



(12) 发明专利申请

(10) 申请公布号 CN 115439376 A

(43) 申请公布日 2022. 12. 06

(21) 申请号 202211367841.7

H04N 5/247 (2006.01)

(22) 申请日 2022.11.03

(71) 申请人 武汉大学

地址 430072 湖北省武汉市武昌区珞珈山
街道八一路299号

(72) 发明人 邹勤 张天任 陈驰 杨必胜
王中元

(74) 专利代理机构 武汉科皓知识产权代理事务
所(特殊普通合伙) 42222

专利代理师 齐晨涵

(51) Int. Cl.

G06T 5/00 (2006.01)

G06N 3/04 (2006.01)

G06V 10/80 (2022.01)

H04N 5/232 (2006.01)

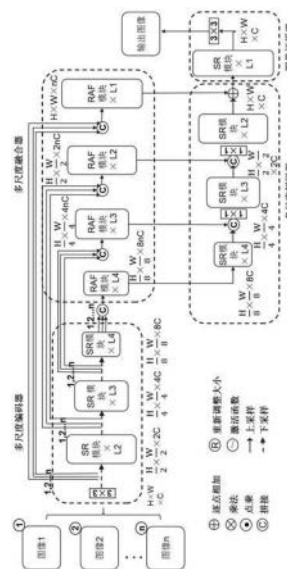
权利要求书2页 说明书7页 附图2页

(54) 发明名称

复眼相机多焦距图像融合模型、方法及装置

(57) 摘要

本文公开了一种复眼相机多焦距图像融合模型、方法及装置,以自参照感知模块SR作为模糊特征提取和特征恢复的基本单元,通过图像内的自参照实现全图特征比对。通过多层自参照提取离焦景深特征,结合自注意力与交叉注意力机制,在图像空间域上融合离焦特征,并在最后阶段采用多层自参照去模糊,从而有效的解决离焦扩散效应与离焦边界线不清晰的问题,并在一定程度上对近焦图像和远焦图像中依然模糊的区域做离焦去模糊,综合而言取得最佳融合效果。



1. 一种复眼相机多焦距图像融合网络模型,其特征在于:该网络模型包括编码器、多尺度统计融合器、解码器和图像还原器;该网络模型采用N张不同焦距的图像作为输入,每张图像单独输入编码器,利用自注意力编码输出四个尺度的特征,其中图像编码器的权重共享;输出为两个分支,其中一条输出由多尺度统计融合器实现,负责特征融合,对四个尺度上的特征进行融合特征,并经过特征最大、最小、均值的统计筛选获得精化的融合特征;另一条输出分支由解码器和还原器实现,负责解码四个尺度上的融合特征并重构图像;

其中,所述多尺度统计融合器基于交叉参照感知融合模块RAF,具体为:输入两个大小为 $H \times W \times C$ 的高维特征 X, Y ,计算 X 对 Y 的参照特征过程,先将 X 经过双层 3×3 的查询卷积并通过矩阵重排获得 $HW \times C$ 的 Q 矩阵 Q_X ,将 Y 经过双层 3×3 的映射卷积并通过矩阵重排获得 $C \times HW$ 的 K 矩阵 K_Y ,同时 Y 经过双层 3×3 的特征卷积并矩阵重排获得 $HW \times C$ 的 V 矩阵 V_Y ; Q 矩阵与 K 矩阵的乘积作为注意力决策矩阵,其大小为 $C \times C$,将 V 矩阵与注意力决策矩阵的乘积再与输入 X 相加,即获得 X 对 Y 的参照特征 \bar{X} ;同理, Y 对 X 的参照特征计算过程与 X 对 Y 参照特征过程同,得到 Y 对 X 的参照特征 \bar{Y} 。

2. 根据权利要求1所述的复眼相机多焦距图像融合网络模型,其特征在于:

所述编码器、解码器和图像还原器均基于自参照感知模块SR进行处理,所述自参照感知模块SR包含多层,每一层经过一个注意力决策子模块A和一个自参照特征生成子模块B,子模块A将输入为 $H \times W \times C$ 的特征图 X 经过批归一化和一个输入通道为 C 输出通道为 $3C$ 的 1×1 卷积,得到三等分 $H \times W \times C$,再分别经过双层 3×3 的卷积获得 Q 矩阵、 K 矩阵和 V 矩阵, Q 矩阵与 K 矩阵的乘积作为注意力决策矩阵,其大小为 $C \times C$,将 V 矩阵与注意力决策矩阵的乘积再与输入 X 相加,即获得 X 的自参照特征 \bar{X} ,子模块B将子模块A获得的 \bar{X} 经过批归一化和一个输入通道为 C 输出通道为 γC 的 1×1 卷积将特征扩充为 $H \times W \times \gamma C$,并将扩充后的特征分别经过双层 3×3 的卷积获得两个输出,将其中一个输出经过激活函数后与另一个输出进行矩阵点乘,将结果经过一个输入通道为 γC 输出通道为 C 的 1×1 卷积,并与最初的输入相加,获得最终结果,维度为 $H \times W \times C$ 。

