

基于改进卡尔曼滤波的增强现实跟踪注册算法^①

刘佳^② 闫冬 郭斌

(南京信息工程大学自动化学院, CICAET, 南京 210044)

摘 要 针对传统应用在特征点匹配技术方向的增强现实系统部分存在鲁棒性不足和实时性较差问题,提出一种新的区域跟踪匹配算法。首先利用改进的 FAST 算法在不同尺度上提取像素信息建立低维模板进行快速定位。其次在跟踪阶段使用加速鲁棒特征(SURF)检测关键点在线实时匹配的同时,借助改进的非线性化的卡尔曼滤波算法预测运动轨迹,缩减匹配区域。接着对使用上述的算法截取部分图像帧,利用上一帧的观测值对下一帧进行预测并估计相机姿态,同时使用随机抽样一致(RANSAC)算法剔除离群值。最后通过得到的相机位姿矩阵进行虚拟物体的注册叠加。实验结果表明,本文提出算法在综合评价中占据优势,在光照变换、旋转变换、尺度变换方面都具有较好的鲁棒性,并具有优越的实时性能。

关键词 卡尔曼滤波; FAST 算法; 加速鲁棒特征(SURF)算法; 增强现实(AR); 姿态估计

0 引言

增强现实(augmented reality, AR)是在用户观察到的真实环境中添加计算机生成的虚拟物体或文字等信息的技术^[1]。建立虚拟环境和现实物理世界的映射关系,智能实时地将相关信息融合呈现在位于现实环境的用户前面,达到增强用户临场感知能力的目的^[2]。该技术目前在手术医疗、军事训练、工业维修、教育、交通导航上有着广泛的应用。增强现实的关键技术有跟踪注册、虚实融合和人机交互,这其中的核心是跟踪注册技术。目前基于视觉增强现实跟踪注册技术大致分为三大类,一类是基于四方黑白标识的跟踪注册,一类是针对自然特征的匹配注册,还有一类是通过建立局部地图进行虚拟注册。其中第1种方法由于临场环境的局限性限制了发展空间;第2种由计算机性能和算法决定系统运行速度,不能保证系统实时处于稳定状态;而第3种方法的代表则是 SLAM/PTAM(simultaneous

localization and mapping/parallel tracking and mapping),通过建立虚拟地图获得实时点的位姿关系进行注册,系统本身复杂度高,并且为了保证系统的稳定性而牺牲了不小的运行速度。

传统基于自然特征点匹配的增强现实系统的实现,是借助检测算法检测预存的目标图像和相机捕获图像之间的点对应位置,计算之间的单应性关系,从而实现虚实融合注册。传统的人工特征如尺度不变特征变换(scale invariant feature transform, SIFT)、加速鲁棒特征^[3](speeded-up robust features, SURF),两者在移动端都不能实时稳定运行。为了能在移动端使用此类特征,Ethan 等人^[4]在 2011 年提出快速特征点提取和描述算法(oriented fast and brief, ORB)。随着神经网络的发展,部分研究者通过训练深度卷积神经网络以估计平面单应性,这种算法在很大程度上节省了检测和匹配所耗费的时间,提高了系统的实时性。例如 Nguyen 等人^[5]提出的无监督深度单应性估计模型。除了这种端对端的训练方式,也有很多人开始使用神经网络结合传统的人工

① 国家自然科学基金(61773219)和江苏省产业前瞻与关键核心技术(BE2020006-3)资助项目。

② 女,1981年生,博士,教授;研究方向:计算机视觉,增强现实,机器人技术;E-mail: liujia@nuist.edu.cn
(收稿日期:2020-01-03)

特征的深度学习算法进行训练,如 Verdie 等人^[6]提出的时间不变检测器(temporally invariant learned detector, TILDE)算法,该特征对变化巨大的场景具备很好的鲁棒性。尽管如此,TILDE 算法却缺少了特征描述符,需要借助传统的方法进行匹配。而 Yi 等人^[7]提出的基于学习的不变特征变换(learned invariant feature transform, LIFT)算法通过 soft-argmax 函数和空间变化网络实现从检测到匹配的完整步骤,因此稳定效果很好。从实时性考虑,如果上述基于深度学习的特征不借助 GPU 的并行运算能力,很难达到实时运行的目的。因而考虑现实条件,传统的特征匹配方法仍然具备很大的优势。

本文基于自然特征识别注册方法考虑,通过改进 FAST(features from accelerated segment test)检测算法建立低维图像空间进行直方图匹配,使其能快速定位,进而利用 SURF 算法较好的鲁棒性结合改进后的卡尔曼滤波^[8]算法进行跟踪注册,实现匹配和跟踪的结合。

