



骞 森

基于改进的 SIFT 特征的图像双向匹配算法

骞 森, 朱剑英

(南京航空航天大学 机械电子工程研究所, 南京 210016)

摘 要: 图像匹配是计算机视觉中许多领域的基础, 特征提取则是图像匹配的基础, 其中不变量特征是一个重要的理论。SIFT 是最有效的尺度、旋转、亮度不变量局部特征之一, 但算法复杂、计算时间长。分析了 SIFT 的计算时间分配, 通过计算关键点的邻域梯度直方图时动态修改采样步长, 大大提高了 SIFT 的计算速度。分析了基于 SIFT 特征的图像匹配算法, 提出了双向匹配算法, 提高了图像匹配的准确率。实验结果表明所提出的方法是有效的。

关 键 词: 图像匹配; 不变量特征; SIFT

中图分类号: TP391.41

文献标识码: A

文章编号: 1003-8728(2007)09-1179-04

Improved SIFT-based Bidirectional Image Matching Algorithm

Qian Sen, Zhu Jianying

(Mechanical & Electronic Engineering Institute, Nanjing University of Aeronautics & Astronautics, Nanjing 210016)

Abstract: Scale invariant feature transform(SIFT) is one of the most effective local feature of scale, rotation and illumination invariant, but its algorithm is complicated and computation time is long. The computation time distribution is analyzed, and the computation speed is greatly improved by dynamically modifying sampling step when computing the gradient histogram of the region around the key point location. SIFT-based image matching algorithm is analyzed, and a bidirectional matching algorithm is proposed to improve the accuracy of image matching. The experimental results show that the proposed algorithms are effective.

Key words: image matching; invariant features; SIFT

图像匹配是计算机视觉中许多研究领域的基础, 包括目标或场景识别、从多幅图像中求解 3D 结构、立体对应和运动跟踪等。特征提取则是匹配的基础, 基于不变量特征的图像匹配是当前的一个研究热点。对不同视点和焦距所产生的图像, 特征不变量对其尺度和旋转具有不变性, 在图像配准、全景图拼接、目标识别跟踪和机器人视觉导航等方面有重要的应用。

用于图像匹配的不变量特征, 可分为全局特征和局部特征。全局特征如互信息、Hausdorff 距离、Hu 不变矩等, 这些方法虽然经典, 但他们要在图像的全部空间内进行搜索, 计算量太大不能反映局部特征。局部特征的提取分为两步: 提取兴趣点和计算特征描述符。传统的兴趣点主要指角点, 发生在灰度水平方向和垂直方向急剧变化的地方。常用的角点有 Harris 角点^[1]

和 K-L 角点^[2]等, 它们在图像匹配和目标跟踪方面有成功的应用, 但不适用尺度和旋转变化的场合。David Lowe 在 2004 年改进了他 1999 年提出的 SIFT(scale invariant feature transform)算法^[3,4], 使其对尺度、旋转、亮度、仿射、噪音等都具有不变性, 在混乱和遮挡情况下也可鲁棒地识别目标。Mikolajczyk 在 2004 年提出了尺度和仿射不变量的兴趣点检测^[5], 但他比较了各种局部特征检测和描述符算法, 指出 Lowe 的 SIFT 算法取得了最好效果^[6]。

SIFT 算法复杂, 时间复杂度高。YanKe 在 2004 年提出了对 SIFT 的特征描述符用 PCA 进行降维的算法, 并取得更好的效果^[7]。但其算法是在 SIFT 计算结果的基础上, 对于可以预先计算的离线情况是有用的, 比如建立图像检索的 SIFT 特征数据库, 但对于实时情

况,反而增加了计算量。Luke Ledwich 用精简的 SIFT 算法进行室内图像的检索和定位^[8],但只适用于特定场景,且对计算复杂度的优化很小,没有工程意义。Elisabetta Delponete 尝试用 SVD(singular value decomposition)匹配算法对 SIFT 计算出的特征描述符进行图像匹配^[9],但其匹配计算复杂,结果也不比 Lowe 的直接匹配方法优越。

本文基于 OpenCV 用 C++ 语言实现了基于 SIFT 的图像匹配,对 SIFT 的计算过程进行了分析和优化,在计算关键点邻域的梯度直方图时,自适应修改采样步长,大大降低了 SIFT 算法的计算量,减少了运算时间。提出双向匹配算法,提高了基于 SIFT 的图像匹配的准确率。

