 硕 士 研 究 生 读 书 报 告



题目：对于跳跃视频的剪辑：

非连续性的模型转换和视频抠图

JumpCut: Non-Successive Mask Transfer and Interpolation for Video Cutout

作者姓名 裴灵晖 （linghui pei）

作者学号 nb15093

指导教师 李启雷 (qilei li)

学科专业 移动互联网与游戏开发 (Mobile Internet and game development )

所在学院 软件学院

提交日期 二○15 年 12 月

摘要

介绍了跳跃视频剪辑，一个新的模型转换和交互式视频抠图的插值方法。给定一个前景掩模的源框架已经可用，我们估计的前景掩模在另一个，通常是非连续的，目标帧。观察背景和前景区域通常表现出不同的运动，我们利用这些差异计算两个近邻领域（split-NNF）从目标到源的框架。然后用这些NNF共同预测相干标记在目标帧的像素。同样的split-NNF，也可以用来帮助一个新的边缘分类器检测轮廓边缘（s-edges），分离从背景前景。然后是一种改进的水平集方法的应用产生一个干净的模型，基于像素的标签和s-edges由前两个步骤计算。由此产生的掩模转移的方法也可用于相干插值的前景模型是吐温两远源帧。我们的研究结果表明，所提出的方法是显着更准确的比现有的国家的最先进的在各种视频序列。因此，它可以减少用户的力量，和一个有效的交互式视频对象分割工具提供了依据。

**关键词**：**视频分割,前景提取、对象断路**

**Abstract**

We introduce JumpCut, a new mask transfer and interpolation method for interactive video cutout. Given a source frame for which a foreground mask is already available, we compute an estimate of the foreground mask at another, typically non-successive, target frame. Observing that the background and foreground regions typ- ically exhibit different motions, we leverage these differences by computing two separate nearest-neighbor fields (split-NNF) from the target to the source frame. These NNFs are then used to jointly predict a coherent labeling of the pixels in the target frame. The same split-NNF is also used to aid a novel edge classifier in detect- ing silhouette edges (S-edges) that separate the foreground from the background. A modified level set method is then applied to produce a clean mask, based on the pixel labels and the S-edges computed by the previous two steps. The resulting mask transfer method may also be used for coherently interpolating the foreground masks be- tween two distant source frames. Our results demonstrate that the proposed method is significantly more accurate than the existing state-of-the-art on a wide variety of video sequences. Thus, it re- duces the required amount of user effort, and provides a basis for an effective interactive video object cutout tool.

**Keywords: video segmentation, foreground extraction, object cutout**

1. 介绍

视频前景提取是经常在电视和电影生产用于合成视觉元素从各种不同的来源到一个新的视频背景。挑战参与准确地从背景中分离复杂的动态对象的自然视频,以及需要创造性的控制,使自动方法不切实际的任务。因此,在过去的十年中,大量研究集中在有效的交互式工具的发展。然而,尽管令人印象深刻的进步互动前景提取、任务仍需要相当大的用户的努力和专业知识。

具体来说,最先进的方法仍然受到颜色和/或纹理模棱两可之间的前景和背景区域,由低对比度边缘分离前景从背景元素,快速运动前景,通常涉及非刚性的变形和翻天覆地的变化在前台区域的拓扑。

在这项工作中,我们描述了一种新的工具,互动前景提取特别大,非刚性的前景运动目标。我们解决这一问题的交互式视频器,任务是提取二进制掩码标签每一帧的像素为前景或背景。柔边α无光,如果需要,用我们的面具trimap为基础

基本计算元素,我们关注的是在这工作是面具转移:给定一个正确分割源框架,我们的目标是计算前景面具,非连续的目标框架(图1)。在实践中,面具传输距离通常是4 - 32帧,根据前景运动和变形的速度出现在视频。正如我们在我们的结果和同伴的视频,演示的能力来执行这样的面具准确传输提供了一种有效的交互式视频抠图工具的基础。

