

SIFT 算法距离比阈值自适应改进研究

万剑华, 孙姝娟, 曾 喆

(中国石油大学地球科学与技术学院, 山东青岛 266580)

摘要:对 SIFT 算法中距离比阈值参数进行分析, 提出自适应距离比阈值改进方法, 并提出用随机取样一致性算法优化点与匹配点的比值作为判断最佳阈值的标准。结果表明, 改进后的方法能够通过较少的迭代确定最佳阈值, 并且不会给计算带来很多负担, 同时能够提高匹配点的准确性。

关键词:测绘; 尺度不变特征变换(SIFT); 距离比阈值; 自适应; 重复率

中图分类号:P 237 **文献标志码:**A

Improvement on adaptive distance ratio threshold of scale invariant feature transform algorithm

WAN Jian-hua, SUN Shu-juan, ZENG Zhe

(School of Geosciences in China University of Petroleum, Qingdao 266580, China)

Abstract: The distance ratio threshold is used for matching in scale invariant feature transform algorithm, which presently is a fixed value in many researches and weakens the algorithm's robustness. The distance ratio threshold was analyzed in detail, and an adaptive method was put forward. In addition, the ratio of random sample consensus optimized points and matching points was used as an approximation of repeatability to identifying the optimum value. The experimental results show that this method can find the optimum threshold value through few iterations, which can improve the matching accuracy while not bring too much calculation.

Key words: mapping; scale invariant feature transform(SIFT); distance ratio threshold; adaptivity; repeatability

SIFT(scale invariant feature transform, 尺度不变特征变换)算法^[1-2]具有尺度不变性, 对于图像缩放、旋转保持不变性, 同时对于仿射变换也有一定的适应能力。Mikolajczyk 和 Schmid^[3-4]通过试验比较了若干具有代表性的描述子(如 SIFT, 矩不变量, 差分不变量等)在图像经过几何变形、光照变化处理、旋转、模糊处理等情况下的性能, 表明 SIFT 描述子在以上各种情况下的性能是最佳的。SIFT 算法性能优越, 应用十分广泛, 已成为图像处理和计算机视觉领域研究的热点之一。但是其实现过程复杂, 需要构建高斯尺度空间并寻找 128 维的特征描述子, 计算时间很长, 并且不完全具有仿射不变性。比较典型的改进算法包括 PCA-SIFT(principle component analysis-SIFT)算法^[5], SURF(speeded up robust features)算法^[6],

GLOH(gradient location-orientation histogram)算法^[4]和 ASIFT(affine-SIFT)算法^[7]等。为提高算法的鲁棒性, 笔者研究一种自适应的距离比阈值确定方法。

1 SIFT 算法简介

SIFT 算法的实现过程可以分为 5 个步骤:

(1) 尺度空间极值点探测。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^[8]。Koenderink(1984)和 Lindeberg(1994)已证实在一系列假设下, 高斯函数是唯一可能的空间尺度核函数。因此, 一幅图像的尺度空间可以定义为可变高斯函数与输入影像的卷积:

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

其中

收稿日期:2012-12-22

基金项目:教育部自主创新科研计划项目(10CX05007A, 12CX04002A);国家自然科学基金项目(41101355)

作者简介:万剑华(1966-),男,教授,博士生导师,主要从事 3S 技术的应用研究。E-mail:wjh66310@163.com。

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

式中, $I(x, y)$ 为输入的图像; * 为卷积; σ 为可变核, 当 σ 连续变化时, $L(x, y, \sigma)$ 构成图像的尺度空间。

高斯差分是归一化拉普拉斯的近似, 而且计算简单, 因此 Lowe 通过对相邻层的卷积结果进行差值构造了差分金字塔:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma). \quad (2)$$

式中, $D(x, y, \sigma)$ 为差分金字塔; k 为常数因子。

在 DOG 中检测极值点时, 将待检测点与和它同尺度的 8 个邻域点以及上下相邻尺度的 18 个邻域点进行比较, 如果是极值点就作为候选点。

(2) 关键点定位。确定候选点之后, 利用尺度空间函数 $D(x, y, \sigma)$ 的泰勒二次展开式进行最小二乘拟合, 通过计算拟合曲面的极值来进一步确定关键点的精确位置和尺度。同时要去掉低对比度点和边缘点, 以增强匹配稳定性。

(3) 特征方向分配。像素点 (x, y) 处梯度的模 $m(x, y)$ 和方向 $\theta(x, y)$ 求取如下:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \quad (3)$$

$$\theta(x, y) = \tan^{-1}(L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)). \quad (4)$$

