

基于 SIFT 的三视图图像特征匹配算法

张连怡, 王爱平, 万国伟, 李思昆

(国防科技大学计算机学院, 长沙 410073)

摘要: 提出一种基于 SIFT 的三视图图像特征匹配算法。采用 SIFT 算法对三视图图像进行特征匹配, 引入计算机视觉中的三线性关系和基于三焦张量的像素转移误差计算方法, 对 SIFT 算法的匹配结果进行筛选, 剔除匹配过程中产生的误配点。真实图像数据实验表明, 该算法有效剔除了 SIFT 算法产生的 97% 以上的误配点, 提高了三视图图像匹配的准确度。

关键词: SIFT 算法; 三线性关系; 三焦张量

SIFT-based Feature Matching Algorithm of Three-view Images

ZHANG Lian-yi, WANG Ai-ping, WAN Guo-wei, LI Si-kun

(Computer School, National University of Defence Technology, Changsha 410073)

【Abstract】 This paper proposes a novel method of SIFT-based feature matching of three-view image. SIFT based feature matching algorithm is used in three-view images matching process. Then the trilinear constraint and trifocal-tensor-based pixel transfer error computation method is introduced to remove the outliers produced by the SIFT algorithm. Experiments are carried out on the real three-view images, and the result shows that the algorithm highly improves the veracity of three-view images matching, with 97 percent of the outliers eliminated.

【Key words】 SIFT algorithm; trilinear constraint; trifocal tensor

1 概述

图像特征匹配是计算机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。图像特征匹配通常仅针对两幅图像, 随着摄影测量和计算机视觉的发展, 多幅图像的特征匹配得到了深入研究, 其中同一场景下 3 幅图像的特征匹配具有较高的研究意义和应用价值。

图像匹配常用的特征主要有封闭的区域、边缘轮廓、角点等^[1], 用于特征匹配的特征描述子和相似性度量标准有很多, 如灰度相关、二值图像相关、链码相关、结构匹配、斜面匹配、不变矩之间的距离、动态规划和松弛法等^[2]。近年来, 基于局部不变量描述子(local invariant descriptor)的方法在目标识别和匹配领域取得了显著进展, 其中 SIFT(Scale Invariant Feature Transform)算法对光照变化、尺度变化和旋转等具有一定的不变性, 稳定性好, 是性能最优的算法之一^[3]。但是, 在图像匹配过程中, SIFT 算法主要依赖图像的灰度信息, 因此存在错误匹配的问题。

三视几何(three-view geometry)包含了 3 幅图像中全部的透视几何关系, 独立于场景结构, 仅依赖于 3 幅图像中的投影关系。三视几何通过三焦张量(trifocal tensor)来表示。三焦张量给出了 3 幅图像中点与点、点与线和线与线之间严格的几何约束关系。

针对同一场景下 3 幅图像的特征匹配问题, 本文将 SIFT 特征匹配和三焦张量的几何约束关系结合起来, 提出一种新的基于 SIFT 特征的三视图图像匹配算法。该匹配算法不仅具有 SIFT 算法的显著优点, 而且可以通过三焦张量的几何约束消除错误的匹配点, 准确完成三视图图像的特征匹配。

2 基于 SIFT 特征的匹配算法

Lowe 在总结了现有的基于不变量技术的特征检测方法的基础上, 提出了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子: SIFT 算子。基于 SIFT 特征的匹配算法^[4]主要分为 2 个部分: (1) 建立尺度空间, 检测并提取关键点, 为每个关键点建立相应的特征描述子; (2) 利用关键点的特征描述子进行匹配。

SIFT 算法在多尺度空间进行关键点(key point)的检测和提取, 保证了特征具有尺度不变性。首先将输入的图像和不同的高斯核作卷积得到二维图像的尺度空间(scale space), 2 个相邻尺度的图像相减得到高斯差分尺度空间(DOG scale space)。在高斯差分尺度空间中, 精确确定关键点的坐标、对应尺度和方向。之后按照如下步骤为每个关键点建立特征描述子: (1) 选取关键点周围一个 16×16 像素大小的邻域, 将再将该邻域均匀地分为 4×4 个子区域(每个子区域大小为 4×4 像素), 对每个子区域计算梯度方向直方图(直方图均匀分为 8 个方向)。(2) 对 4×4 个子区域的 8 方向梯度直方图根据位置依次排序, 得到一个 $4 \times 4 \times 8 = 128$ 维的向量, 该向量定义为 SIFT 特征描述子。(3) 将这个特征描述子归一化, 去除光照变化的影响, 最终得到对光照变化、尺度变化和旋转不敏感的 SIFT 局部特征描述子。

