



(12)发明专利申请

(10)申请公布号 CN 107067452 A

(43)申请公布日 2017. 08. 18

(21)申请号 201710089850.7

(22)申请日 2017.02.20

(71)申请人 同济大学

地址 200092 上海市杨浦区四平路1239号

(72)发明人 尤鸣宇 朱江 沈春华

(74)专利代理机构 上海科盛知识产权代理有限公司 31225

代理人 宣慧兰

(51)Int.Cl.

G06T 15/00(2011.01)

G06N 3/08(2006.01)

G06T 7/50(2017.01)

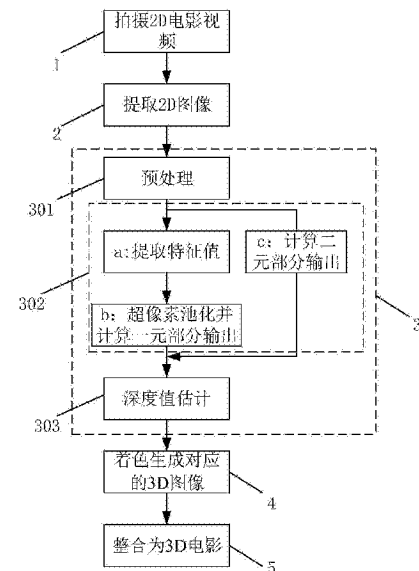
权利要求书2页 说明书4页 附图1页

(54)发明名称

一种基于全卷积神经网络的电影2D转3D方法

(57)摘要

本发明涉及一种基于全卷积神经网络的电影2D转3D方法,该方法包括如下步骤:(1)提取拍摄的电影2D视频中的每一帧2D图像;(2)对每一帧2D图像分别采用全卷积神经网络提取特征值并计算深度值;(3)对每一帧2D图像分别根据其深度值进行着色生成对应的3D图像;(4)将所有3D图像依次整合形成3D电影。与现有技术相比,本发明场景适应性强,深度估计质量高,计算快速,能够高效地用于各种场景下的电影2D转3D。



1. 一种基于全卷积神经网络的电影2D转3D方法,其特征在于,该方法包括如下步骤:

- (1) 提取拍摄的电影2D视频中的每一帧2D图像;
- (2) 对每一帧2D图像分别采用全卷积神经网络提取特征值并计算深度值;
- (3) 对每一帧2D图像分别根据其深度值进行着色生成对应的3D图像;
- (4) 将所有3D图像依次整合形成3D电影。

2. 根据权利要求1所述的一种基于全卷积神经网络的电影2D转3D方法,其特征在于,所述的步骤(2)包括如下子步骤:

- (201) 将2D图像分隔成超像素,形成超像素图像;
- (202) 对超像素图像进行处理和计算得到所述的2D图像的一元部分输出和二元部分输出;
- (203) 根据2D图像的一元部分输出和二元部分输出进行深度估计得到深度值。

3. 根据权利要求2所述的一种基于全卷积神经网络的电影2D转3D方法,其特征在于,步骤(202)中具体为:

- (a) 将超像素图像输入至全卷积神经网络,得到多个卷积特征图;
- (b) 将卷积特征图进行超像素池化,进而得到该2D图像的一元部分输出;
- (c) 对超像素图像进行计算得到二元部分输出。

4. 根据权利要求3所述的一种基于全卷积神经网络的电影2D转3D方法,其特征在于,所述的全卷积神经网络包括多个依次级联的特征提取基本单元,每个特征提取单元包括依次连接的卷积层和池化层。

5. 根据权利要求3所述的一种基于全卷积神经网络的电影2D转3D方法,其特征在于,步骤(b)中超像素池化具体操作为:首先在卷积特征图上进行上采样操作,使得卷积特征图大小与原始2D图像大小相同,然后,对经过上采样操作后的卷积特征图中的每个超像素区域分别计算特征值的平均值,得到n维特征向量,n为超像素区域的个数,最后,将n维特征向量输入到一元部分计算单元进行计算得到一元部分输出矩阵Z,Z为n维行向量,向量中第p个元素记作 Z_p , Z_p 表示第p个超像素区域的一元部分计算值, $p=1,2,\dots,n$ 。

6. 根据权利要求5所述的一种基于全卷积神经网络的电影2D转3D方法,其特征在于,步骤(c)具体为:

对于第i个超像素区域,考虑所有与其相邻的所有超像素区域,设其中一个与第i个超像素区域相邻的超像素区域记作第j个超像素区域,分别计算第i个超像素区域和第j个超像素区域之间的K种相似度,得到相似度向量 $[S_j^{(1)}, S_j^{(2)}, \dots, S_j^{(K)}]^T$,其中 $i=1,2,\dots,n$,对每个超像素区域进行上述操作得到相似度向量,最后将所有相似度向量输入到二元部分计算单元计算得到二元部分输出矩阵R,R为 $n*n$ 维矩阵,其第p行第q列元素记作 R_{pq} , R_{pq} 表示第p个超像素区域相对于第q个超像素区域的二元部分计算值, $p=1,2,\dots,n,q=1,2,\dots,n,n$ 为超像素区域的个数。

7. 根据权利要求6所述的一种基于全卷积神经网络的电影2D转3D方法,其特征在于,所述的一元部分计算单元和二元部分计算单元均包括多个依次连接的全连接层。

8. 根据权利要求6所述的一种基于全卷积神经网络的电影2D转3D方法,其特征在于,步骤(203)具体为:

根据下式求取深度值：

$$y = (I + D - R)^{-1}Z,$$

其中 y 为深度值矩阵， I 为 $n \times n$ 维单位阵， D 为 $n \times n$ 对角矩阵，其对角元素记作 D_{pp} ，

$$D_{pp} = \sum_{q=1}^n R_{pq}, p=1, 2, \dots, n, y \text{ 为 } n \text{ 维行向量, } y \text{ 中的第 } i \text{ 个元素记作 } y_i, y_i \text{ 表示第 } i \text{ 个超像素区}$$

域的深度值， $i=1, 2, \dots, n, n$ 为超像素区域的个数。

9. 根据权利要求3所述的一种基于全卷积神经网络的电影2D转3D方法，其特征在于，所述的全卷积神经网络为针对不同场景下的全卷积神经网络，所述的不同场景包括室内、室外、白天和夜晚，在进行2D图像转3D图像时，根据该图像选择对应场景下的全卷积神经网络。

一种基于全卷积神经网络的电影2D转3D方法

技术领域

[0001] 本发明涉及一种电影2D转3D方法,尤其是涉及一种基于全卷积神经网络的电影2D转3D方法。

背景技术

[0002] 3D视频是传统2D视频的扩展,通过增加图像深度信息,用户能够体验到富有立体感和临场感的视频内容。三维显示技术已经成为当前社会的热点技术,被应用在生活的各种场景中,具有重要的现实意义。3D内容贫乏及其高昂的拍摄成本是制约当前3D视频发展的主要因素。现有3D电影拍摄时需将两台摄影机架在一具可调角度的特制云台上,并以特定的夹角来拍摄。这对摄影师拍摄的要求较高,且工作量成倍增加,需要后期进行双素材的平衡处理,导致3D成本高而资源少。单幅图像2D转3D方法能够在较短时间内以较低成本将现有2D视频资源转换为可立体显示的3D视频,从而解决上述的问题。

[0003] 2D转3D方法的核心任务是从2D图像中提取景物的深度信息。传统深度图像信息输入的应用中,通常以双视差图像作为输入,输出深度信息估计结果,但此类应用难以实现视频2D转3D的应用需求。所以通常需要通过输入单视差图片来进行立体显示,而目前的单目图像深度估计的方法大多基于几何模型或其他的视觉线索,使用手工标记的特征,这些都不适用于2D电影转3D电影的应用。

发明内容

[0004] 本发明的目的就是为了解决上述现有技术存在的缺陷而提供一种基于全卷积神经网络的电影2D转3D方法。

[0005] 本发明的目的可以通过以下技术方案来实现:

[0006] 一种基于全卷积神经网络的电影2D转3D方法,该方法包括如下步骤:

[0007] (1) 提取拍摄的电影2D视频中的每一帧2D图像;

[0008] (2) 对每一帧2D图像分别采用全卷积神经网络提取特征值并计算深度值;

[0009] (3) 对每一帧2D图像分别根据其深度值进行着色生成对应的3D图像;

[0010] (4) 将所有3D图像依次整合形成3D电影。

[0011] 所述的步骤(2)包括如下子步骤:

[0012] (201) 将2D图像分隔成超像素,形成超像素图像;

[0013] (202) 对超像素图像进行处理和计算得到所述的2D图像的一元部分输出和二元部分输出;

[0014] (203) 根据2D图像的一元部分输出和二元部分输出进行深度估计得到深度值。

[0015] 步骤(202)中具体为:

[0016] (a) 将超像素图像输入至全卷积神经网络,得到多个卷积特征图;

[0017] (b) 将卷积特征图进行超像素池化,进而得到该2D图像的一元部分输出;

[0018] (c) 对超像素图像进行计算得到二元部分输出。

[0019] 所述的全卷积神经网络包括多个依次级联的特征提取基本单元,每个特征提取单元包括依次连接的卷积层和池化层。

[0020] 步骤(b)中超像素池化具体操作为:首先在卷积特征图上进行上采样操作,使得卷积特征图大小与原始2D图像大小相同,然后,对经过上采样操作后的卷积特征图中的每个超像素区域分别计算特征值的平均值,得到n维特征向量,n为超像素区域的个数,最后,将n维特征向量输入到一元部分计算单元进行计算得到一元部分输出矩阵Z,Z为n维行向量,向量中第p个元素记作 Z_p , Z_p 表示第p个超像素区域的一元部分计算值, $p=1,2,\dots,n$ 。

[0021] 步骤(c)具体为:

[0022] 对于第i个超像素区域,考虑所有与其相邻的所有超像素区域,设其中一个与第i个超像素区域相邻的超像素区域记作第j个超像素区域,分别计算第i个超像素区域和第j个超像素区域之间的K种相似度,得到相似度向量 $[S_{ij}^{(1)}, S_{ij}^{(2)}, \dots, S_{ij}^{(K)}]^T$,其中 $i=1,2,\dots,n$,对每个超像素区域进行上述操作得到相似度向量,最后将所有相似度向量输入到二元部分计算单元计算得到二元部分输出矩阵R,R为 $n*n$ 维矩阵,其第p行第q列元素记作 R_{pq} , R_{pq} 表示第p个超像素区域相对于第q个超像素区域的二元部分计算值, $p=1,2,\dots,n,q=1,2,\dots,n,n$ 为超像素区域的个数。

[0023] 所述的一元部分计算单元和二元部分计算单元均包括多个依次连接的全连接层。

[0024] 步骤(203)具体为:

[0025] 根据下式求取深度值:

$$[0026] \quad y = (I+D-R)^{-1}Z,$$

[0027] 其中y为深度值矩阵,I为 $n*n$ 维单位阵,D为 $n*n$ 对角矩阵,其对角元素记作 D_{pp} ,

$D_{pp} = \sum_{q=1}^n R_{pq}$, $p=1,2,\dots,n$,y为n维行向量,y中的第i个元素记作 y_i , y_i 表示第i个超像素区域的深度值, $i=1,2,\dots,n,n$ 为超像素区域的个数。

[0028] 所述的全卷积神经网络为针对不同场景下的全卷积神经网络,所述的不同场景包括室内、室外、白天和夜晚,在进行2D图像转3D图像时,根据该图像选择对应场景下的全卷积神经网络。

[0029] 与现有技术相比,本发明具有如下优点:

[0030] (1) 本发明不需要额外的几何视觉线索,仅用单幅图像即可完成2D转3D,从而实现整个电影2D视频的转化,操作简单方便;

[0031] (2) 本发明将单幅2D图像的一元部分输出矩阵和二元部分输出矩阵相结合,从而计算每个超像素区域的深度值,使得图像深度估计结果更加准确,进而提高3D效果;

[0032] (3) 本发明中采用全卷积神经网络和超像素池化,极大提升了电影2D转3D的速度;

[0033] (4) 本发明针对不同场景下设置不同的全卷积神经网络,从而提高3D效果,提升用户的观赏体验。

附图说明

[0034] 图1为本发明基于全卷积神经网络的电影2D转3D方法的流程框图。

具体实施方式

[0035] 下面结合附图和具体实施例对本发明进行详细说明。

[0036] 实施例

[0037] 如图1所示,一种基于全卷积神经网络的电影2D转3D方法,该方法包括如下步骤:

[0038] (1) 拍摄拍摄2D视频:普通的单摄影机拍摄即可;

