



(12) 发明专利

(10) 授权公告号 CN 110111366 B

(45) 授权公告日 2021.04.30

(21) 申请号 201910371605.4

(22) 申请日 2019.05.06

(65) 同一申请的已公布的文献号
申请公布号 CN 110111366 A

(43) 申请公布日 2019.08.09

(73) 专利权人 北京理工大学
地址 100081 北京市海淀区中关村南大街5号

(72) 发明人 陈文颀 孙洋洋 窦丽华 陈杰

(74) 专利代理机构 北京理工大学专利中心
11120

代理人 温子云

(51) Int.Cl.

G06T 7/246 (2017.01)

(56) 对比文件

CN 107292912 A, 2017.10.24

CN 107506712 A, 2017.12.22

CN 109146872 A, 2019.01.04

审查员 宋晓毓

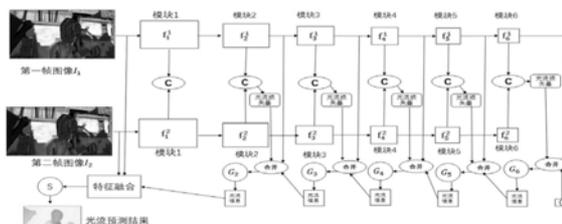
权利要求书2页 说明书8页 附图3页

(54) 发明名称

一种基于多级损失量的端到端光流估计方法

(57) 摘要

本发明公开了一种基于多级损失量的端到端光流估计方法,将两张相邻图像送入相同的特征提取卷积神经网络进行特征提取,获得两帧图像的多尺度的特征图;在每一个尺度下,对两幅图像特征图进行相关性分析操作,从而获得多尺度的损失量信息;将同一尺度下获取的损失量信息、第一帧图像在该尺度下的特征图、上一级预测得到光流信息合并在一起,送入光流预测卷积神经网络中,获取该尺度下的残差流,并与上一级光流信息的上采样结果相加得到该尺度的光流信息;将第二级尺度的光流信息与输入的两帧图像进行特征融合操作,将融合后的信息送入到运动边缘优化网络得到最终的光流预测结果。使用本发明能够提升光流估计算法精度与效率。



CN 110111366 B

1. 一种基于多级损失量的端到端光流估计方法,其特征在于,包括:

步骤一:将两张相邻图像 I_1 和 I_2 ,送入相同的特征提取卷积神经网络进行特征提取,获得两帧图像的多尺度的特征图 f_i^1 和 f_i^2 ;所述特征提取卷积神经网络是一个N级孪生网络, $i=1,2,\dots,N$;

步骤二:在每一个尺度 i 下,对图像 I_1 和 I_2 的特征图进行相关性分析操作,获得该尺度 i 下的损失量信息 cv_i ,从而获得多尺度的损失量信息;

步骤三:针对每一个尺度 i ,将尺度 i 下获取的损失量信息 cv_i 、第一帧图像 I_1 在该尺度下的特征图 f_i^1 、上一级预测得到光流信息 V_{i+1} 合并在一起,送入光流预测卷积神经网络 G_i 中,获取该尺度下的残差流 v_i ,将残差流 v_i 与上一级光流信息的上采样结果 \hat{V}_{i+1} 相加得到该尺度的光流信息 V_i ;

步骤四:将第二级尺度的光流信息 V_2 与输入的两帧图像 I_1 和 I_2 进行特征融合操作,将融合后的信息送入到运动边缘优化网络得到最终的光流预测结果。

2. 如权利要求1所述的方法,其特征在于,所述特征提取卷积神经网络包括6个模块;每个模块由两个卷积层构成,两个卷积层具有相同的宽度,且卷积核大小设置为3,每个卷积层后面使用leaky-ReLU作为激活函数;特征提取卷积神经网络中采样步长代替池化层,第一层卷积使用步长为2,第二层卷积使用步长为1;第一级模块到第六级模块的卷积层通道数分别设置为{16,32,64,96,128,192}。

3. 如权利要求1所述的方法,其特征在于,所述相关性分析操作为:对于特征图 f_i^1 中的某个像素点 x_1 ,以逐像素的形式与特征图 f_i^2 中以 x_2 为中心、面积为 $[2k+1]^2$ 的正方形区域进行相乘操作,得到 $[2k+1]^2$ 个乘积结果后以通道的形式堆叠起来,即为损失量信息 cv_i ;其中, k 为设定的区域边长,尺度越低,采用的 k 值越大。

4. 如权利要求3所述的方法,其特征在于,不同尺度的 k 值不同,所述特征提取卷积神经网络中第2~第6个模块尺度下的特征图分别使用 $k=[14,10,8,6,4]$ 。

5. 如权利要求1所述的方法,其特征在于,所述上采样操作采用双线性插值获得。

6. 如权利要求1所述的方法,其特征在于,所述光流预测卷积神经网络 G_i 包括6个卷积层,没有步长和池化层;每个卷积核的大小设置为 3×3 ,且每个卷积层后面使用leaky-ReLU作为非线性激活层;各个卷积层的通道数设置为{128,128,96,64,32,2}。

7. 如权利要求1所述的方法,其特征在于,所述运动边缘优化网络包括特征提取子网络和ASPP模块;

