Special Problems Report

Ayush Baid

Fall 2019

Contents

Abstract

We are building a 4D spatio-temporal reconstruction system for salad crops. This system will help monitor the growth of the plant in a non-intrusive fashion and provide data and a framework to take automated decisions. In this semester, I tackled the spatial structure-from-motion (SFM) problem to generate 3D reconstructions. The first challenge I faced was that the salad crops present a challenge for feature point matching due to the repetition of texture. I have used a couple of tricks to make the feature matching and essential matrix computation more robust. On top of that, the camera takes numerous macro shots with a small baseline to avoid texture repetition in an image and hence avoid bad feature matches. The point cloud 3D reconstruction look good and we are ready to proceed to time-domain association and reconstruction.

1 Introduction

We are working with Dr. Chen and his team at environmental science and engineering on a hydroponics project. The project aims to analyse the suitability of different waste water treatment strategies for agriculture use. Specifically, we want the capability to measure the plant size as it is highly correlated with the health of the plant. A non-intrusive way to capture the growth in plant size is essential.

The use of computer vision provides an accurate way to measure the plant growth by generating the 3D model of the salads evolving over time. The images are captured by sensors on the robot. The robot's base is suspended using cables, which are controlled using motors to provide granular movements. A robotic arm is attached to the base, which allows finer movements. Out imaging system is mounted on the tip of the arm. In the future, we want the camera system to be one of the end effector. Other end effectors can be tools which help in harvesting the salads or other types of sensors.

The imaging and reconstruction technology will provide faster and more scalable monitoring and analyis capabilities for this project. This method can also be scaled easily to have large hydroponic farms which use the nutrients in the waste water, along with providing food. Also, this project can serve as a blueprint and can be applied to other precision agriculture applications.

2 Related Work

We are trying to emulate the work by Dong et al. [1] which perform the 4D reconstruction for precision agriculture. In that work, the focus is on large farms and not on the individual plants. We plan to extend their techniques to process each plant individually and hence provide more granular results and monitoring capability.

Roy et al. [2] perform reconstruction of apples using a single camera. They face challenges using traditional point feature based front-end, and use segmentation and the fruit contour itself to generate dense matches, and perform SFM incrementally. [3] use Kinect to capture RGB-D images and use them to perform canopy analysis of cotton. An extensive survey of the use of LiDAR based crop-monitoring is presented in [4].

We prefer to use simple RGB camera because of its low cost and compact and light design. A simple camera costs less than 30 USD and hence provide extreme scalability. It is essential if we want to deploy multiple robots. The simple cameras are also light weight and can be supported by a simple robot.

3 Challenges

A common approach to perform feature matching between two images is to get the two nearest neighbor and perform the ratio test [5] to consider matches where the closest neighbor is sufficiently smaller than the second nearest neighbor.

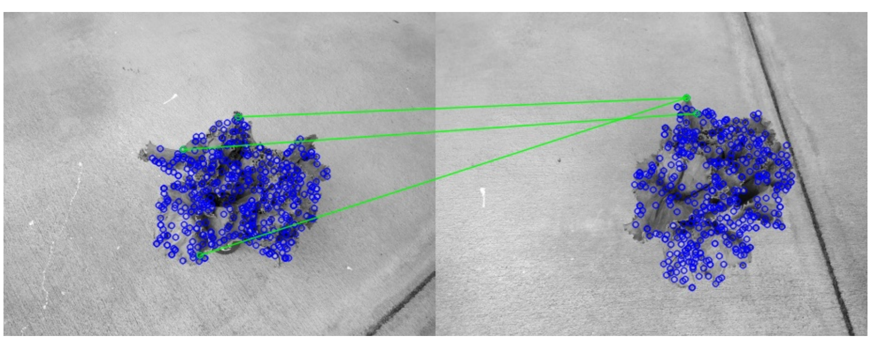

Figure 1: *Challenges in feature point matching.* The blue circles denote the feature points extracted in the individual images, and the green lines are the matches which successfully pass the ratio test.

This was a challenge for lettuces as the ratio test rejected almost all the matches. For example, figure 1 has a pair of images where just 3 matches pass the ratio test. Such a low number of matches are not sufficient for essential matrix computation, and the generated point cloud will be of an abysmally low density and hence not of much practical use.