3. 根据权利要求1所述的复眼相机多焦距图像融合网络模型,其特征在于:所述编码器设置有三层特征提取器,每一层采用像素重排进行降采样,将降采样后的特征经过自参照感知模块SR处理,输出四个尺度的高维特征。

4. 根据权利要求1所述的复眼相机多焦距图像融合网络模型,其特征在于:所述多尺度统计融合器对四层特征进行融合,每一层采用交叉感知处理,再经过特征最大、最小、均值的综合统计,将输入的多个尺度高维特征融合为一个尺度融合特征;即将参照特征 \bar{X} 和 \bar{Y} 同时计算通道层上的均值、最大、最小统计特征,再将三种统计特征叠加并经过 1×1 的卷积,最终输出 $H \times W \times C$ 的融合参照特征;计算公式如下:

$$\bar{X} = X + V_Y \times (Q_X \times K_Y).$$

$$\bar{Y} = Y + V_X \times (Q_Y \times K_X).$$

$$\text{RAF}_{(X, Y)} = \text{Conv}(\text{concat}(\text{Mean}(\bar{X}, \bar{Y}), \text{Min}(\bar{X}, \bar{Y}), \text{Max}(\bar{X}, \bar{Y}))).$$

5. 根据权利要求1所述的复眼相机多焦距图像融合网络模型,其特征在于:所述解码器对四层特征进行解码,每一层采用像素重排进行上采样,将输入的融合特征,经过每一层的

自参照感知模块SR处理,输出为解码后的低维图像特征。

6. 根据权利要求1所述的复眼相机多焦距图像融合网络模型,其特征在于:所述还原器利用自参照感知模块SR,将最终的融合解码特征精细化,用于离焦去模糊的高清重建;解码特征经过6层自参照感知模块重构,输入到卷积核大小为 3×3 步长为1填充为1的卷积,还原图像为 $H \times W \times 3$ 的RGB数据。

7. 一种利用权利要求1-6中任一所述复眼相机多焦距图像融合网络模型进行图像融合的方法,其特征在于,包括以下步骤:

构建仿真多焦距图像数据集,所述数据集中每一个样本包含一张具有景深差异特征的原始图像与K张不同离焦模糊的仿真图像;

利用所构建的数据集对所述复眼相机多焦距图像融合模型进行训练,从K张多焦距图像中随机选取N张不同离焦模糊图作为模型输入,以原始图像作为真值,训练模型学习自动融合聚焦特征并还原清晰图像的能力, $N \leq K$;

利用训练好的复眼相机多焦距图像融合模型进行多焦距图像融合的测试。

8. 根据权利要求7所述的利用复眼相机多焦距图像融合网络模型进行图像融合的方法,其特征在于:构建仿真多焦距图像数据集的具体过程如下:

S2-1,选取N张具有显著景深差异特征的图像作为原始图像,并对其进行深度估计和边缘提取;

S2-2,将深度数据和边缘数据量化为 $(x, y, \text{depth}, \text{edge})$ 的四维元组整列,按照聚簇数K的层次聚类算法,找到聚类中心作为K个多焦距图像的聚焦中心,其中 x, y 代表像素点的位置偏移, depth 代表像素点的深度值, edge 代表边缘梯度值;

S2-3,根据聚焦中心的深度值与图像上任一像素点之间的深度差值,决定图像上该点的高斯模糊核半径,并对该点处进行高斯模糊,从而分别得到K张不同聚焦中心的多焦距图像。

9. 一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序;

当一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求7-8中任一所述的图像融合方法。

10. 一种计算机可读介质,其上存储有计算机程序,其特征在于:所述程序被处理器执行时实现如权利要求7-8中任一所述的图像融合方法。

复眼相机多焦距图像融合模型、方法及装置

技术领域

[0001] 本发明涉及人工智能和机器视觉领域,具体涉及一种复眼相机多焦距图像融合模型、方法及装置。

背景技术

[0002] 复眼相机由多个图像采集单元组成,每个采集单元通常具有独立的相机和数字成像系统。目前,数字图像采集主要依据小孔成像原理,借助光感元件记录图像的成像。由于成像焦距的限制,整个相机系统无法获取全聚焦图像,由此造成在光感元件聚焦范围内的景物清晰,聚焦范围外的景物模糊的问题。当一台相机在拍摄时刻采用某一种焦距进行曝光,只能清晰呈现某一景深的物体,其他景深的物体将出现不同程度的离焦模糊。复眼相机具有多个相机,可以对每个相机设置不同的焦距,从而获得不同景深物体的清晰成像的若干图像。将不同焦点不同景深的多张图像在空间域或变换域上进行处理,得到全景深清晰图像,即在同一张图像上包含更丰富的信息,增强后续图像识别的效果。多焦距图像融合技术,在目标检测、目标分割、人脸识别等众多领域被广泛使用。

