

# Real-Time Planar Tracking and Jitter Reduction in Electricity-Related Public Safety Environments

Zili Yang, Wu Chen, Jiashan Zhu, Yinting Li

Qujing Power Supply Bureau, Yunnan Power Supply Co., Ltd., Qujing Yunnan  
Email: 124434320@qq.com

Received: Jul. 10<sup>th</sup>, 2020; accepted: Jul. 23<sup>rd</sup>, 2020; published: Jul. 30<sup>th</sup>, 2020

---

## Abstract

This paper combines the specific needs of fast plane recognition and poses estimation in augmented reality applications, to achieve a fast and efficient three-dimensional planar object identification and tracking algorithm. In order to solve the jitter problem during tracking, this paper also implements a post-processing module for position and attitude data smoothing, based on the Butterworth filter, which handles data at the expense of very low system latency, to satisfy the needs of practical uses in virtual power station training scenes to achieve a combination of virtual and reality display.

## Keywords

Augmented Reality, Planar Tracking, Jitter Reduction, Filtering

---

# 涉电公共安全场景实时跟踪和防抖算法研究

杨子力, 陈武, 朱家山, 李胤廷

云南电网有限责任公司曲靖供电局, 云南 曲靖  
Email: 124434320@qq.com

收稿日期: 2020年7月10日; 录用日期: 2020年7月23日; 发布日期: 2020年7月30日

---

## 摘要

针对涉电安全工作与治理工作的增强现实展示应用, 需要在移动平台上实现快速平面识别和位置姿态估算, 并且叠加虚拟场景与实际平面联动的需求, 提出一种快速而高效的三维平面物体识别和跟踪算法。

同时,为了解决物体识别和跟踪过程中的数据抖动问题,采用了Butterworth滤波器构建了位置和姿态数据的后处理平滑功能模块,以极低的系统延迟为代价,将输出的结果数据进行了平滑和防抖动处理。实验表明,该方法可以快速、低延迟、无抖动地将虚拟电站培训场景与摄像头呈现的现实画面叠加渲染,实现虚实结合的展示效果,从而满足了实际使用中的需要。

## 关键词

增强现实,平面跟踪,防抖动,滤波

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

增强现实(Augmented Reality),简称AR,已经成为各行各业必不可少的一种展示技术手段。基于AR技术的展示工程,通常需要首先确定一个位置标记(Marker),例如,通过摄像头画面识别和跟踪一张二维码或者图片,计算得到它在空间中的位置和姿态信息,然后,基于这个位置姿态信息去叠加显示虚拟场景。识别和叠加显示的过程一般都是在同一台智能手机上完成的,当用户操作手机通过不同角度来观察标记物体时,虚拟场景也会随之发生位置和角度的变化,并始终保持正确的叠加状态,形成虚实结合的展示效果。

随着AR技术的不断发展,近年来已经逐渐涌现出各种无标记物(Markerless)的跟踪算法,例如,通过摄像头画面的移动来确定目标的空间和几何关系,然后计算得到摄像头的实时位置姿态信息,用于后续的虚实结合应用。但是,与基于标记物的方法相比,无标记物的方案很难保证始终稳定的位置姿态输出,而是存在着数据在一定范围内跳动的问题。在展示效果上会反映为虚拟画面的抖动,这对于用户来说无疑是一种非常不舒服的视觉体验。

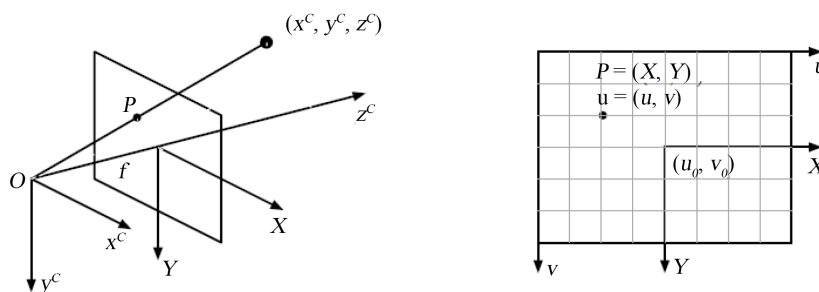
传统的信号过滤和防抖算法是基于Kalman滤波方法实现的,这种方法可以使输出结果的曲线更平滑,但是需要提前记录多帧数据并结合在一起运算出新的结果,因而造成系统的画面延迟和帧率降低的问题。另一种可行的防抖方法是构建一系列预置的关键数据帧,然后,将实时输出的位置姿态与数据帧中的数据进行匹配,在不降低帧速率的前提下得到一个稳定的输出结果。但是,这种方法可能会造成数据的不连续问题,在用户移动摄像头并从其它角度观察虚实叠加的效果时,会造成“跳帧”的问题。

