

一种鲁棒高效的视频运动目标检测与跟踪算法

刘少华¹ 张茂军¹ 熊志辉¹ 陈旺¹

摘要 提出了一种视频运动目标的快速检测和稳定跟踪算法. 目标检测使用减背景法, 用均值法构造背景图像, 提出一种基于熵能和广义高斯分布的局部自适应阈值选取算法, 可有效克服噪声的影响. 采用基于特征匹配的目标跟踪方法, 提出一种 LICS (Logarithm illuminance contrast statistic) 特征, 该特征能够更加充分有效地表征目标, 可在光照和目标姿态变化的情况下实现刚体目标的稳定跟踪. 使用 Kalman 滤波限制搜索匹配范围以减小计算量. 用目标子区域匹配的方法解决目标相互遮挡时的跟踪问题. 实验结果表明, 该算法在运动目标检测效果、跟踪稳定性和运行时间方面都有良好的性能.

关键词 目标检测与跟踪, 熵能, 广义高斯模型, 自适应阈值
中图分类号 TP391

A Robust and Efficient Video Moving Object Detection and Tracking Algorithm

LIU Shao-Hua¹ ZHANG Mao-Jun¹ XIONG Zhi-Hui¹ CHEN-Wang¹

Abstract A simple and efficient moving object detection and tracking algorithm is proposed. The object detection is based on the background subtraction method; an adaptive local threshold selection method on the use of entropy power and GGD (Generalized Gaussian distribution) is proposed to get over the noise influence. Feature based tracking method is used in object tracking. A feature named LICS (Logarithm illuminance contrast statistic) is proposed, which can effectively represent the objects' appearance. Tracking of rigid objects by LICS is stable when the objects' illumination and posture are variable. The Kalman filter is used to restrict the search window and reduce the calculation. A sub-block matching algorithm is used to handle the objects occlusion. The experimental results show that the proposed algorithm has good performance.

Key words Object detection and tracking, entropy power, generalized Gaussian distribution (GGD), adaptive threshold selection

视频运动目标的检测与跟踪是智能视频处理的基础问题之一, 它为后续的高级处理如目标分类、行为理解等提供支持, 在军事和社会安全领域有着广泛的应用前景.

1 相关研究

目标检测的挑战来自于如何克服自然环境变化(光照变化、树叶的摇晃、雨雪、水面等)和成像器材(电子噪声、摄像机的晃动等)的影响. 目前最主要的检测方法为减背景法, 其核心是使用合适的模型来描述场景的变化. 常用的描述模型有混合高斯模型、非参数模型、码本模型等. 总体来说, 减背景法的检测效果与其计算复杂度成正比, 在实际应用时需要在二者间寻找折中方案. 本文采用运行期均值法构造简单的背景模型, 通过更加有效的阈值选取算法来弥补其易受噪声影响的缺点.

阈值选取是目标检测的关键, 既要保证能够检测出重要信息, 又能够抑制噪声的影响. 阈值的选取方法主要有三类: 全局阈值法、局部阈值法和动态阈值法. 香农的熵定理为阈值选取提供了有效的解决思路. 文献 [1] 所选阈值与图像的熵能成正比, 可用于图像边缘提取和运动目标检测. 文献 [2] 对二值化后的差分图像进行分块, 根据每个块的信息熵来确定该块是否为前景块. 文献 [3] 使用最小交叉熵函数求取阈值, 使差分图像和二值图像间的信息量差异最小化. 以上方法中, 文献 [1] 方法对于不同视频必须手动调整阈值选取参数, 以适合不同的噪声强度, 所计算出的阈值应用于整幅图像. 在真实视频里, 噪声并不是均匀分布于整幅图像的, 如背景是树的区域噪声就大于背景为建筑物的区域, 采用相同阈值并不能得到最优的分割. 文献 [2] 虽然计算局部信息熵, 但采用的也是全局阈值, 而且阈值也必须通过实验手工选取. 文献 [3] 方法虽然理论最优但实际上也存在上述问题, 而且计算量很大. 本文采用熵能原理根据背景图像序列获得分割阈值的训练样本, 并使用广义高斯分布描述样本分布, 从而自动计算出与场景中的每个局部区域噪声相匹配的局部阈值, 从而尽可能地抑制噪声得到优秀的分割结果.

运动目标跟踪要解决的主要问题是鲁棒性、准

收稿日期 2008-01-04 收修改稿日期 2008-12-09
Received January 4, 2008; in revised form December 9, 2008
国家自然科学基金 (60705013, 60773023) 资助
Supported by National Natural Science Foundation of China (60705013, 60773023)
1. 国防科学技术大学信息系统与管理学院 长沙 410073
1. College of Information System and Management, National University of Defense Technology, Changsha 410073
DOI: 10.3724/SP.J.1004.2009.01055

准确性和快速性^[4-5]. 其中准确性主要是指目标检测和分割的准确度. 影响跟踪鲁棒性的主要因素来自于目标姿态变化、环境光照变化和目标的相互遮挡. 目前的主要跟踪方法主要有四类^[4,6], 分别是基于模型、基于区域、基于变形模板和基于特征的方法. 其中, 基于变形模板的方法如 ACM (Active contour model) 法相对更加适合单目标的跟踪; 基于 3D 模型的方法可较好解决目标的姿态变化, 但其模型获取困难并且计算量巨大; 基于区域的方法采用目标局部区域的纹理或色彩特征进行跟踪, 难以解决遮挡问题, 同时对光照变化也很敏感; 基于特征的方法与基于区域的方法类似, 但其更关注目标的总体特征, 有一定的抗遮挡性.

