

基于目标识别与显著性检测的图像场景多对象分割

李青¹ 袁家政¹ 刘宏哲²

(北京联合大学计算机技术研究所 北京 100101)¹

(北京联合大学北京市信息服务工程重点实验室 北京 100101)²

摘要 提出了一种基于目标识别与显著性检测的图像场景多对象分割方法。该方法的步骤包括:在图像训练集上训练语义对象的检测器,用来检测输入图像中对象的位置,标定对象的包围盒;对输入的图像进行过分割处理,得到超像素集合,根据包围盒的位置和超像素的语义概率值计算兴趣区域;在3种稠密尺度上进行场景显著性检测,得到输入图像的显著图;在兴趣区域内计算超像素的邻接关系,形成邻接矩阵,构建条件随机场模型,将多对象分割问题转化成多类别标记问题,每一个对象是一种类别;以每个超像素作为场模型的节点,超像素的邻接关系对应场模型中节点之间的连接关系,将显著性和图像特征转化为节点和边的权重值;利用图割算法,在条件随机场模型上进行优化,迭代终止时得到像素的对象标记结果,从而实现对多个对象的分割。实验结果表明该方法效果较好。

关键词 图像分割,语义标记,对象推理

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.05.057

Multi-object Segmentation of Image Scene Based on Object Recognition and Saliency Detection

LI Qing¹ YUAN Jia-zheng¹ LIU Hong-zhe²

(Institute of Computer Technology, Beijing Union University, Beijing 100101, China)¹

(Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China)²

Abstract This paper proposed a multi-object segmentation method of image scene based on object recognition and saliency detection. The object detector is learned on the training set, and then is used to locate the object in the test image with visualization of its bounding box. The test image is over-segmented into a set of superpixels. According to the location of bounding box and the superpixel-level probabilities, the region of interest is fixed. Then, a saliency map is obtained through a three-scale saliency detection. In the region of interest, a CRF model is established among the neighboring superpixels, whose nodes indicate the superpixels and edges indicate their neighborhood. The saliency of a superpixel is embedded into the weight of relative node, and the feature difference between two neighboring superpixels is embedded into the weight of relative edge. Thus, the multi-object segmentation task is transformed into a multi-labeling task. Finally, the CRF formulation is optimized using graph cut algorithm to get the multi-object segmentation result. The experimental results show the good performance of our method.

Keywords Image segmentation, Semantic labeling, Object reasoning

1 引言

在科技发展的推动下,图像场景内容的理解、处理、加工逐渐发展为计算机视觉、模式识别、人工智能、虚拟现实等领域及其交叉方向的国际前沿研究热点。其中,图像场景语义分割是图像场景内容理解的关键问题,它涉及到很多应用行业,如智慧旅游、智能交通、公共安全、影视娱乐等;同时它也是难点问题,表现在:由于图像内容外观特征变化多样,同一

种语义类别的对象有可能表现出不相似的外观特征,不同语义类别的对象也有可能表现出相似的外观特征,这为计算机识别出不同语义类别的对象并准确分割出对象的轮廓区域带来了挑战。

图像场景语义分割的准确定义是为图像中的每一个像素赋予一个语义类别的标记,从而将图像场景中不同的对象内容识别并且分割出来,它是建立在识别的基础上的。2006年,微软剑桥研究院的 Shotton 等提出了语义标记的概念,并

到稿日期:2016-08-07 返修日期:2016-11-13 本文受国家自然科学基金(61502036),北京市教委科技计划一般项目(KM201611417015),北京市信息服务工程重点实验室开放课题(Zk20201502)资助。

李青(1983—),女,博士,讲师,CCF会员,主要研究方向为计算机视觉、图像处理,E-mail:liqing10@bnu.edu.cn(通信作者);袁家政(1971—),男,博士,教授,主要研究方向为图形图像处理、文物遗迹的数字化处理、数字博物馆等;刘宏哲(1971—),女,博士,教授,主要研究方向为图像处理、多媒体语义计算等。

实现了将对象识别、分割与标记集于一体的场景类别级语义分割算法^[1],该算法开创了场景语义分割的里程碑,并形成了基于参数学习的语义分割模式。随后,许多研究工作都着力于这种有参数的方式,涉及到街景^[2]、自然场景^[3]、室内场景^[4]等众多种类的图像场景。

