

基于改进随机蕨的增强现实场景实时跟踪注册算法

赵越,李晶皎,李海鹏,杨丹

(东北大学 信息科学与工程学院,辽宁 沈阳 110819)

摘 要: 针对增强现实场景跟踪注册存在速度较慢等问题,提出了一种基于改进随机蕨的增强现实跟踪注册算法.该算法采用离线训练和在线跟踪两个模块.提出了一种嵌入式蕨分类器以提高特征点匹配精度,该分类器采用有监督的降维方法,并利用了所有可能的信息.通过该分类器进行特征匹配,进而计算摄像机位姿并渲染注册虚拟物体.实验结果证明,提出的嵌入式蕨在平均分类精度上优于其他算法.平均处理每帧图像的时间为 34.22 ms,基本满足实时性.

关 键 词: 随机蕨;降维;分类器;跟踪;增强现实

中图分类号: TP 301 文献标志码: A 文章编号: 1005-3026(2016)05-0614-05

Real-Time Tracking and Registration Algorithm of Scenarios of Augmented Reality Based on Improved Random Fern

ZHAO Yue ,LI Jing-jiao ,LI Hai-peng ,YANG Dan

(School of Information Science & Engineering ,Northeastern University ,Shenyang 110819 ,China. Corresponding author :ZHAO Yue ,E-mail :zy_ky7777@126.com)

Abstract : Because of the slow speed of tracking and registration on scenarios of augmented reality ,an algorithm was proposed based on the improved random fern on tracking and registration of augmented reality. The proposed algorithm has the offline training section and the online tracking section. A classifier based on the embedded fern was proposed. In the developed classifier ,a supervised dimensionality reduction method and all possible information were used. What 's more ,it was also used for the feature matching ,and then camera pose was computed and virtual objects were rendered and registered. Experimental results show that the proposed algorithm is superior to other algorithms on the average classification accuracy. The average processing time of each frame is about 34.22 ms ,which can almost meet the real-time.

Key words : random fern ;dimensionality reduction ;classifier ;tracking ;augmented reality

增强现实(augmented reality ,AR) ,是一种将虚拟信息通过计算机系统叠加到真实世界的技术^[1].对于增强现实技术来说,识别和跟踪自然特征相对来说更加困难.而较大的场景则包含了更多更复杂的情况,其跟踪定位问题更成为学者们近年来的研究热点.

Davison 等^[2]提出了基于定位和三维重建(SLAM)同时进行的场景注册方法,但其时间复杂度为 $O(N^2)$, N 为特征点个数,因此只适用 N 不太大的情况.张运超等^[3]提出了一种通过智能手机实现智慧城市导览的增强现实方法.该方法采用动态区域划分及 BHKM 聚类的大量场景识别算法,能够实现用户位置的精确定位. BRISK 与光流结合的特征跟踪注册方法满足了用户多尺度、多时空的信息需求,但无法处理光照、遮挡等情况. Mooser 等^[4]提出了一种鲁棒的运动中恢复结构的方法应用于较大场景的定位中.该方法离线阶段建立 3D 点云,在线阶段通过近似最近邻匹配方法进行 2D-3D 特征点匹配,从而进行增强现实的注册.但该方法的运行时间也依赖于 3D

点云的大小,当 3D 点云较大时,2D-3D 特征点匹配耗时很长,实时性不够好。

针对以上问题,本文提出了一种基于改进随机蕨的增强现实跟踪注册算法。本文提出的方法在运行时间上不依赖于真实 3D 点云的大小,基本能够满足实时性需求,通过利用所有可能的特征信息使定位更加准确。