我们研究了一种快速的无标记点三维平面识别算法,这样可以迅速帮助我们找到虚拟场景叠加的位置和角度,并实现AR显示的需求。同时,基于该算法实现基于Butterworth滤波器的位置姿态数据防抖算法。与Kalman滤波器兼具跟踪的效果不同,Butterworth滤波器专注于噪声数据的去除,因此,防抖计算的效果更好,系统延迟几乎可以忽略不计。

## 2. 无标记点的三维平面追踪算法

### 2.1. 图像特征的实时提取方法

通过一台普通手机的摄像头画面来完成全部的实时AR跟踪算法功能,它可以被理解为一个针孔相机的模型,如图1所示:



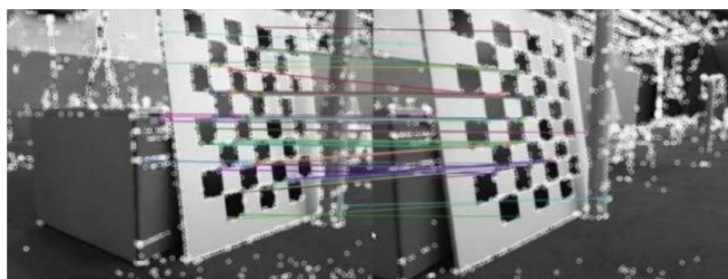
**Figure 1.** Pinhole camera model  
**图 1.** 针孔相机模型

其中  $O$  为摄像机原点,  $P(x^c, y^c, z^c)$  为相机空间坐标系中的任意一点,  $(u, v)$  为该点在图像空间的对应像素点。

首先, 从图像中提取必要的特征信息, 这一过程中需要从图像中提取出一些自身特点突出, 不会因光照、仿射变换和噪音等因素而变化的点, 如角点、边缘点、暗区的亮点及亮区的暗点等。即使是看起来没有什么突出颜色信息的平面表面(例如桌面, 地面等), 也存在着大量符合上述要求的特征点, 即图像空间的多组  $(u, v)$  坐标, 如果可以将它们转换为三维空间的坐标, 就可以从中拟合出空间平面的信息, 进而实现平面。

采用 ORB 算法来实现实时高速的图像特征提取, 它具有较稳定的旋转不变性, 但是对尺度变化的感应能力较弱[1]。该算法最大的优势在于效率很高, 约为传统的 SIFT 提取算法的 100 倍左右, 非常适用于移动平台端的运算和 AR 跟踪算法使用。

通过 ORB 对任意连续的两帧摄像头画面进行图像特征提取和匹配的结果如图 2 所示。图中白色的点即所有被识别出来的特征点, 而两幅图之间的连线表示它两端的两个点在三维空间中可以被认为是一个点。因为, 这两个点在各自的图像空间中有不同的坐标值  $(u^0, v^0)$  和  $(u^1, v^1)$ , 所以我们可以借助图像匹配的结果推算得到两幅图像对应的相机之间的相对运动关系, 即从左图变换到右图过程中的相对位置姿态矩阵[2]。



**Figure 2.** Image feature extraction and matching  
**图 2.** 图像特征提取和匹配

以上步骤可以简化为下面的数学过程:

- 1) 对于摄像头拍摄的任意两个连续帧  $F1$  和  $F2$ , 求解得到两组互相匹配的特征点  $P$  和  $Q$ , 满足:

$$P = \{p_1, p_2, \dots, p_n\} \in F1$$

$$Q = \{q_1, q_2, \dots, q_n\} \in F2$$

- 2) 求取位置矩阵  $t$  和旋转矩阵  $R$ , 使得:

$$\forall x, q_x = R \cdot p_x + t$$

如果匹配的特征点足够多,且误匹配的点数较少,那么上述过程可以顺利得到从  $F1$  变换到  $F2$  的位置矩阵  $t$  和旋转矩阵  $R$ ,即,得到相机在空间运动的相对位置姿态关系,进而实现虚拟场景与实际画面同步运动时的跟踪矩阵计算。

下一步,需要识别并拟合得到一个空间中的平面对象(例如桌面,地面等),从而确定虚拟场景放置的具体位置。

## 2.2. 基于 RANSAC 的平面拟合方法

如果能够正确运算得到两帧图像之间的位置姿态矩阵结果,那么,对于每一个匹配过的特征点,都可以基于三角化的方法求取它在相机空间的  $Z$  坐标。具体方法是,在不同的位置观测同一个三维点  $P(x^c, y^c, z^c)$ ,已知在两帧中各自观察到的三维点的图像坐标为  $(u^0, v^0)$  和  $(u^1, v^1)$ ,利用三角关系,可以恢复得到该三维点的深度信息。

因此,对于每帧图像,都可以得到一个稀疏三维点云的列表  $L$ 。然后,通过 RANSAC 算法来分析点云数据并拟合得到一个空间平面。该算法可以从一组含有“外部点”(outliers)的数据中正确估计数学模型参数,但是,带有一定的随机性,需要多次迭代以提升算法的稳定性和正确性。这里的“外部点”通常指的是数据中的噪声,对于本文的需求而言就是落在平面之外的匹配特征点。

在使用 RANSAC 进行平面拟合的过程中,首先从稀疏点云列表  $L$  选取随机的三个点来构成一个平面,然后,将所有其它的点带入到这个平面的数学模型中进行计算,判断它们是否为“内部点”(inliers)。最后,比较当前模型和之前推出的最佳模型的“内部点”的数量,更新最佳模型的参数信息,直到迭代结束为止[3]。

第一次进行平面匹配时,因为无法判断“内部点”和“外部点”的比例关系,难以推算出最合理的迭代值,因此,可能会消耗较长的时间进行计算。但是,当得到被跟踪的平面信息之后,下一次迭代时可以直接将上一帧的已知平面对应的点作为 RANSAC 的初始参数,从而加速计算和迭代完成。基于 RANSAC 进行拟合的示意结果如图 3 所示:

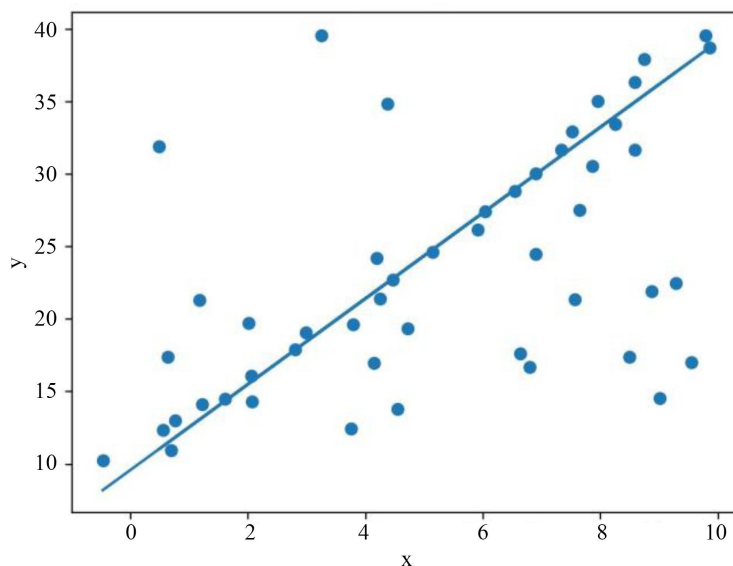


Figure 3. RANSAC-based data fitting results

图 3. 基于 RANSAC 的数据拟合结果

正确计算得到空间平面的信息之后, 可以将虚拟场景对应放置在这个平面上, 即: 确定虚拟场景在相机中的初始位置, 并且, 随着摄像头的运动, 实时计算相机的位置姿态变换矩阵并赋予虚拟场景, 从而完成无标记点辅助的三维平面识别和跟踪的全部过程。