特征提取子网络共含有5个模块,使用步长代替池化操作,除了第一个模块仅使用一个卷积核大小为 7×7 的卷积层外,其余每个模块都含有两个通道数相同且卷积核大小为 3×3 的卷积层,每个卷积层后面使用leaky-ReLU作为非线性激活函数,各个模块的通道数分别设置为[64,64,128,256,512];5个模块均使用空洞率为2的空洞卷积;

ASPP模块包括5个模块,前四个均为卷积核大小为 3×3 的卷积层,不同卷积层的空洞率分别设置为[1,4,8,12];特征提取子网络的输出并行进入这4个模块进行处理,输出的特征进入一个卷积核大小为3、通道数为64、步长为0的卷积层中进行特征的特征融合操作;

输入的光流信息与两帧图像的合并结果经特征提取子网络和ASPP模块的处理后,使用一个卷积核为3、通道数为2、步长为0的卷积层预测光流的值,并进行上采样,得到上采样的

编码特征图；

将特征提取子网络中第二个卷积层得到的特征图进行还原操作，即使用卷积核大小为1、通道数为8的卷积进行卷积操作；之后将还原得到的特征图与上采样的编码特征图进行合并，将合并结果经过两层卷积核大小为3、通道数为64的卷积层进行特征融合，最后经过一个卷积核大小为3、通道数为2的卷积层预测得到最后的光流结果。

8. 如权利要求1所述的方法，其特征在于，对步骤一至步骤四的操作所构成的多级损失量的端到端光流估计网络进行训练时，所采用的L2损失函数L为：

$$L = \sum_x |\varphi(V_i(\theta, x)) - \varphi(\bar{V}_i(x))|_2 + \lambda|\theta|_2$$

$$\varphi(x) = x + \gamma \left| \frac{\ln(x + \alpha)}{e^x} \right|$$

其中， $V_i(\theta, x)$ 为预测得到的光流结果，即光流矩阵， x 为光流矩阵的每个元素， θ 为本端光流估计方法中所使用的所有神经网络的参数， $\bar{V}_i(x)$ 为作为样本训练标签的光流实际值，即样本训练标签， α 为设定的较小正数， γ 为正则系数， $|\cdot|_2$ 表示 L_2 范数， λ 为正则项系数。

一种基于多级损失量的端到端光流估计方法

技术领域

[0001] 本发明涉及计算机视觉中的光流估计领域,特别涉及端到端光流估计,具体为一种基于多级损失量的端到端光流估计方法。

背景技术

[0002] 光流表征着图像亮度模式的表观运动,这种运动通常由观测者和场景的相对运动引起。光流估计是计算机视觉领域中经典的研究课题之一,作为一种低级别的视觉任务,光流在高级别视觉任务,如视频动作识别、视频目标检测及跟踪等,有着广泛的使用;此外,一个高性能的光流估计算法对于基于光流的应用,如视频编辑、机器人导航等有着重要意义。

[0003] 光流场是运动场在二维空间上的投影,是运动的低级别表示。光流算法的目的是解算出给定的若干帧图像间(通常是两帧)各个像素移动的距离。光流估计需要精确的逐像素定位,同时需要找到两张图输入图像的对应关系。这就涉及到图像特征表达的学习,以及学习在两个图像上的不同位置匹配这些特征表达。主流光流估计算法作如下分类:一是传统的非深度学习算法;二是基于深度学习的光流估计算法。非深度学习方法大部分做法是首先将光流计算问题构建为一个优化问题,即求解最小化能量函数,通常使用变分法进行求解。基于非深度学习的光流估计算法需要人工精心设计有效的能量函数,要得到稠密光流需要大量运行时间,此外,这类方法多研究限制条件下小位移光流的预测,限制了该类算法的应用场景。基于深度学习的光流估计算法利用深度CNN提取输入图像的从低级到高级的特征,然后基于深度特征使用传统算法进行光流求解。目前,领先的光流算法通常使用基于双目视觉的研究方法,这样的方法尽管取得了较高的精度,但是其运算速度往往长达数分钟,严重限制了其在实时性应用上的使用。近年来,由于深度学习在计算机视觉领域的领先地位,卷积神经网络(CNN)也开始被引入到光流估计算法之中。在这些方法中,绝大部分是首先使用CNN提取图像高级语义信息,然后使用传统的基于区域匹配的方法进行光流估计,然而这些方法往往又需要复杂的后处理,占据了大量的运行时间,不满足视频检测、跟踪等任务的速度需求。因此设计一种高效率、高精度的基于CNN的端到端光流估计网络显得至关重要。

发明内容

[0004] 有鉴于此,本发明提供了一种基于多级损失量的端到端光流估计方法,能够提升光流估计算法精度与效率。

[0005] 为了解决上述技术问题,本发明是这样实现的:

[0006] 一种基于多级损失量的端到端光流估计方法,包括:

[0007] 步骤一:将两张相邻图像 I_1 和 I_2 ,送入相同的特征提取卷积神经网络进行特征提取,获得两帧图像的多尺度的特征图 f_i^1 和 f_i^2 ;

