



(12) 发明专利

(10) 授权公告号 CN 108846473 B

(45) 授权公告日 2022.03.01

(21) 申请号 201810317425.3

CN 107545586 A, 2018.01.05

(22) 申请日 2018.04.10

CN 105357515 A, 2016.02.24

(65) 同一申请的已公布的文献号

CN 105931190 A, 2016.09.07

申请公布号 CN 108846473 A

EP 2806396 A1, 2014.11.26

(43) 申请公布日 2018.11.20

CN 107578436 A, 2018.01.12

(73) 专利权人 杭州电子科技大学

US 2016350930 A1, 2016.12.01

地址 310027 浙江省杭州市下沙高教园区2号大街

CN 107146248 A, 2017.09.08

CN 107767413 A, 2018.03.06

(72) 发明人 周文晖 梁麟开 魏兴明 周恩慈

赵翔等. “基于交叉熵和空间分割的全局可靠性灵敏度分析”. 《航空学报》. 2018, 第39卷(第2期),

(74) 专利代理机构 杭州君度专利代理事务所

刘畅等. “基于四维光场数据的深度估计算法”. 《中北大学学报(自然科学版)》. 2016, 第37卷(第4期),

(特殊普通合伙) 33240

代理人 朱月芬

Lu Zhang等. “Crowd Counting via Scale-Adaptive Convolutional Neural Network”. 《2018 IEEE Winter Conference on Applications of Computer Vision》. 2018,

(51) Int. Cl.

G06T 7/557 (2017.01)

G06N 3/04 (2006.01)

(56) 对比文件

CN 106340041 A, 2017.01.18

CN 106651943 A, 2017.05.10

CN 104598744 A, 2015.05.06

CN 107545586 A, 2018.01.05

CN 103793911 A, 2014.05.14

CN 104966289 A, 2015.10.07

CN 106257537 A, 2016.12.28

(续)

审查员 刘倩倩

权利要求书3页 说明书9页 附图2页

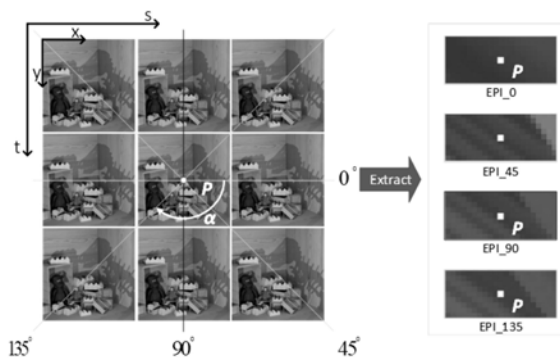
(54) 发明名称

基于方向和尺度自适应卷积神经网络的光场深度估计方法

准确的深度估计结果。

(57) 摘要

本发明公开了一种基于方向和尺度自适应卷积神经网络的光场深度估计方法。本发明包括如下步骤:步骤1. 准备光场数据集, 制作训练集和测试集; 步骤2. 搭建方向和尺度自适应的卷积神经网络SOA-EPN; 步骤3. 使用训练集训练搭建好的SOA-EPN网络; 步骤4. 使用练好的SOA-EPN网络在测试集上进行测试; 本发明借助于尺度和方向感知卷积神经网络来预测光场深度, 不仅利用了多方向, 而且很好的处理了遮挡等问题, 得到



CN 108846473 B

[接上页]

(56) 对比文件

Yaoxiang Luo等.“EPI-Patch Based Convolutional Neural Network for Depth Estimation on 4D Light Field”.

《International Conference on Neural Information Processing》.2017,

Lin L等.“A confidence-aware depth estimation method for light-field cameras based on multiple cues”.《Journal of Computational Methods in Science & Engineering》.2017,第17卷(第9期),

D.Soukup等.“Depth Estimation within a Multi-Line-Scan Light-Field framework”.

《International Symposium on Visual Computing》.2014,

Hao Sheng等.“Occlusion-aware depth estimation for light field using multi-orientation EPIs”.《Pattern Recognition》.2018,第74卷

丁伟利等.“基于先验似然的高分辨光场图像深度重建算法研究”.《光学学报》.2015,第35卷(第7期),

肖照林等.“针对光场图像的多尺度抗混叠绘制方法”.《计算机辅助设计与图形学学报》.2014,第26卷(第7期),

1. 基于方向和尺度自适应卷积神经网络的光场深度估计方法, 其特征在于包括如下步骤:

- 步骤1. 准备光场数据集, 制作训练集和测试集;
- 步骤2. 搭建方向和尺度自适应的卷积神经网络SOA-EPN;
- 步骤3. 使用训练集训练搭建好的SOA-EPN网络;
- 步骤4. 使用练好的SOA-EPN网络在测试集上进行测试;
- 步骤1具体包括下述步骤:

步骤1-1: 使用海德堡图像处理实验室提供的4D光场数据集作为实验数据集, 该数据集包含28个场景, 并提供了高精度的视差和性能评估指标; 每一个场景包含9行9列共81张大小为 512×512 的子光圈图像; 将上述28个场景分为两部分, 利用其中16个场景制作训练集, 12个场景制作测试集;

步骤1-2: 将每一个4D光场场景定义为 $LF(s, t, x, y)$, 其中 (x, y) 是图像空间坐标系, (s, t) 是角度坐标系; 对于中心子光圈图像 $LF(4, 4, x, y)$ 上的任一点P, 提取P点的 0° 、 45° 、 90° 、 135° 四个方向的EPI图像块, EPI图像块高 \times 宽 \times 通道的大小为 $9 \times 21 \times 3$, 这四个EPI图像块构成了数据集的一个样本, 输入到卷积神经网络;

所述步骤2具体实现如下:

步骤2-1: 搭建方向和尺度自适应的卷积神经网络SOA-EPN;