1 改进的增强现实实时跟踪注册算法

本文提出的增强现实场景跟踪注册算法包括离线训练阶段和在线跟踪注册阶段,算法流程图如图 1 所示。

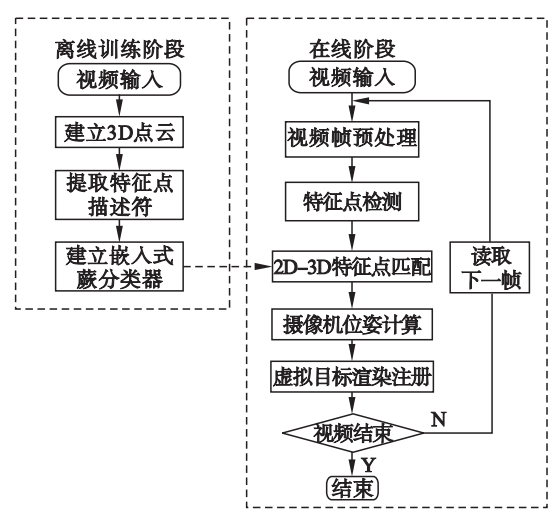


图 1 本文算法流程图
Fig. 1 The flowchart of the algorithm

1.1 离线训练阶段

在离线训练阶段,主要完成以下几个步骤:

- 1) 输入训练视频流。
- 2) 用 Subtrack Optimization^[3] 的 SFM (structure from motion) 方法建立 3D 点云。
- 3) 对 3D 点云中的 3D 点进行特征描述符提取。对于点云,它自己仅仅包含了几何信息,称为每个结构点的 3D 定位。对于每个 3D 点必须和一个可以在现阶段相匹配的视觉描述符相关联。使用 Walsh-Hadamard 描述符提取特征点,每个 32×32 像素大小的图像块被投影在一个 20 维的 Walsh-Hadamard 核上^[5]。
- 4) 对提取的特征点描述符建立分类器,此分类器采用本文基于随机蕨的嵌入式分类器。

1.2 在线跟踪注册阶段

在线跟踪注册阶段,主要完成以下几个步骤:

- 1) 输入非训练视频流。

- 2) 对视频流进行预处理。
- 3) 利用 FAST 方法进行特征点检测,并提取特征点 Walsh-Hadamard 描述符。
- 4) 将提取到的描述符投入到离线阶段训练得到的分类器中,进行 2D-3D 特征点的匹配。
- 5) 利用顺序抽样一致性算法(PROSAC)对已匹配特征点对估计两图像的外极几何关系,计算摄像机位姿。
- 6) 渲染注册虚拟目标,直到视频的最后一帧结束。

2 基于随机蕨的嵌入式分类器

2.1 把特征点匹配看作分类问题

给定一个视频帧图像 Q ,增强现实的目标是关于提供的稀疏 3D 点云而确认精确的 6 自由度摄像机位姿,进而进行虚拟注册。该方法的主要步骤如下:①提取图像 Q 的特征;②通过特征匹配建立 2D-3D 的对应;③摄像机位姿估计。

这里得到的稀疏 3D 点模型由 M 个 3D 点 P^i 组成。对于每一个 3D 点 P^i ,用 $D_i^j \in \mathbb{R}^D$ 表示 3D 点 P^i 的一系列描述符,即包含被用来对 P^i 作三角测量的图像点的所有特征描述符。需要注意的是,一些 3D 点拥有数百个相关联的特征描述符,且这些具有最大量的对应描述符的点对于定位来说可能拥有最大的信息量,例如一个塔上的一点能够从许多视角观察到。目的是将在视频帧图像 Q 中检测到的这 N 个 2D 特征点(其描述符用 $Q_i \in \mathbb{R}^D$ 表示)与 M 个 3D 点 P^i 进行匹配。通过制定把匹配当做分类问题而使用所有可用的信息。在训练分类器后,不必保留训练样本,因此,允许将所有描述符用于训练。其基本思想是,每个三维点由一个唯一的类标记表示(1,..., M),及其相应的描述符被作用于判别分类标记的训练数据。其利用了判别设定的所有描述符(对比于最近邻分配),这将产生一种改进的集的对对应关系。

训练后,把所有查询描述符 Q_i 放到分类器中,并获得一个 $1 \times M$ 的向量类条件概率,表示假设对应到 3D 点的确定性信息。这里的分类方法不需要保存描述符,因为如果最高类条件概率和次高类条件概率间的距离比小于阈值 ε ,那么就确定了对应关系。

