



(12)发明专利申请

(10)申请公布号 CN 109389618 A

(43)申请公布日 2019.02.26

(21)申请号 201711402509.9

(22)申请日 2017.12.22

(30)优先权数据

17184931.8 2017.08.04 EP

(71)申请人 列日大学

地址 比利时列日

(72)发明人 M.范德鲁根布罗伊克 M.布拉哈姆

S.皮伊拉德

(74)专利代理机构 中国专利代理(香港)有限公

司 72001

代理人 周学斌 郑冀之

(51)Int.Cl.

G06T 7/215(2017.01)

G06T 7/277(2017.01)

G06T 7/254(2017.01)

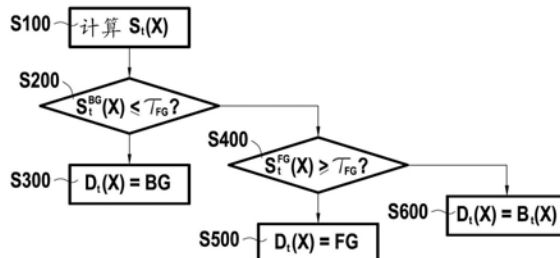
权利要求书1页 说明书10页 附图6页

(54)发明名称

前景和背景检测方法

(57)摘要

本发明涉及一种用于将像素指派给前景像素集和背景像素集之一的方法。在该方法中,如果满足第一条条件,则像素被指派给背景像素集,并且如果未满足第一条条件并且满足第二条条件,则像素被指派给前景像素集。该方法包括按照语义分割算法计算该像素属于前景相关对象的概率的步骤(S100),该第一条条件是,该像素属于前景相关对象的该概率不超过第一阈值,并且该第二条条件是,该像素属于前景相关对象的该概率与针对该像素的基线概率之间的差等于或者超过第二预定阈值。



1. 一种用于将像素指派给前景像素集和背景像素集之一的方法,包括以下步骤:  
按照语义分割算法,计算(S100)所述像素属于前景相关对象的概率;  
如果所述像素属于前景相关对象的概率不超过第一预定阈值,则将所述像素指派(S300)给背景像素集;以及  
如果所述像素属于前景相关对象的概率超过第一预定阈值并且所述像素属于前景相关对象的概率与针对所述像素的基线概率之间的差等于或者超过第二预定阈值,则将所述像素指派(S500)给前景像素集。
2. 按照权利要求1所述的方法,进一步包括基线更新步骤,其中,如果所述像素已经被指派给背景像素集,则使得针对所述像素的基线概率等于所述像素属于按照语义分割算法计算的前景相关对象的概率。
3. 按照权利要求2所述的方法,其中,如果所述像素已经被指派给背景像素集,则基线更新步骤仅按照执行的预定概率随机执行。
4. 按照任何前述权利要求所述的方法,进一步包括以下步骤:如果所述像素属于前景相关对象的概率超过第一预定阈值并且所述像素属于前景相关对象的概率与针对所述像素的基线概率之间的差低于第二预定阈值,则按照将所述像素的像素值与背景模型中的对应像素的像素值相比较的背景减除算法,将所述像素指派给前景像素集或者背景像素集。
5. 按照权利要求4所述的方法,其中,所述像素属于图像,并且背景模型是基于至少另一相关图像的。
6. 按照之前权利要求中的任一项所述的方法,其中,所述像素属于按时间前后顺序的图像序列中的图像。
7. 按照权利要求6所述的方法,其中,针对所述像素的基线概率已被初始化为等于在按时间前后顺序的图像序列中的初始图像中的按照语义分割算法所计算的对应像素属于前景相关对象的概率。
8. 按照任何前述权利要求所述的方法,其中,第一和第二预定阈值已经被预先确定,以使得对测试图像序列优化所述方法的F分数。
9. 按照权利要求1-7中的任一项所述的方法,其中,第一和第二预定阈值已经分别基于背景减除算法的假前景检测率和背景减除算法的真前景检测率被启发式地预先确定。
10. 一种数据处理设备(2),其被编程以使得实行按照权利要求1到9的任一项所述的方法。
11. 一种包括指令集的数据存储介质,所述指令集被配置成由数据处理设备(2)读取,以实行按照权利要求1到9的任一项所述的方法。

## 前景和背景检测方法

### 技术领域

[0001] 本公开内容涉及一种用于将像素指派到前景和背景像素集之一的方法。

### 背景技术

[0002] 在计算机视觉中的主要研究方向是运动检测的领域。运动检测的目标是将像素按照其是否属于这样的移动对象与否而进行分类,从而将可能被错误分类的任何像素过滤,以便检测场景中的移动对象。已证明将该任务(其在大自然中甚至显然被基本的动物视觉系统轻松解决)复制在计算机视觉中是复杂的。

[0003] 在计算机视觉的领域中,图像可以被表达为多个图片元素或者像素。在图像中的每个单个像素可以具有在图像中的位置 $x$ 和像素值 $\vec{I}(x)$ 。

[0004] 位置 $x$ 可具有任何数量的维度。出于该原因,虽然术语“体素”(代表“体积元素”)在3D成像的领域中用来代替“像素”,但是术语“像素”在本公开内容中应该被宽泛地理解为也涵盖这样的体素和具有任何数量的维度的图像中的任何图片元素,所述图像包括3D图像和/或多光谱图像。

[0005] 位置 $x$ 可以被限制到有限域,例如如果其是由固定成像设备捕获的图像的话。然而,可替换地,其可以不被限制到有限域,例如,如果图像是由诸如例如卫星机载相机之类的移动成像设备捕获的图像的话。

[0006] 像素值 $\vec{I}(x)$ 也可以具有任何数量的维度。例如,在单色图像中,像素值 $\vec{I}(x)$ 可以是标量亮度值,但是在多色图像(诸如红绿蓝(RGB)分量视频图像或者色度饱和度值(HSV)图像)中,该像素值 $\vec{I}(x)$ 可以是多维向量值。

[0007] 在过去的二十年中,针对运动检测已提出了大量背景减除算法。这些背景减除算法中的许多已经由P.-M.Jodoin、S. Pierard、Y.Wang和M.Van Droogendroeck在2014年7月的Chapman and Hall/CRC的Background Modeling and Foreground Detection for Video Surveillance的第24章的“Overview and benchmarking of motion detection methods”中和由T.Bouwmans在2014年5月的Computer Science Review的第11-12卷的第31-66页的“Traditional and recent approaches in background modeling for foreground detection: An overview”中进行了回顾。

[0008] 大多数背景减除算法涉及在每个图像中的低水平特征(诸如单个像素值)与背景模型的比较,所述背景模型可以被简化为不具有移动对象和可能地适应性的图像。可假设具有相对于背景模型的显著差异的像素属于移动对象,并且因此可被指派给前景像素集,而剩余像素可以被指派给背景像素集。例如,由C.Stauffer和E.Gimson在1999年6月的IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR)的第2卷的第246-252页的“Adaptive background mixture models for real-time tracking”中和由O.Barnich和M.Van Droogenbroeck在2011年6月的IEEE Trans. Image Process.的第20卷第6号的第1709-1724页的“ViBe: A universal background subtraction algorithm for video

sequences”中公开的背景减除算法按照颜色分量对像素分类,而由V.Jain、B.Kimia、和J.Mundy在2007年9月的IEEE Int. Conf. Image Process. (ICIP)的第6卷的第321-324页的“Background modeling based on subpixel edges”中、S.Zhang、H.Yao、和S.Liu在2008年10月的IEEE Int. Conf. Image Process. (ICIP)的第1556-1559页的“Dynamic background modeling and subtraction using spatio-temporal local binary patterns”中、M.Chen、Q.Yang、Q.Li、G.Wang和M.-H.Yang在2014年9月的Eur. Conf. Comput. Vision (ECCV)的Springer的Lecture Notes Comp. Sci.的第8695卷的第521-534页的“Spatiotemporal background subtraction using minimum spanning tree and optical flow”中、以及M.Braham、A.Lejeune、和M.Van Droogenbroeck在2014年12月的IEEE Int. Conf. 3D Imaging (IC3D)的第1-8页的“A physically motivated pixel-based model for background subtraction in 3D images”中公开的背景减除算法分别使用了边缘、纹理描述符、光学流或者深度来将像素指派给前景或者背景。用于背景建模的特征的综合性回顾和分类由T.Bouwman、C.Silva、C.Marghes、M.Zitouni、H.Bhaskar和C.Frelicot在2016年11月的CoRR的abs/1611.09099卷的第1-131页的“On the role and the importance of features for background modelling and foreground detection”中给出。

