

基于超像素时空特征的视频显著性检测方法

李艳蕊, 徐熙平*

长春理工大学光电工程学院, 吉林 长春 130022

摘要 提出一种基于超像素时空特征的视频显著性检测方法。所提方法可对图像进行超像素分割, 提取颜色梯度和运动梯度特征, 以构建超像素级时空梯度图。用平均加权测地距离来衡量时空梯度图上每一个超像素相对于其邻域的时空显著程度, 形成时空显著图。根据时间域上目标运动的连续性, 并借助熵的概念来表征运动模式的一致程度, 构建运动一致性图。融合时空显著图和运动一致性图, 通过自适应阈值处理定位运动目标。实验从可视化分析和定量评估两个方面将所提方法与其他算法进行对比, 结果表明所提方法具有较强的抗环境干扰能力, 适用于背景纹理复杂或环境随机变化的视频中运动目标的检测, 其检测准确率高达 92%。

关键词 图像处理; 显著性检测; 超像素分割; 测地距离; 运动熵

中图分类号 TP391

文献标识码 A

doi: 10.3788/AOS201939.0110001

Video Saliency Detection Method Based on Spatiotemporal Features of Superpixels

Li Yandi, Xu Xiping*

College of Photoelectrical Engineering, Changchun University of Science and Technology,
Changchun, Jilin 130022, China

Abstract A novel video saliency detection method is proposed based on the spatiotemporal features of superpixels, which is used to the superpixel segmentation of images and extract the features of color gradient and motion gradient for the construction of a spatial-temporal gradient map of superpixels. The average weighted geodesic distance is used to measure the spatiotemporal saliency degree of each superpixel relative to its neighbor on the spatiotemporal gradient map, and thus the spatiotemporal saliency map is formed. In order to obtain the motion coherency map, the motion entropy in the multiple continuous frames is computed to represent the motion coherence of motion object over time. The fusion of spatiotemporal saliency maps and motion coherency maps is applied to locate in the salient motion using adaptive segmentation. In addition, the performance of the proposed method is compared with those of the other algorithms in experiments from two perspectives of visual analysis and qualitative evaluation. The results show that the proposed method is robust and suitable for the detection of moving targets in videos within complex background texture and changeable environment. Moreover, the detection precision is up to 92%.

Key words image processing; saliency detection; superpixel segmentation; geodesic distance; motion entropy

OCIS codes 100.4145; 100.4199; 100.3008

1 引 言

近年来视频显著性检测在多媒体领域引起了广泛关注, 特别是在智能监控、视频检索和目标跟踪与识别等方面^[1-2]。计算机视觉技术越来越多地被用于人类视觉建模, 通过在一定程度上削弱动态环境因素的干扰, 可高效地确定视频中的感兴趣对象, 有利于智能化分析和理解视频内容。运动显著性检测

(运动目标检测)作为视频分析和理解的基础环节, 其检测结果将会对后续特征提取及识别工作产生重大影响。因此, 运动显著性检测算法设计成为视频理解中的关键技术和研究重点。

目前关于显著性检测的研究总体上可以分为三类: 空域显著性分析、时域显著性分析和时空域显著性分析。基于空域特征的目标检测方法的分割精度高, 但该方法对图像的分割过细, 易造成信息冗余;

收稿日期: 2018-07-10; 修回日期: 2018-07-22; 录用日期: 2018-08-13

基金项目: 吉林省科技发展计划 (20160520018JH)

* E-mail: xyp@cust.edu.cn

基于时域特征(运动特征)的目标检测算法的分割速度快,但精度低,受环境因素的干扰严重。因此,兼顾视频序列中目标的运动模式与空间分布的时空分割技术已经成为现阶段主流的运动目标检测方法。基于空域特征的显著性模型在图像显著性检测领域已经取得了较好的成果,因此很多研究将其作为视频显著性的一个先验模型,在此基础上融入运动特征,得到了理想的检测效果。1998年Itti等^[3]提出了利用多尺度和中心周围原则结合颜色、强度和方向特征来计算图像显著性的方法。随后,Zhang等^[4]将傅里叶变换与Itti模型相结合,用于检测视频显著性区域。Lin等^[5]将模糊神经网络与Itti模型相结合,训练一种自学习式ICVA(Intention-Oriented Computational Visual Attention)模型,成功地将其应用于道路视频内容检测领域。而Itti^[6]也在自己最初的图像显著性模型的基础上引入运动特征,通过计算所有像素的颜色、亮度、方向和运动特征的先验概率密度分布和后验概率密度分布的KL(Kullback-Leibler)散度获得视频显著图,并利用贝叶斯原理实现了视频异常行为的检测。Guo等^[7]将谱残差方法应用到视频显著性检测,首先提取颜色、运动和亮度特征,针对每个特征通道计算四元傅里叶变换的相位谱(PQFT),融合各通道结果,最终得到了显著性结果;该方法的处理速度快,但是只能检测到物体边缘,对大尺寸目标的适应能力较差,受背景运动的干扰也比较明显。Li等^[8]在空间域提取颜色特征,并根据颜色直方图计算稀有度、紧密度和中心区域对比度,将三者作为空间显著性量度,提取运动特征强度和对比度作为时间显著性量度,最终通过支持向量机(SVM)将空间显著性量度和时间显著性量度进行融合,得到显著性目标。Liu等^[9]提出了一种基于超像素的视频显著性模型,首先在每一个超像素内提取运动直方图和颜色直方图作为局部特征,级联得到全局特征,用于描述每一帧的图像内容,然后用超像素级的运动特征来预测和调整时间显著性,利用空间特征的全局对比度和稀疏性来计算空间显著性,最后通过自适应权重将二者融合得到最终的显著性结果。Wang等^[10]结合帧内边界和帧间运动两方面信息,提出运动边界特征,并得到光流梯度映射比运动显著性测度更加稳健可信的结论。Singh等^[11]首先将像素聚类成块或时间超像素,提取如颜色、运动和边界等多元化特征,从不同侧面生成显著性模型,并通过线性SVM对这些显著性图进行加权融合。类似地,Rahtu等^[12]利用

