

基于 Kinect 深度数据的视频艺术化处理

汤颖 孙康高

(浙江工业大学计算机科学与技术学院 杭州 310000)

摘要 利用深度摄像机 Kinect for XBOX360 提取视频的深度信息来实现视频前景和背景的分离,并分别对视频前景和背景进行多风格的艺术渲染,从而获取更好的视频风格化定制效果。首先,系统利用 Kinect 深度数据实现视频前景的提取;然后在光流场指导下,利用基于纹理传输的方法对视频前景和背景进行不同风格的艺术化渲染;最后,将风格化后的前景视频和背景视频进行融合,从而得到最终的风格化艺术视频。另外,由于采用纹理传输的方式实现对视频的艺术化处理,因此用户可以选择不同的纹理样本来实现自定义的多风格艺术渲染。经过实验测试,前景和背景视频融合后生成的风格化视频取得了较好的艺术效果,从而证明了该系统具有较好的视频前景提取能力和视频风格化渲染能力。

关键词 视频风格化, Kinect, 深度数据, 纹理传输, 光流场

中图分类号 TP399 文献标识码 A

Video Stylization Based on Kinect Depth Information

TANG Ying SUN Kang-gao

(School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310000, China)

Abstract This paper focused on extracting the depth information from Kinect for XBOX360 to separate the video foreground from background and stylizing the foreground and background with different artistic styles. First, we achieved the extraction of video foreground based on the depth data. Next, the foreground and the background video were stylized using different artistic styles based on texture advection which is guided by optical flow field. Finally, the final stylized video was obtained by combining the above two results effectively. We stylized video with texture advection-based method so that multiple rendering styles are supported. The experimental results show that the video stylization produced by our system achieves a good artistic effect.

Keywords Video rendering, Kinect, Depth data, Texture advection, Optical flow field

1 引言

计算机图形学在最近几年的发展日新月异,视频电影媒体处理技术的进步非常快,人们对于视频的艺术化有了越来越高的要求。基于这些因素,视频风格化渲染技术也随之有了很大的发展。同时,在这段时间内,由于各种体感交互设备的上市和快速发展,一种全新的图像模式——深度图像逐渐被人们所熟知。微软公司 Kinect 的发布更是引发了深度图像处理的狂潮,不少专家学者都开始致力于利用 Kinect 特有的深度数据、骨骼数据等信息来进行相关的研究,优秀的研究成果也是层出不穷。虽然不少专家学者在各自的领域都有很多不错的研究,但把这两个领域结合起来进行研究的还比较少。鉴于此,我们希望把这两个领域做一次有效的结合,在原有的视频风格化技术的基础上加入 Kinect 特有的深度信息提取能力,实现视频两个层次多种风格的艺术化处理,从而提升视频整体的艺术感染力。

本文的主要工作就是利用 Kinect for XBOX 360 深度摄像机摄制视频并进行实时的深度信息提取,同时利用深度信息实现视频前景和背景的分离;然后分别对前景视频以及背

景视频在光流场指引下基于纹理传输的多风格艺术化渲染,渲染的效果取决于用户选取的纹理样本,最后把风格化后的前景视频和背景视频融合成具有良好艺术效果的自定义风格化视频。

本文第 2 节介绍了相关工作;第 3 节给出了用 Kinect 提取视频前景的 3 种方法,同时通过实验分析比较,最终采取效果最优的方法;第 4 节介绍了在光流场指导下的基于纹理传输的视频风格化技术,并介绍通过纹理修补来保证风格化视频的时域连续性;第 5 节通过实验对系统的各部分进行测试,并在前景和背景融合阶段通过一个融合方程来提高视频融合的质量;最后总结全文并对未来的工作进行了展望。

2 相关工作

2.1 Kinect 应用开发的研究现状

利用 Kinect 来开发相关应用的例子有很多,如利用深度数据实现人体的骨架提取,即对经过二值化的深度图的边缘进行平滑和细化处理,就可以得到整体骨架^[1],这种方法不仅方便,而且效果也比其他骨架提取的方法好;或者利用 Kinect

本文受国家自然科学基金(71571160),浙江省自然科学基金(LY14F020021)资助。

汤颖(1977—),女,博士,副教授,硕士生导师,CCF 会员,主要研究方向为计算机图形图像、虚拟现实和信息可视化, E-mail: ytang@zjut.edu.cn;

孙康高(1992—),男,硕士生,CCF 会员,主要研究方向为数据挖掘、信息可视化。

来驱动人脸动画合成,人脸跟踪客户端利用 Kinect 跟踪并识别用户脸部表情,把 Kinect 传回的参数再发送给人脸动画合成服务器,根据人脸动画定义表可实时绘制出与用户表情对应的三维人脸动画^[2];也有研究利用 Kinect 的实时骨骼跟踪技术,结合逻辑回归分类算法对人体的各种姿势进行实时地识别^[3]等。