对于分类来说,本文提出了一种对随机蕨原理的扩展,在分类过程中利用了所有可用的描述符,而几乎保持了相同的运行时间和存储信息。

2.2 基本随机蕨

随机蕨算法是一种机器学习分类方法,是在随机森林算法的基础上改进而来的,是一种半朴素贝叶斯分类器(SNBC ,semi-naive Bayesian classifier)^[6]。随机蕨算法通过对随机森林中树的每层节点选取相同决策而建立一种非层次的分配结构。

随机蕨由全部基分类器组成。在每一个基分类器中,通过做 S 个二元决策使得特征空间被唯一分离到 $L=2^S$ 个箱子中的一个。在训练过程中,计算每个箱子类的发生,定义为类依赖概率。基分类器决策以半朴素贝叶斯方式联合。训练的目的是决定这 S 个分离参数,并估计到蕨 F_m 和类 c_i 的每个箱子 l 的类条件概率 $P(F_m=l|c_i)$ 。简化为

$$p_{l\ c_i}=P(F_m=l|c_i)。(1)$$

其中 l 表示蕨 F_m 的 $L=2^S$ 个箱子中的一个, $p_{l\ c_i}$ 是箱子 l 投票给类 c_i 的概率。因此,把所有类条件概率写进一个 $F \times L \times C$ (蕨的数量 \times 箱子的数量 \times 类的数量)的矩阵 F 中,且 $\sum p_{l\ c_i}=1$ 。在训练过程中,通过简单计算标记的数据点进到对应的箱子的数量来估计概率:

$$p_{l\ c_i}=\frac{N_{l\ c_i}}{N_{c_i}}。(2)$$

其中 $N_{l\ c_i}$ 是估计到蕨箱子 l 的类 c_i 的被标记的训练样本的数量, N_{c_i} 是类 c_i 的全部样本数量。

一旦通过训练数据估计得到蕨矩阵 F ,新的测试数据点被放到所有的蕨 F_m 中,对应箱子 l 的概率,测试点估计以半朴素贝叶斯方式联合,最终分配到类 c^* 为

$$c^*=\arg \max P(x|c_i)=\prod_{m=1}^F p_{l\ c_i}。(3)$$

2.3 嵌入式蕨分类器

本文的方法建立在随机蕨的基础上。假设已经拥有了一组包含 N 个数据点的集合 $\{X_1, X_2, \dots, X_N\}$,该集合在一个 D 维特征空间,即 $X_i \in \mathbf{R}^D$ 。把数据点堆叠到一起生成了 $N \times D$ 训练数据矩阵 $X \in \mathbf{R}^{N \times D}$,其中,单元元素 x_{id} 表示第 i 个数据点的第 d 个特征。

和传统的随机蕨相似,主要目的是把一个测试数据点分配到每个蕨 $L=2^S$ 个箱子中的一个。因此,目标是要找到 D 维数据点到第 m 个蕨的 S 维二进制分配向量 $b^m \in \{0,1\}^S$ 的映射,这唯一地分配了特征空间并允许把每一个特征点分配到 $L=2^S$ 个箱子中的一个。

在矩阵符号中,传统的蕨获得二进制分配向量 b^m 通过随机选择特征 IDs 和通过分析对应数

据点记录:

$$b^m=X^m>0\ X^m=\{x_{id}|i\in\{1,\dots,N\}\ d\in K^m\}。(4)$$

其中 $K^m=(r_1^m\ r_2^m\ \dots\ r_S^m)$ 是一组 S 个随机选择特征 IDs。

本文方法的核心思想是选择更多的 $B>S$ 的特征,然后映射数据点的 B 个特征到最后的 S 维空间。对于每一个蕨随机选择一定数量的特征 $B \leq D$,但是 $B>S$,对第 m 个蕨减小输入训练矩阵 X 到一个蕨特定的 $N \times B$ 输入矩阵 $X^m=\{x_{id}|\forall i\text{ and }d\in K^m\}$,其中 K^m 是一组 B 个随机选择特征 IDs。这里,进一步定义二进制分配向量 b^m 为一组线性组合:

$$b^m=(X^mV^m+J^m)>0。(5)$$

其中 V^m 是 $B \times S$ 映射矩阵, J^m 是一个 $S \times 1$ 位移向量。最直接的定义映射矩阵 V^m 的方式是在一种随机状态下选择它,这就经常产生一种被称为倾斜森林的分类器^[7]。然而,这将忽略有用的训练数据标志,可以被利用找到一个改进的映射。基于这个原因,每个蕨确定一个映射矩阵,可以从各类间最佳的区分不同的类。

