面向多种场景的视频对象自动分割算法^①

余欣纬¹, 柯余洋², 熊 焰¹, 黄文超¹

¹(中国科学技术大学 计算机科学与技术学院, 合肥 230027) ²(合肥学院 计算机科学与技术系, 合肥 230000)

摘 要:针对当前应用于视频对象分割的图割方法容易在复杂环境、镜头移动、光照不稳定等场景下鲁棒性不佳 的问题,提出了结合光流和图割的视频对象分割算法.主要思路是通过分析前景对象的运动信息,得到单帧图像上 前景区域的先验知识,从而改善分割结果.论文首先通过光流场采集视频中动作信息,并提取出前景对象先验区域, 然后结合前景和背景先验区域建立图割模型,实现前景对象分割.最后为提高算法在不同场景下的鲁棒性,本文改 进了传统的测地显著性模型,并基于视频本征的时域平滑性,提出了基于混合高斯模型的动态位置模型优化机制. 在两个标准数据集上的实验结果表明,所提算法与当前其他视频对象分割算法相比,降低了分割结果的错误率,有 效提高了在多种场景下的鲁棒性.

关键词:视频对象分割;光流;图割;测地显著性;混合高斯模型

引用格式: 余欣纬,柯余洋,熊焰,黄文超.面向多种场景的视频对象自动分割算法.计算机系统应用,2017,26(11):152-158. http://www.c-sa.org.cn/1003-3254/6044.html

Automatic Video Object Segmentation Algorithm for Multiple Scenes

YU Xin-Wei¹, KE Yu-Yang², XIONG Yan¹, HUANG Wen-Chao¹

¹(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China) ²(Department of Computer Science and Technology, Hefei University, Hefei 230000, China)

Abstract: Aiming at the problems of poor robustness in the complex environment, lens movement and light instability, a video object segmentation algorithm combining optical flow and graph cutting is proposed. The main idea is to improve the segmentation result by analyzing the motion information of the foreground object and obtaining the prior knowledge of the foreground area on the single frame image. Firstly, the motion information in the video is collected by the optical flow field, and the prior knowledge of the foreground object is extracted. Then, the foreground object segmentation is realized by combining the priori areas of foreground and background. Finally, in order to improve the robustness of the algorithm in different scenarios, this paper improves the traditional geodesic saliency model, and employs the dynamic position model optimization mechanism based on Gaussian Mixture Model based on the intrinsic temporary smoothness of video. Experimental results on two benchmark datasets show that the proposed algorithm reduces the error rate of the segmentation results compared with other video object segmentation algorithms, which effectively improves the robustness in many scenarios.

Key words: video object segmentation; optical flow; graph cut; geodesic saliency; Gaussian mixture model

近年来,随着计算机存储和处理能力、网络带宽 和多媒体显示以及拍摄设备的不断发展,人们拍摄和 获取视频的能力不断增强.视频由于其信息的丰富性和生动性而被越来越多的人所接受,成为人们在生活

① 收稿时间: 2017-02-21; 修改时间: 2017-03-09; 采用时间: 2017-03-13

¹⁵² 软件技术·算法 Software Technique Algorithm

中娱乐、学习、记录的重要传播载体.这些不断增长 的海量视频数据资源带来了信息冗余和如何高效管理 的问题.因此,不断增长的大规模视频数据及其应用也 促使了获取、分析以及理解这些视频数据的需求日益 增长.如何降低视频内容理解难度,提取出视频中的关 键信息成为学术界和工业界关心的焦点.而视频对象 分割因为其具有能够有效处理视频中主要信息的特点 在视频摘要、视频检索、视频活动分析等领域拥有广 泛的应用.视频对象自动分割算法因为不需人工参与, 可以在多种大规模视频数据处理场景中应用的特点, 成为近年来的研究热点.

