

· 自动化技术 ·

Harris-SIFT算法及其在双目立体视觉中的应用

赵钦君¹, 赵东标², 韦 虎²

(1. 南京航空航天大学自动化学院 南京 210016; 2. 南京航空航天大学机电学院 南京 210016)

【摘要】为了降低尺度不变特征变换(SIFT)算法的复杂度,提高算法的实时性,克服算法提取的特征点不是角点的缺点,该文提出了一种新的Harris-SIFT算法。该算法首先用Harris算子提取图像的特征点,然后为每个特征点定义主方向,最后将特征描述子的坐标旋转到与特征点的主方向一致,计算出每个特征点的特征向量描述子。双目立体视觉图像匹配实验结果说明了该算法的有效性。

关键词 双目立体视觉; Harris-SIFT; 图像匹配; 尺度空间

中图分类号 TP391

文献标识码 A

doi:10.3969/j.issn.1001-0548.2010.04.015

Harris-SIFT Algorithm and Its Application in Binocular Stereo Vision

ZHAO Qin-jun¹, ZHAO Dong-biao², and WEI Hu²

(1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics Nanjing 210016;

2. College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics Nanjing 210016)

Abstract In order to reduce the scale invariant feature transform (SIFT) algorithm's complexity, improve the real-time performance of the algorithm and ensure that the feature points are corners, a new Harris-SIFT algorithm is proposed. The feature points of images are detected by using Harris operator, the main orientation for each feature point is calculated, and lastly, the feature point descriptors are generated after rotating the coordinates of the descriptors relative to the feature points' main orientations. Experimental results of image matching in binocular stereo vision demonstrate the effectiveness of the proposed algorithm.

Key words binocular stereo vision; Harris-SIFT; image matching; scale space

图像特征匹配是机器视觉和模式识别等领域研究的基本问题以及物体识别、跟踪等应用的重要基础,匹配算法直接影响后续视觉处理的效果。立体匹配是立体视觉系统的重要组成部分,是近年来数字图像处理 and 计算机视觉领域备受关注的前沿方向和研究热点。双目立体视觉在移动机器人、三维测量、物体建模等领域获得了广泛的应用^[1]。该方法模拟人类的视觉系统,设置距离和方位确定的两台相同的摄像机,通过匹配空间中的一点在两台摄像机中的投影点实现立体视觉中同名点的匹配是该方法中最关键的难题。

当空间三维场景被投影为二维图像时,同一景物在不同视点下的图像会有很大不同,而且场景中的诸多因素,如光照条件、对象的几何形状和物理特性、噪声干扰和畸变以及摄像机特性等,都被综合成单一的图像中的灰度值。因此,要准确地对包含了如此之多不利因素的图像进行无歧义的匹配,

显然是十分困难的。对于任何一种立体匹配方法,其有效性依赖于3个问题的解决,即选择正确的匹配特征^[2-4]、寻找特征间的本质属性^[5-6]及建立能正确匹配所选特征的稳定算法^[5,7]。

近年来,各种匹配算法相继出现,并且结合许多数学理论和方法,不断有新的匹配算法被提出。其中,文献[5,8]提出的SIFT是目前最成功的局部特征提取算子,其提取特征过程为:(1)检测尺度空间极值点;(2)精确定位极值点;(3)为每个特征点选择主方向;(4)生成关键点描述子。研究表明^[9-10],SIFT特征点定位准确,具有很好的尺度、旋转、视角和光照不变性,优于其他局部特征提取算子。目前SIFT已成功应用于目标识别、图像视频检索、全景拼接和视觉定位等领域。不过SIFT提取的特征不是人们视觉意义上的角点,而且SIFT计算量较大,实时性较差,难以应用于对实时性要求较高的系统,如基于双目立体视觉的实时跟踪系统。虽然通过

收稿日期:2009-03-27; 修回日期:2009-07-01

基金项目:江苏省自然科学基金(BK2003005)

作者简介:赵钦君(1981-),男,博士生,主要从事机器视觉、图像处理与模式识别方面的研究。

GPU^[11]加速,可以在一定程度上提高SIFT算法的实时性,但不能从根本上解决问题。Harris^[3]算子是一种简单、有效、稳定的角点提取算法,其原理为:如果某一点向任一方向微小偏移都会引起灰度的很大变化,就说明该点是角点。本文结合Harris提取角点的显著性和SIFT描述子,提出一种新的Harris-SIFT算法,并应用于双目立体视觉匹配中,实验结果表明了该方法的有效性。

