3D Reconstruction of Plant Leaves from Rough Multi-Photos

Abstract

Three Dimensional (3D) reconstruction of leaves has gained a great momentum because of its application effects matched with that of computer game demands. Critical challenges for 3D reconstruction on leaves are the significant properties brought with leaves and rigorous requirements on photo quality. Existing 3D reconstruction methods, however, follow the traditional limits about high claim. We propose a novel reconstruction process specifically adapted to leaves that can overcome this limitation. It can achieve acceptable 3D reconstruction of leaves with multiple photos in bad conditions (e.g., phone-taken). Experimental results show promising advance over direct SIFT tested on randomly taken reddish-green Epipremnum aureum.

1. Introduction

Our reconstruction process specifically aims to deal with leaves with multiple photos in bad conditions. Detailed, we compare the feature detections among several algorithms mentioned and get advancements through filtering and lookup table. In this paper, we will first introduce the backgrounds and related works, then describe our advancements in theory followed by experiments, analysis and conclusion.

1.1. Motivation and Objective

Computer games play an increasing role in daily recreations, which brings higher real-life-similar requirements, such as model sheet, natural environment, social networking, etc. Among these natural elements, delineations about trees, flowers and grasses become incredible crucial. However, compared to other objects, their properties — reflecting surface, view angle, self-symmetry and periodicity—and the spatial distribution induce sequences of difficulties, as L. Quan et al. [1] discussed. The resulting relatively low performance overwhelms sheer 3D reconstructions. The objective of this paper is to develop a novel process that provides a promising 3D reconstruction solution to serve for plant leaves from rough multi-photos.

Two dimensional image acquisition is the information source of three dimensional reconstructions. Commonly used 3D reconstruction is based on two or more images, methods known from J. Phattaralerphong and H. Sinoquet [2], also it may only employ one single image sometimes. There are various types of methods for image acquisition that depends on the occasions and purposes of the specific application. Not only the requirements of the application must be met, but also the visual disparity, illumination, performance of camera and the feature of scenario should be considered. Camera calibration in Binocular Stereo Vision refers to the determination of the mapping relationship between the image points P1(u1,v1) and P2(u2,v2), and space coordinate P(xp,yp,zp) in the 3D scenario.

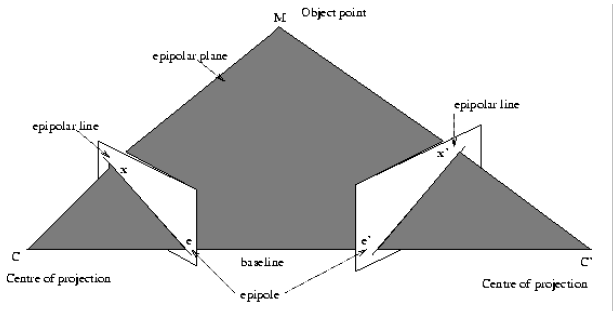


Figure 1: The epipolar constraint

The aim of feature extraction is to gain the characteristics of the images, through which the stereo correspondence processes. As a result, the characteristics of the images closely link to the choice of matching methods. There is no such universally applicable theory for features extraction, leading to a great diversity of stereo correspondence in Binocular Stereo Vision researches. Stereo correspondence is to establish the correspondence between primitive factors in images, i.e., to match P1(u1,v1) and P2(u2,v2) from two images. Certain interference factors in the scenario should be noticed, e.g. illumination, noise, surface physical characteristics and etc. According to the precise correspondence, combined with camera location parameters, 3D geometric information can be recovered without difficulties. Due to the fact that accuracy of 3D reconstruction depends on the precision of correspondence, error of camera location parameters and etc., the previous procedures must be done carefully to achieve relatively accurate 3D reconstruction.

1.2. Overview of our method

To reconstruct 3D point cloud for plant leaves with sets of frames with unknown angles, we aim at solving all the problems including extracting information from pictures, retrieving camera pose, and locating 3D points. A novel process that provides a promising 3D reconstruction solution to serve for plant leaves from rough multi-photos is then introduced in this paper.

