

文章编号:1007-2780(2019)11-1031-06

## 自由立体显示中基于深度卷积神经网络的 虚拟视点生成方法

付傲威<sup>1</sup>, 赵 敏<sup>1</sup>, 罗 令<sup>1</sup>, 邢 妍<sup>2</sup>, 邓 欢<sup>1</sup>, 王琼华<sup>2\*</sup>

(1. 四川大学 电子信息学院, 四川 成都 610065;

2. 北京航空航天大学 仪器科学与光电工程学院, 北京 100191)

**摘要:**传统虚拟视点生成采用像素填充法对生成的虚拟视点图像进行空洞填充和伪影修复,其修复效果无法满足自由立体显示需求。为了获取高质量的虚拟视点图像,提出了一种基于深度卷积神经网络的虚拟视点生成方法。该方法采用随机初始化的深度卷积神经网络作为图像先验,经过卷积神经网络结构的不断迭代,对虚拟视点图像的空洞和伪影进行修复,并将得到的高质量虚拟视点图像合成为自由立体图像,用于自由立体显示。修复后的虚拟视点图像的 PSNR 均值为 25.6,相比传统像素填充方法有明显提升。实验结果表明,所提方法能够实现高质量自由立体显示效果。

**关键词:**自由立体显示;虚拟视点生成;深度卷积神经网络;图像修复

**中图分类号:**TB133 **文献标识码:**A **doi:**10.3788/YJYXS20193411.1031

## Virtual viewpoint image generation method based on deep convolutional neural network in autostereoscopic display

FU Ao-wei<sup>1</sup>, ZHAO Min<sup>1</sup>, LUO Ling<sup>1</sup>,  
XING Yan<sup>2</sup>, DENG Huan<sup>1</sup>, WANG Qiong-hua<sup>2\*</sup>

(1. *School of Electronics and Information Engineering, Sichuan University,*  
*Chengdu 610065, China;*

2. *School of Instrumentation and Optoelectronic Engineering,*  
*Beihang University, Beijing 100191, China)*

**Abstract:** The conventional virtual viewpoint generation methods use the pixel filling algorithm to fill the image holes and restore artifacts on the generated virtual viewpoint images. However, the repaired virtual viewpoint images cannot meet the requirements of the autostereoscopic display. In order to obtain virtual viewpoint images with high quality, we propose a virtual viewpoint image generation method for autostereoscopic display based on deep convolutional neural network (DCNN). The method uses a randomly initialized deep convolutional neural network as an image prior, iterates through the structure of the convolutional neural network, and effectively repairs the image holes and

收稿日期:2019-08-22;修订日期:2019-09-02.

基金项目:国家重点研发计划(No.2017YFB1002900)

Supported by the National Key R&D Program of China(No.2017YFB1002900)

\* 通信联系人, E-mail: qhwang@scu.edu.cn

artifacts of the virtual viewpoint image. The repaired virtual viewpoint images are then synthesized into the autostereoscopic image for autostereoscopic 3D display. The average value of PSNR of the virtual viewpoint images is 25.6, which is higher than that of the conventional pixel filling method. The experimental results demonstrate the proposed method can obtain high quality autostereoscopic 3D images.

**Key words:** autostereoscopic display; virtual viewpoint image; deep convolutional neural network; image inpainting

## 1 引言

近年来,显示技术越来越重要,人们已经不再满足于传统的 2D(Two-dimensional,二维)显示,对于 3D(Three-dimensional,三维)显示的需求越来越迫切。裸眼 3D 显示无需借助眼镜或头盔等辅助设备即可重建出 3D 图像,引起了诸多研究人员的关注<sup>[1-5]</sup>。自由立体显示作为裸眼 3D 显示技术的一种,能够得到清晰的 3D 效果。自由立体显示需要获取 3D 图像,传统 3D 获取方法利用相机阵列对真实的目标场景进行拍摄,得到不同角度的视差信息,但存在成本过高以及标定、传输及存储困难的问题。当目标场景为虚拟场景时,可在 3D 建模软件中搭建相机阵列进行 3D 获取,但存在渲染时间过长和难以实现 3D 实时获取等问题<sup>[6-7]</sup>。利用深度图像绘制(Depth Image Based Rendering,DIBR)技术合成虚拟视点图像的方法具有成本低和绘制速度快等诸多优势,被认为是一种有效的 3D 获取方法<sup>[8-9]</sup>。基于 DIBR 的虚拟视点图像生成利用参考视点的纹理图像和深度图像,通过 3D Warping 变换获得特定位置的虚拟视点图像<sup>[8]</sup>。然而上述两种方法都受到深度图像精度和场景遮挡区域的影响,导致生成的虚拟视点图像存在明显的空洞和伪影,影响了自由立体图像的合成质量和自由立体显示效果。

