

3 维物体 SIFT 特征的提取与应用

熊 英 马惠敏

(清华大学电子工程系图象图形研究所, 北京 100084)

摘 要 SIFT(scale-invariant feature transform)算法自提出以来,就因其优越的性能(尺度不变性、旋转不变性、抗噪声能力强、受光照变化影响小等),而备受图像图形领域研究者的青睐。该算法的核心特征(SIFT特征)基于局部梯度,能够抵抗图像大幅度的伸缩、旋转等,很好地满足了3维物体识别的实际需要。而SIFT特征对投影变换的相对敏感性恰可用于3维模型的视点空间划分,且划分依据与匹配依据一致,能够有效提高匹配准确度。合理设置SIFT算法的阈值还可以有效处理物体背景分割等技术问题。通过充分的预处理,能够有效降低SIFT算法计算复杂度,使得系统基本达到实时匹配。总之,将SIFT特征应用在3维物体识别系统中的视点空间划分、背景物体分割、模式特征匹配等模块,可以有效地提高系统的识别速度与效率,增强系统的稳定性。

关键词 SIFT 3维物体识别 模式匹配 视点空间划分

中图法分类号: TP391.41 文献标志码: A 文章编号: 1006-8961(2010)05-814-06

Extraction and Application of 3D Object SIFT Feature

XIONG Ying, MA Huimin

(Institute of Image and Graphics, Department of Electronic Engineering, Tsinghua University, Beijing 100084)

Abstract The SIFT algorithm is widely adopted by researchers in image and graphic study, with its many advantages such as invariance to scaling, rotation, noise and illumination changes. SIFT feature is based on local gradient, making it invulnerable to large scale of image extension, compression and rotation, and this meets the practical requirements of 3D object recognition. And the sensitivity to homographic transformation of the feature can be applied to partition the view space. Reasonably setting the threshold value, SIFT algorithm can handle the technical problems such as cutting the object from its background. After pre-processing, the high computation complexity can be reduced, making the system run in real time. Therefore, applying SIFT feature in view space partition, cutting the object from its background and pattern matching can effectively enhance the robustness of the system and improve its speed and efficiency.

Keywords SIFT, 3D object recognition, pattern match, view space partition

0 引言

出于在自动导航、自动检测、机器人抓取物件及装配任务等许多实际需要,3维物体识别如今已成为一个非常活跃的研究领域^[1]。3维物体识别的一个难点在于克服因视点位置改变而引起的目标姿态变化^[2]。这种姿态变化一般可以用一个投影变换

来描述。但投影变换的非线性极大地增加了运算的复杂度,因此在实际中往往采用仿射变换来代替投影变换。于是3维物体识别的一个关键问题就是找到物体的仿射不变性特征,并以此进行目标的匹配识别^[3-4]。在小角度近似下,仿射不变性又可以归结为尺度不变性和在平面内的旋转不变性。

SIFT(scale-invariant feature transform)算法由Lowe于1999年提出^[5],并于2004年完善总结^[6]。

基金项目:国家自然科学基金项目(60502013);国家高技术研究发展计划(863)项目(2006AA01Z115)

收稿日期:2009-04-17;改回日期:2009-07-15

第一作者简介:熊英(1987—),男,清华大学电子工程系本科生。主要研究方向为3维物体识别与仿真。E-mail: xiongying0007@gmail.com

该算法通过抽取图像的局部不变性特征(SIFT 特征)进行图像识别匹配。由于具有良好的尺度不变性和旋转不变性、抗噪声能力强等诸多优点,该算法在许多领域得到了广泛的应用,如图像拼接、模式识别等。

目前 SIFT 算法在 3 维物体识别领域的应用并不多见,但实际上,该算法能够有效抵抗图像伸缩、旋转、噪声干扰及光照变化等优点恰恰能够满足 3 维物体识别中的一些实际需要。虽然在大角度下 SIFT 特征的仿射不变性会受到一些限制,但通过合理划分视点空间可以保证算法有较高的识别准确率。另外,本文将证明对 3 维模型做充分的预处理后,系统识别速度也可以大幅提高。