An opposite case of this problem is faced in the blimp project, where there are incorrect feature matches due to repeated texture in the two images, and they do not get filtered in by the ratio test. This can occur due to either the ratio test being not stringent enough, or the absence of texture repetition in the second image.

4 Approach

We intend to use the structure-from-motion (SFM) algorithm [6] for reconstruction. In SFM, we process the two-dimensional image of three-dimensional scenes from different viewpoints and leverage them to reconstruct the scene. We use the *point-based* SFM where we associate 3D landmark points with the feature points in the captured image (measurements) and perform optimizations to generate a point cloud reconstruction.

As we are working with point features, we need a feature point position and descriptor extractor, which will enable us to perform associations between different images and also the 3D landmark point. We use the SIFT [5] algorithm to detect and compute feature points in the image. Once we have the feature points, the next step is to perform matching of feature points between images. We then model the bundle adjustment problem with factor graphs [7] using the matches feature points as the measurements and camera poses and 3D points as variables to be optimized for.

I tried a couple of tricks which makes the essential matrix computation more robust. This is covered in the next subsection (front-end), followed by a subsection on back-end.

4.1 Front End

Repeated texture is traditionally handled by the ratio test. When there are repeated textures, there will be multiple match candidates in close vicinity, and will violate the condition of the ratio test. There might be some cases where the ratio test fails to remove incorrect matches arising due to repeated texture. It is also computationally beneficial if we do not consider the points which correspond to repeated texture for matching.

4.1.1 Feature point filtering

I along with Shicong have introduced a pre-filtering of feature points in the workflow. This filter intents to discard feature points which repeat in the image (i.e. they are the feature points arising due to repeated texture). We use the DBSCAN clustering algorithm [8] which is a non-parametric density based clustering technique. Points which have similar texture will have descriptors which are close to each other and hence will be grouped as a cluster by the DBSCAN algorithm. Points which have a unique texture feature will be marked as outliers, and these are the points which are useful to us. We will reject all the features which are clustered together. Figure 2 presents a sample of the feature classification output. Notice that there is a lot of visual similarity in subsets or points marked red (e.g. along a similar edge/curve) and it makes sense that these points are filtered out.

This filtering could have been achieved by having a 2-way matching and ratio tests, but it is computationally expensive to perform match search over all the points when we can just pre-process the features detected from a single image.

4.1.2 Hypothesis set and Validation set in RANSAC

Due to high degree of repetition of texture in the crops dataset, a large number of matches are rejected by the ratio test. However, there is some information in the rejected matches.

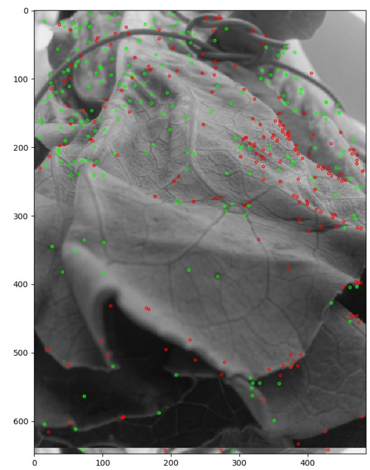

Figure 2: *Feature classification.* The green points represent features which are marked as outliers by the DBSCAN algorithm and is used downstream as a non-repeating feature point. Red points are rejected in this filtering step.

We will use an example to demonstrate the issue. Figure 3 has 9 matches which pass the ratio test. On the other hand, there are 514 matches which fail the ratio test. If we use the 5-point algorithm [9] for computing the essential matrix, there are very few unique samples of points we can use for the essential matrix computation. The candidates for essential matrix, when tested for the number of inliers among the 9 matches, most often output 5 or 6. This does not allow for comparison between different candidates. However, when the essential matrix candidates are evaluated on the 514 failed points, there is one candidate with a maximum count of 6 inliers on the ratio-test-failure set. This serves as a useful tie breaker and demonstrates that there is some useful information in matches which would be otherwise rejected.