Even after utilizing SURF feature detection, noises derive from pictures in bad conditions are still disturbing. Enhancements to the SURF feature detector were made to filter out bad matches and this included process such as KNN ratio test, symmetry test and RANSAC.

The pose estimated could not be applied to add new frames without further calibration even after optimizing with RANSAC because structure from motion is a rough method to predict camera pose. However, it is possible to predict camera matrix with existing cloud point. To find 2D-3D matching pairs, we implemented a Lookup Table which stores 2D coordinated of current frame and its triangulated 3D coordinates and kept tracking when processing each new frame. Correspondingly, we choose to further extend our solution to Iterative-LS method for triangulation.

2. Related Works

We will basically describe related works in this part. The contents include feature detection, structure from motion and PnP problem.

2.1. Feature Detection

Feature detection and matching techniques plays an important role in our application and as well as many computer vision areas. Some early developments, such as L. Wang et al. [3] introduced finding feature points that are corner-like. Interestingness could also be evaluated as high change of intensity using various sliding window functions and this is the backbone of the work of Firstner and then

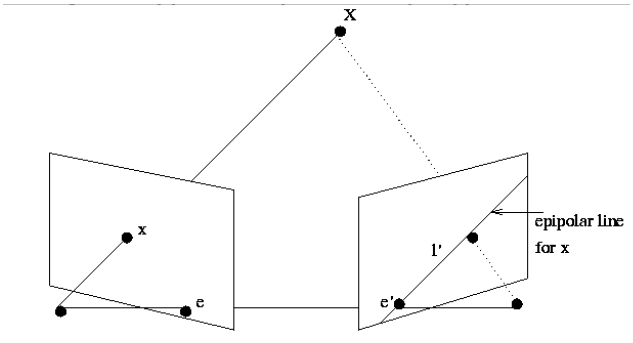


Figure 2: The epipolar line along which the corresponding point for x must lie

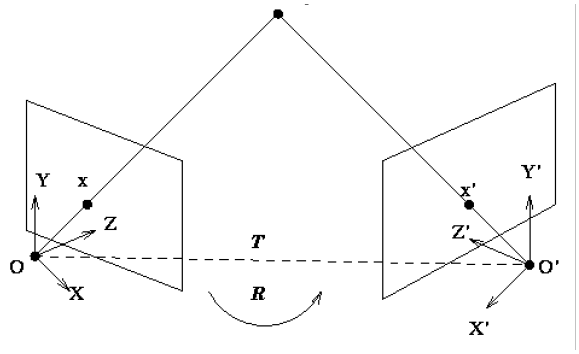


Figure 3: The Euclidean relationship between the two view-centered coordinate

Harris and Stephen. While eigenvalue based detectors like this, such as the Harris detector has rotation invariance, they are non-variant to image scale. More modern techniques such as Lowes Scale Invariant Feature Transform (SIFT) use sampling of scale space to detect points that are invariant to changes in both scale and rotation, which inspired the development of the Speeded Up Robust Feature detector and has some gains in speed and robustness. In the experiment, we also regard SURF as initial detection.

2.2. Structure from Motion

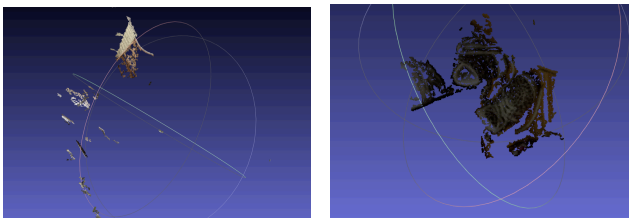
SFM refers to the process of estimating three dimensional structures from two dimensional image sequences which may be coupled with local motion signals. In order to perform reconstruction from correspondences points, there are math calculation involved, especially some matrix techniques. The math covers obtaining the Fundamental matrix from corresponding points, and extracting Camera matrices.

