CMPT 888 Assignment 1: Panorama Stitching

Colin J Brown, Instructor: Ping Tan

April 1, 2015

1 Introduction and Method

The goal of this assignment was to write a program to generate a panorama from a set of images. In particular, the images are assumed to be photos taken from the same position but in different directions with overlapping fields of view. To acheive this, features are extracted from each image to find corresponding points which are used to compute a homography transform between pairs of images.

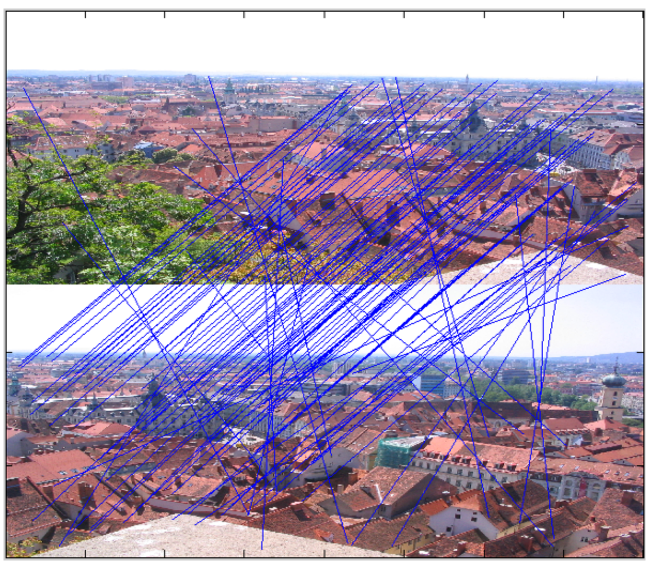

Figure 1: Two images with SIFT features extracted. Lines are drawn between pairs of matching features. The view in the top image is looking slightly to the left of that in the bottom image. Thus, the majority of features in corresponding pairs are on the right side of the top image and left side of the bottom. Clearly there are also some erroneously matched features.

First, for each image, SIFT features are extracted (using a Matlab implementation by Andrea Vedaldi[1]) [2]. A maximum of 5000 features per image are retained for each image to maintain reasonable performance. For each pair of neighbouring images, a set corresponding features are found by comparing the SIFT descriptors. Figure 1 shows two images with lines drawn between matching feature pairs and some incorrectly corresponding features.

Next a homography transform is computed to match one moving image to the reference image. Given a set of at least 4 corresponding points, a homography transform can be computed using a direct linear transform (DLT). Since some feature point-pairs are incorrectly matched, their use in computing DLT would lead to an incorrect image alignment. However, since using subsets of correctly matched point-pairs should lead to similar homographies, the incorrect pairs can be treated as outliers and removed using RANSAC [1].

Here, a point is treated as an inlier if it is within 1% of the image width from its corresponding point after being transformed to the reference space. Each iteration, RANSAC selects a random set of point-pairs, computes DLT and then counts the set of inliers. The largest set of inliers across all iterations is kept. If 4 points are sampled each iteration and 50% of the point-pairs are assumed outliers then 72 iterations of RANSAC must be run to ensure that a sample containing no outliers is found (with 99% confidence).

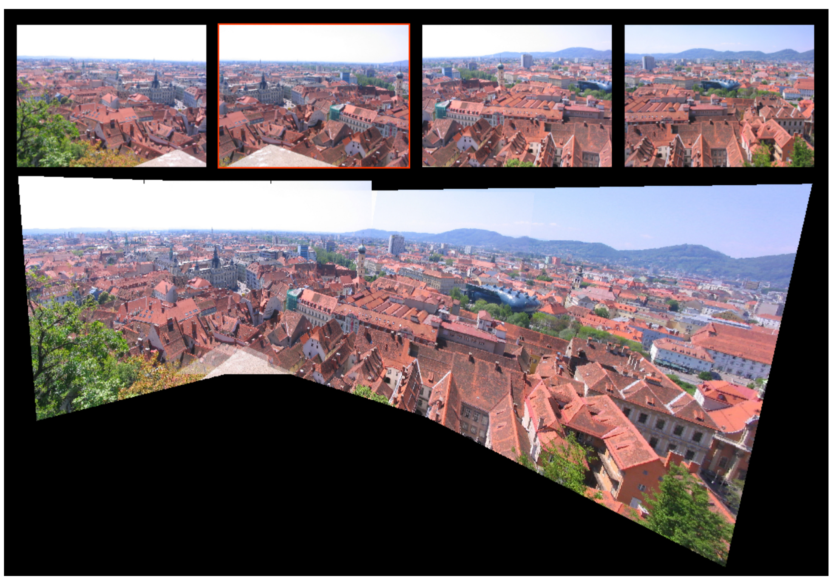

Figure 2: Top: Four images taken from different views at the same loction. Bottom: The resultant panorama using the image second from the right (red border) as reference frame.

