

# 三维监控系统中基于三维重构的交互式标定

邓 磊<sup>1,2</sup>, 陈宝华<sup>1,2</sup>, 赖伟良<sup>1,2</sup>, 陈志祥<sup>1,2</sup>, 周 杰<sup>1,2,3</sup>

(1. 清华大学自动化系, 北京 100084; 2. 清华信息科学与技术国家实验室, 北京 100084;  
3. 智能技术与系统国家重点实验室, 北京 100084)

**摘 要:** 标定多个监控摄像机的位姿是智能监控系统的基础. 传统的标定方法需人工逐一标定每个目标摄像机, 且难以处理非重叠监控视场以及摄像机运动和扰动. 对此本文提出一种基于三维重构的交互式标定框架, 通过引入场景三维特征点云作为中间层, 仅需一次性建立其与参考背景模型间的几何变换关系, 即可通过目标摄像机图像与三维点云的匹配实现自适应标定, 可显著降低工作量. 由于匹配是建立在监控图像与三维点云之间而非监控图像之间, 因此可以处理监控视场非重叠的情况. 对于摄像机运动和扰动, 提出了一种在线相对姿态传递方法, 能够克服摄像机扰动和运动带来的姿态变化问题. 实验证明了本文方法的有效性.

**关键词:** 三维监控系统; 摄像机标定; 三维重构; PTZ 相机

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)03-0527-07

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.03.003

## Interactive Calibrating 3D Surveillance System Based on 3D Reconstruction

DENG Lei<sup>1,2</sup>, CHEN Bao-hua<sup>1,2</sup>, LAI Wei-liang<sup>1,2</sup>, CHEN Zhi-xiang<sup>1,2</sup>, ZHOU Jie<sup>1,2,3</sup>

(1. Department of Automation, Tsinghua University, Beijing 100084, China;

2. Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China;

3. State Key Laboratory of Intelligent Technology and Systems, Beijing 100084, China)

**Abstract:** Calibrating camera is essential for intelligent surveillance systems. Conventional calibrating methods usually calibrate the target cameras one by one and can not handle non-overlapping cases or camera motion/disturbance. In this work, we present an interactive calibration framework based on 3D reconstruction. The reconstructed 3D feature point cloud is treated as the interface between the 3D background model and the target camera. Through 2D-3D matching, a target camera could be automatically calibrated against the 3D feature point cloud. Due to the matching is performed between the target image and the point cloud, non-overlapping cases can be well handled. Also, an online relative pose transfer scheme is proposed to deal with the problem of camera disturbance or motion efficiently. Experiments demonstrate the effectiveness of the proposed framework.

**Key words:** 3D surveillance system; camera calibration; 3D reconstruction; PTZ camera

## 1 引言

三维监控系统是智能监控系统中的一个前沿研究方向<sup>[1-4]</sup>. 可将监控设备的视频画面实时嵌入到统一的参考背景模型中, 整合所有的监控画面信息, 形成对监控场景的整体和自由视角观察, 在智能交通、智能安防、智能社区等领域中应用前景广阔.

在三维监控系统的建立过程中, 标定摄像机在三

维参考背景模型中的位姿是核心环节. 对于这一问题, 一些研究采用基于传感器的方法 (GPS, 惯性导航, 姿态传感器等)<sup>[5]</sup>, 这类方法依赖专门设备且精度不高. 基于计算机视觉的自动标定方法研究较多<sup>[6-9]</sup>, 但通常要求监控图象间有足够的视场重叠, 利用运动匹配或特征匹配来标定摄像机间的相对位姿. 若将这类方法直接用于匹配摄像机图像与参考背景模型图像, 常因为二者差异过大或缺乏对应目标运动信息而

收稿日期: 2015-09-25; 修回日期: 2015-11-26; 责任编辑: 覃怀银

基金项目: 国家自然科学基金 (No. 61225008, No. 61373074, No. 61373090); 国家重点基础研究发展计划 (973) (No. 2014CB349304); 教育部基金 (No. 20120002110033)

失败. 已有的三维监控系统多采用交互式标定<sup>[3,10,11]</sup>的方法, 逐一建立每个摄像机与参考背景模型的对应关系, 结合几何计算得到摄像机的位姿. 这种方法工作量大(正比于摄像机数量), 且仅适合静态摄像机, 无法处理主动相机(如 PTZ 相机)的运动和摄像机扰动带来的影响.

