Akinlawon Solomon
ECE 7866 Computer Vision
HW3

3D Rigid Structure From Motion

Given m images of n points an arbitrary scene, the structure from motion problem (SFM) seeks to recover the three-dimensional shape of objects within the scene. Moreover using these images, the projection matrices associated with the cameras observing these shapes can be obtained. Put simply, the solution to this problem produces the shape of the scene and the corresponding camera locations within some reference frame.

Method:

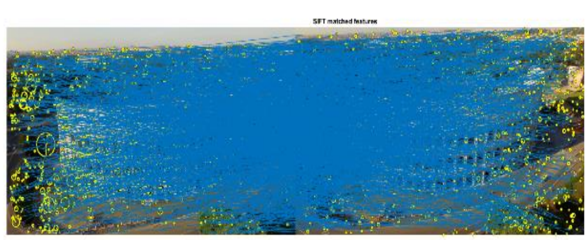
A series of steps were used to first obtain the useful point correspondences across a subset of the images, then the selected points were used to recover the scene shape and structure up to a transformation.

Step 1:

- First a subset of the images were chosen to be used in the recovery process. A total of ten images were used that appear to include interior of the dome. The algorithm worked better when the depth of the coliseum was fully visible in all views. The point correspondences were then obtained using the SIFT algorithm.

Step 2:

- After the images are selected, features must be matched across images. In order to do this, one of the images was selected first and its closest descriptors in all other images were calculated. This set of closest descriptors were chosen as the matched features. A total of 1296 features, obtained by tweaking threshold parameters, were used. Figure 1 shows matched points across two of the images. This matching technique meant that significant point outliers were reduced. Since SIFT also employs pre-filtering, potential noise effects could be ignored.



Akinlawon Solomon
ECE 7866 Computer Vision
HW3

Step 3:

- The co-ordinate locations of these matched features as well as a matrix of observed points W was created. A key issue here is that the points must be normalized so that the mean of all points is 1 and the variance is$\sqrt{2}$. Once this matrix was formed, two different methods were used to estimate the camera matrices M and the scene structure P were used.

- First the projective factorization method, where the matrix W is factorized according to the equation $W = MP$ using SVD. The 4 largest eigenvectors are then chosen such that $M = UD$ and $P = V^T$. Then, using these initial guesses the equation $\| W - MP \|^2$ is minimized in the non-linear least squares sense. This method is known as the bundle adjustment technique, and a number of optimization techniques can be used to solve it. In this project, the Levenberg-Marquardt algorithm was used.

- The second method used was the two-view structure from motion algorithm. In this method, two images are used to recover the scene structure. The camera calibration matrix is first approximated, then used to calculate the fundamental matrix relating the two images. Using SVD, the camera matrices can be recovered. The points are recovered using a triangulation technique, which compares measured 2D points using the calibration matrix to the predicted 2D points using the projected matrices in the least squares sense.

  The recovered 3D shape using the two-view method is depicted below

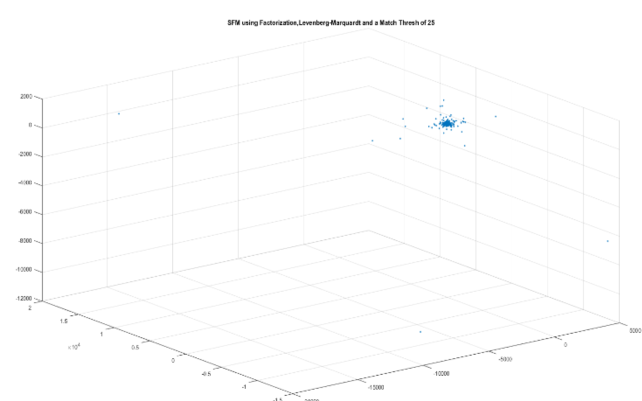
Figure 1. SFM using views image 3000 and 3054 on the left; views from image 3000 and 3036 on the right.

Akinlawon Solomon
ECE 7866 Computer Vision
HW3

Discussion

Several key issues arose in the development of this algorithm.

- The projective factorization method doesn't work well when the number of points is too large across all images. Although a minimum is obtained, the initial guess is critical to getting meaningful results. A final residual of the order of 10^-4 was obtained but the structure is not remotely close to the actual shape. This method also has slow convergence when a large number of points are provided. This tradeoff means that complex shapes will be inaccurate using this method. A plot showing the 3D scene recovered is included in the folder (algorithm was stopped only after a few iterations).
- The two view method recovers some semblance of the 3D structure, but again this is highly dependent on the views specified and the number and strength of the matched points. The choice of views to use as well the points affects the results quite strongly.
- Since the projective method was used, the solution is only correct up to an arbitrary transformation. We can clearly observe that the 3D points are scaled and rotated contrary to what we would expect. This implies that if we can use a Euclidean upgrade in order to get a more meaningful result. Prior knowledge of the camera parameters may also help in producing a more accurate shape.
- The algorithm obtains points along the horizon of the image, which leads to the appearance of greenery within the structure. It seems that these points could be eliminated to yield a more accurate shape.



Akinlawon Solomon
ECE 7866 Computer Vision
HW3