本文采用基于特征的方法实现目标的跟踪, 其关键在于如何确定目标的特征向量. 一些文献使用多种简单特征构成多维特征向量, 如文献 [7] 采用目标的质心、外接椭圆的 12 个控制点和颜色直方图构成特征向量; 文献 [8] 采用目标的位置、速度、颜色直方图和包围盒的大小构成特征向量. 类似的方法还有文献 [9-11] 等. 当特征选取较多时会导致系统效率降低, 而且容易出错, 实际应用中并不必要, 具体特征的选取可参考文献 [12]. 文献 [13-14] 使用边缘特征进行跟踪, 可较好地克服光照的影响, 但边缘的精确运动估计比较困难, 而且在复杂的背景下难以将运动目标和背景区分开来. 文献 [15] 通过计算目标区域的主颜色直方图作为特征进行跟踪, 颜色特征的优势在于受目标形状变化的影响较小, 但对环境光照的变化比较敏感, 要获得好的结果还必须研究色彩稳定算法. 针对以上问题, 本文提出一种 LICS (Logarithm illuminance contrast statistic) 特征, 将光照模型引入到 CCH (Contrast context histogram) 特征^[16] 中, 可减弱光照变化的影响, 同时具备一定的仿射不变特性. 采用 LICS 特征结合 Kalman 滤波对目标进行跟踪, 可有效地克服刚体目标的姿态变化和光照变化对跟踪稳定性的影响, 实现非遮挡情况下的鲁棒跟踪. 使用局部区域特征进行匹配解决遮挡问题.

2 基于熵能的局部自适应阈值选取算法

2.1 基于熵能的阈值选取原理

根据香农定理, 一个离散信号 O 的熵 H 定义为:

$$H = - \sum_{i=i_0}^{i_M} p_i \ln p_i \quad (1)$$

熵能 N 的计算公式为:

$$N(O) = \frac{1}{2\pi e} \exp(2H)$$

高斯信息源 $G(0, \sigma)$ 的信息熵为:

$$H(G) = \ln(\sigma\sqrt{2\pi e})$$

熵能为:

$$N(G) = \sigma^2$$

香农指出, 高斯噪声能吸收其他叠加到他身上的熵能小的信号. 此时, 一个信号的熵能正比于白噪声熵能与信号本身熵能的和. 文献 [1] 将此原理应用到图像处理中, 图像 S 可以看作前景信息 X 和高斯噪声 G 的叠加 $S = X + G$. 则信息熵的标准差可估计为:

$$\sigma = \sqrt{N(S)} = \frac{\exp(H)}{\sqrt{2\pi e}} \quad (2)$$

目标分割的阈值 θ 和信号熵的标准差成正比, 即:

$$\theta = k \cdot \sigma \quad (3)$$

文献 [1] 中, k 的值固定取 4、5 或 6.

2.2 基于自适应局部阈值的目标检测算法

当场景中无运动目标时, 差分图像 S 的像素灰度值 $O(S)$ 应满足: $O(S) < \theta$, 即背景图像差分图中的每个像素值都小于阈值 θ , 由式 (3) 不难得出 $k > O(S)/\sigma$. 根据此原理, 可利用背景图像序列的差分图对 k 进行训练, 以实现阈值的自适应选取. 由于图像中的噪声分布并非均匀, 采用统一阈值并不符合实际情况, 因此将差分图像分成 R 个大小为 $L \times L$ 的像素块, 每个像素块都有自己的局部自适应参数 k . 差分图像的获取采用减背景法, 使用运行期均值法^[17] 来构建参考背景图像.

根据背景图像序列的差分图训练像素块阈值参数 k 的方法如下:

- 1) 统计差分图像 S_i 的灰度直方图, 按照式 (1) 计算 S_i 的熵;
- 2) 按照式 (2) 计算 S_i 的信息熵标准差 σ_i ;
- 3) 对于 S_i 中的第 r 个像素块 B_r , 求取其在第 i 帧的训练参数 $k_i = \max(B_r)/\sigma_i$, 其中 $\max(B_r)$ 表示像素块 B_r 的灰度最大值.

按照以上的步骤, 经过 f 帧 ($f > 30$) 背景图像差分图的训练后, 对每个像素块都可得到一个 k 值序列 (k_1, k_2, \dots, k_f) . 根据噪声的分布特点, 可将 k 值序列看作一个类高斯分布, 为了更好地描述 k 值序列, 我们采用广义高斯分布对其进行描述, 可得到分布 K 的概率分布函数为:

$$F(k) = \int_{-\infty}^k \frac{\alpha}{2\beta\Gamma(1/\alpha)} e^{-|t-\mu|^\alpha/\beta} dt$$

其中, $\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$, $\Gamma(\cdot)$ 是 Gamma 函数. α 为广义高斯分布的形状参数, 当 $\alpha = 2$ 时, 广义高斯分布特例化为高斯分布, 当 $\alpha = 1$ 时, 特例化为拉普拉斯

分布, 当 $\alpha \rightarrow \infty$ 时, 则趋近于平均分布. 由于形状参数的存在, 使得广义高斯分布相比于高斯分布和拉普拉斯分布能更好地描述样本.

