

基于模板更新的自适应 Mean-shift 跟踪算法

李勇勇 谭毅华 田金文

(华中科技大学图像识别与人工智能研究所 武汉 430074)

(华中科技大学多谱信息处理技术国家级重点实验室 武汉 430074)

摘要 提出了一种利用 Mean-shift 算法处理目标跟踪定位,并以 SIFT 特征点匹配结果的最小二乘模型来求解缩放系数和更新目标模型的自适应跟踪方法。该方法实现了目标的快速跟踪,解决了模板更新和目标的尺度缩放问题。实验结果表明,该算法在处理目标尺度变化较大的情况下具有很强的鲁棒性。

关键词 目标跟踪,均值漂移,SIFT点,尺度变化,最小二乘模型

Template Updating Based Adaptive Tracking Algorithm Using Mean-shift

LI Yong-yong TAN Yi-hua TIAN Jin-wen

(Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074, China)

(The State Key Laboratory for Multi-spectral Information Processing Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract This paper proposed an adaptive tracking algorithm in which the Mean-shift algorithm is taken to obtain target position. Among this framework, the target model is updated by analyzing the results of SIFT points matching using least-squares model. This algorithm can fast track target with the advantages that the template updating and the finding of scale coefficients are implemented at the same time. The experimental results show that the algorithm is very robust to the cases that the scales of target vary largely.

Keywords Target tracking, Mean-shift, SIFT point, Scales variant, Least-squares mode

序列图像和视频中的运动目标是人眼最为关注的部分。与静态图像相比,序列图像和视频的最大优点就在于可以捕获运动信息。作为计算机视觉领域的核心问题之一,跟踪视频中的目标可有效地提取目标的运动信息。目前,目标跟踪算法主要包括两类:1)基于状态估计的跟踪方法;2)基于匹配的跟踪方法。基于状态估计的方法主要有 Kalman 滤波和粒子滤波^[1,2,11],基于匹配的跟踪算法主要有 Mean-shift^[3,4], camShift, SIFT^[5]等。Mean-shift 是一种基于模板匹配的跟踪方法,用灰度直方图分布表达目标区域的特征,然后用 Bhattacharyya 系数作为模板区域和当前目标区域的相似性度量,找到和模板区域最相似的区域后,跟踪算法终止。该算法能实现目标模板的快速匹配,但是不能解决模板更新和目标缩放问题。

SIFT 点特征是 D. G. Lowe 1999 年提出、2004 年完善总结的图像的局部特征描述子^[6,7],对旋转变换、尺度变化、仿射变换具有不变性。但是,基于 SIFT 匹配的跟踪算法非常依赖目标图像的内容和质量^[5],当目标本身信息很单一或者目标比较小时,很难提取到足够的特征点。此时无法支持跟踪算法的继续进行,且算法实现复杂,计算量大,时间复杂度高,阻碍了目标实时跟踪的实现。

一方面由于跟踪过程中目标在部分遮挡、光照剧烈变化、

面积较小时 Mean-shift 可以保证目标定位的准确性,而另一方面由于目标在尺度发生变化时 SIFT 特征可以保证其准确匹配,因此为了实现跟踪算法的稳定性,将其二者结合,形成一种新的目标跟踪算法,是一种很自然的想法。文献[8]采用 SIFT 点的相似性度量和基于 Mean-shift 灰度直方图的相似性度量两个度量准则,通过构建 EM 模型,最终找到当前帧中目标的位置来实现跟踪,但其没能很好地解决目标缩放问题。文献[10]中,先由 Mean-shift 算法获得目标的大致位置,再通过 4×4 分块的 SIFT 匹配得到目标的精确位置,但这种分块方法不合理,而且仍然没能适应目标的尺度变化,即缩放问题。

良好的跟踪算法要实现以下两点:1)精确地定位目标;2)跟踪的目标框大小与目标的外接轮廓比较接近。SIFT 算法跟踪能实现目标的精确定位,但是当目标部分遮挡、光照剧烈变化、目标面积较小时,不能保证跟踪继续进行。Mean-shift 基于模板匹配的跟踪算法对目标部分遮挡、光照剧烈变化、目标面积较小情况具有强鲁棒性,但其本身不能解决模板更新问题。模板更新对整个算法的优越性至关重要,一方面目标存在尺度缩放需要更新模板;另一方面运动目标每一帧都存在变化,不考虑这些变化仍然采用第一帧确定的模型去跟踪就会导致目标定位不准确或者目标丢失,所以算法需要有合

