



(12)发明专利申请

(10)申请公布号 CN 109360178 A
(43)申请公布日 2019.02.19

(21)申请号 201811208264.0

(22)申请日 2018.10.17

(71)申请人 天津大学

地址 300072 天津市南开区卫津路92号

(72)发明人 李素梅 薛建伟 刘人赫 侯春萍

(74)专利代理机构 天津市北洋有限责任专利代理事务所 12201

代理人 刘国威

(51)Int.Cl.

G06T 5/50(2006.01)

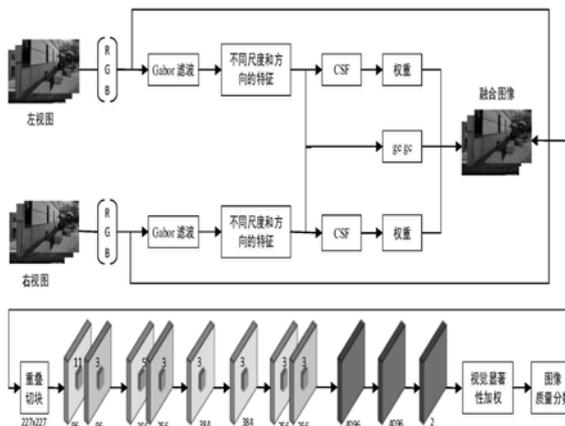
权利要求书2页 说明书9页 附图1页

(54)发明名称

基于融合图像的无参考立体图像质量评价方法

(57)摘要

本发明属于图像处理领域,为提出无参考立体图像质量评价方法,更符合人眼特性,与人类的主观感知保持良好的一致性。为此,本发明采取的技术方案是,基于融合图像的无参考立体图像质量评价方法,将立体图像的左右视图进行融合,使用重叠切块的方法送给神经网络Alexnet进行迁移学习训练,预测得到立体图像的质量,最后再利用人眼的视觉显著特性对融合图像进行加权。本发明主要应用于图像处理场合。



1. 一种基于融合图像的无参考立体图像质量评价方法,其特征是,将立体图像的左右视图进行融合,使用重叠切块的方法送给神经网络Alexnet进行迁移学习训练,预测得到立体图像的质量,最后再利用人眼的视觉显著特性对融合图像进行加权。

2. 如权利要求1所述的基于融合图像的无参考立体图像质量评价方法,其特征是,将立体图像的左右视图进行融合具体步骤如下:首先,左右视图分别在RGB三个通道上进行滤波器Gabor滤波以模拟人眼的视觉多通道特性,获取其不同尺度和方向的结构特征,随后通过对比敏感度函数CSF滤除图像的不重要的频率信息,最后通过公式(1)获得融合图像,在每个通道中,两只眼睛会互相施加增益控制,其施加的增益控制与其能量成正比,并且每只眼睛会对另一只眼睛产生的增益控制进行控制,即增益增强:

$$C(x,y) = \frac{1+\varepsilon_l}{1+\varepsilon_l+\beta_l} I_l(x,y) + \frac{1+\varepsilon_r}{1+\varepsilon_r+\beta_r} I_r(x,y) \quad (1)$$

其中, ε_i 和 β_i 分别用公式(2)和(3)表示为

$$\varepsilon_i = \frac{\sum_{n=0}^{n=jk} I_i^{jk}(x,y) \cdot \omega_{jk}}{ge_i} \quad i \in (l,r) \quad (2)$$

$$\beta_i = \frac{\sum_{n=0}^{n=jk} I_{i'}^{jk}(x,y) \cdot \omega_{jk}}{gc_i} \quad i \in (l,r), i' \in (r,l) \quad (3)$$

$C(x,y)$ 是融合后的图像; $I_l(x,y)$, $I_r(x,y)$ 分别是左视图和右视图; j 和 k 分别是6和8,表示使用6个尺度: $fs \in \{1.5, 2.5, 3.5, 5, 7, 10\}$ (cycles/degree),和8个方向: $\theta \in \{k\pi/8 | k=0, 1, \dots, 7\}$ 的滤波器Gabor提取图像特征; $\varepsilon_l, \varepsilon_r$ 是输入的左视图和右视图用于增益增强的视觉权重对比能量, β_l, β_r 是左视图对右视图的增益控制和右视图对左视图的增益控制; ge_i, gc_i 分别表示左视图对右视图的增益控制门限和右视图对左视图的增益控制门限;

$$C(x,y) = I_l(x,y) + I_r(x,y) \quad (4)$$

存在某一个确定的阈值,当左视图和右视图的对比刺激低于这个阈值时,融合图像是左右视图的线性叠加,用公式(4)来表示;反之,当左右视图的对比刺激高于这个阈值时,融合图像是左右视图非线性叠加形成的,如公式(1)。

3. 如权利要求1所述的基于融合图像的无参考立体图像质量评价方法,其特征是,利用神经网络Alexnet,修改其最后一层,改为二分类,该网络结构一共有十层,第一层为输入层,输入图像块;第二层到第六层为卷积层,输入图像经过卷积层提取一系列特征;第七层到第九层为全连接层,全连接层将学到的特征通过权值的方法映射到样本的标记空间;第十层为网络的输出,即小块的质量分数,该网络输入图像大小227x227。

