Spring 4-2004

Uncalibrated and Self-calibrated Cameras

Avital Braiman
University of Tennessee - Knoxville

Appendix E - UNIVERSITY HONORS PROGRAM
SENIOR PROJECT - APPROVAL

Name: Avital Braman

College: electrical engineering Department: ECE

Faculty Mentor: Dr. Abidi

PROJECT TITLE: uncalibrated and self-calibrated cameras

I have reviewed this completed senior honors thesis with this student and certify that it is a project commensurate with honors level undergraduate research in this field.

Signed: M. A. Abidi Faculty Mentor

Date: 4/28/2004

General Assessment - please provide a short paragraph that highlights the most significant features of the project.

Comments (Optional):

Survey on Reconstruction Methods Using Self-Calibrated and Totally Uncalibrated Cameras

Avital Braiman

ECE 491

Fall 2003

Abstract

The aim of the project is to provide a comprehensive survey on the issue of the 3D scene reconstruction from 2D self-calibrated and totally uncalibrated camera views. Image reconstruction is a very important area in the field of computer vision, robotics, and image processing,

Several approaches has been suggested and tested. This project will survey reconstruction methods for multiple motion scenes containing multiple objects from uncalibrated views. In these cases, neither camera motion, nor the camera settings have to be known. The obtained 3D model is a scaled version of the original object, and the surface texture is obtained from the image sequence as well.

Table of Contents

1. Introduction

There has been considerable interest in recent years in the generation of computer vision algorithms able to operate with uncalibrated cameras. One challenge has been to reconstruct a scene, up to scale, from images obtained by cameras whose internal geometry is not fully known and whose relative orientation is unknown. Remarkably, such a reconstruction is sometimes achievable solely by consideration of corresponding points (that illustrate a common scene point) identified within the images. A key process involved here is that of self-calibration, whereby the unknown relative orientation and intrinsic parameters of the cameras are automatically determined.

Originally, people determined the calibration parameters of a camera by hand. This was done using a calibration object of known metric structure. However, this technique of camera calibration required very detailed scene information that may not be available. Later, researchers discovered a method that made the use of a calibration object unnecessary. When the camera itself was not available, this method also allowed cameras to be calibrated using image sequences. The Researcher that made this discovery was Faugerus at al. He showed that it was possible to calibrate a camera automatically from image data using only scene rigidity constrains. The method he came up with has become known as "self calibration." The topic of self-calibration was researched extendedly using optical algorithms derived for most possible scene types and camera motions.

Hartley researched a stratified approach to calibration. In his approach there is an intermediate affine calibration stage. Other researchers adopted this idea and later Pollefeys et al. developed a method to take into account varying camera parameters. Being able to deal with varying camera parameters is important because camera parameters will vary during a sequence of images if the sequences are taken from different cameras or if zooming position of the same camera has changed. Hartley also showed that self-calibration will be possible for non-translating cameras as well.

If the intrinsic parameters of the camera are known, the extrinsic parameters can be obtained up to a scale factor. After the extrinsic parameters have been found, the 3D locations of the features points can be estimated. However, when the camera is unknown, the intrinsic parameters of the camera are hidden too. Such cameras are called totally uncalibrated. With a totally uncalibrated camera alone, a scene can be reconstructed up to projective ambiguity. Self-calibrated cameras provide metric reconstruction.

Once self-calibration became possible, attempts were made to generate structures in real time. Structure from motion includes the extraction of image features, self-calibration, and then the use of the calibration to retrieve the feature in the scene. The work of Pollefeys et al. resulted in the development of complete systems for structure computation. Applications to this can be seen film and television special effect industries.

1.1 Motivation

The topic of 3D scene reconstruction from 2D camera views is very important for variety of applications including robot navigation, computer vision, image processing, tracking surveillance, and space exploration.