条件随机场(CRF)将运动、光照和颜色等显著性信息进行融合。Fang等^[13]将统计学中的不确定度作为时空信息的显著性量度,分别计算基于运动特征的时间显著性和基于颜色、纹理等特征的空间显著性,然后将熵信息作为权值来融合两部分信息,并取得了较好的结果。以上视频显著性检测方法大多融合了运动特征和单帧图像显著性特征,但是针对背景纹理复杂和多种环境因素干扰的情况,这些方法依然存在稳健性较差的问题。

本文沿用时空特征融合思想,提出了一种基于超像素时空特征的显著运动检测方法。与基于像素的显著性检测方法相比,该方法具有检测速度快、目标完整度较高等优势。该方法首先提取图像的静态边界特征和光流运动特征(方向和幅值),结合运动连续性生成基于超像素的时空梯度图;然后利用超像素之间的平均加权测地距离来计算每一个超像素的时空梯度显著性,进而得到时空显著图;通过在时间域上分析运动目标的运动模式,以超像素为单位计算光流方向及幅值的熵值,得到运动一致性图;最后将时空显著图与运动一致性图进行融合,得到最终的视频显著性结果。

2 视频显著图生成

基于超像素时空特征的显著运动检测方法首先利用线性迭代聚类(SLIC)方法^[14]生成超像素图像,在保留目标完整性的基础上降低图像特征维度。视频显著图主要通过时空显著性和运动一致性两方面进行特征计算得到,具体的生成过程如图1所示。

2.1 时空梯度图生成

考虑到边缘是描述图像空域分布最基本且有效的特征,同时在时间邻域内光流变化较大的像素会携带更多的运动信息。因此在超像素分割的基础上融合单帧图像的颜色梯度特征和运动光流特征(幅值及方向)来表征运动目标的时空梯度。具体过程如图2所示。

设输入视频序列 $I = \{I^1, I^2, \dots\}$, 在Lab色彩空间,第 I^k 帧内像素 x_i^k 处的颜色梯度为

$$G_c^k(x_i^k) = \| (L_i, a_i, b_i), (L_j, a_j, b_j) \|_2, \quad (1)$$

式中: $\| \cdot \|_2$ 表示L2范数。利用(2)式计算得到第 I^k 帧内像素 x_i^k 处的光流幅值为

$$G_{om}^k(x_i^k) = \frac{\sqrt{M_x^2 + M_y^2}}{\max(M)}, \quad (2)$$

式中: M_x^2 和 M_y^2 分别表示 x 、 y 方向上的光流梯度; $M = M_x^2 + M_y^2$; $\max(M)$ 为最大幅值,用于归一化处理。

