SURFACE RECONSTRUCTION OF STEREO IMAGES USING FEATURE EXTRACTION ALGORITHM

[1]Vijayalakshmi B., [2]Sheelaranib B. and [3]Manoharan N.

[1,2,3]Sathyabama University, Chennai-119
Email: vijayalakshmibaba@rediffmail.com.

Abstract

Surface reconstruction and their similarities is a central problem in the field of vision based modeling. It arises in particular from the task of classifying and recognizing objects from their observed silhouette. Defining natural distances between image discontinuities creates a metric space of shapes, whose mathematical structure is inherently relevant to the classification task. One intriguing metric space is identified from using conformal mappings through Feature Extraction Algorithm of 2D stereo images. Surface reconstruction may be performed as unconstrained tracking and 3D data merging or as iterative structure from motion, or through constrained depth recovery using epipolar geometry. Immediate problems with these approaches are aperture problem, variation in stabilizing factor over time and object distortion at different viewpoints. The above said difficulties are avoided by introducing Feature extraction techniques. This approach results a denser depth map from the traits with variable window size to avoid distortion across the composite planar image. This constraint reduces the aperture problem during search. Wide-angle reconstruction of 3D scenes is conventionally achieved by extracting the features from stereo images, have been performed using two CCD Cameras under static and dynamic environment. The result shows the required optimal window requirement to get photorealistic surface reconstruction of physical environment scenes with minimum human intervention.

Key Words: Denser depth map, Feature Extraction, Surface reconstruction, Stereo Image, Optimal Window.

I. INTRODUCTION

In the last years the generation of 3D models of an object has become a topic of interest for several researchers. Particular attention has also been paid on the reconstruction of realistic environment models, which could be employed in a wide range of applications such as medicine, surveillance, video games, virtual reality environments or ergonomics applications. A complete environment model usually consists of the shape and the position of the objects in a cross country terrain. Some available systems consider the two modeling processes as separate, even if they are very close. Considering the techniques that recover the shape of static and dynamic environments, relies on 3D scanners, these sensors are quite expensive but simple to use and various software is available to model the 3D measurements. It refers according to different technologies providing for millions of points, often with related color information. Other techniques try to recover the shape of object with image-based approaches [1]. It utilizes camera stereo-view geometry silhouette extraction or single image measurements. Computer animation software can instead produce realistic 3D environment model by subdividing and smoothing polygonal elements, without any measurements. These spline-based systems are mainly used for movies or games and the created virtual object is animated using similar animation packages or with motion captures data [2]. Concerning the motion of the dynamic systems, the main problem is the great number of degrees of freedom to be recovered. Existing and reliable commercial systems for capturing object motion typically involve the tracking of object movements using sensor-based hardware [3]. Other approaches instead rely on 2D monocular videos of environment as primary input. Through computer vision techniques, image cues, background segmentation, statistics, prior knowledge about object motion, probabilistic approaches and pre-defined articulated object models to recover motions and 3D information the surface reconstruction is obtained.

Many research activities in this area has focused on the problem of tracking a moving object through an image sequence acquired with stereo camera and often using pre-defined 3D models. But little attention has been directed to the determination of 3D information of an object surface directly from an image sequence using a camera model. In this contribution a simple and efficient method to find the poses and the 3D model of an object from the development of feature extraction software for obstacle detection of a cross country terrain stereo images. The paper is organized as follows. Section 2 provides an overview of the methods and contributions of related work while section 3 discusses model issues. Section 4 describes the developed Feature Extraction algorithm for the surface reconstruction of the stereo image. Section 5 introduces the experimental system overview and Section 6 evaluates the quality of the results on the real time images that are grabbed is required for a typical applications. Section 6 concludes with future work.