1 SIFT 速度优化与双向匹配算法

1.1 SIFT 特征

SIFT 图像特征的许多属性适合于对不同图像或场景中同一目标进行匹配。这些特征对于图像尺度、旋转、亮度和 3D 视点都具有不变性,而且有很高的独特性,能使单独一个特征从很大的特征数据库中被高概率正确地匹配出来,减小了由遮挡、混乱或噪声所造成的错误概率。生成 SIFT 图像特征集的主要计算步骤如下^[3,4]:

(1) 尺度空间极值检测。把原始图像放大 2 倍,然后以尺度 2 向下采样,直到某一尺寸,比如 32×32 ,建立图像金字塔,然后对每层的图像以不同标准差的高斯核生成滤波图像组,称为一个 octave,在每一层的每个 octave 中寻找高斯差分图像(different of gaussian, DOG)的局部极值点,作为候选的尺度和方向不变兴趣点。

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

$$L(x, y, \sigma) = G(x, y, \sigma) \times I(x, y) \quad (2)$$

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

式中: $G(x, y, \sigma)$ 为二维高斯函数; σ 为高斯正态分布的标准差; $I(x, y)$ 为金字塔的每层图像; $L(x, y, \sigma)$ 为图像与二维高斯函数的卷积,称为图像的尺度空间; $D(x, y, \sigma)$ 为 DOG 算子,定义为每个 octave 中相邻高斯滤波图像的差分,是归一化 LOG (Laplacian-of-Gaussian) 算子的简单近似。DOG 中每个像素在其 3×3 的 8 邻域和相邻的上层和下层金字塔图像中两个对应的 9 邻域中比较,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。

(2) 关键点定位。在每个极值点,用详细的模型来拟合以去除不稳定的点,包括去除偏移向量的任一分量大于设定阈值的点、去除对比度的绝对值低于设

定阈值的极值点和主曲率不满足一定条件的边缘响应点。

(3) 指定方向。基于局部图像梯度方向,给每个关键点指定一个或多个方向。关键点 (x, y) 的梯度幅值和方向的计算公式为

$$m(x, y) = [(L(x, y+1) - L(x, y-1))^2 + (L(x+1, y) - L(x-1, y))^2]^{1/2} \quad (4)$$

$$\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (5)$$

在关键点的邻域内采样,用直方图统计邻域像素的梯度方向,称为梯度方向直方图,范围为 $0 \sim 360^\circ$,使用 36 个柱(bin),每 10° 一个柱。每个采样点的梯度幅值以高斯函数为权重加入直方图,高斯函数以邻域为窗口, σ 取关键点 (x, y) 所在尺度的 σ 值的 1.5 倍。

梯度方向直方图的峰值则代表了关键点处邻域梯度的主方向,即作为该关键点的方向。在直方图中若有其它小于主峰值 80% 能量的峰值,则为该关键点的辅方向。每个关键点有 3 个信息:位置、所在尺度和方向;其中方向可能是多个:1 个主方向和 1 个以上辅方向。

(4) 关键点描述符:在选择的尺度上,在每个关键点的邻域计算局部图像梯度的方向直方图,最后转换为 128 维的特征描述符。从关键点的方向开始,选取关键点 (x, y) 周围 16×16 的一个矩形块,并分成 16 个 4×4 的子块。按步骤(3)的方法计算每个 4×4 子块的每个点的方向和幅值,把得到的 16 个方向在 $0, 45, 90, 135, 180, 225, 270, 315$ 这 8 个方向上计算直方图,仍以幅度和高斯函数加权,得到 8 个方向描述符。 16 个 4×4 的子块可以得到 128 个方向描述符,将这个 128 维的特征描述符归一化,使其具有亮度不变性。

1.2 SIFT 图像匹配

SIFT 算法以其卓越的性能在全景图拼接等研究领域已获得了成功的应用。基于 SIFT 特征进行图像匹配时,先计算两幅图像的 SIFT 关键点及其特征描述符,然后计算第一幅图像的每个关键点在待匹配图像的关键点集中的最近邻匹配;这里的最近邻定义为关键点的描述符向量间的最小欧氏距离。为了排除因为图像遮挡和背景混乱而产生的无匹配关系的关键点,Low 提出了比较最近邻距离与次近邻距离的方法,距离比率 ratio 小于某个阈值的认为是正确匹配。因为对于错误匹配,由于特征空间的高维性,相似的距离可能有大量其他的错误匹配,从而它的 ratio 值比较高。Low 推荐 ratio 的阈值为 0.8。