The location of points in multiple images can eventually lead to an estimation of its 3D location. Refer to Figure 1, multiple locations and matches call for the need for fundamental geometry to describe relationships between images. The general taxonomy of this was proposed by Scharstein and Szeliski [4], and this involves the concept of epipolar geometry. The illustration below shows the concept of epipolar geometry where a plane could be bounded at the 3D location of the point and two other locations of where this point appears on two pictures. All possible configurations for this plane will include epipolar line segments that intersect at an epipole on each image. This means knowing where the points are on an epipolar line segment means its partner is on the corresponding epipolar line segment of the second pictures. Find 7 or more good matching points will allow the estimation of the Fundamental matrix which would describe the necessary epipolar geometry.

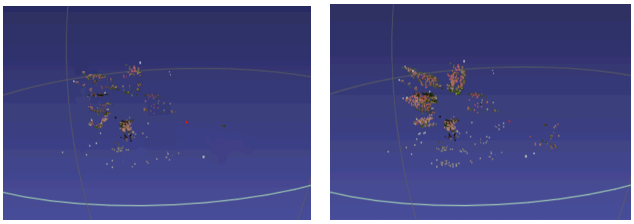
Given a single image, the three dimensional location of any visible object point must lie on the straight line that passes through the center of projection and the image of the object point see the Figure 2. The determination of the intersection of two such lines generated from two independent images is called triangulation. Clearly, the determination of the scene position of an object point through triangulation depends upon matching the image location of the object point in one image to the location of the same object point in the other image. The process of establishing such matches between points in a pair of images is called correspondence. It might seem that correspondence requires a search through the whole image, but the epipolar constraint reduces searches to a single line.

The epipolar is the point of intersection of the line joining the optical centers, that is the baseline, with the image plane. Thus the epipole is the image, in one camera, of the optical center of the other camera. And the epipolar line is the straight line of intersection of the epipolar plane with the image plane. It is the image in one camera of a ray through the optical center and image point in the other camera. All epipolar lines intersect at the epipole. Thus a point x in one image generates a line in the other on which its correspondences is thus reduced from a region to a line.

To calculate depth information from a pair of images we need to compute the epipolar geometry. In the calibrated environment we capture this geometric constraint in an algebraic representation known as the Essential matrix. In the uncalibrated environment, it is captured in the Fundamental matrix. This has fewer degreed of freedom, as now there is only rotation and translation with a scale ambiguity, and it gives rise to fewer solutions when extract



(i) SIFT (ii) SIFT with Corner Harris



(iii) SURF (iv) SURF with proposed filtering

Figure 4: **Different Feature Detection Applied** (i)SIFT (ii)SIFT with Corner Harris (iii)SURF (iv)SURF with proposed filtering

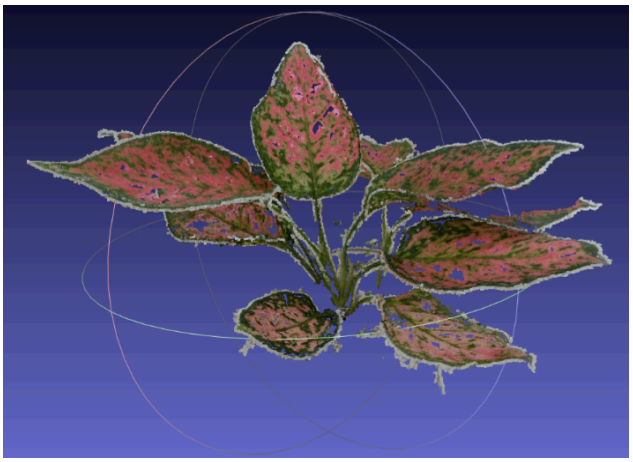


Figure 5: SURF with **proposed filtering** and **Lookup Table**

-ing camera matrices. Setting an initial origin point for the first camera matrix (P) allows the second camera matrix (P) to be computed once the Essential matrix is factored. The factored solution gives two possible solutions for rotation and a choice of sign for translation. These solutions could be interpreted geometrically: Two cameras pointing inwards towards object; Accidentally reversing the input for case 1; Two cameras pointing at an object, but facing away from each other; Accidently reversing the input for case 3.