到稿日期:2010-04-02 返修日期:2010-07-08 本文受航空科学基金(20085179008)资助。

李勇勇(1984—),女,硕士生,主要研究方向为目标跟踪与视频监控系统,E-mail:yydely@hotmail.com;谭毅华(1975—),男,副教授,主要研究方向为智能视觉监控;田金文(1960—),男,教授,主要研究方向为目标检测与识别。

理的模板更新策略。

综合考虑这些情况,本文提出了以 Mean-shift 跟踪算法为主准确定位目标、以 SIFT 算法为辅更新目标模型的自适应跟踪算法。算法实现过程中首先采用 Mean-shift 算法处理目标跟踪,确定目标在当前帧中所在的位置,然后在跟踪的后续处理中,先对 SIFT 向量采用降维处理,用向量夹角的余弦值作为点匹配的相似性度量,再基于 SIFT 点特征匹配结果的最小二乘模型,求解目标的缩放系数。为了保证算法的实时性,算法只在目标中心局部区域内提取 SIFT 点以及进行点特征匹配。在目标模板的更新过程中,采用了将目标模型与当前跟踪结果得到的目标模型进行线性组合的策略,达到一定的帧数就进行一次目标模板的更新,很好地解决了目标缩放和模板更新问题。由于采用了合理的模板更新策略,算法在目标部分遮挡的情况下也表现出了良好的跟踪性能。

1 基于 Mean-shift 算法的目标跟踪

Mean-shift 是一种基于核密度估计的模板跟踪方法,具有无参密度估计和快速匹配的优点。目标跟踪的目的就是实现相邻两帧间的相似性查找^[8],在实现过程中首先在当前帧中给定一个预知的目标位置,后续查找的工作就是找到一个置信区,使得候选区域和模板区域最相似。在本文算法的实现过程中,相似性度量准则采用 Bhattacharyya 系数^[9],其实现公式为

$$\rho(p(y), q) = \sum_{b=1}^m \sqrt{p_b(y) q_b} \quad (1)$$

$$\text{式中, } p_b(y) = \frac{\sum_{i=1}^{N_h} k\left(\left|\frac{(y-x_i)}{h}\right|\right)^2}{\sum_{i=1}^{N_h} k\left(\left|\frac{(y-x_i)}{h}\right|\right)^2}, b=1, \dots, m, \text{ 是目标候}$$

选模型的概率分布。在本文算法中 p_b 为当前帧目标候选模型分灰度直方图的概率分布, q_b 为目标模型分灰度直方图的概率分布。

目标中心位置迭代过程的实现公式为

$$y_1 = \frac{\sum_{i=1}^{N_h} x_i g\left(\left|\frac{y_0 - x_i}{h}\right|\right) w_i}{\sum_{i=1}^{N_h} g\left(\left|\frac{y_0 - x_i}{h}\right|\right) w_i} \quad (2)$$

式中, $w_i = \sum_{b=1}^m \sqrt{\frac{q_b}{p_b(y_0)}} \delta_b$ 为权值, $g(\cdot)$ 是核函数 $k(\cdot)$ 的导数, h 为核函数带宽。

这一过程实现了相邻两帧间由候选区域到目标运动后的真实区域的查找,后续过程需要解决目标的模板更新问题。

2 基于 SIFT 特征点的模板更新策略

在 Mean-shift 跟踪算法中,核函数带宽的选择十分重要,它对应目标实际的宽度。因此在解决模板更新的问题上,首先解决目标尺度的缩放,然后是模板信息的更新。其过程可分为以下几个步骤:1)将模板区域 SIFT 点和当前候选区域 SIFT 点进行匹配,得到匹配的点对;2)利用最小二乘法进行曲线拟合,通过求解曲线的参数得到目标的缩放系数;3)利用缩放系数求得置信区域的真实范围,并求解此范围的目标灰度直方图分布;4)对求得的灰度直方图和模板灰度直方图进行线性组合,以更新目标模型。

