

硕 士 研 究 生 读 书 报 告



题目 无缝的视频的快速计算循环

作者姓名 曾琳

作者学号 NB15126

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○一五年12月

Fast Computation of Seamless Video Loops

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Qilei Li

By

Lin Zeng

Zhejiang University, P.R. China

2015

摘要

本文循环视频短片简明地捕捉自然的活力场景。创建无缝循环通常涉及到最大化时空一致性和应用泊松融合。我们把的端到端视图问题和现在的新技术,共同提高循环质量同时也显著减少处理时间。一个关键概念是放松一致性约束预测随后的混合,从而使循环lowfrequency内容云移动和改变照明。我们也分析了输入视频删除一个不受欢迎的倾向短的循环。直观地演示了收益和质量证实定量使用新的gradient-domain一致性指标。我们通过分类可能提高系统性能loopable像素,屏蔽2 d图切割、修剪graph-cut标签基于主导时期,在粗网格优化

保留更多的细节。在这些技术减少计算乘以几十分钟近实时。

**关键词**：视频纹理, cinemagraph, blend-aware, consisten

Abstract

Short looping videos concisely capture the dynamism of natural scenes. Creating seamless loops usually involves maximizing spatiotemporal consistency and applying Poisson blending. We take an end-to-end view of the problem and present new techniques that jointly improve loop quality while also signiﬁcantly reducing processing time. A key idea is to relax the consistency constraints to anticipate the subsequent blending, thereby enabling looping of lowfrequency content like moving clouds and changing illumination. We also analyze the input video to remove an undesired bias toward short loops. The quality gains are demonstrated visually and conﬁrmed quantitatively using a new gradient-domain consistency metric. We improve system performance by classifying potentially loopable pixels, masking the 2D graph cut, pruning graph-cut labels based on dominant periods, and optimizing on a coarse grid while retaining ﬁner detail. Together these techniques reduce computation times from tens of minutes to nearly real-time.

**Keywords：**video textures, cinemagraphs, blend-aware consistency

1介绍

视频的空间分辨率是接近的数字(如照片。像素的视频和产品照片当前手机)。视频内容是因此变得更加普遍,我们预计,随着存储和带宽继续扩大、视频将取代照片为默认捕获的媒介。本文在计算短视频循环周期运动(如。摇曳,树、水)荡漾在自然场景,因此循环帮助传达比图片更强烈的现场感。我们的目标是创造视频循环无需用户援助,就像自动模式对于消费设备拍摄的照片,和更多比之前的方法有效。几个技术创建循环从短视频输入视频。一般方法是组装内容从原始视频,3 d时空产生的视频社区循环符合输入的视频。通常这是扮演的一个组合优化与客观颜色的一致性。在本文中,我们描述了一个端到端的管道生成视频循环更高质量和更少的计算努力比先前的方法。我们引入了一些新的技术共同应对这两个挑战。

提高循环质量对于许多场景,颜色一致性约束不能完全满意,导致空间缝隙或颞持久性有机污染物。几种方法旨在减少或隐藏这些杂色。羽毛或多分辨率花键加工作为一个后处理。Gradient-domain泊松混合,从而提高结果扩散时空差异。自适应一致性约束认识到不连续不太明显的衰减高频区域。麻烦的场景地区取而代之的是静态(nonlooping)像素,使用辅助分割或作为优化的一部分。然而,对于某些场景可能有少活力了。我们的核心思想是调节循环中的一致性的目标合成算法预测泊松的后续步骤混合,从而为优化提供更大的灵活性。在(如,特别是低频图像差异。光滑的强度由于光照变化或移动云/烟)可以忽略,因为他们很容易通过泊松纠正吗混合。相比之下,不同形状的边界并不容易修好了。我们还表明,视频的质量循环的好处从给予更长时间的偏好。快速循环计算最先进的视频需要很多分钟的循环优化计算。我们描述了一些算法的改进,减少处理时间在桌面PC到7秒5秒钟全高清视频,即。,几乎是实时的。我们回顾之前的工作,包括框架在我们构建,然后提出我们的过程

