3D MODELLING FROM UN CALIBRATED IMAGES – A COMPARATIVE STUDY

Limi V L
Computer Vision And Image Processing
Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham,
Coimbatore, Tamil Nadu – 641 112, India
limijobs@gmail.com
http://www.amrita.edu

D Vekataraman
Computer Vision And Image Processing
Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham,
Coimbatore, Tamil Nadu – 641 112, India,
d_venkat@cb.amrita.edu
http://www.amrita.edu

Abstract

3D modeling is a demanding area of research. Creating a 3D world from sequence of images captured using different mobile cameras pose additional challenge in this field. We plan to explore this area of computer vision to model a 3D world of Indian heritage sites for virtual tourism. In this paper a comparative study of the existing methods used for 3D reconstruction of un-calibrated image sequences was done. The study shows different scenario of modeling 3D objects from un-calibrated images which include community photo collection, images taken from unknown camera, 3D modeling using two un-calibrated images, etc. Hence the different methods available were studied and an overall view of the techniques used in each step of 3D reconstruction was explored. The merits and demerits of each method were also compared.

Keywords**:** 3D reconstruction, un-calibrated images, Scale Invariant Feature Transform, Structure from Motion

1. Introduction

3D reconstruction is one of the challenging areas of research in computer vision which is still in its early stages. The reconstruction becomes more complicated when dealing with photos for which the intrinsic and extrinsic parameters of the camera are unknown such as un-calibrated images. These images may be of the same scene or objects taken from different cameras or using same camera at different time and viewpoints. In 2D images depth information is lost but to model a 3D scene the depth information is needed. It is difficult to estimate the depth of a scene from single image. Multiple images can be used to overcome this problem. A simple triangulation requires at least two images for depth estimation. But it is estimated that only two images may not produce accurate results as information may be insufficient. So a sequence of images may be preferred.

A 3D model can be constructed from calibrated images as well as un-calibrated images. In calibrated images the details of the camera's intrinsic and extrinsic parameter are known in advance. The advantage of using calibrated images is that it produces accurate results. But it requires a lot of complex and time consuming process to accurately calibrate a camera. In case of un-calibrated images the camera parameters are unknown so the tedious task of calibration can be avoided but it comes with the difficulty that the 3D model so generated may not be accurate. Therefore the use of calibrated or un-calibrated images comes with a tradeoff between accuracy and computational time. Taking this into account the images to be used will greatly depend on the type of application for which the 3D reconstruction is to be performed. As we intend to create 3D model of heritage sites for virtual tourism the sources from which images are acquired will differ greatly. Some heritage sites may be nonexistent today due to dilapidation but the old photographs depicting its original structure may be available in archives, which may be used to recreate it virtually. In such cases the images of the structure would be un-calibrated as the details of the camera parameter are unknown. Finding a technique that suits best to such problem is challenging. Taking this into account a study was done on the techniques available for 3D reconstruction using un-calibrated images.

3D reconstruction in general involves the procedure in which initially feature points are estimated. This may be done using Harris corner detection, Scale Invariant Feature Transform (SIFT), etc. These feature points are matched and the outliers are removed. Finally depth of the matched points is calculated and the 3D model reconstructed. Again 3D model may be points cloud, sparse map, dense map, etc. In this paper we describe the various methods that currently exist in modeling 3D from un-calibrated images.

2. Comparative Study

2.1. *3D reconstruction from un-calibrated image sequence collected from community photo collection*

In the paper [Agarwal et al, (2010)] a method has been proposed to leverage the photos uploaded in community photo collection websites such as Flickr.com to model a 3D world. Method such as structure from motion (SFM) has been used to estimate the viewpoint of camera and the 3D scene structure from a set of 2D photos. The common features of a scene were extracted using SIFT to solve the correspondence problem. The outliers in the feature points were removed using Random Sample Consensus (RANSAC). After extracting common features they were tracked across images and then SFM technique were applied on the tracked images to create a sparse 3D model. The dense 3D model was reconstructed from registered photos using the multi view stereo algorithm.

The advantage of the method proposed in this paper is that it makes it possible to reconstruct 3D model from massive collection of un-calibrated photos. Although the paper has some challenges such as complete model of building is hard to construct as the system fails in areas where the image has specular reflections or limited texture, but on the whole it has addressed the problem 3D reconstruction from community photo collection to a great extent.

The paper [Raguram et al, (2011)] proposes a method to model landmarks based on large scale images collected from internet that are highly contaminated. They have used the 2D appearance and 3D geometric constraints used to extract summaries and then construct the 3D models. Initially the input images are clustered based on global appearance descriptors, and then they were refined using 3D geometric constraints. An iconic scene graph was used to represent a cluster of single iconic view. The iconic images were registered using structure from motion and the 3D models were generated. To increase the coverage of scene additional non iconic views were included for 3D modeling. The paper has been implemented in five steps.

First the iconic scene views are clustered using k-Means clustering, which were run on global gist descriptor. In the second step the geometric verification and iconic image selection were performed. For this epipolar geometry of pairs of images from the cluster were estimated and the clusters with inconsistent geometry were rejected. In the third step re-clustering and registration of rejected clusters were performed due to which small iconic scene clusters were obtained. In the fourth step the iconic scene graph was created. In the graph the nodes corresponded to iconic images and edge corresponded to two-view transformation between the iconic images. In the final step sparse 3D models were generated, the reconstructions were performed in an incremental fashion.