[0008] 步骤二:在每一个尺度 i 下,对图像 I_1 和 I_2 的特征图进行相关性分析操作,获得该尺度 i 下的损失量信息 cv_i ,从而获得多尺度的损失量信息;

[0009] 步骤三:针对每一个尺度 i ,将尺度 i 下获取的损失量信息 cv_i 、第一帧图像 I_1 在该尺度下的特征图 f_i^1 、上一级预测得到光流信息 V_{i+1} 合并在一起,送入光流预测卷积神经网络 G_i 中,获取该尺度下的残差流 v_i ,将残差流 v_i 与上一级光流信息的上采样结果 \hat{V}_{i+1} 相加得到该尺度的光流信息 V_i ;

[0010] 步骤四:将第二级尺度的光流信息 V_2 与输入的两帧图像 I_1 和 I_2 进行特征融合操作,将融合后的信息送入到运动边缘优化网络得到最终的光流预测结果。

[0011] 优选地,所述特征提取卷积神经网络包括6个模块;每个模块由两个卷积层构成,两个卷积层具有相同的宽度,且卷积核大小设置为3,每个卷积层后面使用leaky-ReLu作为激活函数;特征提取卷积神经网络中采样步长代替池化层,第一层卷积使用步长为2,第二层卷积使用步长为1;第一级模块到第六级模块的卷积层通道数分别设置为{16,32,64,96,128,192}。

[0012] 优选地,所述相关性分析操作为:对于特征图 f_i^1 中的某个像素点 x_1 ,以逐像素的形式与特征图 f_i^2 中以 x_2 为中心、面积为 $[2k+1]^2$ 的正方形区域进行相乘操作,得到 $[2k+1]^2$ 个乘积结果后以通道的形式堆叠起来,即为损失量信息 cv_i ;其中, k 为设定的区域边长,尺度越低,采用的 k 值越大。

[0013] 优选地,不同尺度的 k 值不同,所述特征提取卷积神经网络中第2~第6个模块尺度下的特征图分别使用 $k=[14,10,8,6,4]$ 。

[0014] 优选地,所述上采样操作采用双线性插值获得。

[0015] 优选地,所述光流预测卷积神经网络 G_i 包括6个卷积层,没有步长和池化层;每个卷积核的大小设置为 3×3 ,且每个卷积层后面使用leaky-ReLu作为非线性激活层;各个卷积层的通道数设置为{128,128,96,64,32,2}。

[0016] 优选地,所述运动边缘优化网络包括特征提取子网络和ASPP模块;

[0017] 特征提取子网络共含有5个模块,使用步长代替池化操作,除了第一个模块仅使用一个卷积核大小为 7×7 的卷积层外,其余每个模块都含有两个通道数相同且卷积核大小为 3×3 的卷积层,每个卷积层后面使用leaky-ReLu作为非线性激活函数,各个模块的通道数分别设置为[64,64,128,256,512];5个模块均使用空洞率为2的空洞卷积;

[0018] ASPP模块包括5个模块,前四个均为卷积核大小为 3×3 的卷积层,不同卷积层的空洞率分别设置为[1,4,8,12];特征提取子网络的输出并行进入这4个模块进行处理,输出的特征进入一个卷积核大小为3、通道数为64、步长为0的卷积层中进行特征的特征融合操作;

[0019] 输入的光流信息与两帧图像的合并结果经特征提取子网络和ASPP模块的处理后,使用一个卷积核为3、通道数为2、步长为0的卷积层预测光流的值,并进行上采样,得到上采样的编码特征图;

[0020] 将特征提取子网络中第二个卷积层得到的特征图进行还原操作,即使用卷积核大小为1、通道数为8的卷积进行卷积操作;之后将还原得到的特征图与上采样的编码特征图进行合并,将合并结果经过两层卷积核大小为3、通道数为64的卷积层进行特征融合,最后经过一个卷积核大小为3、通道数为2的卷积层预测得到最后的光流结果。

[0021] 优选地,对步骤一至步骤四的操作所构成的多级损失量的端到端光流估计网络进行训练时,所采用的L2损失函数 L 为:

$$[0022] \quad L = \sum_x |\varphi(V_i(\theta, x)) - \varphi(\bar{V}_i(x))|_2 + \lambda|\theta|_2$$

$$[0023] \quad \varphi(x) = x + \gamma \left| \frac{\ln(x + \alpha)}{e^x} \right|$$

[0024] 其中, $V_i(\theta, x)$ 为预测得到的光流结果, 即光流矩阵, x 为光流矩阵的每个元素, θ 为本端光流估计方法中所使用的所有神经网络的参数, $\bar{V}_i(x)$ 为作为样本训练标签的光流实际值, 即样本训练标签, α 为设定的较小正数, γ 为正则系数, $|\cdot|_2$ 表示 L_2 范数, λ 为正则项系数, 需要在训练中根据训练数据调整。

[0025] 有益效果:

[0026] (1) 在当前的主流的基于双目视觉的光流估计算法运算速度受限的情况下, 本发明提出一种基于多级损失量的端到端光流估计方法, 本发明利用了基于卷积神经网络的光流估计算法, 在光流估计中利用了多级的特征图计算得到的光流损失量, 可以提升光流估计的准确性。