快速循环计算最先进的视频需要很多分钟的循环优化计算。我们描述了一些算法的改进,减少处理时间在桌面PC到7秒5秒钟全高清视频,即。,几乎是实时的。我们回顾之前的工作,包括框架在我们构建,然后提出我们的处理管道在第三节的贡献。

**2 背景和相关工作**

给定一个输入视频颜色V(x,ti)在每个2 d像素x和输入帧时间ti,目的是计算一个视频循环L(x,t)= V(x,φ(x,t)),0≤t < t,通过确定一个设定的函数φ(x,t)注意循环内容L(x,·)在给定位置x是取自相同的像素位置V(x,·)在输入视频.

技术可以定义时间的对比映射φ。同时过渡区域通过寻找兼容的帧。如果V(·,tA)≈V(·,tA + p),好映射是φ(x,t)= tA + p(t mod)。Kwatra等。[2003]像素在不同的时间过渡。他们获得一个循环时期p通过求解一个二进制的3 d图形切换变量b(x,t)∈{ 0,1 }这样φ(x,t)= t0 +(t mod p)+ b(x(t国防部p))p。从视频创建一个循环通过允许子虚乌有函数φ(x,t)=δ(x(t国防部p))有任意时间偏移量δ稳定输入视频

和解决multilabel 3 d图。几个技术利用用户指导创建循环。的互动的工具Tompkin等。函数静态和循环区域创建cinemagraph。Joshi等。开发一系列的成语(静态、玩、循环和镜子循环)结合几个时空段从一个视频来源,因此强调特定场景元素或形成一个故事。白等。应用空间扭曲的视频内容有选择地动内容。引导用户中风,

他们的方法消除了大规模运动,同时保留高频运动。在手持使用跟踪视频创建cinemagraph画像。使用变形为的情况下创建一个循环连续的前景对象,可以从其背景分割。我们的工作基于自动化技术等。优化一个循环周期px,开始在每个像素帧sx:φ(x,t)= sx +((t−sx)px)。与nonloopable内容区域,一个像素可以转让期px = 1,静态冻结其颜色sx框架。期的目标是确定一组p = { px },开始帧s = { sx },最小化的目标E(p,s)= Econsistency(p,s)+ Estatic(p,s)。(1)E这个词时空一致性= Espatial + Etemporal措施邻近的颜色的一致性在视频中循环的尊重输入视频(《et al . 2005年)。空间项金额一致性对所有相邻像素空间x,z:E空间= Xkx−zk = 1Ψ空间(x,z)γs(x,z)Ψ空间(x,z)= 1

TT−1X t= 0kV kV(x,φ(z,φ(x,t(x,t))))−−VV(x,φ(z,φ(z,t(z,t))))kk 22 +和颞术语一致性循环端架sx和sx + px每个像素:E时间= XxkVk(x,V s(x,x−x1))−−VV(x,s(x,xx + + pp xx)k−21)+k2γt(x)。γs因素,γt减弱高频区域的一致性[Kwatra et al . 2003;白et al . 2012]。最后,这个术语Estatic分配一个点球静态像素来防止一个微不足道的所有静态的解决方案。

最小化是一个二维马尔可夫随机场(MRF)的问题,在每个像素分配一个标签(px,sx)外产品{ p }⊗{年代}候选人的时间,开始帧。一个(近似)解决方案被发现使用multilabel图算法,它多次遍历标签的集合柯尔莫哥洛夫和Zabih 2004]。对于每个标签α,它解决了2 d二进制图像切割来确定像素应该保持目前的标签或被分配标签α——这被称为α扩张。最后,应用时空羽毛面具的缝合由此产生的视频厕所.

