

基于信噪比最大化的目标自适应跟踪

王江涛 杨静宇

(南京理工大学计算机系, 南京 210094)

摘要 针对采用静态目标表达模型容易导致跟踪失败的缺点, 本文提出一种基于自适应特征生成模型的目标跟踪方法。在该方法中, 将目标看作为有效跟踪信号, 而背景则为随机噪声, 在对目标的似然图像进行构建后, 采用局部信噪比对当前目标所处特征空间的可跟踪性进行量化比较, 选用信噪比最大的表达模型作为当前的特征跟踪模型。基于 Mean Shift 的目标跟踪实验表明, 与采用静态目标表达模型相比, 文中的算法具有更好的鲁棒性和可行性。

关键词 自适应特征模型 信噪比 目标跟踪 mean shift

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)01-0089-06

Adaptive Object Tracking Based on SNR Maximizing

WANG Jiang-tao, YANG Jing-yu

(Department of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094)

Abstract According to the poor tracking ability adopting static feature model, an adaptive feature generating model based tracking program is present. In this program, the object is valid tracking signal, on the contrary, the background is noise. Constructing the likelihood maps, a local SNR (Signal Noise Ratio) is computed to evaluate the tracking ability of current feature space, and the feature space with maximal SNR is selected as the optimal tracking feature space. Object tracking results based on mean shift demonstrated that the proposed method is more robust and feasible than the classical one.

Keywords adaptive feature model, SNR, object tracking, mean shift

1 引言

目标跟踪是计算视觉领域中一个极具挑战性的课题, 其已被广泛地应用在人机交互、智能监控、医学图像处理等领域中, 目标跟踪的本质是在图像序列中检测出目标的同时对其进行精确定位。在基于视觉的目标跟踪中, 特征空间的选择对目标跟踪的准确性具有直接的影响, 当前广泛地采用颜色特征对目标进行表达, 在对目标执行跟踪之前, 首先选择一种固定的颜色空间, 用来建立目标的表达模型, 然后在后续序列图像中搜寻与目标表达模型最匹配的位置, 作为目标当前所处的位置; 采用静态颜色表达模型, 具有计算量小的优点, 且有利于目标的实时跟踪; 然而在实际中由于受到光照地变化、物体本身的多变

性等因素的影响, 目标物体的颜色特征难以采用静态颜色模型来精确表达, 甚至会导致目标跟踪失败。

为了提高目标跟踪的鲁棒性, 一些学者提出了在高维空间中对目标进行跟踪的方法, 文献 [1]、[2] 中的特征即为最优跟踪特征, 首先设定可行的特征空间集为 RGB 三原色的 49 种线性组合, 通过量化对不同特征空间的跟踪性能进行比较, 选用性能最优的 5 个空间对目标进行跟踪。文献 [3] 中, 延续了文献 [1] 的思想, 采用 Haar 特征和粒子滤波对最优跟踪特征进行选择, 在文献 [4]、[5] 中, Bohyung Han 等人将特征空间进一步扩展到归一化 RGB 空间, 获得了比文献 [1] 更好的结果。

本文中, 将目标看作为有效信号, 而背景为随机噪声, 提出了一种基于信噪比最大化的目标特征生成准则, 以能够自适应的对特征模型进行更新, 使跟

基金项目: 国家自然科学基金项目 (60632050, 60472060)

收稿日期: 2006-06-26 改回日期: 2006-09-01

第一作者简介: 王江涛 (1979~), 男, 现为南京理工大学模式识别与智能系统专业博士生。主要研究方向为图像分析、视频目标的检测与跟踪等。E-mail: jiangtaoking@126.com

踪系统能始终采用跟踪性能最优的特征空间对目标进行跟踪,同时通过不同视频序列的目标跟踪实验,对提出的算法的性能进行了验证。

2 基于信噪比最大化的特征生成模型

本文的目标是在构建一个自适应跟踪特征生成模型的基础上,使跟踪系统可以自动地选择当前的最优跟踪特征空间,从而连续的对目标特征表达模型进行更新。假定最有效的跟踪特征同时也是对目标和背景最具区分性的特征,给定目标的初始位置,就可以通过一个预先确定的目标表达模型计算出目标和背景的采样特征值的分布,进而对所有特征空间的可分性进行比较。需要指出的是用来进行目标跟踪的特征只是局部区分性的,也就是说目标只需要从当前周围背景中高鉴别性的区分出来即可。由于直方图对不同的视角、遮挡以及非刚体运动均具有较好的鲁棒性,在本文中采用了局部窗口颜色直方图对目标进行表达。众多的实验表明,在通常情况下归一化后的 RGB 颜色空间比 RGB 颜色空间具有更好的鉴别性能^[4],因此本文采用归一化后的 RGB 颜色空间作为目标特征的表达空间。