1 特征提取与匹配技术

1.1 结合改进 FAST 算法作图像快速粗略定位

本文提出了一种通过改进 FAST 检测作图像粗略定位的算法。FAST 算法本身不具备尺度不变性,为了能够减少尺度变化对检测效果的影响,建立图像金字塔进行降采样(见图 1),使其具备尺度不变性。本文旨在提高检测效率以保证在跟踪目标丢失时能快速定位目标所在区域,在建立图像金字塔时以 1.2 倍缩放采样,采样 4 次,当检测到关键点后保存关键点所在位置的 RGB 像素信息,将采集的像素信息重新分配到新建的 3 通道一维矩阵降低图像维度。将新建的矩阵空间转换成 HSV 图像空间矩阵,然后建立 HSV 空间的直方图。除检测目标图

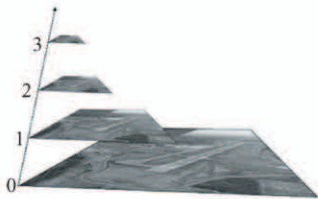


图 1 采样金字塔

像外,还需将相机捕获的图像分割成 5 等份,分别建立颜色直方图,以确定最大相似度^[9](如图 2 所示)。

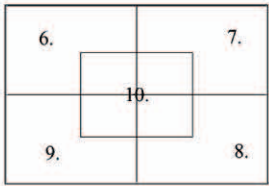


图 2 5 等分切割

本文使用的定位算法主要由以下 3 个要素构成。

(1) 建立尺度空间

- ① 对第 l 层图像进行高斯卷积运算。
- ② 将偶数行和列删除:

$$g_l(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) g_{l-1}(2i + m, 2j + n) \tag{1}$$

(2) 提取特征点

参考 ORB 算法的提取步骤,在粗提取时采用 FAST 特征检测算法提取候选特征点,以提取的候选特征点 p 为圆心,画一个以 3 像素大小为半径的圆,在圆周上的 16 个点内,检测 1、5、9、13 位置上的点的灰度值是否有 3 个或以上都大于或小于 p 点的灰度值,如果不满足则排除这个候选点,如果满足,检测是否存在连续的 12 个点都大于或者小于 p 点的灰度值,如果满足则此候选点为需要的特征点。使用非极大值抑制剔除密集点,以保证特征点的分散性。

(3) 区域相似度检测

将检测到的特征点所在位置的 3 通道像素信息重新写入到 3 通道 n 行 1 列的矩阵中,为了方便比较,将 RGB 空间转换成 HSV 空间,使用如下相关性比较的方法。

$$d(H_1, H_2) = \frac{\sum_i (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_i (H_1(I) - \bar{H}_1)^2 \sum_i (H_2(I) - \bar{H}_2)^2}} \tag{2}$$

当 d 越接近 1,则表明图像越相似;当 $d < 0.5$ 时表示图像不相关(本文将 d 设置为 0.6)。在本

文算法中,建立直方图进行查找相关图像是为了快速定位目标图像的大体区域,检索的图像尺度仅为 320×240 像素大小,为了提高待检测图像的准确率,使用 FAST 算法在多尺度空间检索特征点。将特征点的像素信息写入新的一维图像以加快检测的速度,当图像跟踪失败时,可以更加快速稳定地衔接视频的连续性。

1.2 SURF 算法进行图像匹配

SURF 算法是根据 SIFT 算法改进而来,也具备尺度不变性、旋转不变性和光照一致性,它很好地应用在增强现实和机器人建模等领域。通过构造 Hessian 矩阵,计算特征值。取 Hessian 矩阵为

$$\mathbf{H}(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial^2 y} \end{bmatrix} \quad (3)$$

当 $\det(\mathbf{H}) = 0$ 时,取得局部极值,则认为这个点是关键点。SURF 算法使用盒式滤波器代替高斯滤波,为了削减误差,则在 x 、 y 的二阶导的方向乘上 0.9 得到:

$$\det(\mathbf{H}_{approx}) = \mathbf{L}_{xx}\mathbf{L}_{yy} - (0.9\mathbf{L}_{xy})^2 \quad (4)$$

与 SIFT 算法不同,在尺度空间上,SURF 算法通过改变卷积核实现多尺度图像的并行处理,减少了算法时间复杂度。在选取旋转的主方向时,SURF 算法通过以 $6s$ (s 为兴趣的尺度大小) 邻域内,统计 Haar 小波特征,在 60° 内找到每个响应值赋以高斯权重系数,从而确定主方向 (如图 3 所示)。

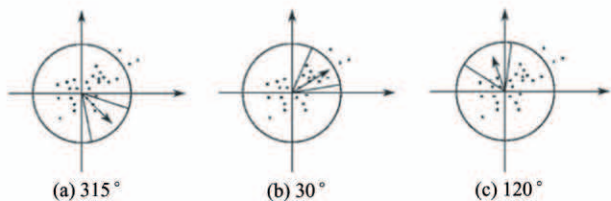


图 3 特征方向选择

计算 SURF 特征点的描述子时,在特征点周围选取 4×4 的区域块,每个块对应 $\sum dx$ 、 $\sum |dx|$ 、 $\sum dy$ 、 $\sum |dy|$ 4 个分向量,因此 SURF 描述子共有 64 维向量。采用 K-means 聚类算法^[10] 建立快速最近邻索引。

