

基于深度传感器的远距增强现实注册技术研究

刘书曼¹, 葛水英¹, 徐士彪²

(1 中国科学院 文献情报中心, 北京 100190; 2 中国科学院 自动化研究所, 北京 100190)

摘 要: 通过标定摄像头和 Kinect 传感器, 结合总体最小误差算法, 提出了基于深度传感器的跟踪注册技术, 以实现远距离无标记的跟踪注册. 实验验证了虚拟物体以不同距离在近景模式注册和远景模式注册中叠加的效果. 实验结果表明该技术比现有跟踪注册技术方便, 且更符合远距离增强现实应用, 具有很强的应用价值.

关键词: 增强现实; Kinect; 无标记跟踪注册; 远距离; 自然交互

中图分类号: TP399

文献标识码: A

文章编号: 1000-7180(2017)05-0089-06

The Research on the Long Distance of Marker-less Registration Based on Depth Sensor in Augmented Reality

LIU Shu-man¹, GE Shui-ying¹, XU Shi-biao²

(1 National Science Library, Chinese Academy of Sciences, Beijing 100190, China;

2 Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: This paper collaborated with camera and Kinect hardware, combined with minimum error method, and proposed a registration technology based on depth sensor, and achieved long distance marker-less registration. Consequently, the experiment completed the virtual contents superposed in near mode and far mode. The results indicate the long distance marker-less registration is better than current registration, and prove this technology has very strong application value.

Key words: augmented reality; Kinect sensor; marker-less registration; long distance; natural interaction

1 引言

增强现实(Augmented Reality, AR)是指在用户所感知的真实环境中, 叠加由计算机生成的虚拟信息, 将原本在一定时空范围内难以体验的实体信息, 比如听觉、视觉、味觉与触觉等信息, 通过建模仿真后再叠加到真实环境中, 从而达到一种超越现实的体验^[1]. 常见的增强现实应用大多使用特定标记或自然图像作为注册的对象, 基于计算机视觉追踪, 虽然达到了一定的注册精度和实时性, 但存在较多局限性. 比如, 图片的识别距离受限, 无法满足大场景的增强现实应用, 易受光照变化和复杂背景的影响,

携带和摆放也存在不便, 同时缺少融入自然交互方式. 因此, 研究远距离的无标记跟踪注册技术和具有自然交互的增强现实具有很强的理论价值和学术价值. Kinect 传感器作为一款出色的体感外部设备, 能够捕获人的骨骼点信息, 不受光照变化和复杂背景的干扰^[2]. 将其应用于增强现实中, 不但能够实现远距离的跟踪注册, 不易受外在环境的干扰, 还可为人机交互带来新的机会.

2 相关研究工作

在增强现实的实现中, 追踪系统尤为重要, 常见的跟踪注册技术主要有机械式跟踪、电磁跟踪、全球

收稿日期: 2016-08-03; **修回日期:** 2016-09-06

基金项目: 国家科技支撑计划(2012BAH42F02); 文化产业重大专项“立体视觉产业技术应用示范基地”; 国家自然科学基金重点项目(61332017)

定位系统、超声跟踪、惯性跟踪、光学式跟踪、视觉跟踪以及混合跟踪^[3]. 按照跟踪原理的不同,可分成三大类:基于跟踪器跟踪、基于计算机视觉跟踪、混合跟踪^[3].

机械式跟踪利用机械装置的节点之间的长度和角度来定位各个节点,最早雏形由计算机图形学先驱 Sutherland 和他的学生在哈佛大学和犹他大学创造^[4]. 其不受光线、电磁、湿度等环境的影响,但设备大且笨重,不利于室外增强现实系统的应用.

电磁式跟踪将接收传感器安置在头盔上,随着使用者在电磁场中运动,计算机通过电磁感应原理可感知使用者运动及方位^[5]. 李启民等人介绍了自适应网格线性插值校正法^[6],对磁力跟踪器进行校正. 一般来说,磁场跟踪设备体积较小、价格适中,并且具备较好的实时性,但较易受其他金属物质或电磁干扰,造成采集空间位置数据的失真,对于大范围的跟踪,效果不佳.

