

The survey of SFM Algorithm

Wenyang Zheng^{1,a}

¹School of Software, Xidian University, Xi'an 710126 ,China;

^azwenyang@outlook.com

Keywords: Kd-tree, RANSAC, SIFT, BA beam adjustment method

Abstract. The purpose of this paper is to describe the SFM algorithm flow. And the algorithm consists of five steps, including feature extraction, matching and creating a track, find the initial pair, initializing the relative orientation of the image pair, adding more images.

Introduction

SFM [1-2] is an algorithm based on the collected unordered images for 3D reconstruction. We first extract the focal length information from the picture, use the feature extraction algorithm such as SIFT to extract the image feature, and use the kd-tree model to calculate the Euclidean distance between the two feature points to match the feature points to find the number of feature points To achieve the desired image pair. For each image matching pair, the F matrix is estimated and the matching pair is optimized by the ransac algorithm. In this way, if a feature point can be passed in such a matching pair, it has been detected, so that a trajectory can be formed.

After entering the structure-from-motion section, the key first step is to select a good image pair to initialize the entire BA process. First of all, select the first two images for the first BA, and then loop to add a new picture to the new BA, and finally until there is no picture that can be added, the BA is finished. Get the camera to estimate the parameters and the sparse 3D point cloud. The bundle between the two images is a sparse beam adjustment method sba software package, which is a nonlinear least squares optimization objective function algorithm.

Details

Feature detection. The use of SIFT [3-4] with scale and rotation invariance improves stability. Suitable for extracting and rotating the angle of a variety of image feature point information, its accuracy is strong, in this algorithm does not need to consider the time cost of the case also more advantages. The SIFT algorithm calculates the position information (x, y) of the feature points by different Gaussian filters (DOG), and also provides a descriptor descriptor information. In a 4 * 4 square histogram around a feature point, a histogram contains the gradient direction of eight bins, that is, a 4 * 4 * 8 = 128 dimension eigenvector.

Feature matching. Once the feature points of each picture are raised, it is necessary to match the feature points between the two pictures, and F (I) is used to represent the feature points around the image I. For each image pair I and J, consider each feature $f \in F(I)$ to find the nearest neighbor eigenvector $f_{nn} \in F(J)$ (Formula 1).

$$f_m = \arg \min_{f' \in F(J)} \left\| f_d - f'_d \right\|_2 \quad (1)$$

In fact, the algorithm using a kd-tree to calculate the nearest adjacent matching. And then we suppose the nearest distance is d1, after that we find out the second nearest matching point and suppose the distance is d2. If the specific value of d1 and d2 less a key value (eg:0.6), we make sure we can accept this matching pair. In this case, the number of matching point between picture I and picture J is not more than 1, but in the picture J, it is possible to match several feature point.

Obviously, we will get the situation of many to one, but in fact the feature points should one-to-one match. And we need an algorithm to remove the repetitive feature points and we use such algorithm to solve the many-to-one match. Finally, if there is not less than 16 feature points between two pictures, we called such image pair initial choosing image pair

However, initial choosing image pair tend to be unreliable, we should use some approach to detect those pictures. We adopt geometric constraint to detect. This detection based on the truth. We suppose a static scene and not all the feature points fit the physical significance in the scene. In this case, we should calculate epipolar geometry. F matrix can associate the pixel between two pictures and it also contains the parameter of camera. Each matching match pair of pixel coordinates needs to satisfy the following Formula 2.

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} F \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \quad (2)$$

The RANSAC (Random Sampling Consistency) algorithm is used to filter the RANSACA algorithm with the 8-point method. The threshold of the number of external points should be less than 0.6% of the image length and width. The

When all the two pairs of matching images have been determined, you can consider connecting the common feature matching points that appear in multiple images to form a trajectory. For example, the feature $f1 \in F(I1)$ matches the characteristic $f2 \in F(I2)$, $f2$ matches the feature $f3 \in F(I3)$, which can form a trajectory $\{f1, f2, f3\}$. And then use the width-first search BFS to find the complete trajectory of each feature point in all image pairs.

