主要算法：

1. 收集各个景点的图片（比如 jpg 图像，每个景点 500 张），然后将所有这些图片的 SIFT 的 keypoints 提取出来；
2. 然后将所有的这些图像的 SIFT keypoints 进行聚类，根据参考文献，可以使用 k-means 进行聚类，k 取值为 100~500 之间，然后得到 100~500 个聚类中心（center）；
3. 得到的所有的 keypoints 其实是属于某一个景点的，而聚类之后每一个 keypoint 又是可以划分到这 k 个 center 中的一个，这样就可以统计每一个景点在每一个 center 上有多少个 keypoints，例如选取玉泉的 5 个标志性的建筑（maoxiang，caozhulou，damen，tushuguan,…）来实验，k-means 的 k 取值 200，则可以使用如下形式的数组来统计这些信息;
   maoxiang[200], caozhulou[200], tushuguan[200],…，假设 maoxiang[45]=1236，则表示所有 maoxiang 图片得到的 SIFT keypoints 中有 1236 个 SIFT keypoints 是属于 center45 的；
4. 得到了每个标志性景点图像的 keypoints 分布情况之后，对于一张输入的待判定图像，首先提取出它的 SIFT keypoints，然后计算出它每一个 SIFT keypoints 是属于哪一个 center（直接通过距离计算，计算每一个 SIFT keypoint 与哪一个 center 的距离最小），这样就可以得到这张待测试图像的 SIFT keypoints 的分布情况，也可以使用类似的数组来描述：
   Image[200]，比如 image[45]=219，则表示这张图像的所有 SIFT keypoints 里面，有 219 个是属于 center45 的；
5. 计算这张图像是属于哪一个标志性景点，还是简单的根据距离来计算，依次计算 dist(landmarki[200], image[200])，即计算待测试图像的 SIFT keypoint 分布向量（image[200]）和之前得到的每个景点的 SIFT keypoint 分布向量（maoxiang[200]，tushuguan[200]，damen[200]等等），找出 dist(landmarki[200], image[200])最小的 i，那么这张待测图像就属于 landmarki。

实验和编程实现的过程：

1. 首先将各个景点的 jpg 的图像（训练图像）转换成灰度的 pgm 图像（有些 sift 程序处理需要 pgm 格式），可以使用 IMAGEMAGICK 来进行处理，使用简单的命令就可以完成图片的批转换：
   mogrify -format pgm *.jpg
2. sift 算法。Sift 算法现在已经很成熟，国外有很多有名的 project 做包括 sift 在内的各种图像处理算法，所以完全不需要也不必要再自己实现 sift 算法来用在该大作业中。Sift 算法比较有名的是 University of California 的 Andrea Vedaldi 教授的一个 c++实现的 sift 算法 siftpp（其实也是被 VLFeat 支持的），具体参见 http://www.vlfeat.org/~vedaldi/code/siftpp.html 。而 VLFeat（http://www.vlfeat.org/）库中也有很好的 sift 的实现。