**3 概述和贡献**

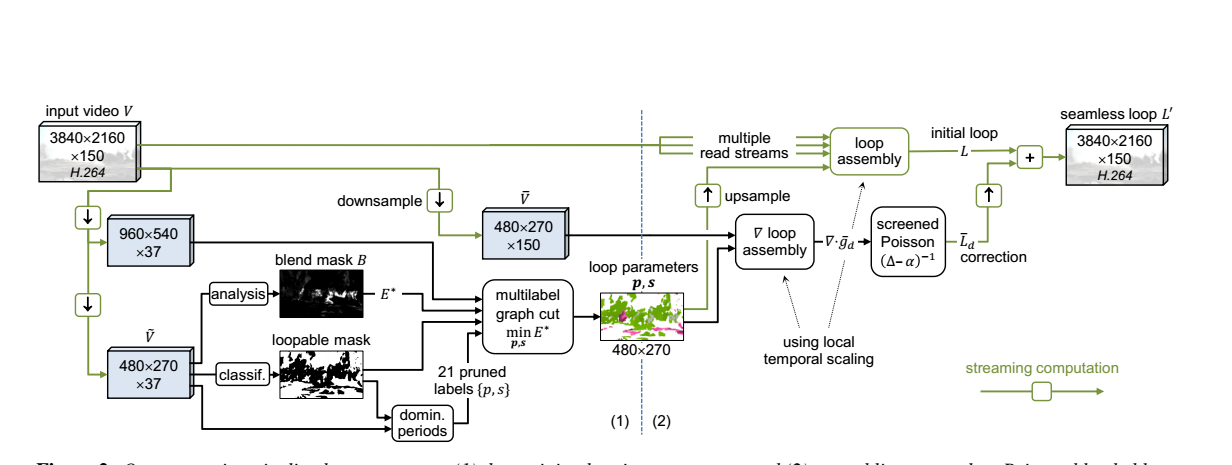


图1

图一总结了我们的处理管道。我们优化每个像素的周期p和启动帧秒。我们采用梯度域（泊松）混合弥漫不一致。我们的贡献包括：

1. 粗化二维优化领域,同时保持了finer-scale细节的准确性。
2. 修改时空一致性使用混合的面具预测泊松混合提供的机会。
3. 分类像素加入影片箱以减少使用2D二元掩模优化领域。

**4 当地时间比例**

优化/ p,s允许像素有不同的循环期。一个缺点是,这样一个表示不支持在常见的视频播放组件。虽然它是可能的创建一个重复循环的长度是最小公倍数所有时期的p,如此循环往往还长。我们的方法是暂时规模等内容循环时间调整到最近的整数的实例数量一个固定大小循环,如。,5秒长。当生成150帧循环,给定一个像素的循环内容是40帧长,我们暂时规模的内容0.9375倍获得完全4循环实例。数学上,我们获得摄动子虚乌有函数φ0(x,t)中增加输出时间t有时重复或跳过一个输入视频帧。时间比例不引入明显的工件因为所有像素同期调整相同因素,所以保存他们的时空关系。它只是在像素之间的界限不同时期空间缝隙可能会恶化。幸运的是,空间一致性Ψ成本两个相邻像素之间的空间(x,z)与不同时间已经使得这些时期最坏的假设都是独立的,所以这些界限躺在像素与相对不变的内容和时间比例没有明显效果。相关的方法是创建独立的循环区域。这也包括像素共享。他们的目标是有选择地冻结这些区域空间控制活力。冻结一个区域可以看作是一个极端的例子时间比例。所有的结果在本文中,我们使用这个地方时间比例方案生成普通5秒循环(T = 150帧)。如图2所示,考虑到循环参数p,s计算在multilabel图切割,当地时间比例是用来定义一个初始循环L和制定泊松融合范围决定了最终L0循环。