本文提出一种基于三维重构的交互式摄像机位姿标定方法, 可用于快速估计多个摄像机在参考背景模型中的位姿, 并且能够克服图像运动(主动相机或者扰动)带来的影响, 流程如图 1 所示. 不同于传统的人工逐一标定摄像机的交互方式, 本文引入了三维特征点云作为中间层, 仅需一次性建立三维点云与参考背景模型间的几何变换关系, 而后借助该三维点云就能自动标定每一个目标摄像机, 可显著降低工作量. 除静态摄像机外, 本文方法还可以处理摄像机运动的情况.

## 2 系统概述

本文方法的系统流程如图 1 所示, 包括离线三维重构及点云嵌入和在线自动标定两个部分.

离线部分中, 首先主动采集足够的场景图像并采用运动恢复结构的方法 SFM (Structure From Motion<sup>[12]</sup>) 进行三维重构, 得到场景的三维特征点云和主动拍摄图像位姿. 随后采用提出的点云嵌入方法将该特征点云嵌入至参考背景模型中. 在线自动部分中, 将目标摄像机视频的当前帧与已知姿态的参考帧做特征匹配并估计相对旋转. 若无参考帧或匹配失败, 则将目标图与离线构建的三维特征点云进行 2D-3D 匹配并估计位姿以及更新参考帧, 最终完成图像嵌入. 若匹配估计成功, 则根据相对旋转传递得到当前帧的姿态, 完成图像嵌入. 此外, 当旋转较大时需更新参考帧.

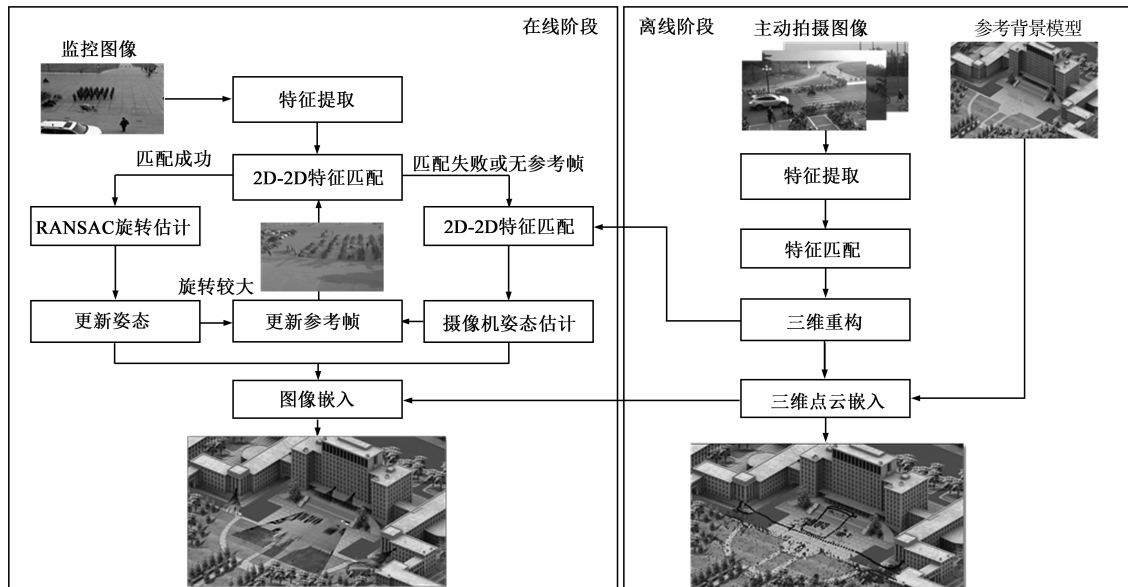


图1 系统总体框架图

RANSAC 是一种计算机视觉中常用的鲁棒估计框架, 它能够有效地克服样本中的外点和假点并正确地估计出最佳模型. 其基本思想是通过随机采样最小样本集并估计模型, 统计其余样本到该模型的距离并计算支持度. 由于当最小样本集均为内点时, 其对应的模型支持度最高. 所以多次采样后支持度最高的模型作为最终选取的最佳模型.

## 3 三维点云重构及嵌入

在离线三维点云重构及嵌入阶段中, 首先对采集的场景图像做三维重构得到三维特征点云. 其次将得到的三维特征点云嵌入到参考背景模型中, 包含两步: (1) 估计三维特征点云与参考背景模型之间的几何关系. 该几何关系主要表现为将参考背景模型某一视角

的视图视为在特征点云坐标系中的虚拟视角仿射摄像机并估计其姿态; (2) 根据手动输入的若干地平面上的 3D 点估计场景地平面方程并用于投影可视区域.