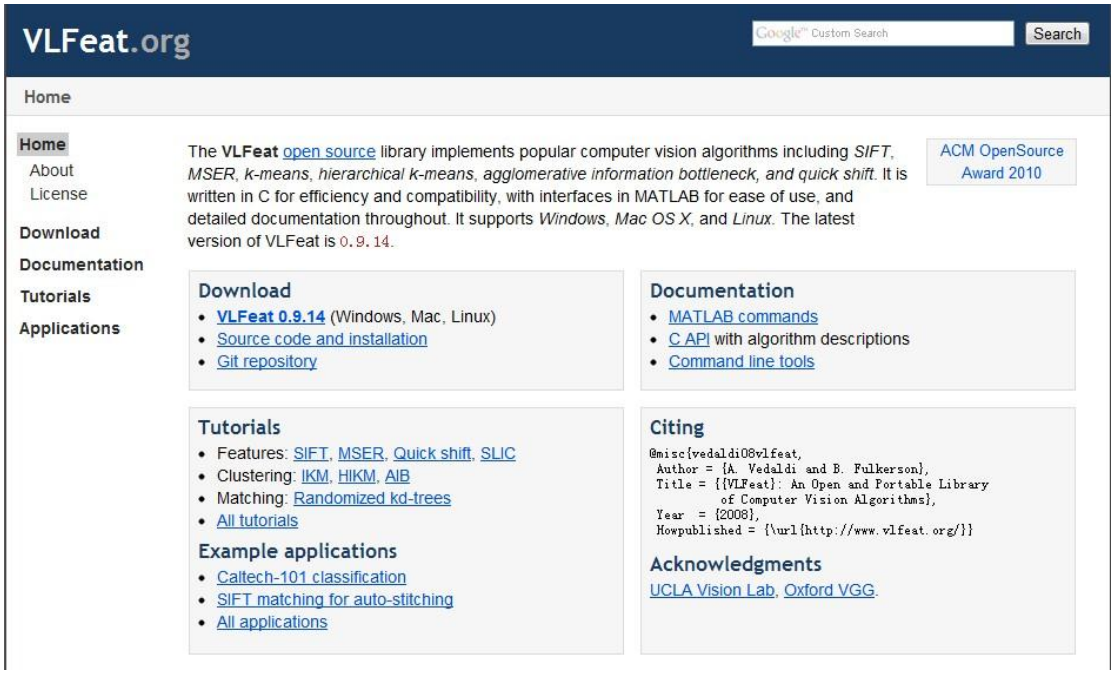


图 1 VLFeat 中的 sitf 实现

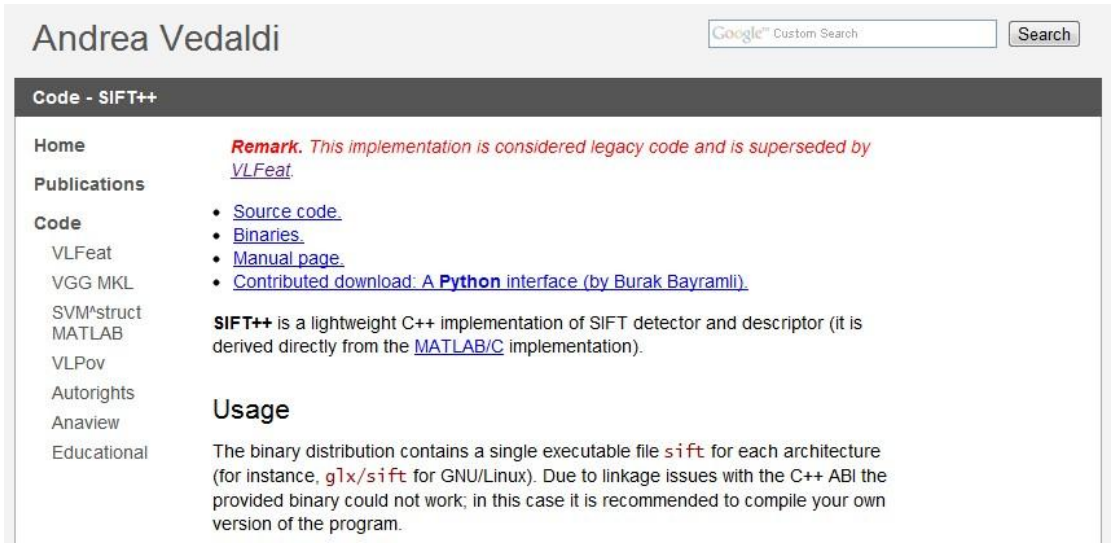


图 2 Andrea Vedaldi 的 sitfpp 实现

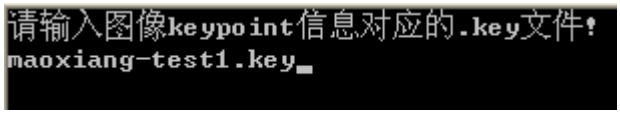
本实验中选取的是 siftpp 的实现；

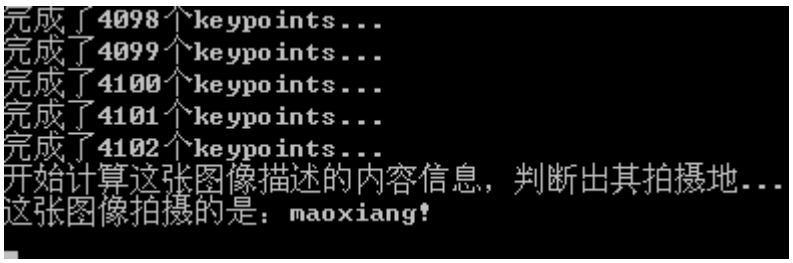
3. 聚类算法。聚类算法一般都是非常需要 cpu 和内存的，本实验中参考了 ICML'03 中的一个快速 k-means(http://www-cse.ucsd.edu/~elkan/fastkmeans.html),它是一个 matlab code 的快速 k-means，本次实验中将其改写成了对应的 c++实现（可能在效率上没有仔细考虑）；
4. 本实验的主要代码（除去 pgm 转换和 SIFT keypoints 的提取是在 linux 下完成的）是在 windows 环境下的 VS2008 中实现的，可能在 linux 环境中运行需要将程序进行少部分的改动；
5. TellALandmark 中是一个完整的实现过程(需要读取所有图像的 keypoint 的.key 文件)，包括对所有 keypoints 进行聚类，计算每个景点的 keypoints 的分布情

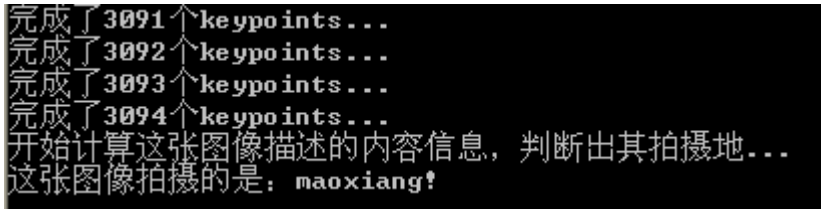
况和计算待测试图像的 keypoints 分布情况，最后判断待测图像是属于哪一个 landmark。完整的运行整个程序可能需要很长时间（根据训练图像的多少和每张图像的 keypoints 的多少）；