[0003] 现有的多焦距图像融合技术大多针对两张不同景深的图像进行融合,并且局限于有明显景深差异的两张图像之间的融合。然而,如图1所示,自然室外环境拍摄的图像中,往往存在以下特征:近景与远景的边界线难以清晰辨别;存在显著的离焦扩散效应;存在部分区域在近焦图像和远焦图像中依然模糊。现有的众多方法中,基于决策图的方法合成图像的像素内容“选取”于多焦距图像,对依然模糊的区域无法优化,并且决策图在聚焦与离焦边界上往往不准确;现有的端到端的方法,已有的有监督方法和基于图像生成的无监督方法,都无法有效处理近焦图像和远焦图像中依然模糊的问题。因而,提出一种能有效融合自然外景多焦距图像的算法具有重要意义。

发明内容

[0004] 基于以上技术问题,本发明提出了一种端到端的基于多尺度参照感知融合的复眼相机多焦距图像融合模型、方法及装置。

[0005] 本发明所设计的复眼相机多焦距图像融合网络模型,其特殊之处在于:该网络模型包括编码器、多尺度统计融合器、解码器和图像还原器;该网络模型采用N张不同焦距的图像作为输入,每张图像单独输入编码器,利用自注意力编码输出四个尺度的特征,其中图像编码器的权重共享;输出为两个分支,其中一条输出由多尺度统计融合器实现,负责特征融合,对四个尺度上的特征进行融合特征,并经过特征最大、最小、均值的统计筛选获得精化的融合特征;另一条输出分支由解码器和还原器实现,负责解码四个尺度上的融合特征并重构图像;

其中,所述多尺度统计融合器基于交叉参照感知融合模块RAF,具体为:输入两个大小为 $H \times W \times C$ 的高维特征 X, Y ,计算 X 对 Y 的参照特征过程,先将 X 经过双层 3×3 的查询卷积并通过矩阵重排获得 $HW \times C$ 的 Q 矩阵 Q_x ,将 Y 经过双层 3×3 的映射卷积并通过矩阵重排获得 C

$\times HW$ 的 K 矩阵 K_Y ,同时 Y 经过双层 3×3 的特征卷积并通过矩阵重排获得 $HW \times C$ 的 V 矩阵 V_Y 。 Q 矩阵与 K 矩阵的乘积作为注意力决策矩阵,其大小为 $C \times C$,将 V 矩阵与注意力决策矩阵的乘积再与输入 X 相加,即获得 X 对 Y 的参照特征 \bar{X} ;同理, Y 对 X 的参照特征计算过程与 X 对 Y 参照特征过程同,得到 Y 对 X 的参照特征 \bar{Y} 。

[0006] 进一步地,所述编码器、解码器和图像还原器均基于自参照感知模块SR进行处理,所述自参照感知模块SR包含多层,每一层经过一个注意力决策子模块A和一个自参照特征生成子模块B,子模块A将输入为 $H \times W \times C$ 的特征图 X 经过批归一化和一个输入通道为 C 输出通道为 $3C$ 的 1×1 卷积,得到三等分 $H \times W \times C$,再分别经过双层 3×3 的卷积获得 Q 矩阵、 K 矩阵和 V 矩阵, Q 矩阵与 K 矩阵的乘积作为注意力决策矩阵,其大小为 $C \times C$,将 V 矩阵与注意力决策矩阵的乘积再与输入 X 相加,即获得 X 的自参照特征 \bar{x} ,子模块B将子模块A获得的 \bar{x} 经过批归一化和一个输入通道为 C 输出通道为 γC 的 1×1 卷积将特征扩充为 $H \times W \times \gamma C$,并将扩充后的特征分别经过双层 3×3 的卷积获得两个输出,将其中一个输出经过激活函数后与另一个输出进行矩阵点乘,将结果经过一个输入通道为 γC 输出通道为 C 的 1×1 卷积,并与最初的输入相加,获得最终结果,维度为 $H \times W \times C$ 。

[0007] 进一步地,所述编码器设置有三层特征提取器,每一层采用像素重排进行降采样,将降采样后的特征经过自参照感知模块SR处理,输出四个尺度的高维特征。

[0008] 进一步地,所述多尺度统计融合器对四层特征进行融合,每一层采用交叉感知处理,再经过特征最大、最小、均值的综合统计,将输入的多个尺度高维特征融合为一个尺度融合特征;即将参照特征 \bar{x} 和 \bar{y} 同时计算通道层上的均值、最大、最小统计特征,再将三种统计特征叠加并经过 1×1 的卷积,最终输出 $H \times W \times C$ 的融合参照特征。计算公式如下:

$$\bar{X} = X + V_Y \times (Q_X \times K_Y).$$

$$\bar{Y} = Y + V_X \times (Q_Y \times K_X).$$

$$\text{RAF}_{(x, y)} = \text{Conv}(\text{concat}(\text{Mean}(\bar{X}, \bar{Y}), \text{Min}(\bar{X}, \bar{Y}), \text{Max}(\bar{X}, \bar{Y}))).$$

[0009] 进一步地,所述解码器对四层特征进行解码,每一层采用像素重排进行上采样,将输入的融合特征,经过每一层的自参照感知模块SR处理,输出为解码后的低维图像特征;

