

# 基于传感器与高效二阶最小化结合的追踪方法

王佳顺,管有庆

(南京邮电大学 物联网学院,江苏 南京 210003)

**摘要:**在室外环境中,由于追踪目标的运动状态可能会随时发生改变,而且手持设备的硬件性能与台式机相比存在局限性,所以可以运用在移动设备上的高效实时追踪技术一直是研究的重点和难点。在移动增强现实(augmented reality, AR)系统中,提出了一种基于手持移动设备的无标记物体追踪方法。该方法能很好地适应目标物体的旋转和运动模糊的情况。首先在特征点匹配阶段,将传感器的旋转参数引入鲁棒独立的基本二进制特征描述子(binary robust independent elementary features, BRIEF)中,使特征描述子具有旋转不变性。在特征匹配成功后采用改进的高效二阶最小化追踪算法(efficient second-order minimization-small grid-motion blur, ESM-SG-MB)来对目标物体进行追踪。该算法引入了摄像头的快门时间,并将模板图像进行网格划分,从中选择高梯度的子网格进行追踪。实验结果表明,与现有的移动增强现实系统中运用的追踪方法相比,该方法具有更好的鲁棒性和实时性。

**关键词:**移动增强现实;特征描述子;目标追踪;高效二阶最小化;传感器

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2018)07-0173-05

doi:10.3969/j.issn.1673-629X.2018.07.037

## Object Tracking Solution Based on Sensors and Efficient Second-order Minimization

WANG Jia-shun, GUAN You-qing

(School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** The movement status of tracking targets may change unpredictably in outdoor environment, and the mobile device has poorer performance than stand-alone device does. Therefore, efficient real-time tracking technology that can be applied to mobile devices has always been the focus and difficulty of research. For this, we propose a mark-less tracking solution based on handheld devices in outdoor augmented reality (AR), which can be well adapted to the target's rotating and motion blur. At the stage of feature matching, the rotation parameters of the sensors are introduced to BRIEF (binary robust independent elementary features) descriptor to achieve a rotational-invariant performance. After a successful match is observed, the ESM-SG-MB (efficient second-order minimization-small grid-motion blur) tracking method starts to serve its role to continue tracking. ESM-SG-MB tracking algorithm introduces camera shutter time and divides the template into sub-grids. The average image gradient in each sub-grid is computed, and only high gradient grids are used in tracking. Experiment shows that the proposed tracking solution has a better robustness and real-time ability than several state-of-the-art tracking method in mobile augmented reality.

**Key words:** mobile augmented reality; feature descriptor; object tracking; efficient second-order minimization; sensor

## 0 引言

增强现实(augmented reality, AR)是目前计算机应用领域的一大研究热点,其本质是在现实场景中添加虚拟的影像,使得现实场景与虚拟物体相结合,以此来增强人们对现实世界的体验<sup>[1]</sup>。增强现实技术融合了数字图像处理<sup>[2]</sup>、计算机视觉<sup>[3]</sup>、智能模式<sup>[4]</sup>识别等

多种技术,其中图像识别与追踪技术是比较重要的技术。

近年来在室外增强现实系统中,通常采用GPS获取相机的位置并通过手持设备的内置传感器来获取设备的方向,以此实现追踪定位,但是该方法有它的局限性<sup>[5]</sup>。首先,GPS的数据更新频率以及准确度并不能

收稿日期:2017-09-15

修回日期:2018-01-25

网络出版时间:2018-03-08

基金项目:江苏省高校自然科学基金项目(05KJD520146)

作者简介:王佳顺(1993-),男,硕士研究生,研究方向为软件技术在通信网络中的应用;管有庆,副研究员,硕导,研究方向为数据库、通信软件和下一代网络等。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180307.1053.004.html>

满足实时追踪定位的需求,其次传感器取得的方向、旋转角度等数据总是存在一定的误差,且这种误差会随着时间而不断累积,对追踪定位的精度产生一定的影响<sup>[6-7]</sup>。另外传感器会受环境中可能存在的电磁场的影响,使得检测到的数据出现误差。