(1) 随机选取 k 个聚类质心点 (本文选取 32 个) 为 $u_1, u_2, \dots, u_k \in R^N$ 。

(2) 重复以下过程直到收敛。

① 对于每个点,计算其应该属于的类:

$$c^{(i)} = \underset{j}{\operatorname{argmin}} \|x^{(i)} - u^{(j)}\|^2 \quad (5)$$

② 对于每个类 j ,重新计算该类的质心:

$$u_j = \sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)} / \sum_{i=1}^m 1\{c^{(i)} = j\} \quad (6)$$

(3) 定义畸变系数。

$$J(c, u) = \sum_{i=1}^m \|x^{(i)} - u_c^{(i)}\|^2 \quad (7)$$

畸变系数表示每个点到其质心点的距离平方和,通过该算法将 J 约束到最小值。由于畸变系数是非凸函数,所以质心位置存在随机性会导致最优结果随机化。但在多次实验中发现,SURF 特征检测算法结合 K-means 匹配算法依然具备很好的鲁棒性和实时性。

1.3 关于单应性矩阵的位姿矩阵计算

在位姿估计方面,对成功匹配的两幅图像 (目标图像与检测到的区域图像) 进行单应性矩阵^[11] 计算,通过随机抽样一致^[12] (random sample consensus, RANSAC) 算法将不正常匹配的点剔除。当确定两幅图像对应点的坐标位置后,可以使用下式进行矩阵计算:

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (8)$$

对未知数 H_{ij} ($i, j \in 1, 2, 3$) 至少存在 3 组不共线的匹配点才能计算出单应性矩阵 $\mathbf{H}$,所以使用尽可能多的匹配点可以保证计算结果更具备鲁棒性。与传统方法一样,通过模板图像的 4 个边界角找到对应当前帧的 4 个点,即 $[x_i \ y_i \ 1]^T = \mathbf{H}[x_j \ y_j \ 1]^T$ 。对于增强现实的位姿计算来说,通过单目相机捕获的二维像素坐标和对应确定的三维世界坐标存在如下关系:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f/\Delta x & 0 & u_0 & 0 \\ 0 & f/\Delta y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (9)$$

其中等号右边的第 1 个矩阵表示相机的内参矩阵, f 为相机的焦距, Δx 和 Δy 表示相机捕获画面的 1 个像素点在真实世界中的尺寸大小, u_0 和 v_0 对应相机的焦距中心点。第 2 个矩阵表示相机的外参矩阵, R 为位姿的旋转矩阵, 表示为 3×3 的矩阵, T 表示平移向量, 表示为 3×1 的向量。式(9)中, 旋转矩阵 R 由世界坐标的 3 个角度构成, 因此其构成的 3 组方程中有 6 个未知量, 其中 1 组不存在未知量, 所以构成了 2 组方程的 6 个未知量。为了求出 6 个未知量的解, 则需要至少 3 组在平面坐标中不共线的点构成 6 组方程进行求解。对于相机内参和畸变系数求解, 可以通过文献[13]的棋盘标定法获得。

2 运动目标跟踪

基于自然特征增强现实的图像识别算法, 一般运用特征检测算法进行全局检测, 通过匹配到相应目标图像的相关特征点, 使用光流法或直接法进行持续跟踪。这样方法的优点是速度快, 但也存在鲁棒性差、虚拟物体易丢失等缺点。并且由于光流法和直接法本身的局限性, 不适合在户外等光照变换大的场景下使用, 所以这类方法不具备普遍适用性。本文针对此类算法的局限性, 提出了新的算法, 可适用于移动端和户外等光线强度变换较大的场景。本文通过联合 SURF 和改进后的卡尔曼滤波算法, 在连续跟踪过程中, 始终以最小的计算代价找到最大满足 SURF 计算条件的区域。

2.1 改进的卡尔曼滤波算法

卡尔曼滤波是一种经典的递归滤波器, 主要思想来自贝叶斯估计理论。本文的算法原理是利用它的误差协方差计算状态变量的预测值, 通过结合检测算法找到观测值并配合卡尔曼增益对该预测值进行校正, 最终获得该变量的最优值。卡尔曼滤波算法本身是一种线性系统, 由于本文中观测到的值就是状态值, 因此利用卡尔曼滤波算法的状态转移矩阵很容易实现从上一个状态到下一状态值的估计。卡尔曼滤波算法在一定程度上只考虑上下两帧之间的关系, 所以卡尔曼滤波算法需要一个很精确的观测效果。可是在实际操作过程中, 不论因为相机还是光照效果影响, 观测到的值却不一定准确。为了

解决此类问题, 本文提出一种改进思路, 借助前 N 组观测值, 建立一个非线性的拟合曲线进行预测下一组的观测值。当然这预测的“观测值”和系统观测到的值, 哪一个更准确需要另一个评估指标去完成。观测过程中由于检测到的特征点会因为环境或者相机抖动效果的影响而产生观测失误, 因此根据前 $N-1$ 组预测和观测的误差绝对值进行线性拟合, 保证了算法实时性的同时, 又可以很明显区分目标是快速移动还是因为观测失误而发生瞬间的移动, 改进原理如图 4 所示($\hat{X}_n$ 为拟合观测值, X_n 为实际观测值)。