2.1 似然图像在特征跟踪性能分析中的应用

由于在实际应用中目标和背景往往具有多种颜色分布,因此无法采用通用的线性鉴别分析来寻找将目标和背景分离的最佳投影空间,此时,可以经验的对每一个候选特征空间进行评价,以决定产生较好的类分离性的特征空间。对于一个给定的特征空间,可以在估计目标和背景在该特征空间中的特征分布的基础上,计算这些分布的对数似然比,然后采用量化手段来对目标和背景的可区分能力进行评价。

为了对不同特征空间中目标和背景的跟踪性能进行比较,首先对目标和背景进行像素采样,在初始图像中人工确定目标的位置,本文中采用矩形区域来对目标所处的位置进行覆盖,然后,在目标矩形区域周围,选择一环绕目标的较大矩形区域内的像素作为背景像素。这种方法可以使鉴别特征将目标从背景中分离出来,而不受机动目标方向的影响。

给定一特征空间 F , 令 $H_{obj}(i)$ 为目标区域在该特征空间下的直方图, $H_{bg}(i)$ 则为背景样本在该特征空间下的直方图, i 为直方图的级别 (包含的元素数), $i = 1, 2, \dots, z$ 。通过对直方图进行归一化来构建目标和背景的经验概率密度 $\phi_{obj}(i)$ 和 $\phi_{bg}(i)$:

$$\phi_{obj} = H_{obj}(i) / N_{obj} \quad (1)$$

$$\phi_{bg} = H_{bg}(i) / N_{bg} \quad (2)$$

这里, N_{obj} 和 N_{bg} 分别为目标和背景的样本数, $\phi_{obj}(i)$ 和 $\phi_{bg}(i)$ 即为在对应的直方图中第 i 个分量所出现的频率。

对直方图进行归一化后,特征值 i 所对应的对数似然比 $L(i)$ 可以由下式给出:

$$L(i) = \max \left(\frac{1}{m}, \log \frac{\max(\phi_{obj}(i), \delta)}{\max(\phi_{bg}(i), \delta)} \right) \quad (3)$$

这里 δ 是为了避免分母为 0 或者似然比为 0 而设的一个较小的数值。在求出所有特征值的对数似然比后,将其投影到相对应的特征值上即得到该特征空间下的似然图像,这样目标所处的高亮度区域即可通过确认似然比的大小来检测。

2.2 特征选择准则

如果将背景看作为噪音,前景作为有效的跟踪信号,则信噪比表征了该特征空间下目标的可跟踪性能。在获得跟踪图像的似然图像后,令在第 k 个特征空间下的信噪比为

$$SNR(k) = \frac{\frac{1}{N_{obj}} \sum_{j \in obj} L(j)}{\frac{1}{N_{bg}} \sum_{m \in bg} L(m)} \quad (4)$$

其中, obj 和 bg 分别表示目标区域和背景区域。在求出所有特征空间的信噪比后对其进行排序,信噪比最大的特征空间即为当前的最优跟踪特征空间:

$$F_{opt} = \arg \max(SNR(k)) \quad k = 0, 1, \dots, r \quad (5)$$

这里 r 为当前特征集合中所包含的特征空间数。

如图 1 所示,目标物体为图中小矩形框所包围的区域而背景区域则为围绕目标区域的大矩形框与小矩形框之间所含的部分,图 2 给出了目标和背景



图 1 目标区域和背景区域

Fig 1 Forward region and background region

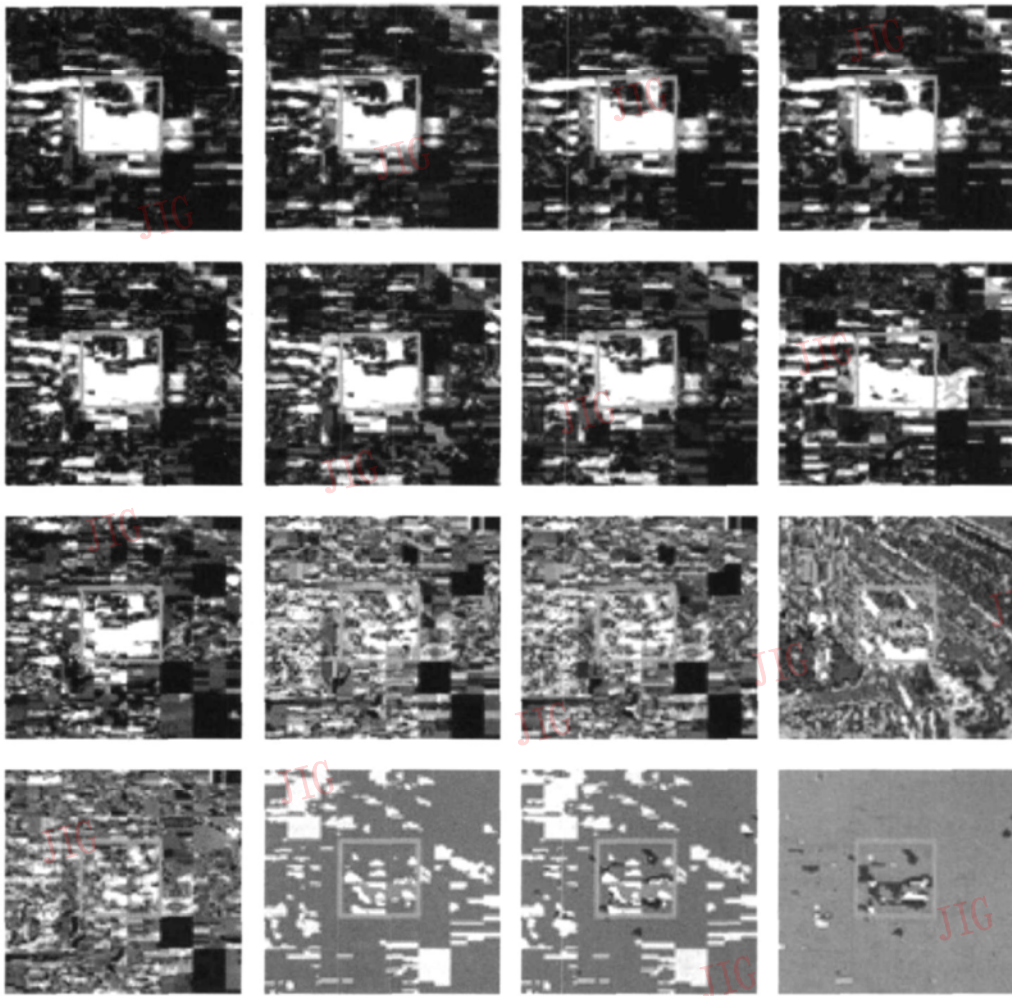


图 2 目标区域的似然图像

Fig 2 Likelihood maps for forward region

区域(即图 1 中大矩形框所包含区域)在 16 种不同的特征空间中所构建的似然图像. 这些似然图像按照信噪比的大小由左至右、由上到下递减排列. 图中高亮度的部分为目标, 低亮度的部分则为背景, 从似然图像可以看出, 不同的特征空间对目标的识别能力差别很大, 因此选择目标表达能力强的特征空间是对目标进行有效跟踪的关键问题.

3 算法的流程和实现

将上述的特征生成模型嵌入到 Mean shift 跟踪系统中, Mean shift(均值偏移)算法^[6]作为一种基于梯度分析的非参数优化算法最先由 Fukunaga 和 Hostetler 提出, 其后 Comaniciu 成功地将其应用到计算机视觉领域中^[7,8], 近年来, Mean shift 算法以其无需参数、快速模式匹配的特性被广泛应用到目标跟踪领域.

采用 Mean shift 算法进行目标跟踪时, 首先在

第 1 帧图像中对目标模型进行初始化, 然后在相邻图像序列中对目标模型所在位置的周围进行搜索, 使相似函数最大化的候选区域即为跟踪目标在当前帧中所处的位置.