确定了关键点并建立了相应的 SIFT 局部特征描述子后, 再进行图像的特征匹配。通常采用关键点特征描述子的欧式距离来作为两幅图像中关键点的相似性判定度量。取一幅图像中某个关键点, 并找出与其匹配的图像中欧式距离最近的前 2 个关键点。在这 2 个关键点中, 如果最近的距离除以次近的距离小于某个阈值, 则接受这一对匹配点。为了进行高效的匹配, 通常采用基于 k-d 树的近似 BBF 搜索算法^[4]进行匹

基金项目: 国家“863”计划基金资助项目(2006AA01Z309)

作者简介: 张连怡(1984-), 女, 本科生, 主研方向: 计算机应用技术; 王爱平、万国伟, 博士研究生; 李思昆, 教授、博士生导师

收稿日期: 2007-07-15 **E-mail:** lianyizhang117@yahoo.com.cn

配。

使用 SIFT 的匹配算法建立起来的匹配集合称为粗匹配集合。在该集合中通常包含以下的错误匹配点：(1)有多个匹配点的特征点；(2)由于灰度相关值较高而被收录的不应匹配上的点。在有重复图案的情况下出现第 2 类误配点的情况特别严重。

3 三焦张量的性质和计算方法

三视几何通过三焦张量 T 来表达 3 幅图像中的透视几何关系。 T 由三个 3×3 矩阵 $[T_1, T_2, T_3]$ 组成。利用三焦张量, 当已知 3 幅图像中两幅之间匹配点的位置时, 第 3 幅图像中对应匹配点的位置按下式确定:

$$x_l'' = x_l' \sum_{k=1}^{k=3} x_k T_k^{jl} - x_j' \sum_{k=1}^{k=3} x_k T_k^{il} \quad (1)$$

其中, $i, j, l = 1, 2, 3$ 。

式(1)共有 9 个方程, 称为三线性关系, 其中有 4 个方程是线性独立的^[5]。该线性关系也被称为像素转移计算公式。

三焦张量的估计值可以通过 3 幅图像的匹配点计算得到。计算方法一般分为七点算法^[6]和基于投影矩阵的六点算法^[7]。七点算法采用同名点的三线性关系的方程组计算三焦张量。基于投影矩阵的六点算法, 需要确定 3 幅图像对应的 3 个摄像机的投影矩阵, 进行矩阵变换后即可求得三焦张量。但在实际计算过程时, 一般不只采用七组点或六组点, 而是尽可能使用多的匹配点以消除噪声影响, 通常采用随机抽样的 RANSAC 策略。RANSAC 策略中抽样次数随着使用的匹配点的增加而成指数增长, 因而达到同样的估计精度, 使用六点算法比七点算法快。但在采用 RANSAC 方法和六点算法计算三焦张量的过程中, 欧几里德空间中任意 4 点到射影空间射影坐标系 4 个规范基的变换矩阵的逆变换存在问题^[8], 因此本文采用了 RANSAC 策略和七点算法来计算三焦张量。

4 基于 SIFT 特征的三视图图像匹配算法

仅采用 SIFT 算法对 3 幅图像进行特征提取和匹配, 匹配结果中会包含一定数量的误配点, 导致匹配结果不准确。同一场景下的 3 幅图像存在确定的透视几何关系, 可以将其作为特征匹配的独立约束条件, 对 SIFT 算法产生匹配结果进行筛选, 以提高三视图图像匹配的准确性。基于 SIFT 特征的三视图图像匹配算法分为 2 个阶段: (1)对 3 幅图像进行 SIFT 特征匹配, 得到初始匹配集合; (2)利用三焦张量对初始匹配集合进行筛选, 得到最终准确的匹配集合。

第 1 阶段, 对 3 幅图像 I_1, I_2, I_3 进行 SIFT 特征提取后, 任取两幅图像进行特征匹配, 得到匹配集合 P_1 。再将第 3 幅图像的特征点集与 P_1 进行匹配, 得到 3 幅图像初始的 SIFT 特征匹配集合 P_2 , 该集合中含有一定数量的误配点。

第 2 阶段, 为了引入 3 幅图像中确定的透视几何关系, 需要计算三焦张量。利用初始匹配特征点集合 P_2 计算三焦张量, 会因其中的误配点而影响三焦张量解的精确性。为此在得到一个三焦张量的解后, 计算 P_2 中全部匹配点的像素转移误差, 当 的中位值 med 小于给定的阈值时, 即可判断该三焦张量的解是精确的, 否则重新计算三焦张量。计算三焦张量的伪码见算法 1。

