

# 适用于 SfM 点云的未标定摄像机注册方法

景彦哲<sup>1</sup> 刘 越<sup>1</sup> 田 鸿<sup>2</sup> 牛春风<sup>1</sup>

(1. 北京理工大学光电学院, 光电成像技术与系统教育部重点实验室, 北京, 100081

2. 北京理工大学计算机学院, 北京 100081)

**摘 要:** 无标注册是增强现实技术(AR)领域的研究热点, 本文提出并实现了一种对未标定摄像机进行内外参数估计的完整方法。该方法采用单幅图像作为输入, 三维物体的 SfM 重建点云作为数据表, 进行匹配并采用 RANSAC 方法去除误匹配得到单应矩阵, 结合 Kruppa 方程得到焦距估计值, 并采用 PnP 问题的 Levenberg-Marquardt 最优化法得到外参矩阵的非线性最小二乘解。为便于观察结果, 另建立基于显卡计算的粒子系统进行验证。实验结果表明, 该方法对于摄像机的内外参数估计具有较好的效果。

**关键词:** 注册; 未标定摄像机; Kruppa 方程; SfM; PnP 问题

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 1003-0530(2013)02-0274-05

## Registration method of uncalibrated camera applied to SfM point cloud

JING Yan-zhe<sup>1</sup> LIU Yue<sup>1</sup> TIAN Hong<sup>2</sup> NIU Chun-feng<sup>1</sup>

(1. Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education of China,

School of Optics and Electronics, Beijing Institute of Technology, Beijing 100081, China;

2. School of Computer Science&Technology, Beijing Institute of Technology, Beijing 100081)

**Abstract:** As a research focus of Augmented Reality, functional technology with respect to markerless registration has few implementations. In order to enrich the development of markerless registration and broaden the scope of application, a novel method based on SfM point clouds is presented to have both calibration and registration done, by which taking a single query image from an uncalibrated camera. A set of SfM point clouds is to match with the features of query image, and directly returns homographic matrix with wrong matches removed by RANSAC, focal length is to be acknowledged by Kruppa equation afterwards. Levenberg-Marquardt algorithm of PnP problem then is employed to calculate the rotation and translation of camera. Also we build a particle system based on GPU to observe our conclusion directly. Experimental results demonstrate the validity of estimation with this method.

**Key words:** calibration; uncalibrated camera; Kruppa equation; SfM; PnP problem

## 1 引言

增强现实技术 (Augmented Reality, AR) 是目前计算机视觉领域的研究热点。AR 技术通过在真实世界场景中叠加具备真实感的虚拟物体, 增强用户沉浸感与交互性。其效果很大程度上取决于注册算法, 即求解摄像机外参数方法的精度。根据是否基于特定标志物, 注册算法可以分为有标增强现实注册 (marker based registration) 与无标增强现实注册 (markerless reg-

istration)。其中无标注册由于算法实现难度大, 计算量大, 不易达到实时效果<sup>[1]</sup>。随着移动设备性能的不断增强, 商业化应用的需求日新月异, 无标 AR 技术开始更多面向大众需求。不仅注册精度的要求不断提高, 在表现形式方面, 平面的无标注册方法也展现了局限性, 基于立体结构的无标注册方法更易于被大众所接受。

Gordon 与 Lowe<sup>[2]</sup> 最早于 2006 年完成了基于多幅图像进行的摄像机姿态估计, 首先使用两两图像估计

空间点位置,并通过非线性最优化的方法求解所有参数,牺牲速度保证了注册的精度;Irschara<sup>[3]</sup>等完成了基于 SfM 点云数据的已标定摄像机快速定位识别,通过 GPU 加速来进行大范围城市建筑中的摄像机姿态实时估计。在国内,北京理工大学<sup>[4]</sup>提出了混合特征跟踪的户外注册方法,利用 GPS 定位缩小检索范围,通过关键帧的匹配计算摄像机姿态。

本文提出一种新的无标注册方法。该方法采用由运动到结构(Structure from Motion, SfM)的三维重建技术得到稀疏点云数据,与未标定摄像机的图像进行特征匹配;得到单应矩阵对内部参数进行快速估计,减少未知参数数量;之后采用非线性迭代算法对外部参数进行估计;建立粒子系统进行仿真演示,对结论加以验证。