[0009] 虽然这些低水平特征中的大多数可以在非常低的计算负荷的情况下来计算,但是它们不能同时解决在真实世界的视频序列中出现的许多挑战,诸如照明改变、伪装、相机抖动、动态背景、阴影等等。仅仅基于RGB颜色分量的对于基于像素的方法的性能的上限由S.Pierard和M.Van Droogenbroeck在2015年9月的Int. Conf. Image Anal. and Process. (ICIAP), Workshop Scene Background Modeling and Initialization (SBMI)的Springer的Lecture Notes Comp. Sci.的第9281卷的第527-534页的“A perfect estimation of a background image does not lead to a perfect background subtraction: analysis of the upper bound on the performance”进行了模拟。特别地,其示出了背景减除算法未能在存在噪声和阴影的情况下提供完美的分割,即便在完美的背景图像可用时。

[0010] 在对于背景减除算法的典型挑战中,我们可以特别考虑伪装的前景对象、“重影”、动态背景和阴影和/或反射效果。

[0011] 前景对象在其相对应的像素值(例如,颜色或者亮度)相似于背景的那些像素值时被认为是“伪装”的。在此情景下,背景减除算法可能会作为漏报错误地将相对应的前景像素指派给背景。这可能例如采取在来自彩色相机的图像上的颜色伪装的形式,或者在来自热相机的图像上的热伪装的形式。例如,雪的覆盖可能导致这样的伪装。

[0012] “重影”是当之前静态的、因此属于背景的对象开始移动时的现象。在该情景中,因为不仅是对应于对象的像素的像素值改变,而且在对象静态时属于之前被对象隐藏的背景的那些像素也改变,所以这些后者的背景像素可能作为误报被错误地指派给前景。

[0013] 动态背景是这样的背景:曾经可能存在像素值改变,诸如例如风吹过的多叶树或者海浪。在该情景中,对应的背景像素可以同样作为误报被错误地指派给前景。

[0014] 相似地,阴影和反射可导致背景像素由于像素值的相关联的改变而作为误报被错误地指派给前景。

[0015] 可能导致背景像素作为误报被错误地指派给前景的其他挑战是噪声图像(例如,由于压缩的人为缺陷)、相机抖动、自动相机调整、慢帧率、平移、倾斜和/或变焦、坏天气、逐步或者突然照明改变、背景对象的运动/插入、在热图像上的残留高温印记(heat stamp)、持续背景改变、由于反射导致的云、烟雾和高光。

[0016] 可导致前景像素被错误指派到背景的其他挑战是快速移动的对象,以及变得不动的前景对象,并且可能因此被错误地合并到背景中。

## 发明内容

[0017] 本公开内容的第一方面涉及一种用于将像素指派到前景像素集和背景像素集之一的方法,该方法比用将像素的像素值与背景模型中的对应像素的像素值相比较的背景减除算法更可靠和鲁棒。特别地,按照该第一方面,本公开内容寻求解决对于背景减除算法的上述挑战。出于此目的,按照该第一方面的方法可以包括以下步骤:按照语义分割算法计算所选图像属于前景相关对象的像素的概率,并且如果该像素属于前景相关对象的概率不超过第一预定阈值,则将该像素指派给背景像素集,如果该像素属于前景相关对象的概率超过第一预定阈值并且该像素属于前景相关对象的概率与针对该像素的基线概率之间的差等于或者超过第二预定阈值,则将该像素指派给前景像素集。

[0018] 人类可以容易地以高精度描绘相关的移动对象,因为他们结合了来自语义水平的知识:他们知道车是什么、认出阴影、在对象运动和相机运动之间进行区分等等。语义分割(也被称为场景标记或者场景解析)的目的在于,通过将图像的每个像素以其包围的对象或者区域的类别进行标记而提供这样的信息。该任务是困难的,并且要求对语义对象和区域的同时检测、定位和分割。然而,在计算机视觉界内出现的深度神经网络和对大型已标记的训练数据集的访问已显著地改进了语义分割算法的性能,如由J.Long、E.Shelhamer、和T.Darrell在2015年6月的IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR)的第3431-3440页的“Fully convolutional networks for semantic segmentation”中、由S.Zheng、S.Jayaysmana、B.Romera-Paredes、V.Vineet、Z.Su、D.Du、C.Huang和P.Torr在2015年12月的IEEE Int. Conf. Comput. Vision (ICCV)的第1529-1537页的“Conditional random fields as recurrent neural networks”中、以及由H.Zhao、J.Shi、X.Qi、X.Wang和J.Jia在2016年12月的CoRR的abs/1612.01105卷的“Pyramid scene parsing network”中描述的。语义分割算法因此已开始用于特定计算机视觉任务,诸如如由L.Sevilla-Lara、D.Sun、V.Jampani、和M.J.Black在2016年6月的IEEE Int Conf. Comput. Vision and Pattern Recogn. (CVPR)的第3889-3898页的“Optical flow with semantic segmentation and localized layers”描述的光学流估计。

[0019] 通过利用来自较高的语义水平的信息,按照本公开内容的该第一方面的方法可以提供一种比由仅将低水平像素值与背景模型相比较的背景减除算法所提供的更鲁棒、可靠的对前景和背景的图像分割。

[0020] 一方面,语义水平可以因此被用来标识前景相关对象,即,属于可预期移动并且因此属于前景的语义类别的对象,并且如果像素属于前景相关对象的概率不超过第一预定阈值则将该知识用于将像素指派给背景像素集的步骤中,以便防止误报,即,由于例如动态背景、重影、阴影和/或反射、相机抖动、平移、倾斜和/或变焦、坏天气、逐步或者突然照明改变

或者背景位移(这通常影响常规背景减除算法的性能)造成的将像素错误地指派给前景像素集。

[0021] 另一方面,语义水平还可以用于在如果像素属于前景相关对象的概率超过第一预定阈值并且在像素属于前景相关对象的概率与针对像素的基线概率之间的差等于或者超过第二预定阈值则将所选图像的像素指派给前景像素集的步骤中,标识像素属于这样的前景对象的概率是否相对于针对该像素(其可能例如对应于在语义背景模型中的对应像素)的基线概率而增大,以便防止漏报,即,由于伪装(即,当背景和前景共享相似像素值时)造成的将像素错误地指派给背景。

[0022] 按照本公开内容的第二方面,以上提及的方法可以进一步包括基线更新步骤,其中,如果像素已经被指派给背景像素集,使得针对该像素的基线概率等于该像素属于按照语义分割算法计算的前景相关对象的概率。因此,使用来自该图像的语义水平的信息,针对该像素的基线概率可以被更新,以供随后相对于其他图像中的对应像素使用。然而,为了避免例如由于间歇性和慢移动对象造成损坏该基线概率,可以应用保守更新策略,其中,如果像素已经被指派给背景像素集,按照执行的预定概率,基线更新步骤仅随机执行。