Once the trajectories are found, the image connection graph is constructed, including the nodes of each image, and the edge of the image with the common trajectory.

The external parameters of the camera are described by the $3 * 3$ rotation matrix R and the $1 * 3$ translation vector (or the camera center coordinate vector). The camera's internal parameters are described by a focal length f and two radial distortion parameters k1 and K2. The geometric scene provides each 3D point X_j in the trajectory, and a 3D point X_j is projected onto the 2D image plane of the camera by the projection equation. The projection error is the distance between the projection point and the real point on the image. As shown Figure 1.

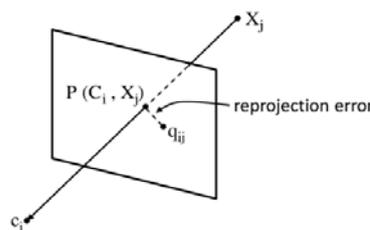


Fig. 1 Illustration of reprojection error

For n perspectives and m trajectories, the target optimization equation for the projection error can be written as Formula 3.

$$g(C, X) = \sum_{i=1}^n \sum_{j=1}^m w_{ij} \left\| q_{ij} - P(C_i, X_j) \right\|^2 \quad (3)$$

First select the appropriate initialization image pair, which is very important, once the wrong initialization, will fall into the local optimal and make the subsequent BA into an infinite loop, cannot correctly solve the global optimal. There are two specific requirements: first, there must be enough matching points; second, there should be enough far from the camera center.

In particular, here we use the uniaxial model between the two image transformations to find the initial image pair. If not well consistent with the unilateral model, indicating that the camera center

or a certain distance. Also use the RANSAC method to reduce noise, improve the reliability of matching, try to choose a low percentage of the internal point, but at least 100 to ensure that the matching point.

The system uses the 5-point method to estimate the initialization of the matching pair, and then the triangulation of the trajectory can provide the initialized 3D point, the initial two frames can start the first bundle adjustment. Sparse bundle adjustment (SBA) is used here.

Finally, add new cameras and 3D points to the BA. This process until the rest of the camera observes no more than 20 points, indicating that the rest of the camera does not have enough points to add, BA ends. Get the camera to estimate the parameters and scene geometry information, which is sparse 3D point cloud.

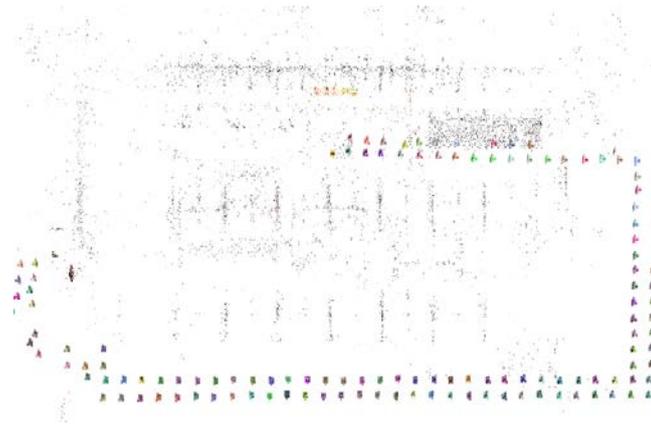


Fig. 2 Sparse point cloud example



Fig. 3 Dense reconstruction of a store

Our experimental results [5-11] have verified the correctness of our algorithmic flow on Figure 2 and Figure 3.

Conclusion

SFM workflow:

The first step is the feature extraction, the general use of SIFT operator, because of its scale and rotation invariance.

The second step is to match and create a track, the image pair match.

The third step is to find the initial pair, the purpose is to find the camera like the largest baseline, using RANSAC four-point method to calculate the single matrix, to meet the single-matrix matching point called the point, not satisfied with the single should be called outside the matrix Point, according to the single matrix formula can be seen when T is smaller, the higher the proportion of points, that is, the more obvious the phenomenon of low parallax.

