Depth Computing based on Stereo Matching with Singular Value Decomposition

Sandar Aung, Nyein Aye
University of Computer Studies Mandalay, Myanmar,
E-mail:sandaraung84@gmail.com, nyeinaye@gmail.com

Abstract

The system is aimed to develop the 3D model reconstruction from multi stereo images. In this paper, an efficient depth computing based on stereo matching algorithm is introduced based on feature extraction with morphological method and efficient matching method. To reconstruct the 3D model, the depth map is created by the algorithm with Gaussian correlation based on the singular value decomposition (SVD) approach is used to project the image point. An efficient surface reconstruction algorithm is developed for generating the 3D model of an object. The objective of the system is to reconstruct the 3D model of multi stereo images to reduce the time complexities and occlusion errors of the scene and object.

1. Introduction

In the last two decades, a tremendous amount of research has been done in the area of reconstructing three-dimensional objects from two-dimensional camera images. One major challenge of the reconstruction problem is to find feature correspondences, that is, to locate the projections of the same three-dimensional geometrical or textural feature on two or more images. Classical approaches to reconstruction focus on estimating structure either from stereo image pairs or from monocular image sequences. Limitations in both of these approaches have motivated a growing interest in computing structure from stereo image sequences; however, most existing techniques in this area assume that feature correspondences are established in a previous step, or that they use domain specific assumptions that are inappropriate in other applications.

This paper presents a robust, incremental 3D reconstruction algorithm using stereo image sequences. The main purpose of this thesis is to study the methodologies of automatically computing a three-dimensional reconstruction of the scene given two or multiple images taken from different viewpoints of the scene.

The next step is to extract the building features that are used to create the fundamental measurement matrix, which are inputs to the singular value decomposition method to calculate the 3D point. The morphological method is useful for extracting image components with more complex operation. The entire measurement matrix is made by gathering the measurement matrix of the feature points of each of the object through the more than two images. The SVD can robustly recover shape and motion form the image under orthographic and affine projection. It computes a reconstruction up to an unknown projective when nothing is known about the camera calibration parameters or the scene.

This paper is organized as follows. The related work of 3D reconstruction from multi stereo images is described in section II. In section III, the proposed depth computing based on stereo matching approach is presented. Section IV sketches the Result and Discussion. Finally, section V provides some concluding remarks and future directions of research.

2. Related Work

Many reports have been already reported for 3D reconstruction of stereo images. [1047] generalized Markov Random Field (MRF) stereo methods to the generation of surface relief (height) fields rather than disparity or depth maps. This generalization enables the reconstruction of complete object models using the same algorithms that have been previously used to compute depth maps in binocular stereo. Their method provides a viewpoint independent smoothness constraint, a more compact parameterization and explicit handling of occlusions.

W. E. L. Grimson [13] presented an approach to surface reconstruction from multiple images. The central idea is to explore the integration of both 3D stereo data and 2D calibrated images. This is motivated by the fact that only robust and accurate feature points that survived the geometry scrutiny of multiple images are reconstructed in space. The density insufficiency and the inevitable holes in the stereo data should be filled in by using information from multiple images.

The idea is therefore to first construct small surface patches from stereo points, then to progressively propagate only reliable patches in their neighborhood from images into the whole surface using a best-first strategy. The problem reduces to searching for an optimal local surface patch going through a given set of stereo points from images. This constrained optimization for a surface patch could be handled by a local graph-cut.

G. Vogiatzis, P. Torr, S. M. Seitz, and R. Cipolla [3] presented 3D reconstruction from stereo images for interactions between real and virtual objects. The 3D reconstruction algorithm in a stereo image pair for realizing mutual occlusion and interactions between the real and virtual world in an image synthesis was proposed. A two-stage algorithm, consisting of disparity estimation and regularization was used to locate a smooth and precise disparity vector. The hierarchical disparity estimation technique increased the efficiency and reliability of the estimation process, and edge-preserving disparity field regularization produces smooth disparity fields while preserving discontinuities that resulted from object boundaries. Depth information concerning the real scene was then recovered from the estimated disparity fields by stereo camera geometry.

G. Zeng, S. Paris, L. Quan, and F. Sillion [4] presented multi-scale 3D scene flow from binocular stereo sequences. Scene flow methods estimated the three-dimensional motion field for points in the world, using multi-camera video data. Such methods combined multi-view reconstruction with motion estimation approaches.