[0027] 同时, 在利用损失量计算光流信息时, 利用后一级的光流估计结果作为输入成分, 通过构造一个从上到下的反向过程将下一级的光流输出结果作为上一级的光流估计的输入, 在对光流信息预测时, 不直接预测该级下的光流信息, 而是预测得到光流残差信息, 使得光流估计网络更易于收敛。

[0028] 其次, 在进行最终光流估计时, 利用运动边缘优化网络提升目标位置信息估计的精确性。

[0029] 此外, 由于基于卷积神经网络的端到端光流估计算法在实际使用中只需要运行前向网络, 因此具有高效率的特点。

[0030] (2) 本发明的特征提取网络没有使用更深和更厚的特征图, 而是在保证特征图提取的语义信息足够丰富的情况下, 使用层数尽可能少, 只有6层卷积层, 进而保证光流估计具有较高的估计效率, 这是光流预测中的一个重要指标。同时为了减少计算量, 特征提取网络的所有的卷积核大小设置为3。在特征提取网络中没有使用池化层, 因为考虑到池化层的简单的最大化或者平均策略会导致信息的丢失。本发明使用卷积中的步长代替池化操作, 第一层卷积使用步长为2, 第二层卷积使用步长为1, 因此每经过一个模块, 特征图的尺度降低为原来的 $\frac{1}{2}$, 即最高层特征图的尺度为原图的 $\frac{1}{64}$ 。在使用步长代替池化的过程中, 保留了特征图的信息, 同时获得了多尺度特征图。

[0031] (3) 本发明对现有的相关性分析操作进行了相应的改进。原始的相关性分析是对特征图上的每一块区域内进行卷积操作来计算特征图的区域间相关性, 本发明将特征图上的块操作改进为对特征图上的每个元素进行卷积操作, 获取特征图间的相关性。改进的相关性分析保证分析过程覆盖到所有的像素点而没有遗漏, 从而提高了损失量信息计算的准确度, 进而提升了光流预测的精度。

[0032] (4) 本发明在进行光流预测时, 利用了多级尺度特征图, 对多级尺度的特征图得到的多级损失量进行预测进而得到每级尺度下的光流预测结果。同时将上一级得到的光流信息进行上采样作为下一级的光流残差流预测网络的输入, 使得光流预测网络更易收敛, 提

升预测精度。同时每级尺度下的光流预测网络不使用池化层,步长设置为0。光流残差流预测网络没有直接预测每个尺度下的光流值,而是训练网络预测上级光流相对于本级光流标签的残差值,进而提升光流预测的准确率。

[0033] (5) 本发明针对光流预测过程中的稠密预测与预测边缘模糊问题,使用了运动边缘优化网络对光流运动边缘进行修复。考虑光流任务对效率的较高要求,运动边缘优化网络的特征提取网络共5个模块,同时使用步长代替池化操作,从而可以减少信息丢失,同时提升光流预测效率。在边缘运动优化网络中使用空洞卷积,同时特征提取网络得到的特征图进行双线性差值上采样,将上采样的结果与原特征图进行特征融合,通过融合使得信息丢失减少,实现边缘的修复,进而提升光流预测的准确率。

[0034] (6) 本发明针对当前光流估计算法训练过程中的损失函数导致的光流网络对小位移或亚像素运动的预测结构不够准确问题,对光流网络的损失函数进行了重新设计,为了加大对亚像素运动的惩罚,我们将预测的光流结果映射到一个运动均衡空间,在空间内,亚像素映射结果与其非亚像素映射结果相近。通过损失函数的改进可以提升小位移的预测精度,进而使得光流网络的预测结果的精准性得到提升。

附图说明

[0035] 图1是本发明中的基于多级损失量的端到端光流估计算法的流程图;

[0036] 图2是本发明中的特征提取卷积神经网络结构图;

[0037] 图3是本发明的光流残差流的预测网络结构图;

[0038] 图4是基于多级损失量的端到端光流估计算法结构图;

[0039] 图5是本发明运动边缘优化网络结构图。

具体实施方式

[0040] 本发明提供了一种基于多级损失量的端到端光流估计方法,其基本思想是:将两张相邻图像 I_1 和 I_2 ,送入图2所示的特征提取卷积神经网络进行特征提取,获得两帧图像的多尺度的特征图;在每一个尺度 i 下,对图像 I_1 和 I_2 的特征图进行相关性分析操作,获得该尺度 i 下的损失量信息,从而获得多尺度的损失量信息;对于获取到的损失量信息,利用图3所示的卷积神经网络获取光流信息;对于获取到的光流信息,利用图5的运动边缘优化网络进行优化,获取最终的光流信息。

[0041] 可见,由于基于卷积神经网络的端到端光流估计算法在实际使用中只需要运行前向网络,因此具有高精度、高效率的特点。

[0042] 进一步地,获取最终光流信息后,设计改进的L2损失函数解决小位移或亚像素运动预测结果不够精确问题,完成端到端光流估计网络的构建,进行网络的训练与测试。

[0043] 下面结合附图并举实施例,对本发明进行详细描述。

[0044] 图1是本发明的基于多级损失量的端到端光流估计算法的流程图。该流程包括如下步骤:

[0045] 步骤一、将两张相邻图像 I_1 和 I_2 ,送入相同的特征提取卷积神经网络进行特征提取,获得两帧图像的多尺度的特征图 f_i^1 和 f_i^2 。

[0046] 如图2所示,从输入视频中获取相邻两帧图像,两帧图像分别记为第一帧图像 I_1 和

第二帧图像 I_2 ，将两帧图像输入到特征提取网络，该模型的特征提取网络采用卷积神经网络。两幅图像的特征提取为孪生网络，即对两帧图像提取特征的卷积神经网络完全相同。

[0047] 将两帧图像 I_1 和 I_2 输入特征提取网络后，在6个模块下获取不同尺度的特征图。特征提取网络的具体结构如图2所示，特征提取网络是一个6级孪生网络，即总共有6个模块。每一个模块由两个卷积层构成，两个卷积层具有相同的宽度，且为了减少计算量，所有的卷积核大小设置为3，每个卷积层后面使用leaky-ReLU作为激活函数，leaky-ReLU可以使输入小于0的部分不被简单的全部丢弃，同时网络没有使用pool（池化）层，因为考虑到池化层的简单的最大化或者平均策略会导致信息的丢失。本发明中使用stride（补偿）代替pool，第一层卷积使用stride为2，第二层卷积使用stride为1，因此每经过一个模块，特征图的尺度降低为原来的1/2，即最高层特征图的尺度为原图的1/64。从第一级模块到第六级模块的卷积层通道数分别设置为{16, 32, 64, 96, 128, 192}。相对于常用的VGG-16或Resnet-101结构，本发明没有使用更深和更厚的特征图，因为光流预测中效率是很重要的一个评测标准。

[0048] 步骤二：在每一个尺度 i 下，对图像 I_1 和 I_2 的特征图进行相关性分析操作，获得该尺度 i 下的损失量信息 cv_i ，从而获得多尺度的损失量信息。

[0049] 参见图4，该步骤包括如下子步骤：

[0050] 步骤S21：将步骤一得到的多尺度特征图为 f_i^1 和 f_i^2 （ $i=1, 2, \dots, 6$ ），其表示两帧图像由第 i 个模块输出的特征图，其中上角标1代表前一帧图像 I_1 ，2代表后一帧图像 I_2 ，即第 i 个尺度下提取得到的图像深度表示。

[0051] 对于得到的特征图 f_i^1 和 f_i^2 ，本发明用相关性分析（correlation）完成损失量信息构建过程，损失量信息构建过程如图4中的C操作所示。原始的相关性分析是用来构建区域间相关性的，即令特征图 f_i^1 中每个patch（块），在特征图 f_i^2 中的一个区域内进行卷积操作，并将每次卷积的结果以通道的方式堆叠。本发明对原始的相关性分析进行修改，把patch降维成一个元素时，即是损失量信息的表示，公式定义为：

$$c(x_1, x_2) = \sum_{o \in [-k, k] \times [-k, k]} \langle f_1(x_1), f_2(x_2+o) \rangle \quad (1)$$

[0053] 其中 x_1 是特征图中的元素， $f(x_1)$ 表示该位置上图像的亮度， x_2+o 定义了一个以 x_2 为中心，面积为 $[2k+1]^2$ 的正方形区域， x_1 与 x_2 在特征图 f_i^1 和 f_i^2 中都以stride=1的步长移动，这样可以考虑所有的像素点而没有遗漏。因此，公式表示对于特征图 f_i^1 中的某个像素点 x_1 ，以逐像素的形式与特征图 f_i^2 中以 x_2 为中心的区域进行相乘操作，得到 $[2k+1]^2$ 个乘积结果后以通道的形式堆叠起来。

[0054] 步骤S22：光流估计中不同模块下的特征图具有不同尺度，每一个像素点对应的感受野也大不相同，因而本发明使用多个特征图尺度来构建多尺度的损失量信息。在步骤S21中，对于尺度为 $w \times h$ 的两个特征图（ w 和 h 分别为特征图的长和宽），得到损失量信息的tensor（张量）大小为 $w \times h \times [2k+1]^2$ 。参数 k 用来度量每个尺度的特征图下考虑的相关范围，对于不同尺度的特征图，每个像素编码的感受野大小是不同的，例如第二个模块下的特征图的每个像素编码对应原图 4×4 的感受野，而最高级尺度的特征图的每个像素可以编码对应原图 64×64 的感受野。因此在低级尺度下采取较大的 k 值，在高级尺度下， k 取较小值，从而保证进行相关性分析操作的范围接近。本发明在不同的特征图尺度下使用不同的 k 值，具体来讲，在图4中的模块2到模块6对应的尺度下的特征图分别使用 $k=[14, 10, 8, 6, 4]$ ，本发明在模块2出得到该阶段最后光流预测，因而不考虑第模块1的损失量信息。在特征提取