我们希望设定阈值 k_r 使得事件 $K < k_r$ 的概率尽可能地接近 1. 令 $\lambda = (k_r - \mu)/\sigma$, $F(\lambda)$ 的值仅与 λ 和 α 的取值有关, 当 $\lambda = 3$ 时, 对任意的 α 值 ($\alpha \geq 0.6$) 都有 $F(\lambda) > 0.99$, 可满足我们的要求. 相应的, 有 $k_r = \mu + \lambda\sigma$. 根据文献 [18] 的推导, 可求得参数 σ 的矩估计为:

$$\sigma = \hat{m} \frac{\sqrt{\Gamma(3/\alpha)\Gamma(1/\alpha)}}{\Gamma(2/\alpha)}$$

其中, \hat{m} 为分布 K 的一阶绝对矩. 在计算时, λ 的值取 3, 形状参数 α 的计算参考文献 [18]. 由于 k_r 的值与 α 成反比, 为了防止得到的阈值过大, 我们强制使 $\alpha \geq 0.6$.

对后续的图片序列, 采用如下的步骤进行目标检测:

- 1) 计算差分图像 S_j 的信息熵标准差 σ_j ;
- 2) 按 (4) 式进行分割, 得到二值化图像 I ;

$$I(s) = \begin{cases} 1, & \text{如果 } s \in B_r \text{ 并且 } O(s) < k_r\sigma_j \\ 0, & \text{否则} \end{cases} \quad (4)$$

3) 求取每个像素块在当前帧的阈值参数, $k_j = \max(B_r)/\sigma_j$;

4) 对满足 $k_j \leq k_r$ 的像素块, 更新其广义高斯分布的相关参数. 更新公式为:

$$\mu_{j+1} = (1 - \tau)\mu_j + \tau\max(B_r)$$

$$\hat{m}_{j+1} = (1 - \tau)\hat{m}_j + \tau|\max(B_r) - \mu_{j+1}|$$

然后根据更新后的参数计算像素块 B_r 新的形状参数 α 和新的阈值参数 k_r .

3 运动目标跟踪

运动目标跟踪采用我们提出的目标 LICS 特征. LICS 特征是对 CCH 特征的改进, CCH 特征由 Huang 等 [16] 提出, 相对于 SIFT (Scale invariant feature transform) 特征, 其计算简单并具有 SIFT 特征相似的仿射不变性. 文献 [19] 对其进行简化并用于目标检测中背景模型的构建, 取得了较好的效果. 本文改进了其计算方法, 使其更加适合目标总体特征的描述, 引入了光照模型, 以弱化光照变化对目标特征的影响, 得到 LICS 特征, 并使用该特征结合 Kalman 滤波进行目标跟踪.

3.1 运动目标 LICS 特征的计算

根据 Wu 等 [20] 提出的基于亮度对数的目标检测方法, 一个目标表面的亮度 $f(x, y)$ 是入射光

强 $i(x, y)$ 与目标表面反射系数 $r(x, y)$ 的乘积, 即 $f(x, y) = i(x, y) \times r(x, y)$. 根据式 (5) 的推导, 可发现通过求取对数, 能够剔除入射光强的影响.

$$\begin{aligned} \ln f(x_1, y_1) - \ln f(x_2, y_2) = & \\ & [\ln i(x_1, y_1) + \ln r(x_1, y_1)] - & (5) \\ & [\ln i(x_2, y_2) + \ln r(x_2, y_2)] = \\ & \ln r(x_1, y_1) - \ln r(x_2, y_2) \end{aligned}$$

根据上述原理, 本文提出 LICS 特征, 通过目标像素灰度对数与平均像素灰度对数差值的统计得到该特征, 求取方法如下文所述.

首先以目标质心为原点, 按角度将目标平均划分为 n 个子区域 (q_1, q_2, \dots, q_n) (图 1 中 $n = 8$, 每个子区域跨越角度为 $\pi/4$), 分别计算每个子区域的像素灰度平均值.

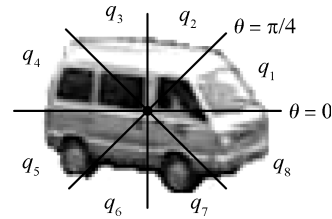


图 1 LICS 特征的目标子区域划分

Fig. 1 The segmentation of object for LICS

然后把目标区域中各像素灰度值 $L(P)$ 取对数, 与像素所在子区域的平均像素灰度 $L(P_{q_i})$ ($i = 1, 2, \dots, n$) 的对数值做差, 得到每个像素的 LICS 特征值, $C(P, P_{q_i}) = \ln(L(P)) - \ln(L(P_{q_i}))$.

最后, 统计每个子区域 q_i 中 LICS 特征值的正数和负数均值:

$$CH_{q_i}^+(P) = \frac{\sum\{C(P, P_{q_i}) | P \in q_i \text{ and } C(P, P_{q_i}) > 0\}}{N_{q_i}^+}$$

$$CH_{q_i}^-(P) = \frac{\sum\{C(P, P_{q_i}) | P \in q_i \text{ and } C(P, P_{q_i}) < 0\}}{N_{q_i}^-}$$

其中, $N_{q_i}^+$ 是在 q_i 子区域中正数的个数, $N_{q_i}^-$ 是在 q_i 子区域中负数的个数.