In order to further constrain the feature matching problem, a multitude of correspondence algorithms have been proposed in the past two decades [5, 9, 1, 6, 13, 8]. The main goal of most of these efforts is to limit the search space or minimize the number of matching candidates for each feature point.

Area-based methods assume that the appearance, that is, the intensity values, of a small neighborhood of pixels remain constant from one viewpoint to another. Hence, for all pixels in an image, a correlation measure such as cross correlation or sum of squared differences [12] can be used to find corresponding pixels in the other image with a similar looking neighborhood. The advantage of this matching technique is that a dense disparity map is achievable, which is an asset for reconstructing the complete surface structure everything in the cameras' viewpoints. However, this technique relies on the images having highly textured regions.

Feature-based methods [1, 6, 8] establish correspondence between similar features in a pair of images. Discrete features such as edges, lines, points with high intensity variation, zero-crossings of gradients, or high level structures are extracted from intensity information, and matching is only performed on these features. A distance metric is used to assess the similarity of features between the images. The advantage of this approach is that features are generally more invariant than actual image intensities under large viewpoint variation.

All the methods outlined in above for solving this problem tend to be computationally expensive and require certain assumptions about the characteristics of the images which do not always apply in any general case. Moreover, single pairs of stereo images by themselves can only provide partial representations of the scene. View registration of multiple stereo pairs would be required to obtain more complete 3D structure information.

3. The Proposed Depth Computing based on Stereo Matching for 3D Reconstruction

The system is mainly developed for 3D reconstruction of multi stereo images. The sixth processing steps are included in this system. These are acquisition of stereo images, preprocessing for image enhancement with morphological processing, feature extraction of feature point, computing depth by Gaussian correlation based on the singular value decomposition (SVD) approach, computing the 3D features and 3D image reconstruction.

3.1 2D Building Feature Extraction

The extraction of foreground object from original image by using manual method is time-consuming process. The system develops automated methods of foreground extraction. Figure 2 shows step by step of the foreground object extraction and image enhancing with morphological processing. Find the different intensity among these color components and the intensity of the low different intensity tend to zero (black). 4 stages of Morphological operations are applied to remove the remaining noise.

Figure 1. Acquiring Multi stereo Images

Table 1. Image Acquisition Constraint for Multi Stereo Images

Camera	NIKON COOLPIX 7600
Dist between camera and object	150 cm
No. of images	4
Resolution	4000x3000

Morphological Processing is performed to remove the noise and un-required small pieces from foreground image. In this system the default threshold of binarization and small disconnect parts removing that is performed by opening method is set to 0.4 and user desired threshold is also available. Filling hole is performed by Hit-miss transformation and complete filling is performed by dilation method. Finally the coordinates of white region are computed from these complete filling images and using the above feature coordinates, the image of foreground object is extracted from the color input image.

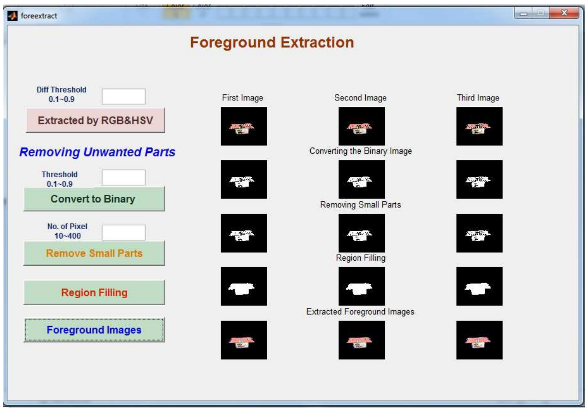

Figure 2. Morphological processing of input images

3.2 Edge Feature Detection

To get the edge feature of a foreground object, edge detection process is performed by using canny edge detection method. The three dimensional depth measurements are computed based on the multi stereo camera system for obtaining the exact 3D measurements [10].Before detecting the edge, gray scale converting has been done. Figure 3 shows the gray scale images and its edge feature detection images.

In many environments, the bulk of the important details of the scene are captured in edge points, enabling us to use only these edge points for registration rather than a full point cloud or a uniformly downsampled point cloud. Edge detection can be viewed as a means of intelligently selecting a small sample of points that will be informative for registration. We will demonstrate that this approach can be both faster and more accurate than alternative approaches. In 2D grayscale images, edges can be detected from image gradient responses that capture the photometric texture of a scene, as is done by the Canny edge-detector.