本文核心思想是为了捕获训练数据集 X^m (由随机选择的特征组成)的结构用于有监督的降维方法,确定最佳二元分离参数($V^m\ J^m$)。为达到上述目的,本文提出使用典型相关分析(CCA)^[8]方法,能够从两个不同视图有效识别一般潜在的空间。这样,关联给定的特征和相对应的标记空间,组成了找到最佳二元分离的基础。

假设每个数据点 x_i 已经关联了相应的标签 $y_i \in \{1\ 2\ \dots\ C\}$,则可以建立一个数据集标签矩阵 $Y \in \{0\ 1\}^{N \times C}$,如果数据点 i 属于类 C ,则相对应的条目 y_{ii} 是 1,否则是 0。典型相关分析(CCA)试图找到两个基础向量 v_x 和 v_y 的集合,以便这些基础向量上的变量映射 X^mv_x 和 Y^mv_y 之间的关联被相互最大化。相关系数 ρ 定义为

$$\rho=\frac{v_x^TXY^Tv_y}{\sqrt{(v_x^TXX^Tv_x)(v_y^TYY^Tv_y)}}。(6)$$

系数 ρ 相对于基础向量按比例不变,因此,CCA 也可以用式(7)表示:

$$\left. \begin{aligned} \max_{v_x\ v_y} \quad & v_x^TXY^Tv_y, \\ & v_x^TXX^Tv_x=1 \\ & v_y^TYY^Tv_y=1. \end{aligned} \right\}。(7)$$

为了找到基础向量,取相关系数 ρ 的最大值,必须解决一个广义特征值问题。用特征值问题的

正则化来避免过度拟合,通过等式(8)解决:

$$XY^T(Y Y^T + o_y I)^{-1} Y X^T v_x = \mu_x (X X^T + o_x I) v_x. \quad (8)$$

其中 o_x 和 o_y 是两个正则化参数, μ_x 是系数, I 是单位矩阵. 正如文献[9]研究所示,通过 o_y 的 Y 的正则化并没有影响 X 的映射,因此,这里设 $o_y = 0$,根据文献[7],使 o_x 趋近于 10^{-6} .

考虑 S 主导的特征向量,从式(8)定义的广义特征值问题的解决而获得,像 $B \times S$ 映射矩阵 V^m ,允许将任何新的数据点放入这个新的嵌入式空间. 在每一维分别定义分离值,通过简单的阈值,投入的值在训练数据值的中值,产生位移向量 J^m . 这样,通过式(5),最后获得了所需要的 $S \times 1$ 二元向量 b^m ,唯一分配数据点到 2^S 个箱子中的一个. 图2说明了本文的嵌入式蕨用于一个单独蕨上的概念.

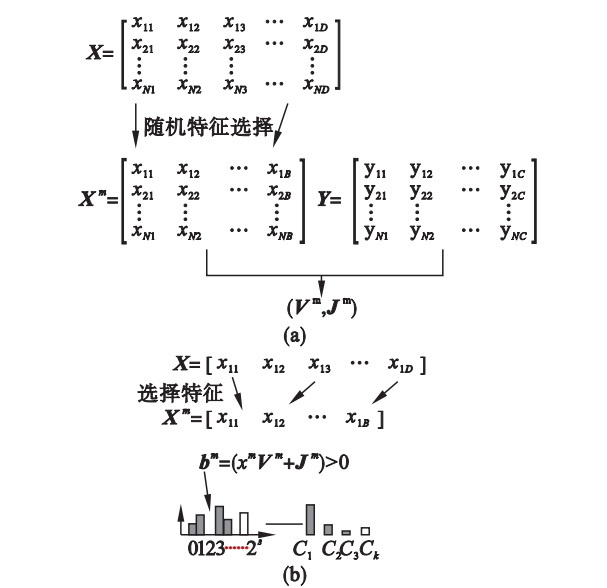


图2 本文方法用于一个单独的基础分类器
Fig. 2 The method for a single base classifier
(a)—训练过程;(b)—测试过程.