深度卷积神经网络(Deep convolutional neural networks,DCNN)是一种新兴的图像生成和修复技术,该技术通过对样本图像的学习,可以实现图像超分辨率重建、图像去噪和图像修复<sup>[10-12]</sup>。Pathak 提出基于语义的深度卷积神经网络,通过学习图像特征,生成图像缺失部分的预测图,最终实现缺失图像的修复。该方法能够修复大面积图像空洞,但纹理信息失真严重,需要大量的样本图像对网络进行训练,训练成本高、耗时长<sup>[13]</sup>。Yang 提出一种

多尺度卷积神经网络,通过对图像特征和纹理特征进行学习,实现了高分辨率的图像修复。该方法能够还原图像纹理,但图像部分内容修复不全,需要大量训练集<sup>[14]</sup>。

本文提出了一种基于深度卷积神经网络的虚拟视点合成方法。该方法给定一个随机初始化神经网络作为虚拟视点图像先验,经过卷积神经网络结构不断迭代对生成的虚拟视点图像进行修复,能够有效去除虚拟视点图像的空洞和伪影,得到高质量的虚拟视点图像。该方法无需大量训练集,且耗时较短。合成的多幅虚拟视点图像可交织生成自由立体图像,用于自由立体显示。实验结果表明,所提方法能够有效提升虚拟视点图像的质量,应用在自由立体显示装置上,具有良好的自由立体显示效果。

## 2 方法原理

本文所提的基于深度卷积神经网络的虚拟视点合成方法的流程如图 1 所示,该方法包括 3 个过程:(1)虚拟视点图像生成过程:输入参考视点纹理图像和深度图像,进行 3D Warping 变换,得到多幅虚拟视点图像;(2)虚拟视点图像修复过

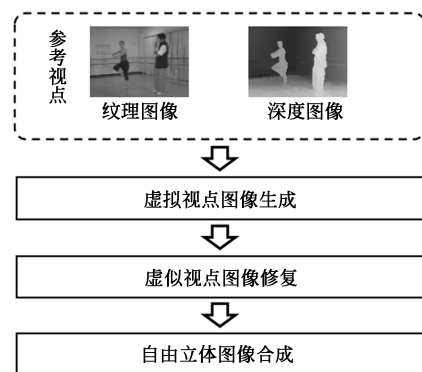


图 1 所提方法的流程图

Fig.1 Principle of the proposed method

程:利用深度卷积神经网络对生成的虚拟视点图像进行修复,得到无空洞和伪影的虚拟视点图像;  
(3) 自由立体图像合成过程:利用像素映射算法,将修复后的虚拟视点图像交织成自由立体图像,用于自由立体显示。

### 2.1 DIBR 虚拟视点图像生成

DIBR 虚拟视点图像生成是通过 3D Warping 变换生成目标视点视差图像的过程,其原理如图 2 所示。3D Warping 变换包括两个步骤:(1)利用已知的深度信息,将参考视点纹理图像反向投影到 3D 空间中;(2)将 3D 空间点重投影到虚拟成像平面,得到对应的虚拟视点图像。

参考视点图像的反向投影和重投影过程表示为:

$$(x, y, z)^T = R_r K_r^{-1} (u, v, 1)^T d(u, v) + t_r, \quad (1)$$

$$(l, m, n)^T = K_v R_v^{-1} \{ (x, y, z)^T - t_v \}, \quad (2)$$

其中,  $(x, y, z)$  表示三维空间点,  $(u, v, 1)^T$  为参考视点图像平面的像素齐次坐标,  $K_r, R_r, t_r$  分别表示参考视点对应摄像机的内部矩阵、旋转矩阵及平移矩阵。  $K_v, R_v, t_v$  分别表示虚拟视点图像对应摄像机的内部矩阵、旋转矩阵及平移矩阵,  $d(u, v)$  表示  $(u, v)$  位置对应的深度值,  $(l, m, n)^T$  表示虚拟视点位置, 对参考视点纹理图像执行反向投影和重投影两个操作, 即可生成目标位置的虚拟视点图像。