[0023] 按照本公开内容的第三方面,该方法可以进一步包括以下步骤:如果像素属于前景相关对象的概率超过第一预定阈值并且在像素属于前景相关对象的概率与针对像素的基线概率之间的差低于第二预定阈值,则按照将该像素的像素值与背景模型中的对应像素的像素值相比较的背景减除算法,以及特别是按照基于至少一个低水平图像特征的背景减除算法,将该像素指派给前景像素集或者背景像素集。因此,以上提及步骤基于语义分割算法未能指派给前景像素集或者背景像素集的任何像素可以使用将该像素的像素值与背景模型中的对应像素的像素值相比较的已知背景减除算法来进行指派。特别地,像素可以属于图像,并且背景模型是基于至少另一个相关图像的,诸如例如在包括像素所属的图像的按时间前后顺序的图像序列中的之前的图像。

[0024] 实际上,按照本公开内容的第四方面,像素可以属于按时间前后顺序的图像序列中的图像,在此情况下,针对该像素的基线概率可以被初始化为等于多个相关图像中的初始图像中的按照语义分割算法所计算的对应像素属于前景相关对象的概率。因此,来自该初始图像的语义知识可以被用于至少初始地设置基线概率,在属于前景相关对象的随后图像中的对应像素的概率在确定那些像素是否要被指派给前景时与所述基线概率相比较。

[0025] 按照本发明的第五方面,第一和第二预定阈值可以已经被预先确定,以使得对测试图像序列优化该方法的F分数。检测方法的F分数可以被定义为在准确度和召回率(recall)之间的调和平均数,其中,准确度是正确肯定与正确肯定和误报的总和的比率,以及召回率是正确肯定与正确肯定和漏报的总和的比率。在本上下文中,被正确指派给前景的像素可以被认为是正确肯定,被不正确指派给前景的像素表示误报,并且被不正确指派给背景的像素表示漏报。因此,预先确定第一和第二预定阈值以使得对测试图像序列优化以上提及的方法的F分数可以确保当该方法随后对所选图像实行时在准确度 $p$ 和召回率 $r$ 之间的良好折衷。

[0026] 然而,在本发明的可替换第六方面,第一和第二预定阈值可以已经分别基于背景减除算法的假前景检测率和背景减除算法的真前景检测率被启发式地预先确定。实际上发明人已经发现,可以利用其在测试图像序列上优化该方法的F分数的第一和第二预定阈值

分别与在该方法中应用的背景减除算法的假前景检测率和真前景检测率紧密关联。因此，如果那些比率从背景减除算法的较早测试中是已知的，则当对所选图像实行该方法时确保在准确度和召回率之间的良好折衷变为可能的，即便在没有实行应用背景减除算法和语义分割算法两者对该方法的F分数进行消耗时间和消耗资源的优化的情况下。

[0027] 本发明还涉及一种数据处理设备，其被编程以使得实行本发明的图像背景识别方法；还涉及一种包括指令集的数据存储介质，所述指令集被配置为由数据处理设备读取，以实行按照本发明的图像背景识别方法；还涉及以磁、电磁、电和/或机械形式的信号集合，其包括用于使得数据处理设备实行按照本发明的图像背景识别方法的指令集；和/或还涉及经由磁、电磁、电和/或机械装置传输用于使得数据处理设备实行按照本发明的图像背景识别方法的指令集的过程。

[0028] “数据存储介质”可以被理解为能够包含可由读取设备在至少特定时间段内读取的数据的任何物理介质。这样的数据存储介质的示例是磁带和磁盘、光盘（只读的以及可记录的或者可重写的）、逻辑电路存储器（诸如只读存储器芯片、随机存取存储器芯片和闪存存储器芯片）、以及更奇异的数据存储介质，诸如化学、生化或者机械存储器。

[0029] “电磁”被理解为具有相干（LASER、MASER）或者不相干形式的、从无线电到UV以及以后（包括微波、红外和可见光）的电磁谱的任何部分。

[0030] “对象”被理解为是真实世界的任何可观察元素，包括动物和/或人类。

[0031] 本发明的一些方面的以上概述不旨在描述本发明的每个所公开的实施例或者每个实现方式。特别地，在本说明书内的任何说明性实施例的所选特征可以被合并到附加实施例中，除非清楚说明是相反情况。

## 附图说明

[0032] 本发明可以连同附图考虑以下各种实施例的详细描述而被更完全地理解，在附图中：

-图1是图示了按照本公开内容的方面的方法的核心例程的流程图；

-图2是图示了语义分割算法和背景减除算法的结果如何被组合在图1的核心例程中的功能方案；

-图3A和3B是绘制了在背景减除算法的误报率FPR和正确肯定率TPR与分别针对图2的方法中的第一预定阈值  $T_{BG}$  和第二预定阈值  $T_{FG}$  的最优值之间的正相关的图；

-图4到6是绘制了通过图1的方法通过其中使用的背景减除算法实现的改进的图；以及

-图7图示了语义分割算法、背景减除算法和将两者组合用于困难场景中的各种视频序列的方法的输出；以及

-图8和9图示了应用于图1的方法的视频系统的潜在实施例。

[0033] 虽然本发明服从与各种修改和可替换形式，但是其详情已经作为示例在附图中示出，并且将详细描述。然而应该理解，本发明并不将本发明的方面限制于所描述的特定实施例。相反，本发明要覆盖落入在本发明范围内的所有修改、等同物和可替换物。

## 具体实施方式

[0034] 对于以下定义的术语，应该应用这些定义，除非在权利要求中或者在本说明书中

的其他地方给出了不同的定义。

[0035] 假定本文的所有数值之前都加上术语“大约”，无论是否明确指示。术语“大约”一般指代本领域技术人员将认为等同于所述值的数字范围（即，具有相同功能或者结果）。在许多实例中，术语“大约”可以指示为包括四舍五入到最近有数字的数。

[0036] 如在本说明书和所附权利要求中所使用的，单数形式“一”、“一个”和“该”包括复数的所指对象，除非内容清楚表示是另外情况。如在本说明书和所附权利要求中所使用的，术语“或者”一般在其包括“和/或”的意义上使用，除非清楚表示是另外情况。

[0037] 应该参考附图阅读以下详细描述，在附图中，在不同附图中的相似元件被相同地编号。详细描述和不一定按照比例的附图描绘了说明性实施例，并且不旨在限制本发明的范围。所描绘的说明性实施例仅旨在是示范性的。任何说明性实施例的所选特征都可以被合并到附加实施例中，除非清楚说明是相反情况。

[0038] 在可能已经例如由成像设备在时间序列之后的时间 $t$ 连续捕获的图像集合中，每个图像可以由多个像素形成，在图像中的每个单个像素具有专用像素位置 $x$ 和像素值 $\vec{I}(x)$ 。为了易于理解，在附图中，像素位置 $x$ 被示为二维的，但是其可具有任何数量的维度。对于3D图像，例如，像素位置 $x$ 可以具有三个维度。在所图示的实施例中的像素值 $\vec{I}(x)$ 是三维向量，具有RGB或者HSV三元组的形式，以用于获取多色图像。然而，在可替换实施例中，其可以具有任何其他数量的维度。

[0039] 语义分割算法可以应用于每个图像，以便针对每个像素位置 $x$ 和时间 $t$ 计算实值向量 $\vec{v}_t(x) = [v_t^1(x), v_t^2(x), \dots, v_t^N(x)]$ ，其中 $v_t^i(x)$ 表示对于具有 $N$ 个不相交对象类别的集合 $C = [c_1, c_2, \dots, c_N]$ 的每个类别 $c_i$ 的分数。在像素位置 $x$ 和时间 $t$ 处的像素属于集合 $C$ 的每个类别 $c_i$ 的概率集合 $p_t(x \in c_i)$ 可以通过向分数 $v_t^i(x)$ 应用softmax函数来计算。