为确定关键点的主方向, 用直方图统计以关键点为中心的邻域窗口内像素的梯度方向。直方图的峰值代表了关键点处邻域梯度的主方向, 即关键点处的主方向。在梯度直方图中, 当存在另外一个相当于主峰值 80% 能量的峰值时, 则将这个方向作为关键点处的辅方向, 以增强匹配的鲁棒性。

(4) 生成特征描述子。首先将坐标轴旋转到特征点的主方向, 以确保旋转不变性。以关键点为中心, 取 16×16 的窗口, 在每个 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。这样对于一个种子点会产生 $4 \times 4 \times 8$ 共 128 维的 SIFT 特征向量, 特征描述子构建完毕。

(5) 特征描述子的匹配。采用特征向量间的欧氏距离作为其相似性度量, 计算两个特征点之间的最邻近欧氏距离与次邻近欧氏距离。如果这两个距离的比值小于给定的距离比阈值, 则认为这是一对匹配点。当距离比阈值参数减小时, 错误匹配会减少, 但是同时要舍弃更多的正确匹配; 反之, 当距离

比阈值增大时, 错误匹配会迅速增加, 影响结果的准确性。

2 改进算法

在 SIFT 算法的 5 个步骤中, 最后一步用到距离比阈值参数, Lowe 根据试验确定的经验值为 0.8, 此时会减少 90% 的错误匹配, 但是同时舍弃了 5% 的正确匹配。这种通过大量试验确定阈值的方法具有一定的合理性, 但是并不能适应所有的影像, 即固定阈值不具备很强的鲁棒性。

大量试验以及前人的研究成果表明, 距离比阈值参数的最佳取值一般比 0.8 小。因此本文中自适应方法为: 从 Lowe 的经验阈值参数开始逐步减小阈值, 通过一定的判断标准来评定每个阈值的优劣, 从而确定最佳距离比阈值。

Schmid 提出用重复率作为判断兴趣点探测器的判断标准^[9]。给定一个点 X 和两个投影矩阵 P_1 和 P_i , 点 X 在影像 I_1 和 I_i 上的投影分别为 $x_1 = P_1 X$ 和 $x_i = P_i X$ 。如果点 x_1 在影像 I_1 上被探测到, 且其在影像 I_i 上的对应点 x_i 也能被探测到, 则称 x_1 和 x_i 是一对匹配点。重复率定义为两幅影像中的匹配点与探测到的特征点数量总和的比值。

为了检测重复率, 在 x_1 和 x_i 之间必须建立一个唯一的映射关系。对于二维平面, 这种关系是由一个单应性矩阵定义的:

$$x_i = H_i x_1. \quad (5)$$

其中 $H_i = P_i P_1^{-1}$ 。

然而, 重复率检测必须考虑到探测的不确定性。匹配点通常不会在 x_i 的准确位置被探测到, 而是在 x_1 的邻域。邻域大小是通过 ε 来指定, 这个邻域内的重复率被称为 ε -重复率。在 ε 邻域的点对 (x_1, x_i) 是通过下式定义的:

$$R_i(\varepsilon) = \{ (x_1, x_i) \mid \text{dist}(H_i x_1, x_i) < \varepsilon \}. \quad (6)$$

式中, $R_i(\varepsilon)$ 表示重复点对; $\text{dist}()$ 表示点对 (x_1, x_i) 之间的距离。

两幅影像中探测到的点数量可能不同。被探测到的匹配点数量最多不会超过产生较少特征点的影像上的特征点数。因此, 影像 I_i 的重复率定义为

$$r_i(\varepsilon) = \frac{|R_i(\varepsilon)|}{\min(n_1, n_i)}. \quad (7)$$

式中, n_1 和 n_i 分别指两幅图像中检测到的特征点的数量。

重复率可以从一定程度上衡量兴趣算子提取的兴趣点在几何畸变、辐射畸变以及噪声影响下的独

特性、不变性和稳定性,然而这种计算重复率的方法涉及两幅影像之间的单应性矩阵及邻域大小参数 ϵ ,过程比较复杂。因此,需要寻找一种比较简便的方法对参数进行寻优,在不增加太多计算负担的同时保证匹配效果良好。通过大量试验发现,引入 RANSAC(random sample consensus,随机抽样一致)算法对匹配点进行优化,可以在很大程度上去除错误匹配点,提高结果的准确性。因此可以近似认为优化后的点为匹配点,所以重复率的近似计算公式为

$$r = \frac{n_o}{\min(n_1, n_2)} \quad (8)$$

其中 n_o 表示 RANSAC 优化点的数量。

用公式(8)来计算重复率,大大简化了计算过程。用这个作为判断标准来比较距离比阈值参数,挑选出使重复率达到最优时的参数作为最佳距离比阈值。

结合前面的分析,距离比阈值参数优化过程如图 1 所示。参数是否最优通过重复率来进行衡量,重复率越大,表示匹配程度越高,说明距离比阈值的选取越合理;重复率越小,表示匹配程度越低,说明距离比阈值参数不够合理,需要继续调整。