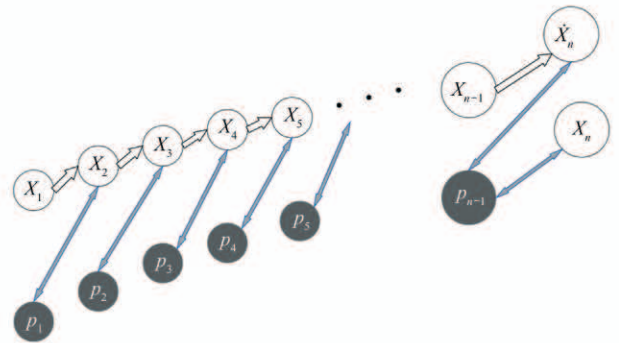


图 4 改进观测值选取原理图

2.2 跟踪算法流程

本文选取的观测值为目标图像 4 个边界角点的中心像素点($\frac{1}{2} \sum_{2i}^2 x_{2i}, \frac{1}{2} \sum_{2i+1}^3 y_{2i+1}$), 输入到卡尔曼滤波系统中得到预测值, 再以预测点为中心, 在上一帧中得到 $\max\{|x_j - x_i|\}$ 的值为矩形的宽, $\max\{|y_j - y_i|\}$ 的值为矩形的长, 重新裁剪一个新的区域。在这个区域内进行检测和计算相机位姿, 得到新的一组测量值, 由测量值和新的边界角点更新下一帧的矩形面积大小和区域。原理如图 5 所示(1 为当前观测情况, 2 为预测值情况)。

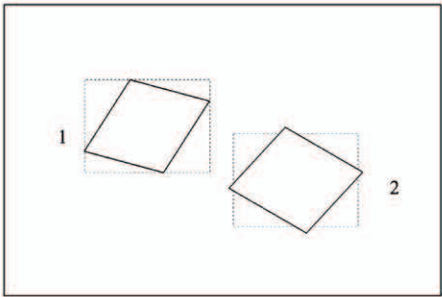


图 5 区域跟踪原理

首先根据卡尔曼滤波满足的线性条件建立一个四维的状态向量和二维的观测向量,四维的状态向量分别表示 x, y 的位置点和其对应的速度值。通过运动方程 $\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{B}_k \boldsymbol{\mu}_k + \mathbf{w}_k$ 表示,由于缺少控制向量,则设 $\mathbf{B}_k$ 为 0 向量,满足状态转移矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T$, 可以理解为上一个状态的位移量加上单位速度量则表现为当前状态的位移量,而考虑系统误差和观测误差,令 $\mathbf{w}_k \sim N(0, \mathbf{Q}_k)$, $\mathbf{v}_k \sim N(0, \mathbf{R}_k)$ 。观测方程根据卡尔曼滤波表示为 $\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$, 因为本文观测方程和状态方程的特殊关系, $\mathbf{H}_k$ 为 2×4 的矩阵,其中观测方程只和当前状态向量的前二维有关,即位移点,所以取 $\mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}^T$ 。为了满足系统对状态方程的最优估计,需要在观测阶段对状态值进行校正,此时引入协方差矩阵进行更新:

$$\mathbf{P}_{k|k-1} = \mathbf{A} \cdot \mathbf{P}_{k-1|k-1} \cdot \mathbf{A}^T + \mathbf{Q} \quad (10)$$

其中, $\mathbf{P}_{k|k-1}$ 表示预测状态值的协方差矩阵。通过当前系统的预测结果和现在状态的测量得到当前状态的最优化估计:

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H} \cdot \mathbf{x}_{k|k-1}) \quad (11)$$

$\mathbf{K}_k$ 表示当前的卡尔曼增益系数,由协方差矩阵 $\mathbf{P}$ 和测量矩阵 $\mathbf{H}$ 表示,即:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \cdot \mathbf{H}^T / (\mathbf{H} \cdot \mathbf{P}_{k|k-1} \cdot \mathbf{H}^T + \mathbf{R}) \quad (12)$$

将此时的卡尔曼增益带入最优估计求解,并利用这个增益计算下一时刻所需要的协方差矩阵值:

$$\mathbf{P}_{k|k} = (1 - \mathbf{K}_k \cdot \mathbf{H}) \mathbf{P}_{k|k-1} \quad (13)$$

此时可以完整地通过观测值预测下一时刻的中心点位置。但是观测值的选取却影响着整个系统的稳定。如果观测值出现较大的误差,而预测值也将不准确,在这里,本文通过改进对观测值的选取来改善整个系统的性能,方法如下。

(1) 初始化阶段

建立非线性的损失函数模型,本文设置模型为 $f(x) = \exp(a \cdot x^2 + b \cdot x + c)$ (14)