随着互联网图像的爆发式增长,对图像场景内容理解的需求也在不断提高。传统的有参数学习方式需要固定的数据集(包括图像数量固定和语义类别固定),在面对海量图像数据时,这种方式存在着一定的局限,对于新的语义类别,需要重新学习模型参数。2009年,麻省理工学院的Liu等提出了一种非参数的场景解析方法,并首次将这种方法定义为语义迁移方法^[5]。它的思想是,利用已经标注语义信息的相似图像将语义信息迁移到待分析的图像上。因此,语义迁移的关键点是如何准确实现待分析图像和已标记图像之间的像素级匹配。后续关于语义迁移的研究工作也主要集中在这一点上^[6-9]。

目前,不管是有参数学习还是非参数解析,大多数场景语义分割方法都还处于场景类别层次的语义分割阶段,即划分出场景中不同语义类别内容的区域,而对场景中对象的语义分割的关注较少。当场景中出现同一类别的多个对象时,场景类别层次的语义分割将这多个对象划分为“一个对象”。目前,在单个对象分割领域,国内外学者提出了一些较为有效的场景对象分割方法,包括对象抠取算法^[10-11]、对象共分割算法^[12-15]等。例如,微软剑桥研究院的Rother等提出了对象共分割的概念^[12-13],即将多张图像中的相似对象同时分割出来;明尼苏达大学的Bai等提出了一系列图像和视频对象的抠取算法^[16-17],它们通过用户交互提供潜在的对象先验信息。但是这些方法基本上都是围绕着图像中的视觉内容进行划分的,缺少语义信息,继而也缺少语义所对应的先验信息,即无法识别和理解所分割对象的语义。

随着行业技术的发展,这种类别级的划分越来越难以满足实际应用对场景内容和对象区域的需求。例如,存在这样的需求:在安全监控系统中,当场景中有多个个人出现时,需要算法把每个人都划分出来,以便于进一步分析。然而对于同一个场景中出现多个同类别对象的情况,目前相关的研究工作还比较少。这种情况已经引起了国内外学者们的关注,因此目前的研究工作开始逐步转向面向场景对象的语义分割。例如,澳大利亚国立大学的Gould等提出了一种基于样例随机场模型的图像多对象分割方法^[18],北卡罗莱纳大学教堂山分校的Tighe等提出了一种基于样例检测的区域级图像解析方法^[19]。由于对象的语义分割难度非常大,该方向的研究工作还处于初始阶段,场景中对象的语义分割还有很大的提升空间。

针对上述研究现状与存在的问题,本文提出一种基于目标识别与显著性检测的图像场景多对象分割方法,旨在提供准确的具有语义类别信息的多对象区域划分,如图1所示,即在识别出对象类别的同时还划分出同一类别不同对象的区域。本文中所使用的对象的定义与文献^[18-19]类似,对象所属类别在场景中所占区域较少,但是对场景理解具有重要的

作用,并且对象通常具有一定的形状,以此与场景材质类别进行区分。



图1 本方法的目标。

2 方法

本方法的总体流程如图2所示。对输入的图像进行过分割处理,得到图像的超像素集合;另一方面,对输入的图像进行目标检测和场景语义识别,得到测试图像的对象识别和定位包围盒,根据包围盒和语义概率值计算对象的兴趣区域;然后,对测试图像进行3种稠密尺度的显著性检测,得到像素级显著图,并将其转化成超像素的显著性值;在兴趣区域内,构建条件随机场模型,将多对象分割问题转化成多类别标记问题,兴趣区域内的超像素对应模型的节点,超像素的邻接关系对应模型的边;利用图割算法,在条件随机场模型上进行优化,迭代终止时得到像素的对象标记结果,从而实现了多个对象的分割。