当前,自动化视频对象分割方法有诸多研究思路, 包括:基于对象候选集筛选的方法、基于点追踪和聚 类的方法以及基于图割的方法.基于图割的方法主要 是通过将图像或视频分割问题转换为图的最小割问题, 实现对图像或视频的前景对象分割.Khoreva等^[1]提出 了使用分类器计算图中边权值的方法,优化并提升了 现有方法的分割结果.Ma等^[2]利用在图的最大团算法, 针对多处前景对象候选区域评分,得到分割结果. Zhang等^[3]构造了一个分层的有向无环图来选取最终 的目标区域,并利用视频帧间的光流信息对目标区域 进行扩张,目标分割精度有所提升.但该方法是以像素 点作为基本单元构建马尔科夫随机场,分割效率较低. 尽管现有的自动化分割方法能够有效地检测并分割出 视频前景对象区域,但在精度和场景鲁棒性上仍有较 大提升空间.

为解决现有图割方法在多种场景中分割能力不佳 的问题,本文提出了一个在多种场景中具有更好准确 性和鲁棒性的视频对象分割算法.首先,通过分析视频 中的运动信息,获得前景对象的先验知识,为之后的分 割过程提供重要信息.在得到前景对象先验知识的基 础上进行分割,能够较好地排除因背景因素导致的干 扰,提高不同场景下的分割精度.然后,本文改进了传 统的测地显著性计算模型^[4],结合前景和背景先验区域 计算得到分割结果.最后,为了解决一些场景中分割结 果不佳的问题,本文基于混合高斯模型建立了动态位 置模型优化机制,利用前景对象本征的位置平滑性优 化分割结果.

1 前景对象先验知识提取

本文利用视频的动作信息分析得到前景对象的动

作轮廓,在动作轮廓中包含的区域即为所求前景区域 先验知识.相较于背景区域,符合人类认知特点的前景 对象一般在图像和动作特征上具有独特性.首先,本文 采用光流 (Optical flow) 计算得到视频序列中的动作模 型.通过对光流向量梯度计算得到前景对象动作轮廓. 针对传统梯度计算模型产生的背景噪声干扰问题,本 文提出考虑向量长度和方向的混合梯度计算模型,能 够较好地排除背景噪声和镜头动作的干扰.然后,本文 改进了点包容性 (Point in Polygon, PIP) 算法^[5,6]解决从 不完整连续动作轮廓中提取前景先验区域的问题.最 后,在某些场景中可能因为镜头抖动、背景遮挡或前 景对象本身动作信息不明显而导致光流计算出现错误. 针对上述问题,本文提出基于时域平滑性的前景区域 优化方案,确保算法在上述场景中的鲁棒性.



图 1 本文算法流程图

1.1 光流

为了精确地提取视频中的动作信息,本文采用了 光流场算法⁽⁷⁾建立视频的动作模型.光流是指时变图像 中的模式运动速度.因为当物体在运动时,它在图像上 对应点的亮度模式也在运动.光流表达了图像的变化, 由于它包含了目标运动的信息,因此可被观察者用来 确定目标的运动情况.由光流的定义可以引申出光流 场,它是指图像中所有像素点构成的一种二维瞬时速 度场,其中的二维速度矢量是景物中可见点的三维速 度矢量在成像表面的投影.因此光流可以较为准确地 描述被观察物体的运动信息.

本文对视频序列中相邻的帧图像计算得到光流场. 光流场由光流向量组成,每个光流向量描述了对应像 素点在相邻帧之间运动的方向和程度.

1.2 梯度计算模型

令 $\overline{o_p^r}$ 是根据相邻单帧图像计算得到的光流图像 *t* 中对应像素点 *p* 的光流向量. 传统的梯度计算求解公

Software Technique Algorithm 软件技术 算法 153

式为:

计算机系统应用

$$g_p^t = 1 - \exp(-\lambda^g \left\| \nabla \overrightarrow{o_p^t} \right\|) \tag{1}$$

式中, $g_p^t \in [0,1]$ 代表在像素点 p的光流向量梯度值; λ^s 代表用于控制斜率的参数.在本文方法中, λ^s 默认设 置为 0.7.