这样可得到一个 $2n$ 维的向量: $LICS(P) = (CH_{q_1}^+, CH_{q_1}^-, CH_{q_2}^+, CH_{q_2}^-, \dots, CH_{q_n}^+, CH_{q_n}^-)$, 即为我们想要的 LICS 特征向量. 实际使用中可根据运动目标面积的大小, 将目标子区域分为 4 个、8 个或 16 个, 从而在目标特征精度和计算量间获得平衡.

在跟踪时, 使用 Kalman 滤波器对已跟踪的运动目标做下一步的预测, 将预测位置 $w \times w$ 邻域内

的观测目标与已跟踪目标做 LICS 特征匹配. 对于运动稳定的目标, Kalman 滤波器的预测误差很小, 为保证对不同视频的适应性, w 通常在 4~8 之间选取. Kalman 滤波器的设定可参考文献 [11].

3.2 遮挡的处理

本文将目标分成多个子块, 通过子块的灰度匹配完成遮挡情况下的跟踪. 子块的划分方法为: 将目标的包围盒划分为 $m \times n$ 个子区域, 分别计算每个子区域中目标像素所占的比例, 如果大于 1/2, 则该区域为目标的一个子块. 在理论上 m 和 n 的取值越大则跟踪精度越高, 实际上若子块太小则其包含的纹理信息过少, 匹配结果容易出错, 因此应保证每个子块的大小不小于 8×8 . 对包含背景像素的子块, 在匹配过程中不计算背景像素.

对每个子块在 Kalman 预测位置的 $w \times w$ 邻域内做灰度模板匹配, 获得每个子块的最优位移量, 将子块中出现次数最多的位移量作为目标的位移量. 位移量与目标位移量在 x 或 y 方向上误差超过 2 个像素的子块被判定为遮挡.

为了减小计算量, 在正常跟踪时不进行分块及匹配操作, 而是采用 Kalman 滤波器的预测结果判断目标的包围盒在下一帧是否发生重叠, 如果是则按照遮挡处理, 进行分块匹配操作. 另外就是在跟踪失败时进行分块匹配来确认是否发生遮挡.

3.3 目标检测跟踪算法的整体流程

跟踪算法的整体流程见图 2. 对于输入的视频图像序列, 第 1 到 $i-1$ 帧背景序列用于训练阈值参数 k . 使用该参数对第 i 帧图像做目标检测, 提取目标区域的 LICS 特征, 将目标位置输入 Kalman 滤波器预测目标在 $i+1$ 帧的位置, 并更新阈值参数 k . 对第 $i+1$ 帧图像, 首先对其做目标检测和 LICS 特征提取, 然后将第 i 帧 Kalman 预测位置 $w \times w$ 邻域内的检测目标与被预测目标做 LICS 特征匹配, 如果匹配成功则正常跟踪, 如果失败有可能发生遮挡, 进行遮挡处理后得到跟踪结果. 最后用跟踪结果对 Kalman 滤波器进行更新.

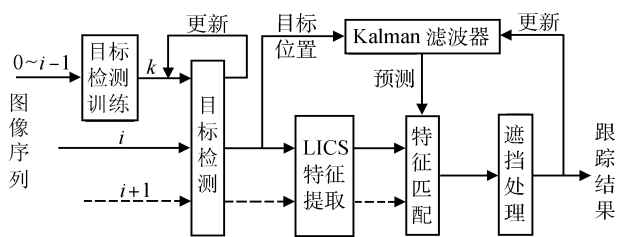


图 2 跟踪算法整体流程
Fig. 2 The whole process of object tracking

4 实验结果及分析

4.1 目标检测实验

目标检测的测试视频采用文献 [19] 中的三个户外动态视频, 视频的分辨率都是 320×240 . 图 3 给出了每个视频的名称以及所列图像在视频中的帧数和视频总帧数 (当前帧/总帧数). 其中, Fountains 中有四个喷泉, 右上角有晃动的树叶; Waving leaves 中有大面积的树叶晃动; Ocean waves 中有不断拍打海岸的海浪. 图 3 列出了不同方法的目标检测结果, 都采用相同的后处理, 去除了小于 5 个像素点的连通区域. 第一行为原始图像; 第二行为本文算法的检测结果, 像素块大小为 8×8 , 三个序列均进行 100 帧的训练; 第三行为文献 [19] 方法的检测结果; 第四行为文献 [1] 方法的检测结果, 阈值参数 k 固定取 8.