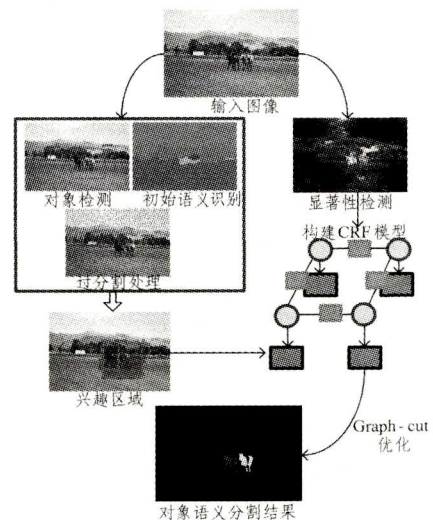


图2 方法总体流程图

对于输入图像,首先进行过分割、对象检测和初始语义估计,确定对象分割的兴趣区域;另一方面,根据图像特征计算对象显著性;结合兴趣区域和显著性构建图像的条件随机场模型,并优化求解,得到带有语义信息的对象分割结果。在初始语义识别的可视化效果中,不同灰度对应不同的语义类别;在对象语义分割结果可视化效果中,不同灰度对应不同的对象。

2.1 基于对象识别的兴趣区域检测

为实现带有语义信息的对象分割,语义类别的识别是不可缺少的,它为对象分割提供了一个初始的语义信息。利用

经典的 TextonBoost 算法^[1]在已经标注了语义信息的训练集图像上学习多类别识别分类器,将这些多个类别的分类器分别作用于测试图像,得到每一个像素点属于每一类别的概率值,即得到测试图像的像素级语义类别概率图,其可视化效果对应图 2 中的初始语义识别图。

为检测图像中对象的数量,并为后续对象分割提供数量信息,通过对对象检测得到图像中对象的数量,为后续对象分割提供数量信息,并缩小计算区域。在训练集图像上,利用 Exemplar-SVM^[20]算法训练不同类别对象的分类器。针对每一种类别的分类器,将该分类器作用于测试图像,输出多个对象包围盒,其可视化效果对应图 2 中的对象检测图,可以看到部分包围盒之间有重叠的区域。对于输出的多个对象包围盒,按照分值进行排序,选择分值最高的前 k 个包围盒作为候选集。 k 的选取方式如下:由于待分割的对象的个数由对象检测器确定,对于检测器所定位的对象包围盒,如果其分值大于设定阈值 $thr1$,则认为场景中存在一个与之对应的对象;否则舍弃该包围盒。通过阈值 $thr1$ 筛选,得到的包围盒数量即为 k 。由于每张图像中对象的数量不同,不同的图像具有不同的 k 值,为确保尽可能的准确,根据数据集的先验信息,在训练集图像上学习得到阈值 $thr1$ 的设定值。

为提高算法的计算效率,对图像进行过分割处理,利用 gPb 算法^[21]得到超像素集合,超像素数量级为 1000 左右。在超像素级别的测试图像上确定对象的兴趣区域,兴趣区域以外的区域不作为计算范围。对于对象 obj 及其类别 C ,兴趣区域的确定应符合以下的原则:

- (1) 语义类别 C 的超像素区域属于兴趣区域;
- (2) 对象 obj 的超像素区域属于兴趣区域。

根据对象检测结果,遍历 k 个包围盒,选择包围盒的超像素集合;根据初始语义概率图中 C 类别的概率分布,选择概率值大于一定阈值 $thr2$ 的超像素区域;这两种区域的合集构成了兴趣区域。阈值 $thr2$ 的取值通过学习得到:在训练集图像的初始语义概率图上学习 C 类别概率值分布,选取满足 95% 以上 C 类别正样本的概率值记为 $thr2$ 。如图 2 所示,在兴趣区域图中,符合条件的包围盒只有两个,分别将其框出,同时对象类别概率值符合条件的区域用半透明状态标识。

由于包围盒的边界与超像素的边界存在一定的间距,为防止包围盒边界将一个超像素划分为两个区域,需要对其进行以下处理:当包围盒的分值大于设定阈值 $thr1$ 时,认为该包围盒是高可靠性的包围盒,确定处于该包围盒范围内的超像素子集。如果一个超像素中超过一定比例的像素点处于该包围盒范围内,则认为该超像素处于该包围盒范围,在本文中,比例值设定为 80%。