### 3. AR 场景跟踪过程中的实时防抖算法

在识别和跟踪三维平面, 以及计算摄像头运动的相对位置姿态矩阵的过程中, 往往会存在着数据不稳定的问题, 因而造成虚拟场景模型在被叠加的实际场景上反复小幅度跳动, 产生一种“画面正在抖动”的观感效果。

这一问题的主要成因来自于两个方面: 一是因为 ORB 特征提取算法自身对尺度信息的匹配能力不强, 误匹配率比较高, 进而造成相机位置姿态计算和三维点云计算的源数据质量较低; 二是因为移动设备摄像头自身的画面质量, 分辨率, 以及运动过程中的模糊带来的数据精度损失, 进而影响了相邻帧进行图像匹配运算的结果[4]。

在无法短时间内改善 ORB 算法的先天性不足, 以及无法直接提升硬件设备的质量的前提下, 本文提出了一种新的基于 Butterworth 滤波器的数据后处理算法, 将画面的抖动感最大限度地降低到用户可以接受的程度, 同时不会因为数据的平滑处理带来过于明显的延迟问题。

对信号进行分析处理时, 会考虑将信号分为低频和高频两个部分。低频成分对应数据变化相对缓慢的部分, 例如一幅图像内容的大致轮廓和颜色, 或者空间坐标变化的整体趋势; 而高频成分则对应着信号变化剧烈的部分, 例如图像的局部细节或者噪声, 或者空间坐标产生剧烈振荡和无法预测的运动的部分[5]。

Butterworth 滤波器是一种低频滤波器, 即, 过滤或者大幅度衰减输入信号的高频成分, 同时让它的低频成分通过[6]。如图 4 所示, 它在通频带内的频率响应曲线非常平坦, 没有起伏, 而在阻频带则逐渐下降为零[7]。这一特性使得它对于 AR 相机运动跟踪矩阵的数据平滑和噪声过滤有很好的效果。

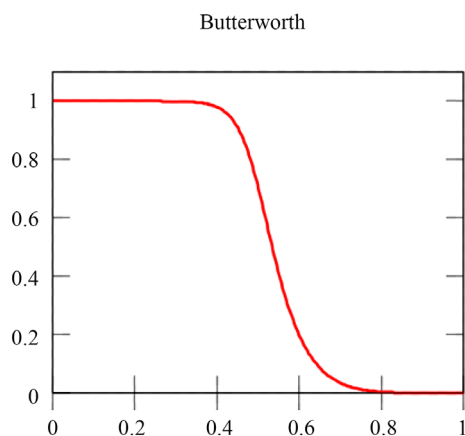


Figure 4. Frequency response characteristics of Butterworth Filters

图 4. Butterworth 滤波器的频率响应特性

$N$  阶 Butterworth 滤波器的高频信号衰减速度是  $20 N$  dB/dec (分贝每 10 倍频), 但是随着滤波器的阶数越高, 系统的响应越慢, 系统延迟越明显。如果数据采样间隔是  $t$ , 那么通过一个  $N$  阶 Butterworth 滤波器对输入信号进行处理之后, 结果和原信号相比会有  $N \cdot t$  的延迟[8]。在测试过程中发现, 使用二阶 Butterworth 滤波器即可基本满足画面去抖动的需求, 而手机图像的发送和数据解析间隔为 30~50 ms 之间,

此时整个后处理系统的信号延迟约为 60~100 ms，对于用户体验的影响很小，可以忽略不计。

#### 4. 实验测试与结果分析

在涉电安全工作与治理工作当中，通过 AR 技术，将虚拟的电力设备、设施模型与实际场景相融合，通过移动平台实现了涉电违章建筑、违章施工、违章种植、外力破坏、社会人员触电风险知识等突出涉电公共安全隐患的虚实结合的可视化效果。这一系统的开发环境主要为 Android 平台，底层算法使用 C++ 和 OpenCV 实现，将摄像头画面解析为三维平面识别的空间坐标和相机位置姿态矩阵之后，传递给 Unity 前端进行虚拟画面的正确对位和叠加渲染，同时通过 Butterworth 滤波器实现了跟踪结果的防抖动处理，并且满足实时系统对于运算速率的需求。