Robotics is an idea that developed quite early in the explosion of technological advancement. If intelligent mechanization could do so many of the tasks that humans previously did, it is feasible that a single machine might one day be able to do everything a human can. The term `mobile robot' refers to a robot capable of true locomotion. So many of those tasks longing to be automated involve getting from one place to another, a problem which has proved to be an important topic for research and development. For at least twenty years industry and researchers have been looking at this issue of automated `routing', yet the presence of mobile robots in industrial or commercial applications is extremely limited.

For NASA space exploration programs it is important to develop machine vision algorithms that enable autonomous exploration and sample return from small bodies through onboard visual feature tracking and landmark recognition. These algorithms will provide estimates of spacecraft motion and position used to guide the spacecraft during autonomous landing and exploration. They will also enable hazard detection by providing estimates of 3-D surface landscape through processing of monocular image streams. Due to the small size, irregular shape and variable surface properties of small bodies, accurate position estimation and hazard avoidance are needed for safe and precise small body landing and sample return. Because of the communication delay induced by the large distances between the earth and targeted small bodies, landing on small bodies must be done autonomously using on-board sensors and algorithms. Current navigation technology does not provide the precision necessary to accurately land on small bodies, so other positioning techniques must be investigated. Optical sensors combined with autonomous machine vision algorithms offer a solution to the precise positioning problem; images can be automatically analyzed to determine the position of a spacecraft with respect to a related body. Since current camera resolutions enable visual positioning with errors on order of centimeters from a range of hundreds of meters, visual position estimation is accurate enough for small body landing.

2. Methods

In the field of scene reconstruction the usage of self-calibrated and totally uncalibrated cameras are important research areas. This paper will review these important research areas and provide a comprehensive approach for each area.

2.1
3D Reconstruction from Sequence of Images Using Self-Calibrated Cameras

In Pollefeys et al. [14], [16], and [17] a system that retrieves a 3D surface model from a sequence of images taken with off-the-shelf consumer cameras is presented. The images are acquired by the user, by freely moving the camera around the object. Neither the camera motion nor the camera settings have to be known. The acquired 3D model is a scaled version of the original object i.e., a *metric* reconstruction. A system in [14], [16], and [17] uses full perspective cameras and does not require prior models or calibration. When compared to existing systems such as Photo-Modeler 2000, this approach is better because it is fully automatic. This comprehensive system combines algorithms of different areas of computer vision, which include: projective reconstruction, self-calibration and dense depth estimation.

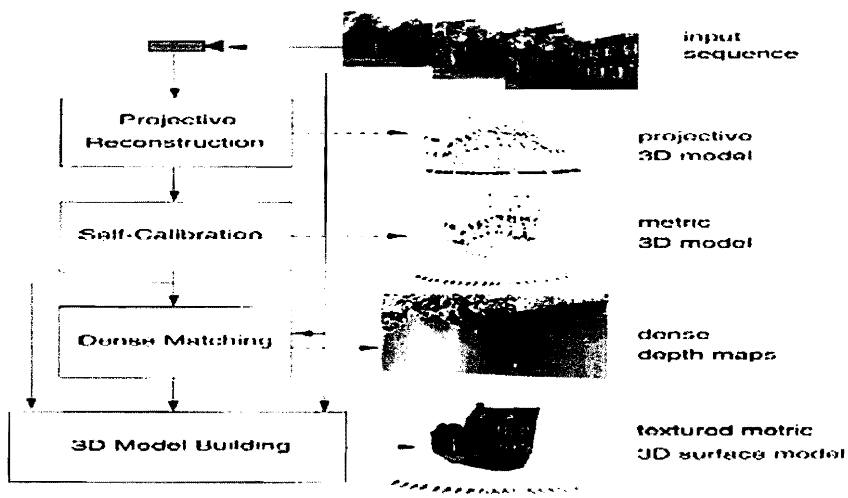

 Figure 1. An outline of the system. The little triangular pyramids represent the camera [14]

In Fig. 1, an outline of the system is given. It is made up of independent components, which pass on the necessary information to the next component. The first component calculates the projective calibration of the sequence together with a sparse reconstruction. The next part computes the metric calibration from the projective camera matrices through self-calibration. In the next part, dense correspondence maps are predicted. Finally, the 3D model is built. All results are combined in a 3D surface reconstruction of the scene [14].