超声跟踪通过计算传感器发出超声波脉冲到接收超声波脉冲的时间可以获得跟踪物的位置与朝向^[7]. 此追踪方法比较适合小范围对实时性要求不是特别高的 AR 应用,比如应用于医疗领域的超声引导下穿刺活检的乳腺手术^[8]. 在大范围跟踪时其精确度相对较低,容易受气压、湿度、温度等随机因素影响.

基于计算机视觉的追踪主要包括有标记和无标记的追踪. 有标记跟踪通过识别特定的标识进行三维注册,其追踪算法成熟,操作简单,实现方便. 但容易受光线、阴影、遮挡等影响,所使用的标识物不够具象,影响用户体验;无标记跟踪通过识别一些自然特征物进行三维注册,需事先捕获特征信息,计算开销较大,对于无特征的场景跟踪有限. 此算法成熟,实现方便,跟踪精度达到像素级别,误识率较低^[9].

混合追踪综合多种跟踪注册技术构成混合系统,适应不同环境以及注册精度要求,以克服单个技术的缺点,也是目前最有希望处理由一般室内或室外的移动 AR 环境带来的困难^[10]. 纽曼等人设计了加速度计的定位传感器结合视觉跟踪的混合跟踪定位系统,以此提高跟踪的有效性和鲁棒性^[11]. 但多传感数据的融合是难点,复杂度高. 可以看出:基于跟踪器的跟踪注册技术需要依靠硬件设备,容易受环境的影响,适用范围有限.

基于标记的跟踪注册技术虽然注册精度较高,但应用场景受限,当标识物被遮挡或是光线有变化,容易注册失败. 而基于自然特征无标记的三维注册,虽然应用范围较大,但存在注册误差,且误差会随时间叠加,对系统的消耗较大. 而且均无法实现远距离的跟踪注册.

混合跟踪技术虽然能结合基于跟踪器跟踪以及基于视觉跟踪的优势,扬长补短,但复杂性高.

本文利用 Kinect 传感器的深度传感,标定 Kinect 传感器与摄像头,提出基于深度传感的跟踪注册技术,弥补以上方式存在的某些缺陷,比如降低跟踪注册的复杂度,对环境的干扰具有相当的鲁棒性,而且能实现远距离的无标记跟踪注册,扩大增强现实系统的应用范围. 同时,本文还将根据 Kinect 传感器的特性,自定义手势交互,改变一般标记的增强现实系统较难引入人的互动的现状.

双目摄像头虽然可以计算深度信息,但实现算法略复杂,运行的稳定性较差,用于增强现实系统尚未成熟. 单目摄像头的增强现实效果,很难加入自然的人机交互方式. 同时,虽然单独使用 Kinect 传感器也可以实现增强现实,但 Kinect 彩色摄像头的分辨率不高,远距离的增强现实会因为图像清晰度不够,影响增强现实效果.

利用 Kinect 传感器与高清摄像头的配准,解决远距离图像清晰度问题、实现无标记的跟踪注册、实现自然的人机交互方式. 同时, Kinect 传感器和摄像头的配准,弥补了单独利用 Kinect 传感器存在纵向场景范围局限的缺陷.

图 1 展示了 Kinect 传感器结合摄像头实现增强现实的关键流程. 实现过程的关键和技术核心在于 Kinect 传感器坐标系与摄像头坐标系之间的标定.

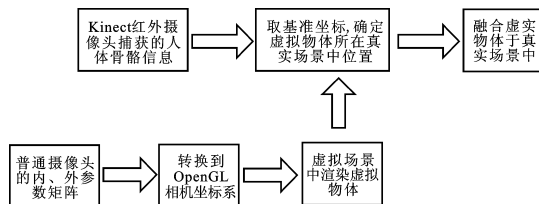


图 1 Kinect 传感器与摄像头实现增强现实关键过程

3 基于深度传感器的远距离跟踪注册

Kinect 传感器是一个外型类似网络摄影机的

装置. 它包括了一个 RGB 彩色摄像机, 一个红外线发射器和红外线 CMOS 摄像机所构成的 3D 结构光深度感应器. 利用这些传感器, Kinect 传感器能同时获取 RGB 图像和深度图像数据. 利用捕获的人体位置的三维数据, 可以进行人体动作的识别和跟踪, 无需佩戴一些像数据手套之类的辅助性工具, 减少人机交互的成本, 提高人机交互的自然性.