当前手机已成为移动 AR 系统的主流运行平台。手机由于内置有传感器、照相机、GPS 等系统组成的机电系统,使得手机成为室外 AR 系统的理想平台。但是移动设备的计算能力和速度都比不上传统的电脑,使得传统的复杂追踪算法不能很好地适用于移动设备,所以对适用于移动设备的追踪算法的研究一直是学者关注的重点。

目前,AR 系统中的识别追踪算法已经取得了许多成果。Bleser 和 Stricker<sup>[8]</sup>研究了多种基于视觉惯性的融合模型并提出了基于多种传感器结合的 3D 追踪模型。但是该模型的生成和维持需要耗费大量的时间,而且该模型是基于静态环境,但是实际上环境可能不断发生变化。Hwangbo 等将惯性传感器与 Kanade-Lucas-Tomasi (KLT) 特征追踪算法相结合<sup>[9]</sup>。但是基于 KLT 的追踪算法覆盖范围比较小,当相机移动范围太大时,追踪算法容易失效。Calonder<sup>[10]</sup>等提出了一个鲁棒独立的基本二进制特征描述子(binary robust independent elementary features, BRIEF),该描述子可以在设备上实时计算,但是 BRIEF 没有旋转不变性。

文中在 BRIEF 二进制描述子的基础上,引入了惯性传感器的旋转参数,使 BRIEF 具有旋转不变性,当追踪目标旋转时能更好地进行识别匹配。系统在识别目标物体后,采用改进的 ESM 算法(efficient second-order minimization-small grid-motion blur, ESM-SG-MB)对运动物体进行追踪。该算法将模板图像划分成若干子网格,剔除了低梯度的子网格,增加了追踪效率,并且在 ESM 算法的基础上引入了摄像头快门时间,以适应运动模糊的情况。

## 1 基于传感器的特征匹配

在基于视觉的追踪系统中,局部特征的检测和匹配是不可缺少的一步。BRIEF 是利用局部图像领域内随机点对的灰度大小关系来建立局部图像特征描述子,得到的二值特征描述子不仅匹配速度快,而且对存储要求低,所以在手机应用中具有良好的发展前景。随机点对的分布可以是任意的也可以是按照一定分布排列的。在 Calonder 等的实验中表明,在特征点中心周围的采样准则服从各向同性的同一高斯分布时,会有较好的识别结果。

以特征点为中心,取  $S \times S$  ( $S$  可变)的邻域大窗口,在大窗口中随机选取  $n$  对点,对这  $n$  对点分别做高

斯平滑,比较  $n$  对像素点的灰度值大小。式 1 定义了一个二进制码,其中  $p(x)$ 、 $p(y)$  分别表示在图像块  $p$  上,  $x$  和  $y$  处像素的灰度值。

$$\tau(p; x, y) = \begin{cases} 1, & \text{if } p(x) < p(y) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

最后把通过式 1 得到的  $n$  个二进制码组成一个  $n$  维向量,如下所示:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i, y_i) \quad (2)$$

由式 2 可知,一个关键点的二进制特征描述子反映了  $n$  对像素点的灰度值,这  $n$  对像素点的取样采用各向同性的同一高斯分布,可以用矩阵  $A$  表示这  $n$  对像素点:

$$A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix} \quad (3)$$

其中,  $x_i$  和  $y_i$  表示图像平面坐标系中每对像素点的位置。

为了使 BRIEF 描述子具有旋转不变性,需要在矩阵  $A$  上加入一个旋转变量  $R_\theta$ ,  $R_\theta$  由移动设备内置的传感器获得。由于惯性传感器会受到噪声影响,所以用卡尔曼滤波器<sup>[11]</sup>对传感器获取的信号进行滤波,得到  $f(R_\theta, K_g)$ ,使得  $A_\theta = f(R_\theta, K_g)A$ 。其中  $K_g$  表示卡尔曼增益,  $f$  表示信号经过卡尔曼滤波器时的变换函数。加入旋转的 BRIEF 二进制描述子的定义如下:

$$f_{n,\theta}(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; [A_\theta]_{x_i}, [A_\theta]_{y_i}) \quad (4)$$

其中,  $p$  表示一个图像块的特征描述子;  $\theta$  表示惯性传感器采集的一个特定的旋转角;  $n$  表示二进制描述子的维数,通常为 128, 256 或者 512。