1 3 维物体识别系统结构

3 维物体识别在当今计算机视觉中已成为一个非常活跃的领域,现今已提出较多的识别技术与算法^[7]。一个典型的识别系统大致可分为 2 个部分,即对 3 维物体建立特征数据库与对目标图像特征提取(见图 1)。

在建立 3 维物体特征数据库过程中,需要完成的核心任务有:

- 1) 视点空间划分 选择代表视点及抽取代表

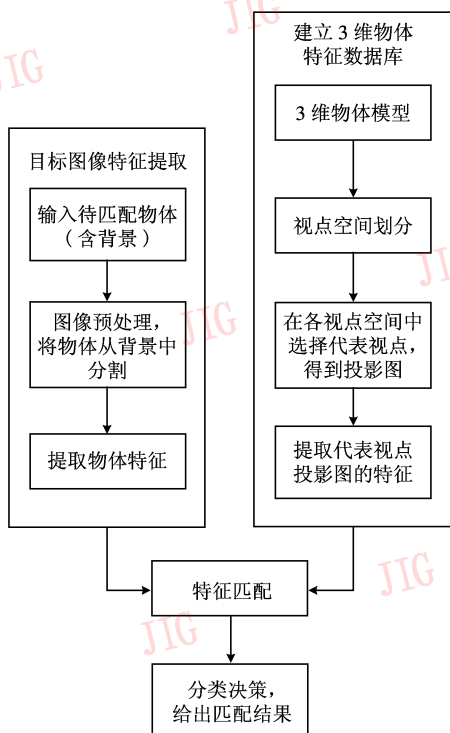


图 1 典型 3 维物体识别系统结构框图

Fig. 1 A typical 3D object recognition system

视点的特征。其中视点空间划分目前较多采用均匀划分后聚类、突变论两种方法,但前者存在划分不够精细准确的问题,而后者需求的海量计算往往无法实现。

- 2) 目标分割 在目标图像特征提取过程中需要将输入图像预处理,如将物体从背景中分割,以便有效消除背景噪声干扰,准确地提取物体特征。

- 3) 特征匹配 最后将输入图像特征与物体模型特征进行比对,进行分类决策并给出识别结果。

2 SIFT 特征在视点空间划分中的应用

SIFT 特征可以应用于一种基于聚类思想的视点空间划分。通过采样得到 3 维物体各个视点下的投影视图,抽取这些视图的 SIFT 特征进行聚类分析。将特征大体相同的视点集聚为一类,用一个代表视点表达该类,本文直接取该类点的公有 SIFT 特征。

首先在高斯球面上均匀且较为密集地采集 3 维物体的投影视图,计算每个视图的 SIFT 特征。接着对所有投影图两两进行匹配分析,记录匹配点个数,以此定义距离并聚类。

两幅投影视图间匹配点越多,它们就越相似,相应的距离也就越小,即距离应与匹配点数成反比。记 num_{ij} 为两幅投影视图的匹配点对数,于是定义距离 $dist\{i,j\} = \frac{1}{num_{ij} + \epsilon}$ 。其中,分母 ϵ 是为了数学上处理方便而引入的,这样距离不会趋于无穷。实验表明,聚类结果受 ϵ 影响不大,不妨取 $\epsilon = 1$ 。

实验采用自下而上的层次聚类方法(agglomerative hierarchical clustering),点到类的距离定义为点到该类最远点的距离,类间距离定义为一类中所有点到另一类距离的最大者:

$$dist\{i,D\} = \max_{j \in D} dist\{i,j\}$$

$$dist\{D,E\} = \max_{i \in D} dist\{i,E\}$$

根据此定义得到的聚类结果可以保证每一类内的任两点都有足够的匹配点数,从而从该类中抽取的代表视点一定能够很好地代表全类特征。

聚类中对距离门限设定的越小,类内点的相似度就越高。但如果门限值设定过小,聚类过程将受噪声影响严重,结果中会出现很多自成一类的孤立点,达不到较好的效果。实验表明,门限值设定在 0.2 到 0.3 左右较为适宜。