与传统的彩色图像不同,深度图像能反映场景的三维结构信息,可以很容易地从深度图像中获知场景中物体到摄像头之间的距离。但是,由于受 Kinect 深度测距原理及其精度所限,Kinect 的深度图像存在以下 4 方面的问题:与距离的平方成正比的噪声、由无效像素组成的空洞、深度图像与彩色图像边缘不匹配和深度图像在时域上的闪烁。这一系列的不足和问题也引发了学者的研究,各种去噪算法、深度图像修复和增强算法的出现在一定程度上解决了这些问题。如,文献^[4]利用基于 Prewitt 算子的分块自适应阈值边缘检测算法得到图像边缘,然后对边缘和非边缘区采取不同的中值滤波算法来进行降噪处理,获得了较好的效果;文献^[5]提出的基于改进的双边滤波深度图像增强算法和文献^[6]提出的基于区域生长和距离变换的错误像素去除方法都能在一定程度上解决上述问题。

与传统的彩色图像相比,深度图像在目标检测、物体识别、人体姿态估计、三维重建等领域有着广泛的应用,但是涉及将图像的深度数据、骨骼数据与视频风格化结合起来的研究还很少。而本文就是要创新性地把这两个领域结合起来进行研究。

2.2 视频风格化技术研究现状

作为非真实感绘制的一个重要分支,视频风格化处理技术大致可归纳为 3 类:基于光流的视频艺术化、基于非线性滤波的视频艺术化以及基于分割的视频艺术化。基于光流的视频艺术化方法又可以分为基于标记(笔刷)和基于纹理的方法^[7]。本文所采用的风格化技术就是基于光流场的视频艺术化方法中的纹理传输的方法。

本系统可以实现油画艺术风格的视频风格化,而基于模拟的绘制技术和基于学习的绘制技术是两种主要的油画风格化绘制方法^[8]。顾名思义,基于模拟的绘制技术就是模拟油画艺术创作者真实的油画绘制过程,该技术可以利用算法在画布上叠加各种层次颜色、尺寸的笔触,从而可以在一定程度上生成具有油画艺术家笔迹的真实油画风格。但是基于模拟的方法比较复杂,需要对每个渲染笔刷的各种属性进行维护,较难实现。而对于一些手绘感艺术效果的渲染如油画水彩画风格等,该方法确实比其他渲染方法更具优势。文献^[9]提出了一个交互式的系统,其通过视频分析来获取语义内容,然后建立稠密特征对应,从而通过定位和传递笔刷来完成最终的油画感动画的风格化绘制。第二种绘制技术则通过纹理合成的方法来模拟模板图像的风格特征,因此使得输出结果兼具目标图像本身风格和模板图像风格,这样就可以利用不同的输入模板实现多风格渲染,实现起来也比第一种方法简单。

每种视频风格化算法都需要面对时域连续性的问题。现有的视频艺术化方法,大部分都是从图像处理算法扩展而来的,使渲染出来的视频或多或少的都会产生不连续的现象。

保持时空一致性是视频风格化处理的一个难点,也是评判这个风格化算法好坏的关键点。文献^[10]提出了基于运动层操作的 Video Painting 算法,该算法较好地解决了直接将

图像风格化算法运用到视频帧上造成的时空上不连续的问题。为了增强视频的时域连续性,文献^[11]提出了高度时域连续性的点画风格视频艺术化算法。文献^[12]选择性地对抽象图形进行颜色量化,通过颜色量化,可以模拟出一些卡通、涂鸦的效果。标准的量化算法,一些微小的亮度变化可能会在输出图中被放大,也就是说其对噪声非常敏感。而该文提出的软量化方法中,以上的那些噪声点会被平均到它的周围区域,让其变化不那么显著,进而保证了时域连续性。而文献^[13]则利用线特征的轮廓分析改进视频中线提取的时间连续性。而本文采用光流场来保证视频的时域连续的方法也能取得较好的效果。对于 3D 模型动画,文献^[14]没有使用光流场来指导色素的运动,而是通过 3D 模型的运动分解获取的运动向量来指导色素运动,从而增加时域连续性。以上各种方法虽然各有优缺点,但都可以在一定程度上解决视频的时域连续性问题。

3 Kinect 实现视频前景的提取

深度数据是 Kinect 的精髓。本节通过对 Kinect 深度数据进行分析,从原理出发,综合列出了 3 种前景视频提取的方法:基于深度距离的前景提取、基于用户索引的前景提取和基于 Background Removal API 的前景提取。通过实验比较 3 种方法的优劣,最终采取效果最优的方法。

3.1 Kinect 深度数据简介

Kinect 摄像机主要由 3 个组件构成:一个彩色摄像机、一个多阵列麦克风和深度传感器。其中深度传感器主要由一个红外投影机和一个红外摄像头(单色 CMOS 传感器)组成^[15]。这 3 个摄像头则相当于 Kinect 的 3 只“眼睛”,它们除了能以 30 帧/s 的频率采集高、中、低 3 种分辨率的彩色视频影像,还能采集到普通的彩色摄像头无法采集的深度数据。同时,深度传感器传回的深度信息构成了深度视频帧,与彩色摄像机传回的彩色视频帧具有一对一的同步对应关系。而深度视频帧的像素值就是平常所说的深度数据或景深数据。这样,就可以很方便地结合深度视频帧的深度信息来对对应的彩色视频帧进行相关的处理,比如实现彩色视频帧前景的提取等。