在此过程中,如果目标非常小或者目标本身信息单一,以

致没有足够的 SIFT 点用以求解缩放系数,则不进行目标缩放的处理,直接以上一帧中目标框的大小进行跟踪。此自适应跟踪策略增强了跟踪算法的鲁棒性。

2.1 SIFT 特征点的提取与匹配

SIFT 点特征的匹配过程包括两个步骤:1)SIFT 特征点的提取;2)特征点的匹配。由于提取 128 维的 SIFT 点非常耗时,而本文仅用 SIFT 点的匹配结果进行曲线拟合来求解目标缩放系数,因此算法仅提取 64 维的 SIFT 向量进行后续处理。

SIFT 特征点提取算法,首先对输入的图片进行不同尺度的高斯核函数滤波,形成高斯金字塔模型,然后将相邻尺度的两个高斯图像相减,得到高斯金字塔的多尺度空间表示。在尺度空间中,求其局部极值,得到图像的 SIFT 特征点。实现过程如下。

首先计算输入图像的尺度空间

$$L(x, y, \sigma) = I(x, y) \otimes G(x, y, \sigma) \quad (3)$$

$$\text{式中, } I(x, y) \text{ 表示输入图像, } G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}, \sigma$$

称为尺度因子,其值越小则表征该图像被平滑得越少,相应的尺度也就越小。 L 代表了图像的尺度空间。

然后计算高斯差分尺度空间的局部极值,即 SIFT 点,计算公式如下

$$\begin{aligned} D(x, y, \sigma) &= I(x, y) \otimes (G(x, y, k\sigma) - G(x, y, \sigma)) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (4)$$

SIFT 特征点的匹配过程:首先,根据当前帧中目标区域的特征点,建立二叉树;其次,根据前一帧中目标区域的特征点,在此二叉树中进行匹配,返回此特征点的二近邻。

在设计匹配准则时,考虑每一个 SIFT 点是由 64 维的 SIFT 向量表示的,而向量的相似性度量可以用余弦值进行表达,并且这一值的范围归一化在 0~1 的范围内,余弦值越大两点越相似。两向量夹角的余弦值的计算公式如下

$$\cos(x, y) = \frac{x \cdot y}{|x| |y|} \quad (5)$$

当两向量夹角的余弦值大于一定阈值(经验值为 0.95)时,则认为两点为相似的匹配点对。

2.2 最小二乘法求解目标缩放系数

在处理目标的缩放问题时,首先需要建立目标尺度变化的仿射模型。由于刚体目标的运动可以用仿射变换来表示两帧之间目标的平移、旋转和比例缩放的变化,求得这些仿射变换的参数,就可以得到目标尺度的缩放比例。仿射模型的构造函数如下

$$\begin{aligned} x_i' &= k \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} x_i + \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \\ &= \begin{pmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{pmatrix} x_i + \begin{pmatrix} a_3 \\ a_4 \end{pmatrix} \end{aligned} \quad (6)$$

式中, x_i' 和 x_i 分别是同一个目标特征点在第 $i+1$ 帧和 i 帧中的位置, k 表示目标大小的伸缩幅度, θ 表示目标的旋转角度, Δx 和 Δy 分别表示该目标特征点的水平和垂直位移。由上式可知, k 即为所求尺度变化系数,且

$$k = \sqrt{(k\cos\theta)^2 + (k\sin\theta)^2} = \sqrt{a_1^2 + a_2^2} \quad (7)$$

所以,只需要得到 a_1 和 a_2 , 就可以求得目标的缩放系数。目标缩放系数的求解过程如下

设随机变量 η 关于变量 ζ 的回归方程为 $y = \mu(x; a_1, a_2, \dots, a_k)$, 用最小二乘法估计参数 $a_1, a_2, \dots, a_k$, 即选择参数 $a_1, a_2, \dots, a_k$ 使得 η 的观测值 y_i 与相应函数值 $\mu(x_i; a_1, a_2, \dots, a_k)_{i=1,2,\dots,n}$ 的离差平方和达到最小。

$$S = \sum_{i=1}^n [y_i - \mu(x_i; a_1, a_2, \dots, a_k)]^2 \rightarrow \min \quad (8)$$