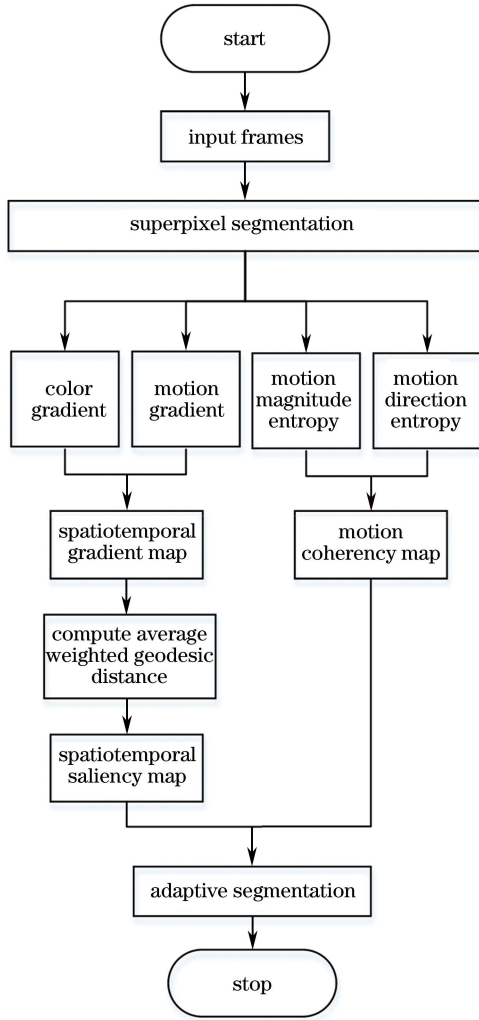


图 1 算法流程图

Fig. 1 Flow chart of proposed method

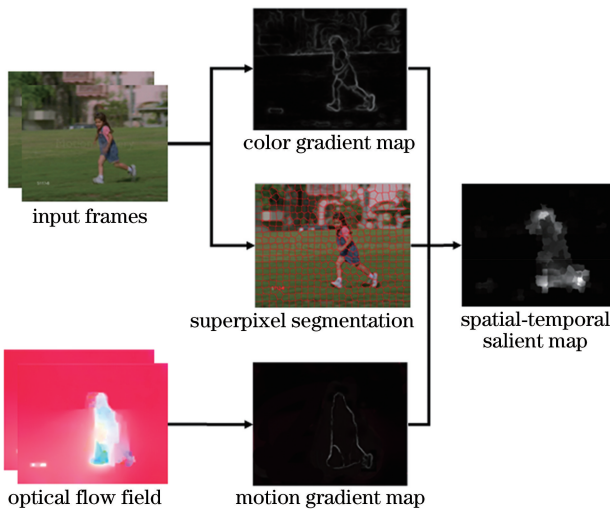


图 2 时空梯度图生成过程

Fig. 2 Illustration of spatiotemporal gradient map generation

将光流方向梯度 $G_{od}^k(x_i^k)$ 定义为某一像素与其邻域 N_s (取值为 8) 内所有像素值之间的差值平均

值,计算公式为

$$G_{od}^k(x_i^k) = \frac{1}{N} \sum_{j=1}^{|N_j|} \|\theta_i^k - \theta_j^k\|, \theta = \arctan(M_y/M_x). \quad (3)$$

设图像 I^k 经过分割处理得到的超像素图像为: $Y^k = \{Y_1^k, Y_2^k, \dots\}$, 对于颜色梯度图 $G_c^k(x_i^k)$, 每一个超像素 Y_n^k 的边界值用该超像素内前 10 个最大像素值的平均值表示, 并生成超像素边界图 $\hat{G}_c^k(Y_i^k)$; 利用文献[15]中的光流计算方法得到图像 F^k 的光流场 V^k , V^k 中每一个像素的光流幅值和光流方向梯度分别为 $G_{om}^k(x_i^k)$ 和 $G_{od}^k(x_i^k)$, 同理可以得到超像素光流幅值图 $\hat{G}_{om}^k(Y_i^k)$ 和光流方向图 $\hat{G}_{od}^k(Y_i^k)$. 由于光流特征对环境噪声和摄像机不规则运动的干扰比较敏感, 在背景纹理复杂和前景、背景颜色差异较小的情况下, 颜色梯度的表征能力下降. 针对这种情况, 将超像素边界图 $\hat{G}_c^k(Y_i^k)$ 、光流幅值图 $\hat{G}_{om}^k(Y_i^k)$ 和光流方向图 $\hat{G}_{od}^k(Y_i^k)$ 以(4)式的形式进行融合, 得到一幅基于超像素且具有时空信息的边界图 G_{ST}^k 为

$$G_{ST}^k = \hat{G}_c^k(Y_i^k) \odot \hat{G}_{om}^k(Y_i^k) \odot \hat{G}_{od}^k(Y_i^k), \quad (4)$$

式中:“ \odot ”代表两个向量(矩阵)中对应元素相乘. 通过计算 G_{ST}^k 可以将具有较大光流变化和颜色梯度较大的区域从背景中检测出来. 图 2 表明, 采用融合单帧图像的颜色梯度特征和运动光流特征来表征运动目标的时空梯度的方式仅得到一个关于前景的粗略定位, 要获得精确定位需利用时空显著关系对该区域进行进一步分割.