对于图像特征点的提取,可以使用中心环绕极值特征(center surround extremas, CenSurE)<sup>[12]</sup>的方法, CenSurE 特征具有特征点坐标精确、计算效率高等特点。然后使用旋转 BRIEF 二进制描述子来描述所提取出的特征点并计算两个二进制描述子之间的汉明距离来进行特征匹配。

## 2 基于传感器与高效二阶最小化结合的追踪方法

### 2.1 ESM-SG-MB 算法

ESM 是 Benhimane 和 Malis 于 2004 年在 IROS 上提出的算法<sup>[13-14]</sup>。该算法采用重构误差平方作为衡量模板图像  $R$  与相机取得的帧图像  $T$  相似性的指标,然后对于姿态空间进行了在离群上的重新构建,使得搜索的步长更为合理,在寻优上使用的是二阶近似的快速算法。ESM 算法最大的优点是不用计算 Hessian 矩阵,并且具有较快的收敛速度。

定义  $R(x)$  为模板图像像素点的灰度,  $T(x)$  为目标图像像素点的灰度,  $x$  表示像素坐标。模板图像  $R$  上的点可以经过某种变换,与目标图像  $T$  进行比较。假设  $w(x;e)$  为在  $x$  处的像素经过变换后得到的新的位置,其中  $e$  为变换参数。因此通过式 5 对变换参数  $e$  进行估计:

$$\hat{e} = \operatorname{argmin}_{x \in R} \sum [R(W(x;e)) - T(x)]^2 \quad (5)$$

然而传统的 ESM 算法模型并没有考虑到运动模糊的情况,借鉴 Tai Y W 等<sup>[15]</sup>提出的适用于非均匀模糊的模型,这里定义物体运动模糊的模型为:

$$T(e, t_0) = \frac{1}{1 - t_0} \int_{t_0}^1 R(W(x;e)) dt \quad (6)$$

其中,  $e$  表示模板图像的变换参数;  $t_0$  为相机快门打开时间;  $x$  表示模糊的像素的坐标。

这里将  $R(w(x;e))$  在  $e = 0$  处用二阶泰勒展开式展开,得到:

$$R(W(x;e)) \approx R(x) + J(0)e + \frac{1}{2}e^T H_R e \quad (7)$$

其中,  $J(0)$  是  $R(w(x;e))$  在  $e = 0$  处的雅可比矩阵;  $H_R$  是它的 Hessian 矩阵。

将式 7 代入式 6,可以得到图像构造模型的近似值:

$$\begin{aligned} T(e, t_0) &\approx \frac{1}{1 - t_0} [R(x) + J(0)e + \frac{1}{2}e^T H_R e] dt = \\ &R(x) + \frac{1}{1 - t_0} \left( \int_{t_0}^1 e dt \right) J(0) + \\ &\frac{1}{2(1 - t_0)} \left( \int_{t_0}^1 e^2 dt \right) H_R \end{aligned} \quad (8)$$

这里为了计算简单,假设模板图像做简单的线性运动,即:

$$e = t\beta \quad (9)$$

将式 9 代入式 8 可得:

$$\begin{aligned} T(\beta, t_0) &\approx R(x) + \frac{1}{1 - t_0} \left( \int_{t_0}^1 t\beta dt \right) J(0) + \\ &\frac{1}{2(1 - t_0)} \left( \int_{t_0}^1 (t\beta)^2 dt \right) H_R = R(x) + \\ &a(t_0)\beta J(0) + b(t_0)\beta^2 H_R \end{aligned} \quad (10)$$

为了避免计算 Hessian 矩阵  $H_R$ ,对  $T(\beta, t_0)$  求一阶导数,得:

$$J_T(\beta) = \frac{\partial T(\beta, t_0)}{\partial \beta} = a(t_0)J(0) + 2b(t_0)\beta H_R \quad (11)$$