分别求上式对 $a_1, a_2, \dots, a_k$ 的偏导数, 并令其为零, 则可以得到如下方程组

$$\begin{cases} \sum_{i=1}^n [y_i - \mu(x_i; a_1, a_2, \dots, a_k)] \frac{\partial}{\partial a_1} \mu(x_i; a_1, a_2, \dots, a_k) = 0 \\ \sum_{i=1}^n [y_i - \mu(x_i; a_1, a_2, \dots, a_k)] \frac{\partial}{\partial a_2} \mu(x_i; a_1, a_2, \dots, a_k) = 0 \\ \dots\dots\dots \\ \sum_{i=1}^n [y_i - \mu(x_i; a_1, a_2, \dots, a_k)] \frac{\partial}{\partial a_k} \mu(x_i; a_1, a_2, \dots, a_k) = 0 \end{cases} \quad (9)$$

求解上述方程, 就可以求出参数 $a_1, a_2, \dots, a_k$ 的估计值。

解得

$$a_1 =$$

$$\frac{M \sum_{i=1}^M (x_i x_i' + y_i y_i') - (\sum_{i=1}^M x_i)(\sum_{i=1}^M x_i') - (\sum_{i=1}^M y_i)(\sum_{i=1}^M y_i')}{M \sum_{i=1}^M (x_i^2 + y_i^2) - (\sum_{i=1}^M x_i)^2 - (\sum_{i=1}^M y_i)^2} \quad (10)$$

$$a_2 =$$

$$\frac{M \sum_{i=1}^M (x_i' y_i - x_i y_i') - (\sum_{i=1}^M x_i')(\sum_{i=1}^M y_i) + (\sum_{i=1}^M x_i)(\sum_{i=1}^M y_i')}{M \sum_{i=1}^M (x_i^2 + y_i^2) - (\sum_{i=1}^M x_i)^2 - (\sum_{i=1}^M y_i)^2} \quad (11)$$

将式(10)和式(11)的计算结果带入式(7)中, 即可得到目标的缩放系数。

2.3 模板更新

求解当前缩放后目标区域的灰度直方图 $curHist$, 则新模板被定义为

$$newHist = (1 - \lambda) preHist + \lambda curHist \quad (12)$$

式中, $preHist$ 为当前的模板分布, $newHist$ 为下一帧处理将要用的模板灰度直方图分布。 λ 为经验值, 在此取 0.05。

3 整体算法描述

本文算法的整体实现过程如下:

- 1) 输入图片序列第一帧, 手动框定要跟踪的目标, 计算目标框内的归一化的灰度直方图, 提取目标区域的 SIFT 点;
- 2) 根据目标的灰度直方图, 在新一帧中, 应用 Mean-shift 算法求解目标的新位置坐标, 实现目标中心位置由前一帧位置 X_{i-1} 到当前帧位置 X_i 的跟踪;
- 3) 每隔 5 帧求解一次目标尺度缩放因子, 计算过程如下:
 - ① 在位置 X_i 处 2 倍于上一帧跟踪框大小的范围内提取 SIFT 特征点, 并根据得到的点建立二叉树;
 - ② 输入上一帧中跟踪框区域的目标特征点, 在此二叉树中进行匹配搜索, 得到匹配的点对;
 - ③ 应用最小二乘法求解目标缩放系数, 得到当前帧中跟踪框的大小, 并计算当前跟踪框中的目标分布归一化灰度直方图;
 - ④ 更新目标模型, 目标模型的更新公式为式(12);
 - ⑤ 在当前帧中, 以跟踪结果 X_i 为中心画框, 框的大小为最新得到的框大小;

- 4) 判断是否是最后一帧, 若是, 则算法结束; 否则, 返回 2) 继续计算。

4 实验结果及分析

为评估本文提出的算法的效果, 实验采用 5 组测试序列。图 2 给出了基于 SIFT 点匹配的最小二乘模型求解目标缩放的结果, 图 3 给出了本文算法跟踪结果与传统 Mmean-shift 跟踪结果的对比, 图 4、图 5、图 6 将本文算法的跟踪结果和 SIFT 跟踪的实验结果进行对比说明, 其中第一行给出的是 SIFT 跟踪结果, 第二行给出的是本文算法的跟踪结果。

图 1 给出了两个盒子序列目标缩放整个过程中目标实际大小和跟踪结果的目标大小的定量比较结果。可以看出整个跟踪过程中跟踪框大小随目标的大小变化而变化, 且变化比例相一致。