网络得到的多个特征图尺度下执行这个操作,因此称之为多级损失量。本发明将获取的多级损失量表示为:

$$[0055] \quad cv_i = \text{corr}(f_i^1, f_i^2) \quad (2)$$

[0056] 其中,corr表示相关性分析操作。

[0057] 步骤三:针对每一个尺度*i*,将尺度*i*下获取的损失量信息 cv_i 、第一帧图像 I_1 在该尺度下的特征图 f_i^1 、上一级预测得到光流信息 V_{i+1} 合并在一起,送入光流预测卷积神经网络 G_i 中,获取该尺度下的残差流 v_i ,将残差流 v_i 与上一级光流信息的上采样结果 \hat{V}_{i+1} 相加得到该尺度的光流信息 V_i 。

[0058] 参见图4,本步骤的具体实施包括:

[0059] 步骤S31:在第*i*个模块下得到的*i*级特征图下,将该级的损失量信息 $cv_i = \text{corr}(f_i^1, f_i^2)$ 、前一帧图像 I_1 在第*i*个尺度下的深度特征图表示 f_{i+1}^1 、*i*+1模块下的*i*+1级预测得到的并经过上采样的光流结果利用合并操作合并到一起作为该级光流估计时的输入 $input_i$,记为:

$$[0060] \quad input_i = \text{concat}\{f_i^1, cv_i, u(V_{i+1})\} \quad (3)$$

[0061] 其中,concat是一个连接函数,是指将三个矩阵通过前后连接的方式合并在一起,合并后矩阵的大小是三个矩阵之和。 $u(\cdot)$ 是一个上采样函数,这里上采样是因为每一级卷积操作后特征图的大小发生了变化,为了使得下一级数据与本级数据能够合并,需要将他们处理成同等大小,这里的上采样就是将下一级数据通过采样方式扩大到本级数据大小; V_{i+1} 为上一级网络得到的光流结果。本发明将光流信息作为输入成份使网络更易于收敛,将上一级得到的光流 V_{i+1} 在经过上采样后作为下一级网络的输入成份。本发明使用双线性插值实现上采样过程,简单的双线性插值比其他上采样方法取得了更好的效果,记为:

$$[0062] \quad \hat{V}_i = u(V_i) \quad (4)$$

[0063] 步骤S32:将步骤S31得到的合并结果输入进图4中的轻量级的光流预测卷积神经网络 G_i 。光流预测卷积神经网络采用浅层光流预测卷积神经网络实现,其网络结构如图3所示,预测该尺度下的残差流,表示为:

$$[0064] \quad v_i = G_i(input_i, \theta) \quad (5)$$

[0065] 其中, v_i 为网络 G_i 预测得到的残差流结果, θ 为光流预测卷积神经网络要训练的网络参数。随后将残差流与上一级上采样后的光流 \hat{V}_{i+1} 相加得到该级的光流输出,表示为:

$$[0066] \quad V_i = v_i + \hat{V}_{i+1} \quad (6)$$

[0067] 光流预测卷积神经网络共有6个卷积层,输入张量的维度已经经过下采样,网络中没有进一步采取下采样操作,即没有stride或pool层。每个卷积核的大小设置为 3×3 ,且每个卷积层后面使用leay ReLU作为非线性激活层。各个卷积层的channel数设置为{128, 128, 96, 64, 32, 2},最后一层通道数为2表示在该层进行预测。本发明没有直接预测每个尺度下的光流值,而是训练网络预测上级光流相对于本级光流实际值(样本训练标签)的残差值,这比直接预测光流取得更好的效果。考虑到对于不同的输入尺度,一个 3×3 卷积核所能采集的感受野是不同的,因此对于不同的尺度下使用不同的轻量光流预测卷积神经网络。

[0068] 步骤四：将第二级尺度的光流信息 V_2 与输入的两帧图像 I_1 和 I_2 进行特征融合(fusion)操作,将融合后的信息送入到运动边缘优化网络得到最终的光流预测结果。

[0069] 图4所示网络中模块1得到的第一级特征图的语义信息过于低级,因此本发明从模块2得到的第2级尺度下的光流预测卷积神经网络得到第一阶段的光流。在得到第一阶段的光流后,本发明将第一阶段光流与输入图像进行特征融合操作。具体来讲,第一阶段得到的光流尺度实际上是输入图像尺度的 $\frac{1}{4}$,首先将光流数值结果扩大20倍,然后上采样到输入图像尺寸,接下来使用上采样的光流结果对第二帧图像进行弯曲操作,得到弯曲后的第二帧图像,表示为:

$$[0070] \quad \bar{I}_2 = w(\hat{V}_2, I_2) \quad (7)$$

[0071] 其中 $w(\cdot)$ 表示warping(弯曲)操作。之后对第一帧图像与弯曲后的第二帧图像求差,并将结果压缩为1维得到brightness(亮度)信息,此信息作为输入成份之一,表示为:

$$[0072] \quad \text{brightness} = \text{ChannelNorm}(I_1 - \bar{I}_2) \quad (8)$$

[0073] 其中,ChannelNorm(\cdot)表示维度压缩操作,将矩阵压缩为1维向量。

[0074] 接下来,将第一帧图像 I_1 、第二帧图像 I_2 、弯曲后的第二帧图像 \bar{I}_2 、上采样光流 \bar{I}_2 、brightness(亮度)信息一起进行合并操作。

