

Large Scene Reconstruction with Local Details Recovery

S. H. Or[†] K. H. Wong[†] Michael M. Y. Chang[‡] C.Y. Ip[§]

[†]Computer Science & Engineering Department
The Chinese University of Hong Kong
Hong Kong

[‡]Information Engineering Department
The Chinese University of Hong Kong
Hong Kong

Abstract

We propose a method to recover the global structure with local details around a point. To handle a large scale of motion i.e. 360 degree around the point, we use an optimization-based algorithm to estimate the structure from the panorama around the fixed camera point. The global structure estimated can thus be used to initialize a structure from motion algorithm to recover the local details through simple camera motion such as panning. Synthetic as well as real data are used to test the validity of the algorithm. Our method can be used in applications such as authoring of virtual environments from real scene.

1. Introduction

Structure from motion {SFM} is a well established technique to recover both the structure and motion information from multiple views of an object. Most algorithms [4, 12] are developed to handle the particular problem of estimating the camera centered 3D coordinates of feature points of the object to be investigated. However to handle the problem of correspondences among all frames, almost all of the algorithms adopt a continuous frames approach which assuming little camera motion is applied to the object. By doing this, it is guaranteed that the feature points could be consistently located throughout the sequence. The drawback is that the structure recovered is limited to local region i.e. part of the object only. To recover a structure with greater coverage such as a building interior, much human effort is needed with the SFM technique.

In this paper, we propose a method to handle the particular problem of recovering both the global structure and the local details around a point. To handle the large range of motion involved i.e. 360 degree around the point, we propose an optimization-based algorithm to estimate the structure from the panorama around the camera point. The global structure estimated can thus be used to initialize a structure from motion algorithm which recovers the local details through simple camera motion such as panning.

To recover the structure within the panorama, we propose a modified line drawing interpretation algorithm which is originally proposed by Marill [7]. For the local details estimation, we used the factorization algorithm by Tomasi and Kanade [12]. The resulting algorithm thus have the benefits of large scale recovery with accurate small details estimation.

*shor@cse.cuhk.edu.hk

†khwong@cse.cuhk.edu.hk

‡mchang@ie.cuhk.edu.hk

§cyip@cse.cuhk.edu.hk

2. Background

3D model acquisition is a fast developing field which attracts much interests recently [6, 8]. Usually the model to be acquired is placed in a controlled environment and active sensing methods are employed to extract the surface of the object examined. Registration is also needed so that a complete model can be recovered due to the limitation that only a partial surface is extracted at each scan [8]. For those models that are of greater sizes, active sensing such as laser range finder may not be optimal.

Recently researchers started to investigate the possibility of using panorama in model recovery [3, 9, 10]. By using cylindrical panorama obtained at different locations and using the 8-point structure from motion algorithm together with multi-base line stereo, [4] reconstructed the depth map from a large field of view. By incorporating human intervention, [9] described a technique to perform the 3D reconstruction from cylindrical panorama. Our proposed method also shared the same inspiration of 3D reconstruction from panorama with the difference that only a single panorama is needed. [6] proposed a method to decouple the non-metric component of the homography matrix and recovering the intrinsic parameters as well as planar structures from a single view. By utilizing the physical constraints such as planarity and perpendicularity and parallelism, Sturm [10] proposed an algorithm to perform reconstruction from a parabolic mirror based omnidirectional camera. Our algorithm also share the same idea of human input with that of Sturm [10] but with the difference that our approach need only the input of planarity i.e. reconstruction can be performed without the parallelism and other constraints.

3. Theory

Our system consists of two stages: Global structure recovery and local details estimation. In global structure recovery, a panorama around the scene is first constructed. Panorama stitching from series of perspective images is a well established technique [2, 11] and can be achieved through various commercial packages. A small set of connected straight line features inside the panorama are then identified. The extraction of line features can be achieved through image processing techniques or manual effort. The set of line features is then passed to an optimization based algorithm which reconstructs a 3D model of the scene. For those parts in the scene where more irregular, complicated surfaces are found, a sequence of images are taken further about that part and traditional SFM algorithm is applied. More accurate estimate of the details in that region can thus be made. The model obtained from the previous stage can

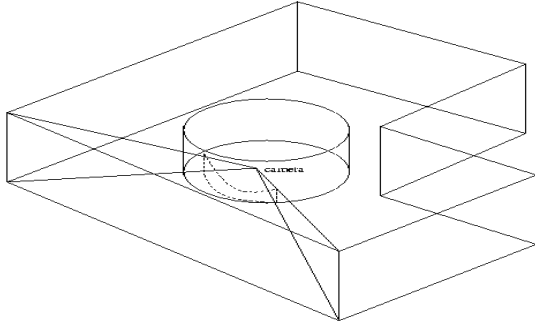


Figure 1: Projection of a model on to a unit cylinder forming a panorama.

be used here as the initialization for the SFM algorithm.

