

学生编号

2022-2023 学年秋季学期

《计算思维实训（1）》(0830A030)

课程报告

成绩 (百分制)	
-------------	--

学 号	21122782	学 院	计算机工程与科学学院
姓 名	陈诺	手工签名	
报告题目	基于 SIFT 的图像拼接功能		
实训报告成绩 50%	实训过程描述：清晰、全面。（5%）		
	报告主要部分：1、计算思维实例叙述清楚、有意义；2、分析到位；3、思路独特或有创意；4、关键实现技术描述清楚详细；5、内容丰富；6、不直接粘贴所有代码；7、代码有注释或说明。（25%）		
	实训收获体会：感受真实、深刻，建议有价值。（10%）		
	书写格式：书写规范、用词正确、无明显的错别字，图表、代码清晰规范，格式协调、效果好，参考文献书写、引用规范合理。（10%）		
工作实绩 50%	1、平时表现、进步情况（25%），2、工作实绩（25%）		

教师对该生工作实绩简要评述：

教师签名：

日期： 年 月 日

Part 1. 本学期实训过程概述

1. 实训总体概况、实训过程
2. 对计算思维的认识
3. 实训体会及建议

Part 2. 综合实训报告

基于 SIFT 算法的图像拼接功能

21122782 陈诺

计算机工程与科学学院

摘要： 通过使用 SIFT 算法，获取两张输入图像的关键点与特征向量，通过 KNN 将两组特征向量配对，并用 RANSAC 筛选外点（噪音）。对处理后特征向量的关键点对求解单应性矩阵，并根据矩阵将右图做透视变换，最后将处理后的右图叠加到左图上，获得拼接后的图像。本文将分析此程序所使用的算法，并对于不同类型输入的敏感性与景深对拼接功能的影响做出评价。

关键字： 计算机视觉，SIFT，关键点检测，RANSAC，KNN，全景拼接

5 一、引言

图像拼接是将多个重叠的图像对齐成一个大的组合，它代表了一个 3D 场景的一部分。拼接可以看作是场景重建的一种特殊情况，其中图像仅通过平面单应性进行关联。图像拼接在运动检测和跟踪、增强现实、分辨率增强、视频压缩和图像稳定等机器视觉领域有很大的应用。

本次实训期望实现的是将同一场景不同视角所拍摄的两张图片实行拼接，将通过尺度不变特征变换（Scale-invariant feature transform, SIFT）检测图像的特征点，而后使用 K 最邻近（K-nearest-neighbor, KNN）算法将两张图片的特征点匹配，并通过随机采样一致（Random Sample Consensus, RANSAC）筛选匹配的特征点实现降噪，最后根据有效匹配的特征点计算 Homography 单应性矩阵，从而对图片进行透视变换，最终完成拼接操作。

6 二、算法分析

1. 尺度不变特征变换（Scale-invariant feature transform, SIFT）^[1]

SIFT 可在图像中获取特征点，且这些特征点具有尺度不变性。由于 SIFT 特征是局部特性，有实验证明，在图像噪声为 10% 的情况下，仍有近 80% 的关键点的位置和尺度可以正确检测，体现了 SIFT 特征的稳健程度；再比如，即使两幅图像的成像视点相差超过 50° 时，最终可以正确检测到的特征仍然在 50% 以上，这说明 SIFT 特征对几何形变具有较强的适应性^[3]，在旋转、尺度缩放、亮度变化时具有不变性，因此被用于相似图像的特征点匹配。

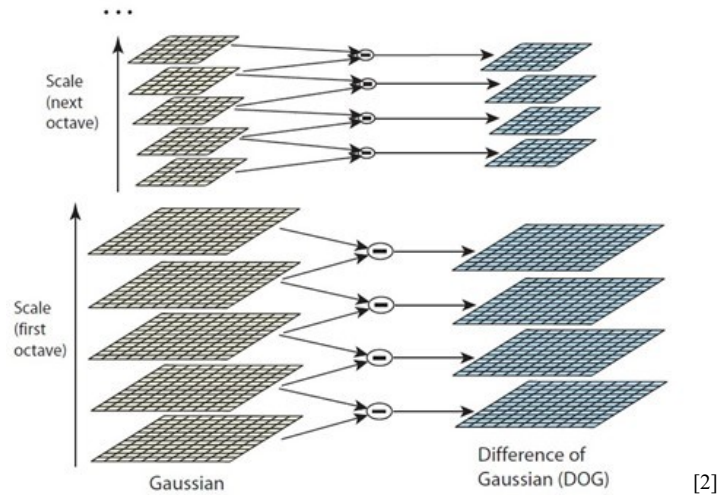
SIFT 的实现主要有四个步骤：尺度空间检测、关键点定位、方向确定与关键点描述。