目前最先进的互动方法,依赖本地和全球的组合分类器。在前台边缘的矩形窗口使用本地分类器,遥远的传播是具有挑战性的,而钟山等人使用长方向的窗户,适当的大小和方向是很难确定的。个快速的方法计算最近邻域(NNF)之间的两个图片,是有效处理大位移光学。PatchMatch发现通讯补丁之间的差异的基础上,同时考虑颜色和质地。计算NNFs分段空间相干,但可能包含遥远通信由于其使用的随机搜索。然而,PatchMatch方法有时会陷入局部最小值,特别是在颜色和纹理的存在歧义。

在这个工作我们的基本见解,位移的背景(BG)和前景(FG)地区通常表现出不同的和不相关的行为。具体来说,BG运动通常是由相机运动的场景,而FG运动是独立的相机,而且通常更大的位移和非刚性的变形特性。因此,我们的方法试图明确解释这种差异,通过跟踪BG和成品区域使用两个单独的NNFs(split-NNF)。这些单独的NNFs更好的本地化,因此局部最小值的影响较小。

我们的方法的另一个关键组件是一个新的分类的过程,利用我们的split-NNF边缘。具体来说,每个凸边不分段目标框架分为三个类别之一:背景区域内的边缘(B-edges),在前台(F-edges),最重要的是,轮廓边缘分开这两个地区(S-edges)。分类是通过非参数,监督分类器,它使用面具的边缘和源帧作为训练数据。

获得最终的,干净,面具为目标框架我们应用基于水平集方法。这最后一步是推波助澜的结果前两个阶段:初始化水平集方法与预测的pixelwise面具split-NNF,并依赖于边缘分类器的结果提前到附近的轮廓边缘,而目标掩模平滑轮廓。

上述整个过程的结果在一个新的面具转移方法,成功地传播面具更远和更准确地比现有最先进的方法。配备这种改进连任的面具转移能力,我们也可以实现一个新的面具插值机制,我们给出两个不相邻分段帧重建中间帧的前景的面具,在一个连贯的和准确的方式。

2相关工作

视频前景提取是经常在电视和电影生产用于将来自不同来源的视觉元素组合成一个视频。在工作室,前景提取可能使用一个常数来完成彩色屏幕从自然中提取动态前景元素视频是一个更具挑战性的问题,吸引了重要研究关注

在他们的开创性工作,壮族等。结合双向光流和背景估计插入trimap在视频体积。贝叶斯席子然后用于计算每一帧的前景无光。描述keyframe-based系统自动全景:跟踪一个前景轮廓的曲线表示视频序列。通过使用时空优化他们传播用户约束序列中向前和向后。

制定前景提取的3 d graphcut问题时空视频体积,进一步细化结果使用2 d graphcuts内部跟踪本地windows。还操作3 d视频体积,并提供一个用户界面直接在本卷画约束。通等。描述更直观和有效的用户界面绘画的约束,基于局部3 d graphcuts。

LIVEcut系统价格等。使用多个线索与2 d graphcut优化为了传播选择向前,逐帧。白的视频SnapCut系统等的分割在每一帧通过合作获得一组重叠的局部分类器,每个集成多个地方特色。这种方法已被纳入Adobe后遗症,Rotobrush工具,后来改进的更好的整合运动和颜色模型。因为这些方法都是基于统计数据收集的一组广场本地windows集中在前景物体的边界,他们很难处理时间不连续出现由于大动作,或与不可分割的视频数据。

推进先进的进一步通过引入无偏定向分类器,旨在应对更大的前景运动,同时保持地方统计数据的优势。他们也提出多个分类器结合在一起的新方法。然而,对于遥远的转移很难确定适当的窗口大小和抽样只有四个方向(0、45、90、135度)并不总是足够的。

因此,尽管令人印象深刻的进步互动前景提取,它仍然是容易找到的视频序列,挑战现有的先进的方法:视频中前景和背景之间的颜色和纹理模糊区域,在摄像机的运动和/或动态前景对象具有重要意义。