1.3 SIFT 速度优化

SIFT 算法的计算复杂度很高,不确定的参数很多,它们决定着检测出的特征点的质量和数量,需要

通过实验确定最佳参数的大致范围,但这些参数对算法的复杂度影响不大。SIFT 要用于实时,必须对计算过程作改进。作者统计了算法各个阶段的时间开销,大致如下:(1) 建立图像的 DOG 金字塔:30% ~ 55%;(2)求局部极值确定关键点:5% ~ 15%;(3) 指定方向:5% ~ 15%;(4) 计算特征描述符:30% ~ 55%。对于 256 × 256 的 lena 图像,在 2.93G、512M、Windows XP SP2 的电脑上 4 个阶段分别需要 359 ms、141 ms、93、516 ms,总时间为 1125 ms。步骤(1) ~ 步骤(3)中的算法优化空间不大,因为微小的改动都会影响结果,步骤(4)中计算特征描述符,大概占总时间的一半左右,可以进行一些优化。

计算特征描述符时,需要计算关键点的邻域梯度直方图时,邻域窗口大小与关键点所在的尺度有关,大部分在 20 到 38 之间。尝试改为自适应修改采样步长:设窗口半径为 radius,步长为 step,则

step = 1 radius < 25
step = 2 25 ≤ radius ≤ 35
step = 3 radius > 35

实验表明,这样使特征描述符的计算时间减少了 45% ~ 65%,匹配结果基本不受影响,有时甚至比原来的结果更好,经分析是不影响误差的情况下,各个尺度互相作用的结果。lena 图像此部分计算时间由原来的 516 ms 减少为 280 ms,使整个 SIFT 的计算时间减少 20% 左右。

1.4 双向匹配算法

SIFT 的关键点描述符有 128 维特征向量,最有效的方法是采用穷尽搜索算法进行图像匹配。对于两个特征集,求第一个特征集中的每个关键点在第二个特征集中的匹配点,即求第一个特征集中的每个关键点在第二个特征集中的最近邻与次近邻的距离比率 ratio, ratio 小于某个阈值的对应点认为是正确匹配,否则抛弃。因此 ratio 的选择非常重要, Lowe 文章推荐 ratio 为 0.8,但作者对大量任意存在尺度、旋转和亮度变化的两幅图片进行匹配,结果表明 ratio 取值在 0.4 ~ 0.6 之间最佳,小于 0.4 的很少有匹配点,大于 0.6 的则存在大量错误匹配点。作者建议 ratio 的取值原则如下:

ratio = 0.4 对于准确度要求高的匹配;
ratio = 0.6 对于匹配点数目要求比较多的匹配;

ratio = 0.5 一般情况下。
也可按如下原则:当最近邻距离 < 200 时 ratio = 0.6,反之 ratio = 0.4。ratio 的取值策略能排除部分错误匹配点。

为了进一步去除错误匹配,提高匹配准确率,根据交集的思想,提出双向匹配算法:基于上述第一次匹配结果,反过来求第二个特征集中已被匹配的关键点在第一个特征集中匹配点,即求已被匹配的关键点在第一个特征集中的最近邻与次近邻的距离比率 ratio, ratio 小于某个阈值的匹配点才认为是正确匹配。这里 ratio 都取 0.6,以找到尽可能多的匹配点。

计算 SIFT 特征向量时,同一点可能有多个方向,因此属于不同的特征点,他们中间的全部或部分可能产生正确的匹配对,但它们实际上是同一点,因此需要根据像素坐标删除这些重复的匹配点。双向匹配可去除大部分的重复匹配点。

2 实验结果

实验用图片为:(1) 256 × 256 的 lena 原图像;(2) 旋转 30° 的图像;(3) 旋转 30°、缩小 0.3 倍的图像;(4) 旋转 60° (裁剪)、放大 2.5 倍的图像、加入均值为 0 方差为 0.01 的高斯噪声;(5) 任意裁剪一个 100 × 100 区域、旋转 30° (不裁剪)、放大 2 倍的图像;(6) Lowe 演示程序中的 box. pgm;(7) Lowe 演示程序中的 book. pgm;(8) Lowe 演示程序中的 scene. pgm。除了 ratio 和 step 外, SIFT 特征提取和匹配的其它参数凭经验确定,且保持不变。

用自适应修改采样步长的方法对上述两两图像对进行匹配,匹配结果和运算时间见表 1,可见总的匹配结果为原来的 92%,而计算时间减少了 17%。

用不同的 ratio 分别进行 SIFT 图像匹配,匹配结果和运算时间见表 2,可见自适应 ratio 可在一定范围内提高匹配的准确率,而经过双向匹配后,所有错误匹配点全部删除,同时保留了尽可能多的匹配点。

