

# 基于 ORB 自然特征的 AR 实时系统实现\*

任 结, 周 余, 于 耀, 都思丹, 王自强

(南京大学 电子科学与工程学院, 南京 210009)

**摘要:** 针对当前基于自然特征的增强现实效率低, 提出一种新的基于 ORB 自然特征的实时注册方法。提取视频帧与基准图像中的 ORB 特征点, 使用汉明距离匹配, 利用 RANSAC 算法筛选得到最佳匹配点对, 确定摄像机的位姿。将三维虚拟物体叠加到真实场景中, 达到虚实结合的效果。实验表明, 在不同尺度角度、一定环境光变化、复杂背景和基准图像部分遮挡的情况下, 该 AR 系统都具有良好的性能, 跟踪定位准确度高, 速度基本达到实时要求。

**关键词:** 增强现实; 自然特征; 汉明距离; 匹配; 实时注册

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1001-3695(2012)09-3594-03

**doi:**10.3969/j.issn.1001-3695.2012.09.106

## Real-time augmented reality based on ORB

REN Jie, ZHOU Yu, YU Yao, DU Si-dan, WANG Zi-qiang

(Dept. of Electronic & Engineering, Nanjing University, Nanjing 210009, China)

**Abstract:** Now augmented reality based on natural features is not efficient. This paper proposed a new method of real-time AR registration based on ORB features. First it extracted ORB natural features from video frames and a reference image, Selected matching features using appropriate Hamming distance and RANSAC algorithm. Then it calculated and determined the camera posed. Finally, superimpose the computer-generated three-dimensional virtual objects on the real-life scene to achieve combination of virtual and real. Experiments show that, in the case of different scales, different angles, some changes in ambient light, complex backgrounds or part of the reference image blocked, the AR system performs well to track and position accurately. And the processing speed of the system approximately meets the requirement of real-time system.

**Key words:** augmented reality(AR); natural feature; Hamming distance; matching; real-time registration

增强现实一般定义为:将计算机生成的虚拟三维对象实时叠加到真实场景中,构造出虚实结合的空间<sup>[1]</sup>。AR 研究的核心问题是准确的三维跟踪注册,即将虚拟物体准确定位在预先设定的真实场景中。

基于特征提取、匹配、定位的注册技术在 AR 研究中占有很大比例,通过建立二维图像特征坐标和三维真实世界特征坐标之间一一对应关系,估算摄像机的姿态,即虚拟物体进行几何变换的参数。

现有基于特征注册技术主要分为基于特殊标志(marker)的注册和基于自然特征(markerless)的注册。基于 marker 的 AR 软件包主要有 ARToolKit<sup>[2]</sup>和 ARtag<sup>[3]</sup>。Marker 注册方法快速稳定,但是标志物必须完整出现在视频帧中,影响视觉效果,有很大的局限性。基于自然特征的 AR 实现是当前的热门研究方向<sup>[4]</sup>。

自然特征描述子大致可以分为基于绝对值的描述子和基于比较的描述子。前者一般是将灰度、梯度等量化,构造直方图,如 Sift<sup>[5]</sup>、Surf<sup>[6]</sup>等算子。这类描述子判别能力强,但计算复杂度高、效率低,要实现实时的 AR 跟踪注册,必须加入跟踪算法,如 KLT 算法<sup>[7]</sup>。后者一般通过比较预先训练的或者随

机点对的特征值大小,来构造描述子,如 BRIEF<sup>[8]</sup>、ORB<sup>[9]</sup>等算子。这类描述子一般都是为了提高计算速度而设计的。BRIEF 算子速度快,但不具备旋转不变性,对噪声敏感,不具备尺度不变性。ORB 算子基于 BRIEF 算子提出,解决了上述 BRIEF 算子的缺点。在计算速度方面,ORB 是 Sift 的 100 倍左右,是 Surf 的 10 倍左右,可以满足实时的特征匹配。在性能上与 Sift 和 Surf 相当,基本可以取代 Sift 和 Surf。