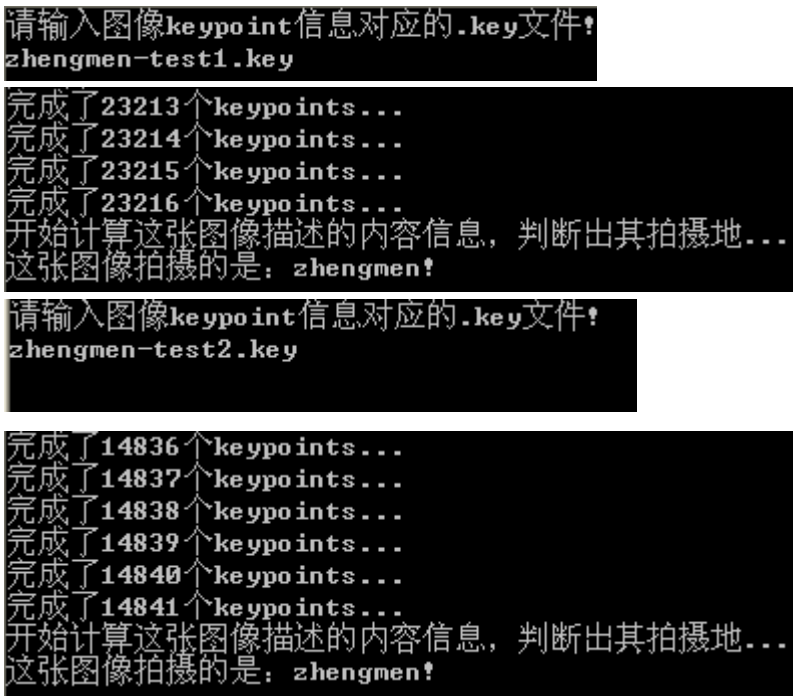
6. TestTellALandmark 是一个测试一张输入图像的程序，该程序直接使用已经提取出的各个景点的 keypoints 的分布数据（使用 TellALandmark 提取），直接输入一个待测试图像的 SIFT keypoint 文件（.key 文件）即可。该程序提取的是玉泉的 5 个景点，分别是曹光彪主楼、玉泉大门、图书馆、毛主席像、以及生仪楼，但是训练的数据集很小，每个景点只选取了 10 张图像，聚类也没有等到完全收敛就停止了（当时选取的原始图像过大，提取的 SIFT keypoints 过多，每一张有 2W+歌 keypoints，当时从时间上考虑，所以没有等到完全收敛就强制停止了 k-means 算法的迭代，建议进行训练的每一张图像的长和宽选取为 500pix 左右，这样每一张图像的 keypoints 不会过多，大概 2k 左右），所以有些照片的测试结果可能与实际情况有差别，若要获得更准确的实验结果，建议重新运行 TellALandmark（即重新选取更多的训练图像，重新进行训练）；
7. 本实验由于时间仓促，代码是几个部分合在一起的，代码没有进行很详细的调整，里面既有 c 代码也有 c++代码，代码风格和规范上做的很差；
8. 本实验重新选取了更大的数据集（分别从互联网上爬取了埃菲尔铁塔、自由女神像、进门大桥、凯旋门、比萨斜塔、上海东方之珠和美国银行中心大厦各 200 张图像）重做实验，结果还没有出来；
9. 该实验所用的代码仅作为计算机视觉的提交作业，不经允许不能上传到互联网上！作者可能会进行后续工作。

部分测试结果









在测试结果中，毛像的每一张都能正确的预测出来，而大门和曹主楼的有时候会预测错误，这主要与两个方面的因素有关：1.最主要的是训练的样本数量不够，每个景点只选取了 10 张，聚类只选取了 100 个 center，keypoints 的聚类也没有完全收敛，而且照片都是在旁晚（吃晚饭的时间）拍摄的，光线不好；2.毛像的照片与其他照片有明显的不同，因为基本上上半部分都是天空，图像的特点很突出。

参考文献

[1] Lew, Sebe, Djeraba, and jian, “Content-base Multimedia Information Retrieval: State of the Art and Challenges”, ACM Transactions on Multimedia Computing, Communications and Applications, ACM Press, 2006:1-19.

[2] J. Hays and A. A. Efros, “Im2gps: estimating geographic information from a single image”, in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008:1-8.

[3] P. Serdyukov, V. Murdock, and R. van Zowl, “Placing flickr photos on a map”, in SIGIR’09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM, 2009:484-491.

[4] D. J. Crandall, L. Backstrom, D. Huttemlocher, and J. Kleinberg, “Mapping the world’s photos”, in WWW’09: Proceedings of the 18th international conference on World Wide Web. New York, NY, USA: ACM, 2009: 761-770.

[5] Carneiro G., Chan A. B., Moreno P. J., and Vasconcelos N., “Supervised Learning if Semantic Classes for Image Annotation and Retrieval”, IEEE Transactions on

Pattern Analysis and Machine Intelligence, March 2007, 29(3):394-410.

[6] J. Smith and S. Chang, “Querying by color regions using the VisualSEEK content-based visual query system”, In: M. Maybury, editor, Intelligent Multimedia Information Retrieval, AAAI Press, 1997.

[7] J. R. Smith, “Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis”, PhD thesis, Graduate School of Arts and Science, Columbia University, February 1997.