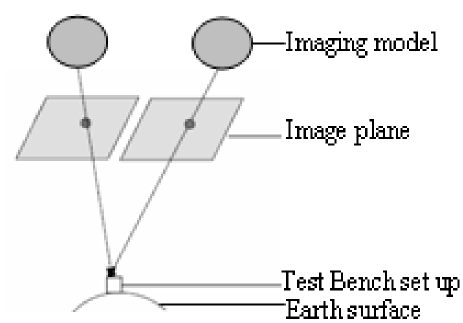
II. METHODS AND CONTRIBUTIONS

A. Triangulation method of reconstruction

Surface reconstruction of scene from stereo framework can be understood naturally as a generalization of traditional passive stereo methods that operate entirely within the spatial domain shown in Fig.1. Traditional stereo finds correspondence pixels by comparing spatial neighborhoods around candidate pairs of pixels in the two images. This current frame work simply adds a depth dimension of a depth map from the triangulation technique to the neighborhoods for stereo matching [4]. Computed depth forms the reconstructed surface model of an object. But the maximized error occurred while solving the correspondence between the images alters the reconstructed surface from the physical system.

Fig. 1. Traditional passive stereo method

B. Colinearity based reconstruction



To recover accurate surface models from images are based on the co linearity equations. They state that a point in object space and its corresponding point in an image and the projective center of the camera lie on a straight line. If a point in stereo-image is more than one frame, its surface coordinates with respect to world coordinate system can be recovered from the camera parameters [5]. Although this method is very accurate, if a point to be imaged is at least in two images and a good baseline between consecutive frames. Therefore it is not possible for a dynamic camera imaging system. A simplification of co linearity equations leads to the perspective projection that relates the image measurements to the world coordinate system (X, Y, Z) from image coordinates (x, y) just through the camera constant c referred as,

$$x = -c\frac{X}{Z} \tag{1}$$

$$y = -c\frac{Y}{Z} \tag{2}$$

To recover surface information from a uncalibrated view, referred as 'ill-posed' problem [6], requires three unknown coordinates, and the camera constant. Therefore the system is underdetermined and some more assumptions need to be introduced. For man made objects like buildings, geometric constraints on the object and image invariant can be used to solve an ill-posed problem. But in case of free form objects like human body and animals, these assumptions are not valid. Therefore (1) can be furthermore simplified, by describing the relationship between the surface of the object coordinates and 2D image measurements with a scaled factor $s = -c/Z$.

The effect of orthographic projection is a simple scaling of the object coordinates. The scaled-orthographic model amounts to parallel projection, with a scaling added to mimic the effect that the image of an object shrinks with the distance. This camera model can be used if we assume the Z coordinate almost constant in the image or when the range of Z values of the object is small compared to the distance between the camera and the object. In those cases the scale factor c/Z will remain almost constant as optimized value. Moreover it is not necessary to recover the absolute depth of the points with respect to the object coordinate system. Furthermore the camera constant is not required and this makes the algorithm suitable for all applications that deal with uncalibrated images or video. But, due to ill-posed problem, still have an undetermined system, as the scale factor s cannot be determined only by means of (2) and a single frame. Complexity of the system may be reduced by extracting the features from the grabbed stereo image.

C. Reconstruction through Feature Extraction

Stereo images are block adjusted by windowing technique from the physical camera model. The block adjusted stereo images are always resampled to epipolar geometry. As a result this can be used directly for stereo feature extraction without any additional adjustment and resampling. The object-to-image relationship for each stereo image is expressed by a sample Rational Polynomial Camera RPC model through least-squares approach.

III. MODEL ISSUES

A cross country terrain model is the elevated model of the landscape. On the other hand, a Digital Surface Model includes the, objects with their heights above as well as below the ground surface. The basic idea of using a feature extraction technique to reconstruct the surface of the man made objects with different heights over the terrain can be detected by applying a threshold to a normalized surface model. The other issue is realizing an accurate point correspondences are difficult to compute in regions with homogeneous color and intensity. View dependent effects, such as specular highlights or reflections, leads to correspondence mismatches. Obtaining dense correspondence for many image points is especially hard. Finally, differences between images due to occlusions are difficult to handle. This is a severe problem for general scene reconstruction where such occlusions happen frequently.