### 2.2 虚拟视点图像修复

受深度图像精度和场景遮挡区域的影响, 利用 DIBR 技术生成的虚拟视点图像存在明显的空洞和伪影。因此, 需要对虚拟视点图像进行修复, 去除空洞和伪影。所提方法利用一个随机初始化的卷积神经网络作为图像先验, 基于虚拟视点图

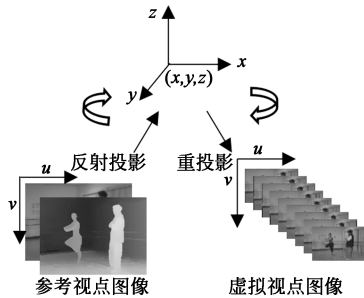


图 2 基于 DIBR 的虚拟视点图像生成原理示意图  
Fig.2 Schematic diagram of the virtual viewpoint image generation based on DIBR

像和任务依赖的观测模型, 通过不断迭代使模型参数逼近最大似然, 最终输出修复后的虚拟视点图像<sup>[10]</sup>。随机初始化卷积神经网络作为图像先验, 通过迭代修复虚拟视点图像, 无需大量的训练图像, 因此大大节省了资源 and 时间。

所提方法从随机分布中采样虚拟视点图像数据, 并基于该数据分布限制受损观测样本分布, 从而对虚拟视点图像进行修复。图像修复任务可以被认为能量最小化问题, 表示为:

$$x^* = \min_x E(x; x_0) + R(x), \quad (3)$$

其中  $E(x; x_0)$  是一个任务依赖的数据项,  $x^*$  表示修复后的虚拟视点图像,  $x$  表示网络输出,  $x_0$  为未修复的虚拟视点图像,  $R(x)$  是一个正则化项, 用来抓取虚拟视点图像的通用先验。所提方法采用卷积神经网络来代替  $R(x)$ , 卷积神经网络的参数化表征形式为:  $x = f_\theta(z)$ ,  $x \in R^{3 \times H \times E}$ ,  $z \in R^{C \times H \times W}$ , 其中  $z$  为编码张量,  $\theta$  为网络参数。

虚拟视点图像恢复过程可以表示为:  $x^* = f_\theta^*(z)$ ,  $\theta^*$  为修复虚拟视点图像的最优网络参数, 寻找最优参数  $\theta^*$  的过程表示为:

$$\theta^* = \min_\theta \| f_\theta(z) - x_0 \|, \quad (4)$$

其值选取如下:

$$\theta^* = \operatorname{argmin}_\theta E(f_\theta(z); x_0). \quad (5)$$

在随机初始化卷积神经网络的基础上, 给定一个随机权重  $\theta_0$ , 采用随机梯度下降算法训练, 经过多次迭代, 得到最优参数  $\theta^*$ , 使生成的虚拟视点图像不断逼近初始的虚拟视点图像, 最终获得修复后的虚拟视点图像。

采用的卷积神经网络为编解码的网络结构, 编码过程包括卷积、下采样、批量标准化以及修正线性单元激活过程, 对生成的虚拟视点图像特定进行

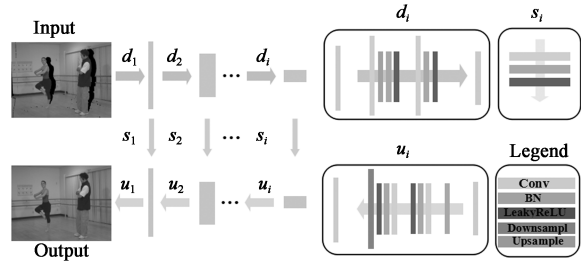


图 3 虚拟视点图像的修复过程示意图  
Fig.3 Schematic diagram of virtual viewpoint image inpainting