2.2 时空显著性计算

要进行时空显著性计算, 首先要建立超像素图像 Y^k 对应的时空显著图 $G^k = \{V^k, E^k\}$, 其拓扑结构为无向带权图, 其中 V^k 代表图中所有节点的集合, E^k 代表两个节点之间的连接线; Y_m^k 和 Y_n^k 之间的连接线 E^k 的梯度权值用 W_{mn}^k 表示, 其值取决于两个相邻超像素 Y_m^k 和 Y_n^k 对应的时空梯度值 $G_{ST}^k(Y_m^k)$ 与 $G_{ST}^k(Y_n^k)$ 之间的差值, 可表示为

$$W_{mn}^k = \|G_{ST}^k(Y_m^k) - G_{ST}^k(Y_n^k)\|. \quad (5)$$

文献[16]表明, 沿着图像边缘的区域很大概率属于背景. 因此通过计算每一个超像素与位于图像边缘的超像素之间的最短加权测地距离来度量该超像素的显著性 $S_{intra,n}^k$, 即属于前景的概率. $S_{intra,n}^k$ 的计算公式为

$$S_{intra,n}^k = \min_{v_m \in Y^k, v_n \in S_b^k} d_{geo}(v_m, v_n, G^k), \quad (6)$$

式中： v_m 和 v_n 为时空显著图 G^k 内两个代表超像素的节点； S_p^k 表示视频帧 F_n^k 中沿 4 个边缘的所有超像素集合； $d_{\text{geo}}(\cdot)$ 表示时空显著图 G^k 中节点 v_m 与 v_n 之间的测地距离，代表沿着超像素边界最短加权路径的累积值^[17]。该加权路径的权值由两部分构成，分别为超像素之间的梯度权值 W_{mn}^k 和超像素之间的距离权值 w_{mn} 。 w_{mn} 可通过计算两个超像素之间的欧氏距离得到：

$$w_{mn} = \| p_m, p_n \|_2, \quad (7)$$

式中： p_m, p_n 分别代表第 m, n 个超像素的中心。对 w_{mn} 进行归一化处理可得：

$$w'_{mn} = 1 - \exp(-w_{mn}), \quad (8)$$

则平均加权测地距离 $d_{\text{geo}}(v_m, v_n, G^k)$ 可表示为

$$d_{\text{geo}}(v_m, v_n, G^k) = \min_{L_{v_m, v_n}} \frac{1}{z} \sum [\omega'_{mn} \cdot W_{mn}^k], \quad (9)$$

式中： z 表示 G^k 内从节点 v_m 到 v_n 的最短路径所经过的所有超像素的总和。

综上，一帧图像 I^k 的时空显著性 $S_{\text{intra},n}^k$ 可以通过(6)~(9)式计算得到，对其结果进行归一化可得到时空显著图 P^k 。但是，仅通过时空显著性获取运动目标并不准确，特别是对目标周围的区域进行过分割，会使目标更容易被误认为是前景，且纹理近似和光流估计产生的误差也会对目标定位造成干扰。考虑到目标在相邻帧之间的运动是连续的，故利用帧间运动一致性完善时空显著图的信息。

2.3 运动一致性计算

视频中具有显著性的运动模式相对于背景中随机发生的小幅度运动(如摇晃的树枝、波动的湖水等)存在明显的差异。由于在时间轴上目标的运动过程具有一定的连续性，借助时间一致性约束可将虚假前景超像素剔除。引入信息熵的概念来评估运动目标所携带光流信息(幅值和方向)的一致程度。计算过程中，将每个超像素中心作为检测对象来取代对单个像素的分析，可有效降低运算量。

运动一致性计算首先是利用当前帧的前 T 帧图像中对应位置的光流幅值来计算超像素 Y_i^k 的光流幅值熵 $E_{\text{mag}}(M')$ ， $E_{\text{mag}}(M')$ 可表示为

$$E_{\text{mag}}(M') = - \sum_{M'} p_{M'} \lg p_{M'}, \quad (10)$$

式中： $p_{M'}$ 表示光流幅值 M' 出现的概率。计算过程中，若 T 值设置较大，可以滤除由较少帧数得到的不可靠光流信息，但同时也容易造成信息冗余，使

$E_{\text{mag}}(M')$ 对目标的识别能力严重下降；若 T 值设置非常小，即只考虑当前帧的运动信息，则可能会因异常光流的干扰而导致检测失效。综合考虑，在本实验中将 T 值设为 15。

根据 2.1 节得到 $\hat{G}_{\text{od}}^k(Y_i^k)$ ，同理计算得到光流方向熵为

$$E_{\text{dir}}(\theta) = - \sum_{\theta} p(\theta) \lg p(\theta), \quad (11)$$

式中： $p(\theta)$ 为光流方向 θ 出现的概率。由此可见，根据光流幅值一致性 $E_{\text{mag}}(m)$ 能够有效检测到视频帧中运动幅值较大的区域，同时方向一致性 $E_{\text{dir}}(\theta)$ 能够检测运动方向变化较随机的区域。因此，通过一定的融合策略来评估运动物体的时间一致性 $S_{\text{inter},n}^k$ ，能够有效剔除仅根据时空显著性得到的虚假前景，从而定位到精确的运动显著性区域。 $S_{\text{inter},n}^k$ 的计算公式为

$$S_{\text{inter},n}^k = E_{\text{mag}}(M)[1 - E_{\text{dir}}(\theta)]. \quad (12)$$