The fourth step is to initialize the relative orientation of the image pair. According to the RANSC eight-point method, the eigen-matrix can be calculated. The R and T of the second image can be obtained by decomposing the intrinsic matrix SVD. In this step, distortion correction is needed, T and corrected the coordinates of the coordinates of the triangle to calculate the three-dimensional point, where the method used is a direct linear transformation DLT, can be understood as mapping in front of the intersection.

The fifth step is to add more images, the third sub-graph, for example, according to the fourth step generated three points and the third sub-graph and the first two graphs track relationship, you can reverse the third sub-graph R, T, And then continue to triangulate the calculation of more three-dimensional points, using the same DLT, so repeat the fifth step, and finally will put all the pictures of POSE (R, T) and three-dimensional point, which is sparse reconstruction of SFM results.

References

- [1] Noah Snavely, Steven M.Seitz and Richard Szeliski, "Modeling the World from Internet Photo Collections," *International Journal of Computer Vision*, November 2008, Volume 80, Issue 2, pp 189-210.
- [2] Roberto Gorrieri, "The Basic Calculus: SFM," *Process Algebras for Petri Nets* pp 77-94, Part of the Monographs in Theoretical Computer Science. An EATCS Series book series(EATCS)
- [3] Born, A.Olbrich, M.Maywald, and R,Wiesendanger "Analysis of electrical breakdown failures by means of SFM-based methods," *Applied Physics A*, March 1998, Volume 66, Supplement 1, pp 1063-1065
- [4] Shimiao Li, and Loong-Fah Cheong "Behind the Depth Uncertainty: Resolving Ordinal Depth in SFM," *European Conference on Computer Vision, ECCV 2008: ComputerVision-ECCV 2008* pp 330-343
- [5] Ruipeng Gao, Mingmin Zhao, Tao Ye, Fan Ye, Guojie Luo, Yizhou Wang, Kaigui Bian, Tao Wang, Xiaoming Li, "Multi-story Indoor Floor Plan Reconstruction via Mobile Crowdsensing," *IEEE Transactions on Mobile Computing (IEEE TMC)*, vol. 15, no. 6, pp. 1427-1442, 2016.
- [6] Ruipeng Gao, Yang Tian, Fan Ye, Guojie Luo, Kaigui Bian, Yizhou Wang, Tao Wang, and Xiaoming Li, "Sextant: Towards Ubiquitous Indoor Localization Service by Photo-taking of the Environment," *IEEE Transactions on Mobile Computing (IEEE TMC)*, vol.15, no.2, pp.460-474, 2016.
- [7] Ruipeng Gao, Mingmin Zhao, Tao Ye, Fan Ye, Yizhou Wang, Guojie Luo, "Smartphone-based Real Time Vehicle Tracking in Indoor Parking Structures," *IEEE Transactions on Mobile Computing (IEEE TMC)*, vol. 16, no. 7, pp. 2023-2036, 2017.
- [8] Ruipeng Gao, Bing Zhou, Fan Ye, and Yizhou Wang, "Knitter: Fast, Resilient Single-User Indoor Floor Plan Construction," in *Proceedings of IEEE INFOCOM*, 2017.
- [9] Ruipeng Gao, Mingmin Zhao, Tao Ye, Fan Ye, Yizhou Wang, Kaigui Bian, Tao Wang, and Xiaoming Li, "Jigsaw: Indoor Floor Plan Reconstruction via Mobile Crowdsensing," in *Proceedings of ACM MobiCom*, 2014.
- [10]K. Bian, and J. Park. Maximizing Rendezvous Diversity in Rendezvous Protocols for Decentralized Cognitive Radio Networks, *IEEE Transactions on Mobile Computing (IEEE TMC)*, Vol. 12(7):1294--1307, July 2013.
- [11]Y Liao, K Bian, L Song, Z Han. Full-duplex MAC protocol design and analysis. *IEEE Communications Letters* 19 (7): 1185-1188, July 2015.