3.1 基于特征骨骼点匹配的图像配准技术

3.1.1 多坐标系统的标定

跟踪注册问题解决的关键在于增强现实系统中各坐标系之间的转换. 本文研究的增强现实系统与基于图片识别的增强现实系统的不同在于增加了 Kinect 传感器坐标系, 所以坐标系之间的转换相对复杂. 本文主要的 9 个坐标系, 包括了世界坐标系、物体坐标系、两个设备坐标系、Kinect 传感器坐标系、摄像头坐标系、虚拟场景坐标系和两个屏幕坐标系. 为实现 Kinect 传感器和摄像头之间信息的实时传递, 需要对二者坐标系进行标定. 图 2 展示了 Kinect 传感器坐标系和摄像头坐标系的标定过程, 即坐标系之间的仿射变换.

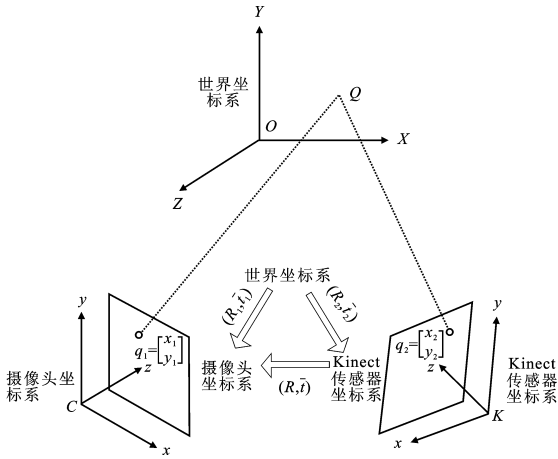


图 2 本研究增强现实系统多坐标系变换

世界坐标系下的三维点 Q , 在摄像头坐标系下和 Kinect 传感器坐标系下映射到屏幕上的二维点分别为 q_1 和 q_2 . 具体变换过程如下所示.

点 Q_c 为点 Q 在摄像头坐标系下的位置, q_1 为 Q_c 投影在屏幕的二维点. 则根据内参矩阵的定义可得到:

$$q_1 = H_{i1} \times Q_c = \begin{bmatrix} f_{x1} & 0 & c_{x1} \\ 0 & f_{y1} & c_{y1} \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$$

式中, $Q_c = (x_c \ y_c \ z_c)^T$, H_{i1} 为摄像头的内参矩阵. 同时, $R_1 = \begin{bmatrix} r_{c1} & r_{c2} & r_{c3} \\ r_{c4} & r_{c5} & r_{c6} \\ r_{c7} & r_{c8} & r_{c9} \end{bmatrix}$ 为其外参旋转矩阵, $t_1 = [t_{cx} \ t_{cy} \ t_{cz}]^T$ 为其外参平移向量, 于是根据外参的定义:

$$H_c = H_{i1}^{-1} \times q_1 \quad (1)$$

$$H_c = R_1 \times Q + t_1 \quad (2)$$

同理, 点 Q_k 为点 Q 在 Kinect 传感器坐标系下的位置, q_2 为 Q_k 的投影坐标.

$$\text{则: } q_2 = H_{i2} \times Q_k = \begin{bmatrix} f_{x2} & 0 & c_{x2} \\ 0 & f_{y2} & c_{y2} \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}.$$

其中, $Q_k = (x_k \ y_k \ z_k)^T$, H_{i2} 为 Kinect 传感器的内参矩阵. 同时, $R_2 = \begin{bmatrix} r_{k1} & r_{k2} & r_{k3} \\ r_{k4} & r_{k5} & r_{k6} \\ r_{k7} & r_{k8} & r_{k9} \end{bmatrix}$ 为其外参