1 尺度不变特征变换算法及分析

SIFT算法是文献[5,8]在总结现有的基于不变量技术的特征检测方法的基础上,于2004年正式提出的一种基于尺度空间的、对图像缩放、旋转甚至仿射变换保持不变性的图像局部特征描述算法,其特征提取过程可以分为4步。

1.1 检测尺度空间极值点

尺度空间理论最早出现于计算机视觉领域,其目的是模拟图像数据的多尺度特征。在文献[12]中证明了高斯卷积核是实现尺度变换的唯一变换核,而且是唯一的线性核。一幅二维图像,在不同尺度下的尺度空间表示可由图像与高斯核卷积得到:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

式中 (x, y) 是空间坐标; σ 是尺度坐标; $G(x, y, \sigma)$ 是尺度可变高斯函数:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

为了有效地在尺度空间检测到稳定的特征点,提出了高斯差分尺度空间(DOG scale-space)。DOG算子计算简单,是尺度归一化的LOG算子的近似。利用不同尺度的高斯差分核与图像卷积生成:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3)$$

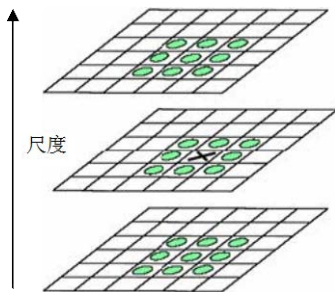


图1 DOG尺度空间极值点检测

$$\begin{cases} m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / ((L(x+1, y) - L(x-1, y)))) \end{cases} \quad (9)$$

然后在高斯空间中特征点的邻域内采样,创建梯度

为了寻找尺度空间的极值点,每一个采样点要和它所有的相邻点比较,看其是否比它的图像域和尺度域的相邻点大或者小。如图1所示,中间的检测点和与它同尺度的8个相邻点以及上下相邻尺度对应的9×2个点共26个点比较,如果该像素在这26个邻域像素中皆为极值,则作为候选的极值点。

1.2 精确定位极值点

通过拟和三维二次函数以精确确定特征点的位置和尺度,同时去除低对比度的特征点和不稳定的边缘响应点(因为DOG算子会产生较强的边缘响应),以增强匹配稳定性、提高抗噪声能力。

利用DOG函数的二阶Taylor展开式 $D(\mathbf{X})$ 插值得到特征点位置和尺度坐标的精确值(亚像素精度):

$$D(\mathbf{X}) = D + \frac{\partial D^T}{\partial \mathbf{X}} \mathbf{X} + \frac{1}{2} \mathbf{X}^T \frac{\partial^2 D}{\partial \mathbf{X}^2} \mathbf{X} \quad (4)$$

式中 向量 $\mathbf{X} = (x, y, \sigma)$, 表示采样点和特征点之间的位置、尺度偏移。令式(4)的一阶导数为0,可得特征点精确位置的偏移向量为:

$$\hat{\mathbf{X}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{X}^2} \frac{\partial D}{\partial \mathbf{X}} \quad (5)$$

将 $\hat{\mathbf{X}}$ 加到原粗特征点的坐标 \mathbf{X} , 即得到特征点的亚像素精确插值估计。将(5)代入式(4)可得:

$$D(\hat{\mathbf{X}}) = D + \frac{1}{2} \frac{\partial D}{\partial \mathbf{X}} \hat{\mathbf{X}} \quad (6)$$

当 $|D(\mathbf{X})|$ 值小于某阈值时,该特征点可以舍去,通常,这种特征点对噪声敏感故而不稳定。

此外还应舍去具有不稳定的边缘响应点,DOG函数的此类极值点通常在边缘切向有较大的主曲率,而在边缘的垂直方向有较小的主曲率。为了检测主曲率是否在某域值 r 下,只需检测是否满足:

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} < \frac{(r+1)^2}{r} \quad (7)$$

式中 \mathbf{H} 为DOG函数的Hessian矩阵:

$$\mathbf{H} = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix} \quad (8)$$

在实验中,取 $r = 10$ ^[5,8]。

1.3 为每个特征点选择主方向

利用特征点邻域像素的梯度方向分布特性,为每个特征点指定方向参数,使算子具备旋转不变性。首先在高斯空间计算特征点的梯度模和方向:

方向直方图。直方图每10度作为一个柱,共36个柱。

然后将邻域内的每个采样点按梯度方向 θ 归入适当的柱,以梯度模 m 作为贡献的权重。最后选择直方图的主峰值作为特征点的主方向,选取量值达到主峰值80%以上的局部峰值作为辅助方向。这样一个特征点可能会被指定具有多个方向,可以增强匹配的鲁棒性。本文定义的主方向还不够精确,从而会影响生成的描述子的性能。

1.4 生成特征点描述子

首先将坐标轴旋转为特征点的方向,以确保旋转不变性。接下来以特征点为中心取 8×8 的窗口。图2a的中央黑点为当前特征点的位置,每个小格代表特征点邻域所在尺度空间的一个像素,箭头方向代表该像素的梯度方向,箭头长度代表梯度模值,图中圆圈区域代表高斯加权的范围(越靠近特征点的像素梯度方向信息贡献越大)。然后在每 4×4 的小块上计算8个方向的梯度方向直方图,绘制每个梯度方向的累加值,即可形成一个种子点,如图2b所示。图中一个特征点由 2×2 共4个种子点组成,每个种子点有8个方向向量信息。这种邻域方向性信息联合的思想增强了算法抗噪声的能力,同时对于含有定位误差的特征匹配也提供了较好的容错性。

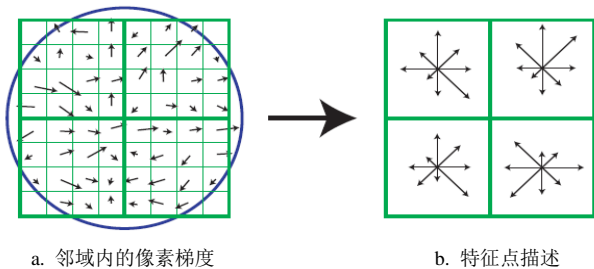


图2 生成SIFT特征点描述子

实际计算过程中,为了增强匹配的稳健性,对每个特征点使用 4×4 共16个种子点来描述^[5,8],对于一个特征点就可以产生128个数据,即最终形成128维的SIFT特征向量。此时SIFT特征向量已经去除了尺度变化、旋转等几何变形因素的影响,再继续将特征向量的长度归一化,则可以进一步去除光照变化的影响。

2 Harris-SIFT算法

当应用于对实时性要求较高的双目立体视觉系统(比如实时跟踪系统)中时,SIFT特征提取和匹配算法主要有3个问题:

(1) 特征提取复杂度太高,计算时间太长。SIFT算法多次使用了卷积平滑操作和加权直方图统计,需要大量浮点数运算,算法复杂度很高,且计算量主要集成在特征描述生成过程,正比于特征点数目。

(2) 生成的特征点太多,影响匹配和搜索速度。SIFT算法可以提取大量特征点,但实际正确匹配的特征点只占很小一部分,对那些没有匹配成功的特征点建立描述子会占用大量时间。

(3) 特征集合整体显著性不高。SIFT特征点不能准确定位角点,大部分特征点不能反映图像的结构。

通过以上分析,为了解决SIFT算法的实时性问题,可以考虑用别的更有效的特征点检测算子取代SIFT中的极值点提取算法。Harris算子是一种有效的点特征提取算子,其值为:

$$R = \det(C) - k \text{tr}^2(C) \quad (10)$$

式中 \det 为矩阵的行列式; tr 为矩阵直迹; C 为相关矩阵,且:

$$C(x) = \begin{bmatrix} I_u^2(x) & I_{uv}(x) \\ I_{uv}(x) & I_v^2(x) \end{bmatrix} \quad (11)$$

式中 $I_u(x)$ 、 $I_v(x)$ 和 $I_{uv}(x)$ 分别为图像点 x 的灰度在 u 和 v 方向的偏导以及二阶混合偏导; k 为经验值,通常取 $0.04 \sim 0.06$ 。当某点的Harris算子 R 大于设定阈值 T 时,该点为角点。



图3 算法流程图

Harris算子的优点有:(1) 计算简单。(2) 提取的特征点均匀合理,能反映图像的结构。(3) 可定量地提取特征角点。(4) 即使存在图像旋转、灰度变化、噪声影响和视点变换,它也是最稳定的一种点特征提取算法。