4. 如权利要求1所述的基于融合图像的无参考立体图像质量评价方法,其特征是,更进一步地,神经网络Alexnet使用ReLU非线性激活函数;在卷积层和池化层后经过局部响应归一化层,实现局部抑制,加快网络收敛速度,提高网络的泛化能力;全连接层使用Dropout层,随机让网络某些隐含层节点的权重不工作,防止过拟合,采用三层全连接层。

5. 如权利要求1所述的基于融合图像的无参考立体图像质量评价方法,其特征是,利用人眼的视觉显著特性对融合图像进行加权具体步骤如下,采用具有各向异性的高斯核函数模拟注意力由中央向四周扩散的中央偏移CB因子:

$$CB(x, y) = \exp\left\{-\left(\frac{(x-x_0)^2}{2\sigma_h^2} + \frac{(y-y_0)^2}{2\sigma_v^2}\right)\right\} \quad (5)$$

$CB(x, y)$ 表示像素点 (x, y) 对中心点 (x_0, y_0) 的偏移信息, (x_0, y_0) 表示失真右视点的中心点坐标, (x, y) 为像素点坐标, σ_h 和 σ_v 分别表示图像水平方向和垂直方向的标准差;

通过公式 (5) 得到图像的中央偏移 CB 因子 $CB(x, y)$, 归一化 $CB(x, y)$ 得到图像对应的权值矩阵 $CB_{normal}(x, y)$, 用公式 (6) 表示, 其中 M 和 N 为图像的长和宽, (x, y) 为图像的像素位置; 将归一化的权值矩阵按原始图像分块的方式进行分块处理并求和得到块归一权值 $CB_{normblock}(i)$, 用公式 (7) 表示, $CB_{normblock}(i)$ 为第 i 个图像块的权值; 将对应位置的权值矩阵与输入图像块质量相乘, 得到图像块的质量值, 再将一幅图像所有图像块的质量值相加, 即得到基于人眼视觉显著特性图像质量值 $value$, 其中 T 为图像分块的数量, 为 $value_{block}(i)$ 为图像块 i 的质量值, 用公式 (8) 表示

$$CB_{normal}(x, y) = \frac{CB(x, y)}{\sum_{x=1}^M \sum_{y=1}^N CB(x, y)} \quad (6)$$

$$CB_{normblock}(i) = \sum_{x \in block-i} \sum_{y \in block-i} CB_{normal}(x, y) \quad (7)$$

$$value = \sum_{i \in 1}^T CB_{normal}(i) \cdot value_{block}(i) \quad (8)$$

6. 如权利要求1所述的基于融合图像的无参考立体图像质量评价方法, 其特征是, 还包括验证步骤: 采用皮尔森线性相关系数 $PLCC$, 斯皮尔曼秩相关系数 $SROCC$ 和均方误差 $RMSE$ 指标作为衡量图像质量评价的尺度, $SROCC$ 和 $PLCC$ 越接近于 1, 表示模型性能越好; $RMSE$ 越接近于 0, 表示性能越好。

基于融合图像的无参考立体图像质量评价方法

技术领域

[0001] 本发明属于图像处理领域,涉及到立体图像质量评价方法的改进优化,以及人类视觉系统中的视觉显著性在立体图像质量客观评价中的应用。具体讲,涉及基于融合图像的无参考立体图像质量评价方法。

背景技术

[0002] 近年来,随着多媒体技术的发展,立体图像受到越来越多的关注,而立体图像在采集、压缩、传输、显示等过程中均会产生降质问题,立体图像的质量会直接影响人们的视觉感受,因此,如何有效地评估立体图像的质量成为立体图像处理和计算机视觉领域的关键问题之一。

[0003] 由于立体图像主观质量评价要求测试者在特定的情况下对立体图像进行打分得到,这类方法费时费力,成本较高,且极易受测试者主观情绪的影响,不适于实际应用;而立体图像客观质量评价方法是通过客观模型给出立体图像的分值,能够有效弥补主观评价方法带来的不足。

[0004] 根据对参考图像的依赖程度,立体图像客观质量客观评价可以分为全参考图像质量评价、半参考图像质量评价和无参考图像质量评价。全参考图像质量评价需要测试图像和参考图像的全部信息,半参考图像质量评价需要获得参考图像的部分信息,但是在实际环境中这些信息可能没有或者很难获得,因此,这两类方法的应用范围有限。相比而言,无参考图像质量评价方法仅利用失真图像来进行评价,更符合实际应用的需要。