定义观测数据 N 组(本文 N 设置为 20),通过观测数据建立最小二乘函数:

$$\operatorname{argmin} \sum_x \|f(x) - \exp(a \cdot x^2 + b \cdot x + c)\|_2 \quad (15)$$

同时对前 N 组的初始值进行赋值求解,如果对

前 N 组全部赋值 0,则最后求得的参数容易陷入局部最优解,其待求参数在初始化阶段被错误求解出来,因此在 $f(x)$ 和 x 的取值为 0 的基础上加上高斯扰动,令其处于波动状态。

(2) 求解阶段

在相机运动或者目标物体运动过程中不完全符合本文所期待的运动方式,为了保证给定拟合的数据不随时增长导致拟合参数不准确并且计算量大幅提升,在接受新的数据同时剔除老旧的数据使其维持在 N 的参数范围内。利用 L-M 方法迭代求解,最后通过已知参数寻找下一时刻单位的解,即求解预测的“观测点”。

(3) 判定阶段

此时本文得到两组观测点,一组是实际观测到的点,另一组是预测的“观测点”。希望实际观测的点是准确的,但是不论是噪声存在或者光照因素影响,观测到的数据总可能是错误的,这里本文引入第三方评测指标来判断哪个值更准确。

假设之前的观测数据都是准确的(或者观测数据被矫正过),而卡尔曼滤波的预测值和下一帧的观测之间也存在误差,误差可能很小。将这些误差的绝对值建立一组拟合数据,根据误差增长率来判断选取哪一个“观测值”更可靠,同样建立 $N-1$ 组的数据来拟合线性方程,通过对第 N 组数据的预测来判断卡尔曼滤波预测的值和其“观测值”的相对误差更接近:

$$\mathbf{z}_k = \min_{\mathbf{z}_k} \| \{ 3 \times \| \hat{\mathbf{z}}_k - \mathbf{p}_{k-1} \| - \mathbf{g}_k \| \}, \{ \| \mathbf{z}_k - \mathbf{p}_{k-1} \| - \mathbf{g}_k \| \} \| \quad (16)$$

其中, $\mathbf{g}_k$ 为误差的预测, $\hat{\mathbf{z}}_k$ 为预测到的“观测值”, $\mathbf{z}_k$ 为系统的观测值, $\mathbf{p}_{k-1}$ 为上一帧的预测值。为了保证系统预测的可靠性,分配权重给两者以避免局部最优化。最后选取更接近的“观测点”作为新的观测点,算法的流程图如图 6 所示。

3 实验结果分析

本实验从实时性、稳定性两个因素进行分析,对 500 帧图像进行尺度变换、旋转变换、光照变换和视角变换这 4 种变化。使用全局 SURF、ORB、FREAK^[14]、

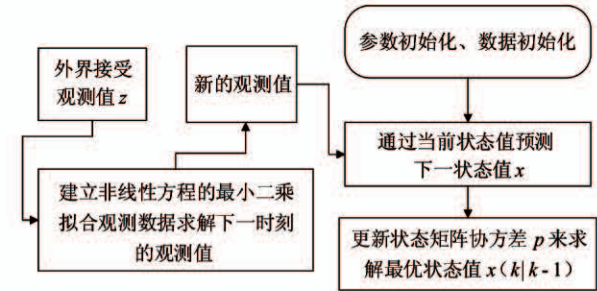


图 6 改进后卡尔曼预测更新流程图

ORB-LK^[15]和本文提出的算法进行比对分析,观察运行帧率和正确姿态注册的图像数目,采用 fps 为单位进行实时性比对,稳定性的单位采用正确姿态的注册百分比。本次实验使用的测试平台是 Inter(R) Core(TM) i7-2600 CPU @ 3.2 GHz、8 G 的 DDR3 内存、罗技摄像头、OpenCV3.3.1 图像视觉库、OpenGL 绘图库、Ubuntu16.04 LTS 操作平台。本文的总体流程图如图 7 所示。

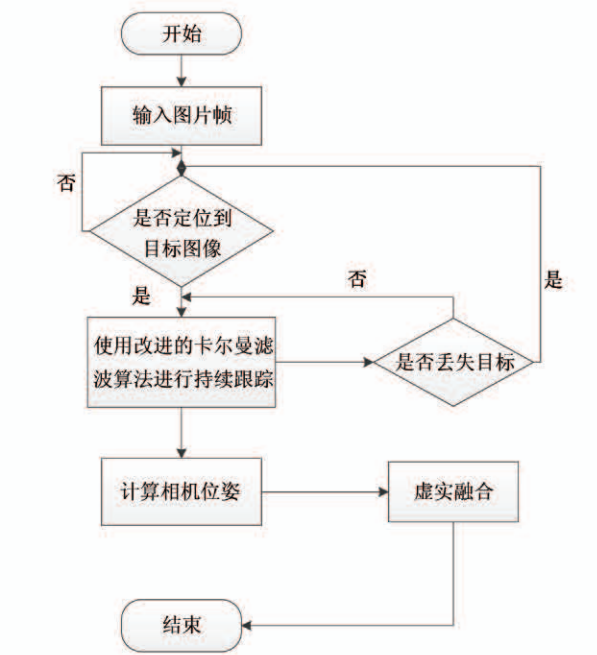


图 7 全局流程图

3.1 全局定位检索结果与分析

本次实验设定摄像头采集的分辨率为 640 × 640 像素、给定的模板图像分辨率为 415 × 585 像素。为了提高稳定跟踪的实时性,在初始化和目标丢失时需要重新定位目标图像的粗略位置以减少不必要的搜索空间。从实时性方向考虑,对给定的