然而,上述传统梯度计算方法更容易受到镜头移 动或晃动、背景噪声的干扰.在图 2 中,源图像序列中 的动作信息包含:前景对象的跑动动作、背景噪声以 及镜头的持续移动.根据(b)中梯度计算结果可以发 现,式(1)虽然能够较好地提取出前景对象的运动轮 廓,但同样容易受到镜头运动和背景噪声的干扰.使得 (1)无法从光流图像中得到正确的前景对象先验知识.



154 软件技术·算法 Software Technique Algorithm

本文将光流向量的方向纳入计算模型中,提出了 混合梯度计算模型,较好地排除了镜头和背景因素的 干扰.基于对像素点光流向量特性的观察,本文认为: 当某一像素点处于不同物体的轮廓上时,由于不同物 体的运动特征不一致,因而该像素点对应的光流向量 应与相邻像素点的向量方向有较大的夹角.反之,如果 某一像素点与相邻像素点都处于同一物体内部,则该 像素点对应的光流向量方向会与相邻像素点一致.因 此,本文提出基于向量方向的梯度计算公式:

$$d_p^t = 1 - \exp(-\lambda^d \frac{1}{C} \sum_{q \in N} (\delta \theta_{p,q}^2))$$
(2)

式中, N 代表像素点 p 的相邻像素点集合; $\delta\theta_{p,q}^2$ 代表光 流向量 \vec{f}_p 和 \vec{f}_q 之间夹角的 L2 范数值; C 表示像素点 p 的相邻像素个数, 在本文方法中, C 的值默认设定为 4. 通过式 (2) 可知, 光流向量与相邻向量夹角越大, 对 应像素点计算得到的梯度值越大.

本文方法中结合了两种计算方法的混合梯度计算 模型如下:

$$b_p^t = \begin{cases} g_p^t, & g_p^t \ge HT \\ g_p^t \cdot d_p^t, & LT < g_p^t < HT \end{cases}$$
(3)

式中, HT 代表较高的阈值, 作用是将强烈的动作信息 与可能出现混淆的像素点区分开. 梯度值低于 HT 的像 素点被认为难以分辨是否处于前景对象轮廓边缘, 因 而需要结合式 (2) 进一步计算. LT 代表较低的阈值, 作 用是判断像素点是否存在动作信息, 梯度值低于 LT 的 像素点将被视为噪声干扰. HT 与 LT 均为自适应阈值. 通过使用混合梯度计算模型, 能够较好地得到前景对 象运动轮廓.

1.3 先验区域标注

在不设约束的任意视频中,人们难以预测前景对 象的动作模式.由于动作模式的不确定性,通过梯度计 算模型得到前景对象轮廓通常是不连续的.为从运动 轮廓中标注出所需的前景先验区域,我们改进了点包 容性算法^[5,6].算法原理如图 3 所示,从平面上某一点向 任意方向发出射线,如果射线与平面上的多边形边的 交点个数均为奇数,则该点处于某一多边形内部;若交 点个数均为偶数,则处于任意多边形的外部.针对动作 轮廓不连续的问题,本文参照点包容性算法原理,对每 个像素点计算水平、垂直、斜 45 度等八个方向的射 线交点数,并采用多数投票法实现了正确的标注.