[0005] 尽管平面图像的质量评价已取得显著成果,但对立体图像质量评价的研究尚不充分。最开始人们将平面图像质量评价方法[1-3]应用于立体图像的质量评价中,如峰值信噪比(PSNR)、均方误差(MSE)、结构相似度(SSIM)[4]等。由于该类方法未考虑立体图像的深度信息,因此不适合直接用于立体图像的质量评价。随后,一些文献提出基于双目视觉特性的立体图像质量评价方法。文献[5]提出了一种融合人类双目特性的立体图像评价方法,使之更符合双目特性。但是,目前对人类视觉系统的认知还很有限,传统的方法难以全面反映人类对于立体图像的视觉感受,因而研究人员采用能模拟人类大脑的神经网络方式进行立体图像质量评价,文献[6]提出了一种基于支持向量机(SVM)的立体图像质量评价方法。但是传统的机器学习方法需要人工选取立体图像特征,选取的特征不一定能够完全反映立体图像的质量情况,限制了立体图像质量技术的发展。2011年后,深度学习网络,特别是卷积神经网络快速发展[7]。卷积神经网络(CNN)能够从图像中选取需要的特征,在图像分类、语音识别等方面能得到更高的正确率,比如2012年Hinton等人在计算机视觉识别挑战赛(ImageNet)上使用神经网络Alexnet[8]图像分类模型中夺得第一名,性能远超其它机器学习方法。基于卷积神经网络(CNN)在图像识别,对象检测和语义理解等领域的巨大成功,文献[9]提出了一种三通道五层卷积神经网络,网络初始化参数通过迁移学习训练二维图像得到,将立体图像的左视图、右视图和差值图分块分别作为网络输入,通过卷积提取立体图像特征,最终全连接加权得出最终的质量分数。文献[10]首先通过主成分分析法将左右视

图融合,然后对融合图像进行减均值和对比度归一化操作,最后采用非重叠切块的方法将图像切成小块送入网络进行训练,通过卷积神经网络(CNN)建立图像特征与主观评分差值之间的关系模型。

发明内容

[0006] 为克服现有技术的不足,本发明旨在提出无参考立体图像质量评价方法,更符合人眼特性,与人类的主观感知保持良好的一致性。为此,本发明采取的技术方案是,基于融合图像的无参考立体图像质量评价方法,将立体图像的左右视图进行融合,使用重叠切块的方法送给神经网络Alexnet进行迁移学习训练,预测得到立体图像的质量,最后再利用人眼的视觉显著特性对融合图像进行加权。

[0007] 将立体图像的左右视图进行融合具体步骤如下:

[0008] 首先,左右视图分别在RGB三个通道上进行滤波器Gabor滤波以模拟人眼的视觉多通道特性,获取其不同尺度和方向的结构特征,随后通过对比敏感度函数CSF滤除图像的不重要的频率信息,最后通过公式(1)获得融合图像,在每个通道中,两只眼睛会互相施加增益控制,其施加的增益控制与其能量成正比,并且每只眼睛会对另一只眼睛产生的增益控制进行控制,即增益增强:

$$[0009] \quad C(x,y) = \frac{1+\varepsilon_l}{1+\varepsilon_l+\beta_l} I_l(x,y) + \frac{1+\varepsilon_r}{1+\varepsilon_r+\beta_r} I_r(x,y) \quad (1)$$

[0010] 其中, ε_i 和 β_i 分别用公式(2)和(3)表示为

$$[0011] \quad \varepsilon_i = \frac{\sum_{n=0}^{n=jk} I_i^{jk}(x,y) \cdot \omega_{jk}}{g e_i} \quad i \in (l,r) \quad (2)$$

$$[0012] \quad \beta_i = \frac{\sum_{n=0}^{n=jk} I_{i'}^{jk}(x,y) \cdot \omega_{jk}}{g c_i} \quad i \in (l,r), i' \in (r,l) \quad (3)$$

[0013] $C(x,y)$ 是融合后的图像; $I_l(x,y)$, $I_r(x,y)$ 分别是左视图和右视图; j 和 k 分别是6和8,表示使用6个尺度: $f_s \in \{1.5, 2.5, 3.5, 5, 7, 10\}$ (cycles/degree),和8个方向: $\theta \in \{k\pi/8 | k=0, 1, \dots, 7\}$ 的滤波器Gabor提取图像特征; $\varepsilon_l, \varepsilon_r$ 是输入的左视图和右视图用于增益增强的视觉权重对比能量, β_l, β_r 是左视图对右视图的增益控制和右视图对左视图的增益控制; $g e_i, g c_i$ 分别表示左视图对右视图的增益控制门限和右视图对左视图的增益控制门限;

$$[0014] \quad C(x,y) = I_l(x,y) + I_r(x,y) \quad (4)$$

[0015] 存在某一个确定的阈值,当左视图和右视图的对比刺激低于这个阈值时,融合图像是左右视图的线性叠加,用公式(4)来表示;反之,当左右视图的对比刺激高于这个阈值时,融合图像是左右视图非线性叠加形成的,如公式(1)。

[0016] 利用神经网络Alexnet,修改其最后一层,改为二分类,该网络结构一共有十层,第一层为输入层,输入图像块;第二层到第六层为卷积层,输入图像经过卷积层提取一系列特征;第七层到第九层为全连接层,全连接层将学到的特征通过权值的方法映射到样本的标记空间;第十层为网络的输出,即小块的质量分数,该网络输入图像大小227x227。

[0017] 更进一步地,神经网络(Alexnet)使用ReLU非线性激活函数;在卷积层和池化层后

经过局部响应归一化层,实现局部抑制,加快网络收敛速度,提高网络的泛化能力;全连接层使用Dropout层,随机让网络某些隐含层节点的权重不工作,防止过拟合,采用三层全连接层。

[0018] 利用人眼的视觉显著特性对融合图像进行加权具体步骤如下,采用具有各向异性的高斯核函数模拟注意力由中央向四周扩散的中央偏移CB因子:

$$[0019] \quad CB(x, y) = \exp\left\{-\left(\frac{(x-x_0)^2}{2\sigma_h^2} + \frac{(y-y_0)^2}{2\sigma_v^2}\right)\right\} \quad (5)$$