Providing both Camera matrices and a list of calibrated 2D points from both images allow triangulation to take place to estimate 3D locations. Due to the ambiguity in the possible solutions, this may not always work. In the case of cameras that are faced away from each other, rays pointing outwards will never intersect or even come close to matching. However, picking the solution where the rays could point in the other direction, where virtually the rays are heading in the direction behind the image planes, will give a perfectly acceptable 3D location as scale is ambiguous either way when working with the Essential matrix.

3. Proposed Approach

In this section, we first describe our suitable feature detection applied to leaf-propertied objects (I. Shlyakhter et al. [5], A. Reche-Martinez et al. [6]). Consequently, we describe the filtering methods. At last, we add Lookup Table to store 2D-3D pairs for solving PnP problem.

3.1. Improved Feature Matching

One of the most widely used feature detectors is SIFT mentioned in section 2.1. It uses the maxima from a Difference-of-Gaussians (DOG) pyramid as features. The

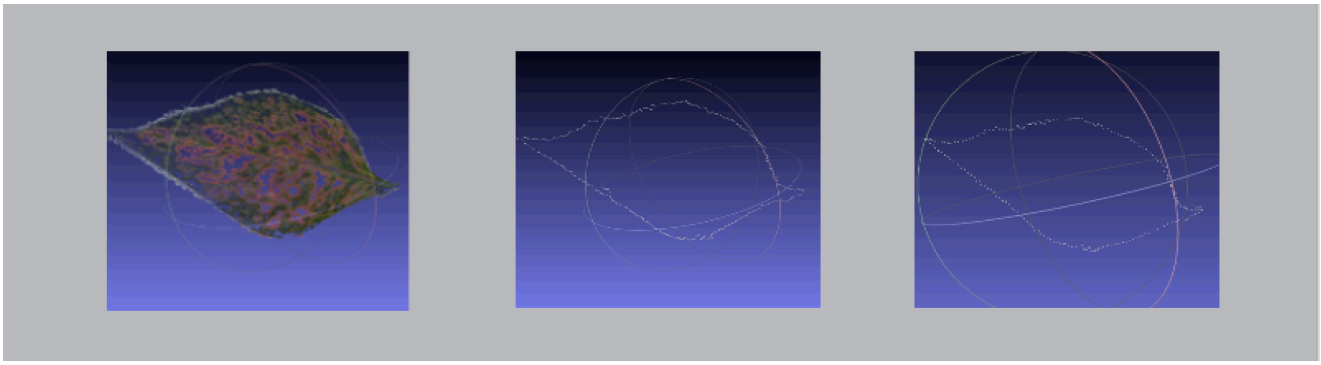


Figure 6: (i) Single leaf reconstruction (ii) **Normal** edge detection (iii) **Convex** applied **edge** detection

first step in SIFT is finding a dominant gradient direction. To make it rotation-invariant, the descriptor is rotated to fit this orientation. What's more, we consider the idea of combining the matches from SIFT and Corner Harris as feature detectors. However, both of them cannot satisfy promising results because of particular characteristics of leaves mentioned in section 1.1. More details are described by L. Juan and O. Gwun [7].

Our proposed method is based on SURF where DOG is replaced with a Hessian matrix-based blob detector. Also, instead of evaluating the gradient histograms, SURF computed for the sums of gradient components and the sums of their absolute values. Since based pictures are taken in bad conditions (normal photography), the feature detectors used provide plenty of good matches, but also had a significant amount of noises. As these points are used to generated 3D coordinated, which in turn act as references for images further along the pipe. Then robustness is significant for feature detector. Enhancements to the SURF feature detector are made to filter out bad matches and this included process such as KNN ratio test, symmetry test, and RANSAC described in detail by R. Raguram et al. [8].

