硕 士 研 究 生 读 书 报 告



题目：对于跳跃视频的剪辑：

非连续性的掩膜转移和视频抠图

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

作者姓名  裴灵晖

作者学号 nb15093

指导教师 李启雷

学科专业 移动互联网与游戏开发

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。鉴于其前景掩模已经可用源帧，我们计算的前景掩模的估计在另一个典型的不连续，目标框架。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.

CR Categories: I.4.6 [Image Processing and Computer Vision]: Segmentation—pixel classification;

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。涉及准确地从自然视频的背景，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，尽管在交互式前景提取令人瞩目的进展、任务仍需要相当大的用户的努力和专业知识。

具体而言，国家的最先进的方法仍然是由颜色和/或前景和背景的地区码之间的纹理模糊性的挑战，以低对比度边缘与背景分离前景元素，并通过快速运动的前景，这典型值为涉及非刚性变形和在前景区域的拓扑剧变。

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。

在这项工作中，我们描述了用于交互式前景提取，专门针对大和非刚性前景运动的新工具。我们处理的交互式视频切口，其中，任务是提取二进制掩码标记各个帧作为前景或背景的像素的问题。软边阿尔法磨砂可以得到，如果需要的话，通过使用我们的面具作为一个三分图。

我们集中于在此工作的基本计算元件是掩模转移：给定一个正确地分割源帧中，目标是计算在另一个前景掩模，非连续对象帧（图1）。在实践中，屏蔽罩传送距离通常介于4和32帧，这取决于运动的速度和前景中存在的视频变形量。正如我们证明我们的结果和同伴的视频，有能力执行这样的掩膜转移准确地提供了一种有效的交互式视频缺口工具的基础。

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,使用长方向的窗户,适当的大小和方向是很难确定的。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发现基于斑块之间的差异的基础上,同时考虑颜色和纹理对应。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。

为了获得最终的洁净，掩码目标框架，我们应用基于水平集的方法。这个最后的步骤变得容易由前两个阶段的结果：水平集方法与我们的分裂NNF预测pixelwise掩模初始化，并依赖于我们的边缘分类的结果，以捕捉到附近的轮廓边，而平滑目标罩轮廓。

整个过程概述了上述结果在成功传播模型更远，比国家的最先进的现有方法更准确地新的掩模传输方法。有了这些改进非连续的面具转移能力，我们也能够实现一个新的面具的插值机制，在给定的两个不相邻的分割帧我们重建前景掩模的中间帧，连贯和准确的方式。

2、相关工作

描述了图像和视频分割的交互系统，面向具有触摸界面的移动设备。他们的方法是基于水平集框架，与采样的触摸的点的附近的外观模型。由于该方法依赖于本地培训参数模型，它是有限的在实践中具有前景和背景区域的非重叠统计的图像，并与小的位移的视频。

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。

描述了一个交互式图像分割方法，它采用了使用边抑制功能，以消除非显着边缘的水平集方法。他们的水平集具有边缘场项，这是使用距离变换的相对于所述显着边缘限定。我们的方法采用了级别设置了类似的术语，但是，我们的边缘分类器区别地使用受过训练，从中我们传播面具分割源帧。

1. 方法

3.1 面具转移

给一个分段的来源IS与二进制掩模框架MS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。

一种方法来掩盖传输到每个像素 It 为背景(BG) or 前景 (FG) u用已训练上对（I，M）不同分类进行分类(Is,Ms). 另一种方法是计算对应'匹配每一个像素×2将其用 It with '(x) 2 Is。 设置Mt(x) = Ms('(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。

如前所述，我们为了计算最近邻域（NNF），用作“使用基于PatchMatch的方法。我们观察到，BG和FG区域中的序列典型地移位以不同的方式。对BG位移通常是由相机相对于所述（主要是静态的）场景的运动引起的，而FG位移通常是更大和更小刚性的。因此，我们的想法是，通过计算一个单独的NNF为两个区域，以利用两个不同的位移模式。

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(湾et al . 2008年),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(Schaefer et al . 2006年)约束下的匹配特征点,导致该校BG-aligned来源。