由于参考背景模型分为三维背景模型和二维背景视图, 因而估计虚拟视角摄像机姿态的方法也分为 3D-2D (三维特征点云-二维背景视图) 与 3D-3D (三维特征点云-三维背景模型) 两种.

### 3.1 3D-2D 下虚拟视角摄像机姿态估计

二维参考背景视图一般来自于卫星地图或者 3D 参考背景模型的某一视角. 输入的对应为三维特征点云到二维视图中的若干 3D-2D 对应点 (数量  $\geq 4$ ). 二维背景视图与三维点云之间的变换关系可以视为在某个角度观测三维点云. 对此提出一种求解该变换的方法,

描述如下。

二维背景视图一般为正交投影,通常将其定义为仿射摄像机或者无穷远摄像机<sup>[14]</sup>,其数学描述为

$$\begin{aligned} \mathbf{P}_A &= \mathbf{K}_A \begin{pmatrix} [\mathbf{R}_A]_{1,2} & [\mathbf{t}_A]_{1,2} \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} f & & & \\ & f & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} [\mathbf{R}_A]_{1,2} & -[\mathbf{R}_A \mathbf{C}_A]_{1,2} \\ 0 & 1 \end{pmatrix} \end{aligned} \quad (1)$$

其中  $\mathbf{R}_A, \mathbf{C}_A$  为该仿射摄像机在世界坐标系下的旋转和平移,  $[\cdot]_{1,2}$  为矩阵或向量的前两行,  $f$  为摄像机焦距。仿射摄像机投影矩阵  $\mathbf{P}_A$  中有 8 个参数, 因为存在两个内在约束  $\mathbf{R}_A(\mathbf{R}_A)^T = \mathbf{I}$ , 故  $\mathbf{P}_A$  只有 6 个自由度, 即 1 个尺度, 3 个旋转和 2 个平移。

输入若干三维特征点云与参考背景视图的 3D-2D 对应点  $(\mathbf{X}_i, \mathbf{x}_i)$ 。则其应满足投影关系

$$[\mathbf{P}_A]_{1,2} \tilde{\mathbf{X}}_i = \mathbf{x}_i \quad (2)$$

其中  $\tilde{\mathbf{X}}$  为齐次坐标。

为求解式(2), 根据 Kronecker 乘积的性质得到

$$(\tilde{\mathbf{X}}_i^T \otimes \mathbf{I}_{2 \times 2}) \text{vec}([\mathbf{P}_A]_{1,2}) = \mathbf{x}_i \quad (3)$$

由于每组对应点能提供两个约束, 对于的 6 个自由度, 最少需要 3 组输入对应点来求解。由于  $\mathbf{P}_A$  有 8 个参量, 将 3 组对应点输入后构造方程组  $\mathbf{A}\mathbf{Y} = \mathbf{b}$ , 其中  $\mathbf{A}$  为  $6 \times 8$  的系数矩阵。基于该欠定方程组得到  $\mathbf{P}_A$  的参数化的解为

$$\mathbf{P}_A = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} + \alpha \mathbf{P}_1 + \beta \mathbf{P}_2 = f(\alpha, \beta) \quad (4)$$

其中  $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$  为特解,  $\alpha, \beta$  为待求参数,  $\mathbf{P}_1, \mathbf{P}_2$  为该欠定线性方程组的基 (即张成数据矩阵  $\mathbf{A}$  的零空间的两个基)。将式(4)带入内在旋转约束  $\mathbf{R}_A(\mathbf{R}_A)^T = \mathbf{I}$  中得到关于  $\alpha, \beta$  的方程组并求解得到 4 个满足约束的解。为了确定 4 个解中的唯一真解, 需要用额外的对应点进一步验证。所以提出的方法至少需要 4 组 3D-2D 对应点 (三组求解 + 一组验证), 结合 RANSAC 鲁棒估计框架便得到最佳的摄像机姿态。

传统的人工标定方法是根据三维背景模型的若干 3D 点与目标摄像机的 2D 点的对应关系来估计摄像机姿态, 通常采用三点几何解法<sup>[15]</sup>, 即通过挖掘空间中 3 个 3D 点构成的三角形与图像平面 2D 点构成的三角形之间的几何关系来确定摄像机姿态。而本文研究的是根据三维点云中 3D 点与二维参考背景模型中的 2D 点对应关系来估计二维参考背景模型的位姿。由于二维参考背景模型为仿射摄像机模型而非传统的透视摄像机模型, 传统方法并不适用此问题。为此本文提出了仿射摄像机姿态估计的代数解法, 即挖掘仿射摄像机矩阵  $\mathbf{P}_A$  应满足的代数约束并建立等式求解。