First, the relative motion between consecutive images needs to be recovered. This process is related to finding corresponding image features between these images. The next step is made up of retrieving the motion and calibration of the camera and the 3D structure of the features. This is done in two stages. At first the reconstruction contains a projective skew such as parallel lines are not parallel, angles are not correct, relative distances are not conserved. This is caused because there was no prior calibration. Using a self-calibration algorithm, Pollefeys et al. [16], this distortion can be removed, resulting in a reconstruction comparable to the original scene. Since the focal length and other intrinsic camera parameters do not have to be measured and calibrated in advance and can change during the acquirement, this uncalibrated approach to 3D reconstruction allows much more flexibility in the attainment process. Therefore, the next step consists of an effort to match all image pixels of an image with pixels in neighboring images. This is done so that these points too can be reconstructed. Knowing all the camera parameters acquired in the previous stage assists this task.

It has been shown by Faugeras et al. [3] and Hartley et al.[11] that a reconstruction up to an arbitrary projective transformation was possible from an uncalibrated image sequence. Since then, many researchers tried getting accurate estimates of the projective calibration of an image sequence. Algorithms were suggested to estimate the fundamental matrix from image pairs Torr and Zhang et al. Later, algorithms that sequentially recover the projective calibration of a complete image sequence were developed Beardsley et al [1].

Unfortunately, a projective calibration is not satisfactory for most applications. Therefore, researchers attempted to automatically upgrade projective calibrations to metric i.e., Euclidean up to scale. This is based on some constraints on the camera's intrinsic parameters. This method is called self-calibration. Usually, it is assumed that the same camera is used throughout the sequence and that the intrinsic camera parameters are constant. One of the main problems with self-calibration is that critical motion sequences exist for which self-calibration does not result in a unique solution. Pollefeys et al. [16] proposed a more practical approach that assumes that some parameters are approximately known but which allows others to vary. Therefore, this approach can deal with zooming focusing cameras.

Once the calibration of the image sequence has been estimated, the depth can be estimated. This is done using stereoscopic depth estimation. The challenging part in stereoscopic depth estimation is to find dense correspondence maps between the images. The correspondence problem is solved by using constraints derived from the calibration and from some assumptions about the scene.

Figure 2. Images of the Arenberg castle sequence. This sequence is used to shows the different steps of the reconstruction system [14].

Pollefeys et al. [14] shows the different steps of the method in detail. An image sequence of the Arenberg castle in Leuven is used to show the different steps of the reconstruction method. Some images of this sequence can be seen in Fig. 2. The full sequence consists of 24 images.