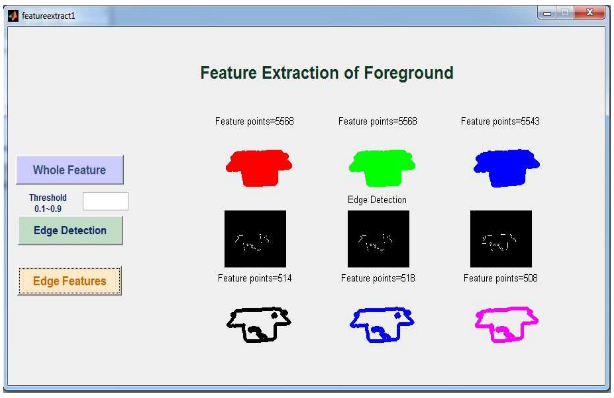

Figure3. Extraction of the whole feature point and edge feature point from stereo images

3.3 Depth Computing based on SVD approach

An essential component of depth computing algorithm is the mathematical procedure of Gaussian correlation based on the singular value decomposition (SVD) approach. Tomasi and Kanade introduced the depth computing algorithm for an affine reconstruction using a measurement matrix [2]. The algorithm assumes that the image points to be measured are present in all image view. The SVD is a simple effective technique for reconstructing the 3D structure of object space by using two images from un-calibrated camera. 3D affine transformations have been widely used in computer vision.

The 3D feature depth map is obtained by using the singular value decomposition under affine projection approach. The extracted features from the image are transformed the fundamental matrix which decompose the three matrix using the factorization method of SVD. SVD can be performed on real (m, n) matrix.Once build the matrix G, decompose the matrix into the multiple of orthogonal matrices T, U and diagonal matrix D as follow.

$$G = TDU^T \qquad \text{Equation (1)}$$

Convert D to a new matrix E obtained by replacing every Diagonal element with 1 and then compute the product

$$P = TEU^T \qquad \text{Equation (2)}$$

P contains the matched pairs of feature points.

Then all corresponding image pairs between two images are computed by the transformation matrix. The transformation matrix T is computed as follow:

$$S_2 = TS_1$$

$$S_2(S_1)' = TS_1S_1'$$

$$S_2S_1'(S_1S_1')^{-1} = T(S_1S_1')^*(S_1S_1')^{-1}$$

$$T = S_2S_1'(S_1S_1')^{-1}$$

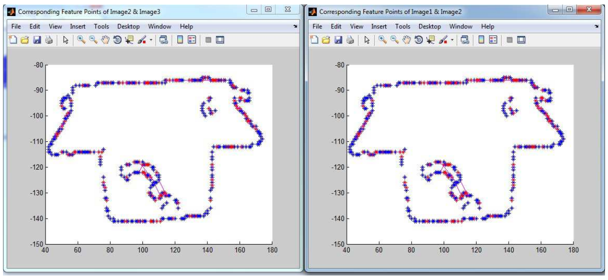

Figure4. Matching pair with SVD approach

To reconstruct the 3D shape of each object, it is necessary to obtain the group of feature points relative to the object separately. Each building on an object is considered as an individual object and the shape of the object. The entire measurement matrix is made by gathering the measurement matrix of the feature points of each object through the more than two images. The 3D cloud points are obtained from Stereo System.

Consider the left and right stereo camera system as following Figure7. If the positions of the image *P* in two cameras are x_0 and x_1 for focal length *f*, the depth measurement *Z* can be described as:

$$Z = f\frac{T}{d} \qquad \text{Equation (3)}$$

$d = x_1 - x_0$ is the *disparity*.

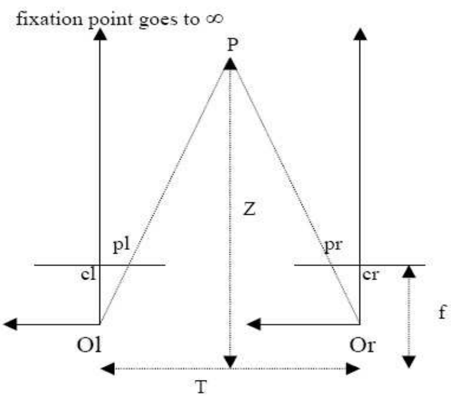

Figure 5. Stereo Camera System

Next, surface reconstruction for 3D building reconstruction using Delaunay triangulation method and wire frame drawing. The triangle is build using the three nodal points which are projected from a planar surface in the 3D scene and the surface reconstruction is achieved by using Delaunay triangulation method and wireframe.