所述的SOA-EPN由4个尺度感知网络SAN构成; 训练集和测试集的每一个样本包含四个EPI图像块, 每一个所述的尺度感知网络SAN是以其中1个EPI图像块作为输入, 预测输出是该图像块中心点P的视差预测向量 logit_i , 其中 $i=0, 1, 2, 3$, 因此会得到4个视差预测向量, 使用一个有229个神经元的全连接层将这4个视差预测向量进行融合得到一个最终视差;

将4个尺度感知网络SAN进行权值共享, 设定视差预测范围是-4到4, 将视差预测定义为一个分类任务, 预测精度是0.035个像素, 则可分为229类;

所述尺度感知网络SAN一共18层, 分为3个子网络, 具体如下:

(1) 深度特征表征网络, 即第1到第8个卷积层:

该部分使用8个卷积核为 2×2 、步长为 1×1 的卷积层对输入EPI图像块进行深度特征表征, 对每个卷积层的输出进行批归一化处理, 再使用ReLU函数对批归一化的输出结果进行激活; 所述的8个卷积层中每个卷积层的输出特征图数量依次是(16, 32, 64, 128, 256, 384, 512, 512), 第8个卷积层输出的张量记为 x_0 , x_0 其大小为 $512 \times 1 \times 13$, 即512个大小为 1×13 的特征图;

(2) 尺度自适应选择网络, 即第9到第16个卷积层:

尺度自适应选择网络SASN包含四个尺度选择单元SSU, 每个尺度选择单元SSU由2个卷积核为 1×2 、步长为 1×1 的卷积层, 一个切片操作和一个跳跃式连接组成; 尺度自适应选择网络SASN的作用是改变SOA-EPN网络的输入EPI图像块的尺寸大小; SOA-EPN在第8个卷积层的感知域大小为 9×9 , 而 x_0 是尺寸为 9×21 的EPI图像块的深度特征, 通过改变 x_0 的尺度间接地改变输入EPI图像块的尺寸, 进而实现尺度自适应选择功能; 通过四次切片操作 $s(\cdot)$ 改变 x_0 的尺寸, 进而获得尺寸为 9×19 、 9×17 、 9×15 、 9×13 的输入EPI图像块的深度特征 x_1 、 x_2 、 x_3 、 x_4 , 尺度自适应选择功能具体来说就是在 9×21 、 9×19 、 9×17 、 9×15 、 9×13 五个尺寸中选择一个或者是它们的某种组合, 选择的策略是在训练SOA-EPN的过程学习得到的; 其中

SASN只以一张尺寸为 9×21 的EPI图像块作为输入,通过尺度选择单元SSU实现尺度自适应选择,而不是输入多张不同尺寸的EPI图像块;

第 i 个尺度选择单元的输出 h_i 定义为:

$$\begin{cases} h_i = F(h_{i-1}) + x_i \\ x_i = s(x_{i-1}) \\ h_0 = x_0 \end{cases}$$

其中, x_0 是尺寸为 9×21 的EPI图像块的深度特征, x_i 代表 x_1, x_2, x_3, x_4 ; $F(\cdot)$ 代表每一个尺度选择单元SSU的2个卷积层的卷积操作;值得注意的是, x_i 的特征图数量可能会小于 $F(h_{i-1})$ 的特征图数量,采用补0的方式使得 x_i 和 $F(h_{i-1})$ 的特征图数量相等; $s(\cdot)$ 是切片操作;

第9层和第10层是卷积核为 1×2 、步长为 1×1 的卷积层,它是第一个尺度选择单元SSU的卷积操作;对每个卷积层的输出进行批归一化处理,再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活;所述的2个卷积层的输出特征图的数量依次是(576, 576),第10个卷积层输出576个大小为 1×11 的特征图;

第11层和第12层是卷积核为 1×2 、步长为 1×1 的卷积层,它是第二个尺度选择单元SSU的卷积操作;对每个卷积层的输出进行批归一化(BN)处理,再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活;所述的2个卷积层的输出特征图的数量依次是(576, 576),第12个卷积层输出576个大小为 1×9 的特征图;

第13层和第14层是卷积核为 1×2 、步长为 1×1 的卷积层,它是第三个尺度选择单元SSU的卷积操作;对每个卷积层的输出进行批归一化处理,再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活;所述的2个卷积层的输出特征图的数量依次是(640, 640),第14个卷积层输出640个大小为 1×7 的特征图;

第15层和第16层是卷积核为 1×2 、步长为 1×1 的卷积层,它是第四个尺度选择单元SSU的卷积操作;对每个卷积层的输出进行批归一化处理,再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活;所述的2个卷积层的输出特征图的数量依次是(640, 640),第16个卷积层输出640个大小为 1×5 的特征图;

(3) 预测网络,即第17、第18个全连接层:

第17层是拥有1024个神经元的全连接层,使用ReLU函数对该全连接层的输出进行激活,使用Dropout技术处理ReLU函数的输出,每一个神经元只有50%的概率参与计算;第17层的最终输出是一个大小为 1×1024 的特征向量;

第18层是拥有229个神经元的全连接层,从而得到大小为 1×229 的视差预测向量 logit_i ;

步骤2-2:使用多层交叉熵损失函数对SOA-EPN进行训练,多层交叉熵损失函数如下:

$$\begin{cases} \text{Loss} = -1/N \sum_{i=0}^4 w_i \sum_{n=1}^N y_{gr}^n \log(y_i^n) \\ y_i^n = \text{softmax}(\text{logit}_i^n) = e^{\text{logit}_i^n} / \| e^{\text{logit}_i^n} \| \end{cases}$$

其中, N 是训练集的一个批大小, w_i 是权重, $\| \cdot \|$ 计算向量的L1距离, y_{gr}^n 是真实视差概率向量;

所述步骤3具体实现如下：

从训练集中随机选取256个样本作为一个批次输入到SOA-EPN网络，网络预测输出 logit_i ，利用前面定义的多层交叉熵损失函数计算预测值 logit_i 和真实值 y_{gt}^n 之间的误差，进而使用反向传播算法和梯度优化算法更新SOA-EPN网络的参数；当Loss达到稳定值则完成训练；