本文通过 ORB 算子进行特征的提取和匹配,使用 RANSAC<sup>[10]</sup>算法筛选最佳特征点对,建立相机坐标和世界坐标之间的变换关系,最后通过 OpenGL 将虚拟物体定位于现实世界中。由于环境光等噪声的影响,本文还使用了简单的滤波器对系统进行了优化,使得健壮性得到提高。

## 1 系统框架

图 1 为系统 AR 实现的设计流程。

首先选取合适的基准图像,该系统中选取计算机桌面的背景图片(如图 2、3 所示)并提取 ORB 特征,对单摄像头视频流的每一帧提取 ORB 特征后与基准图像的特征进行匹配筛选,将得到的匹配点对代入投影方程中求得摄像机旋转矩阵  $R$  和

收稿日期: 2012-02-20; 修回日期: 2012-03-27

**作者简介:**任结(1988-),男,江苏无锡人,硕士研究生,主要研究方向为图像处理与应用、嵌入式系统开发与应用(jieren2006@163.com);周余(1980-),男,重庆人,讲师,博士,主要研究方向为嵌入式开发、图像处理;于耀(1982-),男,贵州人,讲师,博士,主要研究方向为嵌入式开发、图像处理;都思丹(1963-),女,江苏南京人,教授,博导,主要研究方向为嵌入式开发、图像处理、模式识别;王自强(1964-),男,江苏无锡人,副教授,博士,主要研究方向为嵌入式开发。

平移矩阵  $T$ , 最后将 OpenL 生成的三维虚拟物体叠加到真实场景中某个坐标位置上以达到增强现实的效果。

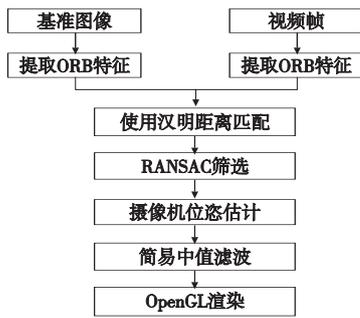


图1 系统框架

## 2 ORB 特征的提取与匹配

### 2.1 ORB 特征描述子

首先介绍 ORB 特征描述子。ORB(oriented FAST and rotated BRIEF)是基于 FAST<sup>[11]</sup> 特征检测和 BRIEF 描述子改良的。

FAST 特征点是不带有方向性的,通过获取特征点邻域的重心确定方向参数<sup>[12]</sup>。文献[12]中定义邻域矩如式(1), $x, y$  是相对 FAST 特征点的位置,圆形邻域半径  $r, x, y \in [-r, r]$ 。

$$m_{pq} = \sum_{x,y} x^p y^q I(x, y) \quad (1)$$

圆形邻域重心计算式:

$$C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (2)$$

特征点与重心的夹角定义为 FAST 特征点的方向:

$$\theta = \arctan\left(\frac{m_{01}}{m_{10}}\right) = \arctan\left(\frac{\sum_{x,y} yI(x, y)}{\sum_{x,y} xI(x, y)}\right) \quad (3)$$

式(3)中分母不允许为 0,在实验观察分析中可以看出此种情况几乎不会发生。

BRIEF 描述子的主要思想是在特征点附近随机选取若干点对,将这些像素点对的灰度值的大小进行比较,组合成一个二进制串,并将这个二进制串作为该特征点的特征描述子。随机点对生成的描述子对噪声比较敏感,为此 ORB 在 FAST 特征点  $31 \times 31$  像素领域内随机选取  $5 \times 5$  的像素块对,使用积分图对若干对像素块进行比较,得到二进制串。这样使用像素块灰度总和和比较的方法有效地降低了随机噪声的干扰。