[0040] 在集合 $C$ 的 $N$ 个不相交对象类别中，子集 $R$ 可以对应于前景相关对象，即，涉及运动检测的对象。例如，如果图像涉及街道场景，则这些前景相关的对象可以包括如车辆、人类和动物的潜在移动对象，但是不包括如树或者建筑物的典型不移动的对象。通过使用语义分割算法，因此有可能计算在像素位置 $x$ 和时间 $t$ 处的像素属于前景相关对象的合计概率（aggregated probability） $p_{S,t}(x) = p_t(x \in R) = \sum_{c_i \in R} p_t(x \in c_i)$ 。可能会有可能考虑不同的子集 $R$ ，可能针对图像的不同区域具有不同数量的不相交的前景相关对象类别。例如，当图像示出道路和人行道两者时，子集 $R$ 可以在人行道区域中仅包括人类和动物作为前景相关对象类别，但是在道路区域中还包括车辆。

[0041] 按照语义分割算法在像素位置 $x$ 和时间 $t$ 处的像素属于前景相关对象的该概率 $p_{S,t}(x)$ 可以用于一种用于在图像集合的每个图像中将像素指派到前景和背景像素集的方法中。图1示出了图示了该方法的核心例程的流程图，其中，在像素位置 $x$ 和时间 $t$ 处的像素被指派给前景像素集或者背景像素集。在第一步骤S100中，使用语义分割算法来计算在像素位置 $x$ 和时间 $t$ 处的像素属于前景相关对象的概率 $p_{S,t}(x)$ 。在第二步骤S200中，确定该概率 $p_{S,t}(x)$ 是否低于或者等于第一预定阈值 $\tau_{BG}$ 。如果该比较的结果是肯定的，并且因此确



定按照语义分割算法在像素位置 $x$ 和时间 $t$ 处的像素属于前景相关对象的概率  $p_{S,t}(x)$  不超过第一预定阈值  $\tau_{BG}$ , 则认为在像素位置 $x$ 和时间 $t$ 处的像素属于潜在移动对象是不太可能的, 并且因此在步骤S300中, 在像素位置 $x$ 和时间 $t$ 处的像素被指派给背景。使用二进制变量  $D \in \{BG, FG\}$ , 其中值BG指示背景像素, 并且值FG指示前景像素, 这可以被表达为第一规则:

$$S_t^{BG}(x) \leq \tau_{BG} \rightarrow D_t(x) = BG$$

其中,  $S_t^{BG}(x)$  表示等于概率  $p_{S,t}(x)$  的信号, 并且  $D_t(x)$  表示针对在像素位置 $x$ 和时间 $t$ 处的像素的二进制变量 $D$ 的值。该第一规则提供了一种用以解决通过将背景像素错误指派给前景像素严重影响常规背景减除算法性能的照明改变、动态背景、重影和强阴影的挑战的简单方法。

[0042] 另一方面, 如果在步骤S200中, 确定该概率  $p_{S,t}(x)$  不低于或者等于第一预定阈值  $\tau_{BG}$ , 则在下一步骤S400中, 确定差  $S_t^{FG}(x) = p_{S,t}(x) - M_t(x)$  是否至少等于第二预定阈值  $\tau_{FG}$ , 其中,  $M_t(x)$  表示针对像素位置 $x$ 和时间 $t$ 的基线概率。该基线概率  $M_t(x)$  对应于针对像素位置 $x$ 和时间 $t$ 的背景的语义模型。其可以已被初始化为等于在像素位置 $x$ 和时间0处的对应像素的概率  $p_{S,0}(x)$ , 即, 在相关图像的集合的初始图像中, 按照语义分割算法属于前景相关对象。其然后可以已经按照以下更新策略在每个随后的时间步长更新:

$$\begin{aligned} D_t(x) = FG &\rightarrow M_{t+1}(x) = M_t(x) \\ D_t(x) = BG &\rightarrow {}_{\alpha}M_{t+1}(x) = p_{S,t}(x) \\ &\rightarrow {}_{1-\alpha}M_{t+1}(x) = M_t(x) \end{aligned}$$

其中,  $\rightarrow_{\alpha}$  表示执行的预定概率 $\alpha$ , 其可以例如被设置为0.00024。因此, 针对像素位置 $x$ 和下一时间步长 $t+1$ 的基线概率  $M_{t+1}(x)$  的值被保持等于针对在时间步长 $t$ 的对应像素的基线概率  $M_t(x)$ , 并且如果  $D_t(x) = BG$  (即, 在像素位置 $x$ 和时间 $t$ 处的像素已经被指派给背景像素集) 则仅按照预定执行概率 $p$ 随机更新, 具有概率  $p_{S,t}(x)$ 。

[0043] 以预定执行概率 $\alpha$ 对针对像素位置 $x$ 的基线概率  $M_t(x)$  是否要被更新的随机确定可以使用随机数生成器实行。然而, 因为这样的随机数不能由确定性计算机提供, 所以可以代替地使用伪随机数生成器, 其具有与真实随机数生成器的那些属性相似的属性。另一可替换项是使用之前生成的随机或者伪随机数的大型查找表。

[0044] 如果在步骤S400的比较结果是肯定的, 并且因此确定差  $S_t^{FG}(x)$  确实等于或者大于该第二预定阈值  $\tau_{FG}$ , 则其认为相对于按照语义模型所预期的, 已存在针对在像素位置 $x$ 和时间 $t$ 处的像素的概率  $p_{S,t}(x)$  的显著增长, 并且在下一步骤S500中, 在像素位置 $x$ 和时间 $t$ 处的像素因此被指派给前景像素集。这可以被表达为第二规则:

$$S_t^{FG}(x) \geq \tau_{FG} \rightarrow D_t(x) = FG$$

该第二规则可以在很大程度上防止伪装的(即, 具有与背景相似的像素值的)前景像素

被错误指派给背景,这也是对于常规背景减除算法的挑战。

[0045] 然而,仅语义分割可能不足以在前景和背景之间区分,例如在前景相关对象(例如,移动的车)在相同语义的背景相关对象(例如,停靠的车)的静态对象前面移动的情况下。因为两个对象都属于相同的前景相关对象类别,所以概率  $p_{s,t}(x)$  将不会随移动对象在像素位置  $x$  和时间  $t$  处的静态对象之前移动而显著增大。

[0046] 为了解决这样的情景,如果在步骤S400的比较结果是否定的,即,如果概率  $S_t^{BG}(x)$  超过第一预定阈值  $\tau_{BG}$  并且差  $S_t^{FG}(x)$  低于第二预定阈值  $\tau_{FG}$ ,则将第三规则应用到下一步骤S600中,按照常规背景减除算法将在像素位置  $x$  和时间  $t$  处的像素指派给前景像素集或者背景像素集,所述常规背景减除算法基于多个相关图像中的至少另一个图像,将在像素位置  $x$  和时间  $t$  处的像素的像素值  $I(x)$  与背景模型中的对应像素的像素值相比较。这可以被表达为第三规则:

$$D_t(x) = B_t(x)$$

其中,  $B_t(x) \in \{BG, FG\}$  表示来自背景减除算法的结果。

[0047] 图2因此图示了使用以上提及的三个规则,三个信号  $S_t^{BG}(x)$ 、 $S_t^{FG}(x)$  和  $B_t(x)$  可以被如何获取和组合应用于前景和背景检测。这些信号如何被组合也可以使用以下表来总结:

$B_t(x)$	$S_t^{BG}(x) \leq \tau_{BG}$	$S_t^{FG}(x) \geq \tau_{FG}$	$D_t(x)$
BG	假	假	BG
BG	假	真	FG
BG	真	假	BG
BG	真	真	X
FG	假	假	FG
FG	假	真	FG
FG	真	假	BG
FG	真	真	X

表1:按照方法的三个规则的前景和背景检测。

[0048] 如果第一预定阈值  $\tau_{BG}$  被设置为低于第二预定阈值  $\tau_{FG}$ ,则不能有效地遇到在以上表1上以“X”表示的两个情景。

[0049] 因为第一规则仅将像素指派给背景像素集,从而增大了第一预定阈值  $\tau_{BG}$  使得第一规则被更频繁地应用仅可能减小正确肯定率TPR(其是像素被正确指派给前景的比率)和误报率FPR(其是像素被错误地指派给前景像素集的比率)。另一方面,因为第二规则集仅将像素指派给前景像素集,减小了第二预定阈值  $\tau_{FG}$  使得第二规则被更频繁地应用仅可能增大正确肯定率TPR和误报率FPR。理想地,第一预定阈值  $\tau_{BG}$  和第二预定阈值  $\tau_{FG}$  因此被设置为实现在最高可能正确肯定率TPR和最低可能误报率FPR之间的最佳折衷的水平。

[0050] 用于设置第一预定阈值  $\tau_{BG}$  和第二预定阈值  $\tau_{FG}$  的一种第一可替换项是使用以

上提及的方法以针对第一预定阈值  $\tau_{BG}$  和第二预定阈值  $\tau_{FG}$  的各种不同值对测试图像序列执行测试,并且针对给定的背景减除和语义分割算法,针对导致最佳总体F分数的这些阈值选择值的二元组,即,在准确度和召回率之间的最高调和平均数,其中,准确度是正确肯定(被正确指派给前景像素集的像素实例)与正确肯定和误报(被错误指派给前景像素集的像素实例)的总和的比率,以及召回率是正确肯定与正确肯定和漏报(被错误指派给背景像素集的像素实例)的总和的比率。这可以被执行为网格搜索优化。

[0051] 本发明人已经对53个视频序列实行了这样的测试,所述视频序列被组织在11个类别中,具有由Y.Wang、P.-M.Jodoin、F.Porikli、J.Konrad、Y.Benezeth、和P.Ishwar在2014年6月于美国俄亥俄州的Columbus的IEEE Int. Conf. Comput. Vision and Pattern Recogn. Workshops (CVPRW) 的第393-400页的“CDnet 2014: An expanded change detection benchmark dataset”所呈现的CDNet数据集,应用了使用34种不同背景减除算法的本公开方法,并且作为语义分割算法,应用了由H.Zhao、J.Shi、X.Qi、X.Wang和J.Jia在CoRR的abs/1612.01105卷的“Pyramid scene parsing network”所公开的深度架构PSPNet,其在由B.Zhou、H.Zhao、X.Puig、S.Fidler、A.Barriuso和A.Torralla在2016年8月的CoRR的abs/1608.05442卷的“Semantic understanding of scenes through the ADE20K dataset”呈现的ADE20K数据集上使用由H.Zhao使公众可用的PSPNet50 ADE20K模型,以NVIDIA® GeForce® GTX Titan X GPU上的473 x473像素图像分辨率以大致7fps运行来训练,以提取语义。该PSPNet架构的最后一层针对N=150个不相交对象类别的集合  $C = \{c_1, c_2, \dots, c_N\}$  的每个类别  $c_i$  向每个像素指派分数。在这些测试中,前景相关对象类别的所选子集是  $R = \{\text{人, 车, 垫子, 箱子, 书, 船, 公交车, 卡车, 瓶子, 货车, 包, 自行车}\}$ , 其对应于CDNet前景对象的语义。

[0052] 在这些测试期间,已经发现,在以上提及的方法的第三规则中使用的背景减除算法的误报率FPR和正确肯定率TPR分别与在应用于相同方法的第一和第二规则中时实现最佳总体F分数的第一预定阈值  $\tau_{BG}$  和第二预定阈值  $\tau_{FG}$  之间存在紧密关联,如图3A和3B所示。

[0053] 因此,第二可替换方法是分别基于要在方法的第三规则中使用的背景减除算法的误报率FPR和正确肯定率TPR来启发式地设置第一预定阈值  $\tau_{BG}$  和第二预定阈值  $\tau_{FG}$ 。例如,第一预定阈值  $\tau_{BG}$  可以被设置为背景减除算法的误报率FPR的一半,而第二预定阈值  $\tau_{FG}$  可以被设置为等于背景减除算法的正确肯定率TPR。因为按照定义,背景减除算法应该比随机分类器执行得更好,所以其误报率FPR应该低于其正确肯定率TPR,因此确保第一预定阈值  $\tau_{BG}$  也低于第二预定阈值  $\tau_{FG}$ 。

[0054] 按照第三可替换方法,第一预定阈值  $\tau_{BG}$  和第二预定阈值  $\tau_{FG}$  可以被设置为默认值,其例如对应于产生针对2014 CDNet排名中最佳性能的五项背景减除算法的每一项的最佳总体F分数的这些阈值的值的算术平均值,并使用相同的语义分割算法。

[0055] 这三个可替换方法中的每一个已经被测试,并且被发现是就其本身而言提供了对于根本的背景减除算法的结果的非常显著的改进。图4图示了该改进,其针对这三个方法的

每一项,被定义为一减去将背景减除与语义分割相组合的方法的错误率除以背景减除算法自身的错误率。更具体地,图4图示了针对34个背景减除算法的整个集合和针对仅5个最佳执行的背景减除算法两者的对总体CDNet数据集测量的平均值改进。如在该图上可见的,第一种方法提供了非常显著的改进,甚至优于已经最佳执行的背景减除算法,并且该改进关于第二和第三种可替换方法几乎没有降低。

[0056] 图5图示了就平均正确肯定率TPR和误报率FPR的改变而言相对于以上提及的34个不同背景减除算法集合的每一个背景减除算法的改进。如此处可见,本方法趋向于显著减小误报率FPR,而同时增大正确肯定率TPR。

[0057] 图6图示了按照视频序列的类别、CDNet数据集的11个类别中的针对所有34个不同背景减除算法和针对5个最佳执行两者的平均改进,所述CDNet数据集的11个类别为:“基线”、“动态背景”、“相机抖动”、“间歇性对象运动”、“阴影”、“热”、“坏天气”、“低帧率”、“夜间视频”、“平移-倾斜-变焦相机”和“湍流”。对于“基线”、“动态背景”、“阴影”和“坏天气”类别,可以观察到特别好的改进。相对于“热”和“夜间视频”类别而言,必须指出,用于教导语义分割算法的ADE20K数据集不包括这些类型的图像,这可以解释对于那些类别的较不好的结果。

[0058] 图7图示了按照本公开内容的在真实世界监视任务的四个不同场景中的方法的益处。从左到右,四列分别对应于具有动态背景、重影风险、强阴影和伪装效果的场景。从上到下,五行图示了对应的视频序列的帧、针对每个像素的概率  $p_{s,t}(x)$ 、由S.Bianco、G.Ciocca、和R.Schettini在2015年的CoRR的abs/1505.02921卷的“*How far can you get by combing change detection algorithms?*”中描述的IUTIS-5背景减除算法的输出、本公开方法的输出(在其第三规则中应用IUTIS-5背景减除算法)和地面实况。如可见,相对于IUTIS-5背景减除算法自身,本公开方法大大减少了由动态背景、重影和强阴影导致的误报前景像素检测的数量,而同时缓解了伪装效果。