图 2 为对某飞机用上述方法聚类的结果,图中

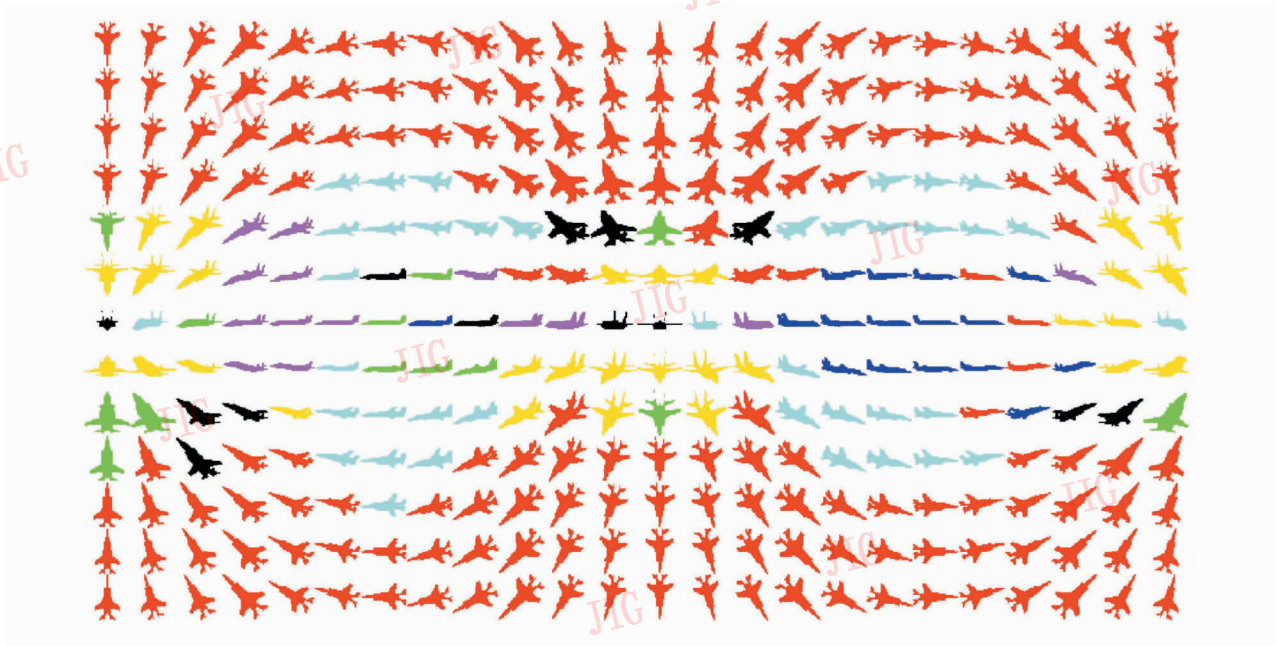


图 2 对某飞机全球面采样聚类结果

Fig. 2 Clustering results of a plane sampled on Gaussian sphere

相邻的同色飞机被聚为一类。可以看到,大量高纬度视点投影图被聚作同一类,这是由于飞机类似一个平板物体,从大角度观测视图自然是相似的;而且,实验中使用的是经纬度等间距采样(采样间隔为 15°),故高纬度采样相对密集,被聚成同类的点也就较为密集(事实上,第 1 行和最后一行的视图时来自同一视点,不同方向)。这些从另一方面印证了结果的正确性。

聚类后可以抽取代表视点。由于本文聚类方法保证每类内两个视图都有很高的相似性,因此从类中任取一个元素都可以作为代表视点。图 2 中飞机被聚成 25 类,每类代表视点如图 3。



图 3 对飞机处理后得到的代表视点投影图

Fig. 3 2D projection figures on representative viewpoints

事实上,一个聚类往往有相同或相似的 SIFT 特征,可以将它们提取出来,这样就不需要重新选取代表视点,而且可以通过统计平均方法增大所选取 SIFT 特征的代表性,提高匹配准确度。

这种视点空间划分方法的另一个好处在于划分标准与后续的匹配标准是一致的。这样可以在确保不会因为聚类而丢失信息的前提下,进行最大限度的数据压缩。处理后数据库占用存储空间非常小,一个 3 维物体经过预处理后将转换成几百个 128 维向量。数据库中只需要存储这些 128 维向量和几个代表视点的 2 维投影图,而其他各个视点的 2 维投影图则不需要进行存储。