500 帧图像进行采样分析,每 10 帧计算其平均帧率,采样 50 次,用本文算法和传统方法如 SURF、ORB、使用 SURF + FREAK 的组合的方式,和深度学习方法如 SuperPoint^[16](self supervised interest point detection and description)分别进行比较,图中的最低值为 0,最高值为 60,结果如图 8 所示。

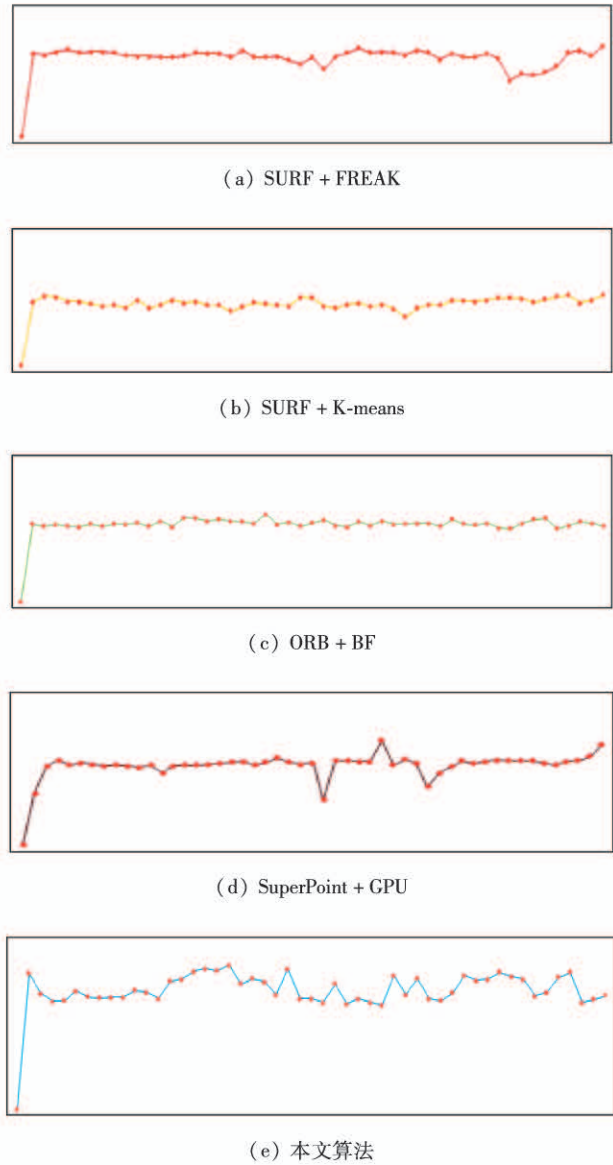


图 8 定位算法帧率

其中前 3 种算法属于直接检索全局图像进行实时匹配分析图像的位姿关系,而本文为了缩减前期的定位时间,使用 FAST 算法进行采集像素信息建立直方图的方法以快速确定目标图像所在位置。本文算法中为了确保定位图像的大部分区域属于目标图像,在相似度检测算法方面设置阈值为 0.6。

3.2 改进后卡尔曼滤波器的跟踪精度分析

本文使用卡尔曼滤波器的预测值进行尺度定位,结合 SURF 算法的观测值不断修正预测值结果。与其他目标跟踪算法不同,增强现实(AR)的跟踪算法不仅需要目标的位置和尺度信息,更依赖相机的姿态信息。目前已知关于位姿跟踪的相关算法有 IC(inerse compositional)^[17]、ESM(improving vision-based control efficient second-order minimization techniques)^[18]、Struck(structured output tracking with Kernels)^[19]、Gracker(graph-based planar object tracker)^[20]等。其中基于模板的 IC 和 ESM 算法对于遮挡不具备鲁棒性,Struck 对于光照性比较差,而 Gracker 算法依赖的 SIFT 描述子,在实时性方面不具备优势。因此本文使用卡尔曼滤波的方法,预测目标的位置和尺度信息,通过观测位姿信息校正目标位置和尺度,跟踪效果如图 9 所示。