这些先前的方法相比,我们的面具传播基于最近邻域(NNFs)框架,并使用PatchMatch(巴恩斯et al . 2009年)来计算。PatchMatch基于补丁之间的差异,从而寻找相似的颜色和纹理。PatchMatch的另一个优点,参数分类器相比,它明确鼓励空间相干NNFs。尽管如此,它仍有能力捕捉大位移和非刚性的运动由于其使用随机搜索。PatchMatch计算大位移的有效性光学流最近被证，不同于这些方法,我们的方法明确占通常不同的运动表现出的背景和前景的地区。具体来说,我们使用两个分别计算NNFs跟踪这两个区域,每一个都是更好的本地化,从而降低PatchMatch易感性的局部最小值。

在计算机视觉界已经有大量视频序列分割跟踪研究。Faktor和伊朗也使用不同的背景和前景运动初始化他们的成品/ BG分类器。他们的方法似乎代表了最先进的自动(无监督)前景提取。然而,结果和其他方法尚未强劲或自动精确到足以取代互动的抠图方法。第四节所示,我们将我们的方法和SeamSeg Ramakanth先生2014,目前顶尖高手SegTrack基准。

水平集方法已广泛应用于自动和交互式图像分割的背景。我们称读者克莱莫等。广泛回顾相关文献。下面,我们只强调几个特别相关的工作。

描述交互式图像分割系统,面向移动设备的触摸界面。他们的方法是基于水平集的框架,以外观模型附近的感动点采样。以来,这种方法依赖于本地训练参数的模型,它是有限的在实践中与非重叠图像前景和背景区域,统计,和小位移视频。

描述交互式图像分割方法,它采用水平集方法,使用边缘抑制去除不凸出的边缘。他们的水平集的优势领域术语,定义使用对凸边的距离变换。我们的方法使用一个水平集与类似的术语,然而,我们边训练分类器是有区别地使用分段源框架我们传播着面具。

3、方法

3.1 面具转移

给一个分段的来源与二进制掩模框架表明前景像素的女士,和目标框架,我们的目标是确定目标前景面具。

面具转移的一个方法是将每个像素的分类为背景(BG)或前景(FG)使用各种分类器训练(是,Ms)。另一个替代方法是计算密集的对应的每个像素匹配x 2”(x)2,和设置太(x)=(女士”(x))。在这项工作中,我们选择后者的方法,因为我们相信这是更适合传输的源和目标帧视频序列相互远离,和BG和成品区域可能强烈流离失所。大位移连续帧之间也会发生在视频序列快速相机和/或FG对象运动。

正如前面所提到的,我们使用一个方法基于PatchMatch为了计算最近邻域(NNF)作为”。我们观察到BG和成品区域的序列通常是流离失所的以不同的方式。BG位移通常是由摄像机的运动(主要是静态)的场景,而FG位移通常是更大的和更少的刚性。因此,我们的想法是利用两个不同位移模式通过计算一个单独的NNF的两个地区。这种方法如图2所示。

具体来说,我们开始调整,它对前景和背景的运动,分别。FG对齐模板匹配计算,蒙面的前景在哪里是模板,里面有平方的总和差异面具作为匹配度量。发现的最佳翻译前景区域,我们把它应用到获得FG-aligned源安全部队。调整的背景,我们执行特征之间的匹配和使用冲浪。与离群值通过RANSAC拒绝。接下来,我们经使用刚性移动最小二乘(MLS)方法约束下的匹配特征点,导致该校BG-aligned来源。

考虑到两个对齐的来源是⇤,⇤2 { f b },然后我们计算一对本地化NNFs是⇤。我们的方法是基于PatchMatch,但只局限于搜索本地为了达到更好的准确性和一致性。PatchMatch计算交错的NNF连贯的传播和随机搜索。我们的本地化PatchMatch不同于原方法在三个方面:第一,因为它⇤大致一致,没有必要随机初始化NNF,与初始NNF零点偏移在每个像素使用。第二,在每个迭代中随机搜索,没有必要样本整个形象;相反,我们随机搜索只在局部区域大小W⇥W在每个像素,其中W将图像对角线的三分之一。虽然看起来是一个大窗户,必须记住,PatchMatch不是一个完整的随机搜索:只有少数随机样本都来自那个窗口。第三,进一步鼓励地方比赛,我们使用一个补丁的加权组合差异与欧氏距离度量。对应的重量是10和3到4帧的传输,和10个1.5,更遥远的转移。