算法 1 计算三焦张量

输入 P

输出 T

Step1 执行 RANSAC 策略, 从匹配点集合 P 中任意抽取 7 组匹配点, 采用规范化的线性方法和代数最小化方法^[5]计算

得到三焦张量估计值 T_i 。

Step2 计算在该三焦张量 T_i 下, 集合 P 中每一个匹配点的像素转移误差 d_i (transfere rror):

$$d_i = \sqrt{(\hat{x}_i'' - x_i'')^2} \quad (i = 1, 2, \dots, n) \quad (2)$$

式(2)中, \hat{x}_i'' 表示由像素转移式(1)得到的空间点 X 在第 3 幅图像中的像点; x_i'' 表示第 3 幅图像中空间点 X 实际的像点; n 表示集合 P 中的元素个数。

Step3 计算集合 P 中全部匹配点的像素转移误差集合 $\{d_i\}$ 的中位值 $median\{d_i\}$, 记为 med 。

Step4 给定阈值 t , 当 $med < t$ 时, T_i 为三焦张量 T 的最终估计值。否则执行 Step1, 重新计算三焦张量。

Step5 return T 。

给定阈值 τ , 当得到一个精确的三焦张量解后, 将初始匹配集合 P_2 中像素转移误差不小于阈值 τ 的匹配点剔除, 最终得到准确的特征匹配点集合。基于 SIFT 特征的三视图图像匹配算法的伪码见算法 2。

算法 2 基于 SIFT 特征的三视图图像匹配算法

输入 I_1, I_2, I_3

输出 Q

Step1 采用 SIFT 算法求出 3 幅图像 I_1, I_2 和 I_3 公共匹配点集合 P , P 中元素个数记为 n 。

Step2 采用算法 1 计算三焦张量估计值 T 。根据 T , 计算集合 P 中全部匹配点的像素转移误差集合 $\{d_i\}$ 。 $\{d_i\}$ 的中位值 $median\{d_i\}$, 记为 med 。

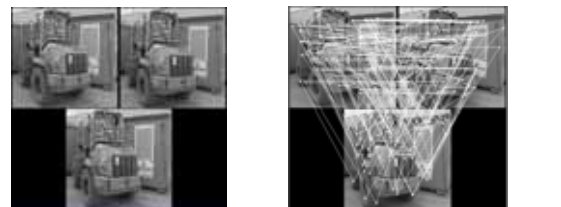
Step3 给定阈值 $\tau = 1.96\sigma, \sigma = 1.4826 \times (1+5/n-7) \times med$ 。剔除集合 P 中像素转移误差大于等于 τ 的匹配点。剩余匹配点的集合即为最终得到的三视图图像的特征匹配点集合 Q 。

Step4 return Q 。

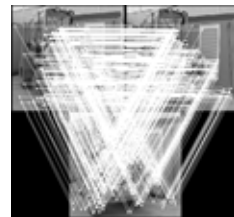
5 实验结果及分析

本文算法在含有误配点的初始集合中采用随机抽样的策略来计算三焦张量, 因此由于误配点的影响, 三焦张量的计算结果不是唯一的, 为此本文采用了多次实验的方法对算法的性能进行分析。

第 1 组实验的 3 幅图像是经摄像机旋转了一定角度后拍摄所得。3 幅图像经过 SIFT 算法匹配后, 得到的初始匹配特征点集合 P 中共含公共匹配特征点 489 对, 其中误配点 67 对(由人眼识别得到)。误配点在整个集合 P 中占 13.70%, 如图 1(b)。



(a) 3 幅原始图像 (b) SIFT 算法运算结果中全部的误配点



(c) 阈值 $t=1.5$ 时不含误配点的 select 集合

图 1 实验 1 原始图像和实验结果

集合 P 经本文算法处理后,得到最终的匹配点集合(下文称为 select 集合)。阈值 t 分别取 1.5 和 2.0 个像素单位,即集合 P 中全部匹配点的像素转移误差的中位值 med 分别小于 1.5 和 2.0 个像素。本文算法在 2 个阈值下分别独立执行了 100 次,得到的 select 集合的匹配点数目和出现的误配点的统计数据如表 1、表 2 所示。

表 1 实验 1 匹配点数的相关参数

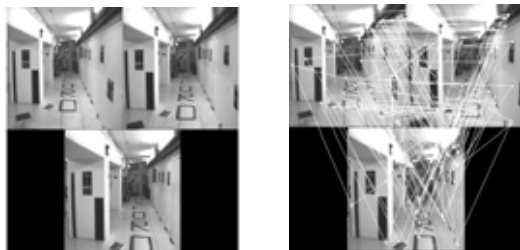
阈值	匹配点数的最小值	匹配点数的平均值	误配点数的最大值
1.5	341	359	1
2.0	334	363	2