将式 11 得到的  $H_R$  代入式 10,图像构造模型为:

$$\begin{aligned} T(\beta, t_0) &\approx R(x) + a(t_0)\beta J(0) + \frac{1}{2}[J(\beta) - \\ &\text{万方数据}(t_0)J(0)]\beta = R(x) + \frac{1}{2}[J(\beta) + \end{aligned}$$

$$a(t_0)J(0)]\beta \quad (12)$$

假设摄像头获取的图像为  $T$ ,可以计算出运动状态  $\beta$  的估计为:

$$\hat{\beta} = 2 \frac{T - R(x)}{J_T(\beta) + a(t_0)J(0)} \quad (13)$$

计算出的运动姿态  $\beta$  被加入当前姿态,该过程被反复迭代,直到收敛。

ESM 算法进行模板匹配追踪时,采用的是模板图像的所有像素点。与 ESM 算法不同,ESM-SG-MB 算法将模板图像分成许多个子网格,从中筛选出符合条件的子网格进行追踪。追踪的目标图像是根据模板图像中心点附近的  $120 \times 120$  个像素区域。将该块区域分为  $5 \times 5$  个小的子网格,每个网格有  $24 \times 24$  个像素点。计算出每个子网格中的平均图像梯度,只有当该子网格的每个像素点的灰度级大于 10 时,该子网格才会被用于追踪。因为实验表明,梯度低的图像区域对 ESM 的收敛性没有影响。

将这种改进的 ESM 算法称为 ESM-SG-MB 算法,可以通过修改快门时间,也就是  $t_0$  的大小,来修改现在实现的 ESM 原型,使其对运动模糊有更好的适应性。ESM-SG-MB 算法的详细过程见算法 1。

算法 1: ESM-SG-MB 算法。

输入:视频帧 frame;模板图像  $R$ ;

输出:更新的姿态估计  $e$ 。

(1) 将模板图像分割成  $5 \times 5$  个小的子网格;

(2) 剔除子网格每个像素灰度级小于 10 的网格;

(3) 选择合适的快门打开时间  $t_0$  以及初始姿态估计  $e$ ;

(4) 已知初始变换为  $e$ ,计算雅可比矩阵  $J(0)$  和  $J_T(e)$ ;

(5) 根据式 13 更新  $e$ ,判断是否收敛,如果不收敛则返回步骤 4,否则结束。

## 2.2 旋转 BRIEF 与 ESM-SG-MB 结合的追踪算法

图 1 阐述了提出的追踪算法流程。当追踪开始时,系统先从摄像头读取视频帧,如果是第一帧的话,系统就用 CenSurE 方法提取图像的特征点并计算出各个特征点的旋转 BRIEF 二进制描述子的值。然后将模板图像的二进制描述子与目标图像相匹配,匹配成功后,进入 ESM-SG-MB 追踪算法继续追踪目标图像。在追踪过程中,如果检测到追踪物体丢失,则重新进行特征点的检测与匹配。最后,当系统检测到当前帧是视频的最后帧时,追踪结束。