令 $\rho(\mathbf{y}) \equiv \rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}]$ 为模型图像直方图 $\hat{\mathbf{p}}$ 和候选图像中中心点为 \mathbf{y} 的直方图 $\hat{\mathbf{q}}$ 间的相似度量函数, $\rho(\mathbf{y})$ 表征了模型图像中和核直方图和候选图像中核直方图的相似性, 从而 $\rho(\mathbf{y})$ 的局部极值所处的位置即为目标所处的位置. 采用 Bhattacharyya 系数作为 $\hat{\mathbf{p}}$ 和 $\hat{\mathbf{q}}$ 间的相似度量函数, 有

$$\rho(\mathbf{y}) \equiv \rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}] = \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}) \hat{q}_u} \quad (6)$$

对 $\rho(\mathbf{y})$ 展开泰勒级数, 取其前两项进行逼近.

$$\rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0) \hat{q}_u} + \frac{1}{2} \sum_{u=1}^m \hat{p}_u(\hat{\mathbf{y}}) \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}} \quad (7)$$

将上式重新写为如下形式:

$$P[\hat{p}(y), \hat{q}] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(\hat{y}_0) \hat{q}_u} + \frac{C_h}{2} \sum_{u=1}^m w_i k \left\| \frac{y - x_i}{h} \right\|^2 \quad (8)$$

其中, w_i 为加权系数。

$$w_i = \sum_{u=1}^m \frac{\hat{q}_u}{\sqrt{\hat{p}_u(\hat{y}_0)}} \delta[b(x_i) - u] \quad (9)$$

由 mean shift 算法可得偏移后的坐标位置如下:

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g \left\| \frac{\hat{y}_0 - x_i}{h} \right\|}{\sum_{i=1}^{n_h} w_i g \left\| \frac{\hat{y}_0 - x_i}{h} \right\|} \quad (10)$$

求得 \hat{y}_1 后, 就完成一次均值偏移, 随后可令 $\hat{y}_0 \leftarrow \hat{y}_1$ 进行下一次均值偏移; 由均值偏移的性质可知, 其总是向 Bhattacharyya 系数的局部极值 (峰值) 移动, 因此当核窗口的位置移动到峰值时 $\hat{y}_0 \approx \hat{y}_1$, 此时核窗口将在目标所处位置收敛, 从而完成对目标的定位。

在 mean shift 跟踪系统中嵌入基于信噪比最大化的特征空间生成模型后, 此时在每一帧对目标进行跟踪前, 首先对当前帧目标所处的位置进行似然图像化, 然后求取信噪比最大的特征空间作为下一帧的跟踪空间, 修改后的整个算法流程如下:

(1) 在当前目标位置, 对特征空间集合包含的所有特征空间求解目标和背景的特征分布, 利用公式 (3) 求取目标和背景的对数似然比, 进而构建其似然图像。

(2) 由公式 (4) 求取所有似然图像内目标和背景的信噪比, 并对其进行排序, 选择使信噪比最大的特征空间作为当前跟踪空间。

(3) 对下一帧图像执行目标跟踪, 帧序号加 1, $N \leftarrow N + 1$ 。

(4) 在当前位置求取所选择特征空间中的特征分布, 执行 mean shift 跟踪算法。

(5) 跟踪是否结束? 是: 转步骤 (6); 否: 转步骤 (1)。

(6) 跟踪结束。

4 实验结果与分析

在进行跟踪实验之前, 首先对背景区域的选择进行分析。显然, 背景区域只有包围在目标区域周围时, 目标和背景之间的可区分性才有意义, 因此, 只需要对背景区域的尺寸进行选择。令背景和背景矩形

区域尺度比为 rat 图 3 给出了不同尺度背景区域情况下的信噪比, 从图 3 可以看出, 背景区域的大小变化对信噪比的影响很大; 当 rat 取 2 时, 具有较大的信噪比且不同特征空间之间的信噪比差别较大, 较好的反映了不同特征空间的局部区分能力; 当 rat 取 6 时, 不同特征空间之间的信噪比维持在 1 附近, 并且各信噪比之间差异较小, 这说明对目标和背景的区分能力减弱。实验中发现, 当背景和背景矩形区域长、宽比为 1:3 时取得了较好的效果。