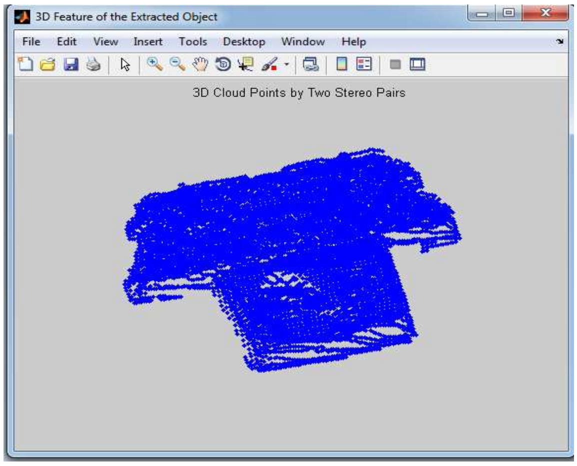

Figure6. Calculated 3D cloud point of the object

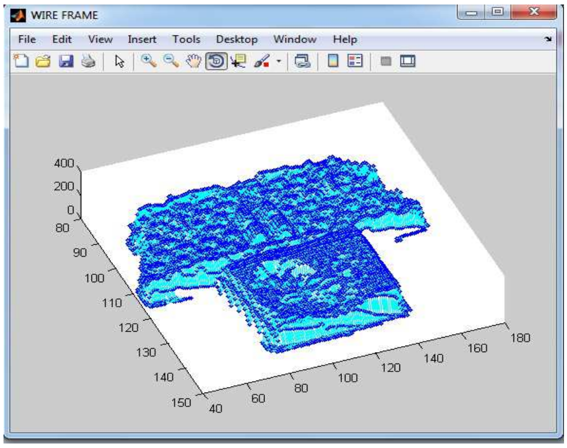

Figure7. Triangle drawing for the object

4. Result and Discussion

MATLAB R2010a is applied for performance of reconstruction the 3D surface of an object from 2D images. NIKON COOLPIX 7600 digital camera is used for image acquisition. An object is put on the turn table and the images can be taken by a camera by rotating the turn table if need. Unlike the other existing reconstruction system, the proposed system used only the edge detection clould point to reduce the complexity of 3D reconstruction. In the experiments, over 20 images were tested with six different types of objects. Generally, many objects have formed by the combination of curve surface, flat surface, convex and concave surface, etc. The 3D surface data is computed by the depth information from the images. The depth is computed by factorization method of SVD [11]. The SVD is a robust mathematical function to employ when reconstructing the affine structure. This method can robustly recover shape and motion from the image under orthographic and affine projection. The 3D feature point and shapes of the building are recovered from the images of an object by multiple objects. It is necessary to obtain the group of feature points relative to the object separately to reconstruct the 3D shape of each object. Each building on an object is considered as an individual object and the shape of the object. Using this information, the dense depth map for the extracted building object is created in the next stage. Once the depth data is obtained, the 3D positions of the object are computed from the relation of the camera viewing angles and the centre coordinate of image. The entire measurement matrix is made by gathering the measurement matrix of the feature points of each object trough the more than two images. In the final stage, 3D building result is described to prove the how does work of the proposed method.

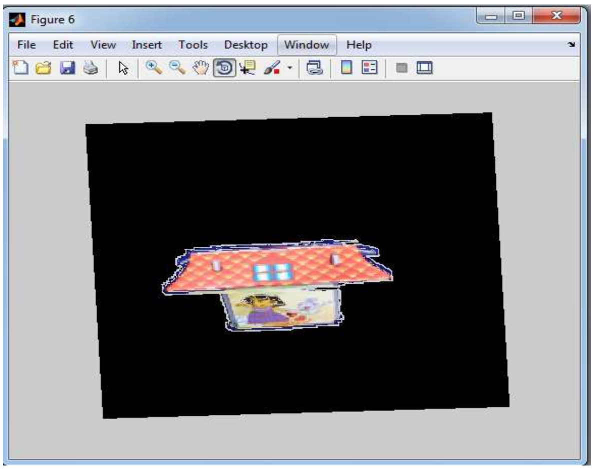

Figure 8. 3D surface reconstruction model from stereo images

To measure the performance of the proposed system, the time complexity of the reconstruction of 3D images using the whole feature point of the images and the proposed reconstruction of 3D images using the edge feature point is calculated as following equation (5).

$$\textbf{O (reduce rate)} = \frac{\textbf{O (TP1 * TP2)}}{\textbf{O (TE1* TE2)}} \qquad \text{Equation (4)}$$

TP1= Total no. of pixel from first image

TP2=Total no. of pixel from second image

TE1=Total no. of edge feature from first image

TE2=Total no. of edge feature from second image

The sample time evalution for the example multistereo images is as shown in Figure 9.