①输入数据矩阵 X ,通过随机选择特征降维集合 K^m 而减少到一个特定蕨矩阵 X^m . 这个矩阵和提供的 CCA 中的标签矩阵一起使用,提出了一个被 V^m 定义的新的嵌入式空间. 这种映射使每一个训练样本能够分配到一个箱子,和类的条件概率 $p_{l \in i}$ 的计算一样. ②在测试同样选择的特征降维时,通过训练学习到的映射被应用到把样本分配到一个箱子(二进制向量 b^m). 最后的基础分类器概率以半朴素贝叶斯方式联合得到. 以上步骤仅仅是一个单独的基础分类器的形成和使用过程. 重复前面的描述步骤,通过映射矩阵和位移向量 (V^m, J^m) 获得了一组随机蕨. 注意,这里获得的映射矩阵,降维特征矩阵 X^m 对于每个蕨是不同

的,而对于所有 CCA 步骤,标签矩阵 Y 是相同的. 因此,和传统随机蕨形成对比,对于每个蕨仅通过选择特征降维来注入随机性,因为每个蕨的映射被估计在一个特定的方式. 最终的映射到二元向量 b^m 被用来定义类的条件概率 $p_{l \in i}$,通过计算估计到相应箱子的类成员的数目来得到. 在测试过程中数据点映射到每个蕨的不同映射空间使用式(5)中的 (V^m, J^m) ,提供了类条件概率,最终分类结果是通过联合如式(3)中描述的半朴素贝叶斯方法得到的蕨的概率.

3 实验结果与分析

实验在台式计算机上进行,用以评估本文的算法,CPU 主频为 3.1 GHz,内存为 4 GB,显卡为 NVIDIA GeForce GT 630 M.

假设已通过任意 SFM 方法生成一个感兴趣区域的三维重构,已得到了所有摄像机位姿,稀疏点云的三维坐标,及被用于 3D 点和所有相应描述符的三角测量的所有图像点列表.

3.1 几种匹配算法的性能对比

这里使用两个常用的数据库 Dubrovnik 和 Rome^[10]来进行算法性能的评估. Dubrovnik 由约 6 千个图像,200 万个点和 1 千万个描述符组成. Rome 由约 1 万 5 千个图像,400 万个点和 21 亿个描述符组成. 两个数据库中,从稀疏 3D 点云确认潜在的视觉集,把它们放入到几个更小的视图单元. 潜在的视觉集决定了 3D 点云和在当前视图单元里必须考虑的对应的描述符子集.

通过使用本文提出的分类器方法进行特征匹配和其他标准最近邻匹配方法、随机蕨方法和随机森林方法相比较. 用 K 折交叉验证($K = 10$)验证每个视觉集. 而最近邻方法采用 $K - D$ 树最近邻方法. 对所有蕨方法使用相同的基分类器数 $F = 30$. 随机蕨中每个基分类器采用 10 个二元决策. 随机森林中的树是完全分裂树,且固定估计节点测试的数量为 \sqrt{D} . 本文提出分类器的 S 取值 16, B 取值 25. 表 1 显示了在两个数据库上的平均分类精度. 从表中可以看出本文提出方法在两个数据库上的实验结果,在平均分类精度上高于其他几种方法.

运行时间上,本文提出方法不依赖于 3D 点的数量,且对于分类器来说是基本稳定的. 本文提出方法比随机蕨方法平均时间长,是其 1.12 倍,这是因为映射的原因. 而随机森林相对更慢一些,是随机蕨方法的 1.41 倍. 最近邻方法的运行时间

依赖于 3D 点的数量 ,所以 ,当数据库中 3D 点的数量超过几百个时 ,其运行时间变得很慢.