将(6)式得到的时空显著性 $S_{\text{intra},n}^k$ 与(12)式得到的运动一致性 $S_{\text{inter},n}^k$ 进行融合，可得

$$S_{\text{final},n}^k = S_{\text{inter},n}^k \odot S_{\text{intra},n}^k. \quad (13)$$

利用最大类间差方法(大津法)对图像 I^k 进行自适应分割，最终得到背景区域 B^k 和前景区域 U^k 为

$$B^k = \{Y_n^k \mid S_{\text{final},n}^k \leq \sigma^k\}, \quad \sigma^k = \mu(S_{\text{final},n}^k), \quad (14)$$

$$U^k = Y^k - B^k. \quad (15)$$

3 实验结果及分析

3.1 实验设计

为了评价所提算法的综合性能，可进行可视化分析和定量评估两部分实验，并在 SegTrackV2 视频数据集^[18]上进行视频显著性测试。SegTrackV2 数据集提供了不同场景下的多种类型的运动视频，挑选其中的女孩、降落伞、飞鸟和非洲猎豹/鹿 4 种情景，分别代表运动过程中外形变化较大、光照干扰严重、背景复杂和运动速度较快 4 种情况，这样能够更全面地验证所提的运动目标检测算法的稳健性和普遍性。实验中选取连续 15 帧图像用于计算显著目标的运动一致性。软件环境为 Matlab2016a，硬件条件采用 8 GB RAM 的 Inter Core i7 处理器，主频为 3.6 GHz，未进行加速处理。

3.2 可视化分析

3.2.1 视频显著性检测

首先需评估显著性检测结果，尽管这不是最终的运动目标分割结果，但是作为所提算法的关键步骤，对其进行性能验证也是十分必要的。图 3 对比

了 5 种算法 [TMP (Approach Using Temporal Superpixels)^[10]、PQFT(Phase Spectrum of Fourier Transform)^[7]、ITTI^[6]、SR (Spectral Residual

approach)^[19] 和 SCUW (Spatiotemporal Cues and Uncertainty Weighting)^[13] 算法]对基于图像/视频显著性模型的检测结果。



图 3 6 种不同显著性检测算法的检测结果。(a)原图;(b)标注图像;(c) TMP 算法;(d) PQFT 算法;(e) ITTI 算法;
(f) SR 算法;(g) SCUW 算法;(h)所提算法

Fig. 3 Detection results by six different saliency detection methods. (a) Original image; (b) annotation image; (c) TMP method; (d) PQFT method; (e) ITTI method; (f) SR method; (g) SCUW method; (h) proposed method

图 3 中每一点的像素值(显著性值)代表该位置的超像素属于前景的概率。其中 TMP 算法和所提算法均采用了基于超像素的显著性检测方法,其余方法均基于像素级。可以看出,所有算法对运动速度较快的运动检测效果均比较理想,这也充分说明运动特征对视频显著性的影响较空间域特征更加明显。除 SR 模型之外,其余的模型对于复杂背景的抑制能力均较弱。比如在检测到降落伞的同时,背景光线也作为显著性区域被检测出来。而基于频域特征的 SR 模型对强光干扰的稳健性较好,但是无法确定目标轮廓及准确的位置。几乎所有的算法

对树林中飞鸟的检测都不是很理想,原因在于:单帧图像的显著性检测算法的基本思想是采用颜色、亮度和轮廓等特征与周围背景进行对比以确定显著区域,而在背景纹理复杂且树叶晃动干扰较大的情况,仅考虑相邻帧之间的时空域变化很难将运动目标的显著性完整描述。PQFT 算法在计算显著性时将空间特征和运动特征共同作为四元数,忽略了两种特征对运动显著性的贡献程度(权重),从而导致在检测到运动目标的同时引入了冗余的空间特征。利用颜色和运动梯度特征可以使空间域特征(如颜色、亮度等)差异较大及运

动特征发生突变的前景、背景得到较好的区分(如奔跑的小女孩和花豹)。此外,借助运动一致性约束,PQFT算法更注重运动目标在时间域上的持续变化情况,可将纹理复杂的背景分割得更细致(如树林中的飞鸟),同时也使得在强光背景干扰下依然保持较好的稳健性(如降落伞)。

3.2.2 运动目标检测

对显著性结果进行自适应阈值化处理可以定位到运动目标,将所提算法的最终分割结果与已有的运动目标检测算法进行对比,可验证其可行性及稳

健性。选取4种包含不同环境因素(包括阴影干扰、动态背景、目标与背景颜色相近和光照变化等)的视频进行测试,并对比了5种目前比较常见的运动目标检测算法的可视化结果,包括三帧差分法(TFD)^[17]、三帧差分与背景减除融合法(TFD+BS)^[18]、视觉背景检测(ViBe)算法^[19]、高斯混合模型(MoG)^[20]和非参数核密度估计(KDE)算法^[21]。为确保对比实验的公平性,所有方法均未进行形态学处理,而是仅剔除了噪声干扰(像素个数小于20的区域),其视觉效果对比如图4所示。