2.2 显著性检测

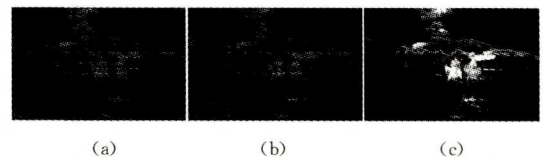
显著性反映了视觉对显著对象的关注,在一定程度上反映了对对象的区域,近年来显著性检测与对象分割的结合越来越紧密^[22-23]。本文使用一种 coarse-to-fine 的三级稠密尺度显著性检测,3 种尺度由粗至细分别记为 $scale1, scale2, scale3$ 。 $scale1$,每隔 4 个像素点取一个像素点作为滑动窗的中心点;

$scale2$,每隔 3 个像素点取一个像素点作为滑动窗的中心点; $scale3$,每隔 2 个像素点取一个像素点作为滑动窗的中心点。滑动窗口的大小为 7×7 ,在图像上进行从左至右、从上至下的滑动检测。对于每一个窗口内的块结构 $patch$,计算其 RGB 颜色空间中 3 个通道的颜色均值(见式(1)),并将其作为该块结构中心点像素的特征值 $F_{patch}(i, j)$ 。

$$\begin{aligned} R(i, j) &= \frac{1}{size(P)} \sum_{i, j \in P} R(i-3; i+3, j-3; j+3) \\ G(i, j) &= \frac{1}{size(P)} \sum_{i, j \in P} G(i-3; i+3, j-3; j+3) \\ B(i, j) &= \frac{1}{size(P)} \sum_{i, j \in P} B(i-3; i+3, j-3; j+3) \\ F_{patch}(i, j) &= \langle R(i, j), G(i, j), B(i, j) \rangle \end{aligned} \quad (1)$$

其中, $size(P)$ 为滑动窗口大小, (i, j) 表示当前 $patch$ 的中心点。

在 3 个尺度上滑动扫描,形成块结构集合 $\{P\}$,对所有块结构进行颜色值归一化处理。如图 3 所示,其中图(a)为 $scale1$ 的显著性可视化效果,图(b)为 $scale2$ 的显著性可视化效果,图(c)为 $scale3$ 的显著性可视化效果。针对当前窗口块结构 $patch(i, j)$,计算所有块结构到 $patch(i, j)$ 在颜色空间的欧氏距离值并排序,在 3 个尺度上选择距离最近的 M 个块结构,以此 M 个块结构的颜色平均值作为该块结构中心点像素的显著性值 $S(i, j)$ 。根据图像的分辨率和块结构的数量级,实验中 M 取值 60。



注:颜色越浅代表对象显著性越高,颜色越深代表对象显著性越低。

图 3 多尺度对象显著性检测示意图

对于滑动窗口未采样的点,它的显著性值是根据它的颜色值以及它周围像素点的显著性值进行线性插值得到的,计算过程如式(2)所示:

$$S(i, j) = \sum_{(x, y) \in N} \|color(i, j) - color(x, y)\| * \{1 - dis(x, y)\} \quad (2)$$

其中, N 为 (i, j) 的邻域集, $color(\cdot)$ 为像素点的颜色值, $dis(x, y)$ 为邻域点 (x, y) 到该点的距离。所有度量都进行了归一化处理。

通过滑动窗口扫描得到的是像素级显著性,根据像素点与超像素的对应关系,以超像素中所有像素的显著性均值作为该超像素的显著性值,将像素级显著性转化成超像素级显著性,为后续构建超像素级条件随机场模型提供了基础。

2.3 条件随机场模型

传统语义分割方法^[1]大多通过构建条件随机场模型(CRF)将语义分割问题转化为多类别标记问题。多对象分割也可看作一个多类别标记问题,不同于语义分割的是:每一个对象是一种标签,非对象的图像区域默认为“其他”标签。在兴趣区域的范围内,构建超像素级条件随机场模型,每一个超像素对应场模型的一个节点,超像素之间的邻接关系对应场