### 3.2 3D-3D 下虚拟视角摄像机姿态估计

三维参考背景模型一般来自于 GIS 数据或者激光

扫描。输入的对对应关系为三维特征点云与三维背景模型间的若干 3D 对应点 (数量  $\geq 3$ )。通过最小化欧式距离误差<sup>[16]</sup> (式(5)) 可得到三维点云到三维背景模型的最佳相似变换  $\mathbf{T} = s\mathbf{R}[\mathbf{I} | -\mathbf{C}]$ 。

$$\min_{s, \mathbf{R}, \mathbf{t}} \sum_i \|\mathbf{X}_i^v - (s\mathbf{R}\mathbf{X}_i^c + \mathbf{t})\|^2 \quad (5)$$

其中  $\mathbf{t} = -s\mathbf{R}\mathbf{C}$ ,  $s$  是相似变换的尺度,  $\mathbf{R}$  是相似变换的旋转矩阵,  $\mathbf{C}$  是相似变换的平移,  $\mathbf{X}_i^v$  是三维背景模型中的第  $i$  个 3D 对应点,  $\mathbf{X}_i^c$  是三维点云中的第  $i$  个 3D 对应点。应用这一变换, 便可将点云嵌入到三维背景模型中并与参考背景模型统一起来。

如果背景参考模型为三维模型, 接下来可以使用渲染技术做三维自由视角观察。对于三维监控而言, 实时监控画面是以地面上动态纹理贴图嵌入至三维模型中。其本质是地平面方程引导的监控摄像机到与地面垂直的仿射摄像机间的单应变换。该仿射摄像机的参数决定了监控画面清晰度的大小, 通过放大比率 (焦距), 获取更高分辨率的纹理贴图, 使用中可以根据分辨率要求设计焦距值。

## 4 摄像机自动标定与图像嵌入

在线自动标定与图像嵌入阶段的任务是将目标摄像机监控视频嵌入参考背景模型, 方法是结合 2D-3D 图像定位方法与 2D-2D 位姿传递方法, 流程如图 1。包含了 2D-3D 图像定位 (第 4.1 节), 2D-2D 姿态传递 (第 4.2 节) 和图像嵌入 (第 4.3 节)。

### 4.1 2D-3D 图像定位

2D-3D 图像定位使用输入图像和三维特征点云的匹配计算该图像在三维特征点云中的位姿。首先提取输入图像特征 (SIFT), 并与三维特征点云进行 2D-3D 特征匹配。随后根据 2D-3D 匹配关系估计输入图像的摄像机姿态<sup>[15]</sup> (如图 2)。在 2D-3D 特征匹配阶段, 用输入的特征描述符在三维点云中通过 Kd-tree 寻找近邻和次近邻, 并计算其距离比率。若该比率小于一定阈值, 则认为该匹配可靠, 并反向进行 3D-2D 匹配再验证近邻与次近邻的距离比。若该比率也小于一定阈值则认为该匹配成功。得到匹配后估计目标摄像机位姿。若姿态估计成功, 则以此姿态创建参考帧。

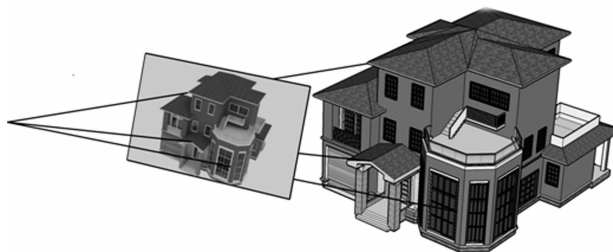
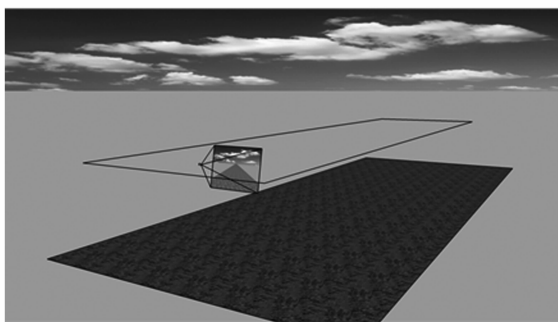