进一步地,所述还原器利用自参照感知模块SR,将最终的融合解码特征精细化,用于离焦去模糊的高清重建。解码特征经过6层自参照感知模块重构,输入到卷积核大小为 3×3 步长为1填充为1的卷积,还原图像为 $H \times W \times 3$ 的RGB数据。

[0010] 基于同一发明构思,本方案还设计了一种利用所述复眼相机多焦距图像融合模型进行图像融合的方法,其特殊之处在于,包括以下步骤:

构建仿真多焦距图像数据集,所述数据集中每一个样本包含一张具有景深差异特征的原始图像与 K 张不同离焦模糊的仿真图像;

利用所构建的数据集对所述复眼相机多焦距图像融合模型进行训练,以 K 张不同离焦模糊图作为模型输入,以原始图像作为真值,训练模型学习自动融合聚焦特征并还原清晰图像的能力;

利用训练好的复眼相机多焦距图像融合模型进行多焦距图像融合的测试。

[0011] 进一步地,构建仿真多焦距图像数据集的具体过程如下:

S2-1,选取N张具有显著景深差异特征的图像作为原始图像,并对其进行深度估计和边缘提取;

S2-2,将深度数据和边缘数据量化为(x,y,depth,edge)的四维元组整列,按照簇数K的层次聚类算法,找到聚类中心作为K个多焦距图像的聚焦中心;

S2-3,根据聚焦中心的深度值与图像上任一像素点之间的深度差值,决定图像上该点的高斯模糊核半径,并对该点处进行高斯模糊,从而分别得到K张不同聚焦中心的多焦距图像。

[0012] 进一步地,训练所述复眼相机多焦距图像融合模型过程中,一个训练样本中,从K张多焦距图像上随机选择N($N \leq K$)张作为网络的输入,这使得网络模型学习到融合多焦距图像的基本特性并具有离焦去模糊的特性。

[0013] 基于同一发明构思,本发明还设计了一种电子设备,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序;

当一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现所述的复眼相机多焦距图像融合方法。

[0014] 基于同一发明构思,本发明还设计了一种计算机可读介质,其上存储有计算机程序,其特征在于:所述程序被处理器执行时实现所述复眼相机多焦距图像融合方法。

[0015] 本发明首先构建一种端到端的多焦距图像融合模型,以自参照感知模块SR作为模糊特征提取和特征恢复的基本单元,通过图像内的自参照实现全图特征比对。通过多层自参照提取离焦景深特征,结合自注意力与交叉注意力机制,在图像空间域上融合离焦特征,并在最后阶段采用多层自参照去模糊,从而有效的解决离焦扩散效应与离焦边界线不清晰的问题,并在一定程度上对近焦图像和远焦图像中依然模糊的区域做离焦去模糊,综合而言取得最佳融合效果。

[0016] 本发明的深度数据和边缘数据量化为(x,y,depth,edge)的四维元组整列,按照簇数K的层次聚类算法,找到聚类中心作为k个多焦距图像的聚焦中心。传统的深度估计未充分考虑物体边缘,容易造成边缘模糊,而本发明在不同离焦区域划分时,结合边缘信息进行处理,可以实现具有边缘敏感性的物体分割,可实现更准确的离焦区域划分。

[0017] 本发明的优点:

首先,针对自然外景多焦距图像的特点,本发明创新性地结合多尺度交叉感知融合模块RAF,在离焦模糊度的鉴别上同时考量了同一图像空间域之间的对比特征和不同离焦图像之间的对比特征,从而更加有效提取离焦图像景深模糊特征并融合特征,对于边界模糊的多焦距图像能更好的融合。其次,图像中的离焦模糊和聚焦清晰,在信息熵上所表示的,是相对的信息量少熵值小与信息量大熵值高,因而在高维特征统计中,往往是统计量最值上涵盖了融合所需的信息。本发明创新性的构造了特征统计的最大、最小、均值融合模块,能从图像高维特征统计量上提取离焦对比度最值,从而保留融合所需的聚焦特征,使得最终融合图像具有最清晰效果。

[0018] 最后,采用多尺度方式融合,是图像处理中常用到的方式,通过在特征表征过程中减小图像尺度并增加特征通道数,能够更深层次的提取高维度的离焦特征,使得离焦模糊的边界判断更准确。

[0019] 对比而言,现有的大多数多焦距图像融合方法,大多基于卷积或者部分注意力机制的构思,缺少了图像之间与图像内部的完全的特征比对,从而在处理边界模糊、离焦扩散等现象中没有本方法效果好。

[0020] 与现有技术相比,本方案离焦模糊的图像边界辨别清晰;公共的模糊区域上,本方法在一定程度上去模糊效果明显优于现有方法,综合视觉效果最佳。

附图说明

[0021] 图1是多焦距图像融合网络模型总体结构。

[0022] 图2是网络中采用的自参照感知模块SR内部结构。

[0023] 图3是结合交叉感知融合的多尺度统计融合器内部结构。

具体实施方式

[0024] 本发明提供的方法设计了一种新型的注意力机制的编解码架构的深度学习网络模型,其总体结构参见图1。一种利用复眼相机多焦距图像融合模型进行图像融合的方法,实施包含以下步骤:

步骤S1,构建如图1所示的网络模型,具体步骤如下:

步骤S1-1,构建基于自参照的多尺度编码器:设置三层特征提取器,每一层采用像素重排PixelShuffle进行降采样,将降采样后的特征经过自参照感知模块SR处理,输出四个尺度的高维特征。利用自参照的多尺度编码器对输入的RGB的3通道图像,首先采用卷积核大小为 3×3 步长为1的卷积进行特征的高维表示,通道数扩充到C,此处C为32,并作为原始尺度的特征输出;随后采用PixelShuffle进行降采样,使得特征尺度减半通道数翻倍,经过一个自参照感知模块SR编码,并输出降尺度特征;此步骤重复两次,最终获得四个尺度的编码。图2所示为本方案采用的自参照感知模块SR结构:模块SR内包含多层,每一层经过一个注意力决策子模块A和一个自参照特征生成子模块B。子模块A将输入为 $H \times W \times C$ 的特征图X经过批归一化和一个输入通道为C输出通道为 $3C$ 的 1×1 卷积,得到三等分 $H \times W \times C$,再分别经过双层 3×3 的卷积获得Q矩阵、K矩阵和V矩阵,Q矩阵与K矩阵的乘积作为注意力决策矩阵,其大小为 $C \times C$,将V矩阵与注意力决策矩阵的乘积再与输入X相加,即获得X的自参照特征 \bar{x} 。子模块B将子模块A获得的 \bar{x} 经过批归一化和一个输入通道为C输出通道为 γC 的 1×1 卷积将特征扩充为 $H \times W \times \gamma C$,并将扩充后的特征分别经过双层 3×3 的卷积获得两个输出,将其中一个输出经过激活函数后与另一个输出进行矩阵点乘,将结果经过一个输入通道为 γC 输出通道为C的 1×1 卷积,并与最初的输入相加,获得最终结果,维度为 $H \times W \times C$ 。

[0025] 步骤S1-2,构建基于交叉感知的多尺度统计融合器,对四层特征进行融合,每一层采用交叉感知,再经过特征最大、最小、均值的综合统计,将输入的多个尺度高维特征融合为一个尺度融合特征。

[0026] 图3所示为本方案中基于交叉感知融合模块RAF的统计融合器模型结构:输入两个大小为 $H \times W \times C$ 的高维特征X、Y,计算X对Y的参照特征过程,先将X经过双层 3×3 的查询卷积并通过矩阵重排获得 $HW \times C$ 的Q矩阵 Q_x ,将Y经过双层 3×3 的映射卷积并通过矩阵重排获得 $C \times HW$ 的K矩阵 K_y ,同时Y经过双层 3×3 的特征卷积并通过矩阵重排获得 $HW \times C$ 的V矩阵 V_y 。Q矩阵与K矩阵的乘积作为注意力决策矩阵,其大小为 $C \times C$,将V矩阵与注意力决策矩阵的乘积

再与输入 X 相加,即获得 X 对 Y 的参照特征 \bar{x} ;同理, Y 对 X 的参照特征计算过程与 X 对 Y 参照特征过程同,得到 Y 对 X 的参照特征。特征最大、最小、均值的综合统计是将参照特征 \bar{x} 和 \bar{y} 同时计算通道层上的均值、最大、最小统计特征,再将三种统计特征叠加并经过 1×1 的卷积,最终输出 $H \times W \times C$ 的融合参照特征。计算公式如下:

$$\bar{X} = X + V_Y \times (Q_X \times K_Y)$$

$$\bar{Y} = Y + V_X \times (Q_Y \times K_X)$$

$$\text{RAF}_{(X, Y)} = \text{Conv}(\text{concat}(\text{Mean}(\bar{X}, \bar{Y}), \text{Min}(\bar{X}, \bar{Y}), \text{Max}(\bar{X}, \bar{Y})))$$

当待融合的特征为 N 个时,对 N 个高纬特征编号为 N_1, N_2, \dots, N_n ,依次对 N_1 和 N_2 按照上述步骤融合,将其融合结果再与 N_3 融合,以此类推,最终将融合结果与 N_n 融合,从而将 N 个高纬特征融合为一个融合特征。

[0027] 步骤S1-3,构建基于自参照感知的多尺度解码器,对四层特征进行解码,每一层采用像素重排进行上采样,将输入的融合特征,经过每一层的自参照感知模块SR处理,输出为解码后的低维图像特征。在 $1/8$ 的尺度上,将输入的融合特征经过2层自参照模块,由PixelUnShuffle进行上采样,将数据尺度翻倍通道数减半,将结果与 $1/4$ 的尺度上输入的融合特征在通道上合并得到 $8C$ 通道,经过卷积核大小为 1×1 的卷积后,由 $8C$ 通道合并到 $4C$ 通道。在 $1/4$ 的尺度上,将输入的融合特征经过2层自参照模块,由PixelUnShuffle进行上采样,将数据尺度翻倍通道数减半,将结果与 $1/2$ 的尺度上输入的融合特征在通道上合并得到 $8C$ 通道,经过卷积核大小为 1×1 的卷积后,由 $4C$ 通道合并到 $2C$ 通道。在 $1/2$ 的尺度上,将输入的融合特征经过4层自参照模块,由PixelUnShuffle进行上采样,将数据尺度翻倍通道数减半,将结果与 $H \times W \times C$ 尺度上输入的融合特征相加。