图 9 预测跟踪效果

如图 9 所示,□位置框表示观测状态值,▼位置框表示选定的预测值,▽位置框表示预测的“观测”值,□对应点表示观测点,▼对应点表示预测点,▽对应点表示预测的“观测点”。每隔 10 帧进行取样,得到一组如上所示轨迹图。随机选取目标图像由于目标小,而预测的“观测点”失误,本系统很好地分析了哪组观测点为正确的观测点。对图 9(b)采样 500 帧图像,判断预测区域和真实区域的交并

表 1 采样跟踪错误率

帧数	50	100	150	200	250	300	350	400	450	500
丢失率	2%	5%	4.6%	5.5%	5.3%	5.6%	6%	6%	6.67%	6.8%

比是否小于 80%,如果小于此值判断为丢失状态。每 50 帧计算丢失次数,共计 10 次。

3.3 跟踪注册结果分析

本文提出算法的实验效果为图 10 所示。



图 10 三维注册效果

为了验证本文算法的有效性,在不同的目标上进行不同的虚拟物体注册,得到如图 11 所示效果。

如图 9 所示,本文提出的算法能实时预测下一

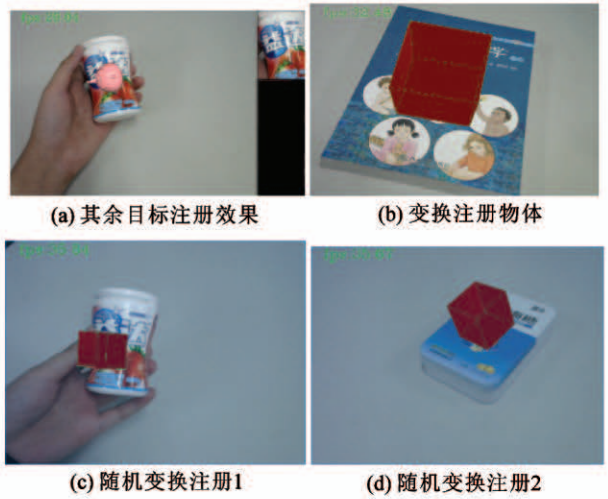


图 11 随机变换注册

帧目标图像大体的位置,并适时地调整检测区域大小,得到效果如图 10 的右侧小窗口所示。通过这样的检测方法,得到的误差点少,合适的匹配点多,因此能尽可能地减少匹配误差,也能通过检索缩减的匹配空间来增加实时性能。本次实验中,采集前

500 帧图像,对第 50 帧、100 帧、200 帧、350 帧和 500 帧进行采样,通过对 SURF、ORB、SURF + FREAK 和 ORB-LK 等几种经典算法和 SuperPoints 基于深度学

习的特征点检测和匹配方法进行实时性和稳定性比较,得到如表 2 和表 3 所示的结果。

表 2 各算法的实时性分析(单位为 fps)

图像帧	SURF + K-means	ORB + BF	SURF + FREAK + BF	ORB-LK	SuperPoint + GPU	本文算法
50	21.3515	32.2883	32.4721	48.3243	26.2601	42.0568
100	21.3621	32.2214	32.5737	48.2101	25.9905	42.555
200	21.7201	32.4663	31.9754	48.0205	28.3223	42.6182
350	21.7454	32.3616	31.9381	48.8001	28.8026	42.5172
500	21.6351	32.3829	31.8698	48.1209	27.3372	42.2314

表 3 各算法的稳定性分析

图像帧	SURF + K-means	ORB + BF	SURF + FREAK + BF	ORB-LK	SuperPoint + GPU	本文算法
50	100%	94%	94%	94%	100%	100%
100	100%	94%	96%	87%	99%	99%
200	95%	88.5%	96%	84%	94.6%	96%
350	94%	84.57%	92%	77.42%	93.23%	96.57%
500	94.8%	82.8%	91%	75.8%	91.1%	92.4%

从表中数据分析,使用 ORB 结合光流法代替匹配,可以大幅度提高算法的实时性,但是由于光流法对光照的敏感性,在第 200 帧之后稳定性大幅度下降。使用 SURF 算法可以提高注册结果的稳定性,但是由于算法本身的时间复杂度高,所以其损耗的时间也比其他对比的算法时间长。本文算法在实时性方面比单纯的 SURF 匹配提高了两倍,其次在稳定性方面和 SURF 全局算法相差无几,因此本文提出的算法在几种变换下都能很好地满足实时性和稳定性要求。

4 结 论

本文提出了一种基于改进卡尔曼滤波的增强现实跟踪注册算法。该算法使用 FAST 检测器在不同的金字塔层进行采样检测并通过使用相关性比较的方法进行快速定位目标区域,以帮助系统在丢失再恢复时提高了稳定性。采用改进后的卡尔曼滤波结合 SURF 对目标区域的位置和大小进行观测从而预测下一帧目标图像的区域大小和位置,以帮助提高