1.4 前景先验区域优化

为解决特殊场景下光流图像错误率较高导致无法 提取先验区域的问题,本文提出了前景先验区域优化 方案.在真实拍摄的视频中,容易出现镜头移动速度过 快、镜头剧烈晃动、对焦不清出现模糊等问题.这些 干扰会对光流计算产生较大影响,导致光流图像错误 率较高,无法准确地得到前景对象先验区域.针对这一 问题,本文提出的优化方案能够利用前景对象的时域 平滑性较好地优化先验区域.由于前景对象在视频序 列的时域上具有平滑性,因而可以利用相邻帧中先验 区域结果补全错误帧的先验区域.本文考虑到在利用 连续性补全的过程中,越是相邻的帧的先验区域越有 更高的置信度.因而本文以需要优化的错误帧为对称 轴建立高斯模型,更接近错误帧的帧图像像素点具有 更高的权重.优化方案计算公式如下:

$$L_{s}^{t} = \frac{1}{C} \sum_{i=-N}^{N} \left[\frac{p_{s}^{i} \cdot \lambda_{1}}{\sigma} \cdot \exp\left(-\frac{\lambda_{2} \cdot i^{2}}{2\sigma^{2}}\right) \right]$$
(4)

式中, L's代表特殊场景帧 t 上像素点 s 属于前景先验区 域的概率. C 代表归一化常数值. N 代表相邻帧的数量, 在本文方法中, N 的值设置为 3. p's表示在相邻帧 i 上 像素点 s 是否属于前景先验区域, 取值为 0 或 1. σ 代 表相邻帧序列上高斯分布的方差. λ₁、λ₂ 为常数系数. 通过应用前景先验区域优化模型, 能够增强算法的鲁 棒性.



(a) 原始图像



(c)优化前的 前景先验区域



(d) 优化后的 前景先验区域



2 测地显著性模型

测地显著性模型属于图割模型的一种. 传统测地 显著性模型[4]通过在图像上建立图结构,计算节点到背 景区域的测地距离,根据距离大小实现前景区域的分 割. 在中心偏移假设^[8]的基础上, 传统模型将图像的边 缘区域标记为背景先验区域,并计算其他节点到边缘 节点的距离.由于前景对象相较于背景区域在外观、 动作和位置等方面具有特征独特性,因而与边缘背景 区域距离越远的节点,其前景显著值越高.本文改进了 传统测地显著性模型,基于前景和背景先验区域分别 计算距离,并提出了新的显著值计算模型.首先,本文 对单帧图像进行过分割处理得到超像素图像,并在此 基础上建立图结构. 然后, 通过计算和比较未标记节点 分别到前景和背景先验区域的距离,实现对前景区域 的分割.为了解决某些场景下中心偏移假设效果不佳 的问题,本文提出了多重显著值计算模型,使得算法在 一些特殊场景下仍有较好的分割结果.

2.1 图的建立

本文采用超像素作为单帧图像中进行前景区域分 割的最小单位.超像素由一系列位置相邻且颜色、亮 度、纹理等特征相似的像素点组成的小区域,这些小 区域最大程度上保证了区域内所有像素点都属于同一 个物体.在一些复杂场景下,以超像素为单位实现前景 对象分割能够较好地保留物体的边缘特性,获得更好 的分割结果.本文采用图割模型实现前景区域分割,以 超像素为节点构图能够显著降低图结构的复杂度,提 高建模和优化的效率.本文采用 SLIC 算法^[9]实现超像 素过分割处理, SLIC 能够较好地保持图像局部的形状 完整性和外观相似性.

在单帧图像 Tⁱ上,本文以超像素为节点,相邻超像 素建立边建立了图结构Gⁱ = {Vⁱ, Eⁱ}.其中 Vⁱ代表 Tⁱ上 的超像素集合, Eⁱ代表图中相邻节点之间边的集合.在 本文中, Eⁱ中每条边的权值由超像素间外观特征差异 决定.权值wⁱ_{ik}计算公式如下:

$$w_{j,k}^{i} = \begin{cases} \left\| x_{j} - x_{k} \right\|, & k \in N(j) \\ 0, & j = k \text{ or otherwise} \end{cases}$$
(5)

式中, x_j和 x_k分别代表在节点 j和 k内部的平均颜色 特征值.本文通过对节点包含的所有像素点 RGB 颜色 空间值求均值得到平均颜色特征值. N(j)代表节点 j的 相邻节点集合.为方便计算,本文建立|Vⁱ|×|Vⁱ|的邻接