因此选用Harris提取特征点取代SIFT算法中的极值点, 然后为每个特征点定义主方向, 为每个特征点生成特征向量描述子, 即为Harris-SIFT算法。

当两幅图像特征点的特征向量生成后, 采用特征点特征向量的欧式距离作为两幅图像中特征点的相似性判定度量。取图像1中的某个特征点, 并找出其与图像2中欧氏距离最近的前两个特征点, 在这两个特征点中, 如果最近的距离除以次近的距离少于某个比例阈值, 则接受这一对匹配点。降低比例阈值, 特征点的匹配点数目会减少, 但更加稳定。在本文的实验中, 最近邻特征点距离与次近邻特征点距离之比取0.49。算法总流程如图3所示。

3 实验结果及分析

在Pentium 4, 2.8 GHz CPU, 1.0 Gb RAM的PC机上用C语言编程实现本文的算法, 并对本文提出的Harris-SIFT算法与原SIFT的算法在双目立体视觉应用中的匹配性能进行了实验分析和比较。实验中, 双目相机的分辨率为1 024×776。

图4~图6为3组图像的匹配实验结果。每幅图的上半部分为左相机拍到的图像, 下半部分为右相机拍到的图像。每个图的a图为用原SIFT算法匹配的结果, b图为用Harris-SIFT算法匹配的结果。由图4~图6可以看出, 原SIFT算法提取的特征点大部分不是角点, 不能反映图像的结构, 而本文提出的Harris-SIFT算法提取的特征点皆为角点, 反映了图像的结构。表1~表3分别为图4~图6匹配的统计结果。由表1~表3可以看出, 本文提出的Harris-SIFT算法降低了特征提取和特征匹配的复杂度, 大大提高了原SIFT算法的实时性, 并保证了图像的正确匹配率。