每一个深度数据为 16 位,其格式如表 1 所列。深度数据的前 13 位表示物体与摄像头之间的距离,后 3 位表示用户索引。对于 Kinect for XBOX360,传感器的有效视距为 1.2~3.5 m,2.25 m 为最佳距离^[16]。关于用户索引位,当它数值为 0 时,表示 Kinect 将该点归为背景点;当它数值为非零时,表示 Kinect 将该像素点归为用户目标。一个数值代表一个用户,故在同一帧中最多可同时追踪 7 个用户,但出于效率和精度的考虑,Kinect 规定只能最多同时识别 6 个用户和同时跟踪 2 个用户。

表 1 深度数据格式

Depth data	Bits of Depth	Bits of index
Bit	D15-D3	D2-D0

为了实现最终的风格化效果,需要完成对视频前景的提取。基于以上对深度数据的分析,很容易想到两种方法来现实这个目标。因为深度数据由深度距离和用户索引两部分组成,因此可以分别根据这两个数值来进行前景的提取。

3.2 基于深度距离的前景提取

由于深度图像帧和彩色视频帧具有一一对应的关系,因

此可以设置一个距离阈值 x , 如把 $1.2\text{m} \sim x\text{m}$ 之间的目标设为前景, $x\text{m}$ 之外的目标设为背景, 这样就可以根据深度数据中的距离信息把前景从视频帧中提取出来。通过下面这一条简单的程序语句, 就可以得到某一像素值的景深距离。

```
var depth = pixel >> DepthImageFrame, PlayerIndexBitmaskWidth
```

其中, `PlayerIndexBitmaskWidth` 是用户索引位的长度, 也就是 3。另外, 为了更加清楚地显示分离效果, 对分离后的彩色视频帧进行二值化处理, 把前景的 RGB 值设置为 255, 背景的 RGB 值设置为 0。把阈值 x 设置为最大距离的一半后得到的结果如图 1 所示。



图 1 基于深度距离的前景提取结果视频帧

从二值化图可见, 提取的前景效果很一般。前景中的一些区域存在明显的黑洞, 前景的边缘也还平滑。如果把这些二值化视频帧整合成一段视频, 则可以发现, 前景的边缘在不断地跳动和闪烁。由于二值化图是根据深度图得到的, 因此可以推断随着时间的变动, 视频帧中某些区域的深度值是不稳定的。

3.3 基于用户索引的前景提取

本小节利用用户索引位来帮助完成提取任务。现在, 直接把视频帧中的人作为前景, 把除人之外的物体作为背景。这样, 可以把用户索引位大于 0 的像素值归为前景像素值, 把用户索引位等于 0 的值归为背景像素值。可以根据以下这条程序语句得到像素值的用户索引位。

```
var player = pixel & DepthImageFrame, PlayerIndexBitmask
```

其中, `PlayerIndexBitmask` 是用户索引位掩码 0000000000000111。此外, 发现 Kinect 的彩色数据流可以实现 BGRA 四通道 32 位图像格式的传输, 而且 Kinect 默认情况下传回的每一个像素的 Alpha 值均为 255, 即完全非透明。于是可以利用 Alpha 值把用户索引位为 0 的像素值的 Alpha 值设为 0, 即完全透明, 这样就可以把人从彩色视频帧中提取出来。利用此方法得到的结果如图 2 所示。



图 2 基于用户索引的前景提取结果视频帧

从图 2 中可以看到, 前景的提取结果依然不太理想, 边缘有明显的锯齿和其他非人体的像素点存在, 头部还有非常明显的像素空洞。

3.4 基于 Background Removal API 的前景提取

微软除了发布了 Kinect for Windows SDK 供编程爱好者来开发 Kinect 相关的应用之外, 同时也发布了 Developer Toolkit, 其中包括了 SDK 的联机文档、示例工程代码和 Kinect Studio 工具^[16]。Kinect Studio 工具包里含有一个 Background Removal API, 该应用程序编程接口可以让 Kinect 实

现蓝幕功能, 但只能在视频中确定和跟踪一个用户。这个 API 中用到了多种图像处理技术来提高和保障每一帧深度视频帧中原始用户码的稳定性和准确性。对于任何角度的人, 该 API 都能较好地识别并去除背景像素。

使用该方法, 不仅要用深度数据和彩色数据, 同时还要用到骨骼数据来跟踪确定同一个用户。因此, 之前的两种方法只需打开 Kinect 的彩色视频流和深度视频流, 而这个方法还需打开骨骼视频流和特定的背景移除彩色视频流。此外, 这个方法实现前景提取的方法与上述的第二种方法类似, 即利用 Alpha 通道把背景改为全透明, 从而分离出前景。但是, 在边缘处, 该方法没有直接把透明度从 255 跳跃到 0, 而是一个渐变的过程, 这样处理就使得边缘不会出现像之前两种方法中有锯齿、不平滑等状况。