[0059] 本公开方法可以在连接到提供连续图像的视频序列的成像设备的数据处理设备的协助下实行,所述数据处理设备诸如例如可编程计算机,如以上提及的NVIDIA® GeForce® GTX Titan X GPU。在这样的情况下,数据处理设备可以使用数据存储介质或者作为以磁、电磁、电和/或机械形式的信号接收用于实行该方法的指令。

[0060] 本公开方法可以例如应用于视频监视、专业和/或消费者数字静态和/或视频相机、使用图像捕获接口的计算机和视频游戏设备、卫星成像和地球观测、自动图像分析和/或医学成像系统或者可以被包括在智能电话中。

[0061] 图8图示了具有成像设备1的本发明的可能应用,成像设备1以数字相机的特定形式,该数字相机具有被编程以实行本发明方法的嵌入式数据处理设备2。图9图示了具有成像设备1的本发明的另一可能应用,成像设备1连接到被编程以实行本发明方法的数据处理设备2。

[0062] 本领域技术人员将认识到,本发明可以以除了本文所描述和设想的具体实施例之外的各种各样的形式来表示。因此,可以在不偏离在如所附权利要求中描述的本发明的范围的情况下做出形式和细节上的偏离。

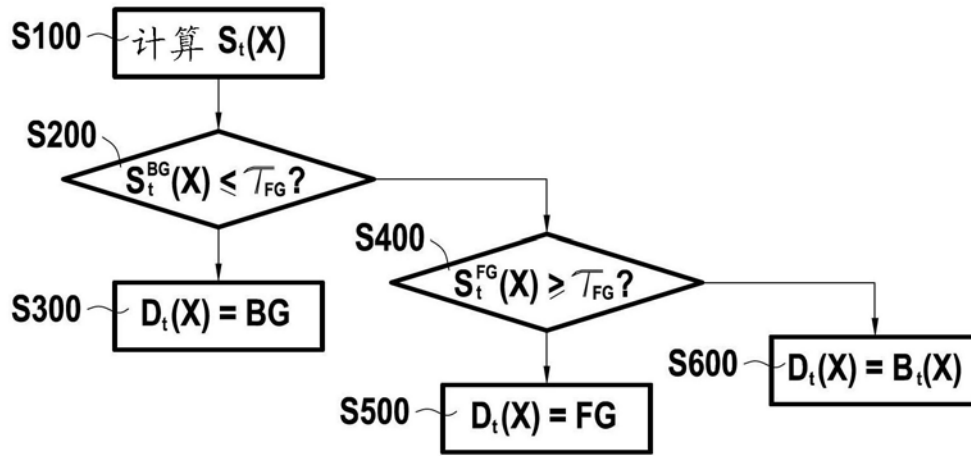


图 1

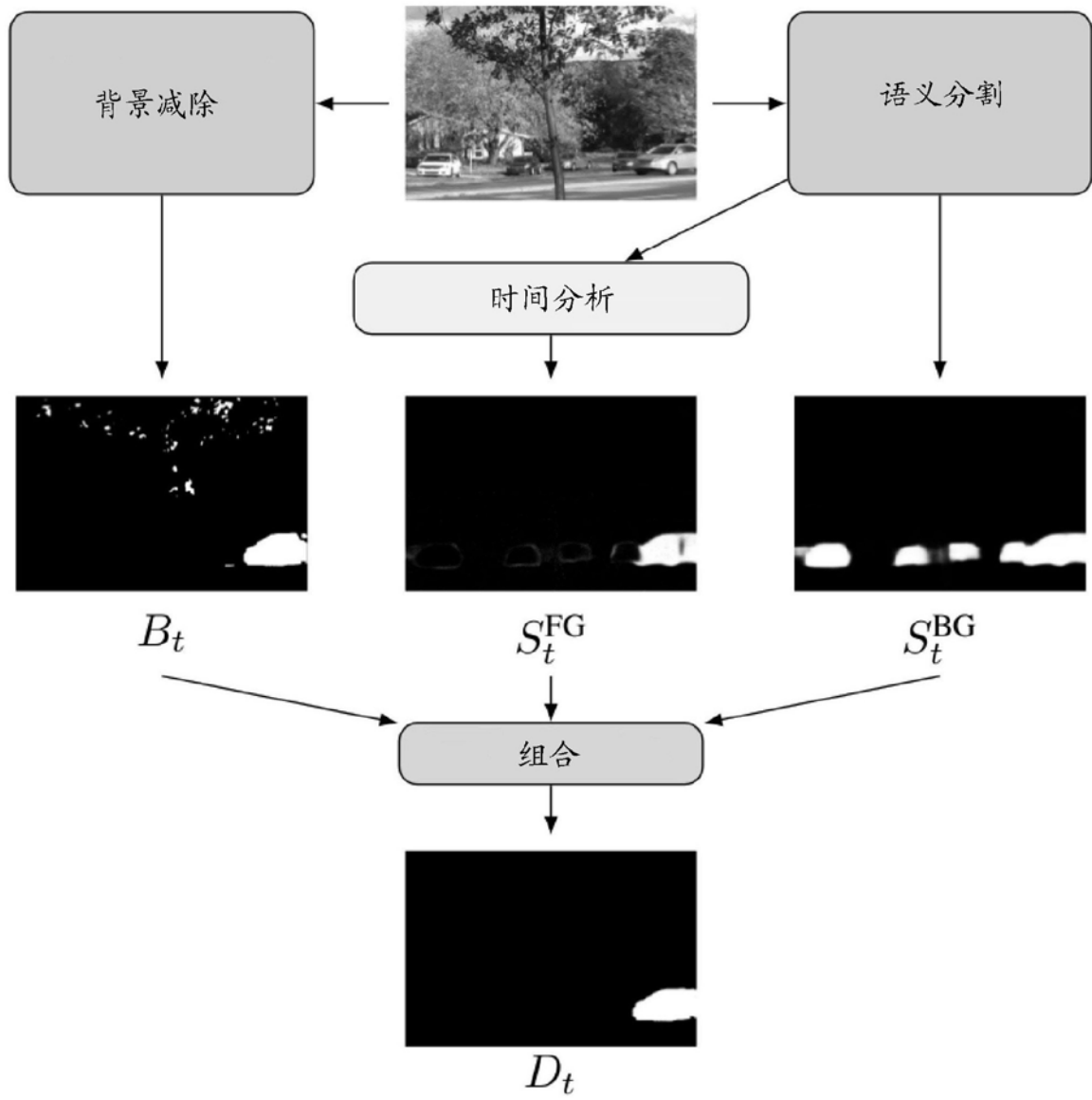


图 2

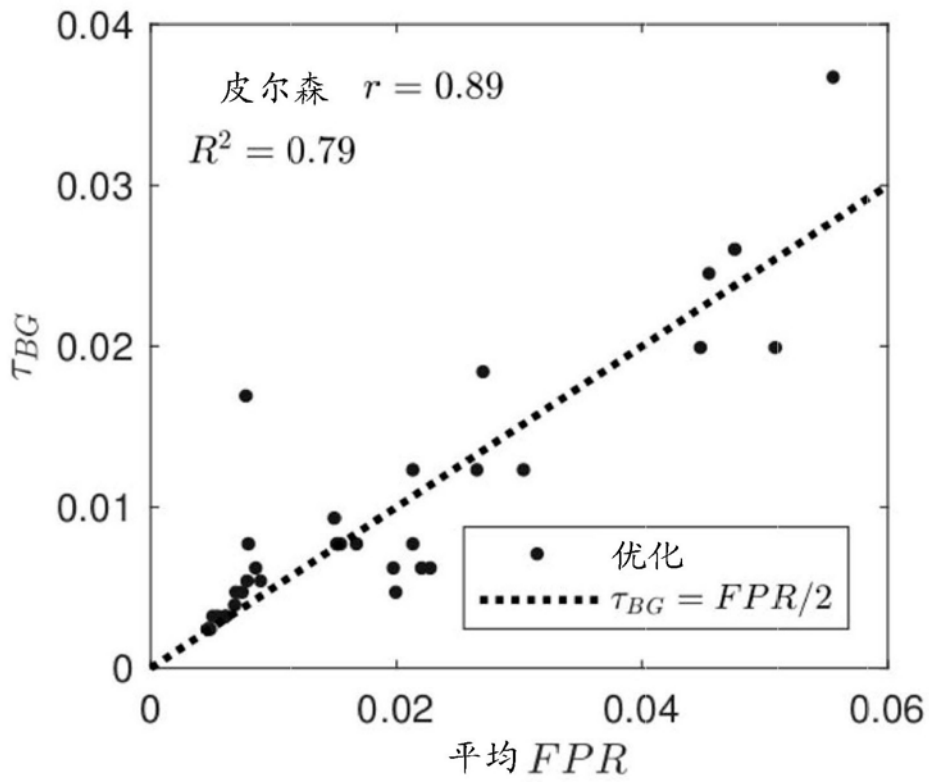


图 3A