[0020] $CB(x, y)$ 表示像素点 (x, y) 对中心点 (x_0, y_0) 的偏移信息, (x_0, y_0) 表示失真右视点的中心点坐标, (x, y) 为像素点坐标, σ_h 和 σ_v 分别表示图像水平方向和垂直方向的标准差;

[0021] 通过公式 (5) 得到图像的中央偏移CB因子 $CB(x, y)$, 归一化 $CB(x, y)$ 得到图像对应的权值矩阵 $CB_{normal}(x, y)$, 用公式 (6) 表示, 其中 M 和 N 为图像的长和宽, (x, y) 为图像的像素位置; 将归一化的权值矩阵按原始图像分块的方式进行分块处理并求和得到块归一权值 $CB_{normblock}(i)$, 用公式 (7) 表示, $CB_{normblock}(i)$ 为第 i 个图像块的权值; 将对应位置的权值矩阵与输入图像块质量相乘, 得到图像块的质量值, 再将一幅图像所有图像块的质量值相加, 即得到基于人眼视觉显著特性图像质量值 $value$, 其中 T 为图像分块的数量, 为 $value_{block}(i)$ 为图像块 i 的质量值, 用公式 (8) 表示。

$$[0022] \quad CB_{normal}(x, y) = \frac{CB(x, y)}{\sum_{x=1}^M \sum_{y=1}^N CB(x, y)} \quad (6)$$

$$[0023] \quad CB_{normblock}(i) = \sum_{x \in block \cdot i} \sum_{y \in block \cdot i} CB_{normal}(x, y) \quad (7)$$

$$[0024] \quad value = \sum_{i \in 1}^T CB_{normal}(i) \cdot value_{block}(i) \quad (8)$$

[0025] 验证步骤: 采用皮尔森线性相关系数 (PLCC), 斯皮尔曼秩相关系数 (SROCC) 和均方误差 (RMSE) 指标作为衡量图像质量评价的尺度, SROCC 和 PLCC 越接近于 1, 表示模型性能越好; RMSE 越接近于 0, 表示性能越好。

[0026] 本发明的特点及有益效果是:

[0027] 本发明提出了一种基于融合图像的无参考立体图像质量评价算法, 取得了较好的分类效果。网络模型通过使用图像融合算法, 更好地模拟了人眼处理立体图像的过程, 减少了数据量; 通过使用迁移学习, 减少了调参这个繁琐的过程, 大大缩短了训练的时间; 最后利用人眼的视觉显著性特性来处理分类出来的图像小块, 使其更符合人眼特性。结果表明本发明所提方法不论在对称和非对称立体图像数据库上较其它方法均取得了较好的结果, 与人类的主观感知保持良好的一致性。

附图说明:

[0028] 图1算法流程图。

[0029] 图2基于视觉显著性的光亮度图。

具体实施方式

[0030] 现有的方法很多都没有考虑到人眼的视觉显著特性,并且在图像进行分块时采用的均是不重叠切块方法,这种方法可能会造成图像结构信息丢失。另外,在机器学习和数据挖掘算法中,迁移学习可以避免从头开始搭建一个网络进行调参的繁琐性,并且使带有标签的数据得到充分利用。基于以上问题,本发明提出了一种基于融合图像的无参考立体图像质量评价方法,通过将立体图像的左右视图进行融合,使用重叠切块的方法送给神经网络(Alexnet)进行迁移学习训练,预测得到立体图像的质量,最后再利用人眼的视觉显著特性对融合图像进行加权。

[0031] 本发明内容主要包括以下3个方面:1)图像融合,分别对左右视图提取图像特征,然后融合来模拟人脑处理图像的过程;2)迁移学习,将神经网络(Alexnet)用于迁移学习,通过使用改进后的神经网络(Alexnet)对数据库进行训练,能得到更为准确的分类模型;3)视觉显著性,权重系数通过中央偏移因子进行确定,进一步模拟人眼视觉系统的显著性特征。

[0032] 本发明实验服务器CPU为3.5GHz的Intel xeon E5-2637v3,64G RAM,使用GPU并行加速,GPU为Titan X,显存12GB,Ubuntu 14.04系统,采用Caffe深度学习框架对网络进行训练。

[0033] 本发明采用LIVE实验库提供的对称和非对称立体图像测试库LIVE3D phase-I、LIVE3D phase-II进行评价实验。LIVE3D phase-I数据库对左右视点图像进行同样的失真处理,共有20种场景,包含5种失真,共有20对参考图像和365对失真图像;LIVE3D phase-II数据库是对LIVE3D phase-I数据库的完善,左右视点图像的失真程度不一定相同,共有8种场景,包含5种失真,共有8对参考图像和360对失真图像。在本发明的模型中,挑选了80%的融合图像作为训练,剩余的20%作为测试,所有的融合图像被切成227x227大小的小块。

[0034] 下面结合技术方案详细说明本方法:

[0035] 一、图像融合

[0036] 与平面图像相比,立体图像包含更多的视觉信息,能给观看者带来沉浸式的视觉体验。立体图像左右视图有所不同,通过视网膜传入人脑后,存在视觉多通道现象,同时也会产生双目融合和双目竞争[11],为了更好地模拟人眼视觉特性,本发明提出了一种融合图像的方法。