## 3 模拟实验结果与分析

### 3.1 模拟实验环境

用一台智能手机(OPPO A57)来试验提出的追踪

方法。手机的 CPU 主频为 1.4 GHz,核心数为八核,操作系统为 Android 6.0.1。用于追踪的后置摄像头

像素为 1 600 万像素,分辨率为 1 280 \* 20P/30 Hz 且曝光时间可以调整。

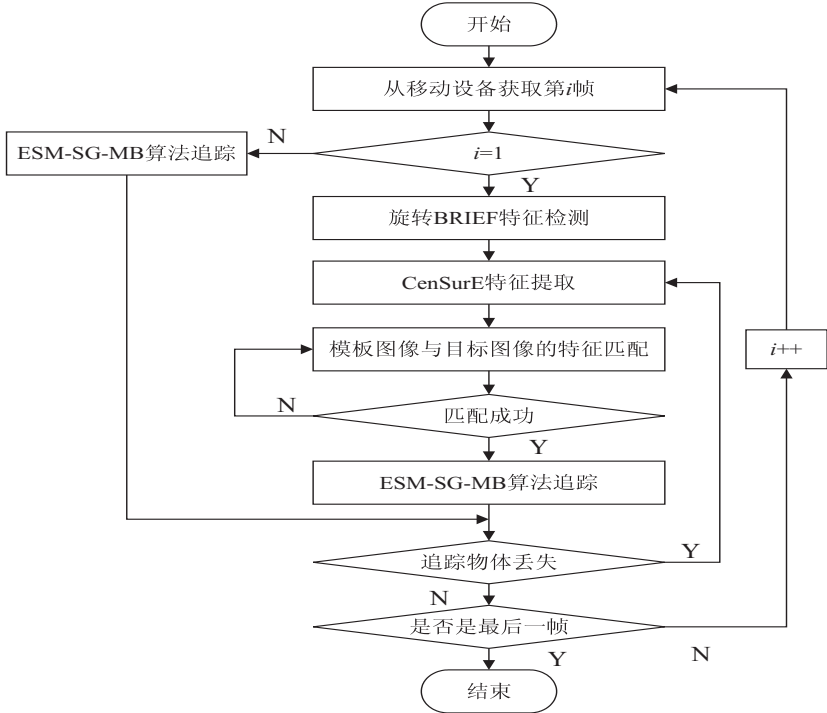


图 1 追踪算法流程

3.2 模拟实验分析

首先取一张经典的 Lena 图片作为目标图像,将图片中点的 120×120 像素大小的图片作为模板图像。对 Lena 图像进行拍摄时,对智能手机进行旋转操作,在随机曝光时间下生成 800 张图像,对于每一张图像,分别用 ESM 追踪算法、Inertial KLT 追踪算法和提出的追踪方法进行追踪,结果如图 2 所示。

图 2 显示出 ESM 算法的旋转不变性较差,与 Inertial KLT 相比,提出的追踪算法的旋转不变性更好。

接下来,采用室外环境下的运动模糊物体追踪的视频流进行实验,如图 3 奔跑中的人的视频序列所示,视频中的人物从清晰到模糊。将模板图像置为 120×120 像素,快门保持打开状态。然后分别使用 ESM 追

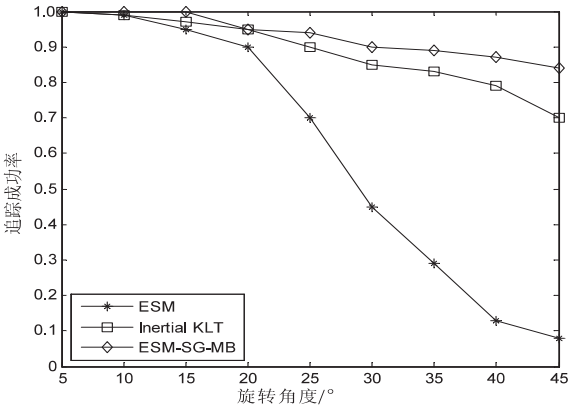


图 2 不同旋转角度下不同算法之间的对比追踪算法、Inertial KLT 算法和提出的追踪方法,跟踪情况如图 3 所示。

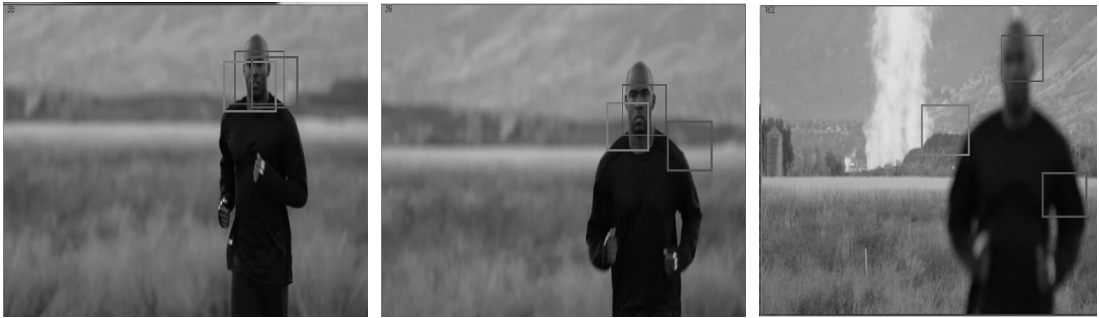


图 3 奔跑中的人

前期视频流中的追踪对象比较清晰,所以三种算法的追踪效果都比较好。但随着目标物体的运动模糊越来越明显,ESM 和 Inertial KLT 追踪算法的追踪效

果越来越差,而提出的追踪算法对运动模糊有较强的抗干扰能力。图 4 显示了 ESM 算法、Inertial KLT 算法、ESM-SG-MB 算法在上述视频中的追踪结果对比