**5 提高循环质量**

在本节中,我们描述几个技术来获得更好的循环,以更大的动态性和改进的时空一致性。解决的弱点minp s E(p,s)在先前的工作不能占混合泊松可能的事实分散的矛盾获得无缝的解决方案,作为一个简单的例子,考虑一个场景的照明照亮慢慢地随着时间的推移。时间一致性比较了颜色在起始帧和结束帧之间循环。因为颜色不同,优化可能会支持短循环或可能甚至完全冻结的场景,而泊松融合消除低频光照变化甚至很长一段循环。尽管对于这种情况可以在全球范围内调整照明进行预处理,泊松融合的好处是,它适用更普遍。例如,它也是有效的空间当地的差异。理想情况下,我们想最小化gradient-domain-basedE0(p,s,10)= Econsistency 0(p,s,10)+ Estatic(p,s)(2)在组合循环参数p,s和决赛L0混合的颜色。然而,这是一个挑战,因为变化在p,年代导致结构性变化所需的梯度场g。相反,我们的方法是减少(1),但使用修改后的客观E∗= Econsistency∗+ Estatic一致性度量在哪里blend-aware。从输入视频中,我们计算空间混合使用的面具B调节时空一致性条件。直观地说,B(x)是小如果不遍历像素x任何锋利的边界输入视频,即。,如果是在一个可混合的地区。从概念上讲,我们要计算着面具B在每个像素的基础最大输入视频的时间导数的高通滤波信号。我们应当看到在接下来的推导过程,这是很好近似仅仅通过的最大时间导数。让六世= V∗G的空间模糊版的输入视频,用一个空间高斯滤波器g .高通滤波信号

因此VH = V−六世。它的时间导数V t + 1H−VH t = V t + 1−t V t + 1∗G−−V t∗G= V t + 1−V t t + 1−−V V t∗G≈V t + 1−V t。近似利用颞衍生品这一事实级较低的空间模糊后,变得可以忽略不计.

**6 快循环计算**

我们探索一些加速技术。为了简化演讲中,我们假设输入视频和计算循环都是5秒长,采样30帧/秒。

**6.1时空downsampling**

我们首先计算一个时空的V˜downsampled版本输入使用3 d视频框过滤器。时间比例因子总是4。空间比例因子是2的幂,产生的垂直大小不超过350。例如,一个输入视频分辨率3840×2160×150是480×270×37了。所有的计算都是在V˜,除了图切定义其目的使用next-finer-level细节(6.5节)在完全的决议和泊松混合输出(6.6节)。

**6.2分类loopable像素**

鉴于downsampled视频,我们快速识别空间区域不可能形成良好的循环,以减少优化

像素的一个小子集。每个分类的方法像素的3类:unloopable不变(V˜常数)(动态但无法循环),或loopable。像素分类为不变的或unloopable是静态的,没有考虑的优化。分类应该保守,犯错标签的一个像素loopable,这在最坏的情况下,优化仍然可以冻结像素。接下来,所述计算初始二进制分数(0,1)独立在每个像素的三个类,空间平滑分数,最后每个像素进行分类根据其最大得分。

空间平滑

我们应用高斯模糊(σ= 7像素)这三个字段。这有助于消除小岛屿静态像素在较大的循环区域以及小岛在静态动态像素区域,这两个视觉让人分心。投票的最后,我们将每个像素根据其最大值平滑的分数。图10显示了平滑的效果最终的分类和产生的二进制的面具loopable像素。我们可以省略的计算混合面具B(5.2节)nonloopable像素,但这并没有导致加速。在

另一方面,我们排除nonloopable像素的估计主要循环时间(在6.4节)和发现极大地提高了质量.