[0028] 步骤S1-4,构建基于自参照感知的图像还原器,利用自参照感知模块SR,将最终的融合解码特征精细化,用于离焦去模糊的高清重建。解码特征经过6层自参照感知模块重构,输入到卷积核大小为 3×3 步长为1填充为1的卷积,还原图像为 $H \times W \times 3$ 的RGB数据。

[0029] 步骤S2. 构建仿真多焦距图像数据集,数据集的构建如下:

步骤S2-1,选取 N 张具有显著景深差异特征的图像作为原始图像,并对其进行单图像深度估计,为了实现更准确的深度估计效果,分别对室外图像和室内图像采用对应的深度估计模型。 N 大于500。本方法实施过程采用Monocular Depth Prediction提供的算法进行深度估计。其中室内图像采用mono_640 \times 192模型,室外图像采用stereo_1024 \times 320模型;采用Sobel算子对图像进行边缘梯度提取,算子模板大小为 3×3 ;

步骤S2-2,将深度数据和边缘数据量化为 $(x, y, \text{depth}, \text{edge})$ 的四维元组整列,其中 x, y 代表像素点的位置偏移,depth代表像素点的深度值,edge代表边缘梯度值,按照聚簇数 K 为3的weighted-kmeans聚类算法,找到聚类中心作为三个多焦距图像的聚焦中心。为同时考虑像素深度信息与像素位置上的关系,聚类维度划分为像素点深度值、像素点 x 偏移、像素点 y 偏移和边缘梯度并固定权重配比,在三个维度上的权重配比为7:1:1:1。最终的结果是将深度图在空间上分层,并将获得三个分类聚簇和三个聚类中心

步骤S2-3,根据聚焦中心的深度值与图像上任一像素点之间的深度差值,决定图像上该点的高斯模糊核,并对该点处进行高斯模糊,从而分别得到三张不同聚焦中心的多

焦距图像。计算仿真高斯模糊核的过程, 主要根据聚焦中心的深度与离焦区域的深度差决定。离焦扩散函数 (PSF) 由模拟深度估计确定, 计算如下:

$$H_{x,y} = \begin{cases} \frac{1}{\pi\alpha(D_i - D_j)^2}, & \text{if } \sqrt{x^2 + y^2} \leq \beta|D_i - D_j| \\ 0, & \text{Otherwise} \end{cases}$$

其中 D_i, D_j 获得分别表示两个深度聚类中心的深度值, x, y 表示图像像素点坐标。由于采用步骤S2-1中的算法估计的深度, 是以范围(0,1)浮点数表示, 采用 $\alpha\beta$ 作为适配参数, 其中 α 取值5.5, β 取值30。对清晰原图做三次高斯层次模糊, 每次层次模糊采取分割区独立高斯核滤波的方式。整体算法流程如下:

Procedure_Synthesis (F, M, K)

```

res = []
for i in K
  for  $M_j$  in M
    if  $i = j$  then
       $I_{M_j} \leftarrow F_{M_i}$ 
    else
       $I_{M_j} \leftarrow F_{M_i} \otimes H_{ij} + Z$ 
   $I \leftarrow I \cup I_{M_j}$ 
  Res. append(I)
return res

```

其中 F 表示步骤S2-1中的清晰图像, M 表示使用深度估计后聚类获得的掩膜图, N 表示需要生成的多焦距图像数量, 此处 K 设为3。 M_j 表示掩膜图上深度聚类第 j 类深度区域, F_{M_i} 表示深度聚类第 j 类深度区域对应到原始图像上的区域, I_{M_j} 表示图像 I 在深度聚类第 j 类深度区域对应的区域生成的离焦模糊图。当 $i=j$ 时表示聚焦中心与当前深度一致, 则图像显示为清晰, 否则, I_{M_j} 像素内容由原始图像根据离焦扩散 H_{ij} 做高斯滤波, 最终生成的图像 I 由各个深度聚类区域的离焦模糊图并联联合。其中 Z 表示均值为0方差为0.0001的随机高斯噪声, 以便更真实的模拟拍摄模糊噪声。算法完成后, 最终返回构建好的 N 张多焦距图像。

[0030] 步骤S3, 利用S2构建的数据集对步骤S1构建的网络模型进行训练。从 K 张多焦距图像中随机选取 N 张不同离焦模糊图作为模型输入, $N \leq K$, 模型的输出为预测的融合全聚焦图像, 标签为给定的原始图像。定义模型的损失函数分为结构相似性损失SSIM Loss与均方误差损失MSE Loss。损失函数公式如下, 其中 α 是平衡MSE损失与结构相似性损失的重要系数。