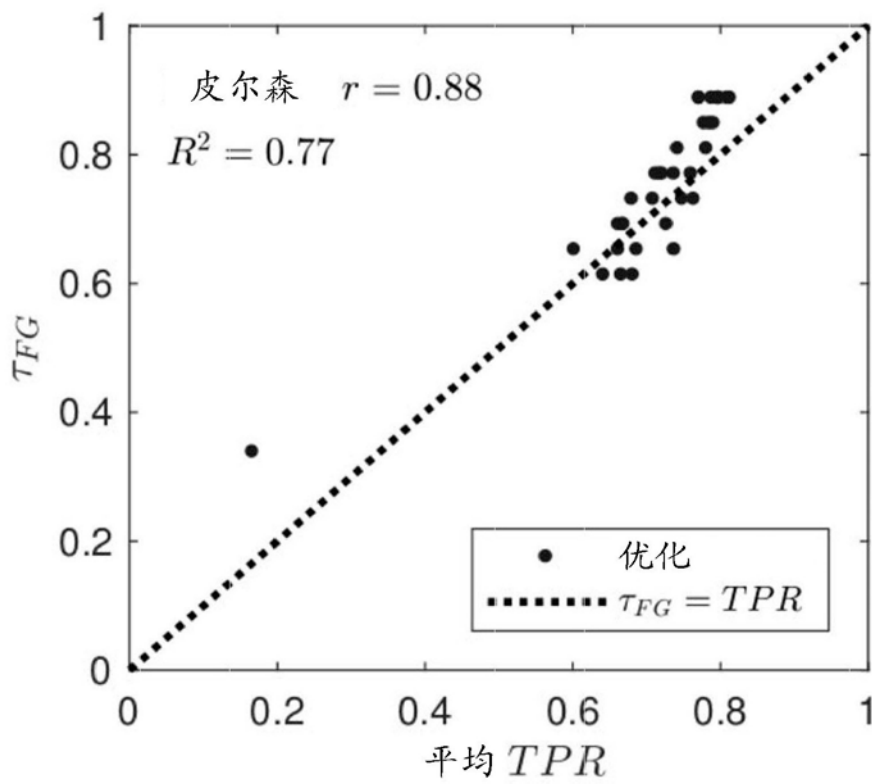


图 3B

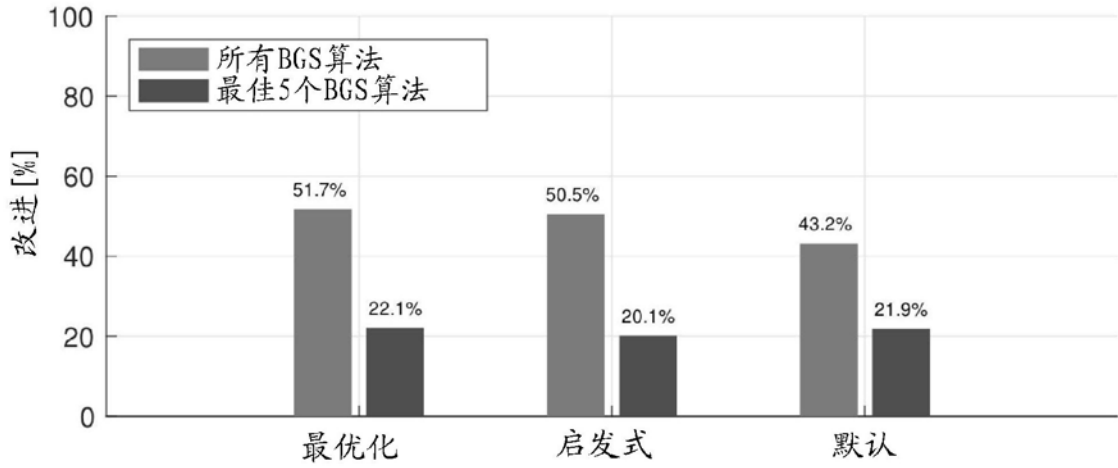


图 4

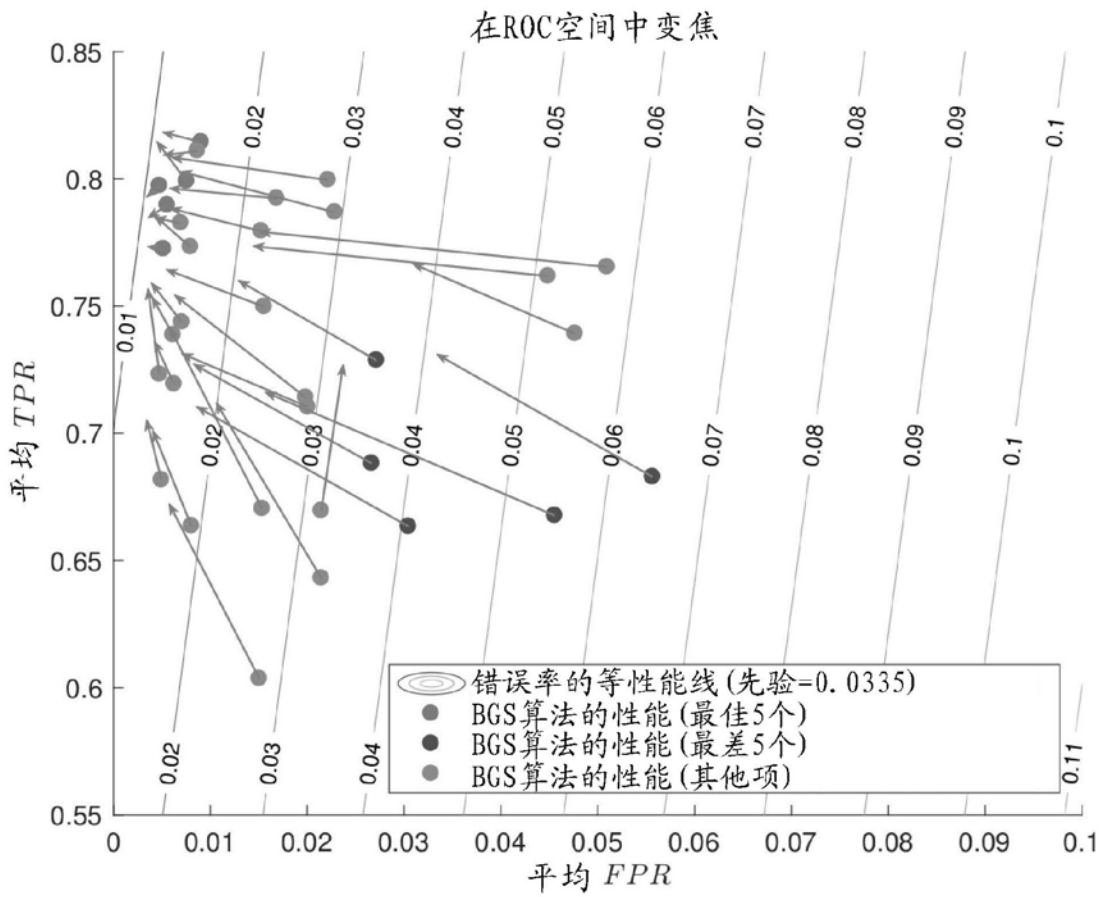


图 5



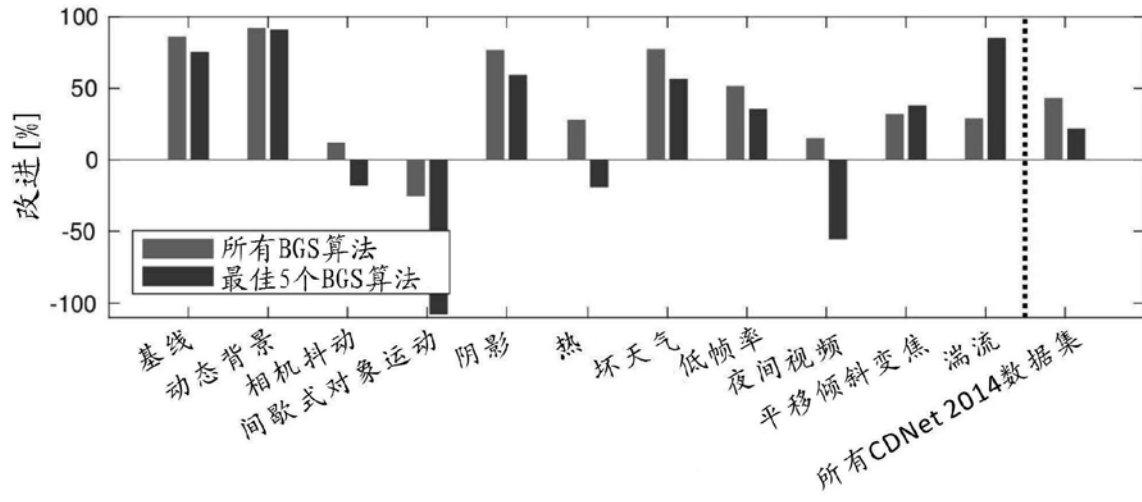


图 6

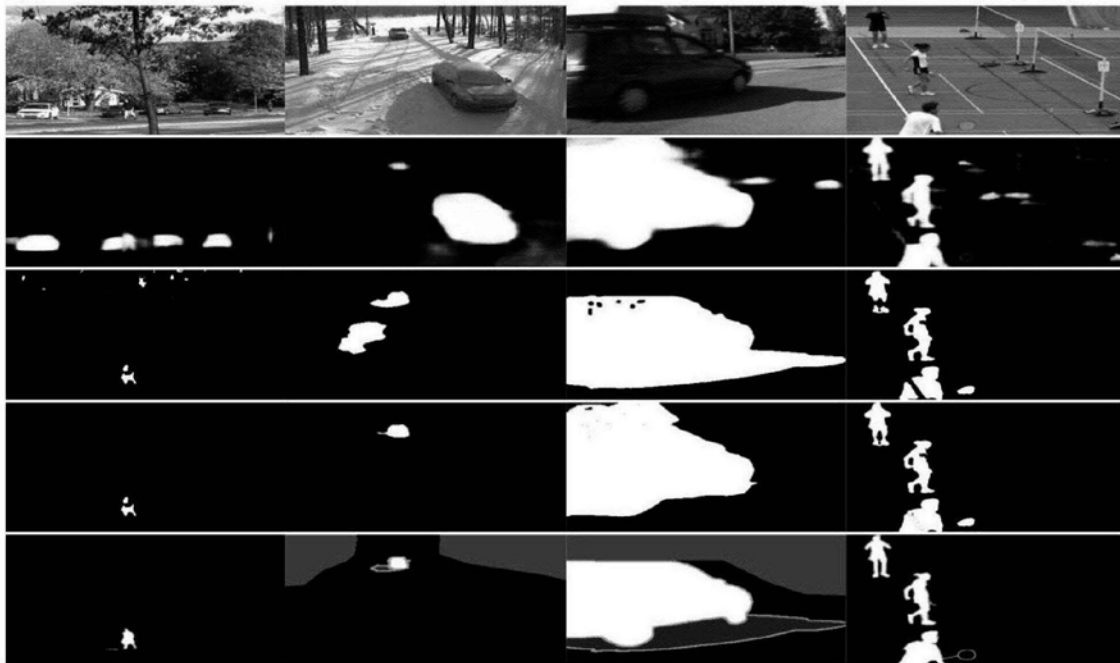


图 7

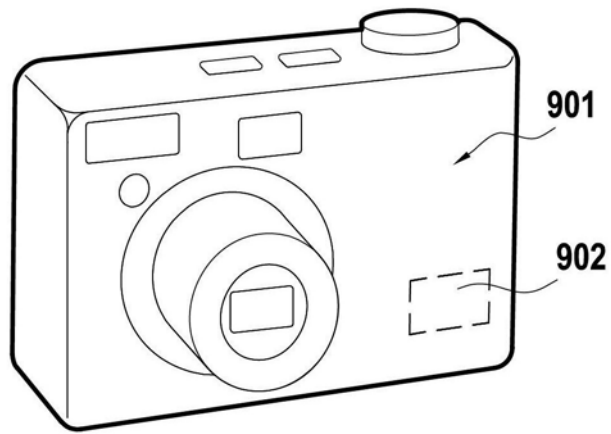


图 8

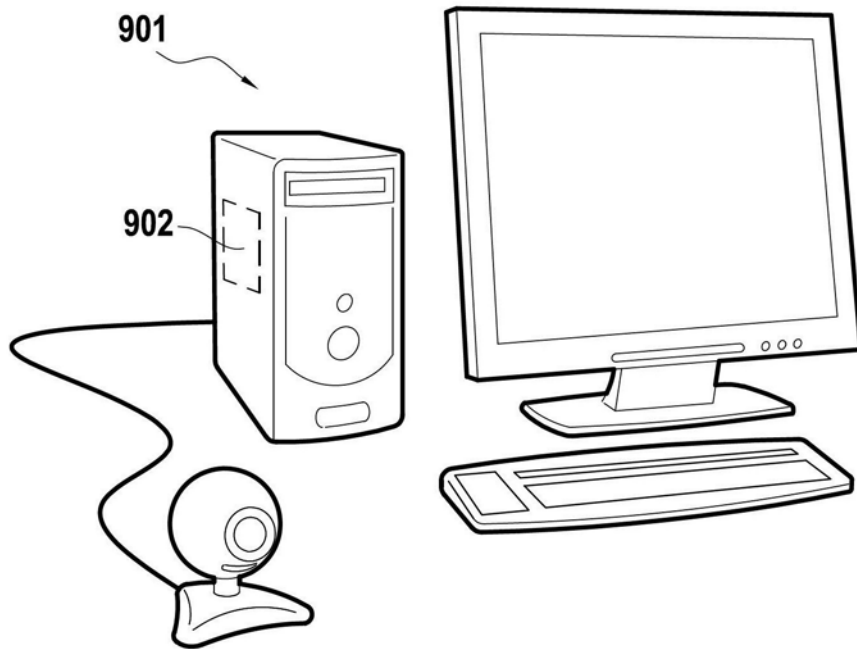


图 9