模型中相应的边。条件随机场模型的能量定义如式(3)所示,其中 U 为能量单一项(Unary Term), B 为能量二元项(Binary Term), U 包括 U_k 和 U_o 。

$$E(s) = U_{(k,o)}(r) + B(r,s) \quad (3)$$

根据对象检测并处理后得到的包围盒能够确定场景中对象的个数,即场模型的多类别标记数。超像素的显著性值反映了它是对象的可能性,但没有反映出它属于哪一个对象。因此,能量单一项的计算方式为:计算每个超像素属于每个对象(包括“其他”对象)的概率值,如果超像素的显著性值大于设定阈值 $thr3$,则该超像素属于每个对象的概率值由3部分构成,即该超像素的显著性值、该超像素是否属于对象的包围盒以及该包围盒的分值,否则该超像素属于每个对象的概率值为零;如果超像素的显著性值小于设定阈值 $thr3$,则该超像素属于“其他”对象,其概率值由它的显著性确定,否则概率值为零,如式(4)所示。阈值 $thr3$ 同样是通过学习选取的,在训练集图像上学习对象样本区域的显著性值概率分布,选取满足90%以上对象区域的显著性值记为 $thr3$ 。

$$U_k(r) = \begin{cases} \exp(S(r) + inbx(r,k) * V(k)), & \text{if } S(r) > thr3 \\ 0, & \text{else} \end{cases}$$

$$U_o(r) = \begin{cases} \exp(-S(r)), & \text{if } S(r) < thr3 \\ 0, & \text{else} \end{cases} \quad (4)$$

其中, $U_k(r)$ 代表超像素 r 属于对象标签 k 的概率值; $U_o(r)$ 代表超像素 r 属于“其他”对象标签 o 的概率值; $S(r)$ 为超像素的显著性; $inbx(r,k)$ 是指示函数,指示超像素 r 是否处于对象 k 的包围盒范围; $V(k)$ 为对象 k 的包围盒分值。

CRF场模型二元项的计算方式如式(5)所示。 $B(r,s)$ 由该边所连接的两个超像素之间的特征差异决定; $Nei(r,s)$ 指示邻接关系,取值为 $\{1,0\}$;超像素特征向量 $Fea(r)$ 的组成包括颜色滤波特征^[1]、边界特征^[21]、HOG梯度特征以及形状实验,这些特征均为像素级;需要将其转化为超像素级;特征差异为两个超像素在特征空间的欧氏距离; λ 和 c 是调节参数。

$$B(r,s) = Nei(r,s) * Smooth(r,s)$$

$$Smooth(r,s) = \exp(-\lambda * \| Fea(r) - Fea(s) \| + \log(\| Fea(r) - Fea(s) \| + 1) + c) \quad (5)$$

利用图割(Graph-cut)算法^[24-25]将CRF模型能量最小化,待标记的对象标签包括对象检测包围盒确定的对象以及一个“其他”标签,“其他”标签用于将兴趣区域中不属于对象的超像素标记“剔除”。当图割优化终止时,每一个超像素都被标记为一个对象标签。

3 实验

3.1 数据集

为验证本方法的有效性,在公共数据集 polo^[9,18]上进行测试。该数据集有317张图像,其中包括80张训练图像和237张测试图像;包含6种语义类别,分别是天空、草地、人、马、地面、树木。其中对象类别是人和马,每一张图像中至少包含一个以上的人或者马的对象实例。该数据集提供了原始图像和对应的语义标记图,如图4(a)和图4(b)所示,不同灰

度表示了不同的语义类别,颜色最深表示不确定类别的区域,记为“void”。由于本方法处理的是对象的语义标记,该数据集没有提供相应的对象标记,因此需要对数据集中的图像进行标注处理,手工标记同一类别的不同对象,如图4(c)所示,不同灰度标记不同的对象,非对象区域为背景。同时,为训练对象识别器,还需要对图像中的对象进行包围盒的标注,如图4(d)所示,矩形包围盒框标记了对象的范围。



图4 训练集图像标注信息

3.2 语义识别

以马这种类别为例,在训练数据集上训练对象识别的分类器来检测对象。根据 Exemplar-SVM^[20],从80张训练图像中选择10张样例图像作为 training 样本,这10张图像包含多个角度、多种颜色的对象实例。70张训练集图像作为 validation,在237张测试图像上进行具有语义识别的对象检测。在对象识别和兴趣区域检测实验中,阈值 $thr1$ 设置为0.12, k 的取值不超过7, $thr2$ 设置为0.4。语义识别的准确率如表1所列,与经典方法^[1]相比,本方法具有一定的优势。