ORB 根据式(3)中求得的方向参数提取 BRIEF 描述子。但是由于环境的因素和噪声的引入,特征点方向会发生变化,随机像素块对的相关性会比较大,从而降低描述子的判别性。ORB 采取贪心算法寻找相关性较低的随机像素块对,一般选取 256 个相关性最低像素块对,构成 256 bit 特征描述子。

FAST 特征和 BRIEF 算子都不具备尺度不变性。ORB 在提取 FAST 特征点之前对图像构造了高斯金字塔,在金字塔的每层提取 FAST 特征点,从而加入尺度特征。

### 2.2 ORB 特征的匹配

上节得到 ORB 256 bit 二进制描述子,假设  $K_1, K_2$  两个描述子:

$$K_1 = x_0 x_1 \cdots x_{255}, K_2 = y_0 y_1 \cdots y_{255} \quad (4)$$

通过汉明距离之间的异或之和表征两个 ORB 特征描述子的相似程度,用  $D(K_1, K_2)$  表示:

$$D(K_1, K_2) = \sum_{i=0}^{255} x_i \oplus y_i \quad (5)$$

$D(K_1, K_2)$  越小代表相似程度越高,反之相似程度低。

使用 brute force 匹配器得到所有可能的匹配特征对,假设特征对最大汉明距离为 MAX\_DIST。为了选取最佳匹配对和提高运行效率,选取合适的阈值(本文选取 0.4),选出小于  $0.4 \times \text{MAX\_DIST}$  的匹配对进行下一步摄像机位姿估计。阈值不能选取太小,否则匹配对太少,影响最终效果,需要通过对视频帧图像进行实验选取最佳阈值。

图 2 是视频中某 2 帧图像与基准图像之间 ORB 特征的匹配。图 2 是在不同尺度和不同视角情况下的匹配结果,连线表示匹配点对。显示的结果是在第三部分使用 RANSAC 算法筛选后得到的最佳匹配点对。

经过使用汉明距离粗筛选和 RANSAC 算法的迭代精筛选后匹配点对很稳定,在尺度、角度、环境、光照变化的情况下,视频帧都能与基准图像找到最佳匹配点对。同时匹配点对选取不少于 30 对,若选取太少会影响后面摄像机位姿计算的精确性。



图2 ORB特征匹配

## 3 摄像机位姿估计

针孔摄像机模型投影方程为

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_{3 \times 3} & T_{3 \times 1} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (6)$$

其中:三维世界坐标  $(X, Y, Z)^T$ , 经过摄像机投影变换后得到二维视图坐标  $(u, v)^T$ ;  $f_x, f_y, C_x, C_y$  是摄像机内参数,实验之前经过标定都为已知值; $R, T$  矩阵就是世界坐标系到相机坐标系的转换参数。

从式(6)可以看出,只要每帧图像与基准图像有足够多的匹配点对,就可以求解得到摄像机的位姿。在此步骤中,RANSAC 算法剔除匹配不好的特征点对,得到最佳的  $R, T$  变换参数。最后将 OpenGL 生成的虚拟物体经过  $R, T$  反变换后定位到真实场景中。

## 4 实验结果与讨论

本文实验使用 Logitech 网络摄像头,  $640 \times 480$  分辨率,计算机系统为 Windows XP(主频 2 GHz,内存 2 GB),开发平台 Visual Studio 2008。

图 3 为本文 AR 实现结果,分别是在不同视角、不同距离、一定环境光变化下的实时注册结果,并且在复杂背景或者基准图像(计算机屏幕显示图像)被部分遮挡的情况下也可以完成准确的实时注册。实验中将 OpenGL 生成的虚拟茶壶和三维坐标系渲染到真实场景中计算机屏幕上。

表 1 是本文 AR 实现每个步骤的平均时间,将 ORB 与 Sift 进行比较。每帧图像处理的总时间约为 61.7 ms,大约为 16 fps,基本上可以达到实时的要求。Sift 在提取特征描述子时