所述步骤4具体包括下述步骤：

从测试集中任取一个场景，提取该场景中心子光圈图像上任一点P四个方向的EPI图像块，输入到训练好的SOA-EPN网络，网络输出一个大小为229的概率向量 logit_4 ，将 logit_4 中概率最大的类记为z，其概率记为a，比较z-1和z+1两个类的概率并取概率较大的类记为s，其概率记为b；则P点的视差disp计算如下：

$\text{disp} = [(a/(a+b)) \times z + (b/(a+b)) \times s] \times 0.035 + \text{offset}$ ，其中，offset为偏移参数，取值4；则P点的深度 $\text{depth} = Bf/\text{disp}$ ，其中f是相机焦距，B是两个相邻子光圈的距离。

2. 根据权利要求1所述的基于方向和尺度自适应卷积神经网络的光场深度估计方法，其特征在于所述的切片操作 $s(\cdot)$ 是指沿宽度通道切除 x_{i-1} 中每一个特征图的两端各一个神经元，进而得到 x_{i-1} 的中心区域，也就是 x_i ；具体来说，是尺寸为 9×21 的EPI图像块经过8个卷积层后输出的特征图； x_0 进行一次切片操作后得到 x_1 ， x_1 是尺寸为 9×19 的EPI图像块经过8个卷积层后输出的特征图；同理， x_1 进行一次切片操作后得到 x_2 ， x_2 是尺寸为 9×17 的EPI图像块经过8个卷积层后输出的特征图；以此类推， x_0 、 x_1 、 x_2 、 x_3 、 x_4 分别是 9×21 、 9×19 、 9×17 、 9×15 、 9×13 五个尺寸的EPI图像块的特征图，尺寸依次是 $512 \times 1 \times 13$ 、 $576 \times 1 \times 11$ 、 $576 \times 1 \times 9$ 、 $640 \times 1 \times 7$ 、 $640 \times 1 \times 5$ 。

基于方向和尺度自适应卷积神经网络的光场深度估计方法

技术领域

[0001] 本发明是关于深度学习与光场深度估计领域,特别涉及基于方向和尺度自适应卷积神经网络的光场深度估计方法。

背景技术

[0002] 深度学习是近年来人工智能领域中的一项重要突破,已在图像识别、语音识别、自然语言处理等多个方向取得了突破性进展。相比于传统机器学习方法,深度学习方法的主要过程为:构建数据集;使用卷积层、全连接层、激活层等定义深度神经网络,定义损失函数;利用数据集训练定义好的深度网络模型,即使用反向传播技术和梯度下降等优化方法更新神经网络的参数。训练完成的深度网络模型可以拟合出某个高维复杂函数,该函数实现网络输入到输出的映射;利用训练好的深度模型在测试数据集上进行预测。卷积神经网络是人工神经网络发展最成功的网络结构之一,它擅长利用卷积层提取输入图像从低维到高维的深度特征,然后利用这些特征完成分类、回归等任务。

[0003] 与此同时,光场成像与计算是近年来计算机视觉领域中的一门新兴技术,通过在相机镜头和图像传感器之间增加一个微透镜阵列,光场相机可以记录下汇聚在传感器像素上的光线的亮度与角度,形成4D光场,为突破传统成像模型和成像机制的维度局限提供了新思路和新方法[]。鉴于4D光场提供的丰富信息,利用光场图像进行场景结构(深度)恢复的研究获得了越来越多的关注,一批基于EPI(Epipolar Plane Image)图像分析的深度估计方法涌现出来,这些方法本质上是一种立体匹配方法。EPI图像是4D光场的一个切片,它由极线(Epipolar Lines)组成,通过计算极线的斜率,可以间接获得深度图,完成深度估计。但是这种方法也面临着难以计算图像无纹理区域、遮挡区域的深度等问题。如何选择大小合适的匹配窗口是解决这些问题的一个关键,例如在计算图像无纹理区域的深度时,需要一个较大的窗口将更多的邻域像素纳入到匹配代价的计算中,增强邻域约束。相反,在计算图像遮挡区域的深度时,需要一个较小的窗口避免纳入过多的遮挡点影响匹配代价的计算。另一方面,多方向EPI图像为深度估计提供了更多的约束和深度线索。综上,我们设计了一种具备EPI图像方向自适应、EPI图像尺度自适应能力的卷积神经网络来解决上述深度估计面临的无纹理和遮挡问题。

发明内容

[0004] 本发明的目的是针对现有技术的不足,设计了一种以 0° 、 45° 、 90° 、 135° 四个方向的EPI图像块作为输入,输出视差图,具备EPI图像方向自适应、EPI图像尺度自适应能力的卷积神经网络(SOA-EPN),并利用SOA-EPN提出了一种光场深度估计方法。

[0005] 本发明解决其技术问题采取的技术方案包括如下步骤:

[0006] 步骤1.准备光场数据集,制作训练集和测试集;

[0007] 步骤2.搭建方向和尺度自适应的卷积神经网络SOA-EPN;

[0008] 步骤3.使用训练集训练搭建好的SOA-EPN网络;

[0009] 步骤4.使用练好的SOA-EPN网络在测试集上进行测试;

[0010] 步骤1具体包括下述步骤:

[0011] 步骤1-1:使用海德堡图像处理实验室提供的4D光场数据集(K.Honauer, O.Johannsen,D.Kondermann,and B.Goldluecke,“A dataset and evaluation methodology for depth estimation on 4d light fields,”in Asian Conference on Computer Vision,2016.)作为实验数据集,该数据集包含28个场景,并提供了高精度的视差和性能评估指标。每一个场景包含9行9列共81张大小为 512×512 的子光圈图像。将上述28个场景分为两部分,利用其中16个场景制作训练集,12个场景制作测试集。

[0012] 步骤1-2:将每一个4D光场场景定义为 $LF(s, t, x, y)$,其中 (x, y) 是图像空间坐标系, (s, t) 是角度坐标系。对于中心子光圈图像 $LF(4, 4, x, y)$ 上的任一点P,提取P点的 0° 、 45° 、 90° 、 135° 四个方向大小为 $9 \times 21 \times 3$ (高 \times 宽 \times 通道)的EPI图像块,这四个EPI图像块构成了数据集的一个样本,输入到卷积神经网络。