学习。解码过程包括批量标准化、卷积、修正线性单元激活和上采样过程,最终输出修复后的虚拟视点图像。虚拟视点图像的修复过程如图 3 所示,其中  $d_i$  表示第  $i$  次迭代的编码过程,  $u_i$  表示第  $i$  次迭代的解码过程,  $s_i$  表示第  $i$  次迭代的残差连接过程,将  $d_i$  层输出和  $u_i$  层输入连接在一起,实现跳跃式传递。经过一次编解码过程,完成一次迭代,多次迭代,最终输出修复后的虚拟视点图像。

### 2.3 自由立体图像合成

自由立体图像由多幅修复后的虚拟视点图像的像素按照视点矢量渲染算法映射生成<sup>[7]</sup>。合成的自由立体图像的子像素按照虚拟视点图像的视点顺序连续分布,自由立体图像的像素结构如图 4 所示,可以看出<sup>[15]</sup>,每幅虚拟视点图像像素已填充到单个柱透镜单元覆盖的全部子像素位置。

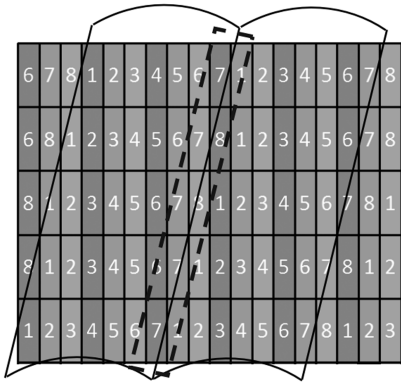


图 4 自由立体图像与虚拟视点的关系示意图

Fig.4 Schematic diagram of the relationship between autostereoscopic image and virtual view

## 3 实 验

实验采用微软亚洲研究院提供的“Ballet”图像集中一个视点的纹理图像和深度图像进行测试,虚拟视点图像生成过程如图 5 所示。

图 5(a)和图 5(b)分别为参考视点的纹理图像和深度图像,图 5(c)为合成的 8 幅虚拟视点图像,每幅图像的分辨率为  $1024 \times 768$ 。由细节放大图可以看出,受场景遮挡关系的影响,生成的虚拟视点图像中,人物和图像边缘存在明显的黑洞,背景区域存在大量细微的黑洞。由于视点 1 和 8 与参考视点 5 距离最远,其对应生成的虚拟视点图像黑洞区域最大。将生成的 8 幅虚拟视点图像分别输入深度卷积神经网络,对虚拟视点图像进

行不断迭代,每 50 次迭代输出一幅对应的虚拟视点图像,迭代过程如图 6 所示。由图中可以看出,当迭代次数  $n=400$  次时,虚拟视点图像的空洞和伪影得到了很好的去除。

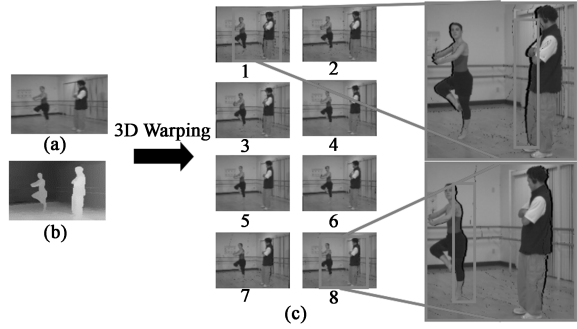


图 5 参考视点和虚拟视点图像。(a)参考视点纹理图像;(b)参考视点深度图像;(c)虚拟视点图像。

Fig.5 Reference view and virtual viewpoint images. (a) Reference view texture image; (b) Reference view depth image; (c) 8 virtual view images.

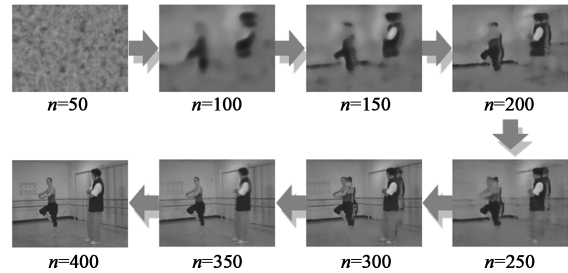


图 6 迭代过程输出的虚拟视点图像

Fig.6 Output virtual viewpoint images during iteration process

8 幅未修复和修复后的虚拟视点图像如图 7 所示。图 7(a)为利用 3D Warping 变换直接生成的虚拟视点图像,图 7(b)为经过深度卷积神经网络修复后的虚拟视点图像。由图 7(a)和 7(b)的细节放大图对比可以看出,修复后的 1 和 8 视点图像的空洞区域得到了很好的修复,同时背景区域中小的空洞也得到了填充。