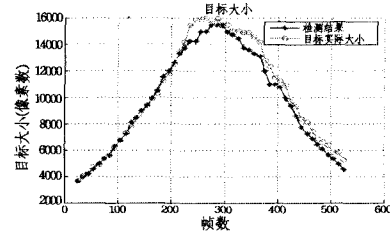


图 1 目标缩放时实际大小和检测结果

图 2 给出了两个盒子逐渐变大和逐渐变小情况的实验结果。算法在处理目标缩放时, 要求匹配到的 SIFT 点的个数大于一定的阈值(实验用的阈值为 10), 因此避免了由于每点对目标缩放系数的影响过大而出现误匹配时造成缩放结果的不稳定。结果显示, 跟踪框的大小随目标的尺度变化而实时变化, 且目标大小和跟踪框大小相吻合。

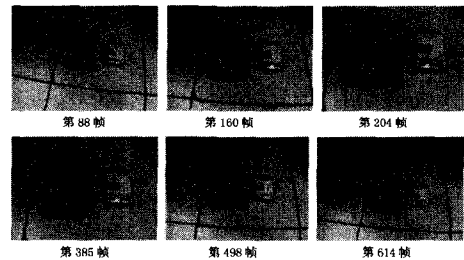
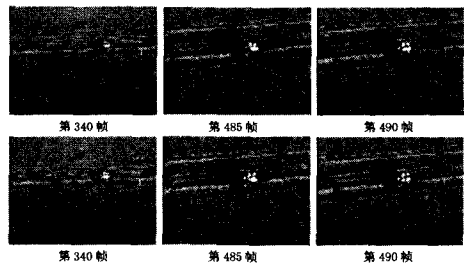


图 2 本文算法处理目标变大(第一行)和目标变小(第二行)的情况

图 3 给出了传统 Mean-shift 算法与本文提出算法的跟踪结果对比。由于传统 Mean-shift 算法没有模板更新策略, 当目标逐渐变大时, 出现了目标丢失现象。而本文算法采用了合理的模板更新策略, 实验效果良好。

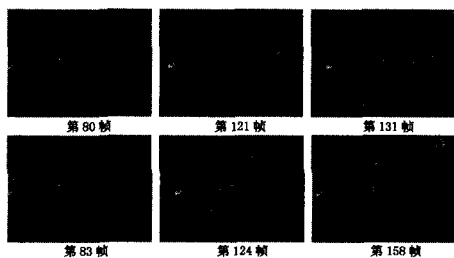


第一行为 SIFT 跟踪结果, 第二行为本文算法跟踪结果。

图 3 传统 Mean-shift 算法和本文算法对比实验

图 4 给出了目标穿越树林时出现部分遮挡时的跟踪效果。从图中可以看出, 采用 SIFT 算法进行跟踪时, 由于第

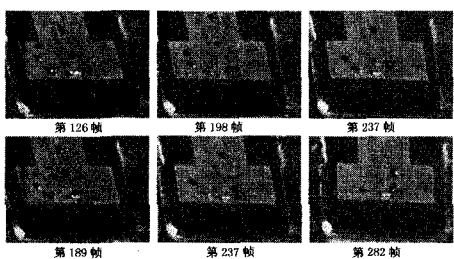
121 帧出现部分遮挡,当前模型变成了树的模型,以致在第 131 帧中出现目标丢失;而本文算法由于系统实时更新了目标模型,在这种情况下跟踪效果很好。



第一行为 SIFT 跟踪结果,第二行为本文算法跟踪结果。

图 4 跟踪目标部分遮挡情况

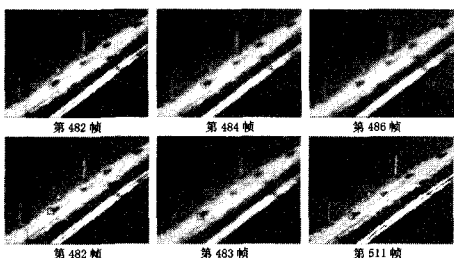
图 5 给出了当目标的颜色和背景色相接近时 6 辆车序列的跟踪效果。可以看到第 237 帧中,出现了光照变化,此时由于强光照射,目标的信息变得非常单一。在用 SIFT 处理跟踪时,由于取得的 SIFT 点数目非常少,已经不能支持跟踪的继续进行,到第 237 帧跟踪终止;而本文算法的跟踪效果良好。由于此序列中目标非常小,目标本身的 SIFT 特征点的个数达不到阈值,因此此情况下没有进行目标的缩放处理。