在系统启动过程中，首先尝试在空间中查找一个静态的平面物体，例如图 5 所示的办公电脑桌面。系统找到平面物体后，会提示用户是否将虚拟场景“放置”到当前平面上。如果用户选择放置，则叠加渲染虚拟场景，构成 AR 虚实结合的效果。

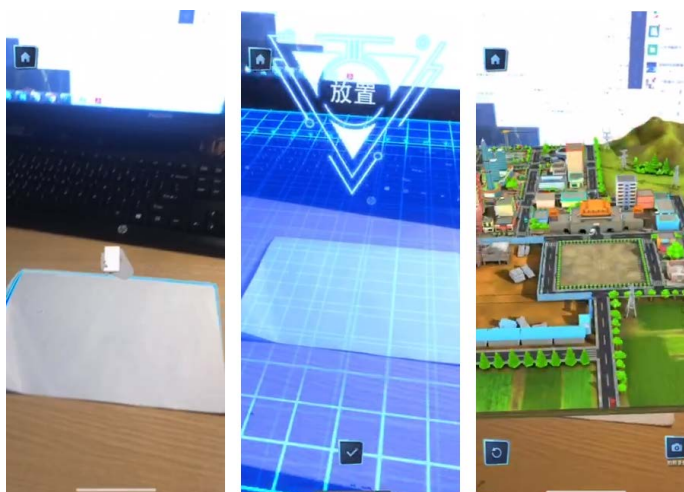


Figure 5. The flow of system startup: find the 3D plane (left), determine current plane (middle), and overlay the 3D scene (right)

图 5. 系统启动的整体流程：查找三维平面(左图)，确定当前平面(中图)，叠加显示三维场景(右图)

系统运行一段时间后的连续多帧显示画面如图 6 所示：

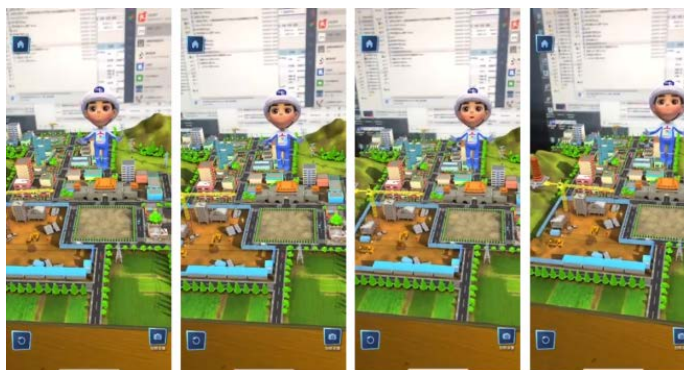
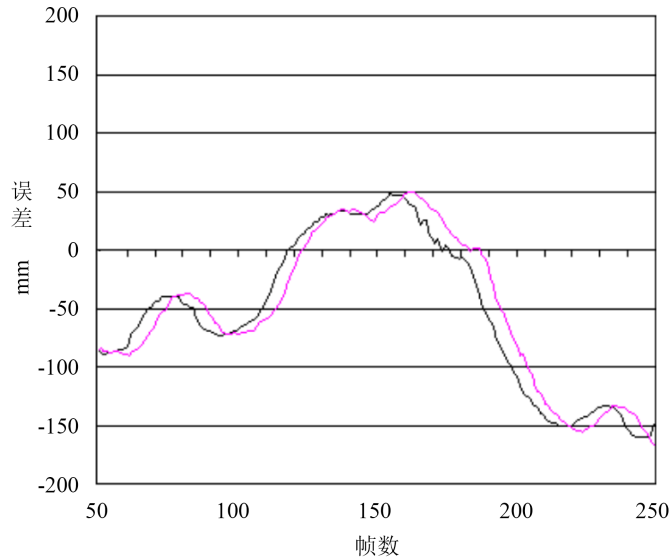