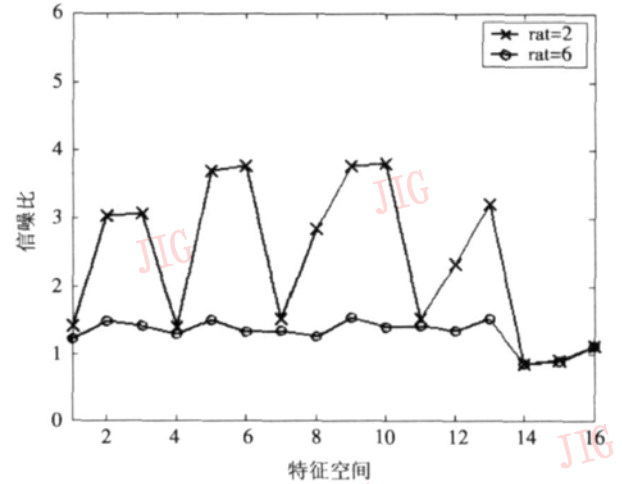


图 3 不同背景区域下的信噪比

Fig 3 SNR with different background regions

为了验证所提出的算法的可行性和有效性, 对不同的图像序列进行跟踪实验, 在每个实验中分别采用本文中的算法 (简称为算法 1) 和采用静态目标表达模型的算法 (简称为算法 2) 对目标进行两次跟踪, 然后对两种算法的跟踪结果进行比较, 以更好地表明算法的性能。其中算法 1 的目标表达特征空间集合为归一化的 RGB 经过 1 维代数组合:

$$U_1 = \{w_1 r + w_2 g + w_3 b \mid w \in [-1, 0, 1]\} \quad (11)$$

以及 2 维组合后颜色空间 rg, gb, rb 。

在这里 r, g, b 为归一化后的 R、G、B 的分量:

$$\begin{aligned} r &= R / (R + G + B) \\ g &= G / (R + G + B) \\ b &= B / (R + G + B) \end{aligned} \quad (12)$$

算法 2 则采用了通用的 2 维 rg 空间作为目标表达空间, 两种算法均采用相同级别的直方图 (255 级), 在跟踪时 mean shift 算法的迭代次数也均相同。算法 1 中目标和背景矩形区域长、宽比均为 1:3。