表1 类别识别的语义标记准确率/%

方法	总体准确率	平均准确率(类别)
本方法	84.1	82.9
Textonboost ^[1]	83.9	77.1

3.3 对象分割

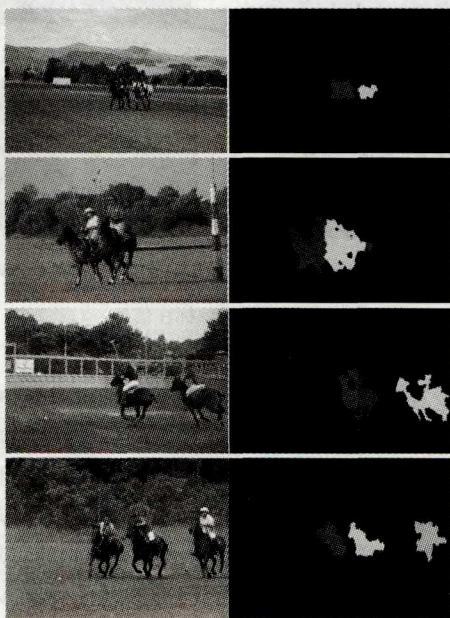
对象标记为每一个像素赋予唯一的对象标记,而语义标记为每一个像素赋予唯一的类别标记,对象标记准确率的计算方式与语义标记不同。将错误的语义标记赋予一个像素时,会导致整个场景的理解出现偏差;但是将错误的对象标记赋予一个对象的像素时,不会改变它是一个对象的事实,对象标记的目的在于将不同的对象划分开。

根据对象标记的特点,采用如下方式对对象标记准确率进行计算:1)在所有基准对象上的像素级平均准确率(Mi-AR);2)在所有识别像素上的总体像素级准确率(Ma-AP);3)在所有基准像素上的总体像素级准确率(Ma-AR)。为计算准确率,基于收益最大化的原则,将对象标记结果按照数据集中的对象标记顺序进行序号转化,以便与 groundtruth 进行匹配,找到对应的匹配对。表2所列为本方法与文献^[18]中方法在准确率方面的对比,可以看到,虽然 Ma-AP 指标比采用文献^[18]中方法所得的低10个百分点,但是在 Mi-AR 和 Ma-AR 这两个指标上本方法优于文献^[18]中的方法。

表2 性能对比与分析/%

方法	Mi-AR	Ma-AP	Ma-AR
文献[18]	53.7	57.4	68.8
本方法	63.6	47.5	77.9

在实验中,条件随机场模型的 λ 和 c 分别设置为10和-1.3, $thr3$ 设置为0.3。图5所示为本方法的部分实验结果,左图为输入测试图像,右图为图割算法迭代1000次后的对象分割结果。图5中,以不同的灰度标识不同的对象,以此类推。可以看到,在多个同类别对象出现并存在遮挡的情况下,或者对于不同分辨尺度的对象,本方法都能较好地进行识别和分割。



注:以“马”这种类别为例,不同的灰度表示不同的“马”对象,其他语义类别作为背景,没有显示。

图5 实验结果图

本文实验所用的计算机性能配置: Intel i5 3.2GHz CPU; 4GB内存; 64位 Windows7 操作系统; 软件开发平台的主程序为 Matlab2010b, 调用 Textonboost^[1] 算法的程序 Visual Studio2008。Polo 数据集图像分辨率约为 500×350 像素, 本方法在语义识别和对象分割过程中平均一张图像的处理时间约为 24s。

结束语 本文提出了一种基于目标识别与显著性检测的图像场景多对象分割方法, 该方法将语义识别和对象分割集成一体, 同时实现了语义标记和同类别多对象的分割。未来工作中, 将会加大对象形状特征的约束作用, 提高语义标记和对象分割的准确率。

参考文献

[1] SHOTTON J, WINN J M, ROTHER C, et al. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context[J]. International Journal of Computer Vision, 2009, 81(1): 2-23.