Figure 6. The actual tracking and running effects of system

图 6. AR 系统实际跟踪和运行的效果

系统实时运行输出的相机位置姿态信息带有一定的振荡，造成叠加显示的画面产生抖动的观感。通过添加 Butterworth 滤波器，可以很好地对输出数据进行平滑处理，消减这种高频的抖动，同时产生的延迟不影响实际体验，如图 7 所示：



**Figure 7.** The smoothing effect and latency profile of the anti-jitter algorithm. The black curve is the X position of the camera, and the magenta curve is the X position of the camera as filtered by Butterworth

**图 7.** 防抖动算法的平滑效果和延迟情况。图中黑色曲线为原始计算的相机 X 位置坐标，洋红色曲线为经过 Butterworth 滤波的相机 X 位置坐标

图中所示的滤波后的结果曲线比原始曲线滞后约 10 帧左右，而手机端运行 AR 程序的渲染帧速率固定为 60 Hz，即每秒钟 60 帧，因此可以判断出系统延迟大约为 160 ms 左右，对于用户体验来说，定位响应延迟的影响基本可以忽略不计，同时有效地改善了画面抖动不稳定的问题，确保 AR 叠加的效果流畅可用。

## 5. 结论

本文研究了一种快速的实时三维平面物体识别和跟踪算法。它基于手机摄像头设备输入的图像信息，通过 ORB 特征提取算法高效地提取和匹配相邻两帧图像的特征数据，进而计算得到两帧图像之间的相对位置和姿态变换关系。通过 RANSAC 方法，将图像的特征点变换到三维空间，并且拟合得到平面物体的位置。将虚拟模型叠加在这个平面物体上，同时基于视频图像连续变换关系矩阵计算相机的位置和姿态变化，即可完成 AR 系统对场景的识别和跟踪过程。

由于硬件设备本身的局限性，以及特征提取算法自身的尺度误差和误匹配问题，会造成结果变换矩阵的微小的计算误差，进而导致画面产生抖动。采用了 Butterworth 滤波器来过滤高频的输入信号，去除数据噪声，并对输出结果曲线进行平滑。这一方法相比传统的 Kalman 滤波方法，具有更好的实时响应能力，对用户体验不会产生明显的影响。

另外，本文研究的方法还有需要进一步改进之处：如果因为摄像头运动过快，或者场景特征信息太少等问题，导致 ORB 特征提取方法失效，那么跟踪算法也会随之失效，并且无法恢复到之前定位的位置。

因此, 研究引入回环检测方法, 并结合移动端的传感器(例如 IMU 等), 可以进一步提高特征匹配和场景跟踪的稳定性, 并且在实际应用中检验其运行效果。

## 参考文献

- [1] 杨炳坤, 程树英, 郑茜颖. 改进的 ORB 特征匹配算法[J]. 传感器与微系统, 2020, 39(2): 136-139.
- [2] 曹恒. 基于单目视觉的 SLAM 算法研究[D]: [硕士学位论文]. 武汉: 华中科技大学, 2016: 1-62.
- [3] 苏云龙, 平雪良, 李楠. 基于 RANSAC 三维点云的平面提取算法[J]. 激光与红外, 2019, 49(6): 780-784
- [4] 杨泽刚, 阮军, 张珊珊. 红外运动目标图像的识别与跟踪技术研究[J]. 武汉理工大学学报, 2011, 35(6): 1293-1296.
- [5] Lin, C. (1999) An Adaptive Estimator for Registration in Augmented Reality. *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, San Francisco, 20-21 October 1999.
- [6] Lu, C., Hager, G. and Mjolsness, E. (2000) Fast and Globally Convergent Pose Estimation from Video Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 610-622. <https://doi.org/10.1109/34.862199>
- [7] Cornelis, K., Pollefeys, M. and Van Gool, L. (2001) Tracking Based Structure and Motion Recovery for Augmented Video Productions. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 17-24. <https://doi.org/10.1145/505008.505012>
- [8] Chia, K.W., Cheok, A.D. and Prince, S.J.D. (2002) Online 6 Dof Augmented Reality Registration from Natural Features. *International Symposium on Mixed and Augmented Reality*, Darmstadt, 1 October 2002, 305. <https://doi.org/10.1109/ISMAR.2002.1115123>