图像序列 1 为采用固定摄像机拍摄的交通场景序列, 该序列中由于受下雪的影响使图像变得模糊, 图 4 给出了序列中第 98、146、200 帧 (从左到右) 采

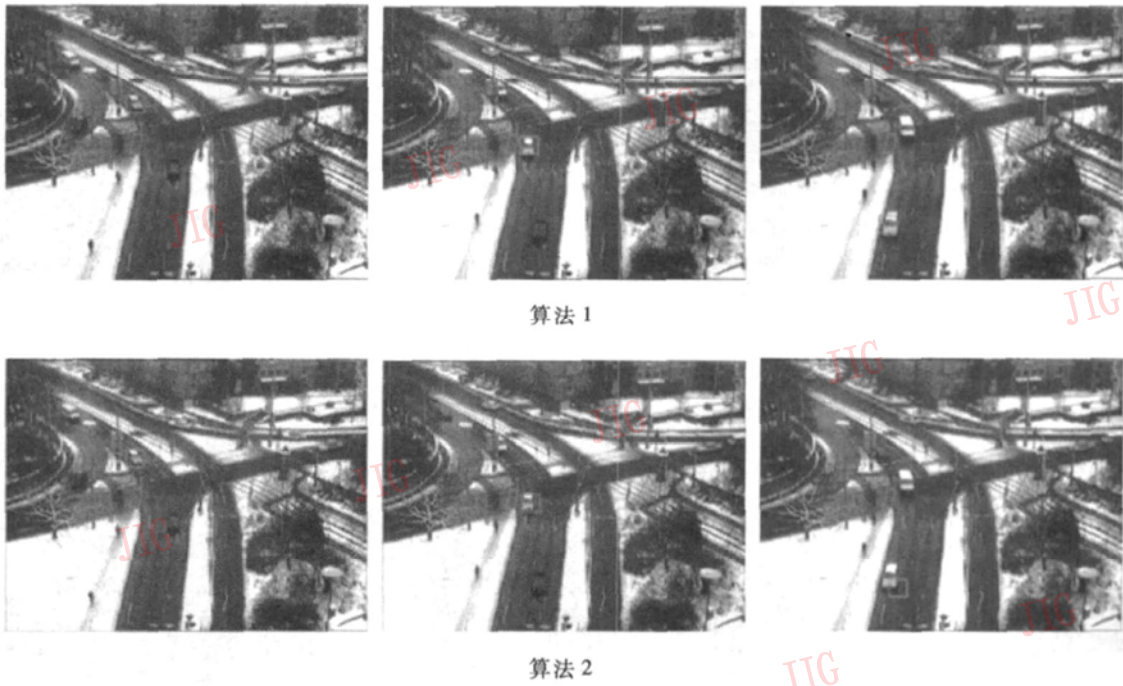


图 4 图像序列 1 的跟踪结果

Fig. 4 Tracking results of sequence 1

用上述两种算法进行跟踪的结果, 其中第 1 列为算法 1 的跟踪结果, 第 2 列为算法 2 的跟踪结果, 从图 4 中可以看出, 随着跟踪时间的不断延长, 算法 1 保持了较为准确的跟踪结果, 而算法 2 则逐渐偏离目标物体。图像序列 2 取自 *pets04* 视频数据库, 跟踪目标为图像中的行人, 图 5 给出了该序列中第 1 866、1 874、1 894、1 962 帧的跟踪结果; 由于行人衣服与背景具有较为相似的颜色分布, 采用算法 2 进行跟踪时跟踪窗口快速偏离目标, 最终造成目标丢失, 而算法 1 能够自动地选取当前跟踪性能最优的

特征空间, 从而较为准确的对目标进行了跟踪。序列 3 则为一航拍视频序列, 在该序列中由于摄像机的位置不断地变化且跟踪目标较小, 增加了跟踪的难度, 图 6 给出了序列中第 1 486、1 566、1 766 帧的跟踪结果, 从跟踪结果中可以看出, 算法 1 维持了较好的跟踪性能, 而算法 2 则出现了较大的跟踪误差。