[0013] 所述步骤2具体实现如下:

[0014] 步骤2-1:搭建方向和尺度自适应的卷积神经网络SOA-EPN。

[0015] 所述的SOA-EPN由4个尺度感知网络(Scale-Aware network简称:SAN)构成。训练集和测试集的每一个样本包含四个EPI图像块,每一个所述的尺度感知网络SAN是以其中1个EPI图像块作为输入,预测输出是该图像块中心点P的视差预测向量 logit_i ,其中 $i=0, 1, 2, 3$,因此会得到4个视差预测向量,使用一个有229个神经元的全连接层将这4个视差预测向量进行融合得到一个最终视差;

[0016] 为了减少SOA-EPN的隐含层数量,减少网络参数量,降低卷积神经网络过拟合的风险,将4个尺度感知网络SAN进行权值共享,达到深度压缩的目的。

[0017] 设定视差预测范围是-4到4(单位:像素),将视差预测定义为一个分类任务,预测精度是0.035个像素,则可分为229类。

[0018] 所述尺度感知网络SAN一共18层,分为3个子网络,具体参看表1:

[0019] 表1:

层号	卷积核	特征图数量	ReLU	步长	标准化	输出张量
第一步：深度特征提取						
1	2×2	16	ReLU	1×1	BN	1×20×16
2		32				1×19×32
3		64				1×18×64
4		128				1×17×128
5		256				1×16×256
6		384				1×15×384
7		512				1×14×512
8		512				1×14×512
第二步：尺度自适应选择(卷积和残差网)						
9	1×2	576	Leaky ReLU	1×1	BN	1×12×576
10		576				1×11×576
11		576				1×10×576
12		576				1×9×576
13		640				1×8×640
14		640				1×7×640
15		640				1×6×640
16		640				1×5×640
第3步：预测(全连接层)						
17	全连接层 1024 个神经元, ReLU					1024
18	全连接层 229 个神经元					229

[0021] 所述的3个子网络,具体如下:

[0022] (1) 深度特征表征网络(第1到第8个卷积层)

[0023] 该部分使用8个卷积核为2×2、步长为1×1的卷积层对输入EPI图像块进行深度特征表征,对每个卷积层的输出进行批归一化(BN, batch normalization)处理,再使用ReLU函数对批归一化的输出结果进行激活;所述的8个卷积层中每个卷积层的输出特征图数量依次是(16, 32, 64, 128, 256, 384, 512, 512),第8个卷积层输出的张量记为 x_0 , x_0 其大小为512×1×13(特征图数量×高×宽),即512个大小为1×13的特征图(特征向量)。

[0024] (2) 尺度自适应选择网络(第9到第16个卷积层)

[0025] 尺度自适应选择网络(Scale-Adaptive-Selection-Network, SASN)是一个具有长跳跃和短跳跃连接(skip connection: 跳跃式连接)的多层前馈网络,包含了四个尺度选择单元(Scale-Selection-Unit, SSU),每个尺度选择单元SSU由2个卷积核为1×2、步长为1×1的卷积层,一个切片操作和一个跳跃式连接组成。尺度自适应选择网络SASN的作用是改变SOA-EPN网络的输入EPI图像块的尺寸大小。实现原理是:SOA-EPN在第8个卷积层的感知域大小为9×9,而 x_0 是尺寸为9×21的EPI图像块的深度特征,通过改变 x_0 的尺度可以间接地改变输入EPI图像块的尺寸,进而实现尺度自适应选择功能(利用感知域的概念,通过切片操作改变特征向量的尺寸,进而间接改变输入图像的尺寸,达到尺度自适应选择的目的,是本专利的核心思想)。可以通过四次切片操作 $s(\cdot)$ 改变 x_0 的尺寸,进而获得尺寸为9×19、9×

17、9×15、9×13的输入EPI图像块的深度特征 x_1 、 x_2 、 x_3 、 x_4 ，尺度自适应选择功能具体来说就是在9×21、9×19、9×17、9×15、9×13五个尺寸中选择一个或者是它们的某种组合，选择的策略是在训练SOA-EPN的过程学习得到的。值得注意的是，SASN只以一张尺寸为9×21的EPI图像块作为输入，通过尺度选择单元SSU实现尺度自适应选择，而不是输入多张不同尺寸的EPI图像块。

[0026] 第 i 个尺度选择单元的输出 h_i 定义为：

$$[0027] \quad \begin{cases} h_0 = F(h_{i-1}) + x_i \\ x_0 = s(x_{i-1}) \\ h_0 = x_0 \end{cases}$$

[0028] 其中， x_0 是尺寸为9×21的EPI图像块的深度特征， x_i 代表 x_1 、 x_2 、 x_3 、 x_4 。 $F(\cdot)$ 代表每一个尺度选择单元SSU的2个卷积层的卷积操作。值得注意的是， x_i 的特征图数量可能会小于 $F(h_{i-1})$ 的特征图数量，采用补0的方式使得 x_i 和 $F(h_{i-1})$ 的特征图数量相等。 $s(\cdot)$ 是切片操作。

[0029] 所述的切片操作 $s(\cdot)$ 是指沿宽度通道切除 x_{i-1} 中每一个特征图(特征向量)的两端各一个神经元，进而得到 x_{i-1} 的中心区域，也就是 x_i 。具体来说，是尺寸为9×21的EPI图像块经过8个卷积层后输出的特征图； x_0 进行一次切片操作后得到 x_1 ， x_1 是尺寸为9×19的EPI图像块经过8个卷积层后输出的特征图；同理， x_1 进行一次切片操作后得到 x_2 ， x_2 是尺寸为9×17的EPI图像块经过8个卷积层后输出的特征图。以此类推， x_1 、 x_2 、 x_3 、 x_4 分别是9×21、9×19、9×17、9×15、9×13五个尺寸的EPI图像块的特征图，尺寸依次是512×1×11、576×1×11、576×1×9、640×1×7、640×1×5。