[0037] 首先,左右视图分别在RGB三个通道上进行滤波器(Gabor)滤波以模拟人眼的视觉多通道特性,获取其不同尺度和方向的结构特征,随后通过对比敏感度函数(CSF)滤除图像的不重要的频率信息,最后通过公式(1)获得融合图像。在每个通道中,两只眼睛会互相施加增益控制,其施加的增益控制与其能量成正比,并且每只眼睛会对另一只眼睛产生的增益控制进行控制,即增益增强。

$$[0038] \quad C(x, y) = \frac{1 + \varepsilon_l}{1 + \varepsilon_l + \beta_l} I_l(x, y) + \frac{1 + \varepsilon_r}{1 + \varepsilon_r + \beta_r} I_r(x, y) \quad (1)$$

[0039] 其中, ε_i 和 β_i 分别用公式(2)和(3)表示为

$$[0040] \quad \varepsilon_i = \frac{\sum_{n=0}^{n=jk} I_i^{jk}(x, y) \cdot \omega_{jk}}{g \varepsilon_i} \quad i \in (l, r) \quad (2)$$

$$[0041] \quad \beta_i = \frac{\sum_{n=0}^{n=jk} I_i^{jk}(x, y) \cdot \omega_{jk}}{gc_i} \quad i \in (l, r), i' \in (r, l) \quad (3)$$

[0042] $C(x, y)$ 是融合后的图像; $I_l(x, y)$, $I_r(x, y)$ 分别是左视图和右视图; 本发明中 j 和 k 分别是 6 和 8, 表示使用 6 个尺度 ($fs \in \{1.5, 2.5, 3.5, 5, 7, 10\}$ (cycles/degree)) 和 8 个方向 ($\theta \in \{k\pi/8 | k=0, 1, \dots, 7\}$) 的滤波器 (Gabor) 滤波器提取图像特征; $\varepsilon_l, \varepsilon_r$ 是输入的左视图和右视图用于增益增强的视觉权重对比能量, β_l, β_r 是左视图对右视图的增益控制和右视图对左视图的增益控制; ge_i, gc_i 分别表示左视图对右视图的增益控制门限和右视图对左视图的增益控制门限。

$$[0043] \quad C(x, y) = I_l(x, y) + I_r(x, y) \quad (4)$$

[0044] 本发明提出的融合图像方法是自适应的。存在某一个确定的阈值, 当左视图和右视图的对比刺激低于这个阈值时, 融合图像是左右视图的线性叠加[11], 可用公式(4)来表示; 反之, 当左右视图的对比刺激高于这个阈值时, 融合图像是左右视图非线性叠加形成的, 如公式(1), 不过, 自然场景下的所有图像都是高于这个阈值的。无论左视图和右视图的对比度刺激如何, 总能找到合适的 ge_i 和 gc_i , 使其具有显著的增益控制和增益增强, 所以本发明所提出的方法既适合对称失真也适合非对称失真。

[0045] 二、神经网络 (Alexnet) 用于迁移学习

[0046] 机器学习技术在分类、回归和聚类上取得了巨大的成功, 但使用此类方法有一个重要的前提是: 训练和测试数据必须从相同的特征空间去取得, 且必须具有相同的分布。当分布改变时, 就需要重新构造训练数据模型。在现实生活中, 重新收集训练数据或重新构造一个数据模型的成本很高甚至是不可能的, 而迁移学习可以充分利用带有标签的数据, 将已经学习到的模型参数通过某种方式来分享给新模型从而加快并优化模型的学习效率, 不用像大多数网络那样从零学习。迁移学习首先在[12]中被提出, 用于解决多任务学习的问题。

[0047] 基于卷积神经网络的立体图像质量评价方法通过在输入图像和输出质量值之间建立“端到端”的映射, 并以轻型结构达到较好的性能效果。本发明利用神经网络 (Alexnet), 修改其最后一层, 改为二分类。该网络结构一共有十层, 第一层为输入层, 输入图像块; 第二层到第六层为卷积层, 输入图像经过卷积层提取一系列特征; 第七层到第九层为全连接层, 全连接层将学到的特征通过权值的方法映射到样本的标记空间; 第十层为网络的输出, 即小块的质量分数。该网络输入图像大小 227x227, 各层参数如表1所示。

[0048] 表1网络参数

[0049]

Layer name	Configurations
Conv-1	Fm-96, kernel-11x11, stride-4
Pooling	Max, kernel-3x3, stride-2
Conv-2	Fm-256, kernel-5x5, pad-2, group-2
Pooling	Max, kernel-3x3, stride-2
Conv-3	Fm-384, kernel-3x3, pad-1
Conv-4	Fm-384, kernel-3x3, pad-1

Conv-5	Fm-256, kernel-3x3, pad-1, group-2
Pooling	Max, kernel-3x3, stride-2
Full-Connected-1	FC-4096
Full-Connected-2	FC-4096
Full-Connected-3	FC-2