具体而言，我们首先调整Is It 相对于前景和背景分别的议案、将FG对准是通过模板匹配，其中， Is 掩蔽前景是模板，用作为匹配度量面罩内平方差的和来计算。 已经找到了最佳的翻译前景区域，我们应用Is 来获得 FG对准源Isf . 要调整的背景下，我们执行的是特征匹配在Is a和 It 之间,和它使用SURF. 通过RANSAC异常排斥，接下来，我们经线采用刚性 Is 移动最小二乘（MLS）法下的匹配特征点的限制，造成了BG对准源 Isb。

考虑到两个对齐的来源Is⇤, ⇤ 2 {f, b} , 然后我们计算一对本地化 NNFs 从It到 Is⇤. 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It and Is⇤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。

两个分别对准NNFs计算，如上所述，导致在NNF更少的错误，由于使用本地搜索和改性距离项。具有两个计算NNFs，我们现在它们融合到单个NNF通过选择，对于它的每个像素 It, 为NNF偏移产生较小的补丁差， 掩膜 Mt 预测在本节前面所述(Mt(x) = Ms('(x))). 中也示出了具有水平集方法最后的改进后的转印掩模。

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 It k and It+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 It k and It+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It k to It 。一次是从m It+k to It.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。

要执行形状匹配，我们开始通过提取和简化两源掩码contours1。下一步，我们计算对于两个轮廓点之间的对应关系。两个轮廓点pi和PJ的相似度测量为三个换算的合计：形状差dij , 色差 dij , 空间距离 dspace.我们使用掩模补丁中心在各轮廓点 (of size 31 ⇥ 31)来表示它的局部形状，并计算两个轮廓点之间的形状差异（其相应的贴片之间的SSD通过在像素的数量除以面具补丁）。与每个点相关联的颜色被计算为平均前景色在一个15⇥15图像补丁中心在它， dcolor i这些颜色之间的距离L2. 请注意，只有前景内的颜色的使用， 空间距离 dspace 被定义为 ij kpi + m pjk2, m 在轮廓的中心（意味着点的位置）in It k and It+k是不同的 。最后的距离dij 的计算公式为as dij = dshape + dcolor + dspace .。

3.3边缘分类

人们早已认识到，边在我们的视觉感知起到形象地分离了至关重要的作用，并且许多图像分割算法合并入边缘的过程中，推理该区域边界应与显着边缘重合。然而，前面描述的掩模转移技术是基于NNFs和不明确地获取边考虑。此外，转移的面具是最终清理使用水平集方法，如在下一节中所述。因此，掺入的边缘进的过程中，并防止掩模从与可能存在在其附近不相关的显着边缘对准本身，我们尝试在目标帧中的所有显着边缘进行分类。

具体地讲，我们试图在图像中的显着边缘分类为三种等级：B-边缘（完全包含在背景中），F-边缘（完全包含在前台），和S-边缘（轮廓边分离前景从背景）。应当指出，然而，在我们的管道的剩余部分，我们目前只利用S-边缘。B超边和F-边缘多余的分类来为免费体验，无需任何额外的计算工作，因为每个边缘被归类通过考察，双方将在后面解释。我们用一个简单的非参数监督三通分级机，其训练数据是由源帧及其掩码给出。边缘分类器在原则上是相似的像素分类器，因为它被施加在单独边缘点，但它包含的边缘的方向，并且利用所述两种颜色的在它的两个侧面。

我们首先从源和目标帧采用先进设备，最先进的快速DOLLA r和Zit-缺口的方法提取边缘图。对于每一个点，其边缘响应超过某一阈值，我们关联的两个像素，一个在边缘的每一侧。对于每个这样的像素我们提取7维的特征向量，由（R，G，B，X，Y，sin✓，cos✓），其中的r，g，b是颜色，x，y是空间坐标和✓是梯度的取向。每个维度是标准化为[0，1]的范围内，然后将色彩，空间坐标和方向是由一个因子分别10.0，2.0和5.0，缩放。然后，我们在两个像素分别进行分类。双方的这种分离是很重要的，因为我们主要正确地分类的S边缘在感兴趣，并且由于前景正在移动横跨背景，点在S-边缘很少同时保留其颜色。

让 pi and pj 与边缘点相关联的两个像素, 如上所述。其中每个点处是使用k近邻分类器分类为前景或背景。应当指出的是，分类器增加了空间（X，Y）坐标的特征矢量提供的NNFs偏移。边缘点，然后确定为一个S-边缘点。边缘点，然后确定为一个S-边缘点，如果 pi and pj .的标签不同， 我们当前实现使用FLANN加速最近邻搜索，并在我们所有的结果K =15。需要注意的是已经列为轮廓边缘点，我们也知道在其一面之缘所在的前景。