[2] XIAO J, QUAN L. Multiple view semantic segmentation for street view images [C]//The 12th International Conference on Computer Vision. 2009: 686-693.

[3] YAO J, FIDLER S, URTASUN R. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2012: 702-709.

[4] REN X, BO L, FOX D. RGB-(D) scene labeling: Features and algorithms [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2759-2766.

[5] LIU C, YUEN J, TORRALBA A. Nonparametric scene parsing: Label transfer via dense scene alignment [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 1972-1979.

[6] ZHANG H, XIAO J, QUAN L. Supervised label transfer for semantic segmentation of street scenes [C]// Proceedings of European conference on Computer Vision. 2010: 561-574.

[7] TIGHE J, LAZEBNIK S. Superparsing: Scalable nonparametric image parsing with superpixels [C]// Proceedings of European Conference on Computer Vision. 2010: 352-365.

[8] CHEN X, LI Q, SONG Y, et al. Supervised geodesic propagation for semantic label transfer [C]// Proceedings of European conference on Computer Vision. 2012: 553-565.

[9] ZHANG H, FANG T, CHEN X, et al. Partial similarity based nonparametric scene parsing in certain environment [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2011: 2241-2248.

[10] PRICE B L, MORSE B S, COHEN S. Geodesic graph cut for interactive image segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2010: 3161-3168.

[11] WU J, ZHAO Y, ZHU J, et al. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2014: 256-263.

[12] ROTHER C, MINKA T P, BLAKE A, et al. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrf [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2006: 993-1000.

[13] VICENTE S, KOLMOGOROV V, ROTHER C. Cosegmentation revisited: Models and optimization [C]// 11th European Conference on Computer Vision. 2010: 465-479.

[14] VICENTE S, ROTHER C, KOLMOGOROV V. Object cosegmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2011: 2217-2224.

[15] BATRA D, KOWDLE A, PARIKH D, et al. Interactively cosegmenting topically related images with intelligent scribble guidance [J]. International Journal of Computer Vision, 2011, 93(3): 273-292.

[16] BAI X, WANG J, SAPIRO G. Dynamic Color Flow: A Motion-Adaptive Color Model for Object Segmentation in Video [C]// Proceedings of European Conference on Computer Vision. 2010: 617-630.

[17] BAI X, WANG J, SAPIRO G. Towards temporally-coherent video matting [C]// International Conference on Computer Vision/ Graphics Collaboration Techniques and Applications. Mirage 2011: 63-74.

- [18] HE X, GOULD S. An exemplar-based CRF for multi-instance object segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2014; 296-303.
- [19] TIGHE J, LAZEBNIK S. Finding things; Image parsing with regions and per-exemplar detectors [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2013; 3001-3008.
- [20] MALISIEWICZ T, GUPTA A, EFROS A A. Ensemble of exemplar-SVMs for object detection and beyond [C]// IEEE International Conference on Computer Vision. 2011; 89-96.
- [21] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour Detection and Hierarchical Image Segmentation [J] IEEE TPAMI, 2011, 33(5); 898-916.
- [22] GOFERMAN S, ZELNIK-MANOR L, TAL A. Context-Aware Saliency Detection [J]. IEEE Trans. Pattern Anal. Mach. Intell., 2012, 34(10); 1915-1926.
- [23] CHENG M, MITRA N J, HUANG X, et al. Global Contrast Based Salient Region Detection [J]. IEEE Trans. Pattern Anal. Mach. Intell., 2015, 37(3); 569-582.
- [24] BOYKOV Y, VEKSLER O, ZABIH R. Efficient Approximate Energy Minimization via Graph Cuts [J]. IEEE TPAMI, 2001, 20(12); 1222-1239.
- [25] BOYKOV Y, KOLMOGOROV V. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision [J]. IEEE TPAMI, 2004, 26(9); 1124-1137.