尺度空间检测需要对图像建立高斯金字塔，并生成高斯差分金字塔以实现图像特征点（差分金字塔上的极值点）的提取。通过对原图取特定的 n 个 σ 做高斯模糊，将其称为一组，而后对第 $n-2$ 张图片降采样，并重复 n 次，我们将得到高斯金字塔以描述

不同尺度的特征；而后将高斯金字塔每组相邻的图片两两做差，得到 n 组有 $n-1$ 张图片的高斯差分金字塔（Difference of Gaussians, DoG）。

其中， n 的选取满足如下计算公式， M 、 N 为图像的大小：

$$n = \log_2 \min(M, N) - t, t \in [0, \log_2 \min(M, N)]$$



关键点定位需要对 DoG 中每组图片求导，得出极值点（关键点）。关键点是由 DOG 空间的局部极值点组成的，关键点的初步探查是通过同一组内各 DoG 相邻两层图像之间比较完成的。为了寻找 DoG 函数的极值点，每一个像素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。由于要在相邻尺度进行比较，只能在中间 $n-2$ 层中进行两个尺度的极值点检测，其它尺度则只能在不同组中进行。为了在每组中检测 S 个尺度的极值点，则 DOG 金字塔每组需 $S+2$ 层图像，而 DOG 金字塔由高斯金字塔相邻两层相减得到，则高斯金字塔每组需 $S+3$ 层图像，实际计算时 S 在 3 到 5 之间。在多尺度空间采用 DOG 算子检测关键点，相比传统的基于 LoG(Laplacian of Gaussian)算子的检测方法,运算速度大大加快 [4]。

关键点的方向确定是为了使描述符具有旋转不变性，需要利用图像的局部特征给每一个关键点分配一个基准方向，一般采用 DoG 金字塔中关键点 3σ 邻域窗口内像素的梯度和方向分布特征，将其按 10° 一个方向分类，即有 36 个方向，并统计邻域窗口内像素的梯度方向，将最高频的方向作为该关键点的主方向，将大于主方向 80% 的方向作为辅方向以增强鲁棒性。最后为了提高精确性，可将统计到的离散的梯度方向进行插值拟合处理以获得更精确的角度方向值。至此，我们获取了含有位置、尺度与方向的 SIFT 特征点信息。

SIFT 算法的最后一步是关键点特征描述。通过用一组向量描述关键点，使其不随光照、视角的改变而发生变化。通常采用关键点尺度空间内 4×4 的窗口中计算 8 个方向的梯度信息，共 $4 \times 4 \times 8 = 128$ 维向量以表征。具体实现步骤为：

- (1) 确定计算描述子所需的图像区域
- (2) 将坐标轴旋转为关键点方向，以确保旋转不变性
- (3) 将邻域内的采样点分配到对应的子区域内，将子区域内的梯度值分配到 8 个方向上，计算其权值
- (4) 插值计算每个点八个方向的梯度

至此，SIFT 特征点获取完成。

在实现图像拼接的过程中，由于基于 python 的 SIFT 算法效率过低，因此实际使用时调用了 opencv 的 `xfeatures2d.SIFT_crreat()` 方法创建 SIFT 容器以计算特征点与特征向量。

2. K 最邻近 (K-nearest-neighbor, KNN)

SIFT 算法能有效地获取图片特征点以及特征向量，为了有效利用这些数据，实现图像拼接功能，需要使用一个高效的分类方法。通过使用 KNN，我们能迅速的匹配特征向量，从而找到对应的特征点。

本次实训将通过调用 opencv BFMatcher 的 `knnMatch` 方法实现 KNN 分类。

3. 随机采样一致 (Random Sample Consensus, RANSAC) [5]

由于两张图片的 SIFT 的部分特征点不能一一匹配，所以还需要筛选算法去除外点（噪音）。RANSAC 算法可以在一组包含“外点”的数据集中，采用不断迭代的方法，寻找最优参数模型，不符合最优模型的点，被定义为“外点”。

以下是 RANSAC 的实现思路：

- (1) 随机从数据集中随机抽出 4 个样本数据（此 4 个样本之间不能共线），计算出变换矩阵 H ，记为模型 M ；
- (2) 计算数据集中所有数据与模型 M 的投影误差，若误差小于阈值，加入内点集 I ；
- (3) 如果当前内点集 I 元素个数大于最优内点集 I_best ，则更新 $I_best = I$ ，同时更新迭代次数 k ；
- (4) 如果迭代次数大于 k ，则退出；否则迭代次数加 1，并重复上述步骤；