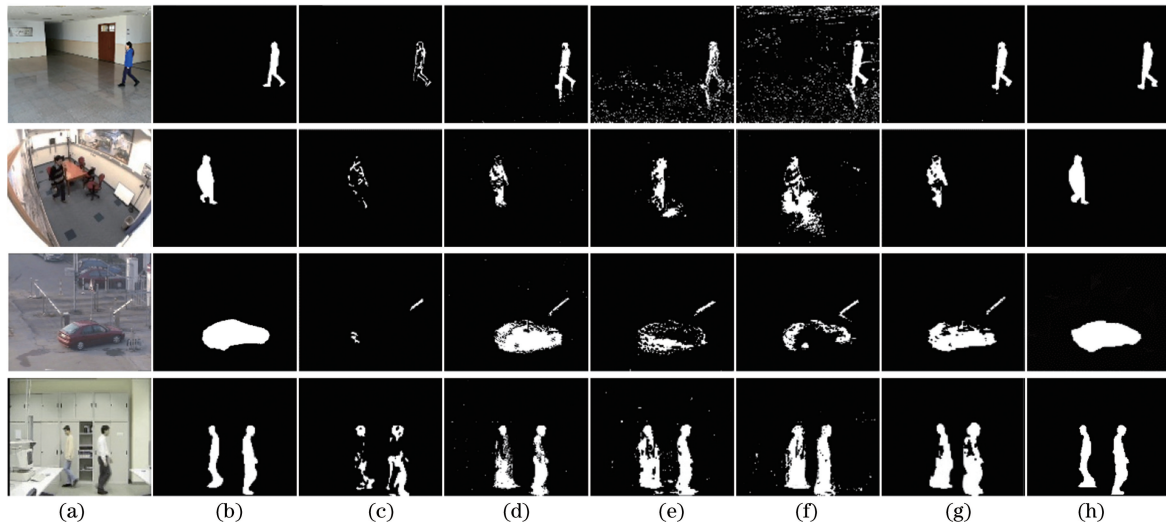


图4 6种不同运动检测算法的检测结果。(a)原图;(b)剪影;(c)TFD^[20];(d)ViBe^[21];(e)KDE^[22];(f)MoG^[23];(g)TFD+BS^[24];(h)所提算法

Fig. 4 Detection results by six different motion detection methods. (a) Original image; (b) silhouette; (c) TFD; (d) ViBe; (e) KDE; (f) MoG; (g) TFD+BS; (h) proposed method

从图4可以看出,对于室内窗边光线变化的情况,TFD能够较好地反映光线的变化。ViBe算法采用邻域像素来创建背景模型,通过比对背景模型和当前输入像素值来检测前景,可以避免由大面积的光照变化和视频细微抖动导致的误判,使得目标检测更加准确。MoG算法由于参数更新的复杂度较大,在光线突变的短时间邻域内很难及时恢复背景信息。对于包含阴影的室内场景,由于阴影与目标运动状态几乎同步,只注重相邻帧之间目标轮廓变化的TFD算法对阴影不太敏感。MoG和KDE算法将大部分阴影检测为前景,所得检测效果较差。由于其余4种方法均在子全局背景模型下进行目标检测,对动态背景更加敏感,作为背景的挡车杆的移动均被检测为前景,而所提算法利用运动一致性约束较好地剔除了这一虚假前景。由于TFD算法对运动较慢的目标不敏感,所以检测结果易出现空洞,不适合单独使用。

3.3 定量评估

通过定量分析进一步证明所提算法的优越性和时空信息融合的必要性。采用两种常用的评价指标对实验结果进行评估:准确率-召回率($P-R$)曲线和 F_β 。 F_β 表示 P 和 R 的加权调和平均数。 $P-R$ 曲线涉及准确率 P 和召回率 R 两个参数,其中 P 指被正确检测到的显著性(或前景)像素的百分比,而 R 指检测到的显著性像素的百分比。 P 和 R 可分别表示为

$$P = \frac{T_P}{T_P + F_P}, \quad (16)$$

$$R = \frac{T_P}{T_P + F_N}, \quad (17)$$

式中: T_P 表示被正确检测到的目标像素个数; F_P 表示背景被误判成运动目标的像素个数; F_N 表示运动目标被误判成背景的像素个数。

本实验中的阈值范围为0~255,对每种算法得

到的显著图进行二值化处理,并以每帧图像的地面实况(Ground Truth)为基准,通过(16)、(17)式分别计算 P 和 R ,即可绘制 P - R 曲线。考虑到 P 和 R 会出现矛盾的情况(图 5 中出现了两条 P - R 曲线发生交叉的现象),在此引入评价指标 F_β ,该值有效表征了显著性算法在提取显著性目标的同时又抑制非显著性目标的能力。 F_β 的计算公式为

