据位置转角的取值范围, 令 3 个带宽参数 h_x , h_y , h_θ 分别等于目标长宽尺寸 h_{ox} , h_{oy} 的一半以及 $\pi/2$, 则目标区域中大量像素将不参与计算, 跟踪结果只依赖于少数大权值像素, 这对提高算法鲁棒性极为不利。为此, 本文将带宽适当放大, 取:

$$\mathbf{H}^{1/2} = \text{diag}\{h_{ox}/\sqrt{2}, h_{oy}/\sqrt{2}, \pi/\sqrt{2}\} \quad (12)$$

限于篇幅, 本文对带宽放大系数及其鲁棒性不作详细讨论。

算法的执行步骤如下:

(1) 初始化 从目标检测模块获取初始时刻目标在图像坐标系中的位置和转角 $\mathbf{S}_0^t = [(\mathbf{Z}_0^t)^T, \phi_0^t]^T$, 并据此初始化局部坐标系, 并令 $\mathbf{r}_0 = [0, 0, 0]^T$, 按式(5)生成目标模型 \mathbf{q} , 并设置跟踪帧数 $t = 0$, 迭代次数 $k = 0$;

(2) 处理下一帧, $t \leftarrow t + 1$, $k = 0$;

(3) $k \leftarrow k + 1$, 以 \mathbf{r}_{k-1} 为初值, 根据式(6)计算候选目标模型 $\mathbf{p}(\mathbf{r}_{k-1})$, 根据式(10)计算权重 $\{w_i\}_{i=1 \dots N_s}$;

(4) 根据式(9)计算新的目标位置和转角 \mathbf{r}_k ;

(5) 移动、转动局部坐标系到新的位置: $\mathbf{S}_k^t = \mathbf{S}_{k-1}^t + \mathbf{r}_k$;

(6) 若 \mathbf{r}_k 中的 3 个分量同时满足 $|x_k| < \varepsilon_p$, $|y_k| < \varepsilon_p$, $|\theta_k| < \varepsilon_\theta$, 将 \mathbf{S}_k^t 作为当前帧目标位置和转角输出, 并令 $\mathbf{S}_0^{t+1} = \mathbf{S}_k^t$, $\mathbf{r}_k = [0, 0, 0]^T$, 并转到第(2)步; 否则令 $\mathbf{r}_k = [0, 0, 0]^T$ 转到第(3)步。

4 实验结果及分析

为了验证提出的算法的有效性, 本文在多种场景下对其进行了实验研究, 并与经典的 Camshift^[3]算法进行了性能对比。实验时选用 Epanechnikov 核来计算目标表示模型。转角收敛判别条件中的 ε_θ 取 0.02rad, 位置收敛判别条件中的 ε_p 取 1。将特征转角按照局部坐标系的坐标轴以及各象限角平分线分成 8 个量化级, 同时为了消除光照变化的影响, 利用 HSV 颜色空间的 H 和 S 通道结合特征转角来计算加权直方图, 两个颜色通道均被分成 8 个量化级。跟踪框初始化是在第 1 帧用手工的方式完成的。

在图 1 所示的序列中, 作为跟踪对象的电话听筒在复杂背景环境下做平移、旋转运动。由于二自由度 MS 算法不能

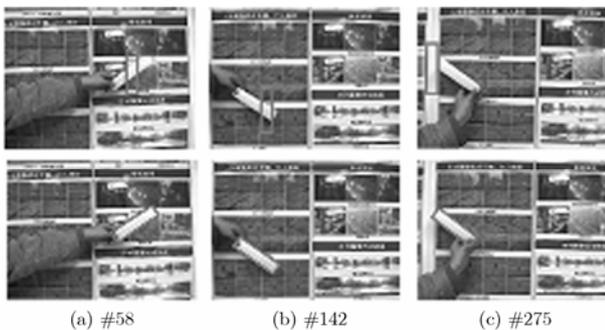


图 1 二自由度 MS 和本文算法对细长体的跟踪结果

驱动其跟踪框随目标一起转动(图 1 第 1 行), 当目标倾斜时, 跟踪框和目标的重叠区域变得很小, 引起大量背景像素融入其中, 这大大增加了算法丢失目标的风险。在第 270 帧左右, 目标斜着穿越了一个与其具有相似颜色模式的背景区域, 这个区域对跟踪框产生“牵拉”作用, 而立即引起跟踪失败。相比之下, 本文的三自由度 MS 算法能使跟踪框紧随目标平移、旋转, 在很大程度上能够抵御相似背景模式的吸引, 因而鲁棒性更好(图 1 第 2 行)。