分别计算两个NNFs对齐,如上所述,结果在更少的错误NNF由于使用本地搜索和修改后的距离。在计算两个NNFs,我们现在融合到一个NNF通过选择,每个像素,NNF抵消,收益率小补丁的区别。面具太然后预测如前所述女士在本节(Mt(x)=((x)))。整个过程如图2所示,这也显示了最后的细化后的转移面具的水平集方法,如3.4节所述。

3.2面具插值

上述技术可以扩展到面具插值,或双向转移面具,我们希望预测前景掩模在一个框架,给出两个已经分割框架,例如。,它k + k。利用两个分割源帧,双向传输通常是能够更准确地预测目标的面具。这个过程的另一个优点是,鉴于双方前景轮廓,之前我们可以使用形状插值作为一个连贯的形状。我们的系统执行插值分层次使用二进制细分;也就是说,考虑到分段源帧它k + k,我们第一次插入中间帧,然后递归直到k = 1每一半。

我们的方法是应用在前一节中描述的技术两次,一次转移面具从k它一旦从+ k。因此,我们实际上利用四NNFs分开。但是,在这里,而不是依赖于模板匹配大致对齐分割前景与目标框架,我们执行形状匹配两个分段面具,并使用形状插值预测中间帧的位置和形状的面具。用于更精确地预测形状使每个两个分割前景区域与目标框架,导致更准确的前景NNFs。

形状匹配是一个深入研究的领域,我们最初考虑使用的最先进的方法和雅各布[2007],它使用inner-distances,显示与整体相似度匹配完整的形状。然而,这种方法的一个缺点是,它使用动态规划的全局最优对应,呈现闭塞和拓扑变化敏感,经常发生在我们的背景。此外,除了几何形状,我们还想利用颜色和纹理特征,这使我们能够找到可靠的通讯没有约束。因此,我们使用下面描述的方法。

在每个轮廓点代表当地的形状,和计算之间的形状差异两个轮廓点相应的补丁之间的SSD(除以面具像素块的数量)。计算每个点的颜色意味着前景颜色的15⇥15图像补丁为中心,这些颜色一栏是L2距离。注意,只有在前台使用的颜色,因为前景形状在背景的位置可能会改变显著不同的帧之间。空间距离dspace被定义为ij kpi + m pjk2,其中m之间的区别是中心(点)的平均位置的轮廓在它k + k。最后计算距离dij dij = dshape +一栏+ dspace。

两两对应计算使用上面描述的相似点可能包含许多局外人,被两个额外的步骤。对于每个轮廓点我在k,我们找到最佳匹配轮廓点Mi + k,并在另一个方向做同样的事情。信件被丢弃,除非比赛在两个方向上都同意。此外,我们坚持匹配点的顺序是单调的;换句话说,我> j意味着Mi >乔丹。比赛违反单调性也丢弃的离群值。

给结果集之间的配对它k和k +的轮廓,我们预测的位置对应点在目标框架通过线性插值。接下来,每个它k + k是扭曲对它使用刚性MLS(Schaefer et al . 2006年),导致两个扭曲的图像如果和前- t k t + k地面区域的大致预测前,地面区域的位置和形状。接下来,我们也调整的背景区域k和k +与目标,如3.1节所述,产生另一条扭曲的图片我b和b。最后,我们计算4 t k t + k单独NNFs从目标框架上面的四个扭曲的图像,然后把面具使用NNF产量的最佳匹配每个像素。

3.3边缘分类

它一直认识到边缘f地面图分离中发挥至关重要的作用在我们的视觉感知鲁宾。许多图像分割算法和边缘合并到这个过程中,推理,区域边界应配合突出的边缘。然而,前面所述的面具传输技术是基于NNFs和没有明确考虑边缘。此外,转移面具最终清理使用水平集方法,在下一节中描述。因此,将边缘到流程,防止面具将自己与无关紧要的凸边附近可能存在,我们试图把所有的凸边在目标框架。