The advantages of the system are that robust and efficient 3D models were generated. The disadvantage of the system is that, since iconic views were taken into consideration the merging of non iconic views generated problems. A challenge described in the system is illumination changes.

2.2. *3D modeling of un-calibrated image sequence using projective transformation*

The paper [Liu et al, (2010)] proposes a practical method of 3D reconstruction based on un- calibrated image sequence. Projective reconstruction by linear transformation has been used in the method. Initially the corresponding points between two neighboring images were obtained by the feature extraction technique SIFT. Then RANSAC approach was used to eliminate false correspondence points. The SIFT technique extracted the key points in all images and then a matching was performed between two images. RANSAC was used to eliminate the outliers, which helped to compute exact fundamental and camera matrix.

First the projective reconstructions between consecutive views were taken to obtain initial projective view, and then it combined the neighboring projective reconstruction with common 3D points. The common 3D points were the same point in the Euclidean space that has different forms in different projective spaces. Then from the common view, matches were found using search algorithm. After finding the match, transformation matrix was found and the projective transformation merged for reconstruction.

2.3. *3D reconstruction from multiple un-calibrated images using close range photogrammetric*

In the paper [Zakaria and Said, (2010)] a workflow of 3D reconstruction from multiple view images has been proposed. A set of un-calibrated images were taken as input and the depth map constructed. The workflow has been divided into three major parts as initially finding the correspondence points between images, then depth estimation and finally the 3D model generation. Initially two images were taken and the correspondence points were estimated using epipolar geometry and along with it the depths were estimated. To estimate the depth the pixels between pairs of images are matched and graph cut method with dynamic programming was used.

The advantage of this method is that from a chain of un-calibrated images the 3D view can be reconstructed with the only disadvantage that the 3D so constructed is not very accurate as the solution of 3D reconstruction is defined up to a projective transformation.

2.4. *Relative 3D reconstruction from point correspondences of multiple un-calibrated images using reference points*

In the paper [Mohr et al, (1995)] a method for relative 3D reconstruction from point correspondence of multiple un-calibrated images using reference point is proposed. In the paper the authors have assumed the imaging system to be a perfect perspective projection, which suggests that the camera is perfect pinhole. The paper has been implemented in two major steps. First relative reconstruction is estimated as a least squares estimation method for which epipolar geometry is not required. Levenberg-Marquardt algorithm is used which works well for noisy images. Secondly a geometric method is proposed to identify the reference points. Through epipolar geometry coplanar points were checked. The fundamental matrices were estimated by non linear optimization method and corner using canny like contour extractor with B-Splines fitted using least square approximation. Maximum curvature points on the B-Splines were taken as corners.

The assumptions made in this paper were that the noises in the images were uncorrelated, Gaussian and centered. The advantage of the method proposed in this paper is that it is less sensitive to bad motion of the camera than algebraic method.

3. Conclusion

In this study we have found that there are various methods available currently that help in the generation of 3D model from a sequence of un-calibrated images depending on the scale of images available. It has been noticed that while dealing with un-calibrated images there is no limitation on the number of views required but we may have to compromise on the quality of the 3D model generated. Also it was observed that when dealing with community photo collection, the feature extractor commonly used in the Scale Invariant Feature Transform (SIFT) that is to detect corresponding feature in different views that has huge viewpoint variation.

Based on the study made we are proposing a new method to implement the 3D modeling of a sequence of images that would help in developing archaeological and heritage sites, promote tourism by creating virtual tourist sites. In our method a sequence of images collected from mobile camera is taken as input. Before using any feature extractor method the image are pre-processed individually depending on the type of constraint such as illumination and noises. After the pre-processing stage the feature points extraction performed using Harris Corner detector. Next the featured points between pairs of images may be mapped the depth is estimated and a sparse 3D map may be generated from these mapped set of images using structure from motion technique.

References

[1] Agarawal Sameer, et al (2010): Reconstructing Rome, Computer Society IEEE, 2010.
[2] Goesele Michael, et al (2010): Scene Reconstruction From Community Photo Collections, Computer Society IEEE
[3] Liu Dawei, et al (2010): A Practical Method for 3D Reconstruction with Un-calibrated Image Sequences, IEEE
[4] Lowe G David (2004): Distinctive Image Features from Scale Invariant Keypoints, International Journal For Computer Vision
[5] Mohr R., Quan L., Veillon F (1995): Relative 3D Reconstruction Using Multiple Un-calibrated Images, International Journal For Robotic Research, Vol.14, No. 6, PP.619-632
[6] Raguram Rahul, et al (2011): Modeling and Recognition of Landmark Image Collections Using Iconic Scene Graphs, International Journal for Computer Vision
[7] Snavely Noah, Seitz Steven M, Szeliski(2008): Modeling The World from Internet Photo Collections, International Journal For Computer Vision
[8] Snavely Noah, et al (2010): Scene Reconstruction And Visualization From Community Photo Collections, IEEE
[9] Zakaria Abdoulaye Mahamat, Said Abas Md (2010): 3D Reconstruction Of a Scene from Multiple Un-calibrated Images Using Close Range Photogrammetry, IEEE