[0030] 第9层和第10层是卷积核为1×2、步长为1×1的卷积层，它是第一个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(576,576)，第10个卷积层输出576个大小为1×11的特征图。

[0031] 第11层和第12层是卷积核为1×2、步长为1×1的卷积层，它是第二个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(576,576)，第12个卷积层输出576个大小为1×9的特征图。

[0032] 第13层和第14层是卷积核为1×2、步长为1×1的卷积层，它是第三个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(640,640)，第14个卷积层输出640个大小为1×7的特征图。

[0033] 第15层和第16层是卷积核为1×2、步长为1×1的卷积层，它是第四个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(640,640)，第16个卷积层输出640个大小为1×5的特征图。

[0034] (3) 预测网络(第17、第18个全连接层)

[0035] 第17层是拥有1024个神经元的全连接层(Fully Connected Layer,简称FC)，使用ReLU函数对该全连接层的输出进行激活，使用Dropout技术处理ReLU函数的输出，每一个神

经元只有50%的概率参与计算。第17层的最终输出是一个大小为 1×1024 的特征向量。

[0036] 第18层是拥有229个神经元的全连接层,从而得到大小为 1×229 的视差预测向量 logit_i 。

[0037] 步骤2-2:为了增强了训练过程的可控性,减少深度学习的黑箱效应,使用多层交叉熵损失函数对SOA-EPN进行训练,多层交叉熵损失函数如下:

$$[0038] \quad \begin{cases} \text{Loss} = -1/N \sum_{i=0}^4 w_i \sum_{n=1}^N y_{gt}^n \log(y_i^n) \\ y_i^n = \text{softmax}(\text{logit}_i^n) = e^{\text{logit}_i^n} / \|e^{\text{logit}_i^n}\| \end{cases}$$

[0039] 其中,N是训练集的一个批大小, w_i 是权重, $\|\cdot\|$ 计算向量的L1距离, y_{gt}^n 是真实视差概率向量。

[0040] 所述步骤3具体实现如下:

[0041] 从训练集中随机选取256个样本作为一个批次输入到SOA-EPN网络,网络预测输出 logit_i ,利用前面定义的多层交叉熵损失函数计算预测值 logit_i 和真实值 y_{gt}^n 之间的误差,进而使用反向传播算法和梯度优化算法更新SOA-EPN网络的参数。当Loss达到稳定值则完成训练。

[0042] 所述步骤4具体包括下述步骤:

[0043] 从测试集中任取一个场景,提取该场景中心子光圈图像上任一点P四个方向的EPI图像块,输入到训练好的SOA-EPN网络,网络输出一个大小为229的概率向量 logit_4 ,将 logit_4 中概率最大的类记为z,其概率记为a,比较z-1和z+1两个类的概率并取概率较大的类记为s,其概率记为b。则P点的视差disp计算如下:

[0044] $\text{disp} = [(a/(a+b)) \times z + (b/(a+b)) \times s] \times 0.035 + \text{offset}$,其中,offset为偏移参数,取值4。

[0045] 则P点的深度 $\text{depth} = Bf/\text{disp}$,其中f是相机焦距,B是两个相邻子光圈的距离。

[0046] 与现有技术相比,本发明主要贡献是:

[0047] (1).尺度感知结构.它将切片操作应用于多尺度特征提取,然后通过多层前馈网络进行长跳和短跳连接的自适应尺度选择。

[0048] (2).多方向图像块共享权重网络.它在多尺度感知结构中共享参数,实现多方位特征融合。

[0049] (3).一种用于深度估计的图像块学习模型.为了达到规模和方向感知,它包含上述两种网络结构,并以多方向图像块作为输入。

附图说明

[0050] 图1:多方向EPI patch截取效果图。

[0051] 图2:SOA-EPN的网络结构图。

[0052] 图3:SAN网络结构详细展示图。

[0053] 图4:SAN网络各层详细参数介绍。

具体实施方式

[0054] 下面结合附图与具体实施方式对本发明作进一步详细描述:

[0055] 如图1-4所示,一种新的基于方向和尺度自适应卷积神经网络的光场深度估计方法,用于对多方向和尺度的光场数据,包括下列过程:

[0056] 本发明方法的具体包括下述过程:

[0057] 本发明解决其技术问题采取的技术方案包括如下步骤:

[0058] 步骤1.准备光场数据集,制作训练集和测试集;

[0059] 步骤2.搭建方向和尺度自适应的卷积神经网络SOA-EPN;

[0060] 步骤3.使用训练集训练搭建好的SOA-EPN网络;

[0061] 步骤4.使用练好的SOA-EPN网络在测试集上进行测试;

[0062] 步骤1具体包括下述步骤:

[0063] 步骤1-1:使用海德堡图像处理实验室提供的4D光场数据集(K.Honauer, O.Johannsen,D.Kondermann,and B.Goldluecke,“A dataset and evaluation methodology for depth estimation on 4d light fields,”in Asian Conference on Computer Vision,2016.)作为实验数据集,该数据集包含28个场景,并提供了高精度的视差和性能评估指标。每一个场景包含9行9列共81张大小为 512×512 的子光圈图像。将上述28个场景分为两部分,利用其中16个场景制作训练集,12个场景制作测试集。

[0064] 步骤1-2:如图1所示将每一个4D光场场景定义为 $LF(s, t, x, y)$,其中 (x, y) 是图像空间坐标系, (s, t) 是角度坐标系。对于中心子光圈图像 $LF(4, 4, x, y)$ 上的任一点P,提取P点的 0° 、 45° 、 90° 、 135° 四个方向大小为 $9 \times 21 \times 3$ (高 \times 宽 \times 通道)的EPI图像块,这四个EPI图像块构成了数据集的一个样本,输入到卷积神经网络。

[0065] 所述步骤2具体实现如下:

[0066] 步骤2-1:搭建方向和尺度自适应的卷积神经网络SOA-EPN。

[0067] 如图2所示,所述的SOA-EPN由4个尺度感知网络(Scale-Aware network简称:SAN)构成。训练集和测试集的每一个样本包含四个EPI图像块,每一个所述的尺度感知网络SAN是以其中1个EPI图像块作为输入,预测输出是该图像块中心点P的视差预测向量 \logit_i ,其中 $i=0, 1, 2, 3$,因此会得到4个视差预测向量,使用一个有229个神经元的全连接层将这4个视差预测向量进行融合得到一个最终视差;

[0068] 为了减少SOA-EPN的隐含层数量,减少网络参数量,降低卷积神经网络过拟合的风险,将4个尺度感知网络SAN进行权值共享,达到深度压缩的目的。

[0069] 设定视差预测范围是-4到4(单位:像素),将视差预测定义为一个分类任务,预测精度是0.035个像素,则可分为229类。

[0070] 如图3和4所示,所述尺度感知网络SAN一共18层,分为3个子网络,具体参看表1:

[0071] 表1:

层号	卷积核	特征图数量	ReLU	步长	标准化	输出张量
第一步：深度特征提取						
1	2×2	16	ReLU	1×1	BN	1×20×16
2		32				1×19×32
3		64				1×18×64
4		128				1×17×128
5		256				1×16×256
6		384				1×15×384
7		512				1×14×512
8		512				1×14×512
第二步：尺度自适应选择(卷积和残差网)						
[0072] 9	1×2	576	Leaky ReLU	1×1	BN	1×12×576
10		576				1×11×576
11		576				1×10×576
12		576				1×9×576
13		640				1×8×640
14		640				1×7×640
15		640				1×6×640
16		640				1×5×640
第3步：预测(全连接层)						
17	全连接层 1024 个神经元, ReLU					1024
18	全连接层 229 个神经元					229

[0073] 所述的3个子网络,具体如下:

[0074] (1) 深度特征表征网络(第1到第8个卷积层)

[0075] 该部分使用8个卷积核为2×2、步长为1×1的卷积层对输入EPI图像块进行深度特征表征,对每个卷积层的输出进行批归一化(BN, batch normalization)处理,再使用ReLU函数对批归一化的输出结果进行激活;所述的8个卷积层中每个卷积层的输出特征图数量依次是(16, 32, 64, 128, 256, 384, 512, 512),第8个卷积层输出的张量记为 x_0 , x_0 其大小为512×1×13(特征图数量×高×宽),即512个大小为1×13的特征图(特征向量)。

[0076] (2) 尺度自适应选择网络(第9到第16个卷积层)

[0077] 尺度自适应选择网络(Scale-Adaptive-Selection-Network, SASN)是一个具有长跳跃和短跳跃连接(skip connection: 跳跃式连接)的多层前馈网络,包含了四个尺度选择单元(Scale-Selection-Unit, SSU),每个尺度选择单元SSU由2个卷积核为1×2、步长为1×1的卷积层,一个切片操作和一个跳跃式连接组成。尺度自适应选择网络SASN的作用是改变SOA-EPN网络的输入EPI图像块的尺寸大小。实现原理是:SOA-EPN在第8个卷积层的感知域大小为9×9,而 x_0 是尺寸为9×21的EPI图像块的深度特征,通过改变 x_0 的尺度可以间接地改变输入EPI图像块的尺寸,进而实现尺度自适应选择功能(利用感知域的概念,通过切片操作改变特征向量的尺寸,进而间接改变输入图像的尺寸,达到尺度自适应选择的目的,是本专利的核心思想)。可以通过四次切片操作 $s(\cdot)$ 改变 x_0 的尺寸,进而获得尺寸为9×19、9×

17、9×15、9×13的输入EPI图像块的深度特征 x_1 、 x_2 、 x_3 、 x_4 ，尺度自适应选择功能具体来说就是在9×21、9×19、9×17、9×15、9×13五个尺寸中选择一个或者是它们的某种组合，选择的策略是在训练SOA-EPN的过程学习得到的。值得注意的是，SASN只以一张尺寸为9×21的EPI图像块作为输入，通过尺度选择单元SSU实现尺度自适应选择，而不是输入多张不同尺寸的EPI图像块。

[0078] 第 i 个尺度选择单元的输出 h_i 定义为：

$$[0079] \quad \begin{cases} h_0 = F(h_{i-1}) + x_i \\ x_0 = s(x_{i-1}) \\ h_0 = x_0 \end{cases}$$

[0080] 其中， x_0 是尺寸为9×21的EPI图像块的深度特征， x_i 代表 x_1 、 x_2 、 x_3 、 x_4 。 $F(\cdot)$ 代表每一个尺度选择单元SSU的2个卷积层的卷积操作。值得注意的是， x_i 的特征图数量可能会小于 $F(h_{i-1})$ 的特征图数量，采用补0的方式使得 x_i 和 $F(h_{i-1})$ 的特征图数量相等。 $s(\cdot)$ 是切片操作。

[0081] 所述的切片操作 $s(\cdot)$ 是指沿宽度通道切除 x_{i-1} 中每一个特征图(特征向量)的两端各一个神经元，进而得到 x_{i-1} 的中心区域，也就是 x_i 。具体来说，是尺寸为9×21的EPI图像块经过8个卷积层后输出的特征图； x_0 进行一次切片操作后得到 x_1 ， x_1 是尺寸为9×19的EPI图像块经过8个卷积层后输出的特征图；同理， x_1 进行一次切片操作后得到 x_2 ， x_2 是尺寸为9×17的EPI图像块经过8个卷积层后输出的特征图。以此类推， x_1 、 x_2 、 x_3 、 x_4 分别是9×21、9×19、9×17、9×15、9×13五个尺寸的EPI图像块的特征图，尺寸依次是512×1×11、576×1×11、576×1×9、640×1×7、640×1×5。

[0082] 第9层和第10层是卷积核为1×2、步长为1×1的卷积层，它是第一个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(576, 576)，第10个卷积层输出576个大小为1×11的特征图。