Software Technique Algorithm 软件技术 算法 155

矩阵 W 以存储 E 中的边权值.

$$W^{i}(k,j) = w^{i}_{k,j} \tag{6}$$

2.2 多重显著值计算模型

基于中心偏移假设[8],本文将图像的边缘节点结合 标记为背景先验区域 B. 结合先前对动作信息分析得 到的前景先验区域F,传统测地显著性模型^[4]通过计算 和比较剩余未标记节点集合 U 分别与 B 和 F 之间的 测地距离,实现对前景对象的分割,单帧图像 i 中节点 j的显著值 S^{i}_{i} 计算公式如下:

$$S_{j}^{i} = \frac{1}{C} \left(\min_{k \in B} \left(d_{geo}\left(j,k\right) \right) - \lambda \cdot \min_{k \in F} \left(d_{geo}\left(j,k\right) \right) \right)$$
(7)

式中, C 代表归一化的常数参数. dgeo (j,k)代表节点 $j,k \in V^i$ 之间的测地距离. λ 是常系数, 用于平衡两项距 离值对所得显著值的贡献比,在本文方法中被设置为 1. 从式(7)中可以看出, Sⁱ的大小与到达 B的距离成正 比,与到达F的距离成反比.因而如果图中节点与背景 先验区域特征相差较大且与前景先验区域特征接近, 则该节点会获得较高的显著值,符合前景区域的特点. 节点 s 和 t 之间的测地距离dgeo(s,t)计算公式如下:

$$d_{geo}(s,t) = \min_{p \in P(s,t)} \left(\sum_{e^i(n1,n2) \in p} \left| W^i(n1,n2) \right| \right)$$
(8)

式中, P(s,t)代表从源节点 s 到目标节点 t 的所有可达 路径. eⁱ(n1,n2)代表路径 p 上的相邻节点 n1 和 n2 之间 的边.



(a) 原始图像



(c) 基于四条边缘的 显著值图像

图 5 多重显著值模型对比图

(d) 多重显著值模型

在实际场景中,一些视频中的前景对象会部分出 现在单帧图像边缘上.对于包含这些场景的视频,应用

156 软件技术·算法 Software Technique Algorithm

传统模型无法得到正确的分割结果,针对这一问题,本 文提出了多重显著值计算模型. 通过分别计算未标记 节点到每一条边缘的测地距离,得到四幅显著值图像, 将其融合后得到最终的显著值图像.因而本文改进了 式(7)中背景测地距离计算项.

$$S_{final} = S_{up} \times S_{down} \times S_{left} \times S_{right}$$
(9)

式中, Sup、Sdown、Sleft、Sright 分别代表以单帧图像四条 边缘为先验背景区域而计算得到的显著值图像,这些 图像通过 (7) 计算得到.

此外,本文观察到图像边缘中的前景区域一般来 说并不大.因而在多重显著值计算之前,本文加入了对 图像边缘的预处理,以提高背景先验区域的准确性.本 文基于颜色和空间特征对每条图像边缘进行聚类,并 将一部分颜色特征相异于大多数节点的簇舍去,取剩 下的节点作为先验背景区域,在本文方法中,采用了基 于密度的 DBSCAN 聚类算法实现.