表 1 不同算法采用 10 折交叉验证法的平均分类精度

Table 1 Mean classification accuracies for 10-fold cross validation in different methods %

算法	Rome	Dubrovnik
随机蕨法	87. 85	92. 14
随机森林法	88. 76	96. 21
最近邻法	94. 28	95. 87
本文的法	97. 33	98. 76

3. 2 增强现实注册结果

实验采用室内视频序列 room(长度为 786 帧)和室外视频序列 building(长度为 1 372 帧) , 视频序列均为 320 × 240 像素、帧率为 30 帧/s. 图 3 为本文算法的增强现实跟踪注册结果. 从图 3 可以看出 ,无论是室内场景还是室外场景 ,注册效果令人满意 ,证明了本文提出算法的有效性.



图 3 本文算法增强现实跟踪注册结果
Fig. 3 The results of tracking and registration of the proposed algorithms
(a)—为视频序列 room 的跟踪注册结果 ;
(b)—为视频序列 building 的跟踪注册结果.

表 2 是不同算法每帧图像的增强现实平均处理时间. 将本文提出的算法与文献 [2] 的算法进

表 2 不同方法每帧图像的增强现实平均跟踪注册时间

Table 2 The average processing time of augmented reality of each frame in different algorithms ms

步骤	本文算法	文献 [2] 算法
图像预处理	1. 20	1. 20
特征点检测和描述	4. 10	7. 40
特征点匹配	19. 40	143. 30
摄像机位姿计算	8. 30	8. 30
OpenGL 绘图	1. 22	1. 22
总时间	34. 22	161. 42

行比较 ,本文提出算法的每帧图像处理的平均总时间为 34. 22 ms ,基本可以满足实时性要求. 而文献 [2] 的算法的平均总时间为 161. 42 ms ,是本文提出方法的 4. 7 倍之多 . 这是因为 3D 的数量太多从而导致匹配过程消耗了过多时间.

4 结 论

- 1) 本文提出的嵌入式蕨在平均分类精度上有所提高.
- 2) 本文提出的算法的每帧图像处理的平均总时间为 34. 22 ms ,基本可以满足实时性要求.
- 3) 本文提出的算法在室内外场景下 ,注册效果均令人满意.

参考文献 :

[1] Azuma R ,Baillot Y ,Behringer R ,et al. Recent advances in augmented reality [J]. *IEEE Computer Graphics and Applications* 2001 21(6) :34 – 47.

[2] Davison A ,Reid I ,Molton N D ,et al. Mono-SLAM :realtime single camera SLAM[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* ,2007 ,29(6) :1052 – 1067.

[3] 张运超 ,陈靖 ,王涌天 ,等. 基于移动增强现实的智慧城市导览[J]. *计算机研究与发展* 2014 51(2) :302 – 310.
(Zhang Yun-chao ,Chen Jing ,Wang Yong-tian ,et al. Smart city guide using mobile augmented reality[J]. *Journal of Computer Research and Development* ,2014 ,51(2) :302 – 310.)

[4] Mooser J ,You S ,Neumann U ,et al. Applying robust structure from motion to markerless augmented reality[C]// *Applications of Computer Vision*. Snowbird 2009 :1 – 8.

[5] Hel-Or Y ,Hel-Or H. Real-time pattern matching using projection kernels [J]. *Pattern Analysis and Machine Intelligence* 2005 27(9) :1430 – 1445.

[6] Zheng F ,Webb G I. A comparative study of semi-naive Bayes methods in classification learning [C]// *Proceedings of the Fourth Australasian Data Mining Conference*. Sydney , 2005 :141 – 156.

[7] Heath D ,Kasif S ,Salzberg S. Induction of oblique decision trees[J]. *Journal of Artificial Intelligence Research* ,1993 , 135 :3 – 12.

[8] Hotelling H. Relations between two sets of variates[J]. *Biometrika* ,1936 28 :312 – 377.

[9] Sun L ,Ji S ,Ye J. Canonical correlation analysis for multilabel classification : a least-squares formulation , extensions ,and analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2011 33(1) :194 – 200.

[10] Li Y ,Snavey N ,Huttenlocher D P. Location recognition using prioritized feature matching[C]//*Proceedings of 11th European Conference on Computer Vision*. Heraklion 2010 : 791 – 804.