综合应用自适应修改采样步长算法和双向匹配算法,匹配结果见表 2 的最下面 3 行,匹配结果图像见图 1,可见减少运算时间的同时,取得了最佳的匹配性能。

表 1 固定 step 和自适应修改 step 的匹配结果 (ratio = 4)

匹配方法/图像对	1,2	1,3	1,4	1,5	6,8	7,8	总计
step = 1 的匹配点 (个)/时间 (ms)	144/2734	19/1391	67/6172	11/2359	25/5437	33/6172	299/24265
自适应修改 step 的匹配点 (个)/时间 (ms)	141/2218	19/1125	60/5375	11/1985	21/4406	24/5063	276/20172

表2 不同 ratio 和双向匹配算法的匹配结果 (step = 1)

匹配方法/图像对	1,2	1,3	1,4	1,5	6,8	7,8
ratio = 4 的匹配点 (个)	144	19	67	11	25	33
ratio = 5 的匹配点 (个)	172	23	91	15	48(2 个一对多错误匹配)	58(1 个孤立点错误匹配)
ratio = 6 的匹配点 (个)	182	25	106	23(1 个孤立点错误匹配)	62(3 个一对多错误匹配)	81(1 个孤立点错误匹配)
自适应修改 ratio 的匹配点 (个)	180	25	103	13	47(2 个一对多错误匹配)	61(1 个孤立点错误匹配)
双向匹配 (个) (ratio = 6)	174	22	103	20	52	63
自适应修改 step + 双向匹配 (个)	172	22	103	20	48	61



(a) Lowe 的匹配结果 :5453 ms, 62 个匹配, 含 3 个错误匹配



(b) 自适应步长的匹配结果 :4515 ms, 62 个匹配, 含 3 个错误匹配



(c) 自适应步长和双向匹配的匹配结果 : 4437 ms, 48 个匹配, 无错误匹配

图1 应用自适应修改采样步长算法和双向匹配算法匹配结果

3 结论

SIFT 是一种很好的尺度、旋转、亮度不变量,但计算复杂度高,关键点的特征描述符是 128 维高维向量,用于图像匹配有其特殊性。基于 OpenCV 用 C++ 语言实现了 SIFT 算法,在计算关键点的特征描述符时,邻域梯度直方图的计算采用自适应采样窗口步长的方法,大大提高了 SIFT 的计算速度,整体平均时间可减少 20% 左右。基于 SIFT 特征进行图像匹配时,分析了关键点与关键点集之间的最近邻距离与次近邻距离的比率 ratio 对匹配结果的影响,提出了 ratio 的选择原则;并提出了双向匹配算法,在尽量保持正确匹配结果的情况下,能消除更多的错误匹配,为 SIFT 算法在图像配准、摄像机标定、目标跟踪等领域的应用提供了更鲁棒的方法。

[参考文献]

- [1] Harris C, Stephens M. A combined corner and edge detector[A]. *Proc. of The Fourth Alvey Vision Conference*[C], Manchester, UK, 1988: 147 ~ 151
- [2] Tomasi C, Kanade T. *Detection and Tracking of Point Features*[R]. Carnegie Mellon University Technical Report CMU - CS - 91 - 132, Pittsburgh, USA, Apr 1991
- [3] Lowe D G. Object recognition from local scale - invariant features [A]. *International Conference on Computer Vision*[C], Corfu, Greece, Sep 1999: 1150 ~ 1157
- [4] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91 ~ 110
- [5] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors [J]. *International Journal of Computer Vision*, 2004, 60(1): 63 ~ 86
- [6] Mikolajczyk K, et al. Local features for object class recognition [A]. *Proc. ICCV'05*[C], Beijing, China, 2005, Volume 2: 1792 ~ 1799
- [7] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors[A]. *Proc. Conf. Computer Vision and Pattern Recognition*[C], 2004: 511 ~ 517
- [8] Ledwich L, Williams S. Reduced SIFT features for image retrieval and indoor localisation[A]. *Australasian Conference on Robotics and Automation*[C], Canberra, Australasian, 2004
- [9] Delponte E, et al. SVD-matching using SIFT features[A]. *Proceedings of the International Conference on Vision, Video and Graphics*[C], Edinburgh, UK, July 2005: 125 ~ 132