图2 根据2D-3D匹配估计摄像机位姿示意图

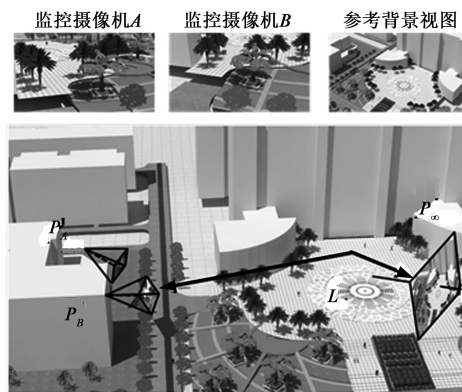
## 4.2 2D-2D 姿态传递

2D-2D 姿态传递方法通过匹配当前帧和带有位姿的参考帧来估计当前帧的位姿. 方法主要包括四步: (1) 对当前帧和参考帧进行 2D-2D 特征匹配; (2) 利用匹配关系使用 RANSAC 鲁棒估计框架估计当前帧与参考帧之间的相对运动; (3) 估计当前帧的姿态, 若当前帧相对于参考帧旋转角度过大, 则使用当前帧更新参考帧; (4) 根据当前帧姿态, 将当前帧图像嵌入参考背景模型.

本文考虑的摄像机运动主要有两种情况, 一是由于自然因素(风、碰撞等)造成的摄像机扰动. 在现实监控场景中通常监控摄像机位置被固定, 且距监控场景较远, 这种运动近似为纯旋转运动. 二是主动 PTZ 相机的运动情况, 由于其可以进行旋转和变焦, 该运动可视为带尺度的旋转. 纯旋转可以视为尺度为 1 的带尺度旋转. 求解如下, 当前帧与参考帧的第  $i$  对应点  $(\mathbf{x}_i^k, \mathbf{x}_i^r)$ , 应满足  $\mathbf{x}_i^k = \mathbf{R}_{kr} \mathbf{x}_i^r$ , 其中  $\mathbf{R}_{kr}$  二者的相对旋转. 通过最小化传递误差  $\min \sum_i \|\mathbf{x}_i^k - \mathbf{R}_{kr} \mathbf{x}_i^r\|^2$  得到最佳的旋转矩阵.



(a) 消隐线与图像内容投影关系



(b) 三维模型与不同摄像机的映射关系

图3 图像嵌入几何示意图

## 5 实验结果及分析

本文实验包含两部分, 静态监控摄像机嵌入和 PTZ 监控摄像机嵌入(摄像机扰动可视为运动较小的 PTZ 相机). 实验场景来自于某大学校区, 参考背景视图来自于与地平面夹角分别为 90 度的卫星地图和 45 度的三维地图, 如图 4(a), 分辨率为  $640 \times 480$ . 它们均可视为真实三维空间中特定视角的仿射摄像机. 用于三维重构的场景图像来自于 GOPRO 运动相机拍摄的视频, 分辨率为  $1280 \times 720$ . 图像数量为 360 张. 覆盖约  $200\text{m} \times 80\text{m}$  的范围. 4 个待嵌入的静态监控摄像机视频同步时长约为 5min, 帧率为 24frame/s. 待嵌入的主动 PTZ 摄像机视频为一段时长为 3min 的旋转缩放运动视频. 2D-3D 特征匹配相互验证的阈值为 0.6. 为加速特征提取, 本文使用 SiftGPU<sup>[17]</sup> 方法.

为了验证本文标定方法的标定工作量及精确度,

将得到的相对旋转  $\mathbf{R}_{kr}$  作用在参考帧的位姿  $\mathbf{P}_r = s_r \mathbf{R}_r [I | -\mathbf{C}_r]$  上得到当前帧的姿态  $\mathbf{P}_k = s_k s_r \mathbf{R}_{kr} \mathbf{R}_r [I | -\mathbf{C}_r]$ . 当相对运动足够大(尺度变化足够大  $s_k > \delta_s$  或  $\mathbf{R}_{kr}$  的欧拉角足够大  $\frac{1}{2}(\text{trace}(\mathbf{R}_{kr}) - 1) > \delta_\theta$ ) 时, 用当前帧更新参考帧  $(I_r \leftarrow I_k, P_r \leftarrow P_k)$ . 保证参考帧能够尽可能地匹配后续帧.

## 4.3 图像嵌入

得到当前帧在点云坐标系中的位姿  $\mathbf{P}_k$  后, 结合得到的虚拟视角摄像机位姿  $\mathbf{P}_A$  (第 3 节), 以及点云坐标系中的平面方程  $L$  便可以进行图像嵌入. 图像嵌入分为两步, 一是获取需要投影的图像区域, 即图像平面消隐线以上的图像区域(如图 3(a)). 二是计算其到参考背景模型的单应变换(三维背景模型或二维视图), 该单应变换是由地平面引导出的, 由当前帧摄像机与虚拟视角摄像机所参与的, 如图 3(b). 得到该单应变换后, 结合投影的图像区域, 便可将监控画面投影嵌入至参考背景模型中.