Apart from this image processing methods must be extended in order to describe the imaging geometry, which is characterized by nearly parallel projection in along-track direction and perspective projection in cross-track direction as well as the mathematical description of the sensor model. The image sensors are modeled with the camera calibration techniques. Sensor calibration defines the intrinsic and extrinsic parameters equivalent to the world model. The difficulty in this process was the separation of the buildings from the trees as both features have heights above the ground. However, their separation can be achieved using different structural matching technique during image processing phase.

IV. FEATURE EXTRACTION ALGORITHM

Stereo images are grabbed from the cross country terrain with the help of test bench module. Gray scale conversion is performed in order to reduce the computational complexity. Wavelet based denoising technique has been adopted for removing the impulse noise. Pre processing is done to reduce the complexity of correspondence matching and camera modeling. A seed pixel is selected with respect to the threshold value, finding out the difference between every image pixel from the solution of the correspondence problem of the stereo image. If a selected seed pixel difference is less than or equal to the threshold, then that pixel is appended, to give independency from image parameter and image specific. The motion field can be described as the projection of the 3D velocity field in a scene onto a 2D image plane and is analogous to optical flow for many situations. Let

$$P = [X, Y, Z]^T \tag{3}$$

be a 3D point in the camera reference frame, Z the optical axis and f the focal length, then the image of a scene point, P can be mapped to the image point p through the following equation:

$$P = f\frac{P}{Z} \tag{4}$$

From this required basic equations are,

$$v_x = \frac{T_Z x - T_X f}{Z} - w_y f + w_Z y + \frac{w_Z xy}{f} - \frac{w_y x^2}{f}$$

$$v_y = \frac{T_Z x - T_y f}{Z} - w_x f + w_Z x + \frac{w_y xy}{f} - \frac{w_x y^2}{f} \tag{5}$$

where v is the velocity vector in image space, T is the translational velocity vector in camera space, f is focal length, Z the depth to the motion point in camera space and ω is the rotational velocity vector in camera space. Detecting discontinuity parameter C is an important process for obtaining the enhanced features to extract from optical flow images,

$$C = \begin{bmatrix} \Sigma E_Z^2 & \Sigma E_X E_Y \\ \Sigma E_X E_Y & \Sigma E_Y^2 \end{bmatrix} \tag{6}$$

where the sums of image brightness E are taken over the neighborhood pixels p. The discontinuities are found by extracting the two eigen values of the matrix that encode the intensity strengths thus producing a new matrix C:

$$C = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad (7)$$

Thus when $\lambda_1 \geq \lambda_2$ and λ_2 is over a given threshold value, it corresponds to two strong discontinuity in the same image window thus specifying a feature clouds. The features are extracted from the region of interest with respect to camera modeling through image processing steps like Image enhancement, Dilation. Region growing and Erosion are performed to extract the discontinuity available in the image. Image enhancement is done with averaging and Gaussian filters to highlight the object features of stereo image. Dilation is performed on the edge detected image to bridge the gap between the edge pixels. Region growing is performed on the dilated image to differentiate the region of required pixels from the background pixels. The surface reconstruction Map in order to provide a different representation of object information from the enhanced feature clouds with the range threshold function is obtained as,

$$prox_i = d\,(r_i) = \frac{K_p}{r_i^2} \quad (8)$$

where r_i is the range in metres to the point of interest i, K_p the threshold constant and $prox_i$ representing range information converted to a $0-1$ proximity scale, where 1 representing the closest surface. Basic blocks to perform this proposed technique of a reconstruction is given in Fig.2.

The complete system can be divided into an offline data acquisition and an online rendering part. In the offline part, the images are preprocessed to estimate calibration and feature extraction through the identification of discontinuity from depth maps of each view. In the online rendering of the image data set is created from novel views of CCD cameras at an interactive rate.