注：迭代次数 k 在不大于最大迭代次数的情况下，是在不断更新而不是固定的；本次实训通过调用 opencv 的 RANSAC 实现。

4. 计算 Homography 单应性矩阵

通过处理筛选配对后的关键点，我们可以计算两张图片之间的映射关系，通过一个单应性矩阵 H 来描述，而后将需要处理，即拼接的图片 $P2$ 与 H 相乘，即可得到与被拼接图 $P1$ 视角一致的 $P2'$ 。最后只需要将 $P2'$ 叠放至 $P1$ 上，就可以完成初步的图像拼接功能。

本次实训通过调用 opencv 的 `findHomography()` 与 `warpPerspective()` 方法实现计算单应性矩阵与图像透视变换。

7 二、实例分析

(1) 处理同源图片平移、旋转、透视变换的拼接

以下是测试图片 2.1:



图 2.1 测试图片

a. 裁剪测试图片，将其分为有重合的两部分



图 2.1.1 Left



图 2.1.2 Right

分别获取裁剪后图片的关键点和特征向量:



图 2.1.3 Left 的关键点



图 2.1.4 Right 的关键点

通过 KNN 将特征向量匹配，并用 RANSAC 筛选去除外点，将匹配成功的特征向量对应的关键点连线：



图 2.1.5 Left、Right 对应匹配关键点连线



图 2.1.6 Left、Right 合并

计算单应性矩阵 H 并将 Right 图做透视变换，左右图合并。



b. 裁剪测试图片，并将一部分旋转



图 2.1.7 Left



图 2.1.8 Right

分别获取裁剪后图片的关键点和特征向量：



图 2.1.9 Left 的关键点



图 2.1.10 Right 的关键点

通过 KNN 将特征向量匹配，并用 RANSAC 筛选去除外点，将匹配成功的特征向量对应的关键点连线：



图 2.1.11 Left、Right 对应匹配关键点连线

图 2.1.12 Left、Right 合并
计算单应性矩阵 H 并将 Right 图做透视变换，左右图合并。

c. 裁剪测试图片，并将一部分水平透视



图 2.1.13 Left



图 2.1.14 Right

分别获取裁剪后图片的关键点和特征向量：

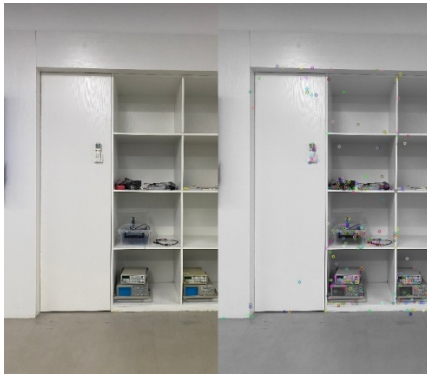


图 2.1.15 Left 的关键点



图 2.1.16 Right 的关键点

通过 KNN 将特征向量匹配，并用 RANSAC 筛选去除外点，将匹配成功的特征向量对应的关键点连线：



图 2.1.17 Left、Right 对应匹配关键点连线



图 2.1.18 Left、Right 合并

计算单应性矩阵 H 并将 Right 图做透视变换，左右图合并。

d. 裁剪测试图片，并将一部分垂直透视



图 2.1.19 Left



图 2.1.20 Right

分别获取裁剪后图片的关键点和特征向量：

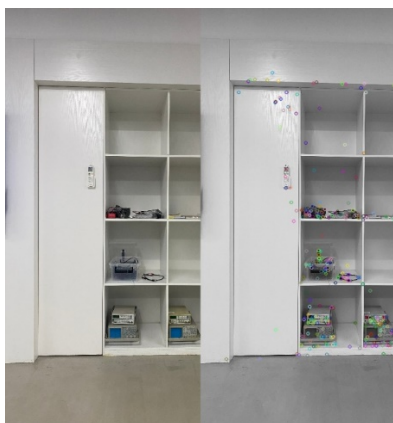


图 2.1.21 Left 的关键点



图 2.1.22 Right 的关键点

通过 KNN 将特征向量匹配，并用 RANSAC 筛选去除外点，将匹配成功的特征向量对应的关键点连线：

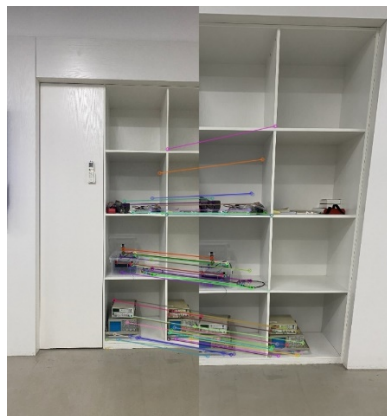