4、结果

在本节中，我们介绍一些我们能够获得使用我们的方法，并评估其性能，其中包括比较几种现有的先进设备，最先进的methods.We实现我们的方法在C ++中的结果，用GPU加速实施PatchMatch算法。我们当前的实现塔季翁通常在1.4秒要求尺寸960⇥540的两帧之间传输的面具，在英特尔的Core i7-2700K 3.50GHz CPU与32G内存和一块GeForce 680 GPU拥有4GB的内存。目前，时间大约45％用于跳线设置匹配，30％的上边缘分类，以及20％的水平集细化。虽然我们的实施可进一步优化，目前运行时间已经启用交互的响应时间。这是一个多线程实施，这可以传播一些掩模当用户与另一帧进行交互，尤其如此。交互式会话的实时采集包括在补充视频。

为了彻底评估我们的方法，我们收集了五套视频短片，具有如图4所示表示帧的SNAPCUT集包含三个例子的动物，人类和静态集是从忠等人的训练数据集。 [2012年]。快速集是我们收集的，具有非常快的动作和显著前景变形了一套新的。有22的视频剪辑中总，和各实施例中，我们获得从原始数据集的基础事实前景掩模（用于动物，人类，静态）或通过仔细的手工分割。

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。

能够处理在前景大的变化有两个立即的应用程序。首先，它允许处理快速运动序列，其中连续帧表现出与大大小的光流的大的变化。其次，它允许遥远的帧之间高高跃起，然后插在中间帧，使用减少的进展。该插值方案是有些违反直觉的，因为它似乎更类似于帧之间的渐进的步骤应该产生更少的错误。然而，正如我们已经表明，在渐进的方式积累更多的错误，比跳跃式，然后插入。

一个途径为今后的工作是向的置信度值，以掩模转移相关联，允许系统自动采用自适应步幅尺寸为掩模转移和内插。另一种可能的改进是采取环状运动，其中，相似前景的姿势可以发现许多帧分开的优点，并且极其长步幅可以有效地使用。一个有趣的方向发展，这是我们还要考虑，是为了进一步扩大这个想法，并把视频序列作为一个无序的帧。类似的帧然后可以聚集并一起处理，而不依赖于沿视频的自然时间相干性。

**References**

AGARWALA, A., HERTZMANN, A., SALESIN, D. H., AND SEITZ, S. M. 2004. Keyframe-based tracking for rotoscoping and ani- mation. *ACM Trans. Graph. 23*, 3 (Aug.), 584–591.

BAI, X., WANG, J., SIMONS, D., AND SAPIRO, G. 2009. Video SnapCut: Robust video object cutout using localized classifiers. *ACM Trans. Graph. 28*, 3 (July), 70:1–11.

BAI, X., WANG, J., AND SAPIRO, G. 2010. Dynamic color flow: A motion-adaptive color model for object segmentation in video. In *Proc. ECCV*, Springer-Verlag, vol. V, 617–630.

BAO, L., YANG, Q., AND JIN, H. 2014. Fast edge-preserving patchmatch for large displacement optical flow. *IEEE Trans. Im- age Processing 23*, 12 (Dec.), 4996–5006.

BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLD- MAN, D. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph 28*, 3, 24.

BAY, H., ESS, A., TUYTELAARS, T., AND VAN GOOL, L. 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding 110*, 3, 346–359.

BROX, T., BRUHN, A., PAPENBERG, N., AND WEICKERT, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *Proc. ECCV’04*. Springer, 25–36.

CASELLES, V., KIMMEL, R., AND SAPIRO, G. 1997. Geodesic active contours. *Intl. J. Comp. Vision 22*, 1, 61–79.

CHEN, Z., JIN, H., LIN, Z., COHEN, S., AND WU, Y. 2013. Large displacement optical flow from nearest neighbor fields. In *Proc. CVPR’13*, 2443–2450.

CHUANG, Y.-Y., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2001. A bayesian approach to digital matting. In *Proc. CVPR 2001*, IEEE Computer Society, vol. 2, 264–271.