图 3 所示为基于 Background Removal API 方法提取的前景图像。从以上结果来看, 用这种方法提取的前景效果是非常好的。



图 3 基于 Background Removal API 前景提取的结果视频帧

3.5 对 3 种提取方法的比较

从实验结果可知, 前两种视频帧前景的提取方法均没有很好地达到本文预期的标准, 提取出来的前景图像噪声都比较大, 而且还存在边界不连续、不稳定、有比较明显的锯齿等情况, 某些前景区域甚至还存在一些大型的像素空洞。这些现象主要是由 Kinect 本身深度图像成像质量较低造成的, 而且视频摄制现场的环境同样也会干扰成像质量。Kinect 作为一种廉价的深度图像获取设备, 深度图像成像质量不高, 深度图像中的深度值会溢出物体的边界, 它产生的随机误差和精度也会随跟踪目标和摄像头之间距离的变化而发生改变。当两者距离较远时, 目标的深度值会随时间而随机地跳变^[15]。而这些缺陷都要靠一些图像修复算法和去噪算法来进行修复。就如文献^[17]提出的基于改进双边滤波法的深度图像修复算法, 它先将深度图像进行二值化处理, 再利用直方图确定阈值生成空洞 Mask 掩膜, 然后用改进后的双边滤波算法进行深度填充, 最后用中值滤波平滑图像, 整个过程比较繁琐。

而对于第三种提取方法, 其提取效果明显比前两种方法要好得多, 并且后期有对彩色前景视频帧的优化处理, 比深度图像的修复方法简单得多。考虑到实验的效率和进度以及修补后的深度图或多或少都会存在误差和失真的情况而不利于保证时域连续性, 最后决定采用第三种提取方法, 即使用微软发布第三方工具包 Developer Toolkit 中与背景移除相关的 API(应用程序编程接口)来实现视频前景的提取。

4 基于纹理传输的视频风格化

为了使本文系统最终能得到多种不同风格的渲染效果, 我们决定采取基于纹理传输的视频风格化技术, 本文所用方法与文献^[7]和文献^[18]类似。同时, 本文在纹理层是利用方向场实现各向异性的纹理合成, 对于一些明显具有方向性纹

理的视频具有较好的渲染效果。另外,为了保证视频的时域连续性,本文将通过计算原视频中每两帧之间的光流场,使纹理层随着光流场传输,让纹理层与原视频融合起来完成最终的风格化。但是,纹理层经过传输后某些区域还是会发生拉伸走样的现象,本文将采取局部纹理修补技术来对这些区域进行快速修补,从而进一步保证视频的时域连续性。

因此,本文的风格化系统主要划分为三大步骤:首先利用方向场完成对第一帧的纹理层的合成;接着利用光流场进行纹理层的传输,并对纹理传输层进行修补以保证视频的时域连续性;最后完成视频帧与纹理层帧的融合,从而得到最终的风格化视频帧。

4.1 基于方向场的纹理传输层合成

4.1.1 方向场计算

很多视频中都有很明显的物体轮廓信息、边缘信息、走向信息等,这些都是视频携带的反映视频内容中非常重要的信息。而评判一个风格化算法的好坏,除了要看它最终能达到的艺术效果,还要看它是否能尽量多地保留视频的原始信息。我们希望本文的风格化系统能够具有这样的能力,从而使风格化后的视频帧仍然保持高质量。

因此,采用与文献[19]相类似的方法,即使用方向场进行第一帧视频纹理层的合成,另外,利用边缘切向场(ETF)[20]来进行方向场的生成,该方向场可以指导样本纹理合成一个各向异性的纹理层。这可以帮助我们实现带方向各向异性的纹理层的合成,能够保留较多的视频中的重要信息,如轮廓走向等。

4.1.2 纹理层合成过程

上述方向场的计算是纹理层合成过程中的关键技术,有了这项技术,对纹理层的合成就会轻松许多,效果也能得到保证。在本系统中,把流场定义为 24 个方向,每个方向 15 度。故对应地,把输入的纹理样本(像素为 64×64 的灰度图)也旋转成 24 个方向,以供之后纹理层的合成使用。图 4 示出一组 24 方向油画风格的纹理样本。

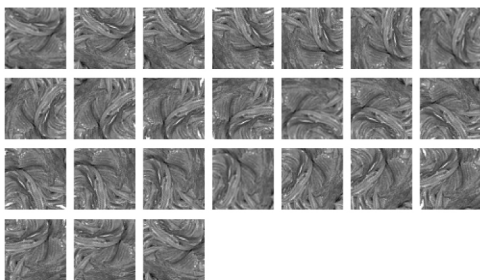


图 4 24 个方向的纹理样本

接下来阐述完整的第一帧纹理层的合成过程。

1) 设置系统参数,同时确定合成窗口的尺寸及邻域大小,并计算视频帧方向场。

2) 根据起初设置的合成窗口的大小,在对应帧的方向场内按由左至右、由上至下的顺序进行全图扫描;通过计算待合成窗口内像素点的平均方向场来确定该窗口的整体方向。

3) 得到待合成窗口的最终方向后,利用基于块的纹理合成的方法[21]在 24 个纹理候选样本中选择对应的纹理样本来完成最后的纹理层的合成。

对于一些方向感明显的视频,用这种基于方向场的风格化技术渲染会得到非常好的效果。图 5 示出了对应的线条感纹理样本合成的纹理传输层。但对于一些方向感不明显的视频,可直接采用基于块的纹理合成方法。