[0075] 将合并操作得到的结果输入到运动边缘优化网络S中,进行运动边缘修复。运动边缘优化网络结构如图5所示,其中的特征提取网络共含有5个模块,使用步长(stride)代替池化(pooling)操作,除了第一个模块仅使用一个卷积核大小为 7×7 的卷积层外,其余4个模块都含有两个通道数相同且卷积核大小为 3×3 的卷积层,每个卷积层后面使用leaky-ReLu作为非线性激活函数,各个模块的通道数分别设置为[64, 64, 128, 256, 512];5个模块均使用空洞率为2的空洞卷积。

[0076] 由于在模块中使用了空洞率为2的空洞卷积,因而此时得到的张量尺度为原图的 $\frac{1}{16}$,将此张量输入进ASPP(atrous spatial pyramid pooling)模块,ASPP模块包括5个模块,前四个均为卷积核大小为 3×3 的卷积层,这4个卷积层的空洞率分别设置为[1, 4, 8, 12],通道数为64。特征提取子网络的输出并行进入这4个模块进行处理,输出的特征进入一个卷积核大小为3、通道数为64的卷积层中进行特征的特征融合操作。

[0077] 输入的光流信息与两帧图像的合并结果经特征提取子网络和ASPP模块的处理后,之后再使用一个卷积核为3、通道数为2的卷积层预测光流的值,这两个卷积层的步长为0,因此此时特征图大小依然为原图的 $\frac{1}{16}$,之后执行解码操作:首先将编码得到的特征图进行

上采样,本发明使用双线性插值将特征图上采样4倍,即为原图的 $\frac{1}{4}$ 。接下来为了补充底层特征编码的像素位置信息,将特征提取模块的第二个卷积得到的特征图进行还原操作,即使用卷积核大小1、通道数为8的卷积进行卷积操作。之后将还原得到的特征图与上采样的编码特征图进行合并,将合并结果经过两层卷积核大小为3、通道数为64的卷积层进行特征融合,最后经过一个卷积核大小为3、通道数为2的卷积层预测得到最后的光流结果。

[0078] 上述步骤一~步骤四构建了如图4的多级损失量的端到端光流估计网络。

[0079] 步骤五、对损失量的端到端光流估计网络进行训练与测试。

[0080] 由于本发明的光流估计网络为端到端网络,因此在训练时直接对图4所示的网络整体进行训练。训练时输入训练集,训练数据执行图4网络的前向过程,得到一个预测结果,同时训练集包含输入训练数据的ground truth(标签),利用损失函数对预测值和标签值计算网络误差,然后利用计算的误差对网络进行反向传播,实现网络的训练。

[0081] 在得到预测的光流结果后,本发明针对光流估计中小位移或亚像素预测不准的问题。对训练过程中采用的损失函数进行了仔细思考,目前普遍使用的L2损失函数L结构如下:

$$[0082] \quad L = \sum_x |V_i(\theta, x) - \bar{V}_i(x)|_2 + \lambda|\theta|_2 \quad (9)$$

[0083] 其中, $V_i(\theta, x)$ 为预测得到的光流结果(光流矩阵), x 为光流矩阵的每个元素, θ 为网络参数, $\bar{V}_i(x)$ 为作为样本训练标签的光流实际值(ground truth),即样本训练标签, $|\cdot|_2$ 表示 L_2 范数; λ 为正则项系数,需要在训练中根据训练数据调整。在损失函数下,若某个像素点对应的光流值小于1(即亚像素运动情况),经过平方后得到贡献的损失则变得更小,因此很难对亚像素运动进行惩罚。

[0084] 为了加大对亚像素运动的惩罚,本发明将预测的光流结果映射到一个运动均衡空间,在空间内,亚像素映射结果与其非亚像素映射结果相近,表示为:

$$[0085] \quad \varphi(x) = x + \gamma \left| \frac{\ln(x+\alpha)}{e^x} \right| \quad (10)$$

$$[0086] \quad L = \sum_x |\varphi(V_i(\theta, x)) - \varphi(\bar{V}_i(x))|_2 + \lambda|\theta|_2 \quad (11)$$

[0087] 其中 α 为一个接近0的较小正数,是为了避免真数位置出现0, γ 为一个正则系数。

[0088] 训练时使用多监督训练方法,即首先将光流ground truth乘以系数0.05,然后下采样至各个尺度,对该尺度对应的浅层网络预测得到的光流求取损失。对于不同尺度的损失函数,使用不同的系数来对训练施加不同的权重,具体来讲,5个尺度下的损失函数系数分别为[0.005,0.01,0.02,0.08,0.32],优化网络输出尺度与第二级尺度相同,因此使用系数0.005.这样的安排使得第二级和第三级损失权重变得稍大可以使网络对于图像细节更好学习。使用Adam算法执行模型的训练,其中 β_1 设置为0.9, β_2 设置为0.999。训练步数设置为120w步,前40w步使用基础学习率0.0001,之后每经过20w步迭代,学习率下降至原来的 $\frac{1}{2}$,直至网络完全收敛。

[0089] 训练完成后,利用视频或者相邻的两帧图像进行测试,测试过程中执行光流网络的前向过程即可。

[0090] 步骤六、采用训练和测试完成的网络进行光流估计。

[0091] 综上所述,以上仅为本发明的较佳实施例而已,并非用于限定本发明的保护范围。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