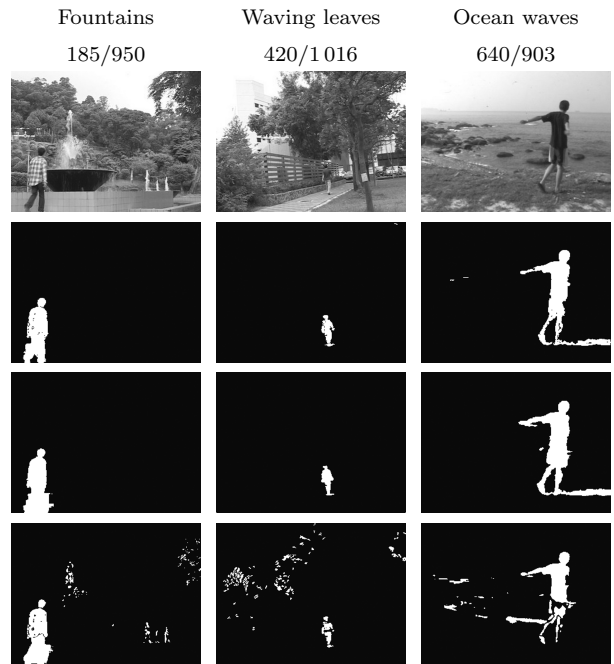


图 3 目标检测视频及检测结果
Fig. 3 The testing videos and the object detection results

为了评价比较本文的目标检测算法, 采用文献 [21] 提出的评价指标: A 为检测出的目标区域, B 为真实的目标区域, 则二者的相似度定义为 $S(A, B) = \frac{A \cap B}{A \cup B}$. $S(A, B)$ 的值在 0 和 1 间, 当 A 和 B 完全相同时取 1, 反之, A 和 B 差别越大, 相似度越接近 0. 目标的真实区域采用人工方法手动获得, 由于视频序列较长, 为减小工作量, 采用与文献 [21] 类似的简化方法, 从每个视频序列中抽取 40 帧来进行评价, 这 40 帧图像平均分布在整个视频序列中. 表 1 显示了三种方法的评价结果.

表 2 显示了三种方法对三个视频进行目标检测时的处理速度 (fps), 其中包含了后处理. 本文方法和文献 [1] 方法的实验环境为 Celeron 2.66 GHz CPU、512MB 内存, 在 Matlab 7.1 下完成, 文献 [19] 方法的实验环境为 3.4 GHz CPU、768 MB 内存.

表 1 目标检测算法评价结果

Table 1 Evaluation results of object detection algorithms

序列名称	本文方法	文献 [19] 方法	文献 [1] 方法
Fountains	0.7224	0.6422	0.5021
Waving leaves	0.7849	0.6084	0.3277
Ocean waves	0.6507	0.4840	0.5753
平均结果	0.7127	0.5782	0.4684

表 2 目标检测算法处理速度 (fps)

Table 2 Processing speeds of object detection algorithms (fps)

序列名称	本文方法	文献 [19] 方法	文献 [1] 方法
Fountains	27.73	22.03	45.24
Waving leaves	27.62	21.09	42.84
Ocean waves	27.32	21.59	46.35
平均结果	27.56	21.57	44.81

4.2 目标跟踪实验

对于目标跟踪而言, 目标检测结果的好坏对跟踪结果有很大的影响, 第 4.1 节的实验证明本文提出的目标检测算法有着出色的性能, 因此, 在跟踪实验里主要验证本文提出的 LICS 特征用于目标跟踪的有效性, 由于 LICS 特征并不适合用于人体的跟踪, 所以采用了另外的视频进行以下三个实验. 计算 LICS 特征时目标都被划分为 8 个子区域, w 取值都为 6.

实验 1. 在目标光照条件发生变化时进行跟踪. 采用图 4 所示视频, 分辨率 320×240 , 长度 220 帧, 共有 3 辆汽车先后出现. 目标从阳光充足的区域进入到建筑物阴影区, 目标的整体亮度发生了较大的变化, 同时目标的尺度也发生了变化, 本文算法均能实现准确的跟踪.



图 4 目标亮度发生变化时的跟踪效果

Fig. 4 Tracking objects with illumination changing

为了测试 LICS 特征对于环境光照的鲁棒性, 分别截取了 3 个目标在图 4 所示的 3 个不同位置时的图像, 并计算目标特征的相似度. 设第 i 个目标在 3 个位置时的特征分别为 A_i 、 B_i 、 C_i ,

使用 Bhattacharyya 距离计算同一目标在任意两个不同位置特征 $A_i B_i$ 、 $A_i C_i$ 、 $B_i C_i$ 间的相似度: $Sim = \frac{1}{N} \sum_{n=1}^N \frac{\min(L_n, K_n)}{\max(L_n, K_n)}$, 其中 L_n 、 K_n 分别为要比较的两个 N 维向量 L 、 K 的第 n 个元素, L 、 K 越接近则计算得到的相似度越接近 1. 计算结果见表 3, 其中的 LICS' 与 LICS 的区别是前者不包含亮度对数的计算. 可看出, 本文提出的 LICS 特征更加适合目标整体特征的计算, 而且对于环境光照的变化有更好的鲁棒性.

实验 2. 多目标的跟踪. 采用 PETS'2001^[11] 的测试视频, 截取的视频长度为 350 帧, 分辨率 320×240 , 共有两车两人共四个目标出现. 检测结果见图 5. 图中汽车都在前进和倒车间发生转换, 其运动速度、运动方向以及姿态都发生了较大变化, 同时有新的运动目标出现, 本文算法能实现每个目标的准确跟踪. 对视频中的人采用如下的区域划分方式来统计 LICS 特征: 将人体从上到下划分为 4 个部分, 最上部分的纵向长度占总体长度的 $1/5$, 剩余 3 个部分在纵向上平均划分, 这样可得到一个 8 维的 LICS 特征向量. 此种方法只适合直立状态人的跟踪, 而且事先还要进行一个人车的分类操作, 可通过计算目标包围盒的高宽比和目标的占空比来进行区分.