## 2 算法介绍

基于 SfM 技术进行三维重建得到的点云数据(简称 SfM 点云)包含两部分内容,第一部分为图像序列对应的各摄像机参数信息,第二部分为点云的世界坐标、颜色、所在图像的编号、所在图下岗特征的编号以及特征点的二维亚像素图像坐标。

SfM 点云具备以下特征,使基于点云的无标注册成为可能:

- 点云数据中每一个点都是来自图像序列中 2 幅或更多图像的特征点,可以通过图像序列获得特征描述子;
- 点云数据包含建立模型时所采用图像序列中所有摄像机内部与外部参数;
- 点云数据的稀疏特性,保证了特征匹配过程消耗的时间较少。

因此,如果将一个未知摄像机采集的单幅图像与点云数据进行匹配,就能得到输入图像与建模图像序列的两视图关系,以及输入图像特征点二维图像坐标与真实物体三维世界坐标的一一对应,并由此计算该未知摄像机的简化内参矩阵与外参矩阵。

本文采用 Wu changchang<sup>[5]</sup>的 Visual Structure from Motion System 对图像序列进行 SfM 三维重建,得到稀疏点云数据与图像对应摄像机参数。之后,对需要求解的未标定摄像机采集的图像提取 Sift 特征,利用点云数据与输入图像的 Sift 特征描述子进行匹配,得到输入图像与最佳匹配图像中点的 2D-2D 特征匹配

对集合,以及输入图像与三维点云中点的 2D-3D 特征匹配对集合。将 2D-2D 匹配关系采用 RANSAC 算法进行处理,去除其中的误匹配,并得到单应矩阵,结合 Kruppa 方程解得内参矩阵的快速估计。利用解得的焦距估计值与已知的点云三维坐标及对应图像二维坐标,采用 PnP 问题的 Levenberg-Marquardt 最优优化法得到外参矩阵的非线性最小二乘解。为获得直观的演示效果,另建立基于 HLSL 加速的粒子系统对所得结果加以验证。

## 3 算法实现

### 3.1 读取点云数据

在点云数据之中,第一部分为建立模型的图像序列中,每幅图像分别对应的摄像机参数,其中包含以像素单位表示的摄像机焦距  $f$ ,以四元数表示的旋转矩阵  $\mathbf{R}$ ,以及摄像机位于世界坐标系(即点云模型坐标系)下的是三维位置  $\mathbf{C}$ 。第二部分为点云数据,即重建点在世界坐标系下的坐标  $\mathbf{X}$ ,在图像序列中出现的次数  $n$ ,以及在每幅图像中所在的特征点序号和二维图像坐标  $\mathbf{m}$ (以图像中心为主点)。将点云中每个三维点与图像序列中对应每幅图像的特征点进行标记,取得每个点对应的 Sift 特征描述子。

### 3.2 特征匹配

将未知摄像机采集的二维图像  $I_0$  作为输入,采用 Sift 算法获得图像 Sift 特征点在图像  $I_0$  上的二维图像坐标集合,并建立对应的 Sift 特征描述子集合。这里 Sift(Scale Invariant Feature Transform,尺度不变特征变换)算法以及下面进行特征匹配的 KD 树(k-dimension tree,k 维搜索树)算法采用了 Rob Hess<sup>[6]</sup>维护的 Sift 库加以实现。

使用 KD 树来进行点云与输入图像的特征描述子之间建立匹配关系。以点云特征描述子集合建立 KD 树,并将输入图像特征进行遍历,得到  $n$  个特征匹配对。

至此,得到了图像  $I_0$  的 Sift 特征点二维图像坐标与点云中对应点的三维世界坐标之间的 2D-3D 匹配关系,并且由于点云中所有点均来自于图像序列的 Sift 特征点,也得到了图像  $I_0$  的 Sift 特征点二维图像坐标与图像序列中各图像的 Sift 特征点二维图像坐标的 2D-2D 匹配关系。