[0039] (2) 提取拍摄的电影2D视频中的每一帧2D图像;

[0040] (3) 对每一帧2D图像分别采用全卷积神经网络提取特征值并计算深度值;

[0041] (4) 对每一帧2D图像分别根据其深度值进行着色生成对应的3D图像;

[0042] (5) 将所有3D图像依次整合形成3D电影。

[0043] 步骤(3)包括如下子步骤:

[0044] (301) 预处理:将2D图像分隔成超像素,形成超像素图像,超像素是由一系列位置相邻且颜色、亮度、纹理等特征相似的像素点组成的小区域。这些小区域大多保留了进一步进行图像分割的有效信息,且一般不会破坏图像中物体的边界信息;

[0045] (302) 对超像素图像进行处理和计算得到2D图像的一元部分输出和二元部分输出;

[0046] (303) 根据2D图像的一元部分输出和二元部分输出进行深度估计得到深度值。

[0047] 步骤(302)中具体为:

[0048] (a) 将超像素图像输入至全卷积神经网络,得到多个卷积特征图,卷积层通过多个卷积核对图像逐行逐列做卷积,可以学习到图像的特征;池化层是在卷积特征提取的基础上,对每个 2×2 区域的卷积特征进行取平均等。多种卷积核提取出来一个区域的多种特征,这样全卷积神经网络的输出为多通道的卷积特征图;

[0049] (b) 将卷积特征图进行超像素池化,进而得到该2D图像的一元部分输出;

[0050] (c) 对超像素图像进行计算得到二元部分输出。

[0051] 全卷积神经网络包括多个依次级联的特征提取基本单元,每个特征提取单元包括依次连接的卷积层和池化层。

[0052] 步骤(b)中超像素池化具体操作为:首先在卷积特征图上进行上采样操作,使得卷积特征图大小与原始2D图像大小相同,然后,对经过上采样操作后的卷积特征图中的每个超像素区域分别计算特征值的平均值,得到 n 维特征向量, n 为超像素区域的个数,最后,将 n 维特征向量输入到一元部分计算单元进行计算得到一元部分输出矩阵 Z , Z 为 n 维行向量,向量中第 p 个元素记作 Z_p , Z_p 表示第 p 个超像素区域的一元部分计算值, $p=1,2,\dots,n$ 。

[0053] 步骤(c)具体为:

[0054] 对于第 i 个超像素区域,考虑所有与其相邻的所有超像素区域,设其中一个与第 i 个超像素区域相邻的超像素区域记作第 j 个超像素区域,分别计算第 i 个超像素区域和第 j 个超像素区域之间的 K 种相似度,得到相似度向量 $[s_{ij}^{(1)}, s_{ij}^{(2)}, \dots, s_{ij}^{(K)}]^T$,其中 $i=1,2,\dots,n$,对每个超像素区域进行上述操作得到相似度向量,最后将所有相似度向量输入到二元部分计算单元计算得到二元部分输出矩阵 R , R 为 $n \times n$ 维矩阵,其第 p 行第 q 列元素记作 R_{pq} , R_{pq} 表示第 p 个超像素区域相对于第 q 个超像素区域的二元部分计算值, $p=1,2,\dots,n$, $q=1,2,\dots,n$, n 为超像素区域的个数。

[0055] 一元部分计算单元和二元部分计算单元均包括多个依次连接的全连接层。

[0056] 步骤(303)具体为:

[0057] 根据下式求取深度值：

$$[0058] \quad y = (I + D - R)^{-1} Z,$$

[0059] 其中 y 为深度值矩阵, I 为 $n \times n$ 维单位阵, D 为 $n \times n$ 对角矩阵,其对角元素记作 D_{pp} ,

$$D_{pp} = \sum_{q=1}^n R_{pq}, p=1, 2, \dots, n, y \text{ 为 } n \text{ 维行向量, } y \text{ 中的第 } i \text{ 个元素记作 } y_i, y_i \text{ 表示第 } i \text{ 个超像素区}$$