为防止因为寻找最佳距离比阈值过程带来过多的计算负担,人为设定当循环次数大于一定值时也跳出循环。

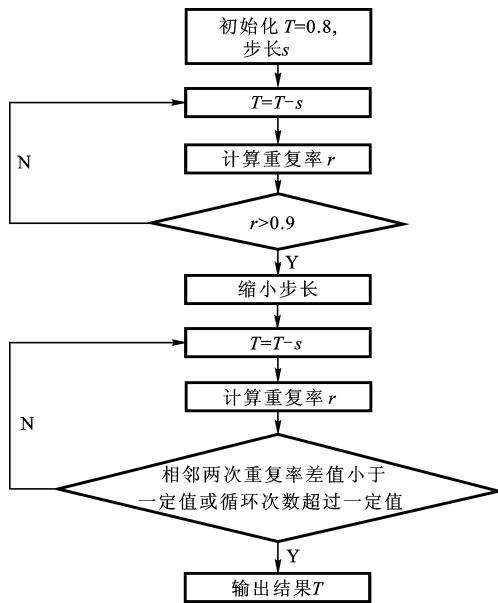


图 1 距离比阈值参数优化过程

Fig. 1 Optimizing process of adaptive distance ratio threshold

3 试验与分析

3.1 近似重复率标准分析

RANSAC 算法通过迭代方式估计数学模型,能从包含大量局外点的数据集中估计出高精度的参数。选取河北某地区用 Canon EOS 5D Mark II 相机获取的多张无人机影像作为试验数据,原始无人机影像数据大小为 5736×3776 ,分辨率为 $0.0064 \text{ mm} \times 0.0064 \text{ mm}$ 。由于 SIFT 算子具有多量性特征,而无人机影像分辨率又比较高,所以一张影像能提取出数以万计的特征点,使得计算量很大。为减少计算负担,首先对影像建立金字塔。建立金字塔后第三层金字塔影像大小为 717×472 ,仍然比较清晰,可以产生大量特征点,完全可以满足试验需要,所以这里选取第二层金字塔影像来进行试验。

先用 SIFT 算法提取特征点并用一定的距离比阈值进行匹配,然后用 RANSAC 算法对匹配结果进行优化;选取多组距离比阈值参数,同样用 RANSAC 算法对匹配结果进行优化。对距离比阈值与匹配点及 RANSAC 优化点之间的关系进行分析,结果如图 2 所示。

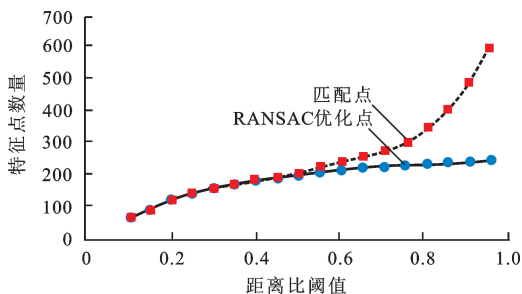


图 2 距离比阈值与匹配点、RANSAC 优化点的关系
Fig. 2 Relationship between distance ratio threshold, matching points and RANSAC optimized points

从图 2 可以看出,当距离比阈值比较小时,匹配点曲线与 RANSAC 优化点曲线基本重合;当距离比阈值增大时,匹配点数量与 RANSAC 优化点的数量都会增加,但是当阈值增大到一定程度之后,只有匹配点数量继续迅速增加,RANSAC 优化点的数量则逐渐趋于稳定。RANSAC 优化点曲线最终趋于平缓说明了给定足够多的匹配点,经这个算法优化后可以将几乎全部正确匹配点找出来,而且数量是稳定的,不会随匹配点数量增加而变化;而匹配点数量一直增加也反映出到一定程度之后增加的基本是错误点。因此,可以近似认为优化后的点是正确的匹配点。该试验验证了可以使用 RANSAC 优化的点与匹配点的比值来作为重复率的近似值,简化了运算。