旋转矩阵, $t_2 = [t_{kx} \ t_{ky} \ t_{kz}]^T$ 为其外参平移向量, 于是:

$$Q_k = H_{i2}^{-1} \times q_2 \quad (3)$$

$$Q_k = R_2 \times Q + t_2 \quad (4)$$

由于刚体变换具有良好的旋转不变性和平移不变性, 通过旋转矩阵和平移向量, 可以将空间中任意一个坐标系转换到另一个任意坐标系, 则 Kinect 传感器坐标系下的点 Q_k 可以通过某个特殊的旋转矩阵和平移向量转换到摄像头坐标下的点 Q_c , 即:

$$Q_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \times Q_k + t = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \times \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (5)$$

R 为旋转矩阵, t 为平移向量, 同时根据式(4)可以得到式(6), 根据式(6)和式(2)得到式(7):

$$Q = R_2^{-1} \times (Q_k - t_2) \quad (6)$$

$$Q_c = R_1 \times R_2^{-1} \times (Q_k - t_2) + t_1 \quad (7)$$

根据式(5)和式(7)的对应关系, 得到式(8)和式(9):

$$\mathbf{R} = \mathbf{R}_1 \times \mathbf{R}_2^{-1} = \begin{bmatrix} r_{c1} & r_{c2} & r_{c3} \\ r_{c4} & r_{c5} & r_{c6} \\ r_{c7} & r_{c8} & r_{c9} \end{bmatrix} \times \begin{bmatrix} r_{k1} & r_{k2} & r_{k3} \\ r_{k4} & r_{k5} & r_{k6} \\ r_{k7} & r_{k8} & r_{k9} \end{bmatrix}^{-1} \quad (8)$$

$$\bar{\mathbf{t}} = \bar{\mathbf{t}}_1 - \mathbf{R}_1 \times \mathbf{R}_2^{-1} \times \bar{\mathbf{t}}_2 = \begin{bmatrix} t_{cx} \\ t_{cy} \\ t_{cz} \end{bmatrix}$$

$$\begin{bmatrix} r_{c1} & r_{c2} & r_{c3} \\ r_{c4} & r_{c5} & r_{c6} \\ r_{c7} & r_{c8} & r_{c9} \end{bmatrix} \times \begin{bmatrix} r_{k1} & r_{k2} & r_{k3} \\ r_{k4} & r_{k5} & r_{k6} \\ r_{k7} & r_{k8} & r_{k9} \end{bmatrix}^{-1} \times \begin{bmatrix} t_{kx} \\ t_{ky} \\ t_{kz} \end{bmatrix} \quad (9)$$

于是, Kinect 传感器和摄像头坐标系之间的相对位置关系可以通过 $\mathbf{R}, \bar{\mathbf{t}}$ 确定。

理论上, Kinect 传感器与摄像头的相对位置保持不变, $\mathbf{R}, \bar{\mathbf{t}}$ 就保持不变。本文根据两摄像头采集的棋盘格图像对应点, 基于张正友法, 计算出 Kinect 传感器和摄像头的内外参数, 并根据式(8) 和式(9) 计算出原始 $\mathbf{R}, \bar{\mathbf{t}}$ 矩阵。由于误差的不可避免性, 每组对应图像所获得 $\mathbf{R}, \bar{\mathbf{t}}$ 值并不相同, 为了获得更好的精度, 为后续进一步迭代得到最优 $\mathbf{R}, \bar{\mathbf{t}}$, 需要设计初始 $\mathbf{R}, \bar{\mathbf{t}}$ 矩阵。本文提出总体误差最小算法, 见式(10), $\tilde{q}_i$ 为 q_i 的估计值。根据式(10) 得到初始迭代的 $\mathbf{R}, \bar{\mathbf{t}}$ 矩阵。

$$\text{Min} \{ \text{Min} \{ (\sum \text{sqr}t (q_i - \tilde{q}_i)^2) \} \} =$$

$$\text{Min} \left\{ \text{Min} \left\{ \left(\sum \text{sqr}t \left[\begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \end{bmatrix} \right]^2 \right) \right\} \right\} \quad (10)$$