[0050] 其中Layer name为每层的名字,包括卷积层(Conv-1、Conv-2、Conv-3、Conv-4、Conv-5)、池化层(Pooling)、全连接层(Full-Connected-1、Full-Connected-2、Full-Connected-3); Configurations为参数配置,Fm为特征图的数量,kernel为用于运算的核的大小,stride为运算的步长的大小,pad为对特征图尺寸扩展的大小,group-2为采用GPU并行加速,Max为池化层池化方法,FC为每个全连接层神经元的个数。

[0051] 该网络使用ReLU非线性激活函数,加快网络的收敛速度,防止了梯度消失的问题。在卷积层和池化层后经过局部响应归一化层,实现局部抑制,加快网络收敛速度,提高网络的泛化能力。全连接层使用Dropout层,随机让网络某些隐含层节点的权重不工作,有效的防止过拟合,采用三层全连接层实现更好的拟合效果。

[0052] 三、视觉显著性

[0053] 视觉心理物理学研究发现,在观看图像时人眼会不自觉的关注某些区域,并优先处理该区域的信息[13],这些区域就是显著性区域,这种特性称为视觉显著性。人眼对图像中央区域关注度比较高的成像机制,即中央偏移(Center Bias,CB)特性。人眼在观看图像时总是倾向于从图像的中心开始寻找视觉注视点,然后其注意力由中央向四周递减[14]。如图2所示的光亮度图,像素的坐标位置越处于图像的中间位置,该像素越容易受到关注;中间区域亮度最亮,表示人们对此区域最敏感,所分配的权重最高,亮度向四周逐渐变暗,权重也依次递减。

[0054] 可采用具有各向异性的高斯核函数[15]模拟注意力由中央向四周扩散的中央偏移(CB)因子:

$$[0055] \quad CB(x, y) = \exp\left\{-\left[\frac{(x-x_0)^2}{2\sigma_h^2} + \frac{(y-y_0)^2}{2\sigma_v^2}\right]\right\} \quad (5)$$

[0056] $CB(x, y)$ 表示像素点 (x, y) 对中心点 (x_0, y_0) 的偏移信息。 (x_0, y_0) 表示失真右视点的中心点坐标, (x, y) 为像素点坐标, σ_h 和 σ_v 分别表示图像水平方向和垂直方向的标准差。

[0057] 通过公式(5)得到图像的中央偏移(CB)因子 $CB(x, y)$,归一化 $CB(x, y)$ 得到图像对应的权值矩阵 $CB_{normal}(x, y)$,用公式(6)表示,其中M和N为图像的长和宽, (x, y) 为图像的像素位置;将归一化的权值矩阵按原始图像分块的方式进行分块处理并求和得到块归一权值 $CB_{normblock}(i)$,用公式(7)表示, $CB_{normblock}(i)$ 为第i个图像块的权值;将对应位置的权值矩阵与输入图像块质量相乘,得到图像块的质量值,再将一幅图像所有图像块的质量值相加,即得到基于人眼视觉显著特性图像质量值value,其中T为图像分块的数量,为 $value_{block}(i)$ 为图像块i的质量值,用公式(8)表示。

$$[0058] \quad CB_{normal}(x, y) = \frac{CB(x, y)}{\sum_{x=1}^M \sum_{y=1}^N CB(x, y)} \quad (6)$$

$$[0059] \quad CB_{nornblock}(i) = \sum_{x \in block-i} \sum_{y \in block-i} CB_{normal}(x, y) \quad (7)$$

$$[0060] \quad value = \sum_{i \in 1}^T CB_{normal}(i) \cdot value_{block}(i) \quad (8)$$

[0061] 四、性能比较与分析

[0062] 本发明采用皮尔森线性相关系数 (PLCC), 斯皮尔曼秩相关系数 (SROCC) 和均方误差 (RMSE) 指标作为衡量图像质量评价的尺度, SROCC和PLCC越接近于1, 表示模型性能越好; RMSE越接近于0, 表示性能越好。

[0063] 为了验证本发明方法的有效性, 将本发明方法与3种全参考立体图像质量评价方法和4种无参考立体图像质量评价方法进行了比较。对比的3种全参考立体图像质量评价方法为文献[16-18]。4种无参考立体图像质量评价方法为文献[10, 19, 21-22]。

[0064] 表2所示为本发明方法与对比方法在LIVE3D phase-I和LIVE3D phase-II数据库上的各项性能指标。

[0065] 表2各种评价方法的总体性能比较

Algorithm	phase-I			phase-II		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Shao [16]	0.932	0.927	5.941	0.836	0.819	6.196
MAD [17]	0.942	0.939	5.498	0.854	0.842	5.869
Jian [18]	0.953	0.947	4.991	0.911	0.905	4.665
Chen[19]	0.895	0.891	7.247	0.880	0.880	5.102
文献[21]	0.9446	0.9377	5.4569	0.9446	0.9019	5.5698
文献[22]	0.953	0.945	-	0.943	0.937	-
Sang[10]	0.9532	0.9444	-	0.9526	0.9467	-
Proposed	0.9861	0.993	0.0687	0.9863	0.9912	0.0767