图 5 多目标的跟踪效果

Fig. 5 Multi-object tracking results

实验 3. 目标发生遮挡时的跟踪. 采用图 6 所示视频, 分辨率 320×240 , 长度 150 帧, 共有四辆车出现. 其中两辆汽车相向而行, 在中途发生完全的遮挡后又分开. 为了减小计算量, 首先利用 Kalman 滤波预测目标在下一帧是否会发生遮挡, 然后根据 LICS 特征匹配情况确认是否真的发生遮挡, 只有在发生遮挡时才进行目标的分块匹配操作. 在本实验中, 将目标平均分为四个子块.



图 6 遮挡情况下的跟踪效果

Fig. 6 Tracking objects with occlusion

对跟踪结果的评价采取文献[22]提出的基于帧的评价方法, 它评价的是目标检测与跟踪的综合结

表 3 光照变化时的目标 LICS 特征相似度
Table 3 Similarities of the same object with different illumination

特征	A ₁ B ₁	A ₁ C ₁	B ₁ C ₁	A ₂ B ₂	A ₂ C ₂	B ₂ C ₂	A ₃ B ₃	A ₃ C ₃	B ₃ C ₃	平均
CCH	0.8140	0.5916	0.6969	0.0853	0.4904	0.2154	0.6506	0.3567	0.2882	0.4655
LICS'	0.7975	0.7101	0.7185	0.7720	0.7763	0.7396	0.8876	0.7996	0.7712	0.7747
LICS	0.9165	0.8218	0.8681	0.8055	0.7907	0.8598	0.8716	0.8424	0.9004	0.8530

表 4 目标跟踪的评价结果
Table 4 Evaluation results of the object tracking algorithm

评价指标	计算公式	实验 1	实验 2	实验 3	平均	备注
TRDR	TP/TG	0.8868	0.9129	0.9667	0.9221	越接近 1 越好
FAR	FP/(TP+FP)	0.0070	0.0222	0	0.0097	越接近 0 越好
Specificity	TN/(FP+TN)	0.9836	0.9541	1	0.9792	越接近 1 越好
Accuracy	(TP+TN)/TF	0.9136	0.9257	0.9800	0.9398	越接近 1 越好
PP	TP/(TP+FP)	0.9930	0.9778	1	0.9903	越接近 1 越好
NP	TN/(FN+TN)	0.7692	0.8320	0.9524	0.8512	越接近 1 越好
FNR	FN/(FN+TP)	0.1132	0.0871	0.0333	0.0779	越接近 0 越好
FPR	FP/(FP+TN)	0.0164	0.0459	0	0.0208	越接近 0 越好

果,也是目前大多数跟踪评价算法采用的方式. 计算方法如下: 首先统计整个视频序列的 TN (True negative)、TP (True positive)、FN (False negative) 和 FP (False positive) 帧数, 分别定义为: TN, 场景中和跟踪结果中均没有运动目标; TP, 跟踪结果中至少有一个目标与真实目标相一致; FN, 场景中有运动目标, 跟踪结果中没有或者跟踪结果都与真实目标不一致; FP, 跟踪结果中有运动目标, 而实际上没有或者真实目标与跟踪结果都不一致. 其中, 一致性指一个目标的质心位于另一个目标的包围盒之内. TG 为存在真实目标的视频帧数, TF 为视频总帧数, 根据以上参数, 可以计算以下几项指标: TRDR (Tracker detection rate)、FAR (False alarm rate)、Specificity、Accuracy、PP (Positive prediction)、NP (Negative prediction)、FNR (False negative rate)、FPR (False positive rate). 表 4 列出了每项指标的计算公式及评价结果. 表 5 列出了本文算法对本节 3 个视频进行目标检测 + 跟踪的处理速度, 实验环境与第 4.1 节相同.

表 5 目标检测 + 跟踪的处理速度
Table 5 Processing speeds of object detection and tracking

实验视频	实验 1	实验 2	实验 3	平均
速度 (fps)	25.58	16.44	15.12	19.05

4.3 目标跟踪算法复杂度分析

与其他的跟踪算法相比, 如粒子滤波器 (Particle filter) 算法, 是基于序贯 Monte Carlo 滤波方法

中最具代表性的方法之一, 它的收敛性在很大程度上依赖于粒子数的多少, 计算复杂度大, 而本文算法计算简单, 能准确进行目标跟踪; 与均值漂移 (Mean shift, MS) 算法相比, 本文算法利用 Kalman 滤波预测结果有效限制了跟踪区域, 跟踪效果相对良好, 而且计算也更加简单. 文献 [23] 对经典的 MS 算法进行了复杂度分析, 在此我们忽略 MS 算法的多尺度搜索过程, 则其复杂度为:

$$C_m = N_r(c_h + N_h c_w) + N_b c_s \quad (6)$$

其中, N_r 为平均迭代次数, N_h 为目标的像素数量, N_b 为直方图数目, c_h 为加权直方图的计算代价, c_w 是迭代过程中为每一像素付出的计算代价, c_s 为计算相似性代价.