具体来说,我们试图将凸边im -年龄分为三个类别之一:B-edges(完全包含在背景),F-edges(完全包含在前台),和S-edges(轮廓边缘分离前景从背景)。然而,值得注意的是,在我们管道的其余部分,我们目前只利用S-edges。免费的额外分类B-edges和F-edges没有任何额外的计算工作,因为每条边是机密通过检查其双方后来解释说。我们使用一个简单的非参数三方监管分类器的训练数据是由源框架及其面具。原则上边缘分类器是类似于一个像素分类器,因为它是应用于单独的边点,但它包含边缘的方向,利用双方的两种颜色。

我们首先从源和目标帧中提取一个边缘地图使用的快速先进的方法。为每个点的边缘响应超过某个阈值我们联系两个像素,一人一边的边缘。对于每一次这样的像素提取7-dimensional特征向量,包括(r,g,b,x,y,sin✓,cos✓),r,g,b是颜色,x,y是空间坐标,✓梯度的方向。每个维度是规范化的范围[0,1],然后是颜色、空间坐标,和取向是按比例缩小的10.0倍,分别为2.0和5.0。然后我们两个像素分别进行分类。双方的这种分离是很重要的,因为我们主要是对正确分类S-edges感兴趣,因为前台的背景,点S-edges很少保留他们的颜色。

π,pj要与优势相关联的两个像素点,如上所述。每一个点都是使用事例分类器分类为前景或背景。应该注意的是,提供的分类器添加补偿NNFs空间(x,y)坐标的特征向量。然后确定边缘点作为S-edge点,如果π和pj的标签是不同的。我们当前的实现使用FLANN(Muja和劳2009)加速最近邻搜索,和k = 15在我们所有的结果。注意,在分类轮廓边缘点,我们也知道哪一边的边缘是前台。图3演示了我们的边缘分类器的结果,对随后的水平集轮廓及其影响。

3.4 水平集面具细化

面具转让和面具插值的结果通常包含一些噪音和并不总是正确遵循前景物体的轮廓边缘。他们也可能包含一些小型孤立的错误分类的岛屿。我们使用一个基于水平集方法的方法清理面具边界,通过平滑和附近S-edges拍摄它。我们的方法是流行的测地线活动轮廓的扩展方法(Caselles et al . 1997年)。让:⌦!R表示水平集函数,⌦im -年龄域。我们设计这样的零水平的收益率预期前景轮廓C:

Ct = {x 2 ⌦| t(x) = 0},

其中t介绍演化过程。C0初始化的结果NNF-based面具转移(3.1节)和0将签署距离变换应用于C0,与及有效信号在室内(前景)地区,和消极的。

有0,初始化轮廓C是进化迭代更新:t = 1 + d,d定义为:d =↵Eg + Ee + Em

第一项如测地线活动轮廓的功能:

Eg =kr kdiv( )

结合图像边缘和平滑之前。g(我)产生的边缘响应Dollár和Zitnick的方法,和⌧设置为100。

对于每个分类S-edge e,我们计算e在其附近的L2距离变换。边缘词然后定义为Ee = s e,在年代2 { 1 + 1 }是积极的室内一侧的边缘和负外部一侧。哪一方的信息是内部和外部来自边缘的分类器。Ee边缘词的目的从而吸引零水平集轮廓向附近的轮廓边缘。类似的边缘领域术语以前使用的刘和Yu[2012],但是没有分类的优势边缘。

最后,面具之前他们是为了惩罚大偏离初始NNF-based面具的结果,因为我们的目标是只删除小错误和光滑的轮廓。它被定义为Em = !s c s c在哪里签署了L2距离变换的初始轮廓C0面具,和!是一个重量字段计算的残差NNF:!= exp(u)。这是最近邻和u的色差是一个比例因子(0.01我们所有的结果)。

4、结果

在本节中,我们讨论的结果,我们可以获得使用我们的方法,并评估其性能,包括对比几种现有最先进的方法。

我们实现了我们的方法在c++中,用GPU-accelerated im - PatchMatch算法的实现。我们当前的实现通常需要在1.4秒之间转移面具两帧大小960⇥960,在一个英特尔酷睿i7 - 2700 k的3.50 ghz CPU与32 g RAM和GeForce 680 GPU 4 gb的RAM。目前,大约有45%的时间花在补丁——匹配,边分类,30%和20%的水平细化。虽然我们的实现可以进一步优化,目前运行时间已经使互动响应时间。尤其是对于多线程实现,它可以传播一些面具在用户交互与另一个框架。实时捕获一个交互式会话中包含补充视频。