上述视点空间划分的方法是十分精细准确的,但也存在一个问题,即需要在全高斯球面密集采集视点和投影视图,而且需要对每幅投影视图计算 SIFT 特征,并对每两个视图进行特征匹配,这个过程需要的计算相对来说是比较繁杂的。但上述过程是在建立特征数据库过程中进行的,因此不会影响系统匹配的实时性。事实上,这相当于通过前期大量的后台准备工作以保证输入图像到来时的快速匹配。

3 SIFT 算法在物体背景分割与特征提取中的作用

在 3 维物体识别系统中需要将输入图像预处

理,以便对其进行特征提取及后续的匹配识别。预处理过程中要解决的一个关键问题是将图像中的物体与背景区分开来,避免背景对识别结果的影响。一般来讲,当不规则物体以任意位置的姿态嵌入在背景中时,将其完全从背景中分离出来是比较困难的,至少需要一定量较复杂的计算,从而增长处理时间,破坏系统的实时性。

而由于 SIFT 特征描述子的特点使其在处理物体背景分割与特征提取过程中有一些特别的优势,可以很好地解决上述问题。这些优势主要表现在以下几方面:

首先,SIFT 特征点取自输入图像差分高斯空间的极值点。而对图像的差分高斯运算相当于一个高通滤波器,能够有效滤除背景中的低频分量。也就是说,在纯色背景以及变化缓慢的渐变背景中一般不会出现 SIFT 特征点,也就不会影响特征匹配。

另一方面,SIFT 算法在提取特征点时有一个门限参数 T (Threshold),可以剔除对比度较低的极值点,因为这些极值点往往源自背景噪声。实验证明,通过适当提高门限参数,可以有效剔除背景中的特征点,而不会影响对物体的特征计算,实验结果如图 4 所示,图中红色箭头为特征点及其主方向。

最后,SIFT 算法进行的是分层 (Octave) 计算,大多数背景中的特征点都会出现在尺度层次较低的视图中(对应文献[6]中参数 σ 较小的视图),而物体中的特征点往往出现在高尺度层次的视图中。通过适当地选择计算层次,也可以有效地滤除背景对特征选取的影响。

总之,SIFT 算法的上述一些特点使其在处理带有背景的输入图像时有相当好的稳定性。本文实验证实了 SIFT 算法能够较好地解决物体在纯色背景、渐变背景以及白噪声背景中的物体分割及特征提取问题。