域的深度值, $i=1, 2, \dots, n, n$ 为超像素区域的个数。

[0060] 全卷积神经网络为针对不同场景下的全卷积神经网络,不同场景包括室内、室外、白天和夜晚,在进行2D图像转3D图像时,根据该图像选择对应场景下的全卷积神经网络。上述全卷积神经网络的参数是在训练数据集上通过CRF损失层中SGD反向传播算法来训练的。针对不同场景(室内、室外、白天、夜晚等)下的数据集,可以训练得到不同的模型。选择适应场景下的模型来进行精调,而精调过程选取的是该2D电影中具有代表性、出现次数多的一些图像。

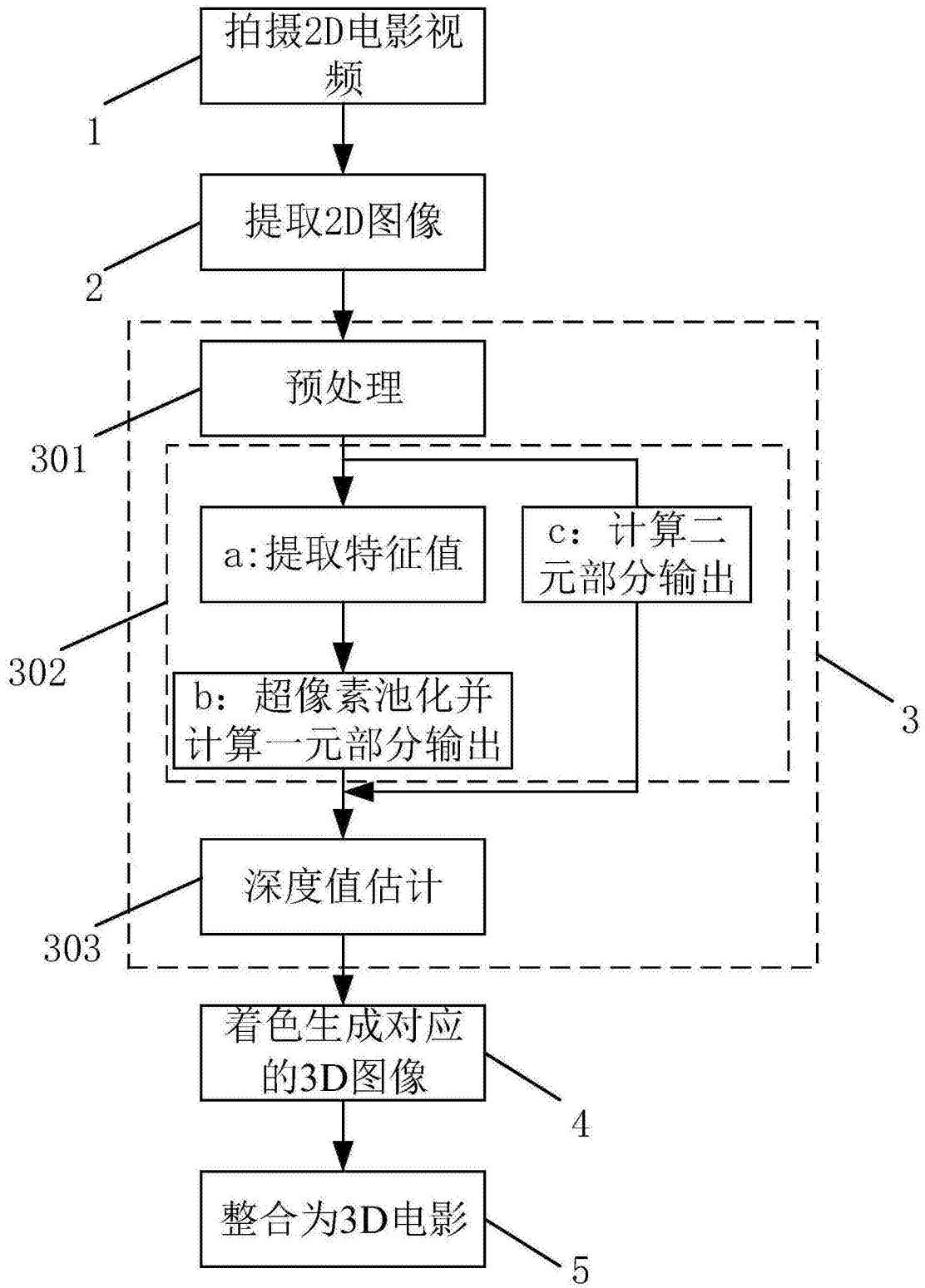


图1