本文跟踪算法的复杂度可表示为:

$$C_e = (N_t + 1)(c_f + N_h c_l) + N_t N_l c_s + c_k \quad (7)$$

其中, N_t 为 Kalman 滤波器预测结果 w 邻域内出现的目标个数, c_l 为单个像素 LICS 特征值的计算代价, c_f 为 LICS 特征向量的计算代价, N_l 为 LICS 特征向量的维数, c_k 为 Kalman 预测的计算代价, 这里假定目标所占的平均像素数目为 N_h , 采用与式 (6) 相同的相似度计算方法, 计算相似性的代价也为 c_s .

对式 (6) 和 (7) 进行比较, 在式 (6) 中, N_r 的值根据文献 [24] 的实验结果约为 4, 式 (7) 中 N_t 的值对于普通视频通常为 1~3, 这里我们取最大值 3, 使

$N_r = N_t + 1$. 则将式 (6) 和 (7) 相减得到式 (8).

$$C_m - C_e = N_r((c_h - c_f) + N_h(c_w - c_l)) + (N_b - N_t N_l)c_s - c_k \quad (8)$$

在文献 [23] 中, $N_b = 2^{12}$, 假设采用 32 维的 LICS 特征向量, 则 $N_l = 32$, $N_t N_l = 96$, 有 $N_b \gg N_t N_l$. c_w 的计算至少包含 3 次乘法、1 次除法和加减法各 1 次, c_l 的计算为 1 次对数和 1 次减法, 可假定 $c_w \approx c_l$. c_h 的计算量与搜索的像素个数成正比, 假设搜索像素数目为 N_s , 有 $N_s > N_h$, 设每个像素权重的计算量为 c_p , 则 $c_h = N_s c_p$, c_p 至少包含 1 次加法、2 次乘法、1 次除法和 1 次开方. c_f 的计算包含 N_h 次加法和 N_l 次除法, 通常 $N_h \gg N_l$, 因此有 $c_h \gg c_f$. Kalman 预测的计算量很小, 相对于整个算法可以忽略不计, $c_k \approx 0$.

通过上述分析, 可发现本文算法的计算量远小于 MS 算法. 一些改进的 MS 算法, 如文献 [23], 其计算量稍大于经典 MS 算法. 文献 [24–25] 将 Kalman 滤波与 MS 算法相结合, 可减小 MS 算法的迭代次数, 根据文献 [24] 的实验结果, N_r 值由原来的 4 降为 3.5, 其计算量仍然远大于本文算法.

5 结论

本文提出的运动目标检测与跟踪方法在检测与跟踪精度、处理速度等方面都有较好的结果. 算法特点有: 1) 目标检测运用熵能自适应方法判定局部阈值, 对不同的视频均可实现全自动检测, 局部自适应阈值可有效抑制噪声的影响, 精确地提取目标; 2) 采用广义高斯分布描述阈值样本, 可更加准确地描述样本分布, 通过计算样本落于阈值内的概率来得到合适的分割阈值, 最大限度地抑制噪声; 3) 提出 LICS 特征用于目标的跟踪, 并结合 Kalman 滤波器的预测结果缩小匹配范围, 减少计算量. 可较好地克服目标的光照、姿态变化对跟踪效果的影响, 增强跟踪的鲁棒性.

该方法也存在一些不足: 一是当目标出现在背景中的混乱区域时目标检测的效果不够理想, 易产生缺失; 二是对非刚体目标的跟踪, 还需进一步研究 LICS 特征的应用.

References

- 1 Franck L, Marc L, Francis F. On the use of entropy power for threshold selection. *Signal Processing*, 2004, **84**(10): 1789–1804
- 2 Zuo J Y, Zhao C H, Pan Q, Lian W. A novel binary image filtering algorithm based on information entropy. In: Proceedings of the 6th World Congress on Intelligent Control and Automation. Dalian, China: IEEE, 2006. 10375–10379