### 3.3 求解内参

点云中与图像  $I_Q$  特征相匹配的点可能来自于多幅图像,选取出现次数最多的某幅图像作为最佳匹配图像  $I_p$ ;将来自于  $I_p$  的特征匹配对作为最小抽样点集,利用 RANSAC 算法进行筛选,去除误配点以提高精度,并计算出表示两幅图像  $I_p$  与  $I_Q$  所在平面的映射关系,即单应矩阵  $\mathbf{H}$ ,使两图像上的点满足关系:  $\mathbf{m}_Q = \mathbf{H}\mathbf{m}_p$ 。

根据基本针孔相机模型<sup>[7]</sup>,摄像机坐标  $\mathbf{X}_{cam} = (x, y, z)^T$  与二维图像坐标  $\mathbf{x} = (u, v, 1)^T$  的映射关系为  $\mathbf{x} = \mathbf{P}\mathbf{X}_{cam}$ ,其中,针孔模型的摄像机矩阵为  $\mathbf{P} = \mathbf{K}(\mathbf{R} | \mathbf{T})$ 。

$\mathbf{R}$  与  $\mathbf{T}$  为摄像机外参矩阵,分别表示摄像机模型在世界坐标系下的旋转矩阵与平移矩阵; $\mathbf{K}$  为摄像机标定矩阵,即摄像机内参矩阵,有限射影摄像机标定矩阵通常具备如下表示:

$$\mathbf{K} = \begin{bmatrix} f_x & s & p_x \\ & f_y & p_y \\ & & 1 \end{bmatrix} \quad (1)$$

其中,  $f_x$  与  $f_y$  分别表示以像素为单位的水平方向与垂直方向焦距,  $p_x$  与  $p_y$  分别表示主点(摄像机主轴与图像平面的交点)在图像中的横向与纵向坐标,  $s$  为扭曲参数,可以解释为摄像机中 CCD 阵列的横轴与纵轴“不垂直”的程度,通常可以认为  $s=0$ 。

另外  $f_x$  与  $f_y$  的比值为像素长宽比,通常非常接近 1,在这里认为  $f_x = f_y$ ;将主点偏置设置在图像中心位置,并且规定图像坐标系中心位置为原点,就有  $p_x = p_y = 0$ ,得到简化标定矩阵:

$$\tilde{\mathbf{K}} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} = \text{diag}(f, f, 1) \quad (2)$$

记  $\omega^* = \tilde{\mathbf{K}} \tilde{\mathbf{K}}^T$ ,  $\mathbf{e} = (e_1, e_2, e_3)^T$  表示对极点,在相差一个常数因子下,有 Kruppa 方程的矩阵表示<sup>[8]</sup>:

$$[\mathbf{e}_Q]_x \omega_Q^* [\mathbf{e}_p]_x^T = \mathbf{F} \omega_p^* \mathbf{F}^T \quad (3)$$

其中  $[\mathbf{e}]_x = \begin{bmatrix} e_3 & -e_2 \\ -e_3 & e_1 \\ e_2 & -e_1 \end{bmatrix}$  为反对称矩阵。

根据基础矩阵与单应矩阵的关系  $\mathbf{F} = [\mathbf{e}_Q]_x \mathbf{H}$ ,可以得到:

$$[\mathbf{e}_Q]_x = \mathbf{F}\mathbf{H}^{-1} \quad (4)$$

将(4)式代入(3)式,可以得到具备非常简单形式的解:

$$\begin{aligned} \omega_Q^* &= (\mathbf{F}\mathbf{H}^{-1})^{-1} \mathbf{F} \omega_p^* \mathbf{F}^T (\mathbf{F}\mathbf{H}^{-1})^{-T} \\ &= \mathbf{H}\mathbf{F}^{-1} \mathbf{F} \omega_p^* \mathbf{F}^T \mathbf{F}^{-T} \mathbf{H}^T \\ &= \mathbf{H} \omega_p^* \mathbf{H}^T \end{aligned} \quad (5)$$

$$\text{即} \begin{bmatrix} f_Q^2 & a & b \\ a & f_Q^2 & c \\ b & c & 1 \end{bmatrix} = \mathbf{H} \omega_p^* \mathbf{H}^T \quad (6)$$