**6.3 标记2d图形**

当nonloopable像素被排除在图表,图不再是普通的2 d网格。尽管可以调用版本的图切为通用图设计,我们发现它是更有效的保护常规2 d的连通性图,而是修改graph-cut考虑解决方案二进制的面具。具体来说,我们省略计算成本方面的数据对于任何nonloopable像素(因为它不能改变)。而且,对于任何一个nonloopable x毗邻loopable像素z,我们转移

平滑成本E∗空间(x,z)的数据成本loopable像素z。并行性,我们的实现建立在multithreade

刘和太阳的图割方法[2010]。

**6.4修剪候选标签**

考虑所有时间{ p }和所有可能的开始为每个时期帧{年代}。对于一个4×暂时downsampled5秒(即输入视频。37岁的帧),最小循环周期8帧,这导致multilabel图削减502标签(s = 0,1,。37 | p = 1;s = 0,1,。29日| p = 8;s = 0,1,。28日| p =9,。;s = 0 | p = 37)。在所有这些表演α扩张标签是昂贵的。我们一些修剪这组21候选人如下。

我们发现它是有用的识别输入两个占主导地位的时期视频中,长期提供更大的内容不同,短作为后备地区之一,没有好的长循环。我们使用调整同步时间成本(p,s)博士

**6.5识别最有前途的同步循环**

(p1,s1)= argmin(p,s)博士(p,s)。我们只计算这些成本6.2节中确定的loopable像素。同时,我们不允许循环时间大于4秒(p > 30)因为循环长度方法输入视频的长度有各种不足

开始框架允许创造良好的循环。然后我们找到最承诺(p2,s2),(1)p1和p2时期不同至少25%的最大循环期间,和(2)的两个循环重叠,即。,(s1,s1 + p1)∩(s2,s2 + 6 =∅p2)。为每一个占主导地位的两个标签(π,si),我们也选择9最近的开始帧作为额外的候选人,总共20标签。21标签是一个静态框架(p = 1),这是选择作为中间帧重叠的两个循环(π,si)。21的设置标签的两阶段优化[廖et al . 2013]是不必要的。我们解决一个multilabel图。所有的像素都初始化标签,对应的时间两个循环(π,si)发现,当我们发现优化有一个容易改变的短比亦然。在表4所示,修剪的标签可以减少循环质量我们的例子的结果(目标E0增加从48.851.9),但幸运的是很小的变化。

**6.5没图切割优化**

随着图形削减是一个计算瓶颈,是很重要的在粗空间分辨率(如执行它。V˜,480×270)。然而,在这项决议,我们发现能细微细节的损失大大降低了估计的时空一致性。我们的解决方案是在双重评价一致性的目标V˜的空间分辨率。图11说明了建设对E∗空间。我们定Etemporal∗同样使用精细的总和在每个块像素差异。实际上,我们正在解决问题在更高的分辨率(例如。960×540),但限制标签2×2空间粒度。描述不同的多尺度策略。他们逐步解决方案和优化节点只有在社区的接缝

发现的粗分辨率。具体来说,对于每个alpha-expansion,他们只考虑节点在距离10的有特定的α标签。我们最初尝试了类似的策略为我们的2 d图形切割和发现解决方案更容易受到局部最小值差继承粗尺度。

**7减少内存使用**

存储3840×2160 30 fps 5秒钟视频使用4个字节/像素需要5.0 GB的内存。我们使用NV12表示(halfresolution浓度)减少到1.9 GB。输入视频V是马上downsampled流,输出

循环L0也以流媒体的方式生成和压缩。

因此,内存瓶颈的计算混合循环L0 = L + Upsample(Ld),因为循环L(x,t)= V(x,φ0(x,t))需要随机访问输入的全分辨率帧视频。起初,似乎整个输入视频V必须常驻内存。但是,因为我们只有两个限制候选标签占主导地位的时期p1,p2,内容需要生成L(·t)来自于只有少数{ k1 },{ k2 }的输入框:0≤(t mod p1)+ k1p1 < t和0≤(t mod p2)+ k2p2 < t(参见图12)。因此,一个有效的解决方案是进步多个同时读取流等输入视频V只有这些框架内存驻留为每个输出帧t。当处理一个5秒钟UHD视频,使用多个读流降低最大内存使用3.4 GB 2.1 GB。我们希望看到一个更大的储蓄,似乎使用的Windows媒体视频处理引入了一个重要的基础每个流读写内存开销。另一方面,这个制定开销是独立的视频长度。