To use this extra information in the computation of the essential matrix itself, we modify the RANSAC [10] algorithm to have two inputs, *hypothesis set* and *test set*. Samples are drawn from the hypothesis set, and the total inlier count is a weighted sum of inlier count on the two sets.

The example demonstrates an extreme case. In normal practise, this algorithm refines the essential matrix slightly to generate higher number of matches as inliers.

4.2 Back End

As we use the robotic arm to capture the images, we get good initialisation for the all the camera camera poses. Hence we add pose priors for all the camera poses. The landmarks are filtered based on a minimum number of measurements for them, and are tested for chireality. For the initialisation, camera pose estimates are used as-is and landmark initial estimates are generated using triangulation of measurements. The factor graph is generated using robust measurement noise and is optimized to generate the final poses and landmark positions.

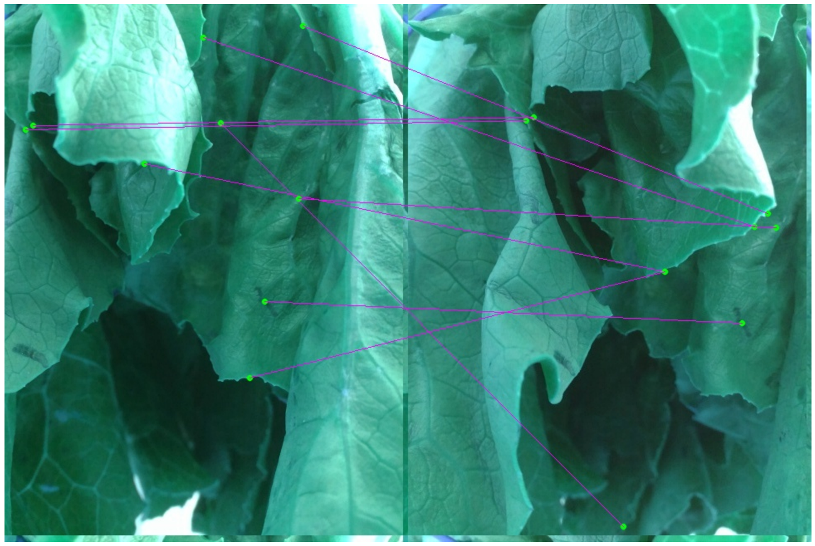

Figure 3: *Pair of images with few matches which pass the ratio test.* Notice that there are incorrect matches which remain after the ratio test.

5 Results

We captured 126 images for a single leafy lettuce plant in stacks aranged in a circular fashion around the plant. The sector angles between stacks is 15 degrees. The initialisation and final result are shown in Figure 4.

6 Discussion

The front end seems to perform well for the crop dataset. However, the generalization of the algorithm is to be tested. For the back end, it results seem to perform good with strict pose priors. It will help if we capture more images for capture the images in a zig-zag pattern such that there is no sudden large changes in camera poses.

7 Future Work

The immediate work to be done is to fix some issues with the code and run it again for a 3d reconstruction. I hope to complete that within a week. The next step is to benchmark the reconstruction. A good benchmarking and evaluation strategy for the generated 3D maps is essential as we will use the map for the robot's next pass planning and visual odometry to resolve errors in the robot's position.

Once the spatial reconstruction is complete and tested, we will start working on the temporal reconstruction problem.

8 Meta Learning

Working on this project made me realise that the structure from motion problem can be challenging depending on the dataset. Plants in particular have highly repetitive texture are not normally found