本文与经典的基于特征点估计单应变换的方法进行对比. 这类方法<sup>[3]</sup>首先通过人工交互获取待嵌入的静态图像与参考背景中地平面上的若干对应点, 一般不少于 4 对. 其次根据这些对应点估计单应变换. 最后根据该变换对待嵌入图像作透视变换并叠加到参考背景模型中. 具体实验中, 本文方法与对比方法中人工交互的结果均为非本文作者的第三方提供.

### 5.1 静态摄像机嵌入

离线部分中由主动拍摄的图像(图 4(c)) 重构的三维点云(图 4(d)) 包含了 360 张主动拍摄图像的位姿和用于定位的 14238 个 3D 点. 重构结果比较完整地表达了场景结构. 点云嵌入结果如图 4(d)、4(e) 所示, 白色点为点云与参考背景视图间的对应点, 由人工标定, 黑色点为重构得到的摄像机位置. 实验结果表明, 通过较少的人工交互(4 组对应点), 重构的三维模型(点云及主动拍摄的摄像机)便可以较好地嵌入至参考

背景视图中。

在线部分中,使用本文方法定位出的静态摄像机及嵌入的视频画面(图 5(a))所示。多个监控摄像机的位姿及其监控的区域以黑框标注。整体上监控画面与参考背景视图较好地对应起来。传统的人工标定方法<sup>[3]</sup>结果见图 5(b),从图上观察,二者嵌入效果接近。但传统人工标定的方法需要的标定点(白色点)与摄像机数量  $n$  成正比,至少需要  $4n$  个。当需要标定的摄像机数量较大时,本文方法优势明显。另一方面,传统通过人工标定对应点只能得到摄像机图像与背景视图之间

的单应变换,无法获取摄像机位姿。而本文方法不仅可以得到该变换,还可以得到摄像机在背景视图中的位姿。标定一个摄像机的时间小于 2s。

本文采用的定量评价指标为最少标定点数和标定误差平均值。标定误差的计算过程如下,用角点检测器提取参考背景视图和原始摄像机图像的 Harris 角点,并建立对应关系。嵌入后,原始图像上的角点坐标变换为参考背景视图坐标系下的点,该点与参考背景模型中对应角点的像素距离为标定误差,其平均值作为评价指标。实验结果如表 1 所示。

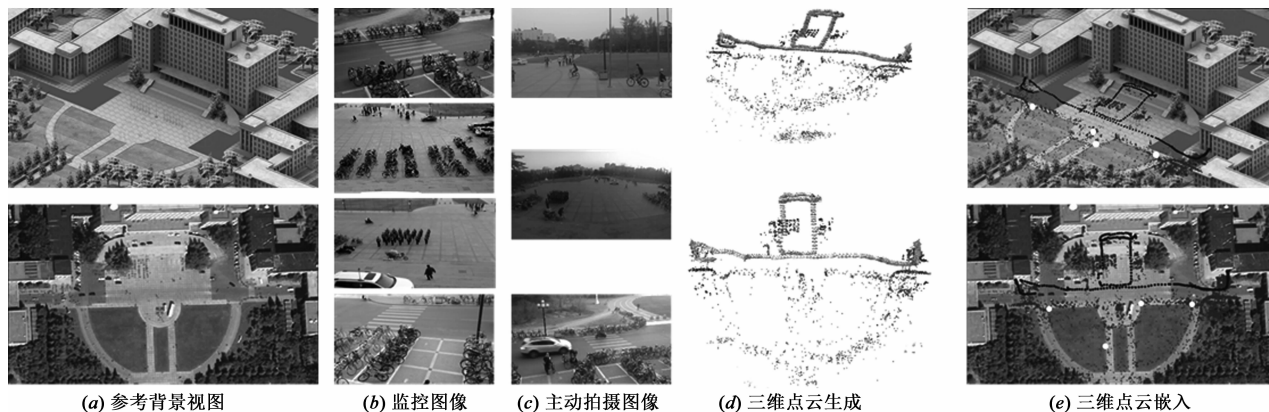
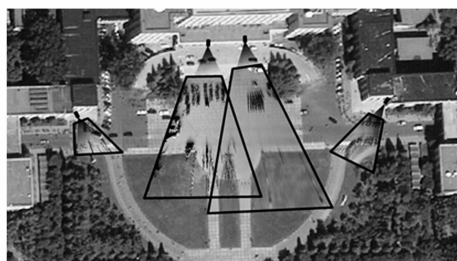