其中,杂项  $a, b, c$  在理想情况下(即图像采集无噪声干扰,单应矩阵无误差的表示两图像所在平面中点的映射关系,标定矩阵  $\mathbf{K} = \tilde{\mathbf{K}}$ )均为零;而在实际情况中,图像采集存在噪声影响;Sift 特征点求取存在误差;以经过 RANSAC 处理的特征匹配对求取的单应矩阵近似表示两平面中点的映射关系;标定矩阵的简化忽略了部分参数的影响等,因此杂项不为零,得到的焦距估计值与实际焦距值近似却并不相等。同理,在之后求取外参数矩阵的过程中,得到的也是与实际摄像机不同的旋转矩阵与平移矩阵,但是具备与所求摄像机相同的 2D-3D 映射关系:

即 
$$\begin{cases} \mathbf{m} = \mathbf{K}(\mathbf{R} | \mathbf{T})\mathbf{X} \\ \mathbf{m} = \tilde{\mathbf{K}}(\mathbf{R}' | \mathbf{T}')\mathbf{X} \end{cases} \quad (7)$$

由于仅关心可以被观察到的屏幕坐标,这两个摄像机矩阵可以认为是等价的,即得到一个标定矩阵为  $\tilde{\mathbf{K}}$ ,外参矩阵为  $\mathbf{R}'$  与  $\mathbf{T}'$  的等价摄像机。

### 3.4 求解外参数矩阵

经过 RANSAC 算法去除误配点的特征匹配对已经得到,包含了点云的世界三维坐标与  $I_Q$  图像二维坐标,并且一一映射;在之前的工作中又完成了简化标定矩阵  $\tilde{\mathbf{K}}$  的计算,求解外参数矩阵  $\mathbf{R}$  与  $\mathbf{T}$  的问题被称为 PNP 问题(Perspective -N-Point)<sup>[9]</sup>。

使用 Levenberg-Marquardt 非线性最优化算法(简称 LM)来进行求解<sup>[10]</sup>。LM 是一种利用梯度求解使目标函数取得极值的参数向量的方法。这种优化方法要求使用一个近似的初始解作为起点,根据函数的梯度信息,找到变化率最大的方向,不断“爬升”直到函数极值点,经过多次迭代求精,得到精度较高的全局最优解。

在这里,未知数为外参数矩阵  $\mathbf{R}$  与  $\mathbf{T}$ ,其中  $\mathbf{R}$  的自由度为 3,  $\mathbf{T}$  的自由度也为 3,因此相当于求解未知数的数目为 6,方程数目为  $n$  的非线性方程组。使用线性方法将像点与物点的成像几何关系在齐次坐标下化为线性方程进行求解,将得到的解作为非线性优化的初始解,根据:

$$\mathbf{X}_{cam} = s\mathbf{K}^{-1}\mathbf{m} = (\mathbf{R} \mid \mathbf{T})\mathbf{X} \quad (8)$$

$$\text{即 } z_c \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & \mathbf{1} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (9)$$

建立以下优化函数作为 LM 算法的目标函数:

$$\sum_{i=1}^n \left[ \left( x_i - \frac{\mathbf{r}_1^T \mathbf{X} + t_x}{\mathbf{r}_3^T \mathbf{X} + t_z} \right)^2 + \left( y_i - \frac{\mathbf{r}_2^T \mathbf{X} + t_y}{\mathbf{r}_3^T \mathbf{X} + t_z} \right)^2 \right] \quad (10)$$

其中,  $x_i$  与  $y_i$  为像点坐标经过内参矩阵反变换得到的摄像机三维坐标,  $r_1, r_2, r_3$  分别为  $\mathbf{R}$  的列向量,  $t_x, t_y, t_z$  分别为  $\mathbf{T}$  的元素。经过多次迭代求精后, 得到关于  $\mathbf{R}$  与  $\mathbf{T}$  的全局最优解。

#### 4 仿真结果及讨论

仿真演示使用了微软基于 DirectX 框架的 XNA 建立 billboard(公告板, 即在三维空间绘制二维图像, 并保证其总面向摄像机) 粒子系统以观察。由于点云包含较大数据量, 如图 1 所示点云包含 4309 个点, 需绘制 4309 个图元, 若摄像机改变姿态, 就要重新计算各图元四顶点位置使其面向摄像机, 按照刷新率 60Hz, 则 CPU 每秒需要计算超过百万次, 并且每帧需要创建 8000 多个三角形, 并把顶点数据传送至 GPU, 将极大的拖慢运行速度。因此, 通过 HLSL(High Level Shader Language, 高级着色语言) 使用 GPU 代替 CPU 进行坐标的矩阵运算及四维颜色向量的计算<sup>[11]</sup>, 保证大数据量下的运行速度。