[0031] $\mathcal{L} = (1 - \mathcal{L}_{SSIM}) + \alpha \mathcal{L}_{mse}$

训练过程根据参考系数评估来调参。主要根据每一轮迭代训练后损失下降, 并用这一轮训练的模型在步骤2构建的数据集上验证相关系数。主要参考系数包含两个指标: 结构相似性系数SSIM和峰值信噪比PSNR。计算步骤如下公式:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

其中MSE计算了生成融合图像与原始图像之间的像素均方误差,公式中*i*、*j*表示像素点的横纵坐标,*m*、*n*表示了图像宽度和高度像素值,*I*、*K*表示了原始图像与融合后的图像。PSNR在均方误差的基础上做了对数计算,值越大说明图像质量越好,通常接近40就基本与原图接近,公式中MAX_I表示像素颜色值的最大值。SSIM是结构相似性系数,评估了生成图像与原图像之间的内容相似度,数值范围在0~1,约接近1说明与原图约接近。其中参数

$\mu_x\mu_y$ 分别代表*x*、*y*的平均值, σ_x^2 和 σ_y^2 分别代表*x*、*y*的方差, σ_{xy} 代表*x*和*y*的协方差。

而 $c_1c_2c_3$ 分别为常数,避免分母为0带来的系统错误。

[0032] 步骤S4,利用S3训练好的网络模型进行多焦距图像融合的测试:利用复眼相机采集K张不同焦距的图像,利用SIFT算法对采集图像进行配准,得到配准后的K张多焦距图像,将配准后的K张图像输入模型,输出为融合图像。

[0033] 本文中所描述的具体实施例仅仅是对本发明精神作举例说明。本发明所属技术领域的技术人员可以对所描述的具体实施例做各种各样的修改或补充或采用类似的方式替代,但并不会偏离本发明的精神或者超越所附权利要求书所定义的范围。