With a set of inliers computed DLT is performed one last time on this entire set to generate the homography used to transform the image. To actually

[1]http://www.cs.ucla.edu/ vedaldi/

transform the image, the inverse homography is computed and pixels are filled in the reference space based on their source value from the moving image. To increase efficiency, the bounding box of the moving image in reference space is computed and only pixels within this bounding box are sampled. If pixels from two images are overlapping, a linear blend of the two values is used.

2 Experiments and Discussion

The panorama implementation was tested on two different datasets. Figure 2 shows the result on the given rooftops dataset. There is a misalignment artefact on the visible edge of the stone platform at the bottom of the left two images. This artefact is caused by changing camera position as the camera was rotated. This effect is most pronounced on objects near to the camera, so objects like the platform in the foreground can become more obviously misaligned. Using a proper panorama tripod mount or ensuring that no foreground objects are visible in the scene would avoid this problem. Using a better error function such as the the reprojection error instead of the algebraic error (used here) might also find a better homography transform and alleviate this problem somewhat.

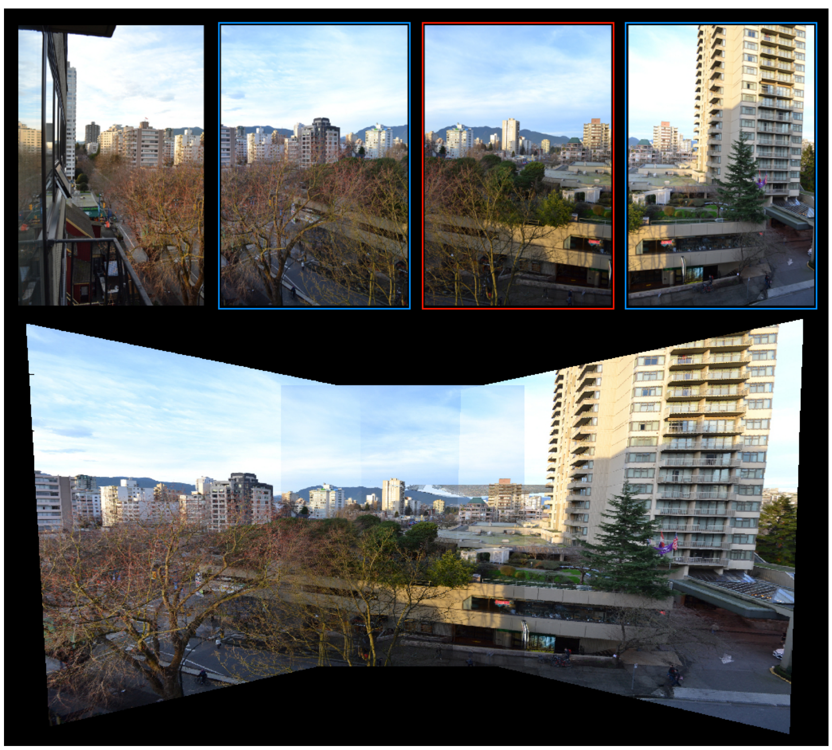

Figure 3: Panorama constructed from 3 right-most images using second from right as reference.

Figure 3 shows the result on 3 images taken from a balcony. Certain artefacts are more clear in this example. For instance, edges are visible in the sky denoting pixel intensity differences at the same region of the two images. This is likely due to different exposure times caused by different scene brightnesses as the camera was rotated. To fix this, either the camera could be set to a fixed exposure setting or the images could be normalized or blended in overlapping regions.

Another issue visible in this panorama is that of the misaligned tree in the foreground similar to the problem with the stone platform in the previous example.

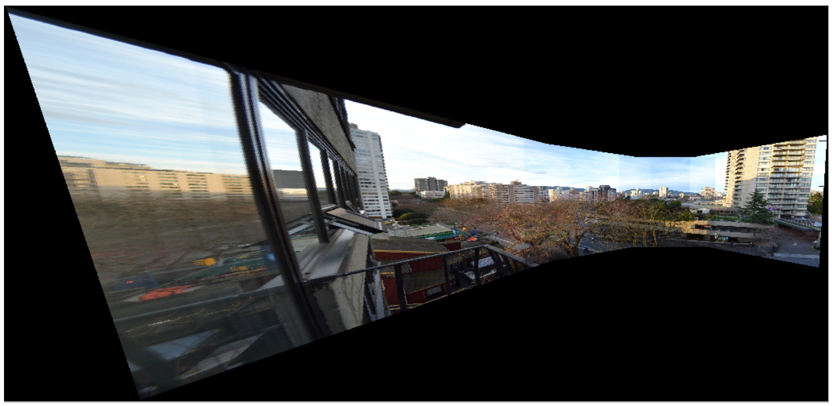

Figure 4: Panorama constructed from all 4 of the images in Figure 3.

In Figure 4 a panorama is shown that was generated from all 4 images shown in Figure 3 again using the image second from the right as the reference image. In this case, the left-most image becomes more distorted as it has to be deformed more severely to be aligned to the second left-most image since the second left-most image is already deformed.

References

[1] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[2] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.