[0067] PLCC, SROCC, RMSE三个指标均取得了非常好的结果, PLCC值超过0.986, SROCC值超过0.99, 这与神经网络 (Alexnet) 在分类任务上展现出的卓越的泛化能力不无关系。因为表1所示为本发明方法与对比方法在LIVE3D phase-I和LIVE3D phase-II数据库上的各项性能指标。PLCC, SROCC, RMSE三个指标均取得了非常好的结果, PLCC值超过0.986, SROCC值超过0.99, 这与AlexNet在分类任务上展现出的卓越的泛化能力不无关系。因为神经网络 (Alexnet) 在由数百万张图像组成的计算机视觉识别挑战赛 (ImageNet) 数据库上接受训练, 提取了几乎完整的基向量, 而构成图像的基向量是普适的, 所以该网络也可作为迁移学习的网络用于立体图像质量评价。另外, 无论是只有对称失真的LIVE3D phase-I数据库还是既有对称失真也由非对称失真的LIVE3D phase-II数据库, 本发明所采用的方法均适用。最后, 本发明的实验结果中的RMSE比文献的结果小很多, 这与图像的预处理过程有很大关系。本发明将所实验的图像库按照图像的好坏分为两类, 因此, 本发明的RMSE的值比其他文献实验结果更小。且现有文献并未给出其在实验过程中对图像处理的具体过程, 但不同的处理过程对实验结果有着重要的影响。

[0068] 表3为在LIVE3D phase-I和LIVE3D phase-II上分开失真类型做的实验, 该数据库包含5种失真, 分别是高斯模糊失真 (BLUR)、快速衰落失真 (FF)、JPEG压缩失真及JPEG2000压缩失真 (JP2K) 和白噪声失真 (WN)。从表中所列的实验结果数据来看, 本发明所提出的方法在不同失真类型上均展现出了非常好的效果, 有些失真类型的PLCC和SROCC甚至达到了

1, RMSE达到了0。但LIVE3D phase-I上的JPEG失真和LIVE3D phase-II上的FF失真,没有达到像其它失真一样的效果,说明本发明所提网络对于这两种失真的学习能力还有待提升。

[0069] 表3不同失真类型的PLCC、SROCC以及RMSE

PLCC						
phase-I	BLUR	FF	JPEG	JP2K	WN	ALL
Proposed	1	1	0.9161	1	1	0.9861
phase-II	BLUR	FF	JPEG	JP2K	WN	ALL
Proposed	1	0.874	1	1	1	0.9863
SROCC						
phase-I	BLUR	FF	JPEG	JP2K	WN	ALL
Proposed	1	1	0.9376	1	1	0.993
phase-II	BLUR	FF	JPEG	JP2K	WN	ALL
Proposed	1	0.9169	1	1	1	0.9912
RMSE						
phase-I	BLUR	FF	JPEG	JP2K	WN	ALL
Proposed	0	0	0.2085	0	0	0.0687
phase-II	BLUR	FF	JPEG	JP2K	WN	ALL
Proposed	0	0.2425	0	0	0	0.0767

[0070] 参考文献

[0072] [1]Carnec M,Le Callet P,Barba D.An image quality assessment method based on perception of structural information[C]//Image Processing,2003.ICIP 2003.Proceedings.2003International Conference on.IEEE,2003,3:III-185.

[0073] [2]Wang,Zhou,and Eero P.Simoncelli."Reduced-reference image quality assessment using a wavelet-domain natural image statistic model."Electronic Imaging 2005.

[0074] [3]Wang Z,Bovik A C.A universal image quality index[J].IEEE Signal Processing Letters, 2002,9(3):81-84.

[0075] [4]You H,Xing L,Perkis A,et al.Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis [C].Proceedings of International Workshop on Video Processing and Quality Metrics for Consumer Electronics,Scottsdale,AZ,USA,2010.

[0076] [5]Lin Y H,Wu J L.Quality assessment of stereoscopic 3D image compression by binocular integration behaviors[J].IEEE transactions on Image Processing,2014,23(4):1527-1542.

[0077] [6]Jung Y J,Sohn H,Lee S I,et al.Predicting visual discomfort of stereoscopic images using human attention model[J].IEEE Transactions on Circuits and Systems for Video Technology,2013, 23(12):2077-2082.

[0078] [7]Lee H,Pham P,Largman Y,et al.Unsupervised feature learning for audio classification using convolutional deep belief networks[C]//Advances in neural information processing systems. 2009:1096-1104.

[0079] [8]Alex Krizhevsky,Ilya Sutskever,and Geoffrey EHinton,"Imagenet classification with deep convolutional neural networks,"in Advances in neural information processing systems,pp. 1097-1105,2012