以下事项需要注意:

- 由于 XNA 与 VSfM 采用坐标系不同, 因此载入点云数据时需要注意符号;
- XNA 中摄像机旋转矩阵转为四元数表示;
- 摄像机在世界坐标系中的位置  $\mathbf{C}$  通过公式  $\mathbf{C} = -\mathbf{R}^{-1}\mathbf{T} = -\mathbf{R}^T\mathbf{T}$  进行计算;
- 在 DirectX 中, 摄像机的焦距并不直接给定, 而是通过视锥体(view frustum)来表示, 这种表示方法建立的摄像机焦距实际取决于程序窗体的大小, 因此需要根据窗口大小计算出仿真摄像机的焦距与算法解得的焦距之比值, 然后对图像或场景进行缩放。(根据公式

$s\mathbf{m} = \mathbf{K}(\mathbf{R} \mid \mathbf{T})\mathbf{X}$  并且  $\mathbf{K} = \tilde{\mathbf{K}} = \text{diag}(f, f, 1)$ , 在旋转矩阵与平移矩阵已经解出的情况下, 不同的焦距在屏幕上带来的表现等效于缩放效果。)

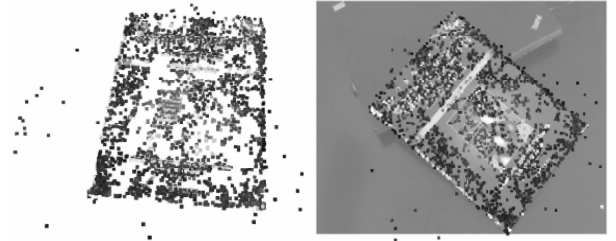


图 1 点云及摄像机参数仿真验证

Fig.1 Point cloud(left) and emulation of camera parameters recovery(right)

根据本文提出算法解得的参数为:

焦距估计: 2495. 343678

旋转矩阵:

$\mathbf{R} =$

$$\begin{bmatrix} 0.9917738416 & 0.0764514811 & -0.1026636168 \\ -0.1208453881 & 0.8236647383 & -0.5540512531 \\ 0.0422023621 & 0.5618999643 & 0.8261279506 \end{bmatrix}$$

平移矩阵与摄像机位置坐标:

$$\mathbf{T} = \begin{bmatrix} 0.2111585441 \\ 0.7022612246 \\ 0.9705216465 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} -0.1655147963 \\ -1.1399072698 \\ -0.3910080475 \end{bmatrix}$$

如图 1 所示, 结果与实际情况相匹配的程度较高, 与人眼直观观察结果相符, 可以满足应用需要。

以图 1(分辨率 1600x1200, 格式为 jpg, 大小为 971KB, 来自 Panasonic FX-180) 为例, 算法中各步骤时间代价如下表所示(单位: ms):

表 1 各步骤时间代价

Tab.1 Time consumption of each procedure

* 载入点云数据	4696
* 建立 KD 树	951
载入输入图像	234
#计算 Sift 特征点	11872
#特征匹配	8268
求解基础矩阵	<1
求解 RT 矩阵	15
杂项操作开支	16

带有 \* 标记的步骤仅需在程序初始化时载入; 而消耗时间最多的步骤中, 带有 # 标记的步骤均可在简单修改下, 通过使用 GPU 加速的 Sift 特征提取及匹配<sup>[12]</sup>来进行加速计算, 因此本文提出的方法在保证精度的前提下, 有望改良达到实时性的要求。

本文中除额外注明外, 均采用 OpenCV2.0 计算机视觉处理库实现。

## 5 结论

本文在研究前人成果的基础上,根据对点云数据使用的实际需要,提出了一种以未标定摄像机采集的单幅图像作为输入,点云数据作为数据库,对摄像机内部参数与外部参数进行估计的完整方法并加以验证。新的方法使用单应矩阵以及 Kruppa 方程来对内参矩阵进行快速估计,减少了摄像机参数的未知数数量,提高了迭代求解外参矩阵的速度;经过结论分析与仿真验证,无标注册的精度较高,并具备算法改良的空间,提高速度以满足实时性的可能,因此本文提出的算法流程能够较好的解决基于 SfM 点云的未标定摄像机注册问题,也是一种新的无标注册方法,为无标增强现实技术提供了一种新的解决方案,具备一定学术研究及商业应用价值。