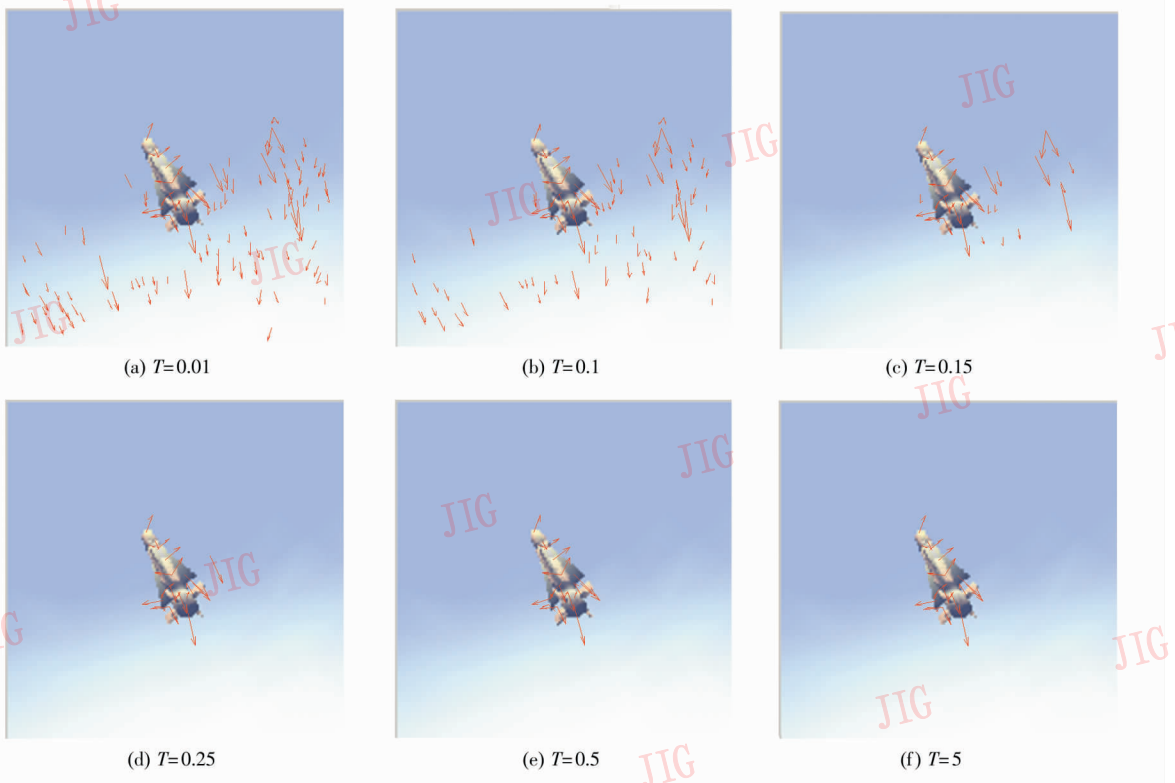


图 4 调节门限参数剔除 SIFT 背景特征点

Fig. 4 Adjusting threshold parameter to remove keypoints in background

4 SIFT 特征匹配

得到待匹配物体的特征后,就可以与数据库中

的特征进行比对,通过分类决策得到匹配结果。为了加快匹配速度,还可以对特征数据库建立索引或哈希。由于 SIFT 特征描述子非常规整(128 维向量),因此建立索引或哈希也相对容易。

SIFT 特征的另一好处在于它是基于局部梯度特征的,因此对尺度、旋转和光照强度都有非常好的不变性。只要局部形态一致,甚至可以进行彩色图像与二值图像之间的匹配(见图 5)。一方面,这

可以简化 3 维物体建模;另一方面,也可以优化输入图像的预处理过程,省去二值化过程。(事实上,图 5 左边图像二值化后将会导致物体和背景融合,无法分离。)

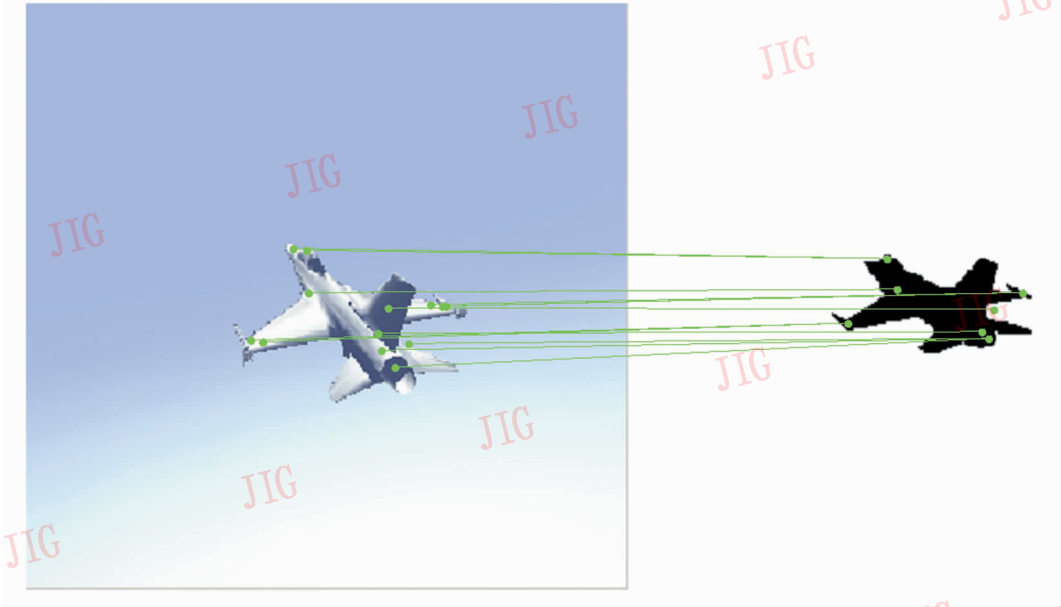


图 5 SIFT 特征在彩色图像和二值图像之间的匹配

Fig. 5 SIFT feature matching between color images and binary ones

但是这种局部特征的匹配方法也存在一定的局限性,即容易出现大量错误匹配点。因此需要对结果进行一些校正。本文中采用 RANSAC 算法根据匹配点位置进行校正,保证所有匹配点满足大致相同的仿射变换方程,误差过大的点为匹配虚警,应被舍去。