(上接第298页)

在得到初步检测结果之后,利用基于区域窗口的鬼影消除了提高了算法的准确率。实验表明,与传统 PBAS 算法相比,改进算法具有更好的检测效果,并具有良好的鲁棒性,适用于检测前景的实时监控系统。

下一步工作将围绕通过减少参数和设计区域背景复杂度以获得更好的检测结果展开。

参 考 文 献

- [1] SOBRAL A, VACAVANT A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos [J]. Computer Vision & Image Understanding, 2014, 122(5); 4-21.
- [2] YE F F, XU L. Real-time detection and discrimination of static objects and ghosts [J]. Journal of Zhejiang University (Engineering Science), 2015, 49(1); 181-185. (in Chinese)
叶芳芳, 许力. 实时的静止目标与鬼影检测及判别方法 [J]. 浙江大学学报(工学版), 2015, 49(1); 181-185.
- [3] WU D P, CHENG W P, YU S L. Camshift Object Tracking Algorithm Based on inter-frame Difference and Motion Prediction [J]. Opto-Electronic Engineering, 2010, 37(1); 55-60. (in Chinese)
邬大鹏, 程卫平, 于盛林. 基于帧间差分和运动估计的 Camshift 目标跟踪算法 [J]. 光电工程, 2010, 37(1); 55-60.
- [4] BARNICH O, VAN DROOGENBROECK M. ViBe: a universal background subtraction algorithm for video sequences [J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2011, 20(6); 1709-1724.
- [5] CRISTANI M, MURINO V. Background Subtraction with Adaptive Spatio-Temporal Neighborhood Analysis [C]// The 3rd International Conference on Computer Vision Theory and Applications. 2008; 484-489.
- [6] WEI G J, HOU Z Q, LI W, et al. Object tracking algorithm fused with optical flow detection and template matching [J]. Application Research of Computer, 2014, 31(11); 3498-3501 (in Chinese)
魏国剑, 侯志强, 李武, 等. 融合光流检测与模板匹配的目标跟踪算法 [J]. 计算机应用研究, 2014, 31(11); 3498-3501
- [7] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection [C]// IEEE Conference on Computer Vision & Pattern Recognition. 2005; 886-893.
- [8] LIU X, LIU H, QIANG Z P, et al. Adaptive Background Modeling Based on Mixture Gaussian Model [J]. Journal of Image and Graphics, 2008, 13(4); 729-734. (in Chinese)
刘鑫, 刘辉, 强振平, 等. 混合高斯模型和帧间差分相融合的自适应背景模型 [J]. 中国图象图形学报, 2008, 13(4); 729-734.
- [9] FAN W C, LI X Y, WEI K, et al. Moving Target Detection Based On Improved Gaussian Mixture Model [J]. Computer Science, 2015, 42(5); 286-288, 319. (in Chinese)
范文超, 李晓宇, 魏凯, 等. 基于改进的高斯混合模型的运动目标检测 [J]. 计算机科学, 2015, 42(5); 286-288, 319.
- [10] ILYAS A, SCUTURICI M, MIGUET S. Real Time Foreground-Background Segmentation Using a Modified Codebook Model [C]// Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE Computer Society, 2009; 454-459.
- [11] HOFMANN M, TIEFENBACHER P, RIGOLL G. Background segmentation with feedback; The Pixel-Based Adaptive Segmenter [C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition Workshops. IEEE, 2012; 38-43.
- [12] WANG H, SUTER D. Background Subtraction Based on a Robust Consensus Method [C]// International Conference on Pattern Recognition. 2006; 223-226.
- [13] 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.
- [14] LI P, WU X H, HE X H, et al. An Improved PBAS For Moving Object Detection [J]. Science Technology and Engineering, 2015, 15(27); 48-53. (in Chinese)
李棚, 吴晓红, 何小海, 等. 一种改进的 PBAS 运动目标检测算法 [J]. 科学技术与工程, 2015, 15(27); 48-53.
- [15] KRYJAK T, KOMORKIEWICZ M, GORGON M. Real-time Foreground Object Detection Combining the PBAS Background Modelling Algorithm and Feedback from Scene Analysis Module [J]. International Journal of Electronics & Telecommunications, 2014, 60(1); 53-64.
- [16] WANG H, SUTER D. Background Subtraction Based on a Robust Consensus Method [C]// International Conference on Pattern Recognition. IEEE Computer Society, 2006; 223-226.