间约是 ORB 的 70 倍,可以看出仅仅用 Sift 这类描述子实现实时 AR 系统是行不通的。如果系统的配置更高,摄像头分辨率选取  $320 \times 240$ ,相信能够达到 30 fps。

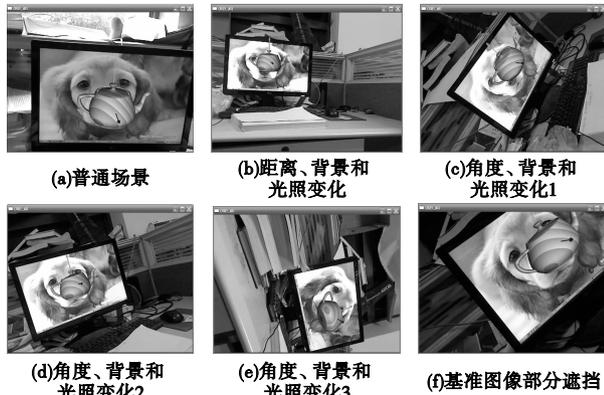


图3 系统运行结果

表1 平均处理时间比较

	ORB/ms	SIFT/ms
提取描述子	9.9	703.1
匹配	8.7	65.3
位姿估计	20.9	18.7
OpenGL 绘图	22.2	22.1
总时间	61.7	809.2

由于外界环境噪声等因素的引入,为了加强系统的抗干扰能力,实验中加入了简单的中值滤波进行系统的优化,增强稳定性。但是如果摄像头快速剧烈运动,系统在跟踪注册方面无法达到良好的效果,这有待下一阶段改进。

## 5 结束语

增强现实是当前图像处理与应用的一个热门话题,本文提出一种新的 AR 实时系统方案——基于特征描述子 ORB 的方案。系统在摄像头不是快速剧烈运动的情况下,跟踪注册可以达到准确稳定的效果。

## 参考文献:

- [1] AZUMA R, BAILLOT Y, BEHRINGER R, *et al.* A survey of augmented reality [J]. *Teleoperators and Virtual Environments*, 1997,6(4):355-385.
- [2] ABAWI D F, BIENWALD J, DORNER R. Accuracy in optical tracking with fiducial markers: an accuracy function for ARToolKit [C] //Proc of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality. Washington DC: IEEE Computer Society, 2004:260-261.
- [3] FIALA M. ARTag, a fiducial marker system using digital techniques [C] //Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2005:590-596.
- [4] HUANG Fen, ZHOU Yu, YU Yao, *et al.* Piano AR: a markerless augmented reality based piano teaching system [C] //Proc of Intelligent Human-Machine Systems and Cybernetics. 2011:47-52.
- [5] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2):91-110.
- [6] BAY H, TUYTELAARS T, Van GOOL L. Speeded-up robust features (SURF) [J]. *Computer Vision and Image Understanding*, 2008, 110(3):346-389.
- [7] TENG Fei, LIANG Xiao-hui, HE Zhi-ying, *et al.* A registration method based on nature feature with KLT tracking algorithm for wearable computers [C] //Proc of International Conference on Cyberworlds. Washington DC: IEEE Computer Society, 2008:416-421.
- [8] CALONDER M, LEPETIT V, STRECHA C, *et al.* Brief: Binary robust independent elementary features [C] //Proc of the 11th European Conference on Computer Vision. Berlin: Springer-Verlag, 2010:778-792.
- [9] RUBLEE E, RABAU D V, KONOLIGE K, *et al.* ORB: an efficient alternative to SIFT or SURF [C] //Proc of IEEE International Conference on Computer Vision. 2011:2564-2571.
- [10] FISCHLE M A, BOLLES R C. 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.
- [11] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection [C] //Proc of the 9th European Conference on Computer Vision. Berlin: Springer-Verlag, 2006:430-443.
- [12] ROSIN P L. Measuring corner properties [J]. *Computer Vision and Image Understanding*, 1999, 73(2):291-307.