- 3 Shi Hua, Li Cui-Hua, Wei Feng-Mei, Wang Hua-Wei. Moving object tracking based on location and confidence of pixels. *Journal of Computer Research and Development*, 2005, **42**(10): 1726–1732
(施华, 李翠华, 韦凤梅, 王华伟. 基于像素可信度和空间位置的运动目标跟踪. *计算机研究与发展*, 2005, **42**(10): 1726–1732)
- 4 Hou Zhi-Qiang, Han Chong-Zhao. A survey of visual tracking. *Acta Automatica Sinica*, 2006, **32**(4): 603–617
(侯志强, 韩崇昭. 视觉跟踪技术综述. *自动化学报*, 2006, **32**(4): 603–617)
- 5 Moeslund T B, Granum E. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 2001, **81**(3): 231–268
- 6 Wang Liang, Hu Wei-Ming, Tan Tie-Niu. A survey of visual analysis of human motion. *Chinese Journal of Computers*, 2002, **25**(3): 225–237
(王亮, 胡卫明, 谭铁牛. 人运动的视觉分析综述. *计算机学报*, 2002, **25**(3): 225–237)
- 7 Kumar P, Ranganath S, Sengupta K, Huang W M. Cooperative multitarget tracking with efficient split and merge handling. *IEEE Transactions on Circuits and Systems for Video Technology*, 2006, **16**(12): 1477–1490
- 8 Han M, Xu W, Tao H, Gong Y H. Multi-object trajectory tracking. *Machine Vision and Applications*, 2007, **18**(3): 221–232
- 9 Lin S P, Chen Y H, Wu B F. A real-time multiple-vehicle detection and tracking system with prior occlusion detection and resolution, and prior queue detection and resolution. In: Proceedings of the 18th International Conference on Pattern Recognition. Washington D. C., USA: IEEE, 2006. 828–831
- 10 Zhou Q M, Aggarwal J K. Object tracking in an outdoor environment using fusion of features and cameras. *Image and Vision Computing*, 2006, **24**(11): 1244–1255
- 11 Xu L Q, Landabaso J L, Lei B. Segmentation and tracking of multiple moving objects for intelligent video analysis. *BT Technology Journal*, 2004, **22**(3): 140–150
- 12 Kaneko T, Hori O. Feature selection for reliable tracking using template matching. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, USA: IEEE, 2003. 796–802
- 13 Chang Fa-Liang, Ma Li, Qiao Yi-Zheng. Target tracking algorithm under occlusion based on feature correlation matching. *Journal of Image and Graphics*, 2006, **11**(6): 877–882
(常发亮, 马丽, 乔谊正. 遮挡情况下基于特征相关匹配的目标跟踪算法. *中国图象图形学报*, 2006, **11**(6): 877–882)
- 14 Zheng Shi-You, Fei Shu-Min, Yuan Xiao-Hui, Long Fei. Occlusion detection and tracking in image sequence based on the affine invariant. *Signal Processing*, 2006, **22**(3): 434–439
(郑世友, 费树岷, 袁晓辉, 龙飞. 基于仿射不变量的运动目标遮挡检测和跟踪. *信号处理*, 2006, **22**(3): 434–439)
- 15 Cheng E D, Piccardi M. Track matching by major color histograms matching and post-matching integration. In: Proceedings of International Conference on Image Analysis and Processing. Berlin, German: Springer, 2005. 1148–1157

- 16 Huang C R, Chen C S, Chung P C. Contrast context histogram — a discriminating local descriptor for image matching. *Pattern Recognition*, 2008, **41**(10): 3071–3077
- 17 Cucchiara R, Grana C, Piccardi M, Prati A. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(10): 1337–1342
- 18 Wang Tai-Yue, Li Zhi-Ming. A fast parameter estimation of generalized Gaussian distribution. *Chinese Journal of Engineering Geophysics*, 2006, **3**(3): 172–176
(汪太月, 李志明. 一种广义高斯分布的参数快速估计法. *工程地球物理学报*, 2006, **3**(3): 172–176)
- 19 Chen Y T, Chen C S, Huang C R, Hung Y P. Efficient hierarchical method for background subtraction. *Pattern Recognition*, 2007, **40**(10): 2706–2715
- 20 Wu Q Z, Jeng B S. Background subtraction based on logarithmic intensities. *Pattern Recognition Letters*, 2002, **23**(13): 1529–1536
- 21 Li L Y, Huang W M, Gu I Y H, Tian Q. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 2004, **13**(11): 1459–1472
- 22 Bashir F, Porikli F. Performance evaluation of object detection and tracking systems. In: *Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*. New York, USA: IEEE, 2006. 1–8
- 23 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)
- 24 Zhu Sheng-Li, Zhu Shan-An, Li Xu-Chao. Algorithm for tracking of fast motion objects with mean shift. *Opto-Electronic Engineering*, 2006, **33**(5): 66–70
(朱胜利, 朱善安, 李旭超. 快速运动目标的 Mean Shift 跟踪算法. *光电工程*, 2006, **33**(5): 66–70)
- 25 Chang Fa-Liang, Liu Xue, Wang Hua-Jie. Target tracking algorithm based on mean shift and Kalman filter. *Computer Engineering and Applications*, 2007, **43**(12): 50–52
(常发亮, 刘雪, 王华杰. 基于均值漂移与卡尔曼滤波的目标跟踪算法. *计算机工程与应用*, 2007, **43**(12): 50–52)



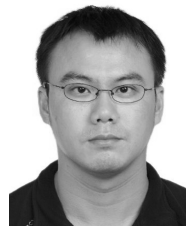
刘少华 国防科学技术大学信息系统与管理学院博士研究生. 主要研究方向为视频运动目标检测与跟踪. 本文通信作者. E-mail: lshh37@yahoo.com.cn
(LIU Shao-Hua Ph.D. candidate at the College of Information System and Management, National University of Defense Technology. His research interest covers video moving object detection and tracking. Corresponding author of this paper.)



张茂军 国防科学技术大学信息系统与管理学院教授. 主要研究方向为虚拟现实与系统仿真技术.
E-mail: maojun@iscas.com.cn
(ZHANG Mao-Jun Professor at the College of Information System and Management, National University of Defense Technology. His research interest covers virtual reality and system simulation.)



熊志辉 国防科学技术大学信息系统与管理学院讲师. 主要研究方向为图像处理与嵌入式系统.
E-mail: xzhnudt@vip.sina.com
(XIONG Zhi-Hui Lecturer at the College of Information System and Management, National University of Defense Technology. His research interest covers image process and embedded system.)



陈旺 国防科技大学信息系统与管理学院博士研究生. 主要研究方向为图像处理与立体视觉.
E-mail: chenwang@nudt.edu.cn
(CHEN Wang Ph.D. candidate at the College of Information System and Management, National University of Defense Technology. His research interest covers image process and stereo vision.)