图4 静态摄像机实验中三维点云嵌入结果

表 1 静态摄像机标定结果

场景	摄像机数量	方法	最少标定点数	平均误差(像素)
场景 A (45 度视图)	4	人工方法 <sup>[3]</sup>	16	5.8
		本文方法	4	4.0
场景 B (卫星图)	4	人工方法 <sup>[3]</sup>	16	4.5
		本文方法	4	3.8



(a) 本文方法嵌入



(b) 传统人工嵌入<sup>[3]</sup>

图5 静态监控摄像机的标定嵌入结果

从实验结果可以看出,一方面本文方法所需最小标定点数明显少于人工标定所需的点数,而且待标定摄像机数量越多,优势越明显。另一方面,本文方法的标定精度优于人工方法,其主要原因是由于三维点云内涵了多张主动相机拍摄的图像信息,而这些内涵的信息是通过自动算法精确重建整合而成。与此同时相比直接标定单张图像,点云分布更广阔,可选的标定点覆盖范围更大,所以在减少人工标定的工作量的同时还能够提高最终的标定精度。

## 5.2 PTZ 摄像机嵌入

对于主动 PTZ 摄像机的标定问题,由于其不断运动,传统的人工标定方法无法处理。本文提出的自动标定框架可以较好地解决。离线点云重构部分与静态摄像机实验相同,实验结果如图 6 所示,从实验结果可以看出 PTZ 视频帧之间,以及视频帧与参考背景视图之间有较好的对应关系。基于姿态传递标定一个摄像机的时间在 50ms 以内,可做到实时嵌入。

由于传统人工标定方法针对静态摄像机设计,当 PTZ 摄像机参数变化发生运动时,需要重新标定。为定量评价本文方法与传统方法<sup>[3]</sup>,本文针对两个应用场景分别选用 18 和 23 个不同参数的 PTZ 采集的图像作为待标定样本,逐一采用人工标定。样本图像的 pan 参数覆盖了水平 70 度,tilt 参数覆盖了俯仰 40 度,zoom 参

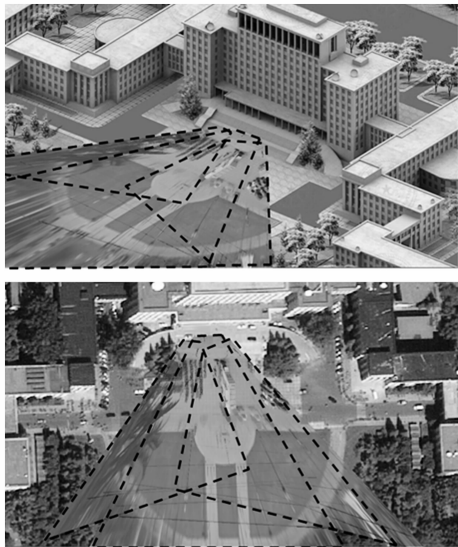


图6 PTZ相机监控画面叠加结果. 不同颜色的框为不同的PTZ参数对应的视场范围

数覆盖 1 ~ 3 倍变焦. 所有样本图像的总视场覆盖了较大范围的场景, 如图 6. 对于这些样本图像, 采用与标定静态摄像机相同的实验方法和指标进行评测, 结果如表 2. 可以看出当需要标定的样本增多, 人工标定代价急剧上升. 并且由于人工标定并未考虑样本之间的几何约束关系, 导致标定结果一致性降低, 误差变大. 本文提出的方法由于 3D 点云内涵了更多的主动拍摄的信息并可以自动完成对目标摄像机的标定, 且相比于视场受限的单张图像的标定方法, 标定点可以更分散, 使得最终标定结果精度更高.

表 2 PTZ 摄像机标定结果

场景	样本数量	方法	最少标定点数	平均误差 (像素)
场景 A (45 度视图)	18	人工方法 <sup>[3]</sup>	72	5.2
		本文方法	4	3.9
场景 B (卫星图)	23	人工方法 <sup>[3]</sup>	92	4.7
		本文方法	4	3.1

## 6 总结与展望

三维监控系统是智能监控系统中的重要研究内容. 现有的三维监控系统中摄像机标定主要采用人工交互方法, 工作繁杂, 而且无法处理摄像机运动的情况 (主动 PTZ 相机和扰动). 对此, 本文提出了一种基于三维重构的交互式标定框架, 仅需进行一次离线采集场景图像并三维重构, 以及简单的人工标注对应点 ( $\geq 4$  个点), 便可以在线自动标定多个目标摄像机并将其监控画面嵌入至参考背景模型中. 提出的在线姿态传递方法能够较好地解决主动 PTZ 相机运动和摄像机扰

动下的标定问题. 实验证明了提出方法的有效性. 未来我们将综合利用多个监控摄像机视频的嵌入结果, 研究高层的协同智能分析.