图,追踪成功率定义为成功追踪的帧数占当前总帧数的比例。可以明显看出,文中提出的算法在运动模糊的情况下,追踪效果更好。

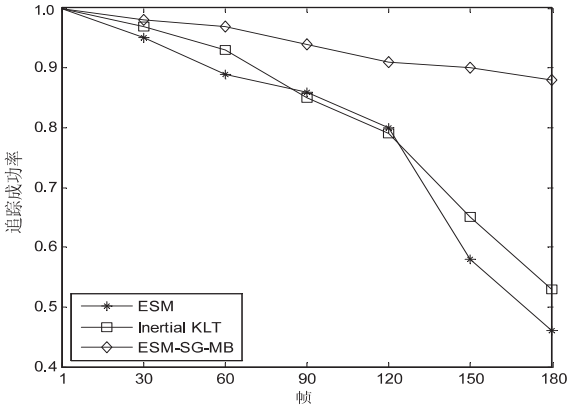


图 4 算法在视频流中的对比

4 结束语

提出了一种运用在移动设备的室外 AR 追踪方法。在 BRIEF 描述子的基础上,加入了旋转变量,使 BRIEF 二进制描述子具有旋转不变性。又改进了 ESM 算法,使得系统在追踪过程中能很好地适应运动模糊。实验结果表明,该追踪方法对目标物体旋转和运动模糊的情况有较强的抗干扰能力。在未来的工作中,可以考虑从降低移动设备在追踪时的计算量、提高实时性的方面进行研究。

参考文献:

[1] 赵 越. 实时增强现实中运动目标及场景的跟踪注册关键问题研究[D]. 沈阳:东北大学,2015.

[2] 桂振文. 面向移动增强现实的场景识别与跟踪注册技术研究[D]. 北京:北京理工大学,2014.

[3] 周国众. 移动增强现实关键技术及应用[J]. 测绘与空间地理信息,2012,35(9):140-144.

[4] 孙 源,陈 靖. 智能手机的移动增强现实技术研究[J]. 计算机科学,2012,39(6A):493-498.

[5] YU L,ONG S K,NEE A Y C. A tracking solution for mobile augmented reality based on sensor-aided marker-less

tracking and panoramic mapping[J]. Multimedia Tools and Applications,2015,75(6):3199-3220.

[6] LI Zhiyong, GAO Song, NAI Ke. Robust object tracking based on adaptive templates matching via the fusion of multiple features[J]. Journal of Visual Communication & Image Representation,2017,44:1-20.

[7] WALIA G S,KAPOOR R. Recent advances on multicue object tracking: a survey[J]. Artificial Intelligence Review, 2016,46(1):1-39.

[8] BLESER G,STRICKER D. Advanced tracking through efficient image processing and visual-inertial sensor fusion [C]//IEEE virtual reality conference. Reno, NE, USA: IEEE,2008:137-144.

[9] HWANGBO M, KIM J, KANADE T. Inertial-aided KLT feature tracking for a movingcamera[C]//IEEE/RSJ international conference on intelligent robots and systems. St. Louis,MO,USA:IEEE,2009:1909-1916.

[10] CALONDER M V,LEPETIT V,STRECHA C,et al. BRIEF: binary robust independent elementary features [C]//11th European conference on computer vision. Heraklion:[s. n.],2010:778-792.

[11] 张 彤,孙玉国. 卡尔曼滤波在 MEMS 惯性姿态测量中的应用[J]. 光学仪器,2015,37(1):28-30.

[12] AGRAWAL M,KONOLIGE K,BLAS M R. CenSurE:center surround extremas for realtime feature detection and matching [C]//European conference on computer vision. Marseille:[s. n.],2008:102-115.

[13] CHEN Lin,ZHOU Fan,SHEN Yu,et al. Illumination insensitive efficient second-order minimization for planar object tracking[C]//2017 IEEE international conference on robotics and automation. Singapore:IEEE,2017:4429-4436.

[14] TAHRI O,MEZOUAR Y. On the efficient second order minimization and image-based visual servoing[C]//IEEE international conference on robotics & automation. Pasadena: IEEE,2008:3213-3218.

[15] TAI Y W,TAN Ping,BROWN M S. Richardson-Lucy deblurring for scenes under a projective motion path[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011,33(8):1603-1618.