3.2 距离比阈值参数自适应方法寻优

随机选取河北某地区的两张无人机影像作为试验数据。同样选取第三层金字塔影像来进行试验,大小为717×472。

指定一张影像为参考影像,另一张为待配准影像。经过 SIFT 特征检测,参考影像共检测到 782 个特征点,待配准影像检测到 868 个特征点。利用本文提出的方法对距离比阈值参数进行优化,迭代过程如图 3 所示。图中横轴表示距离比阈值,纵轴表示 RANSAC 优化点的数量 n_o 与匹配点数量 n_m 的比值。可见算法仅仅进行了 10 次迭代就找到了最佳距离比参数,最终确定的参数值为 0.44,在这个距离比阈值下共找到 192 个匹配点。

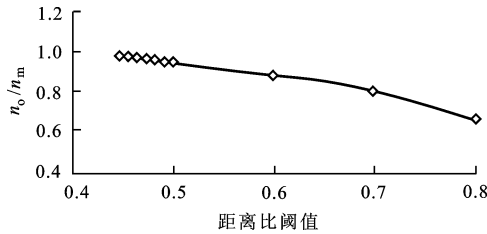
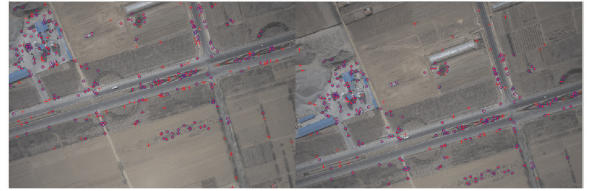


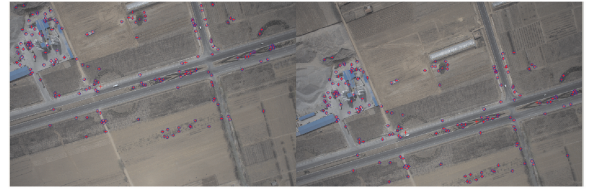
图 3 距离比阈值优化迭代过程

Fig. 3 Iteration of distance ratio threshold

匹配结果如图 4 所示。其中,用圆圈和十字丝表示的为正确的特征点位置,只用十字丝表示的为错误的特征点位置。通过对比可以发现用经验固定阈值匹配后仍存在大量的错误匹配(共匹配 349 对特征点,正确匹配点 236 对),而用自适应距离比阈值方法匹配后错误匹配点数量明显减少(共匹配 192 对特征点,正确匹配 187 对)。



(a) 经验距离比阈值参数匹配结果



(b) 自适应距离比阈值参数匹配结果

图 4 无人机影像匹配结果

Fig. 4 Matching results of UAV images

为进一步验证自适应距离比阈值方法的鲁棒性,再选取 3 组不同场景的图像数据进行试验,对传统 SIFT 算法和改进之后的算法效果进行比较,结果如表 1 所示。其中第一组数据为河北某区 6 张连续的无人机影像组成的 5 个像对,第二组数据为单反相机拍摄的 3 张远距离风景图像组成的 2 个像对,第三组数据为手机拍摄的 3 张近距离静物图像组成的 2 个像对。拍摄的照片都有一定的重叠度以保证可以进行匹配。从表 1 可以看出,改进后的 SIFT 算法根据每个像对都找出了不同的距离比阈值参数,相对于经验阈值,找到的匹配点数量减少很多,但是正确率都在 90% 以上;而传统的 SIFT 方法采用固定经验阈值,找到的匹配点数量很多,但是正确率平均不到 70%。

表 1 算法改进前后 3 组影像试验结果对比

Table 1 Comparison of three groups of images by traditional and improved SIFT algorithm

| 影像 对序号 | 改进 SIFT 算法 | | | | 传统 SIFT 算法 | | | |
|-----------|------------|---------|---------|---------|------------|---------|---------|---------|
| | 距离比阈值 | 匹配点数量/个 | 正确点数量/个 | 正确率/% | 距离比阈值 | 匹配点数量/个 | 正确点数量/个 | 正确率/% |
| 1 | 0.57 | 111 | 108 | 97.2973 | 0.8 | 189 | 119 | 62.9630 |
| 2 | 0.46 | 132 | 124 | 93.9394 | 0.8 | 254 | 163 | 64.1732 |
| 3 | 0.44 | 192 | 187 | 97.3958 | 0.8 | 349 | 236 | 67.6218 |
| 4 | 0.52 | 213 | 201 | 94.3662 | 0.8 | 341 | 238 | 69.7947 |
| 5 | 0.58 | 188 | 172 | 91.4894 | 0.8 | 298 | 193 | 64.7651 |
| 6 | 0.49 | 159 | 150 | 94.3396 | 0.8 | 225 | 158 | 70.2222 |
| 7 | 0.39 | 87 | 84 | 96.5517 | 0.8 | 183 | 90 | 49.1803 |
| 8 | 0.64 | 97 | 91 | 93.8144 | 0.8 | 135 | 93 | 68.8889 |
| 9 | 0.38 | 79 | 74 | 93.6709 | 0.8 | 125 | 82 | 65.6000 |

