1. **Image Alignment with RANSAC**

A. Show the best point correspondence results with different distance thresholds.

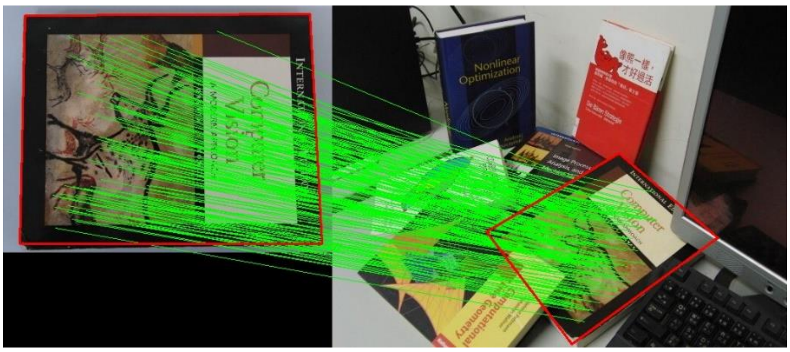
(1) Setting nOctaveLayers = 7, contrastThreshold = 0.005, edgeThreshold = 40, sigma = 2.1, SIFT detector can extract 2517, 1394, 1763, 1666 feature points from 1-image.jpg, 1-book1.jpg, 1-book2.jpg and 1-book3.jpg, respectively.

(2) Then, by setting threshold = 300, ratio = 0.85, 2NN algorithm used in HW1 can match 250, 302, 95 feature points in 1-book1, 1-book2 and 1-book3 with the same number of feature points in 1-image, respectively.

B. Compare the parameter setting in SIFT feature and RANSAC and discuss the result.

With iterations = **iter**, threshold = **t**, subsetSize = **s**, RANSAC will randomly select **s** pairs of feature point correspondences and use them to form a homography matrix H. Then, points on the single-book image of each pair of correspondences will be transformed by H and compared with their corresponding feature points, if distance between a transformed point p and its corresponding points is shorter than **t**, p will be regarded as an inlier. This procedure will be repeated for **iter** iterations and the homography leading to the most inliers will be returned as the final result.

Setting the value of iterations, threshold and subsetSize as 1000, 18 and 4, we can obtain three good alignment results:



As three pictures shown above, book on the left image are perfectly aligned to book on the right image for all of three pairs of images, and the mean lengths of deviation vectors between transformed points and corresponding points for all inliers are 1.906045, 1.044174 and 1.149743. However, the mean lengths of deviation vectors for all points are 85.669673, 66.811534 and 132.901647. The large differences between these values are caused by some miss-matched correspondence not eliminated in the process of 2NN-algorithm. Since the alignment results are not affected by these miss-matched correspondence as the numbers of well-matched correspondences are large enough, we can ignore the large differences between mean lengths of deviation vectors. But, we can still modified the parameters input into SIFT detector and lower the value of threshold and ratio in 2NN-algorithm to detect less feature points (with smaller nOctaveLayers, edgeThreshold and larger contrastThreshold) and eliminate more poorly-matched correspondences to lower the difference between mean length of deviation vectors (e.g. setting threshold = 200 and ratio = 0.6, mean lengths of deviation vectors for all points are lowered to 19.062609, 7.938926 and 43.639283), but it should be also noticed that eliminating too many correspondences in 2NN-algorithm may result in bad alignment result since some well-matched correspondences may be eliminated.

2. **Image Segmentation**

A. Discuss the difference between the results for different K's.

The parameter K in K-means algorithm is the number of cluster centers. Each pixel in an image will be put into a cluster depending on their distance to each cluster center in the color space. According to the results for k = 3, 5, 7, it can be indicated the larger K is, the more complex and colorful the segmented image will be, that is because each cluster can be regarded as a color, and color of each pixel will be set as one of these K color finally. Also, the sum squared distance of each point to closest center $\sum_{clusters\ i}\sum_{p\ in\ cluster\ i}|\boldsymbol{p-ci}|^2$ will be smaller for larger K since each pixel has more choice for its ideal cluster center.

B. Discuss the difference between K-means and K-means++

K-means will always converge to some solution, but sometimes it may lead to a local minimum, which means the sum squared distance of each point to closest center is not globally minimal. The reason for this circumstance is that sometimes several points should be classified to the same segment are selected as centers simultaneously, and then they will converge to some similar solutions. To solve this problem, after randomly choosing the first center, K-means++ will select k - 1 new centers from all points depending on probability proportional to their distance (in this problem, distance means Euclidean distance between RGB values) to all selected centers. Therefore, points should be in the same segment will be less likely to be chosen as centers simultaneously and the global minimum are more likely to be reached.

For problem 2, K-means and K-means++ are both executed for 50 iterations. According to the best result of 50 iterations, we can tell that the centers (colors) chosen by K-means++ are more representative.

C. Mean-shift algorithm on RGB color space.

Setting two parameter threshold and bandwidth, color value of each point p on the image will move to the mean of color values of all points in the scope of a sphere centered at p in the RGB space until the distance to move is shorter than threshold.

D. Mean-shift algorithm on RGB color space and x-y space.

In addition to RGB space, mean-shift algorithm in this problem also considers the x-y coordinate space for each point p, that is, only points whose distances to p are shorter than bandwidth in both color space and coordinate space will be brought in the computation of shift destination.

E. Discuss the segmentation results for different bandwidth parameters.

40, 80 and 100 are set to bandwidth in this problem. According to the result images, a smaller bandwidth generates more small segments with different colors while the segments generated by a larger bandwidth are smoother and flatter and with less type of colors. The reason for difference between bandwidths is that the bandwidth determines the size of shifting window used in mean-shift algorithm. The bigger the shifting window is, the shifting destination of a point will be more easily influenced by the majority of all points since there may be more points with the dominant value in the range of bandwidth; in the extreme case, all points will shift to the global mean if the bandwidth is as large as the size of image. Besides, an excessively small bandwidth may also cause some problems since it will restrict the calculation of shift destination to a small area. In practice, both of too small bandwidth and too large bandwidth cannot segment an image ideally since the former will generate too many tiny and trivial parts and another may miss lots of important information.

F. Compare the segmentation results by using K-means and mean-shift algorithms and their computational cost.

The number of segments in results by K-means is limited by the parameter K and sometimes local minimum may happen, while we do not need to make any model assumption for mean-shift and it is more generalized for complex models with nonconvex shape.

As for computational cost, according to many experiments, the running time of K-means is usually a few minutes while the running time of mean-shift ranges from one hour to tens of hours depending on the parameter bandwidth. The reason for this considerable difference is that K-means only needs to consider the convergence of k centers while mean-shift need to shift all of n points in an image to their convergence points, where n is much larger that k.