图 2.1.23 Left、Right 对应匹配关键点连线 图 2.1.24 Left、Right 合并

计算单应性矩阵 H 并将 Right 图做透视变换，左右图合并。

(2) 处理远景图片的拼接



图 2.2.1 Left



图 2.2.2 Right

分别获取图片的关键点和特征向量：



图 2.2.3 Left 的关键点



图 2.2.4 Right 的关键点

通过 KNN 将特征向量匹配，并用 RANSAC 筛选去除外点，将匹配成功的特征向量对应的关键点连线：

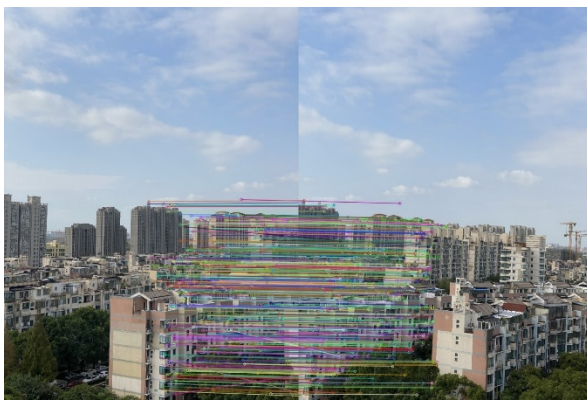


图 2.2.5 Left、Right 对应匹配关键点连线



图 2.2.6 Left、Right 合并

计算单应性矩阵 H 并将 Right 图做透视变换，左右图合并。

(3) 处理近景图片的拼接



图 2.3.1 Left



图 2.3.2 Right

分别获取图片的关键点和特征向量：



图 2.3.3 Left 的关键点



图 2.3.4 Right 的关键点

通过 KNN 将特征向量匹配，并用 RANSAC 筛选去除外点，将匹配成功的特征向量对应的关键点连线：

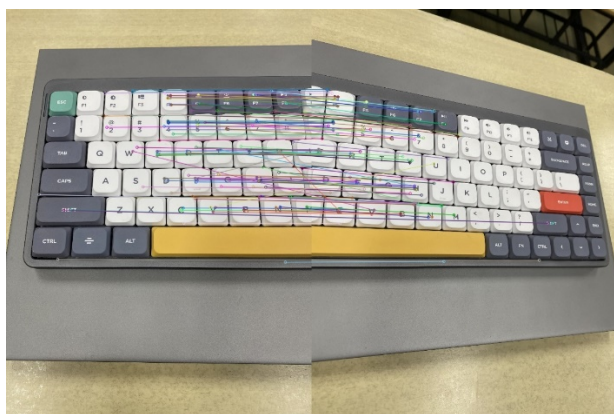


图 2.3.5 Left、Right 对应匹配关键点连线



图 2.3.6 Left、Right 合并

计算单应性矩阵 H 并将 Right 图做透视变换，左右图合并。

8 三、小结

本文通过使用 SIFT 算法实现了图像拼接功能，根据测试用例可看出：本程序对于图像关键点的裁剪、旋转、水平与垂直透视变换都不具有敏感性，面对不同输入有较高的鲁棒性；在拼接景深较大的图片时，拼接效果较佳，主要问题在于拼接后的图片的拼接缝与色差问题，预测可以通过使用图像融合算法避免此现象；而在处理近景图片的拼接时，出现了明显的拼接错位现象，而图像中心已对齐，预测为本程序选择的透视函数或 RANSAC 筛去外点时出现异常，或可以通过将 Left 图像同时做透视变换以优化效果。

参考文献

- [1] David G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
- [2] SIFT 算法详解 <https://blog.csdn.net/zddb1992/article/details/7521424>
- [3] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints.[J]. International Journal of Computer Vision, 2004, 60(2) : 91-110.
- [4] Krystian Mikolajczyk and Cordelia Schmid. Scale & Affine Invariant Interest Point Detectors[J]. International Journal of Computer Vision, 2004, 60(1) : 63-86.
- [5] Fischler M A, Bolles R C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Automated Cartography[J]. Communication of the ACM, 1981, 24(6): 381—395.