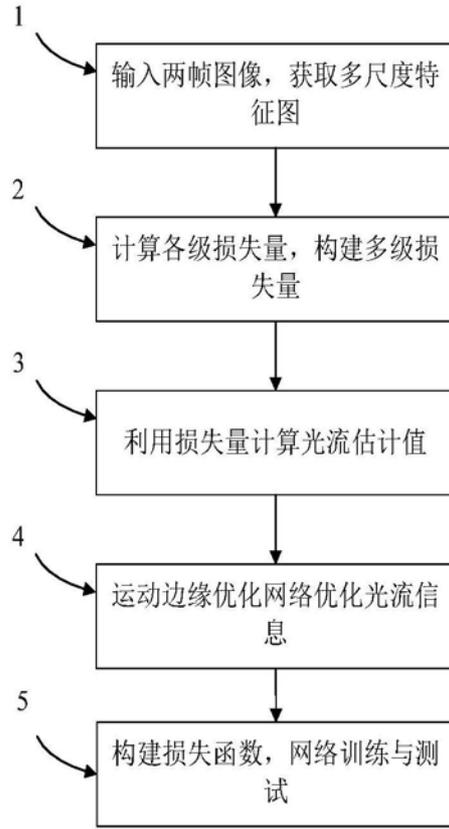


图1

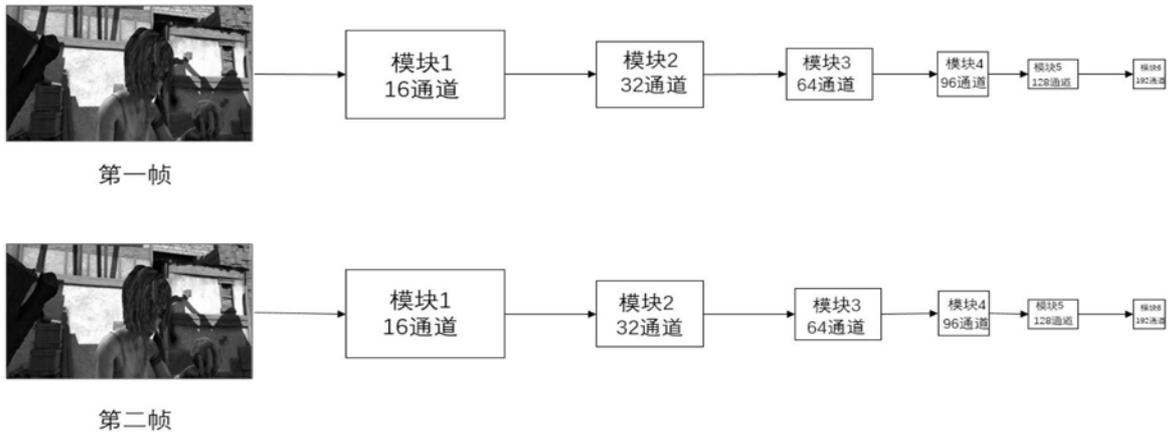


图2

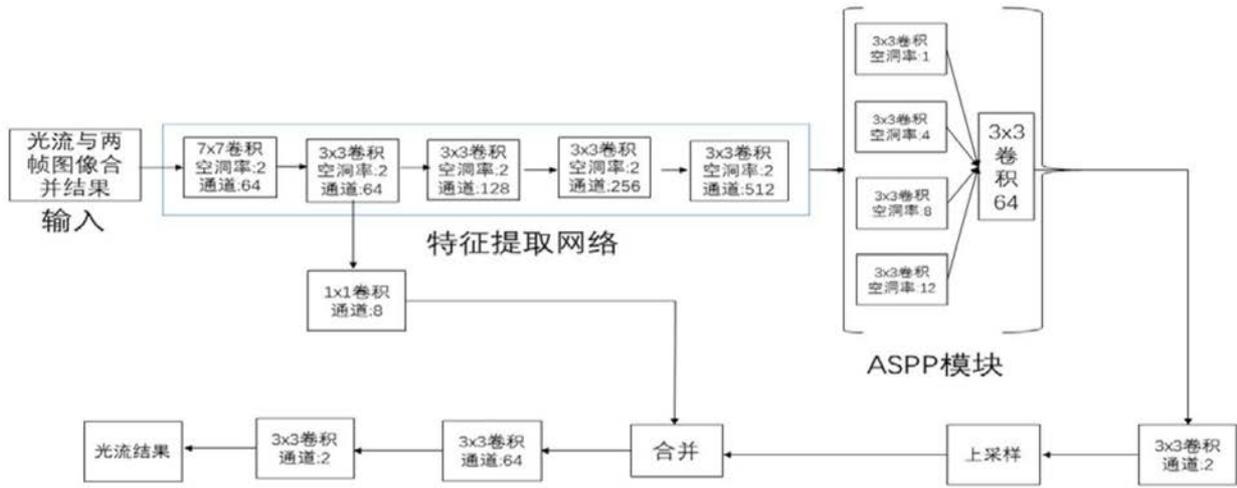


图5