$$F_\beta = \frac{(1 + \beta^2)PR}{\beta^2 P + R}, \quad (18)$$

式中: β 为权重系数,由于在算法的性能评估中更注重准确度的高低,故 $\beta < 1$ 且 $\beta^2 = 0.3$ 。

图 5 显示了上述 6 种显著性检测算法的 P - R 曲线。每条曲线包含了 256 对 P 和 R ,分别表示 0~255 之间所有阈值对应的检测结果,从中能够分析出不同分割阈值对每种显著性检测算法的影响。图 5 表明,当阈值接近 255(即 R 趋近于 1)且所有算法的 P 均取得最小值时,所提算法的最小 R 依然高于其他算法。这是因为所提算法通过融合时空信息能够有效地检测出强响应的运动区域。从曲线走势可以看出所提算法的 P - R 曲线几乎完全包住了其他算法的曲线,最优 P 大于 92%,且检测效率也表现出了明显的优势,这充分证明了所提算法的有效性。

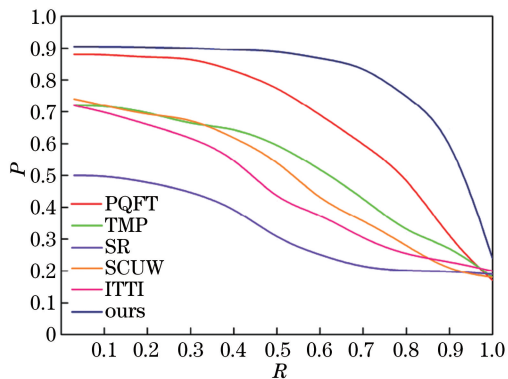


图 5 不同显著性检测方法的 P - R 曲线

Fig. 5 P - R curves by different saliency detection methods

利用自适应阈值分割方法对上述用于对比的显著性检测结果进行分割处理,并对 Ground Truth 图像和每种方法(包括显著性检测和运动目标检测)的分割结果图像均以像素为单位进行统计。利用(16)~(18)式计算分割后的每一帧二值图像的 P 、 R 和 F_β ,结果如图 6 所示。可以看出,所提算法在运动检测方面也表现出了较大的优势,这再次证明了基于时空特征的显著性检测算法的有效性。

4 结 论

提出了一种基于超像素时空特征融合的视频显

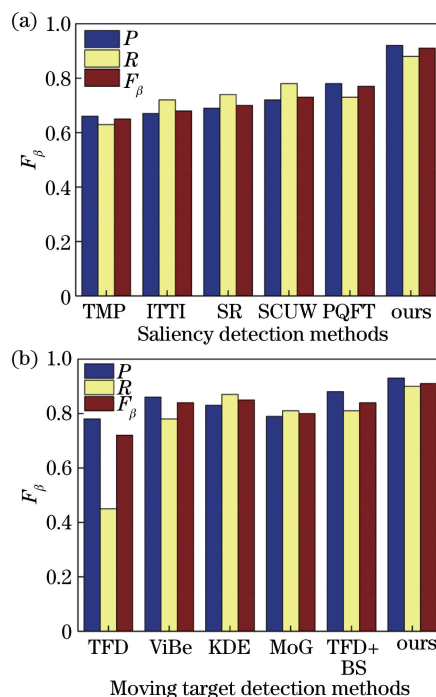


图 6 不同方法的 F_β 。(a)显著性检测;(b)运动目标检测

Fig. 6 F_β by different methods. (a) Saliency detection;

(b) moving target detection

著性检测算法,通过结合时空显著性和运动一致性,生成对背景干扰稳健性较强的视频显著性映射模板。具体过程为:首先计算像素级的静态边界特征和运动梯度特征,并将二者融合生成超像素级的时空梯度图;然后根据背景先验知识将加权测地距离作为显著性量度来估计每一个超像素相对于邻域的时空梯度显著性,得到时空显著图,对时间域上目标的运动模式进行分析,利用光流方向及幅值的熵值来估计超像素之间的运动一致程度,从而构建运动一致性图;最后将时空显著图与时间一致性图进行融合,并通过自适应阈值处理定位到视频中的运动目标。对 SegTrackV2 视频集及其他运动目标检测视频进行测试,检测精度优于 92%。将所提算法与主流的检测方法进行对比,结果表明所提方法在可靠性、稳健性和实时性等方面均具有优异的性能,且具有潜在的应用价值。

参 考 文 献

- [1] Borji A, Itti L. State-of-the-art in visual attention modeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 185-207.
- [2] Lee W F, Huang T H, Yeh S L, et al. Learning-based prediction of visual attention for video signals [J]. IEEE Transactions on Image Processing, 2011, 20(11): 3028-3038.