### 参考文献

- [1] 周忠, 周颐, 肖江剑. 虚拟现实增强技术综述[J]. 中国科学: 信息科学, 2015, 45(2): 157 - 180.
- [2] 李波, 姚春莲, 李炜, 郑锦. 利用相邻帧和背景信息的运动对象检测[J]. 电子学报, 2008, 11(11): 2154 - 2159.  
Li Bo, Yao Chun-lian, Li Wei, Zheng Jin. A novel motion object detection algorithm using adjacent frame and background frame[J]. Acta Electronica Sinica, 2008, 11(11): 2154 - 2159. (in Chinese)
- [3] Ripolles O, Simó J E, Benet G, et al. Smart video sensors for 3D scene reconstruction of large infrastructures[J]. Multimedia Tools & Applications, 2014, 73(2): 977 - 993.
- [4] Wang X. Intelligent multi-camera video surveillance: A review[J]. Pattern Recognition Letters, 2013, 34(1): 3 - 19.
- [5] Sebe I O, Hu J, You S, et al. 3D video surveillance with augmented virtual environments[A]. Proceedings of the ACM SIGMM 2003 Workshop on Video Surveillance[C]. New York, USA: ACM, 2003. 107 - 112.
- [6] Lv F, Zhao T, Nevatia R. Camera calibration from video of a walking human[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(9): 1513 - 1518.
- [7] Zhang Z, Li M, Huang K, et al. Practical camera auto-calibration based on object appearance and motion for traffic scene visual surveillance[A]. Proceedings of IEEE Computer Vision and Pattern Recognition[C]. Anchorage, USA: IEEE, 2008. 1 - 8.
- [8] Stauffer C, Tieu K. Automated multi-camera planar tracking correspondence modeling[A]. Proceedings of IEEE Computer Vision and Pattern Recognition[C]. Madison, USA: IEEE, 2003. 259 - 266.
- [9] 全红艳, 张田文. 一种新的利用模板进行摄像机自标定的方法[J]. 电子学报, 2005, 11(11): 1941 - 1945.  
Quan Hong-yan, Zhang Tian-wen. A new camera self-calibration method using template[J]. Acta Electronica Sinica, 2005, 11(11): 1941 - 1945. (in Chinese)
- [10] Harpreet S, Aydin A, Rakesh K, et al. Video flashlights: real time rendering of multiple videos for immersive model visualization[A]. Proceedings of the 13th Eurographics Workshop on Rendering Techniques[C]. Pisa, Italy: IEEE, 2002. 157 - 168.
- [11] Peter M, Volker S, Peter W, et al. Next-generation 3D visualization for visual surveillance[A]. Proceedings of the 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance[C]. Klagenfurt, Austria: IEEE, 2011. 343 - 348.

- [12] Wu C. Towards linear-time incremental structure from motion[A]. Proceedings of the 2013 International Conference on 3D Vision[C]. Seattle, USA; IEEE, 2013. 127 – 134.
- [13] David L. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
- [14] Hartley A, Zisserman A. Multiple View Geometry in Computer Vision[M]. Cambridge: Cambridge University Press, 2006. 28 – 30.
- [15] Martin A, Robert B. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381 – 395.
- [16] Umeyama S. Least-squares estimation of transformation parameters between two-point patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1991, 13(4): 376 – 380.
- [17] Sinha S N, Frahm J M, Pollefeys M, et al. GPU-based video feature tracking and matching[J]. Workshop on Edge Computing Using New Commodity Architectures, 2006, 2000(Oct 21): 189 – 196.

### 作者简介



邓 磊 男, 1988 年生于山西侯马. 现为清华大学自动化系博士研究生. 主要研究方向为计算机视觉及模式识别.  
E-mail: dengl09@mails. tsinghua. edu. cn



陈宝华 男, 1978 年出生于内蒙古兴安盟. 现为清华大学自动化系博士研究生. 主要研究方向为计算机视觉及模式识别.  
E-mail: cbh10@mails. tsinghua. edu. cn