就计算量而言, 二自由度 MS 算法的计算复杂度约为 nN_sC_p ^[1], 其中 n 为单帧平均迭代次数, N_s 为目标区域的像素个数, C_p 是计算单个像素的权值所付出的代价。不难证明, 本文算法的计算复杂度也是 nN_sC_p , 不过单帧平均迭代次数较前者略多。图 2 给出了二自由度 MS 和本文算法的迭代次数随帧数的变化曲线, 前者的平均迭代次数为 2.8 次/帧, 后者为 3.2 次/帧, 这比文献[4]的迭代次数要少得多, 按文献[4]的算法若仅实现转角自适应, 则平均迭代次数多达 8.2 次/帧。因此本文算法的计算量和二自由度 MS 相当, 比文献[4]的算法要小。

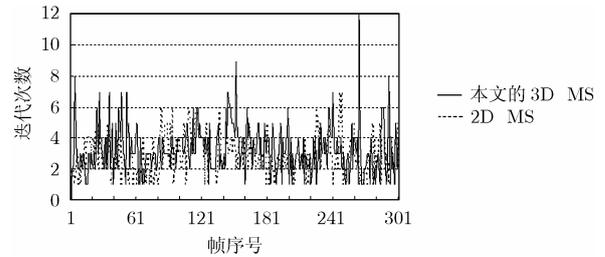


图 2 二自由度 MS 和本文算法的迭代次数

图 3 给出了一个胳膊序列的跟踪实例, 序列长 614 帧, 图像大小为 320×240 , 调用 Opencv 中的标准库函数、执行 Camshift 算法对其进行三自由度跟踪, 其结果如图 3 第 1 行, 第 2 行是本文算法的执行结果。将两种算法对胳膊转角的定位结果绘制成图 4, 结果发现两者跟踪精度的差别不大, 而 Camshift 算法所对应的曲线上的“毛刺”略多, 这和观测到的一种实验现象相一致: Camshift 给出的转角值总是在真值附近出现小幅度的抖动, 这种抖动在图 5 的视频序列中表现得更为严重。相比之下, 本文算法给出的结果平滑性较好。

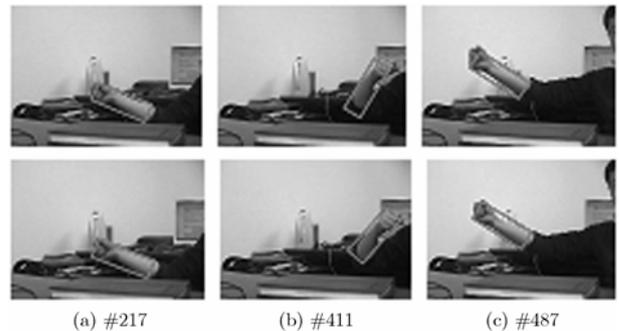


图 3 胳膊序列的三自由度 MS 跟踪

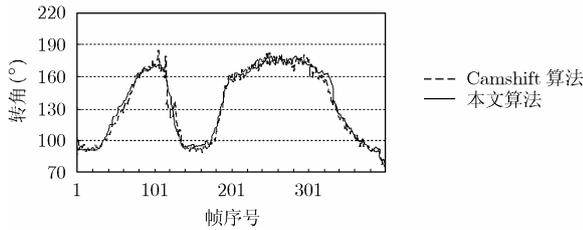


图 4 两种算法对胳膊转角的定位结果

图 5 的视频序列中人脸在做平移和图像平面内的旋转运动。与胳膊序列相比，人脸序列的跟踪框更接近于正方形，这对角度定位带来了很大挑战。由于图中人脸较为圆胖，Camshift 很难通过二阶矩的方法精确捕捉到真实的主轴方向，角度定位精度差(图 5 中第 1 行)。我们用手工方式标注出人脸转角的真值，并计算 Camshift 的角度定位误差，得到误差曲线(图 6 中的虚线)，误差均值达 18.2°，误差峰值超过 35°，可见精度很差。另外，Camshift 的误差曲线出现上下大幅度跳变，这说明角度定位的稳定性也不好，转角频繁在临帧之间发生突变，这不符合转角变化具有连续性这一事实。图 5 第 2 行是本文算法的实验结果，对应的误差曲线(图 6 中的实线)，误差均值为 5.2°，误差峰值为 14°，可见本文算法的角度定位结果无论从精确性还是平稳性上讲都比 Camshift 算法好。这主要是因为本文的目标表示模型中引入了空间角度信息，使目标模型与候选目标的匹配度对角度变化更敏感，而 Camshift 没有充分利用空间信息，造成跟踪精度降低。