- [3] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [4] Zhang Q R, Xiao H M. Biologically motivated salient regions detection approach [C]. Second International Symposium on Intelligent Information Technology Application, 2008: 1100-1104.
- [5] Lin W S, Huang Y W. Intention-oriented computational visual attention model for learning and seeking image content [C]. IEEE Conference on Industrial Electronics and Applications, 2009: 1250-1255.
- [6] Itti L, Baldi P. A principled approach to detecting surprising events in video [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005: 631-637.
- [7] Guo C L, Zhang L M. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression [J]. IEEE Transactions on Image Processing, 2010, 19(1): 185-198.
- [8] Li X H, Lu H C, Zhang L H, *et al.* Saliency detection via dense and sparse reconstruction [C]. IEEE International Conference on Computer Vision, 2013: 2976-2983.
- [9] Liu Z, Zhang X, Luo S H, *et al.* Superpixel-based spatiotemporal saliency detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(9): 1522-1540.
- [10] Wang W G, Shen J B, Shao L. Consistent video saliency using local gradient flow optimization and global refinement [J]. IEEE Transactions on Image Processing, 2015, 24(11): 4185-4196.
- [11] Singh A, Henry C H, Pratt M A. Learning to predict video saliency using temporal superpixels [C]. International Conference on Pattern Recognition Applications and Methods, 2015: 201-209.
- [12] Rahtu E, Kannala J, Salo M, *et al.* Segmenting Salient Objects from Images and Videos [M]//Rahtu E, Kannala J, Salo M, *et al.* eds. Heidelberg: Springer Berlin Heidelberg, 2010: 366-379.
- [13] Fang Y M, Wang Z, Lin W S. Video saliency incorporating spatiotemporal cues and uncertainty weighting [C]. IEEE International Conference on Multimedia and Expo (ICME), 2013: 3910-3921.
- [14] Li Q W, Zhou Y Q, Ma Y P, *et al.* Salient object detection method based on binocular vision [J]. Acta Optica Sinica, 2018, 38(3): 331-343.
李庆武, 周亚琴, 马云鹏, 等. 基于双目视觉的显著性目标检测方法 [J]. 光学学报, 2018, 38(3): 331-343.
- [15] Brox T, Malik J. Large displacement optical flow: descriptor matching in variational motion estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(3): 500-513.
- [16] Wei Y C, Wen F, Zhu W J, *et al.* Geodesic saliency using background priors [M]//Wei Y C, Wen F, Zhu W J, *et al.* eds. Heidelberg: Springer Berlin Heidelberg, 2012: 29-42.
- [17] Yang X, Zhang L L, Liang Y M. Salient region detection based on super-pixels and shortest geodesic distance after L_0 smoothing [J]. Journal of Optoelectronics·Laser, 2017, 28(6): 657-662.
杨鑫, 张雷雷, 梁艳梅. 基于 L_0 平滑的超像素块最短测地距离的显著区域提取方法 [J]. 光电子·激光, 2017, 28(6): 657-662.
- [18] Tsai D, Flagg M, Nakazawa A, *et al.* Motion coherent tracking using multi-label MRF optimization [J]. International Journal of Computer Vision, 2012, 100(2): 190-202.
- [19] Hou X D, Harel J, Koch C. Image signature: highlighting sparse salient regions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 194-201.
- [20] Li Q L, He J F. Vehicles detection based on three-frame-difference method and cross-entropy threshold method [J]. Computer Engineering, 2011, 37(4): 172-174.
李秋林, 何家峰. 基于三帧差法和交叉熵阈值法的车辆检测 [J]. 计算机工程, 2011, 37(4): 172-174.
- [21] Barnich O, van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences [J]. IEEE Transactions on Image Processing, 2011, 20(6): 1709-1724.
- [22] Wang X B, Liu C P, Fei L Y, *et al.* Foreground object detection method using kernel density estimation of a local spatio-temporal model [J]. Journal of Image and Graphics, 2012, 17(7): 813-820.
王兴宝, 刘纯平, 费兰英, 等. 局部时空域模型的核密度估计目标检测方法 [J]. 中国图象图形学报, 2012, 17(7): 813-820.
- [23] Suo P, Wang Y J. An improved adaptive background modeling algorithm based on Gaussian Mixture Model [C]. 9th International Conference on Signal Processing, 2008: 1436-1439.
- [24] Zhang Z H, Xu Y, Su Q, *et al.* Moving object detection based on background subtraction and hybrid frame differencing [J]. Radio Engineering, 2012, 42(8): 14-17.
张正华, 许晔, 苏权, 等. 基于背景差分和混合帧差的运动目标检测 [J]. 无线电工程, 2012, 42(8): 14-17.