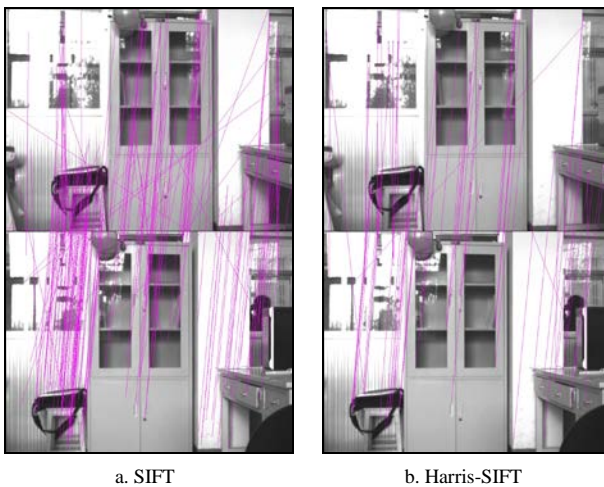


图4 实验一的匹配结果

表1 实验一的匹配结果统计

算法	SIFT	Harris-SIFT
总时间/s	15.8	2.0
匹配点数	199	83
误匹配点数	19	6
匹配率/(%)	90.5	92.8

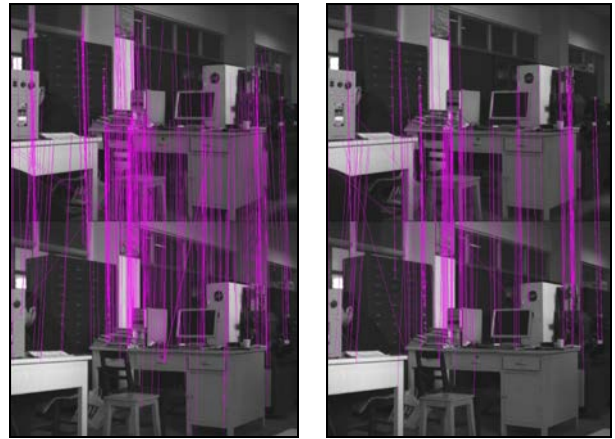


图5 实验二的匹配结果

表2 实验二的匹配结果统计

算法	SIFT	Harris-SIFT
总时间/s	9.9	1.4
匹配点数	271	177
误匹配点数	5	3
匹配率/(%)	98.2	98.3

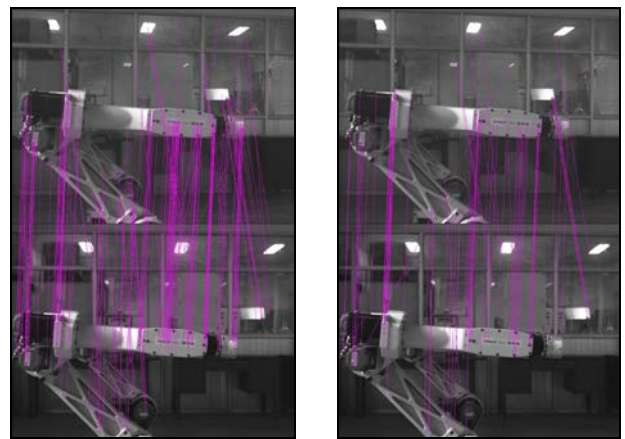


图6 实验三的匹配结果

表3 实验三的匹配结果统计

算法	SIFT	Harris-SIFT
总时间/s	9.9	0.8
匹配点数	169	81
误匹配点数	5	2
匹配率/(%)	97.0	97.5

由以上实验结果可见,在一定程度上, Harris-SIFT算法成功解决了SIFT算法的3个问题:

(1) 由以上算法分析和实验结果可见, SIFT算法检测特征点时要多次用高斯核和DOG算子对图像进行运算, 占用了大量时间。Harris-SIFT算法在生成SIFT特征描述之前用Harris算子检测角点, 而Harris算子计算量很小, 并去除了大量不显著特征点, 同时也降低了特征描述生成阶段的计算量, 即减少了特征描述生成的次数。因此, Harris-SIFT算法大大提高了实时性。

(2) Harris-SIFT算法生成的特征点数量远少于SIFT算法, 从而减少了数据库容量和待匹配特征点数, 因而缩短了匹配时间。

(3) SIFT算法提取的特征点不是图像的角点, 不能反映图像的结构; 而Harris-SIFT算法提取的特征点即为图像角点, 反映了图像的结构, 特征集合性显著, 有利于正确匹配。

4 结 论

SIFT算法可以提取大量不变特征点, 具有很好的尺度、旋转、视角和光照不变性, 优于其他局部特征提取算子, 在图像匹配和目标识别中得到了广泛应用。但当应用于对实时性要求较高的双目立体视觉系统中时, 该算法的复杂度较高, 实时性很差, 且算法的尺度不变特性的优势也难以体现出来。本文提出的Harris-SIFT算子是对SIFT的一个成功改进, 在保证特征正确匹配率的情况下, 降低了特征提取和特征匹配的复杂度, 大大提高了算法的实时性。通过双目立体视觉的匹配实验结果可以看出本文算法的有效性。

参 考 文 献

[1] JIA S, SHENG J, CHUGO D, et al. Obstacle recognition for a mobile robot in indoor environments using RFID and stereo vision[C]//Proceedings of International Conference on the Mechatronics and Automation (ICMA). New York: IEEE, 2007: 2789-2794.
 [2] 干 可, 谢 明, 罗代升. 激光共聚焦序列图像基于特征的分割方法[J]. 电子科技大学学报, 2008, 37(4): 630-633.

GAN Ke, XIE Ming, LUO Dai-sheng. A segmentation based on feature for LSCM sequence images[J]. Journal of University of Electronic Science and Technology of China, 2008, 37(4): 630-633.
 [3] HARRIS C, STEPHENS M. A combined corner and edge detector[C]//Proceedings of the Fourth Alvey Vision Conference. [S.l.]: [s.n.], 1988: 147-151.
 [4] ZHONG Jian-wei, XIE Mei. A fingerprint minutiae matching method based on line segment vector[J]. Journal of Electronic Science and Technology of China, 2007, 5(3): 260-263.
 [5] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
 [6] KE Y, SUKTHANKAR R. PCA-SIFT: a more distinctive representation for local image descriptors[C]//IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2004: 511-517.
 [7] MOREELS P, PERONA P. Evaluation of features detectors and descriptors based on 3D objects[C]//International Conference on Computer Vision. [S.l.]: [s.n.], 2005: 800-807.
 [8] LOWE D G. Object recognition from local scale-invariant features[C]//Proceedings of the International Conference on Computer Vision. Corfu, Greece: [s.n.], 1999: 1150-1157.
 [9] MIKOLAJCZYK K, SCHMID C. A performance evaluation of local descriptors[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.
 [10] MIKOLAJCZYK K, TUYTELAARS T, SCHMID C, et al. A comparison of affine region detectors and descriptors[J]. International Journal of Computer Vision, 2005, 65(1): 43-72.
 [11] HEYMANN S, MULLER K, SMOLIC A, et al. SIFT implementation and optimization for general purpose GPU[C]//15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision. Plzen: Czech Republic, 2007.
 [12] LINDEBERG T. Feature detection with automatic scale selection[J]. International Journal of Computer Vision, 1998, 30(2): 79-116.

编辑 漆 蓉