为验证所提方法的有效性,将修复后的虚拟视点图像与传统像素填充方法修复后的虚拟视点图像进行了对比。不同虚拟视点个数下,两种方法的 PSNR(峰值信噪比)值对比结果如图 8 所示。本文所提方法中,修复后的虚拟视点图像的 PSNR 均值为 25.6,而传统像素填充方法修复后

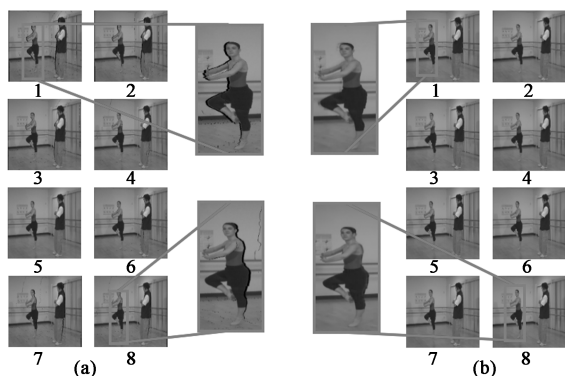


图 7 未修复和修复后的虚拟视点图像。(a)未修复的虚拟视点图像;(b)修复后的虚拟视点图像。

Fig.7 Unrepaired and restored virtual viewpoint images. (a) Unrepaired virtual viewpoint images; (b) Restored virtual viewpoint images.

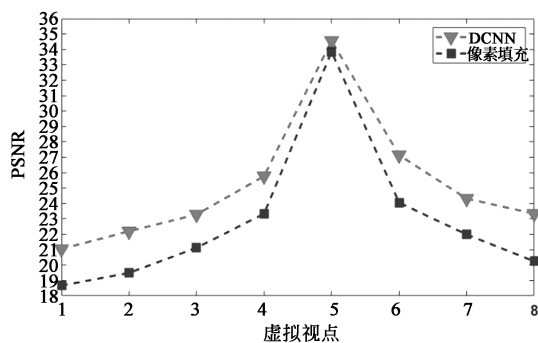


图 8 本文所提方法与传统方法修复的虚拟视点图像 PSNR 对比图

Fig.8 Comparison of PSNR of virtual viewpoint images inpainting by the proposed method and traditional method

的虚拟视点图像的 PSNR 均值为 22.5 之间。由实验结果可知,所提方法对虚拟视点图像的修复效果优于传统方法。

#### 参 考 文 献:

- [1] 王琼华.3D 显示技术与器件[M].北京:科学出版社,2011.  
WANG Q H. *3D Display Technology and Device* [M]. Beijing: Science Press, 2011. (in Chinese)
- [2] 王琼华,王爱红.三维立体显示综述[J].计算机应用,2010,30(3):579-581,588.  
WANG Q H, WANG A H. Survey on stereoscopic three-dimensional display [J]. *Journal of Computer Applications*, 2010, 30(3): 579-581, 588. (in Chinese)
- [3] ZHANG Y Q, WANG Q H, WANG A H, *et al.* Depth-map generation for multi-view autostereoscopic 3-D displays based on the sift algorithm constrained by boundary [J]. *Journal of the Society for Information Display*, 2010, 18(7): 513-518.

实验采用视点矢量渲染算法,将修复后的 8 幅虚拟视点图像合成为分辨率  $3840 \times 2160$  的自由立体图像,并在斜率为 0.181 8 的 8 视点自由立体显示器上进行显示,不同视角的再现 3D 图像如图 9 所示。可以看出,随着视角从左向右进行移动,芭蕾舞者手部与墙上贴图之间的距离逐渐增大。实验结果表明,所提方法生成的虚拟视点图像应用在自由立体显示器上,具有较好的 3D 显示效果。