3.1.2 基于特征骨骼点的匹配

图 3 是利用棋盘格图像的对应点逆向验证 Kinect 传感器坐标系与摄像头坐标系标定的初始 $\mathbf{R}, \bar{\mathbf{t}}$ 准确度, 结果证明了该算法的有效性。但棋盘格具有平面性, 难以说明 $\mathbf{R}, \bar{\mathbf{t}}$ 的精度对三维的坐标信息适用。对此本文做了相关的实验, 见图 4: Kinect 传感器采集的骨骼数据匹配到摄像头坐标系下的结果。此时的结果仍然存在较大的误差, Kinect 传感器坐标系下骨骼点的位置不能完全准确地匹配到摄像头

坐标系下, 数据点存在一定的漂移问题。为了达到具有参考意义的效果, 需要对坐标系间的匹配结果进行修正。



(a) Kinect 传感器采集

(b) 摄像头采集

图 3 Kinect 传感器坐标系与摄像头坐标系标定结果



(a) Kinect 传感器采集

(b) 摄像头采集

图 4 Kinect 传感器坐标系下骨骼数据匹配在摄像头坐标系

经过后期算法的改进, 将初始 $\mathbf{R}, \bar{\mathbf{t}}$ 矩阵进一步迭代, 修补局部最优的缺陷, 使得误差收敛到一个极小的数值, 此时得到的 $\mathbf{R}, \bar{\mathbf{t}}$ 矩阵是近似全局最优。修正算法后可得到较为精准的骨骼点匹配结果, 最终匹配的结果见图 5。此时匹配误差已经降低, 对于增强现实的注册来说, 将人的位置作为虚拟物体叠加的基准位置, 精度是满足的。



(a) Kinect 传感器采集

(b) 摄像头采集

图 5 Kinect 传感器坐标系下骨骼数据匹配在摄像头坐标系下

3.2 基于深度信息识别的远距离跟踪注册

根据上述修正的 $\mathbf{R}, \bar{\mathbf{t}}$ 矩阵, 将 Kinect 传感器采集的骨骼点深度信息匹配到摄像头坐标系下, 将作为虚拟物体叠加到真实场景中的基准位置。根据人

体特征骨骼点位置的变动,可改变虚拟物体注册的基准位置,实现对虚拟物体的灵活注册.本文只选取四个骨骼点的深度信息作为关键的注册位置,其思想来源于 ARToolkit 的识别机制.至于选取哪四个骨骼点位置,根据方便选择.

本文设计了两种模式的注册方式:近景模式注册和远景模式注册.二者注册模式的不同主要在于跟踪的距离,以验证本文提出的基于深度信息识别的远距离跟踪注册技术的鲁棒性.

3.2.1 近景模式注册

在近景模式中可以实现虚拟物体叠加于手部位置,以手的骨骼点坐标位置作为注册中心,根据手部位置的变动,改变注册的位置,同时能控制虚拟物体的姿态.

3.2.2 远景模式注册

在远景模式中可以实现虚拟物体叠加于身体某块区域,以人的脊椎骨骼点坐标位置作为注册中心,根据脊椎位置的变动,改变注册的位置.同时,加入自定义的手势操作,控制虚拟物体的姿态.

4 实验结果分析

4.1 实验运行环境

本文的实验运行平台是 64 位 Windows8.1 操作系统,开发平台借助于 Visual Studio 2012 C++,实现过程使用了 OpenCV 2.4.8 版开源函数库、OpenGL 三维图形库以及微软提供的 Kinect 接口文件.选取的摄像头和 Kinect 传感器分别为 Logitech HD Webcam C920, Kinect Xbox 2.0.实验过程中采集的影像分辨率为 1 920×1 080,视频刷新率每秒 30 帧,保证了影像的清晰度和系统运行的实时性.

4.2 实验运行结果

通过不断改变人体与摄像头的距离,验证远距离增强现实的稳定性.实验结果证明:在 Kinect 传感器识别的范围内,均可实现虚拟物体的叠加.

4.2.1 近景模式

近景模式的实验运行结果如图 6 所示.人距离摄像头 1~2 m,四张图像中人距离普通摄像头的距离分别为 1.0 m,1.3 m,1.7 m,1.9 m.虚拟的雪人叠加在人的手掌上.图中可以看到虚拟雪人的大小也随着距离发生变化,保证了虚拟物体的深度信息.