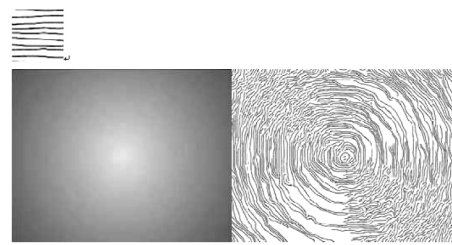


图 5 基于线条纹理样本合成的纹理传输层

4.2 基于光流场的纹理传输与修补

4.2.1 基于光流场的纹理层传输

光流场(flow field)可以追踪视频中每个像素的运动轨迹,这样纹理层就可以跟着这些运动轨迹进行逐帧传输,从而纹理层可以附着在像素点上,使之后的画面看起来更加自然。如图 6 所示,连续相邻的两帧生成一帧对应的光流图,图中的光流信息即表示两帧之间对应像素点的运动轨迹。



图 6 连续视频帧及其对应的光流图

采用文献[22]所提出的计算光流场的方法,该方法结合全局方法以及局部方法,利用共轭梯度求解大型线性系统,能够快速取得稠密光流场。利用计算得到的光流场,就可以很容易地取得视频帧中任何像素点在任意时刻的运动向量 $V(X_t)$, $X_t = (x_t, y_t)$ 表示 t 时刻 X 像素点在视频帧中的位置,则像素 X 下一时刻的位置为 $X_{t+1} = X_t + V(X, t)$ 。于是,可以使用以下公式来完成纹理传输:

$$C_{t+1}(X_{t+1}) = C_t(X_t) = S(U_t(u_t, v_t))$$

其中, C_t 表示像素点 X_t 的纹理颜色, $U_t(u_t, v_t)$ 表示像素点 X_t 的纹理坐标, S 表示纹理样本。这样,就可以利用光流场提供的像素点的运动向量,通过计算 $C_{t+1}(X_{t+1})$,把当前帧的纹理传递到下一帧,如此重复来完成全部帧的传输。本文在 GPU 的 CUDA 框架下完成了纹理层传输的计算工作。

4.2.2 纹理层修补

在光流场的指导下进行纹理层的传输。随着时间的变化,纹理层的某些区域会出现拉伸走样的问题,从而破坏视频的时域连续性。我们希望通过利用相关方法来解决这个问题,文献[23]提出的基于纹理合成的纹理修补方法给了我们一个很好的借鉴。为了保证视频的连续性以及最后的修复效果,可以根据以下 3 个步骤来完成纹理层的修补:1) 定位走样的纹理像素点;2) 选取合适的纹理修补的邻域;3) 确定局部纹理的合成顺序。

对走样纹理点的定位是非常容易的,只需在每帧纹理传输完成后,即纹理坐标传递映射结束后,在纹理帧每一个像素点的 3×3 邻域内查找是否存在相同的坐标值,就能判断此像素点是否发生了拉伸走样。用一张 mask 表来记录其位置。

对于纹理修补邻域的选取,为了能得到好的修复效果,不能像纹理合成时那样直接选取一个由左至右、由上至下的邻域来进行修补。本文先判断纹理走样的方向,再根据其走样方向选取适当的邻域来进行纹理合成,以较好地修复拉伸走样的区域。

选择好纹理修补的邻域后,根据选择的邻域类型来确定局部纹理的合成顺序。对于一些使用特殊邻域来进行纹理修

补的情况,借助堆栈等数据结构,在不改变全局合成顺序的情况下改变局部合成顺序,使用已合成区域作为约束邻域,最终可以完成对不同拉伸方向纹理的有效修复。

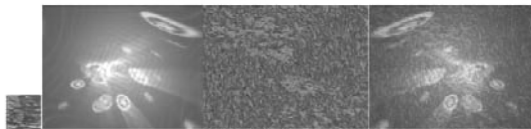
4.3 视频帧与纹理层融合

利用光流场对纹理层进行传输,通过纹理层修补保证视频的时域连续性,从而得到了每一帧的纹理图。最后,需要把视频帧与对应的纹理层一一进行融合,这样原视频不仅有了样本纹理的特征,同时还保留了原视频的视觉信息。

为了完成对应帧之间最终的融合,把原视频帧和纹理层颜色空间由 RGB 空间转换到 YIQ 空间,然后对每一原视频帧的 Y 分量进行重新映射。公式如下:

$$Y(t') = k \times (Y(s) - u_s) + Y(t)$$

其中, $Y(t')$ 为最终风格化视频帧的 Y 分量; $Y(s)$ 为纹理层帧的 Y 分量; $Y(t)$ 为对应的原视频帧的 Y 分量; u_s 表示纹理层帧 Y 分量的均值; k 表示原视频帧 Y 分量的方差/纹理层帧 Y 分量的方差。图 7 分别示出了传入的样本纹理、原视频帧、纹理层合成帧以及最后融合后的风格化帧。