第一行为 SIFT 跟踪结果,第二行为本文算法跟踪结果。

图 5 跟踪目标与背景相似情况

图 6 给出了降分辨率时跟踪的效果。可以看到在此情况下,SIFT 算法跟踪在图像分辨率很低(第 486 帧)即到了左上角时丢失了目标;而本文提出的算法仍呈现出良好的跟踪性能。



第一行为 SIFT 跟踪结果,第二行为本文算法跟踪结果。

图 6 跟踪降分辨率情况

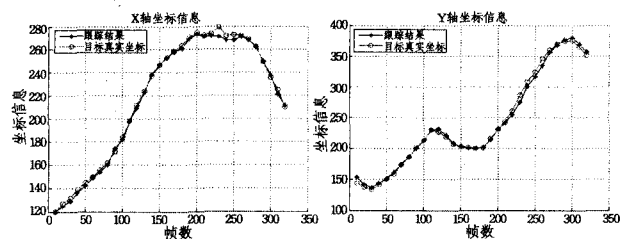


图 7 对应于图 5 序列,目标和背景相似情况下,跟踪结果和目标实际坐标比较

图 7 给出了 6 辆车序列跟踪到目标的坐标位置和目标实际坐标位置的对比。可以看出目标中心和跟踪框中心基本重合,实现了跟踪目标的准确定位。

结束语 本文提出了自适应的目标跟踪策略,解决了传统基于 Mean-shift 跟踪算法不能进行模板更新的问题。通过相邻两帧间匹配的 SIFT 点对,建立最小二乘模型,求解目标的缩放系数,然后将前一帧的目标模型和当前跟踪到的目标模型进行线性计算,得到新的目标模板。在求解匹配点对时,用向量夹角的余弦值进行相似性度量。从实验结果的定量分析中可以看出,算法实现了跟踪目标的准确定位,且跟踪框的大小和目标外接轮廓大小十分接近。实验结果表明,在处理部分遮挡、图像低分辨率、光照变化序列时,改进后算法跟踪性能良好,同现有跟踪算法相比,具有一定的优越性和鲁棒性。

由于 Mean-shift 算法求解的是一个局部极值,而非全局最优值,当目标被完全遮挡时容易出现目标丢失的现象。当目标比较小时,SIFT 特征点非常少,不能进行目标缩放的处理。如何解决目标遮挡情况下的跟踪和较小目标的缩放问题,是下一步研究的重点。

参考文献

- [1] Leputra W, Tan T, Lim F L. Non-overlapping Distributed Tracking Using Particle Filter[C]//International Conference on Pattern Recognition(ICPR). 2006:181-185
- [2] Fazli S, Pour H M, Bouzari H. Particle Filter-based Object Tracking with Sift and Color Feature[C]//International Conference on Machine Vision(ICMV). 2009:89-93
- [3] Comaniciu D, Meer P. Mean shift; a robust approach toward feature space analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619
- [4] Fashing M, Tomasi C. Mean-shift is a bound optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 471-474
- [5] Hu Xue-long, Tang Ying-cheng, Zhang Zheng-hua. Video object matching based on SIFT algorithm[C]// International Conference on Neural Networks and Signal Processing (ICNNSP). 2008:412-415
- [6] Lowe D G. Object recognition from local scale-invariant features [C]// International Conference on Computer Vision, ICCV. 1999:1150-1157
- [7] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60 (2): 91-110
- [8] Zhou Hui-yu, Yuan Yuan. Object tracking using SIFT features and mean shift[J]. Computer Vision and Image Understanding, 2009, 113(3): 345-352
- [9] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 2003, 25(5): 564-577
- [10] 杨心力, 杨恢先. 基于尺度不变特征变换的 Mean-Shift 目标跟踪[J]. 计算机应用, 2009, 28(7): 2678-2680
- [11] 吴刚, 唐振民, 耿烽, 等. 引入前帧加权采样的粒子滤波目标跟踪算法[J], 计算机科学, 2009, 36(8): 215-216