A SURF feature detector is first utilized to detect interesting points in both images, and then it provides a list of Key Points for both images. Descriptors were then computed for either images or Key Points. For the ratio test, a Brute Force Matcher (i.e. mentioned as G. Bradski and A. Kaehler [9]) obtained two nearest neighbors through comparing the descriptors from both descriptor 1 to descriptor 2, and from descriptor 2 to descriptor 1. This would be used for the symmetry test later, but in the meantime, features that have an extremely distinct nearest neighbor, or where the first neighbor is much better than the second neighbor, are desired over with similar neighbors. The ratio test removes matches that are deemed bad, and the rest move through to the symmetry test. The two sets of DMatch obtained through the Brute Force Matcher that survived the ratio test are compared to see if a match from one image to another also has a pair in the other direction. Similarly, non-symmetrical matches are removed. The resulting matches are passed through a RANSAC test and these returns the Fundamental Matrix. Outliers are then filtered out. Furthermore, we assume that if results of the matches of two images are used to be good references for further images, they must have plenty of matches. We set the threshold to be 30 matches (after experiments). Otherwise, better photos have to be provided here.

3.2. Solving PnP Problem

As discussed in D. Nister et al. [10], projective ambiguity can be avoided by converting a fundamental matrix into an essential matrix. This can be done using a camera calibration matrix K, generated from a provided frame. The matrix is constructed to represent reasonable scaling and assumes the focus is at the very center of the same cameras, the conversation can be simplified to M at <double> E = K.t () * F * K. The model is checked in order to verify validity before lookup entries added to the table.

Although the Structure from Motion implementation could recover camera matrix between different frames as we know from D.J. Crandall et al. [11], the method is not applicable to other frames, which are not initial frames. As a result, using SFM only leads to the recovered point cloud shifted and rotated after each iteration (i.e. high projection error). Even though RANSAC is applied to optimize the result, the pose estimated still cannot be applied to adding new frames without further calibration.

In order to predict camera matrix P not violated to the existing point cloud, we use predicted cloud point to predict new camera pose. A set of 2D-3D matching points, Calibration Matrix and Distortion Matrix are needed to finally take Camera Matrix. Therefore, it is not reasonable to predict camera matrix with existing cloud point. To find the 2D-3D matching pair, we implement a Lookup Table which stores 2D coordinated of current frame and its triangulated 3D coordinates and keep track of 2D-3D pair during each iteration of processing new frames. In summary, when new frames come in, feature mapping id performed with the previous frame. Key points are then searched up in the Lookup Table to get known 3D

coordinated. Triangulation then generates new 3D co-ordinates where we utilize Iterative-LS instead of Linear-LS method by assigning weight in each iteration.

4. Experiments

In order to evaluate the performance of proposed 3D reconstruction feature detection and entry lookup, we build a 3D reconstruction program. Test data base on multiple pictures of reddish-green Epipremnum aureum (one type of plant leaf) in bad conditions. Our experiments compare original techs and proposed filtering in three main changes: Structure from Motion (SFM and look-up); Improved Feature Matching (SIFT, SIFT and Corner Harris, SURF, filtering); Triangulation (Linear-LS and Iterative-LS), and extract the contour by specifically analyzing a single reconstructed leaf.

4.1. Results

We directly evaluate the original feature matchings in 3D reconstruction for plant leaves, including SIFT, SIFT with Corner Harris and SURF. As Figure 4 shows, the leaf properties indeed count a lot, and even the result effect cannot reflect itself as leaf standard. Furthermore, by contrasting Fig.4(i) and Fig.4(ii) we know that even if more feature points are introduced by two detectors of SIFT and Corner Harris, not dealing well with correspondences can resulting dramatic noises. However, we also find that some advances added to SURF may achieve promising consequences. Till now, we implement pure SFM after feature detection for estimating three dimensional structures from two dimensional image sequences which are coupled with local motion signals, and we merge the utilization of 2D-3D pair Lookup Table as in Figure 5. The improvement effect is apparent as shown in Fig.4 and Fig.6. The proposed process can suitably deal with the plant leaf reconstruction where the reflecting surface, view angle, self-symmetry and periodicity are disturbing.