[0083] 第11层和第12层是卷积核为1×2、步长为1×1的卷积层，它是第二个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(576, 576)，第12个卷积层输出576个大小为1×9的特征图。

[0084] 第13层和第14层是卷积核为1×2、步长为1×1的卷积层，它是第三个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(640, 640)，第14个卷积层输出640个大小为1×7的特征图。

[0085] 第15层和第16层是卷积核为1×2、步长为1×1的卷积层，它是第四个尺度选择单元SSU的卷积操作。对每个卷积层的输出进行批归一化(BN)处理，再使用斜率为0.01的Leaky ReLU函数对批归一化的输出结果进行激活；所述的2个卷积层的输出特征图的数量依次是(640, 640)，第16个卷积层输出640个大小为1×5的特征图。

[0086] (3) 预测网络(第17、第18个全连接层)

[0087] 第17层是拥有1024个神经元的全连接层(Fully Connected Layer, 简称FC)，使用ReLU函数对该全连接层的输出进行激活，使用Dropout技术处理ReLU函数的输出，每一个神

神经元只有50%的概率参与计算。第17层的最终输出是一个大小为 1×1024 的特征向量。

[0088] 第18层是拥有229个神经元的全连接层,从而得到大小为 1×229 的视差预测向量 logit_i 。

[0089] 步骤2-2:为了增强了训练过程的可控性,减少深度学习的黑箱效应,使用多层交叉熵损失函数对SOA-EPN进行训练,多层交叉熵损失函数如下:

$$[0090] \quad \begin{cases} \text{Loss} = -1/N \sum_{i=0}^4 w_i \sum_{n=1}^N y_{gt}^n \log(y_i^n) \\ y_i^n = \text{softmax}(\text{logit}_i^n) = e^{\text{logit}_i^n} / \|e^{\text{logit}_i^n}\| \end{cases}$$

[0091] 其中,N是训练集的一个批大小, w_i 是权重, $\|\cdot\|$ 计算向量的L1距离, y_{gt}^n 是真实视差概率向量。

[0092] 所述步骤3具体实现如下:

[0093] 从训练集中随机选取256个样本作为一个批次输入到SOA-EPN网络,网络预测输出 logit_i ,利用前面定义的多层交叉熵损失函数计算预测值 logit_i 和真实值 y_{gt}^n 之间的误差,进而使用反向传播算法和梯度优化算法更新SOA-EPN网络的参数。当Loss达到稳定值则完成训练。

[0094] 所述步骤4具体包括下述步骤:

[0095] 从测试集中任取一个场景,提取该场景中心子光圈图像上任一点P四个方向的EPI图像块,输入到训练好的SOA-EPN网络,网络输出一个大小为229的概率向量 logit_4 ,将 logit_4 中概率最大的类记为z,其概率记为a,比较z-1和z+1两个类的概率并取概率较大的类记为s,其概率记为b。则P点的视差disp计算如下:

[0096] $\text{disp} = [(a/(a+b)) \times z + (b/(a+b)) \times s] \times 0.035 + \text{offset}$,其中,offset为偏移参数,取值4。