- [0080] [9]Zhang W,Qu C,Ma L,et al.Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network[J].Pattern Recognition,2016,59:176-187.
- [0081] [10]Q Sang,T Gu,C Li,X Wu.Stereoscopic Image Quality Assessment via Convolutional Neural Networks.International Smart Cities Conference,2017:1-2
- [0082] [11]Jian Ding,Stanley A.Klein and Dennis M.Levi.“Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model,”in Journal of Vision,2013.
- [0083] [12]Sinno Jialin Pan,and Qiang Yang.“A Survey on Transfer Learning.”IEEE Transactions on Knowledge andData Engineering,vol.22,no.10,pp.1345-1359,2010.
- [0084] [13]J.K.Tsotsos,S.M.Culhane,W.Y.K.Wai,Y.H.Lai,N.Davis,and F.Nuflo,“Modelling Visual Attention via Selective Tuning,”Artificial Intelligence,vol.78,no.1-2,pp.507-545,Oct. 1995.
- [0085] [14]P.Tseng,R.Carmi,I.G.M.Cameron,et al.Quantifying center bias of observers in free viewing of dynamic natural scenes[J].Journal of Vision,2009,vol.9,no.7.
- [0086] [15]Le Meur O.,Le Callet,P.,Barba,et al.A coherent computational approach to model bottom-up visual attention[J].Pattern Analysis and Machine Intelligence,IEEE Transactions on, 2006,28(5):802-817.
- [0087] [16]Shao F,Jiang G,Yu M,et al.Binocular energy response based quality assessment of stereoscopic images[J].Digital Signal Processing,2014,29:45-53.
- [0088] [17]Larson E C,Chandler D M.Most apparent distortion:full-reference image quality assessment and the role of strategy[J].Journal of Electronic Imaging,2010,19(1): 011006-011006-21.
- [0089] [18]Ma J,An P.Method to quality assessment of stereo images[C].Visual Communications and Image Processing (VCIP),2016.IEEE,2016:1-4.
- [0090] [19]Chen M J,Cormack L K,Bovik A C.No-Reference Quality Assessment of Natural Stereopairs[J].IEEE Transactions on Image Processing,2013,22(9):3379-3391.
- [0091] [20]Shao F,Lin W,Wang S,et al.Blind Image Quality Assessment for Stereoscopic Images Using Binocular Guided Quality Lookup and Visual Codebook [J].IEEE Transactions on Broadcasting,2015,61(2):154-165.
- [0092] [21]田维军,邵枫,蒋刚毅,等.基于深度学习的无参考立体图像质量评价[J].计算机辅助设计与图形学学报,2016,28(6):968-975.
- [0093] [22]李苗苗,桑庆兵.一种基于图像融合的无参考立体图像质量评价方法[J].光学技术,2017(1):25-32.

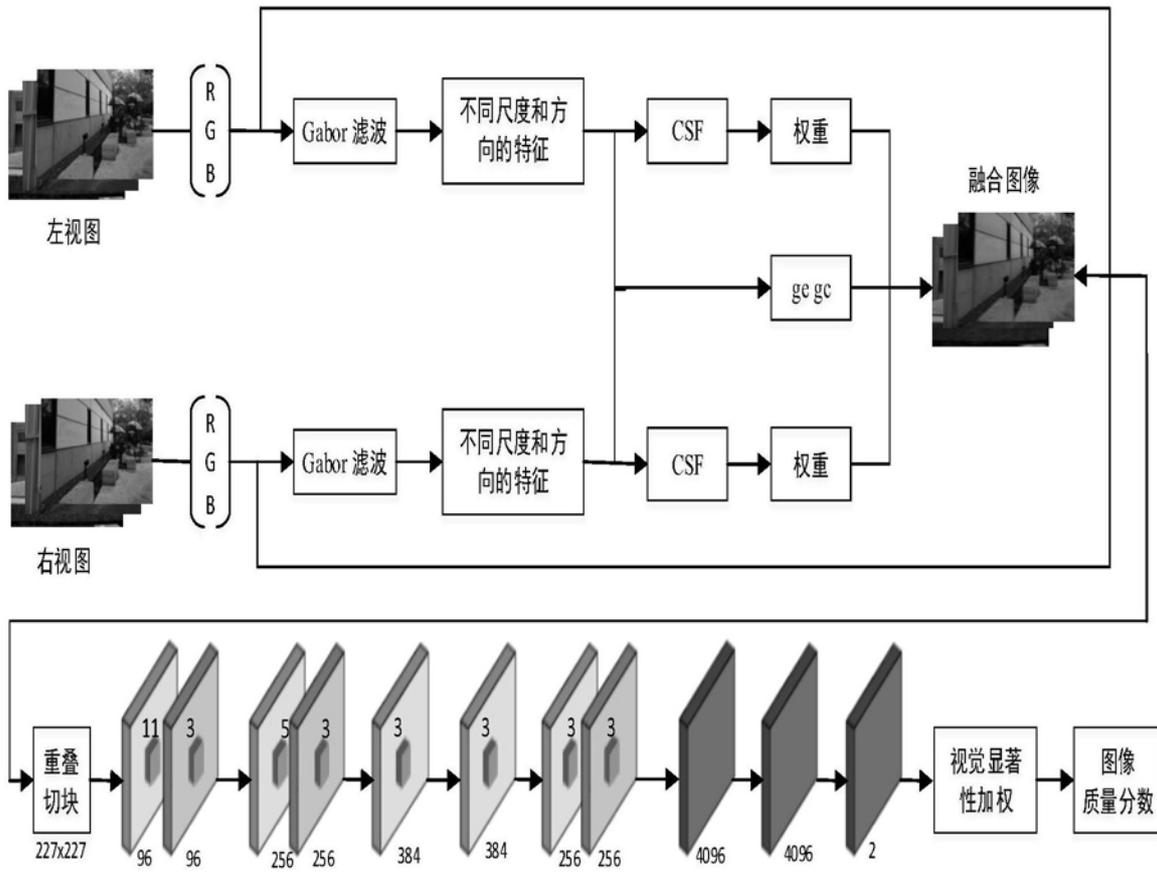


图1



图2