上述不同视频序列的跟踪结果表明, 通过自适应地选取最优跟踪特征空间, 本文提出的算法提高了跟踪系统的鲁棒性和可行性, 较好地解决了不同背景状态下的目标跟踪问题。

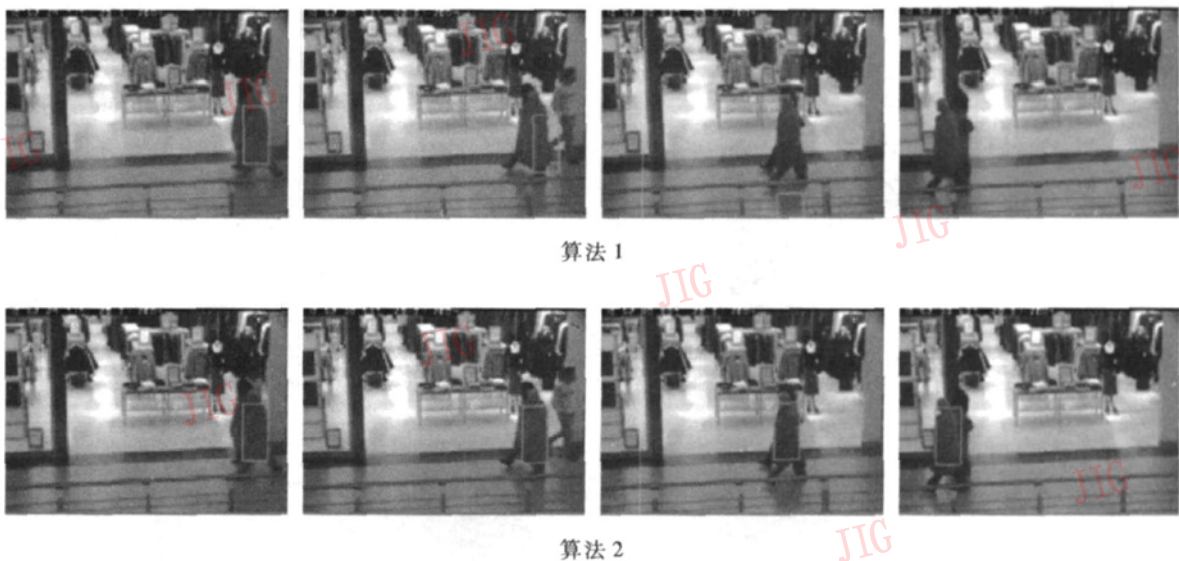


图 5 图像序列 2 的跟踪结果

Fig. 5 Tracking results of sequence 2

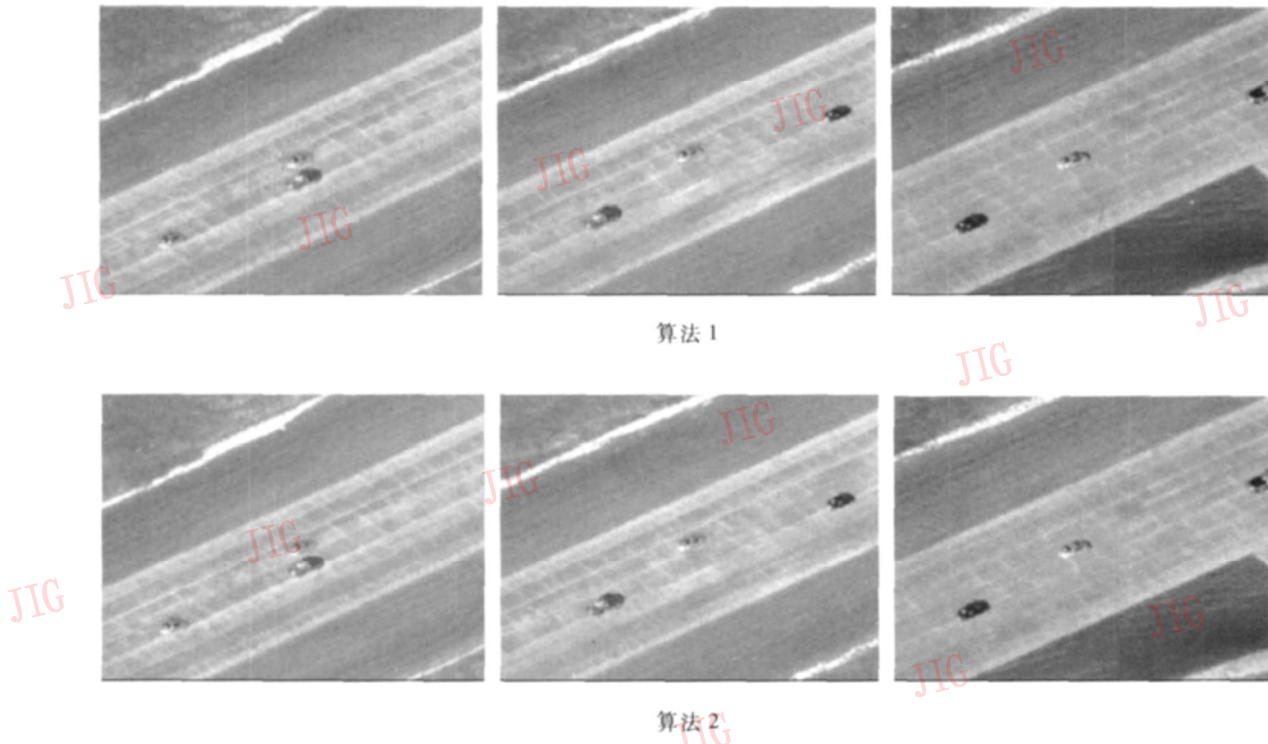


图 6 图像序列 3 的跟踪结果
Fig.6 Tracking results of sequence 3

5 结 论

本文提出了一种新的基于自适应特征选择的目标跟踪方法。通过引入信噪比对不同特征空间的跟踪性能进行比较,使目标能够随着背景或自身特征的变化自适应地对当前跟踪特征模型进行更新,以选择当前的最优特征空间对目标进行跟踪。实验结果表明,该算法很好地解决了采用静态目标表达模型难以对目标进行准确跟踪的问题,提高了跟踪算法的鲁棒性,使其可以适应于更多的不同场合。

参考文献 (References)

- Collins R T, Liu Yan-xi, Leordeanu M. Online selection of discriminative tracking features[A]. In Proceedings of Ninth IEEE International Conference on Computer Vision[C], Nice, France, 2003: 346~352.
- Collins R T, Liu Yan-xi, Leordeanu M. Online selection of discriminative tracking features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1631~1643.
- Wang Jian-yu, Chen Xin-lin, Gao Wen. Online selecting discriminative tracking features using particle filter[A]. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C], San Diego, CA, USA, 2005, 2: 1037~1042.
- Bohyung Han, Davis L. Object tracking by adaptive feature extraction [A]. In Proceedings of the 11th International Conference on Image Processing[C], Singapore, 2004, 3: 1501~1504.
- Bohyung Han, Davis L. Robust observations for object tracking[A]. In Proceedings of the 12th International Conference on Image Processing[C], Genoa, Italy, 2005, 2: 442~445.
- Fukunaga K, Hostetler L D. The estimation of the gradient of a density function, with applications in pattern recognition[J]. IEEE Transactions on Information Theory, 1975, 21(1): 32~40.
- Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564~577.
- 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.