第一组数据的试验结果与图 4 类似,第二、三组试验的一个像对的匹配结果分别如图 5、图 6 所示。同前,仍然用圆圈和十字丝表示正确的特征点位置,用十字丝表示错误的特征点位置。从图 5 和图 6 可以看出,无论对于哪种场景的影像,改进后的 SIFT

算法相对于传统的方法都减少了很多错配点,这说明改进的方法能够适应不同的影像场景,提高了匹配结果的正确性。

由于改进算法要对距离比阈值进行寻优,相对于传统的 SIFT 算法多了一个寻优的过程,因此用时

比算法改进之前多。表 2 中统计了 3 组数据寻优过程所用时间以及算法改进前后从寻找 SIFT 特征点到影像匹配所用总时间。从表 2 可以看出第一组数据距离比阈值寻优过程用时仅为 2 s 左右,第二、三组数据距离比阈值寻优过程用时为 1 s 左右,未增加太多运算负担。这是因为寻优方法比较简单,通常只需几次迭代就可找到最优值。寻优过程用时在整体用时中所占比重较小,算法改进前后总用时差距并不大,但大幅度提高了匹配点的准确性,因此本文中的方法总体上是有效的。



(a) 经验距离比阈值参数匹配结果



(b) 自适应距离比阈值参数匹配结果

图 5 第 6 影像对匹配结果

Fig. 5 Matching results of the sixth image pair



(a) 经验距离比阈值参数匹配结果



(b) 自适应距离比阈值参数匹配结果

图 6 第 9 影像对匹配结果

Fig. 6 Matching results of the ninth image pair

综上所述,本文提出的自适应改进方法能够根据不同的影像特点自动确定最佳的距离比阈值,而

且简便易行,参数优化过程不会产生太多的运算负担,能够迅速确定最佳阈值,提高匹配点的准确性。

表 2 算法改进前后计算用时对比

Table 2 Comparison of time cost by traditional and improved SIFT algorithm

| 影像对 序号 | 改进 SIFT 算法 | | | 传统 SIFT 算法 | |
|-----------|------------|--------------|-----------|------------|-----------|
| | 距离 比阈值 | 阈值寻优 用时/s | 总用 时/s | 距离 比阈值 | 总用 时/s |
| 1 | 0.57 | 1.734 | 13.406 | 0.8 | 12.624 |
| 2 | 0.46 | 2.015 | 12.905 | 0.8 | 10.905 |
| 3 | 0.44 | 2.734 | 14.077 | 0.8 | 11.376 |
| 4 | 0.52 | 2.890 | 14.64 | 0.8 | 13.269 |
| 5 | 0.58 | 2.015 | 13.765 | 0.8 | 11.750 |
| 6 | 0.49 | 1.375 | 11.108 | 0.8 | 11.547 |
| 7 | 0.39 | 1.453 | 11.407 | 0.8 | 10.109 |
| 8 | 0.64 | 1.235 | 4.204 | 0.8 | 3.032 |
| 9 | 0.38 | 0.844 | 3.673 | 0.8 | 2.891 |

4 结 论

(1) SIFT 算子在特征匹配阶段的距离比阈值参数采用固定阈值往往不能适应不同的影像。在大量试验基础上提出了自适应改进方法,使其能够根据不同的影像特点自动确定最佳的距离比阈值,提高了 SIFT 算法的鲁棒性。通过对不同场景影像的试验很好地证明了改进方法的鲁棒性。

(2) 用 RANSAC 优化的点与匹配点数量的比值作为重复率是可靠的。该方法简便易行,数量参数优化过程不会产生太多的运算负担,能够迅速确定最佳阈值,提高匹配点的准确性。

参考文献:

- [1] LOWE D G. Object recognition from local scale-invariant features [C/OL]//Proceedings of the Seventh IEEE International Conference on Computer Vision (Volume 1), Kerkyra, Greece, Sept 20-27, 1999:1150-1157. [2012-11-02] http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=790410&tag=1.
- [2] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004,60(2):91-110.
- [3] MIKOLAJCZYK K, SCHIMID C. A performance evaluation of local descriptors [C/OL]//Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Volume 2), June 18-20, 2003: 257-263. [2012-11-05] http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1211478.