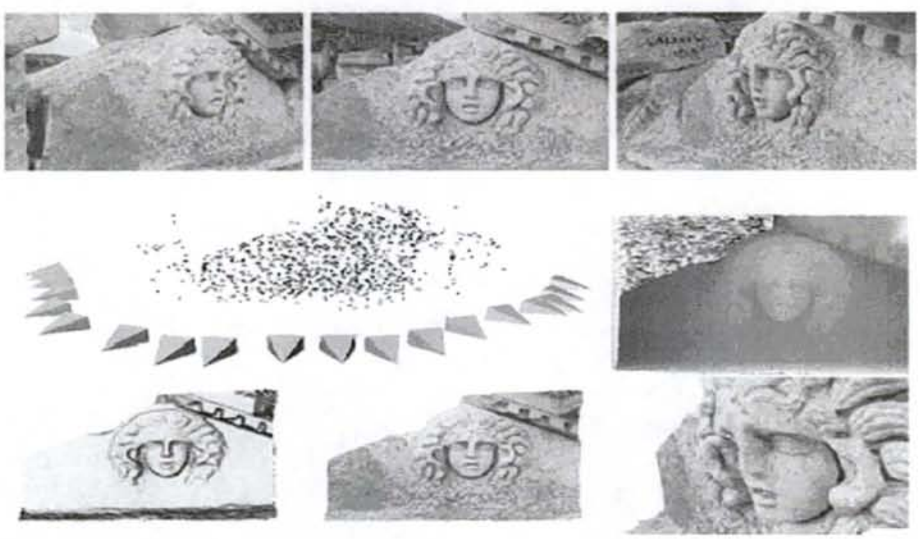

Figure 3 shows a 3D model of a decorative Medusa head discovered at the ancient site of Sagalassos in Turkey. The top three views are of the original video. The middle views are the reconstruction of the 3D feature points with computed camera motion done for the key frames and one of the depth images. At the bottom the shaded and textured views of the recovered 3D model are shown [17].

Pollefeys et al. [17] has shows a video sequence from a Medusa head decorating an ancient fountain in Sagalassos (an ancient city in Turkey). Each 20^{th} frame was used as a key-frame by the video-to-3D processing pipeline. Three frames are seen in the top part of Figure 3. The computed structure and motion are also seen in this figure in the middle left. Small pyramids represent the camera viewpoints. The depth map used to construct the 3D model is shown at the middle right of the figure. The final model is seen at the bottom of the figure.

Figure 4. Three images of a Jain temple in Ranakpur [16].

Figure 5. A viewpoint of the reconstruction [16].

These images were taken during a trip in India [16]. A sequence of 11 images was taken of some details of one of the smaller Jain temples at Ranakpur, India. These images were taken with a standard camera and scanned in. Three of them can be seen in Fig. 4. The reconstruction that was obtained from this sequence is shown in figure 5.

2.2
Multiple Motion Scene Reconstruction Using Totally Uncalibrated Cameras

When nothing is known about the camera intrinsic parameters, the extrinsic parameters, or the object, it is only possible to compute a reconstruction up to an unknown projective transformation. In order to obtain a Euclidean reconstruction from the projective reconstruction some supplementary information about either the camera or the object is needed Han and Kanade [4]. Hartley retrieved the Eulidean shape applying the technique of global optimization and presuming that the camera intrinsic parameters are constant [9]. Heyden and Astrom applied an adjustment algorithm to estimate principal points, the focal lengths, the camera motion, and the object shape [13].

Han and Kanade [4], [5], and [6] propose factorization-based method for multiple motion scene reconstruction from uncalibrated views. Their method reconstructs the scene structure, the camera motion, the trajectories of the moving objects, and the camera focal lengths simultaneously. The number of the moving objects is detected automatically without prior motion segmentation. The assumption in the manuscript Han and Kanade [5] is that the objects are moving linearly with constant speeds. Han and Kanade [4] first used an iterative algorithm to get a projective reconstruction and then proposed three normalization algorithms to enforce metric constrains on the projective reconstruction. The normalization algorithms recover the unknown intrinsic parameters and translate the projective solution to a Euclidean one simultaneously. The first of the three algorithms concentrates on the case that the focal lengths are the only unknown parameters. The second algorithm concentrates on the case that the focal lengths and the principal point are unknown, while the principal point is fixed. These two algorithms mentioned above are linear. The third algorithm is bilinear and is used in the case that the focal lengths, the principal points, and the aspect ratios are all unknown. Applications are used in building modeling and terrain recovery.