3 动态位置模型优化

本文采用多帧差分的混合高斯模型 (Gaussian Mixture Model, GMM) 建立了前景对象的动态位置模 型,以提高算法在多种场景下的鲁棒性. Stauffer 等^[10] 提出了基于混合高斯模型的背景建模方法,通过在每 个像素点上建立 K 个高斯分布, 实现图像分割. 在某一 时刻,像素点{x0,y0}的历史像素值集合为:

$$\{X_1, X_2, \cdots, X_t\} = \left\{ I(x_0, y_0^i), 1 \le i \le t \right\}$$
(10)

式中, i 为视频序列, X; 为像素点{x0,y0}在 i 时刻的像素 值. 当前像素点观测值的概率为:

$$P(X_t) = \sum_{j=1}^{K} \omega_{j,t} \cdot \eta(X_t, \mu_{j,t}, \Sigma_{j,t})$$
(11)

式中,K代表混合高斯模型中分布数量; ω_{it}代表第 j个 分布在时刻 t 的权重值; μ_{it} 和 $\Sigma_{it} = \sigma_{it}^2 E$ 代表第 j 个高 斯模型的均值和协方差, $\sigma_{j,t}$ 代表对应分布的标准差; $\eta(X_{t},\mu_{it},\Sigma_{it})$ 代表高斯概率密度函数, 计算公式为:

$$\eta(X_t, \mu_{j,t}, \Sigma_{j,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{j,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_{j,t})^T \Sigma_{j,t}^{-1}(X_t - \mu_{j,t})}$$
(12)

式中,n代表X,的维度.

首先, 文献[10]将混合高斯模型中的若干个分布按 照优先级 $\rho_{i,t} = \omega_{i,t}/\sigma_{i,t}$ 从大到小排列. 然后, 用像素点 当前值 X,按序与混合高斯模型中所有分布相比较.若 满足公式(13),则认为该像素点与其中某一分布匹配, 并对匹配分布的参数 ω_{it} 、 μ_{it} 和 σ_{it} 进行更新.其他不 匹配的分布只改变自身权重值 ω_{it} .

$$\left|X_{t} - \mu_{j,t}\right| < \delta \cdot \sigma_{j,t} \tag{13}$$

式中,δ一般取值为2.5~3.5.

若与混合高斯模型中所有分布均不匹配,则算法 将根据当前像素值建立一个新的高斯分布(均值为X₁, 初始化标准差及权重). 如果当前分布数小于 K. 则新的 高斯分布直接添加到混合高斯模型中.如果当前分布 数等于 K,则新的分布将取代优先级ρ_{it}最小的分布.

本文采用混合高斯模型对视频序列建模,用单个 或多个高斯分布表示前景对象的动态位置.此外,本文 加入多帧差分优化建模结果,保证了动态位置模型的 时域平滑性.

4 实验结果及分析

为了验证本文算法的有效性和鲁棒性,在两个标 准数据集 SegTrack 和 SegTrack v2 上对本文算法和其 他若干最优算法做了对比实验. SegTrack 和 SegTrack v2 共包含 14 个不同的视频序列及超过 1000 幅帧图 像,并涵盖了多种不同场景的视频,如:镜头快速移 动、复杂环境和光照不佳等.能够较为全面地衡量算 法在不同场景下的分割能力.实验中,本文选取文献 [3]和[11]中算法的结果作对比. 上述算法均为基于图 割的视频对象分割算法并有较好分割能力,本文实验 环境为: Intel i5-3450 @3.10 GHZ, 8 GB 内存, Windows 10 环境下使用 MATLAB 2015a 实现算法.



图 6 中,前景对象处于复杂环境中,且背景环境中 存在外观和纹理特征与前景物体相似的区域. 文献[11] 中算法严重受到了背景噪声的干扰,无法正确定位前

景对象的位置. 文献[3]算法结果未能完整分割出前景 物体轮廓.本文算法分割结果较为完整准确.

图 7 和图 8 中,视频序列的拍摄镜头随前景对象 快速移动,会产生大量的背景噪声. 文献[11]的结果受 到了噪声的较多影响,在前景对象周边产生了模糊.相 较于文献[3],本文的分割结果更为完整,且较好地排除 了噪声的干扰.



(c) 文献[11] (d) 文献[3] (e) 真值图 (b) 本文

图 7 镜头快速移动场景的分割结果



图 8 前景对象动作特征复杂的分割结果

图 9 中,视频中整体光照条件不佳,且不同的区域 明暗对比较为明显.本文算法在光照条件变动较为剧 烈的情况下,仍保持了较好的准确性和鲁棒性.