Then we extract one single leaf, and try to get the contour of this object simply by taking into convex. The contour cannot be directly extracted in this reconstruction method; therefore, we additionally provide such an application discussion (i.e. edge detection). When we extract a single leaf, we can observe 79766 3D points, as shown in Figure 6 above, after normal edge detection 315 points left. At the same time, noises can still occur, which means several points in a limited area may all be regarded as contour points. Then we have to apply convex search and finally 100 contour points are conserved.

4.2. Overhead Analysis

The results reconstructed this way can achieve promising effects, however, the overhead instead is somewhat costing. Because the timing penalty is incurred for which three filtering measures and large Lookup Table account. Compared to regular reconstruction of ideal objects with multiple images whether or not in excellent conditions, after features detected, the filtering and the lookup entries actually take some time. As described in section 3.1 in detail, we have known good neighbors are extracted, also taken to neighbor direction checking and symmetry test each frame. Therefore, each time when we run our reconstruction program, time penalty cannot be neglected.

4.3. Conclusion

The 3D reconstruction of leaves remains raising significant concentrations especially in computer game applications. We propose filtering steps in the reconstruction which can present promising advancements for plant leaves compared to normal methods. The experimental results show high improvements tested on randomly taken reddish-green Epipremnum aureum.

References

[1] L. Quan, P. Tan, G. Zeng, L. Yuan, J.D. Wang and S.B. Kang, “Image-based plant modeling,” ACM Transl J. NY USA, vol. 25, Issue 3, pp. 599-604, July 2006.

[2] J. Phattaralerphong and H.Sinoquet, “A method for 3D reconstruction of tree crown volume from photographs: assessment with 3D-digitized plants,” Oxford J. OX1 2JD UK, doi 10.1093, 2005.

[3] L. Wang, W. Wang, J. Dorsey, X. Yang, B. Guo and H.-Y Shum, “Real-time rendering of plant leaves,” SIGGRAPH LA USA, vol. 24, Issue 3, pp 712-719, 2005

[4] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms,” IJCV, vol. 47, Issue 1, pp. 7-42, April 2002..

[5] I. Shlyakhter, M. Rozenoer, J. Dorsey and S. Teller, “Reconstructing 3d tree models from instrumented photographs,” IEEE Computer Graphics and Application, vol. 21, Issue 3, pp. 53-61, May/June 2001.

[6] A. Reche-Martinez, L. Martin and G. Drettakis, “Volumetric reconstruction and interactive rendering of trees from photographs,” SIGGRAPH, vol. 23, Issue 3, pp. 720-727, August 2014.

[7] L. Juan and O. Gwun, “A comparison of SIFT, PCA-SIFT and SURF,” IJIP, vol. 3, Issue 4, pp. 143-152, 2009

[8] R. Raguram, J.M. Frahm and M. Pollefeys, “A comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus,” ECCV, vol. 5303, pp. 500-513, 2008

[9] G. Bradski and A. Kaehler, Learning OpenCV: Computer vision with the OpenCV library, vol. 2. O’Reilly Media, Inc., 2008, pp.405-458.

[10] D. Nister, H. Stewenius and E. Grossmann, “Non-parametric self-calibration,” ICCV, doi: 10.1109/ICCV.2005.170, Oct. 2005.

[11] D.J. Crandall, A. Owens, N. Snavely and D.P. Huttenlocher, “sfm with MRFs: Discrete-Continuous Optimization for Large-Scale Structure from Motion,” IEEE Transl J. NY USA, vol. 35, Issue 12, pp. 2841-2853, December 2013.