手部位置的变动控制了虚拟雪人的左右旋转,融入自然的交互行为.

4.2.2 远景模式

远景模式的实验运行结果如图 7 所示,此时人距离摄像头 4 m 左右.图 7 还表明了手势变化,可以改变 OpenGL 绘制的虚拟茶壶的姿态,让虚拟茶壶实现上下翻转,融入自然的交互.若实现更远距离的跟踪注册,可以通过扩展 Kinect 传感器的个数.



图 6 近景模式增强现实跟踪注册结果



图 7 远景模式增强现实跟踪注册结果

5 结束语

本文在分析增强现实原理的基础上,以 Kinect 传感器和摄像头构成跟踪系统,不同于基于标记或基于自然特征的无标记增强现实系统单纯摄像头的跟踪系统.实现的远距离的跟踪注册技术,能够摒弃纸质标识物带来的不便,并且拓宽增强现实的应用场景,为大场景的增强现实应用提供了解决方案.同时,省去了基于视觉追踪的增强现实对于图像的分割、特征提取、匹配过程和跟踪的过程,能较容易地

实现增强现实的应用. 同时, 引入人的手势操作, 对虚拟物体进行自然控制, 提高了增强现实体验的沉浸感.

最终实验结果表明本文提出的跟踪注册方法具有很好实时性和稳定性. 本文所探讨的跟踪注册技术适应于大场景的增强现实系统, 未来的工作也将更加具体的研究大场景增强现实系统的设计与实现.

参考文献:

- [1] Graham M, Zook M, Boulton A. Augmented reality in urban places: contested content and the duplicity of code [J]. Transactions of the Institute of British Geographers, 2013, 38(3):464-479.
- [2] Sanchez Sanchez J, Cschi J, Horaudr. Action recognition robust to background clutter by using stereo vision [C]//European Conference on Computer Vision. Berlin Heidelberg: Springer, 2013:332-341.
- [3] Li Yibo, Kang Shaopeng, Qiao Zhihua, et al. Development actuality and application of registration technology in augmented reality[C]//Computational Intelligence and Design, International Symposium. IEEE, 2008.
- [4] Tamura H. Steady steps and giant leap toward practical mixed reality systems and applications[C]//VAR'02: Proc. Intl Status Conf. on Virtual and Augmented Reality. Leipzig, 2002:1-10.
- [5] 金剑华, 陈一民. 增强现实中基于视觉与磁力跟踪器的三维注册方法 [J]. 计算机应用, 2006, 26 (6): 1485-1489.
- [6] 李启民, 陈一民, 黄诗华, 等. 增强现实系统中磁力跟踪注册的精准性研究 [J]. 计算机工程与设计, 2009, 30 (24):5785-5789.
- [7] Sato Y, Nakamoto M, Tamaki Y, et al. Image guidance of breastcancer surgery using 3D ultrasound images and augmented reality visualization [J]. IEEE Trans. Med. Imag, 1998, 17 (5):681-693.
- [8] Kim S, Dey A K. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation[C]// CHI09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2009:133-142.
- [9] Prabha Shreeraj Nair, Vijayalakshmi S, Durgadevi P. Registration technologies and their classification in augmented reality the knowledge-based registration, computer vision-based registration and track-based registration technology [J]. International Journal of Computer Science, Engineering and Application, 2016, 6(1):23-33.
- [10] George Papagiannakis, Gurminder Singh, Nadia Magnenat-Thalmann. A survey of mobile and wireless technologies for augmented reality systems [J]. Computer Animation and Virtual Worlds, 2008, 19 (1): 3-22.
- [11] Newman J, Schall G, Barakonyi I, et al. Wide-area tracking tools for augmented reality [C]// Pervasive Computing 2006, Adjunct Proceedings of Pervasive 2006. Dublin, 2006.

作者简介:

刘书曼 女, (1991-), 硕士研究生. 研究方向为计算机虚拟视觉、增强现实技术. E-mail: shumanl@163.com.

葛水英 女, (1975-), 研究员. 研究方向为虚拟视觉、增强现实.

徐士彪 男, (1986-), 助理研究员. 研究方向为计算机视觉、机器学习及增强现实.