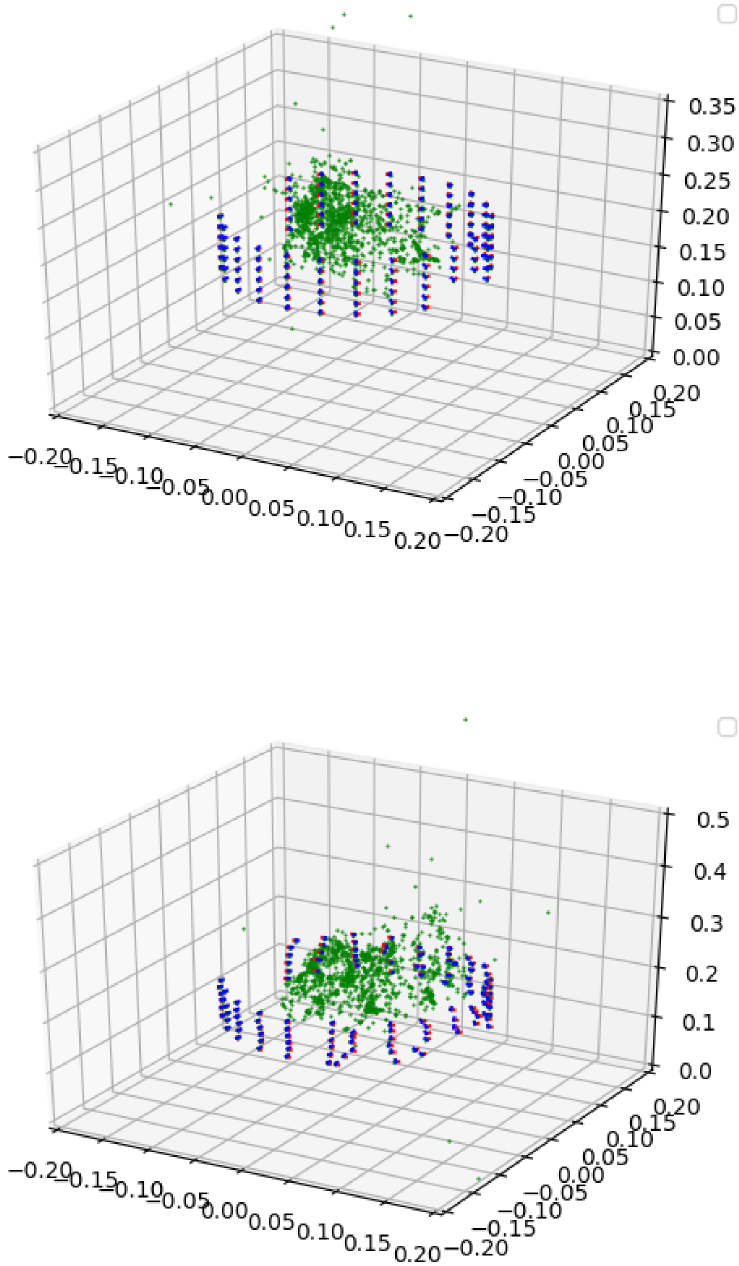

Figure 4: *Initialisation (top) and Results (bottom) of landmarks and camera poses*

in common datasets. Learning about factor graphs and its power to model almost anything was really beneficial. Working with the SLAM subgroup, I learnt about a lot of different areas related to perception. I was helped a lot by the team and could contribute to some ideas too. While the bundle adjustment equation is a simple one line equation, I understood the challenges and problems to get it to work accurately.

References

[1] J. Dong, J. G. Burnham, B. Boots, G. Rains, and F. Dellaert, "4d crop monitoring: Spatio-temporal reconstruction for agriculture," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3878–3885, IEEE, 2017.

[2] P. Roy and V. Isler, "Surveying apple orchards with a monocular vision system," in *2016 IEEE international conference on automation science and engineering (CASE)*, pp. 916–921, IEEE, 2016.

[3] Y. Jiang, C. Li, A. H. Paterson, S. Sun, R. Xu, and J. Robertson, "Quantitative analysis of cotton canopy size in field conditions using a consumer-grade rgb-d camera," *Frontiers in plant science*, vol. 8, p. 2233, 2018.

[4] M. Bietresato, G. Carabin, R. Vidoni, A. Gasparetto, and F. Mazzetto, "Evaluation of a lidar-based 3d-stereoscopic vision system for crop-monitoring applications," *Computers and Electronics in Agriculture*, vol. 124, pp. 1–13, 2016.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[6] S. Ullman, "The interpretation of structure from motion," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 203, no. 1153, pp. 405–426, 1979.

[7] F. Dellaert, M. Kaess, *et al.*, "Factor graphs for robot perception," *Foundations and Trends® in Robotics*, vol. 6, no. 1-2, pp. 1–139, 2017.

[8] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *Kdd*, vol. 96, pp. 226–231, 1996.

[9] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 6, pp. 0756–777, 2004.

[10] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.