此外,以识别误差为标准在 SegTrack 数据集上比 较了本文算法与文献[3]和[12]中算法的结果. 识别误 差计算公式如下:

$$error(S) = \frac{|XOR(S,GT)|}{F}$$
 (14)

Software Technique Algorithm 软件技术 算法 157

式中, S 代表任一算法的最终分割结果; GT 代表数据集 中提供的真实分割对照; F 代表视频序列中帧图像的 数量.通过计算每种算法在单个帧图像上的平均错分 类像素个数, 识别误差能够较好地衡量不同算法的分 割能力.在同等实验条件下比较结果如表 1 所示.

表1 SegTrack 数据集上平均每帧错误率对比

视频序列	本文	文献[3]	文献[12]
Birdfall	256	155	189
Cheetah	628	633	806
Girl	966	1488	1698
Monkeydog	330	365	472
Parachute	162	220	221
Avg.	386	452	542

实验结果表明,通过前景先验区域提取和改进后 的测地显著性模型相结合,能够在前景对象动作模式 较复杂的情况下较好地获得前景信息,也增强了算法 在镜头移动或晃动场景中的鲁棒性,同时提高了算法 在复杂背景环境中的分割能力.

综上所述,本文提出的算法相较于以前提出的图 割算法具有更好的准确性.同时对多种视频中复杂环 境、镜头移动以及光照变化的场景有较好的鲁棒性.

5 结语

提出了一种面向多种场景的视频对象自动分割算 法.针对当前图割方法容易受到复杂环境、背景噪声 等因素影响的问题,通过分析视频动作信息获得前景 先验区域,为图割建模提供了重要的先验知识.同时改 进了传统测地显著性模型,使算法能够应用于更多场 景中.建立了动态位置模型,减少了背景因素对分割结 果的干扰.实验结果证明了本文所提算法准确性相较 于当前图割算法有了较好的提升,并在多种场景的视 频中有更鲁棒的分割结果.

参考文献

1 Khoreva A, Galasso F, Hein M, et al. Classifier based graph construction for video segmentation. Proc. of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015. 951-960.

- 2 Ma TY, Latecki LJ. Maximum weight cliques with mutex constraints for video object segmentation. Proc. of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA. 2012. 670–677.
- 3 Zhang D, Javed O, Shah M. Video object segmentation through spatially accurate and temporally dense extraction of primary object regions. Proc. of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA. 2013. 628–635.
- 4 Wei YC, Wen F, Zhu WJ, *et al*. Geodesic saliency using background priors. Proc. of the 12th European conference on computer vision. Florence, Italy. 2012. 29–42.
- 5 Sutherland IE, Sproull RF, Schumacker RA. A characterization of ten hidden-surface algorithms. ACM Computing Surveys, 1974, 6(1): 1–55. [doi: 10.1145/356625.356626]
- 6 Shimrat M. Algorithm 112: Position of point relative to polygon. Communications of the ACM, 1962, 5(8): 434.
- 7 Liu C. Beyond pixels: Exploring new representations and applications for motion analysis[Ph. D. Thesis]. Cambridge, MA: Massachusetts Institute of Technology, 2009.
- 8 Tatler BW. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision, 2007, 7(14): 4. [doi: 10.1167/7.14.4]
- 9 Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274–2282. [doi: 10.1109/TPAMI.2012.120]
- 10 Stauffer C, Grimson WEL. Learning patterns of activity using real-time tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747–757. [doi: 10.1109/ 34.868677]
- 11 Papazoglou A, Ferrari V. Fast object segmentation in unconstrained video. Proc. of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, NSW, Australia. 2013. 1777–1784.
- 12 Wang WG, Shen JB, Porikli F. Saliency-aware geodesic video object segmentation. Proc. of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015. 3395–3402.