## 致谢

本项研究工作受到国家自然科学基金(61072096)及国家科技部重大专项(2012ZX03002004)的资助,特此致谢。

## 参考文献

- [1] 李玉. 基于计算机视觉的增强现实实时跟踪注册算法及其性能的研究[D]. 北京:北京理工大学光电学院,2008.
- [2] Iryna Gordon, David G. Lowe, "What and where: 3D object recognition with accurate pose," in *Toward Category-Level Object Recognition*[M]. Springer-Verlag, 2006: 67-82.
- [3] Irschara A., Zach C. From Structure-from-Motion Point Clouds to Fast Location Recognition[C]. *IEEE: CVPR 2009*, 2009: 2599-2606.
- [4] 陈靖, 王涌天等. 适用于户外增强现实系统的混合跟踪定位算法[J]. *计算机辅助设计与图形学学报*, 2010, 22(2): 204-209.  
Chen Jing, Wang Yongtian. Hybrid Tracking for Outdoor Augmented Reality System[J]. *Journal of Computer-Aided Design&Computer Graphics*, 2010, 22(2): 204-209. (in Chinese)
- [5] Changchang Wu. VisualSFM: A Visual Structure from Motion System [OL]. <http://www.cs.washington.edu/homes/ccwu/vsfm/index.html>, 2011.
- [6] Rob Hess. An Open Source SIFT Library [C]. *ACM MultiMedia*, 2010.
- [7] Richard Hartley, Andrew Zisserman. *Multiple View Geometry in Computer Vision*[M]. Cambridge: Cambridge University Press, 2000:151-174.

- [8] Manmohan Krishna Chandraker. Two-View Focal Length Estimation for All Camera Motions Using Priors[R]. University of California, San Diego, CSE 252C December 2004.
- [9] 徐文立. 计算机视觉的 PNP 问题的最优解[J]. *自动化学报*. 1992, 18(5): 522-531.
- [10] Pradit Mittrapiyanuruk. A Memo on How to Use the Levenberg-Marquardt Algorithm for Refining Camera Calibration Parameters[OL]. [https://engineering.purdue.edu/kak/courses-i-teach/ECE661/HW5\\_LM\\_handout.pdf](https://engineering.purdue.edu/kak/courses-i-teach/ECE661/HW5_LM_handout.pdf). Nov 14, 2006.
- [11] Riemer Grootjans. XNA 3.0 Game Programming Recipes: A Problem-Solution Approach[M]. Apress, 2009: 231-252.
- [12] Changchang Wu. SiftGPU: GPU Implementation of Scale Invariant Feature Transform (SIFT)[OL]. <http://www.cs.unc.edu/~ccwu/siftgpu/>. 2011.

## 作者简介



景彦哲(1986-),男,山西太原人。北京理工大学光电学院,硕士研究生,主要研究方向为计算机视觉重建与人机交互技术。E-mail:kevin-ss30@163.com



刘越(1968-),男,博士,北京理工大学光电学院教授,光电信息技术与颜色工程研究所所长,光学工程、软件工程博士生导师,兼任中国图形图像学会理事、青年工作委员会委员,北京图形图像学会常务理事、青年工作委员会主任,中国兵工学会光学技术专业委员会委员,中国光学学会光电技术专业委员会委员,中国仪器仪表学会光电技术与系统集成分会理事。主要研究领域包括虚拟现实与增强现实人机交互技术、数字图像处理、微弱光电信号检测技术、计算机视觉等。E-mail:liuyue@bit.edu.cn



田鸿(1986-),男,山东青州人。北京理工大学光电学院,硕士研究生,主要研究方向为虚拟现实与增强现实。E-mail:tnhgcool@163.com



牛春风(1982-),男,天津人。北京理工大学光电学院,硕士研究生,主要研究方向为虚拟现实与增强现实。E-mail:334077842@qq.com