彻底地评估我们的方法,我们收集了五套视频剪辑,代表框架如图4所示。吸附-割集包含三个例子来自白et al。一个——IMAL、人力和训练数据集的静态集钟等。快速组是一套新的我们收集,以非常快的运动前景和显著的变形。总共有22个视频片段,为每一个例子中,我们获得了地面真理前景面具从原始数据集(动物、人类、静态)或通过仔细的手工分割。

4.1 比较和评价

我们比较方法的面具转移精度与几个选择,包括最先进的discontinuity-aware钟方法等。[2012](Z12)和Adobe后遗症Rotobrush工具(RB),基于视频SnapCut方法。为了比较的状态——艺术分割在计算机视觉跟踪结果,我们也包括在这个比较最近SeamSeg方法,已证明了自己作为一个顶级表演者在SegTrack基准。不同的方法错误率的报道在表1所示。错误的比率计算错误分类ar -东亚峰会的转移面具和前台区域的地面真理面具。对于每个测试序列我们使用128帧,计算地面真相掩盖的自动转换框架我我+ d,因为我= 0,16日。,96年为几种不同的传输距离d 2(1、4、8、16、32)。每个视觉检查产生的面具表明,我们的方法比Z12产生更一致的结果,有时倾向于介绍假背景前景地区的孔(见“夫妇”的例子在图5中),这是因为Z12不执行任何一致性约束在每个方向窗口。此外,由于缺乏一致性和使用单一像素颜色分类、Z12更敏感的前景和背景之间的相似的颜色比我们patch-based分类方法(见图5中“马”)。我们的研究结果也明显比由SeamSeg更连贯。相同的表也报道的错误率插值帧中间帧我从两个来源(与地面实况面具)4帧(预告)或8帧(i8)。可以看出插值(双向传输)导致错误率低于单方面转移同样的步伐在所有情况下只有一个(一个特别具有挑战性的序列)。这些结果说明我们遥远的面具传播非常适合一个工作流,它结合了遥远的面具与插值转移。请注意,我们没有测量真实的预告/ i8错误一侧和传播面具另一方面,自从在我们提出交互式工作流用户将修复错误之前我+ 8 + 4 /帧插值应用于生成中间帧。一个工作流的例子所示的实时交互式会话在同伴视频和补充材料。

5、结论和未来的工作

我们提出了一种视频抠图技术专门擅长快速运动序列。面具传递到框架的挑战表现出大运动技术基于局部统计可能不那么有效,非本地更改。我们的方法是基于先进的效率的方法计算两个图像之间的最近邻域(NNF)。关键是这些方法非本地,但同时他们生成一个连贯的映射。通常,在快速运动序列,是前景对象变化快,值得注意的是,虽然背景更稳定,只反映摄像机的运动。因此,我们的NNF前景和背景分离,避免了通常有害努力保持两个连贯的。同一split-NNF分级机也用于我们的优势,进而进一步提高基于水平集的断路器的性能。

能够处理大型前景的变化有两个直接应用。首先,它允许处理快动作序列在连续帧展览光学级流量大的巨大变化。第二,它允许遥远的帧间跳跃,然后插值中间帧,使用减少的步伐。这种插值方案有点反直觉的,因为似乎渐进步骤更相似帧之间产生更少的错误。然而,正如我们已经表明,增量方法积累更多的错误,比向前跳跃,然后插值。

未来工作的一个方法是把信心值面具转让,允许系统自动采用一种自适应步大小掩盖转移和插值。另一个可能的改进是利用循环动作,类似的前景提出了可能会发现许多帧,和超长的进步可以有效地使用。一个有趣的方向,我们也认为,是进一步扩展这一想法,并将视频序列作为一组无序的帧。类似的帧可以集群和加工在一起,不需要依赖自然时间相干视频。

视频序列帧的无序集。类似的帧可以集群和加工在一起,不需要依赖自然时间相干视频。