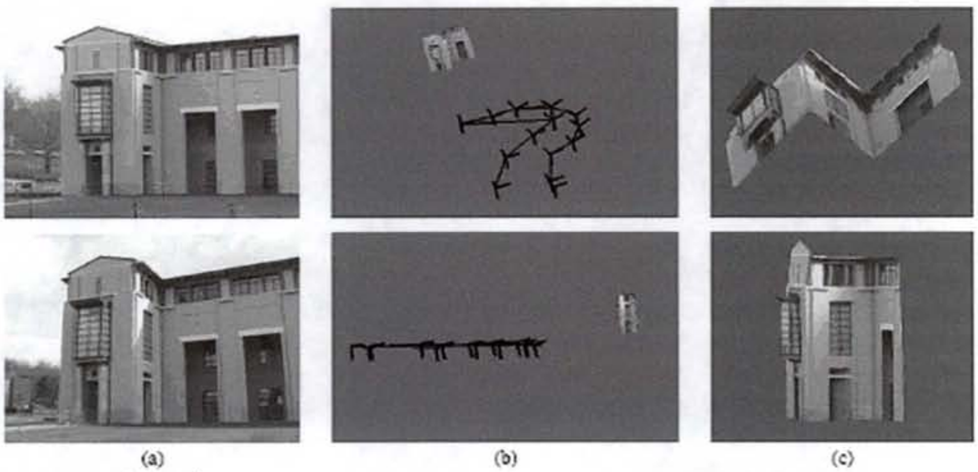

Figure 6. (a) 1st and 9th images of the building sequence. (b) Top and side view of the reconstruction, 3-axis figures show the retrieved cameras. (c) Side and bottom view of reconstructed building includes texture mapping [4].

Han and Kanade [4] show a sequence that includes 14 frames. Two frames are shown in Figure 6(a). 50 feature points were selected manually along the building windows and the corners. The focal lengths are assumed unknown while the principal points are given, and the aspect ratios are 1. Figure 6(b) shows the camera trajectories and a reconstructed building model. The top view shows that the retrieved camera moves toward the building and then away again from the building. The retrieved camera positions that can be seen in the side view reveal that all the cameras have the same height and tilt upward a little bit. Figure 6(c) demonstrates the reconstructed building with texture mapping. To qualify the results, the orthogonality and parallelism of the lines composed of the retrieved feature points are measured.

Figure 7 is an in-flight image sequence taken from an airplane flying over the Grand Canyon. The plane changes its altitude and it view angles during the sequence. The sequence is made up of 97 images, and 86 feature points were tracked through the sequence. Two frames from the sequence are shown in figure 7(a). The focal lengths and the principal point are assumed unknown, and the principal point is assumed fixed over the sequence. The normalization algorithm is used here. Figures 7(b) and (c) demonstrate the reconstructed camera trajectories and landscape map.

In Figure 8 the shapes of the books, the box, the starting positions of the toys, and the motion velocities are retrieved and shown in Figure 8a. The motion trajectories are superimposed in the images. It can be seen that there still exist projective distortions in the reconstruction. This can be seen in the top view of the box. Figure 8b shows the recovered camera position and direction with the scene reconstruction.