从左至右分别为传入的样本纹理、原视频帧、纹理层合成帧以及最后融合后的风格化帧

图 7

5 系统测试与优化

通过使用高效的并且能渲染出极强艺术效果的视频风格化技术以及 Kinect for Windows SDK,借助微软发布的 Kinect for XBOX360 深度摄像机,在 Microsoft Visual Studio 2010 和 MATLAB R2013a 平台上实现一个特殊的自定义视频风格化系统。本文目标是希望这个系统既可以具有较好的视频前景提取功能,又可以让视频得到较好的艺术渲染效果,同时还能保障视频的时域连续性。

5.1 基于 Kinect 的视频前景提取测试

利用第 3 节所述方法,使用 Kinect for XBOX360 完成视频的录制,并利用深度信息实时完成视频前景的提取,得到的前景视频帧如图 8 所示。



图 8 提取的视频前景帧

5.2 视频风格化系统测试

将得到的视频前景帧序列整合成视频,并利用 Matlab 计算该视频的光流场,得到每一帧光流场的 TXT 格式文件,然后再把 TXT 格式的光流图转化为对应的 JPEG 图像。如图 9 所示,连续两帧相邻视频帧计算得到一帧对应的光流图。

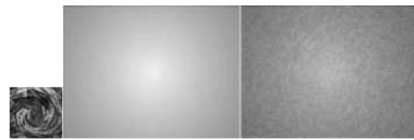


图 9 连续的两帧视频帧及对应的光流图

把光流图和原视频帧输入到风格化系统中,同时选择合适的纹理样本输入到系统中,进行基于纹理传输的视频风格化渲染,如图 10 所示。同时,自定义一个背景视频,用相同方法进行处理,得到的结果如图 11 所示。



图 10 传入的纹理样本和风格化后的前景效果帧



从左至右为传入的纹理样本、原背景视频帧和风格化后的背景效果帧

图 11

5.3 前景和背景视频帧融合测试及优化

基于以上的工作,把风格化后的前景视频帧和背景视频帧进行融合,从而得到最后的风格化效果帧。值得注意的是,这里利用 Kinect 提取的原始视频前景帧来完成最后的融合,因为原始视频前景帧是 4 通道图,并且风格化后的前景视频帧与原始前景视频帧的像素点是一一对应的,所以可以根据原始视频帧中像素点的 Alpha 值来判断在风格化后的视频帧中对应像素点是否属于前景像素点,从而决定是否要在最终的风格化视频帧中显示出来。

前面已介绍,Background Removal API 在进行背景移除的过程中,在处理用户边缘问题时,其透明度会有一个 255 到 0 的过渡,而不是直接跳跃式的处理,就是出于把边缘处理得更加平滑的考虑。因此在进行两帧的融合过程中,如果仅仅考虑把透明度为 255 即完全不透明的像素点归为前景点,则得到的融合效果如图 12 所示。如果把透明度为 0 的像素点当作背景,其他像素点都归为前景,则得到的融合效果如图 13 所示。



图 12 把透明度为 255 的像素点归为前景像素点融合后的效果帧 (方法一)



图 13 把透明度非 0 的像素点归为前景像素点融合后的效果帧 (方法二)

从图 12 可见,用户对对象明显被“削弱”了,用户边缘很多像素点被当作了背景像素从而未得到显示,这是一种非常不好的融合效果。从图 13 来看,很明显在用户的边缘存在一层亮边,这些点其实是背景像素点,但方法二把它们均归类为了前景像素点。相对于方法一,虽然方法二可以把用户对象完

整地保护下来,但它依然有其他像素点干扰,从而降低了融合质量。本文需要把这一层亮边去掉。

结合原 4 通道图的特性,决定把原视频帧的 Alpha 值作为一个融合的权重,采取如下函数遍历全图来完成两视频帧的融合:

$$C=B \times (255-x) / 255+A \times x / 255$$

其中, C 代表融合后视频帧中的某一个像素点, B 代表风格化后背景视频帧中对应的像素点, A 代表风格化后前景视频帧中对应的像素点, x 代表原前景视频帧中对应像素点的 Alpha 值。这样,得到的融合效果如图 14 所示。



图 14 透明度权重方法融合后的效果帧(方法三)

可见,使用这种方法能很好地保留前景视频帧中用户的像素点,同时又不会有过多的背景像素点进入,而且通过 Alpha 作为融合权重,反而使得前景帧与背景帧边缘处有了一个平滑的过渡,使融合效果更加自然,提高了融合后视频的质量。