[0097] 则P点的深度 $\text{depth} = Bf/\text{disp}$,其中f是相机焦距,B是两个相邻子光圈的距离。

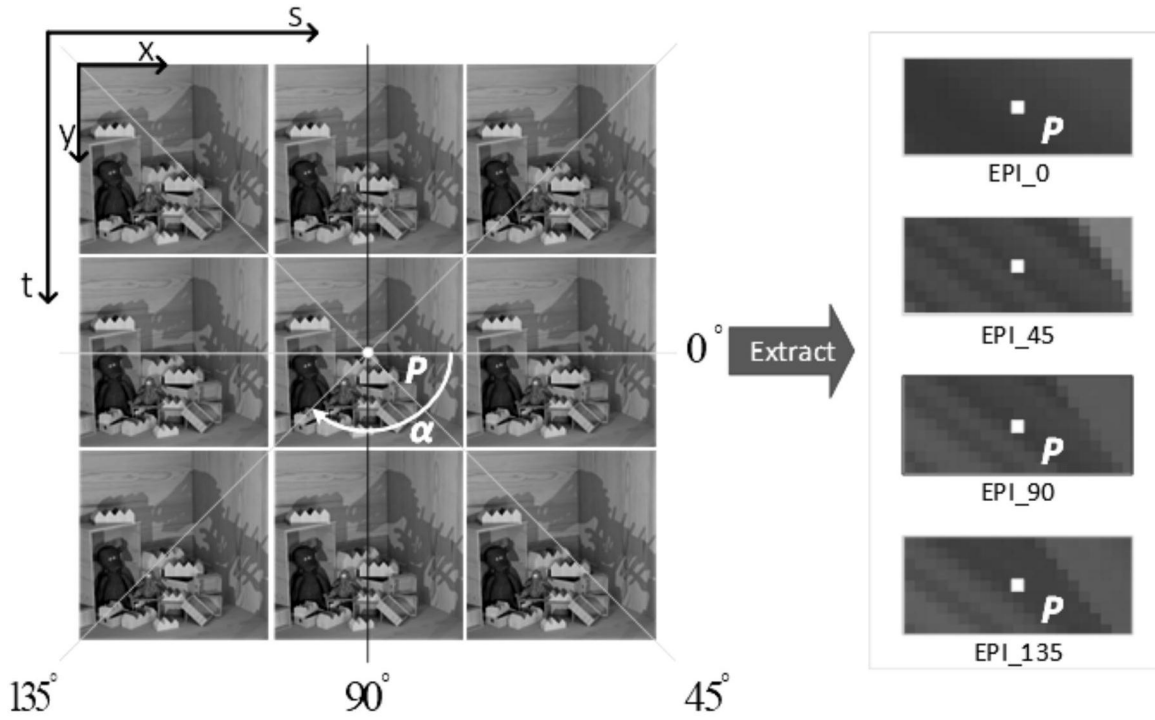


图1

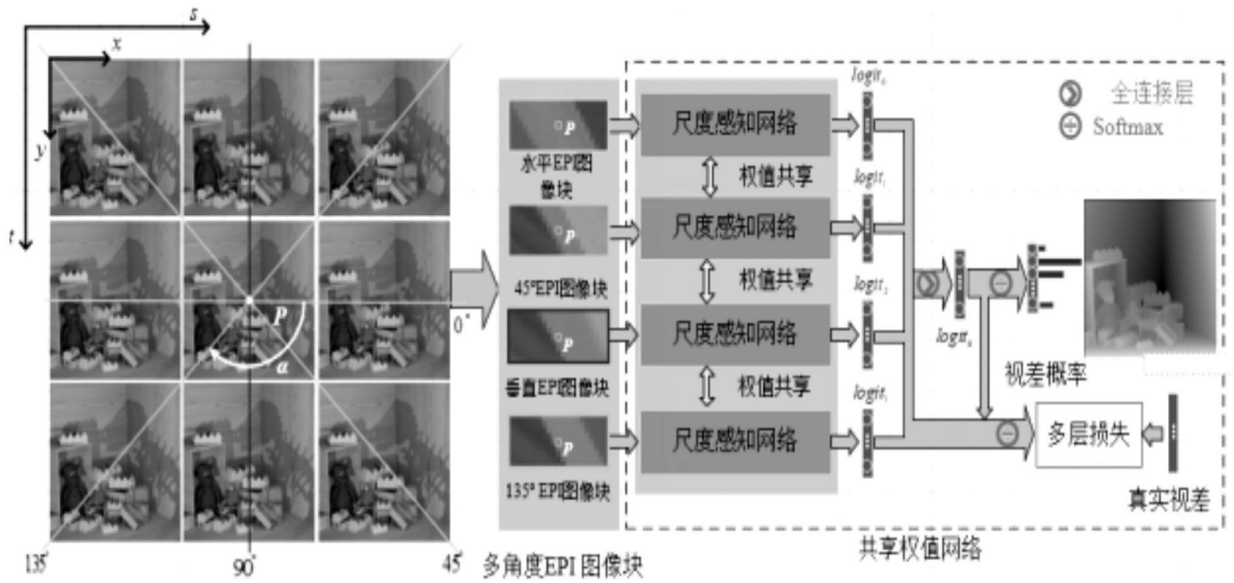


图2

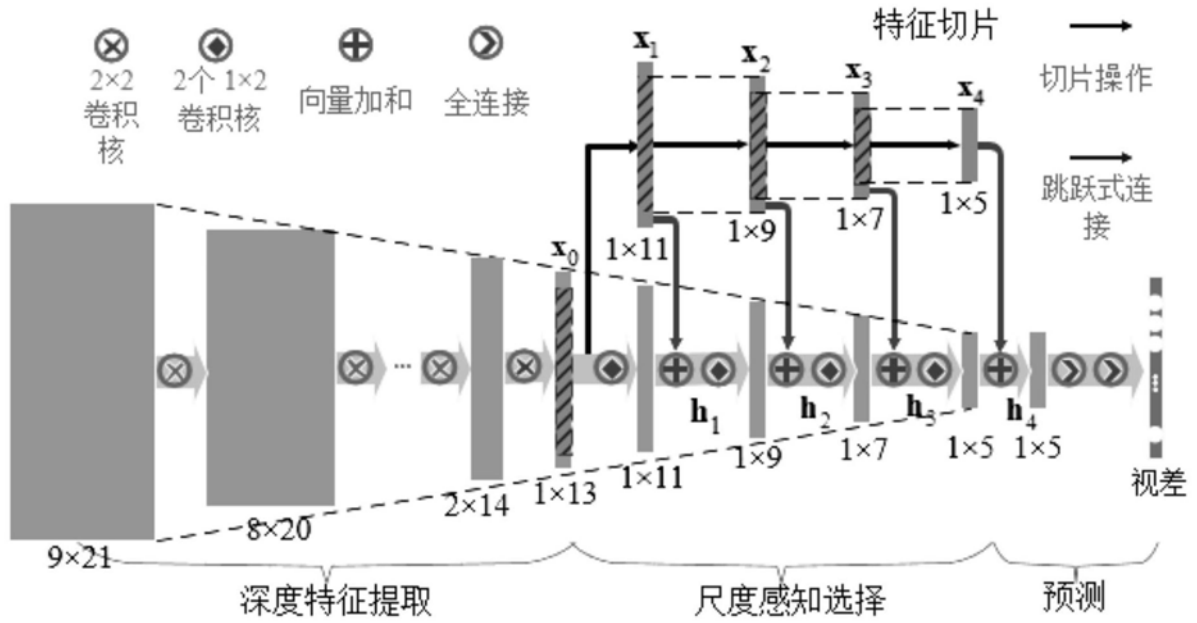


图3

ARCHITECTURES OF SCALE-AWARE STRUCTURE

Input: EPI-patch ($H \times W \times C: 9 \times 21 \times 3$)

	Layer Description	Output Tensor
Step1: Deep feature representation (convolution)		
1 – 8	eight 2×2 conv, (16,32,64,128,256,384,512,512) features, stride 1×1 , BN, ReLU	$1 \times 13 \times 512$
Step2: Adaptive scale selection (convolution & skip-connection)		
9 – 10	two 1×2 conv, (576,576) features, stride 1×1 , BN, Leaky ReLU(0.01)	$1 \times 11 \times 576$
11 – 12	two 1×2 conv, (576,576) features, stride 1×1 , BN, Leaky ReLU(0.01)	$1 \times 9 \times 576$
13 – 14	two 1×2 conv, (640,640) features, stride 1×1 , BN, Leaky ReLU(0.01)	$1 \times 7 \times 640$
15 – 16	two 1×2 conv, (640,640) features, stride 1×1 , BN, Leaky ReLU(0.01)	$1 \times 5 \times 640$
Step3: Predict (fully connection)		
17	FC1024, ReLU, Dropout(0.5)	1024
18	FC229	229

图4