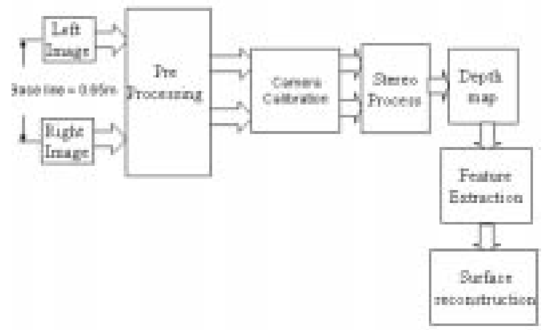


Fig. 2. Functional blocks of a reconstruction technique

V. SYSTEM OVERVIEW

A model was designed with cameras mounted on a wheeled wood ware with a provision of PC with power supply, Hub/Ethernet switch, and Ethernet cable as in Fig.3.

Fig.3. Test bench module

The image acquisition is performed on a two video sequences of 540 frames in 52 sec, of size $640 \times 480 \times 24$ BPP which are grabbed under the developed terrain model inside the campus. The programs are performed on a Pentium IV 2.4 GHz computer using Matlab language V7.0. As the number of cameras is two, a fast Ethernet switch was used to connect both of them simultaneously to one PC as in Fig.4.

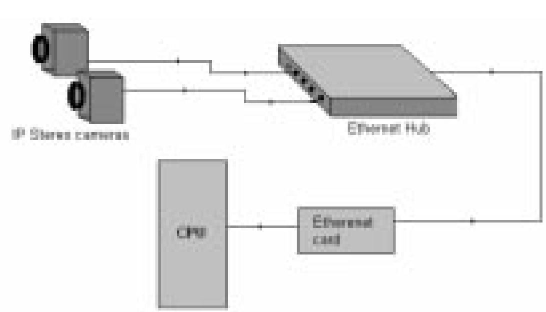


Fig. 4. Image acquisition system

The cameras were mounted on the test bench with an arrangement of dynamic baseline, from 0.6m to 1m in steps. Video images are grabbed in and around of the University campus, on that some of the images is shown in Table 1 as sample data base. The original images have been taken from different environment conditions and different orientations with respect to the test bench. To view them in stereo, it will not be able to align them so that corresponding features come together. The best is image rectification included in the stereo process to appear the images have been taken from approximately the same look direction.

An additional feature of epipolar resampling is also performed to achieve stability on scene alignment which allows comfortable stereo viewing. The object-to-image relationship of each stereo image is expressed by a sample camera modeling parameter through camera calibration technique and surface reconstruction is performed with the developed stereo feature extraction technique. This experiment is conducted for both static and dynamic image conditions

Table I Sample Data Base Of A Grabbed Stereo Image

Sl.No	Image ID	Sample Real time Stereo Test Set		Type of the image
		Left Image	Right Image	
1	DSC02011			Static indoor image
2	DSC02032			Dynamic dense indoor image
3	DSC02171			Dynamic outdoor image
4	DSCO2231			Static and dynamic outdoor images

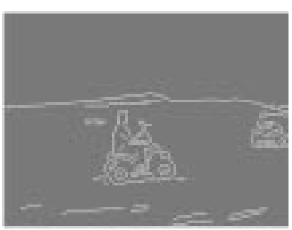
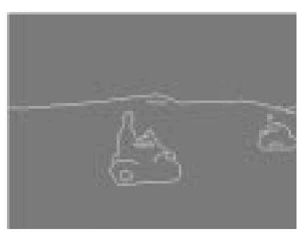
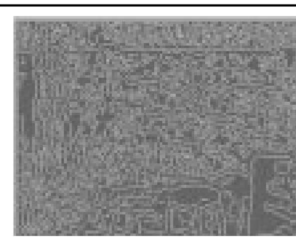
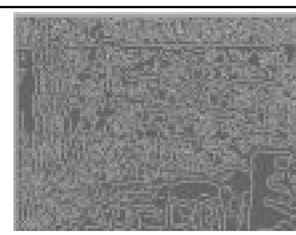
and it is compared with the existing feature extraction techniques.