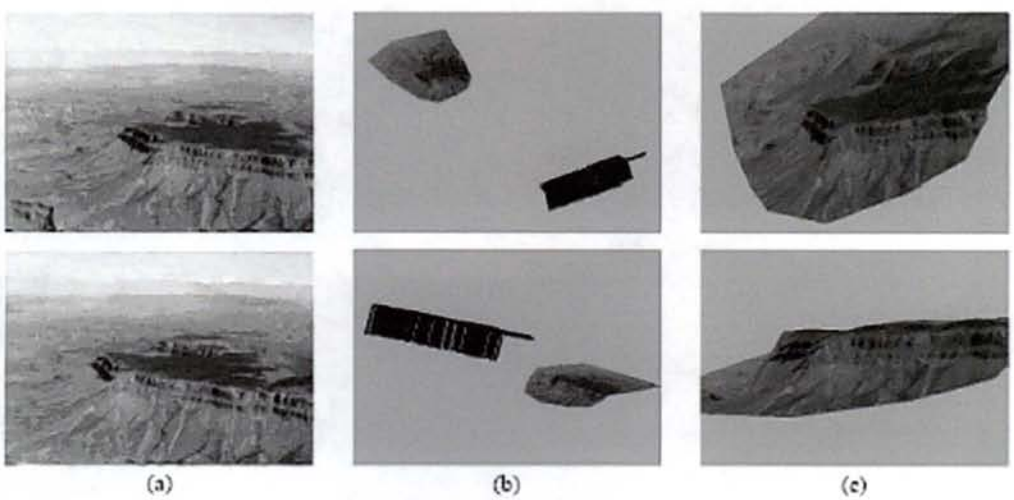

Figure7. (a) 1st and 91st images of the Grand Canyon sequence. (b) Side and top view of the reconstruction, 3-axis figures is the retrieved cameras. (c) Reconstructed Grand Canyon (top and side views) with texture mapping [4].

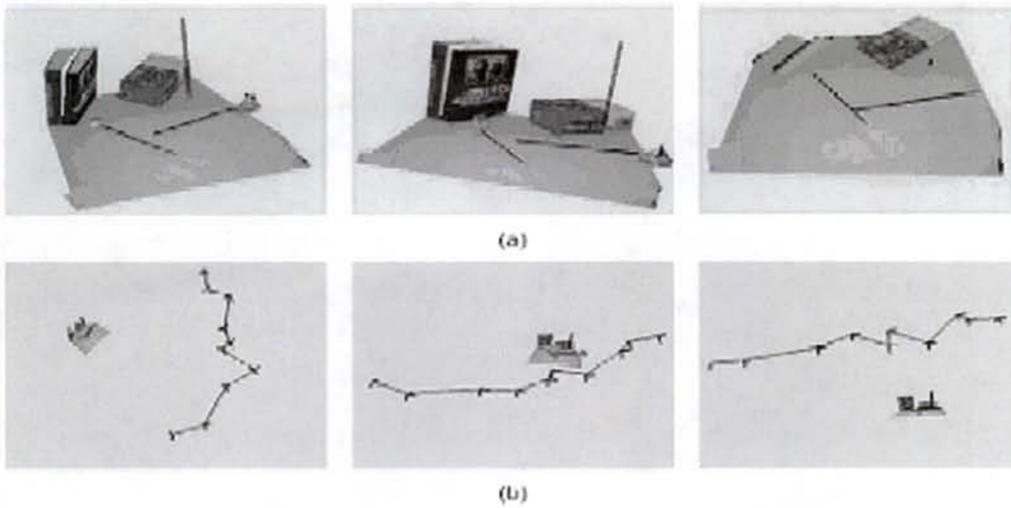

Fig. 8. (a) Three different views of the scene reconstruction with texture mapping. The black lines represent the retrieved motion trajectories. (b) Three different views of the scene reconstruction and the camera placement and orientation. The 3-axis figures represent the retrieved cameras [5].

8. Conclusion

In this review, a comprehensive survey was presented on the issue of the 3D scene reconstruction from 2D self-calibrated and totally uncalibrated camera views. Several approaches has been suggested and tested.

This paper presented an overview of the research on self-calibration and metric reconstruction in the presence of varying and unknown intrinsic camera parameters. It was shown that self-calibration is achievable using only the most general constraint such as image rows and columns are orthogonal. Of course, if more constraints are accessible, this will in general yield better results. An automatic 3D scene modeling method, which is capable of building models from uncalibrated image sequences, is presented and discussed. The method is able to extract detailed metric 3D models without prior knowledge about the scene or the camera. The advantages are numerous: the on-site acquisition time is constrained, the creation of the models is automatic and the generated models are realistic.

For image sequences taken with totally uncalibrated cameras, this paper discussed methods to create 3D models of the scene and to retrieve the extrinsic and intrinsic parameters of the cameras at the same time. This paper also presented a method to reconstruct a scene containing multiple moving objects from totally uncalibrated views. The method is an optimization process to estimate the scene structure and the camera calibration jointly when image correspondences are available. The method has partial knowledge of the scene do to the fact that the points are moving with constant velocities and some intrinsic parameters of the camera are known.