本文将系统在国际通用的普林斯顿数据库^[8]中进行了测试。选取库中 6 种飞机模型建立特征数据库,并对输入的 20 张不同姿态的 2 维投影视图进行了识别匹配,各模型识别率见表 1。

表 1 6 种 3 维飞机模型识别率

Tab. 1 Recognition rate of 6 3D plane model

m1200	m1201	m1206	m1210	m1213	m1216
13/20	15/20	11/20	14/20	8/20	12/20

SIFT 算法的一个问题在于其计算复杂度高,处理时间长。但在本文系统中,可以将耗时部分放在预处理过程进行(视点空间划分,特征库建立等),而识别过程速度很快,可以达到实时。各部分运行的时间见表 2(硬件平台 Intel® Core™ 2 Duo CPU, 2.00 GHz;软件平台 MATLAB R2008a)。

表 2 系统各模块运行时间

Tab. 2 Running time of each module in the system

系统模块	运行时间/s
均匀采样得到投影视图	3.25
各视图特征计算	105.47
特征聚类	7.34
特征数据库建立	5.24
输入图像匹配	0.19

5 结 论

本文分析了目前 3 维物体识别技术中的主要目标,即通过仿射不变性特征进行识别匹配,讨论了 SIFT 算法在图像模式匹配中的特点,并论述了将其应用在 3 维物体识别中的可行性与优势。

使用 SIFT 特征作为模式匹配的特征,可以保证识别系统有非常好的尺度不变性与旋转不变性,经过合理的视点空间聚类划分后可以克服仿射不变性限制,达到较高的准确率。在处理将物体从背景中分割的问题中,通过合理设定门限参数滤除背景中的特征点,就可以直接处理输入图像而不受背景噪

声的影响。最后,SIFT 特征规整,占用存储空间小,且易于建立索引和哈希,这些都有助于提高匹配速度。

参考文献 (References)

- [1] Li Qing, Zhou Manli, Liu Jian. A review on 3D objects recognition[J]. Journal of Image and Graphics, 2000, 5A(12): 985-993. [李庆, 周曼丽, 柳健. 三维物体识别研究进展[J]. 中国图象图形学报, 2000, 5A(12): 985-993.]
- [2] Lü Jing, Su Xianyu, Wang Haixia. Three-dimensional object recognition with rotation - invariant by structure light[J]. Journal of Optoelectronics · Laser, 2004, 15(12): 1492-1497. [吕静, 苏显渝, 王海霞. 旋转不变的三维物体识别[J]. 光电子·激光, 2004, 15(12): 1492-1497.]
- [3] Sun Jie, Li Fengting. Research on 3D Object Recognition Based on Affine Invariance[D]. Beijing: Tsinghua University, 2008. [孙洁, 李凤亭. 基于仿射不变性的三维目标识别研究[M]. 北京: 清华大学, 2008.]
- [4] Sun Yigang, Yang Liyong, Sun Chengqi. Study on approach of 3D object recognition across multiple views in affine coordinate system[J]. Journal of Civil Aviation University of China, 2005, 23(6): 53-56. [孙毅刚, 杨立勇, 孙承琦. 仿射坐标系下多视图的三维物体识别方法研究[J]. 中国民航学院学报, 2005, 23(6): 53-56.]
- [5] Lowe D G. Object recognition from local scale-invariant features[C]//Proceedings of International Conference on Computer Vision. Washington, DC, USA: IEEE Computer Society, 1999: 1150-1157.
- [6] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [7] Silvio Savarese, Li F F. 3D generic object categorization, localization and pose estimation [C]//Proceedings of the IEEE International Conference on Computer Vision. New York, NY, USA: IEEE Press, 2007: 1-8.
- [8] Princeton. Princeton is a standard 3D model library[EB/OL]. (2007-10-05)[2009-04-17]. <http://shape.cs.princeton.edu/search.html>.