最后,只需将融合后的视频帧转化为视频文件输出,就可以得到最终的风格化艺术视频。

使用同样的方法,利用不同的纹理样本又得到了一组样例,如图 15—图 17 所示。这说明,可以选择不同的纹理样本进行传输,从而获得不同的视频风格化效果。

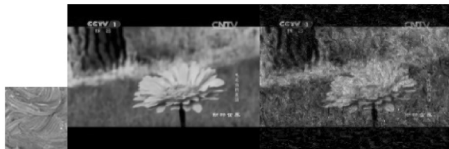


图 15 传入的纹理样本以及渲染前后的背景帧



图 16 传入的前景帧不做任何风格化处理



图 17 前背景视频帧融合后的视频帧

6 总结与展望

6.1 完成的工作及不足

本文完成的主要工作包括:借助微软发布 Kinect for XBOX360 体感摄像机,利用其提取的视频深度信息,实时完成视频前景的提取;另外,利用基于光流场的视频风格化技术,通过纹理传输的方法对前景和背景视频实现多风格的艺术化渲染;最后把风格化后的前景视频和背景视频融合成具

有最终艺术效果的风格化视频。最终,我们实现了这个系统,该系统具有一定的稳定性,可以较好地实现视频前景提取工作,也能通过传输不同的纹理完成多风格自定义的视频艺术化渲染。该系统使最后得到的风格化视频具有了较高的艺术感染力。

但本系统也存在一些不足:1)整个处理过程比较繁琐,步骤比较复杂;2)在 Kinect 实现视频前景的提取过程中,环境因素对前景视频提取的质量会产生一定的影响,从而影响最后风格化质量;3)在风格化阶段,计算光流场耗时过长,每一帧的转化大概需要 9s 的时间,同样,系统在纹理传输层的合成和修补耗时也比较长;4)利用纹理修补的方法来解决纹理层随传输过程出现拉伸走样问题,但处理效果依然不是非常完美,仍然会出现块状纹理的跟随运动,从而降低了视频风格化的质量。

6.2 未来的工作

为了进一步完善系统,可以利用各工程项目之间的时序性,把系统进行再一次的有效整合和精简,使其更加简单便捷。

另外,在本次实验中利用微软发布的 Developer Toolkit 里的相关 API 完成前景视频的提取,以后也可以自己开发一个相关的提取方法,通过各种图像处理技术,能够有效地增强深度图像的质量,削弱环境因素对实验的影响,从而保证最后的前景提取质量。另外,在视频风格化阶段,由于纹理合成的邻域依赖关系以及光流传输的延续性,相关计算只能在 CPU 端进行,以致处理速度缓慢,以后可以改进纹理合成算法,实现通过 GPU 加速并行纹理合成,这将大大提高系统的风格化渲染的效率。为了保障系统对任意视频的风格化都可以得到较好的艺术效果,以后要在纹理修补这一部分进行优化升级,采用更好的纹理修补技术。

参考文献

- [1] 严利民,李跃.基于 Kinect 深度数据的人体骨架提取[J].电子测量技术,2015,38(3):39-42.
- [2] 李俊龙,章登义,黄珺.Kinect 驱动的人脸动画合成技术研究[J].计算机工程,2015,41(3):237-241.
- [3] 刘开余,夏斌.基于 Kinect 的实时人体姿势识别[J].电子设计工程,2014,22(19):31-34.
- [4] 邹星星,钟莎,马增强.基于边缘检测的 Kinect 深度图像去噪算法[J].湖南工业大学学报,2013,27(6):36-39.
- [5] 刘金荣.深度图增强算法研究及其在自由视点视频中的应用[D].湘潭:湘潭大学,2013.
- [6] 陈理.Kinect 深度图像增强算法研究[D].长沙:湖南大学,2013.
- [7] TANG Y,ZHANG Y,SHI X Y,et al. Multi-style video stylization based on texture advection[J]. Science China Information Sciences,2015,58(11):1-13.
- [8] 黄华,臧彧,张磊.图像和视频油画风格化研究[J].计算机科学,2011,38(6):1-6.
- [9] LIN L,ZENG K,LV H,et al. Painterly animation using video semantics and feature correspondence[C]//Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering. ACM,2010:73-80.
- [10] 郑松.视频抠图算法及其风格化绘制研究与实现[D].昆明:云南大学,2014.