References

[1] P. Beardsley, and A. Zisserman, D. Murray, "Sequential Updating of Projective and Affine Structure from Motion." *Int. Journal of Computer Vision*, vol. 23, no. 3, pp. 235-259, 1997.

[2] J.P. Costeira, and T. Kanade, "A Multibody Factorization Method for Independently Moving-Objects," *Int'l J. Computer Vision*, vol. 29, no. 3, pp. 159-179, 1998.

[3] O. Faugeras, "What Can Be Seen in Three Dimensions with an Uncalibrated Stereo Rig." *Computer Vision ECCV'92. Lecture Notes Computer Science*, vol. 588, Springer-Verlag, Berlin, pp. 563-578, 1992.

[4] M. Han, and T. Kanade, "Creating 3D Models with Uncalibrated Cameras," *Proc. Workshop Applications of Computer Vision*, pp. 178- 185, 2000.

[5] M. Han, and T. Kanade, "Multiple Motion Scene Reconstruction with Uncalibrated Cameras," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol 25, no.7, July. 2003.

[6] M. Han, and T. Kanade, "Multiple Motion Scene Reconstruction from Uncalibrated Views," *Proc. Int'l Conf. Computer Vision*, 2001.

[7] M. Han, and T. Kanade, "Reconstruction of a Scene with Multiple Linearly Moving Objects," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 542-549, 2000.

[8] R.I. Hartley, L. de Agapito, I.D. Reid, and E. Hayman, "Camera Calibration and the Search for Infinity," *Proc. Int'l Conf. Computer Vision*, pp. 510-515, 1999.

[9] R.I. Hartley, "Euclidean Reconstruction from Uncalibrated Views," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 908-912, 1994.

[10] R. Hartley, and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[11] R. Hartley, and R. Gupta, and T. Chang, "Stereo from Uncalibrated Cameras," *Proc. International Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Press*, Los Alamitos, CA, pp. 761-764,1992.

[12] Heyden, R. Berthilsson, and G. Sparr, "An iterative factorization method for projective structure and motion from image sequences," *Image and Vision Computing,* vol. 17, pp. 981–991, 1999.

[13] A. Heyden, and K. Astrom, "Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point," *In CVPR97*, pp. 438-443, 1997.

[14] M. Pollefeys , R. Koch , M. Vergauwen , and L. Van Gool, "Automated reconstruction of 3D scenes from sequences of images," *ISPRS Journal of Photogrammetry & Remote Sensing,* vol. 55 pp. 251–267,2000.

[15] M. Pollefeys, L. Van Gool, and M. Proesmans, "Euclidean 3D reconstruction from image sequences with variable focal lengths," *In Proc. European Conference on Computer Vision, Cambridge, UK*, vol. 1, pp. 31–42, 1996.

[16] M. Pollefeys, R. Koch, and L. VanGool, "Self-Calibration and Metric Reconstruction Inspite of Varying and Unknown Intrinsic Camera Parameters," *Int'l J. Computer Vision*, vol. 32, no. 1, pp. 7-25, Aug. 1999.

[17] M. Pollefeys, and L. VanGool, "Visual modeling: from images to images," *J. Visualization and Computer Animation*, vol. 13, pp. 199–209,2002.

[18] L. Quan, "Invariants of 6 Points and Projective Reconstruction from 3 Uncalibrated Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 34-46, 1995.

[19] Structure From Motion. [Online].
Available: Vision.cis.udel.edu./~cv

[20] P. Tissainayagam, and D. Suter, "Performance measures for assessing contour trackers", *Int. J.Image Graphics*, vol. 2, pp. 343–359, 2002.

[21] P. Tissainayagam, and D. Suter, "Visual tracking with automatic motion model switching", Int. *J. Pattern Recognition*, vol. 34, pp. 641–660,2001.