VI. RESULTS AND DISCUSSIONS

The preliminary results show that the proposed feature extraction method is quite promising for detecting the objects from its surface reconstruction is given in Table.2. The results of various experiment conducted refers success of Discontinuity based extraction technique is dependent on number of parameters such as choice and size of image smoothing operators, the choice of edge detecting mask, threshold chosen for edge detection, size and shape of the structuring elements chosen for morphological image processing. Hence this technique cannot be standardized and is not suitable for real time dense object detection. Object complexity is also very high. But it suits for unocculed static minimal object conditions.

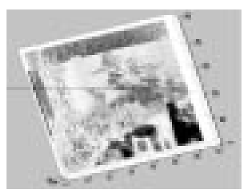
In the case of region growing based extraction large discontinuity only identified as features and it lost the finer details available in the image hence improper surface reconstruction. The solution for the above said

Table I1 Feature Extracted Transform

Image ID	Comparisons of Feature Extraction Techniques		
	Discontinuity based extraction	Region Growing based extraction	Wavelet based extraction
DSC02011			
DSC02032			
DSC02171			
DSCO2231			

Thus a software-based surface reconstruction of stereo images using feature extraction algorithm is implemented. The proposed approach is satisfactory to reconstruct the surface objects and its virtual model is shown in Table III.

Table III Surface Reconstruction

Sl.No	Image ID	Surface Reconstruction
1	DSC02011	
2	DSC02032	
3	DSC02171	
4	DSCO2231	

problem is given with the developed wavelet based feature extraction technique. It is clearly shown in the last column of Table II. The developed algorithm gives better approximation during the surface reconstruction of different image conditions as well as dense object conditions also. Equivalent surface reconstruction of different image condition is implemented; it gives the depth information between the objects with respect to the imaging device orientation, which is not available in the case of 2D image model.

VII. CONCLUSIONS

Feature extraction system is developed for the automatic reconstruction of uncalibrated image sequences. The camera path was calibrated and nearly dense feature extraction was computed. Two different existing methods of extraction and the developed algorithm were discussed. The reconstruction results show that the generated quality is still not sufficient for seamless rendering from arbitrary extrapolated image positions for dense medium like trees. One issue to investigate further is the proper handling of different environment conditions and the problem of global seamless integration of local models.

REFERENCES

[1] R. Szeliski and H.-Y. Shum. "Creating full view panoramic image mosaics and environment maps", *Proceedings SIGGRAPH '97,* pp. 251-258, 1997.

[2] Chris McCarthy and Nick Barnes, "Performance of optical flow techniques for indoor navigation with a mobile robot", *Proceedings of the 2004 IEEE*, April 2004

[3] T. Schamm, S. Vacek, J. Schröder, J.M. Zöllner, and R. Dillmann. "Obstacle detection with a Photonic Mixing Device-camera in autonomous vehicles". *International Journal of Intelligent Systems Technologies and Applications*, 5:315–324, Nov. 2008.

[4] E. Rosten and T. T. Drummond, "Machine learning for high-speed corner Detection". In *Proc. of the ECCV*, volume 1, pp. 430–443, 2006.

[5] M.P. Parsley and S.J. Julier. "Avoiding Negative Depth in Inverse Depth Bearing-Only SLAM", In *Int. Conf. on Intelligent Robots and Systems, IROS*, 2008.

[6] R. Hartley and A. Zisserman, "*Multiple View Geometry in ComputerVision". Cambridge University Press*, ISBN: 0-521-54051-8, second edition, 2006.

[7] H. Hirschmüller, P. Innocent, and J. Garibaldi, "Real-Time Correlation-Based Stereo Vision with Reduced Border Errors", *International Journal of Computer Vision*, 47, pp. 229–246, 2002.

[8] W. van der Mark and D.M. Gavrila, "Real-time dense stereo for intelligent Vehicles"., In *IEEE Trans. on Intelligent Transportation Systems*, 2006

[9] A. Wedel, U. Franke, J. Klappstein, T. Brox, and D. Cremers, "Realtime Depth Estimation and Obstacle Detection from Monocular Video", *DAGM Symposium on Pattern recognition*, 4174, pp. 475–484, 2006.