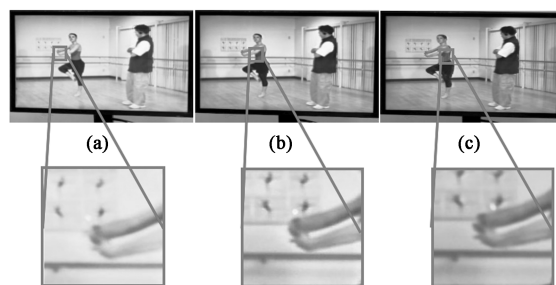


图 9 不同视角的再现 3D 图像。(a)左视角;(b)正视角;(c)右视角。

Fig.9 Reconstructed 3D images with different views. (a) Left view; (b) Positive view; (c) Right view.

## 4 结 论

本文提出了一种基于深度卷积神经网络的虚拟视点生成方法。该方法采用深度卷积神经网络对虚拟视点图像的空洞和伪影进行修复,并将修复后的虚拟视点图像通过视点矢量渲染算法合成自由立体图像,应用在自由立体显示器上,实现了较好的 3D 显示效果。实验结果表明,所提方法生成的虚拟视点图像质量相比传统方法有明显提升。

- [4] XIONG Z L, WANG Q H, LI S L, *et al.* Partially-overlapped viewing zone based integral imaging system with super wide viewing angle [J]. *Optics Express*, 2014, 22(19): 22268-22277.
- [5] DENG H, WANG Q H, LI D H. Method of generating orthoscopic elemental image array from sparse camera array [J]. *Chinese Optics Letters*, 2012, 10(6): 061102.
- [6] XIAO F, XIONG Z L, DENG H, *et al.* P-100: Approach on two-step 3D warping with background filling for multi-view autostereoscopic display [J]. *SID Symposium Digest of Technical Papers*, 2017, 48(1): 1624-1626.
- [7] 赵悟翔, 王琼华, 李大海, 等. 阶梯光栅多视点自由立体显示的子像素排列[J]. *四川大学学报(工程科学版)*, 2009, 41(6): 216-218, 236.
- ZHAO W X, WANG Q H, LI D H, *et al.* Sub-pixel arrangement for multi-view autostereoscopic display based on step parallax barrier [J]. *Journal of Sichuan University (Engineering Science Edition)*, 2009, 41(6): 216-218, 236. (in Chinese)
- [8] FEHN C. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV [C]//*Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*. San Jose, California, United States: SPIE, 2004: 93-104.
- [9] MORI Y, FUKUSHIMA N, YENDO T, *et al.* View generation with 3D warping using depth information for FTV [J]. *Signal Processing: Image Communication*, 2009, 24(1/2): 65-72.
- [10] LEMPITSKY V, VEDALDI A, ULYANOV D. Deep image prior [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 9446-9454.
- [11] CHAWLA N V, BOWYER K W, HALL L O, *et al.* SMOTE: synthetic minority over-sampling technique [J]. *Journal of Artificial Intelligence Research*, 2002, 16(1): 321-357.
- [12] BERTALMIO M, VESE L, SAPIRO G, *et al.* Simultaneous structure and texture image inpainting [J]. *IEEE Transactions on Image Processing*, 2003, 12(8): 882-889.
- [13] PATHAK D, KRÄHENBUHL P, DONAHUE J, *et al.* Context encoders: feature learning by inpainting [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016: 2536-2544.
- [14] YANG C, LU X, LIN Z, *et al.* High-resolution image inpainting using multi-scale neural patch synthesis [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 4076-4084.
- [15] 黄冠婷, 韩学辉, 龚晓婷, 等. 基于图像分割和区域匹配的灰度图像彩色化算法[J]. *液晶与显示*, 2019, 34(6): 619-626.
- HUANG G T, HAN X H, GONG X T, *et al.* Gray image colorization algorithm based on image segmentation and region matching [J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(6): 619-626. (in Chinese)

#### 作者简介:



付傲威(1995—),男,湖北襄阳人,硕士研究生,2017年于三峡大学获得学士学位,主要从事三维显示方面的研究工作。E-mail: Aowei\_fu@163.com



王琼华(1969—),女,四川巴中人,博士,教授,博士生导师,2001年于电子科技大学获得博士学位,主要从事信息显示方面的研究工作。E-mail: qhwang@scu.edu.cn