表 2 实验 1 误配点的频率

阈值	select 集合中误配点的数目		
	2	1	0
1.5	0 次	16 次	84 次
2.0	4 次	60 次	36 次

从表 1 中可以看出,本文算法得到的 select 集合中匹配点的数目约为初始集合 P 的 75% 左右,如图 1(c)所示。从表 2 中可以看出(以次数计算),当阈值 $t=1.5$ 个像素单位时,100 次运算结果中有 84 次得到的 select 集合中不含任何误配点,另外 16 次运算得到的 select 集合中只包含 1 个误配点,98.5% 的误配点被剔除。当阈值 t 变为 2.0 个像素单位时,100 次运算结果中有 96 次得到的 select 集合中最多只含有 1 个误配点,97.0% 以上误配点被剔除。

第 2 组中的 3 幅图像是摄像机由远及近拍摄所得。3 幅图像经初始 SIFT 算法匹配后得到公共匹配特征点 389 对,其中误配点 44 对(由人眼识别得到)。误配点在集合 P 中占 11.31%,如图 2(b)所示。



(a)3 幅原始图像 (b)SIFT 算法运算结果中全部的误配点



(c)阈值 $t=1.5$ 时不含误配点的 select 集合

图 2 实验 2 原始图像和实验结果

阈值 t 分别取 1.5 和 2.0 个像素单位。本文算法在 2 个阈值下分别独立执行了 100 次,得到的 select 集合的匹配点数目和出现的误配点的统计数据如表 3、表 4 所示。从表 4 中可以看出(以次数计算),当阈值 $t=1.5$ 个像素单位时,100 次运算得到的全部 select 集合中均不包含任何误配点。当阈值 $t=2.0$ 个像素单位时,100 次执行结果中只有 1 次得到的 select

集合中出现了误配点,且只有 1 个,97.7% 的误配点被剔除。从表 3 中可以看出,阈值取 1.5 和 2.0 个像素单位时,本文算法得到的 select 集合中的匹配点的数目约占集合 P 的 77% 左右,如图 2(c) 所示。

表 3 实验 2 匹配点数的最小值、均值和误配点数的最大值

阈值 t (像素单位)	select 集合中的匹配 点数的最小值	select 集合中的匹配 点数的平均值	select 集合中的误配 点数的最大值
1.5	273	297	0
2.0	271	301	1

表 4 实验 2 误配点的频率

阈值	select 集合中误配点的数目	
	1	0
1.5	0 次	100 次
2.0	1 次	99 次

另外还对计算三焦张量的抽样次数做了统计:阈值 $t=2.0$ 时,第 1 组图像平均需 160 次抽样,第 2 组图像平均需 39 次抽样;阈值 $t=1.5$ 时,第 1 组图像平均需 450 次抽样,第 2 组图像平均需 130 次抽样。从抽样的次数上来看,随着阈值精度的提高,计算三焦张量的时间开销随之增大。如何进一步减少误配点,如何降低算法的时间复杂度是下一步的工作。

6 结束语

本文将图像识别和匹配领域中性能优越的 SIFT 算法和三视几何中的几何约束关系结合起来,提出并实现了一种基于 SIFT 特征的三视图图像匹配算法。在匹配过程中,该算法保留了正确的 SIFT 特征匹配点,并有效地剔除了 SIFT 算法产生的误配点,显著地改善了 SIFT 算法的匹配结果,实现了 3 幅图像特征的正确匹配。

感谢 澳大利亚阿德莱德(Adelaide)大学理论数学系的 Tony Scoleri 和英国剑桥(Cambridge)大学视觉几何研究小组为本文提供的真实图像数据。

参考文献

- [1] Yilmaz A, Javed O, Shah M. Object Tracking: A Survey[J]. ACM Comput. Surv., 2006, 38(4): 45-50.
- [2] Zitová B, Flusser J. Image Registration Methods: A Survey[J]. Image and Vision Computing, 2003, 21(11): 977-1000.
- [3] Mikolajczyk K, Schmid C. A Performance Evaluation of Local Descriptors[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(1): 1 615-1630.
- [4] Gordon I, Lowe D G. What and Where: 3D Object Recognition with Accurate Pose[J]. Toward Category-level Object Recognition, 2006, 30(12): 67-82.
- [5] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision[M]. [S. l.]: Cambridge Univ. Press, 2000.
- [6] Hartley R I. Lines and Points in Three Views and the Trifocal Tensor[J]. International Journal of Computer Vision, 1997, 22(2): 125-140.
- [7] Torr P H S, Zisserman A. Robust Parameterization and Computation of the Trifocal Tensor[J]. IVC, 1997, 15(8): 591-605.
- [8] 郑 敏. 三视张量在序列影像特征点匹配中的应用研究[D]. 武汉: 武汉大学, 2004.