**8结果与讨论**

像素分类的百分比根据内容nonloopable千差万别,做占主导地位的时期p1,p2预处理和识别分数像素分配给是静态的。一列分析提高活力(循环的分数像素)的放松约束blend-aware一致性度量。作为分配的均匀性的测量回路参数标签,我们报告的熵分布对所有像素。的计算时间得到3.2 ghz的英特尔四核至强W3565(2009年收购)。这些时间与视频仅略有不同决议,因为大部分的计算发生在V˜coarserscale视频。resolution-dependent是将采样工作输入和最终循环的装配包括求和与upsampled gradient-domain修正。表2显示了一部分时间花在循环创造的每一步。的成本压缩输入视频和压缩输出循环具有重要意义。当报道我们忽略这两个步骤计算时间。计算的瓶颈是图优化和泊松融合是紧随其后。如图所示在表3中,我们的计时结果约两个数量级比先前的技术快,甚至在处理视频更高的分辨率。

我们提供相应的视频的比较。结果表明blend-aware一致性使更大的活力(例如,更大比例的循环像素),循环质量不是加速技术负面影响的第六节。

我们继承的许多限制了之前在视频自动循环工作。的一般方法是最有效的视频类的例子的结果,即静止的观点自然场景和有机运动。我们观察到一些自然的动作,比如波浪海滩,有时间超过5秒输入视频短片我们获得。对于这些我们必须捕获再输入并允许延长最大循环期与移动场景人或不同的对象通常是有问题的,因为这些锋利的划定边界和可能缺乏重复运动。在这种情况下,我们的技术可能会冻结的对象,只留下一个动态背景。在某些方面,这是相反的效果寻求在cinemagraph[Beck和伯格2012年],通常在哪里巧妙地动画前景对象的静态背景。创建有效的cinemagraph通常需要用户或指导控制环境。空间内的自然环境,是我们的贡献扩展场景的类别,可以成功地循环而不需要用户输入,认识到许多低频的场景改变(如。、移动云、烟、蒸汽、阴影)通过泊松融合平滑了。本地时间扩展(第四节)会影响现场的速度运动。为了更好地约束这种时间扭曲,我们可以选择视频的长度循环作为一个函数的两个时期p1和p2确定施工。特别是,循环长度可能是MAX(p1,p2)。

**9总结和未来工作**

我们已经提出的技术计算无缝视频循环时，以提高质量和效率。总之，这些技术允许更高质量的结果计算幅度约两个数量比以前的工作速度快。有几个未来工作的方向。我们假设输入视频稳定。这将是有趣的重温视频稳定的特定情况中循环。或许特征跟踪可制成用于周期性场景运动更加坚固，从而使这些运动是更容易从相机抖动分辨。对于用来衰减的可混合区的一致性混合掩码B，这将是有趣的探索这种结构的考虑单独Bspatial和Btemporal掩码的推广。我们设计我们的技术过程输入视频短片并创建相应的短循环。对于一些场景有用的考虑再输入序列识别更好短循环或再循环形式。计算成本我们的技术和内存需求应该呈线性比例关系输入的大小。事实上,如果一个人认为固定数量候选人(循环参数)的标签,计算成本次线性增加。循环计算现在可能不够快是实用的移动设备如智能手机和相机。这将是有趣的探索是否可以进一步加速专门的硬件。

参考文献

[1] Fast Computation of Seamless Video Loop. Jing Liao, Mark Finch, Hugues Hoppe．ACM Transactions on Graphics (Impact Factor: 4.1). 10/2015; 34(6):1-10.