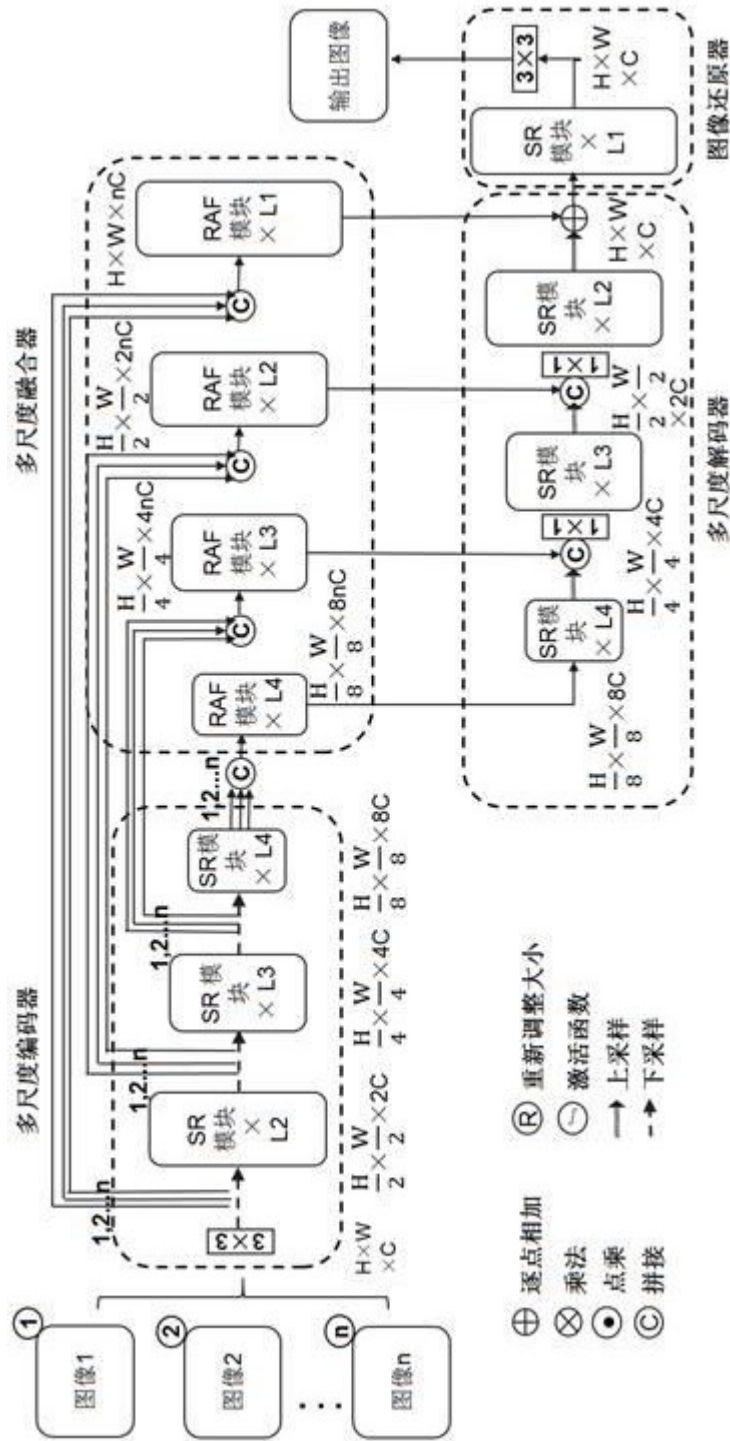


图1

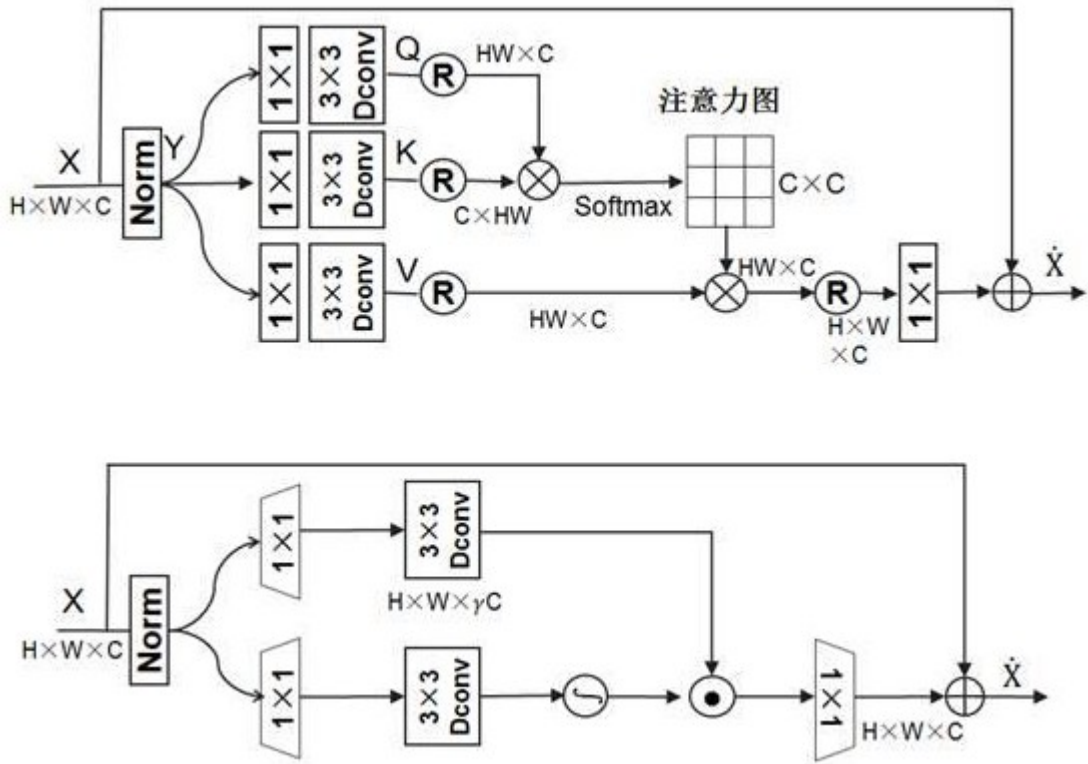


图2

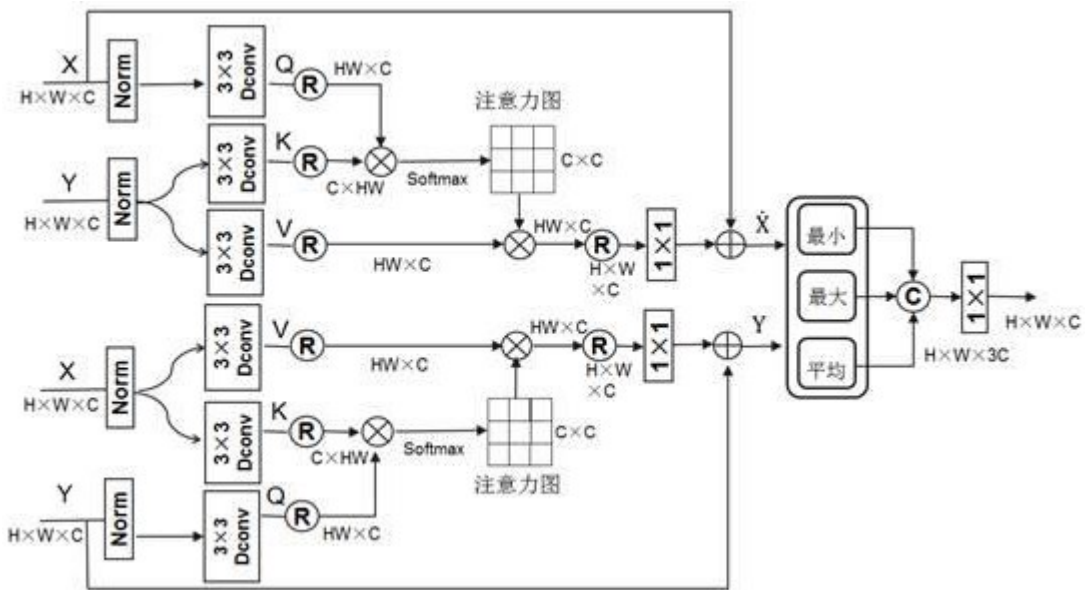


图3