- entropy[J]. *Expert Systems with Applications*, 2014, 41(7): 3538-3560.
- [16] RAJINIKANTH V, COUCEIRO M S. RGB Histogram Based Color Image Segmentation Using Firefly Algorithm[J]. *Procedia Computer Science*, 2015, 46(2015): 1449-1457.
- [17] SANDIP D, SIDDHARTHA B, UJJWAL M. New quantum inspired meta-heuristic techniques for multi-level colour image thresholding[J]. *Applied Soft Computing*, 2016, 46: 677-702.
- [18] PARE S, KUMAR A, BAJAJ V, et al. A multilevel color image segmentation technique based on cuckoo search algorithm and energy curve[J]. *Applied Soft Computing*, 2016, 47: 76-102.
- [19] BHANDARI A K, KUMAR A, CHAUDARY S, et al. A novel color image multilevel thresholding based segmentation using nature inspired optimization algorithms[J]. *Expert Systems with Applications*, 2016, 63(C): 112-133.
- [20] KUMAR S, PANT M, KUMAR M, et al. Colour image segmentation with histogram and homogeneity histogram difference using evolutionary algorithms[J]. *Int. J. Mach. Learn. & Cyber*, 2015: 1-21.
- [21] KRISHNANAND K N, GHOSE D. Detection of multiple source locations using a glowworm metaphor with applications to collective robotics [C] // *IEEE Swarm Intelligence Symposium, USA*, 2005: 84-91.
- [22] KRISHNANAND K N, GHOSE D. Theoretical foundations for multiple rendezvous of glowworm-inspired mobile agents with variable local-decision domains [C] // *American Control Conference*, 2006: 14-16.
- [23] KRISHNANAND K N, GHOSE D. A glowworm swarm optimization based multi-robot system for signal source localization [J]. *Studies in Computational Intelligence*, 2009, 177(177): 49-68.
- [24] KRISHNANAND KN, GHOSE D. Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions[J]. *Swarm Intelligence*, 2009, 3(2): 87-124.
- [25] KRISHNANAND K N, GHOSE D. Glowworm swarm based optimization algorithm for multimodal functions with collective robotics applications[J]. *Multagent and Grid Systems*, 2006, 2(3): 209-222.
- [26] KRISHNANAND K N. Glowworm swarm optimization: a multimodal function optimization paradigm with applications to multiple signal source localization tasks[D]. Indian; Department of Aerospace Engineering, Indian Institute of Science, 2007.
- [27] LIAO W H, KAO Y C, LI Y S. A sensor deployment approach using glowworm swarm optimization algorithm in wireless sensor networks[J]. *Expert Systems with Applications*, 2011, 38(10): 12180-12188.
- [28] HORNG M H. Multilevel Image Thresholding with Glowworm Swam Optimization Algorithm based on the Minimum Cross Entropy [J]. *Advances in Information Sciences and Service Sciences*, 2013, 5(10): 1290-1298.
- [29] LUO Q, OUYANG Z, CHEN X, et al. A multilevel threshold image segmentation algorithm based on glowworm swarm optimization[J]. *J. Comput. Inf. Syst*, 2014, 10(4): 1621-1628.
- [30] GAO H, XU W, SUN J, et al. Multilevel thresholding for image segmentation through an improved quantum-behaved particle swarm algorithm[J]. *IEEE Transactions on Instrumentation and Measurement*, 2010, 59(4): 934-946.
- [31] SATHYA P D, KAYALVIZHI R. Modified bacterial foraging algorithm based multilevel thresholding for image segmentation [J]. *Engineering Applications of Artificial Intelligence*, 2011, 24(4): 595-615.
- [32] KRISHNANAND K N, GHOSE D. Glowworm swarm optimization: A new method for optimising multi-modal functions[J]. *The International Journal of Computational Intelligence Studies*, 2009, 1(1): 93-119.
- [33] 陈贵敏, 贾建援, 韩琪. 粒子群优化算法的惯性权值递减策略研究[J]. *西安交通大学学报*, 2006, 40(1): 53-56.
- (上接第 197 页)
- [11] SEO S H, OSTROMOUKHOV V. Pointillist video stylization based on particle tracing[J]. *Multimedia Tools and Applications*, 2014, 71(1): 279-292.
- [12] GUO Y, YU J, XU X, et al. Example based painting generation [J]. *Journal of Zhejiang University-Science A*, 2006, 7(7): 1152-1159.
- [13] 陈佳舟. 增强现实和图像风格化中的图像结构分析算法研究[D]. 杭州: 浙江大学, 2012.
- [14] BOUSSEAU A, KAPLAN M, THOLLOT J, et al. Interactive watercolor rendering with temporal coherence and abstraction [C] // *Proceedings of the 4th International Symposium on Non-photorealistic animation and Rendering*. ACM, 2006: 141-149.
- [15] 赵旭. Kinect 深度图像修复技术研究[D]. 大连: 大连理工大学, 2013.
- [16] 余涛. Kinect 应用开发实战: 用最自然的方式与机器对话[M]. 北京: 机械工业出版社, 2012.
- [17] 李应彬, 冯杰, 张华熊, 等. 基于改进双边滤波的 Kinect 深度图像空洞修复算法研究[J]. *工业控制计算机*, 2013, 26(11): 105-106.
- [18] CAO C, CHEN S, ZHANG W, et al. Automatic motion-guided video stylization and personalization [C] // *Proceedings of the 19th ACM International Conference on Multimedia*. ACM, 2011: 1041-1044.
- [19] WANG B, WANG W, YANG H, et al. Efficient example-based painting and synthesis of 2d directional texture[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2004, 10(3): 266-277.
- [20] KANG H, LEE S, CHUI C K. Flow-based image abstraction [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(1): 62-76.
- [21] EFROS A A, FREEMAN W T. Image quilting for texture synthesis and transfer [C] // *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 2001: 341-346.
- [22] LIU C. Beyond pixels: exploring new representations and applications for motion analysis [D]. Massachusetts Institute of Technology, 2009.
- [23] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by example-based image inpainting [J]. *IEEE Transactions on Image Processing*, 2004, 13(9): 1200-1212.