实验还表明，由于本文算法不需要计算图像二阶矩，计算量明显小于 Camshift 算法。在 P IV 3.0GHz 的试验平台

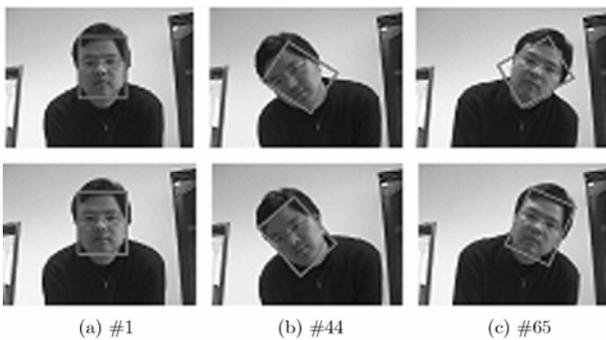


图 5 人脸序列的三自由度 MS 跟踪

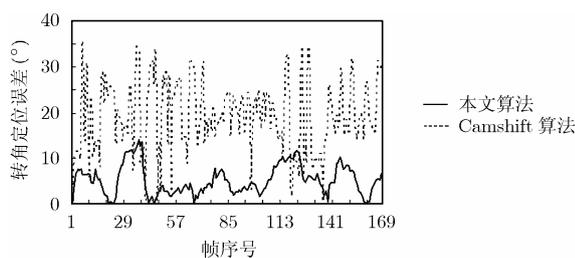


图 6 两种算法对人脸转角的定位误差

上针对图 5 的人脸序列进行实验，跟踪框尺寸为 82×106 ，二自由度 MS 算法的单帧平均执行时间是 15ms，本文算法是 18ms，而 Camshift 算法是 28ms。

5 结束语

本文通过修改目标表示模型，成功地将二自由度 MS 跟踪算法推广到三自由度跟踪情况，以很小的计算成本同时实现了对目标的形心定位和转角定位，目标三自由度运动信息的实时获取明显地提高了 MS 算法的鲁棒性。实验表明，该算法比一些经典的三自由度跟踪算法精度更高，计算量更小，因而更具实用价值。

参考文献

- [1] Comaniciu D, Ramesh V, and Meer P. Kernel-based object tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003, 25(5): 564-577.
- [2] Zhao Q and Tao H. Object tracking using color correlogram. *The 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, Beijing, 2005: 263-270.
- [3] Bradski G R. Computer vision face tracking for use in a perceptual user interface. *IEEE Workshop on Applications of Computer Vision*, Princeton, 1998: 214-219.
- [4] 贾静平, 柴艳妹, 赵荣椿. 一种健壮的目标多自由度 Mean Shift 序列图像跟踪算法. *中国图象图形学报*, 2006, 11(5): 707-713.
- [5] Jia J-P, Chai Y-M, and Zhao R-C. Robust tracking of objects in image sequences using multiple degrees of freedom mean shift algorithm. *Chinese Journal of Image and Graphics*, 2006, 11(5): 707-713.
- [6] Yang C J, Duraiswami R, and Davis L. Efficient spatial-feature tracking via the mean-shift and a new similarity measure. *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA, 2005: 176-183.
- [7] Zhang H H, Huang W M, and Huang Z Y, et al. Affine object tracking with kernel-based spatial-color representation. *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, 2005: 293-300.
- [8] Comaniciu D and Meer P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603-619.

左军毅：男，1975 年生，博士生，研究领域为图像与视频处理、目标跟踪等。
 梁彦：男，1971 年生，副教授，主要研究方向为自适应滤波、多目标跟踪、信息融合、图像处理。
 赵春晖：男，1973 年生，博士生，讲师，研究领域为计算机视觉、智能信息处理等。
 潘泉：男，1961 年生，教授，博士生导师，主要研究方向为多目标跟踪理论及应用、C³I 系统建模与仿真、动态系统建模、数据融合、图像与视频处理等。