系统的实时性。算法使用最小二乘的方法对卡尔曼滤波算法进行改进,通过收集前 N 帧的先验信息对下一帧进行约束矫正提高了系统的稳定性。利用图像的二维坐标信息结合三维空间的坐标信息,使用 PnP 算法获得相机的位姿,利用 OpenGL 绘制虚拟物体,加载视图矩阵以达到注册的目的。实验表明,该算法可以稳定在 40 帧以上的均率运行,如果在稳定性方向降低要求,使用 ORB 算法代替用来观测的 SURF 算法还可以提高 10 帧左右的工作效率。总体上,本文提出的算法在光照变换、旋转变换、尺度变换方面都具备鲁棒性,并且具有优越的实时性能,同样适用于相关的位姿跟踪领域。

参考文献

[1] 高翔,安辉,陈为,等. 移动增强现实可视化综述[J]. 中国计算机辅助设计与图形学学报, 2018, 30(1):1-8

[2] 鲍虎军,章国锋,秦学英,等. 增强现实:原理、算法与应用[M]. 北京:科学出版社, 2019: 1-2

[3] Bay H, Tuytelaars T, Van Gool L. SURF: speeded up robust features[C] // European Conference on Computer Vision, Graz, Austria, 2006:404-417

[4] Rublee E, Rabaud V, Konolige K, et al. ORB: an effi-

- cient alternative to SIFT or SURF [C] // 2011 International Conference on Computer Vision, Barcelona, Spain, 2012; 6-13
- [5] Nguyen T, Chen S W, Shivakumar S S, et al. Unsuper-vised deep homography: a fast and robust homography estimation model [J]. *IEEE Robotics and Automation Letters*, 2018, 3(3): 2346-2353
- [6] Verdie Y, Yi K M, Fua P, et al. TILDE: a temporally invariant learned detector [C] // Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 5279-5288
- [7] Yi K M, Trulls E, Lepetit V, et al. LIFT: learned invariant feature transform [C] // Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 467-483
- [8] 董辉, 王亚男, 童辉, 等. 基于自适应 Kalman 滤波的移动机器人人体目标跟踪 [J]. 高技术通讯, 2019, 29(3): 249-256
- [9] 徐少平, 刘小平, 李春泉, 等. 基于区域最大相似度的快速图像分割算法 [J]. 光电子·激光, 2013, 24(5): 990-998
- [10] 熊忠阳, 陈若田, 张玉芳. 一种有效的 K-means 聚类中心初始化方法 [J]. 计算机应用研究, 2011, 28(11): 4188-4190
- [11] 许金山, 王一江, 程徐, 等. 单源性矩阵自适应估计方法 [J]. 计算机工程与应用, 2016, 52(5): 160-164
- [12] Matas J, Chum O. Randomized RANSAC with d test [J]. *Image and Vision Computing*, 2004, 22(10): 837-842
- [13] Zhang Z. A flexible new technique for camera calibration [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334
- [14] Ortiz R. FREAK: fast retina keypoint [C] // IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012: 510-517
- [15] 闫兴亚, 薛红红, 李乐. 基于 ORB 与光流法的增强现实注册算法研究 [J]. 电子设计工程, 2017, 25(13): 40-43
- [16] De Tene D, Malisiewicz T, Rabinovich A. Super point: self-supervised interest point detection and description [C] // Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, USA, 2018: 2160-2168
- [17] Baker S, Matthews I. Lucas-kanade 20 years on: a unifying framework [J]. *International Journal of Computer Vision*, 2004, 56(3): 221-255
- [18] Malis E. Improving vision-based control using efficient second-order minimization techniques [C] // IEEE International Conference on Robotics and Automation, New Orleans, USA, 2004: 1843-1848
- [19] Hare S, Saffari A, Torr P H S. Efficient online structured output learning for keypoint-based object tracking [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012: 1894-1901
- [20] Wang T, Ling H. Cracker: a graph-based Planar Object Tracker [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(6): 1494-1501

Augmented reality tracking registration algorithm based on improved Kalman filter

Liu Jia, Yan Dong, Guo Bin

(School of Automation, Nanjing University of Information Science and Technology, CICAET, Nanjing 210044)

Abstract

In order to improve robustness and real-time performance of traditional augmented reality systems applied in feature point matching, a new region tracking matching has been proposed. Firstly, features from accelerated segment test (FAST) algorithm is used to extract pixel information on different scales to establish a low-dimensional template for fast positioning. Secondly, in the tracking stage, speeded-up robust features (SURF) algorithm is used to detect the key points on-line and real-time matching. At the same time, the improved non-linear Kalman filter algorithm is used to predict the motion trajectory and reduce the matching area. Part of the image frame from the above algorithms is intercepted, and the observed value of the previous frame is used to predict the next frame and estimate the camera attitude. At the same time, random sample consensus (RANSAC) algorithm is used to eliminate outliers. Finally, the registration of virtual objects is superimposed through the camera position matrix gained. In this paper, the algorithm is equipped with advantages in comprehensive evaluation. It has robustness in illumination transformation, rotation transformation and scale transformation, as well as superior real-time performance.

Key words: Kalman filtering, features from accelerated segment test (FAST) algorithm, speeded-up robust features (SURF) algorithm, augmented reality (AR), attitude estimation