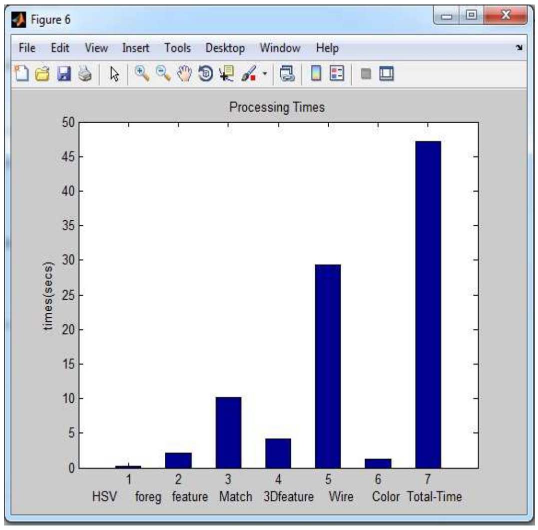

Figure 9. Evaluation of Time Performance of 3D Reconstruction of Figure 1

5. Conclusion

3D reconstruction system from multivew stereo images is developed to reduce the time complexities and occlusion errors of the scene and object that cloud point is find by edge point rather than whole featrue point. The reconstruction of 3D building model can be achieved using more than two images without knowing the camera calibration. The efficient building features are extracted using colour segmentation and Morphological processing. Stereo matching algorithm is developed using single value decomposation approach and cloud edge point is detected by canny method. The surface reconstruction is achieved by using Delaunay triangulation method and wireframe. It provides to reduce the processing cost and complexity errors instead of computing the corresponding pairs for whole object directly.

References

[1] D. Marr and T. Poggio, “A computational theory of human stereo vision,” Proc. Royal Soc. London B, 204:301–328, 1979.

[2] E. Trucco and A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, 1998.

[3] G. Vogiatzis, P. Torr, S. M. Seitz, and R. Cipolla. Reconstructing relief surfaces. In BMVC, pp. 117–126, 2004.

[4] G. Zeng, S. Paris, L. Quan, and F. Sillion. Progressive surface reconstruction from images using a local prior. In ICCV, pp. 1230–1237, 2005.

[5] Hansung Kim, and Kwanghoon Sohn: *3D Reconstruction from Stereo Images for Interactions between Real and Virtual Objects*, Signal Processing: Image Communication 20 (2005) 61–75.

[6] J. E. W. Mayhew and J. P. Frisby, “Psychophysical and computational sutdies towards a theory of human stereopsis,” Artificial Intelligence, 17:349–385, 1981.

[7] J. Weng, N. Ahuja, and T. S. Huang, “Matching two perspective views,” IEEE Trans. Pattern Analysis and Machine Vision, 14(8):806–825, 1992.

[8] N. Ayache and B. Faverjon, “Efficient registration of stereo images by matching graph description of edge segments,” Int. J. of Computer Vision, 1(2):107–131, 1987.

[9] Rui Li and Stan Sclaroff: *Multi-scale 3D Scene Flow from Binocular Stereo Sequences*, IEEE Workshop on Motion and Video Computing, January (2005).

[10] S. Aung, Computing the Three Dimensional Depth Measurement by the Multi Stereo Images, International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 3, No. 5, October 2013. pp.405-408.

[11] S. Aung, Stereo Matching with Gaussian Correlation Approach, International Journal of Core Engineering and Management, Volume 1, Issue 7, October 2014..

[12] S. T. Barnard, “Stochastic stereo matching over scale,” Proc. DARPA Image Understanding Workshop, 769–778, 1988.

[13] W. E. L. Grimson, “Computational experiments with a feature based stereo algorithm,” IEEE Trans. Pattern Analysis and Machine Vision, 7(1):17–34, 1985.