3.1 Global Reconstruction

Marill [7] presented a simple but effective algorithm to recover the 3D model from a 2D line drawing. The algorithm works by associating a depth value with each vertex in the drawing. For N vertices in the drawing, the depth values form a N vector. The initial value of the depth for all points in the line drawing will be zero since they are currently on the same plane.

Each vertex in the drawing would have its own connections to other vertices in the drawing, forming a line each. For a vertex having more than one vertex connected to it, each line pair would produce an angle. Marill observed that for a physically valid object, the standard deviation of these angles $std(ang)$ should be at a minimum. The cost function which describe a valid 3D model is thus

$$\Phi_0 = \sum_{i=1}^N std(ang). \quad (1)$$

By adjusting the depth values for different vertex so that the resulting cost function is a minimum, a 3D model representing the 2D line drawing can be obtained.

The algorithm works under the assumption that the recovered model composed of planar surface. The planarity constraint which is developed by Leclerc and Fischler [5] should also be considered. To implement the planarity constraint, the vertices that forming a face in the resultant model need be specified. Suppose n is the number of sides in the polygon and θ_i be the angle at the i^{th} vertex, and l_j is the j^{th} line of the polygon. The planarity of the estimated model can be enforced by adding the following two terms in the objective function:

$$\Phi_1 = \left((n-2)\pi - \sum_i \theta_i \right)^2, \quad (2)$$

and

$$\Phi_2 = \sum_i \left[1 - \left(\frac{(l_{j-1} \times l_j) \cdot (l_j \times l_{j+1})}{\|l_{j-1} \times l_j\| \|l_j \times l_{j+1}\|} \right) \right]^2. \quad (3)$$

In the above, Φ_1 corresponds to the interior angle restriction of a convex polygon. Φ_2 represents the torsion factor in a concave polygon.

The resulting objective function is therefore given by

$$obj = \alpha\Phi_0 + \beta\Phi_1 + \gamma\Phi_2, \quad (4)$$

where α , β , and γ are weighting constants.

The algorithm proposed by Marill handles only the situation of perspective view. In our application which is to recover the structure around a point in space, a panorama i.e. a cylindrical projection of a scene is obtained. The situation is shown in fig. 1. In the figure, all vertices of the original model are projected on the unit cylinder i.e. the panorama. Note that we can still recover the original 3D model by converting the normalized cylindrical coordinates into Cartesian counterpart, and multiplying each vertex by its corresponding depth values. The problem of recovery is thus casted as finding the *depth vector* i.e. N depth values of the vertices, of the scene. The problem domain is now different from that of Marill as the situation is now in a cylindrical system.

We propose a modified version of Marill's algorithm to determine the *depth vector* of the scene. Firstly feature points which corresponding to the corners of the planes inside the scene together with their connectivity are determined. The data are then passed to the algorithm below to estimate the structure.

Input :

- N feature points in cylindrical coordinates $(1, \theta_i, h_i)$ where θ_i is the panning angle of point i and h_i is the vertical coordinate of point i in the panorama, and $i = 1 \dots N$.
- Depth vector D of dimension N , initialized to 1 for all elements D_i .
- *stepSize* initialized to 0.99

Algorithm to recover the depth vector from the panorama

1. Calculate the current objective function value c.f. eq.4, of the structure in unit cylinder projection together with planarity constraint applied.
2. for $scale = 1 \dots ScaleMax$
 - repeat forever
 - for each *line* in the drawing
 - compute new objective (cost) function value for $D(i) = D(i) + \frac{stepSize}{scale}$, and $D(j) = D(j) + \frac{stepSize}{scale}$ where i and j are the index of points connecting the selected line,
 - restore $D(i)$ and $D(j)$ to their original values.
 - for each *point* i in the drawing
 - compute new cost value for $D(i) = D(i) - \frac{stepSize}{scale}$,
 - restore $D(i)$ to its original value.
 - Pick the minimal cost.

```

if the minimum is smaller than
the current function value,
    make the corresponding
    changes and set the current
    cost as minimum.
else
    no improvement can be made,
    so break from repeat loop
end repeat

```

end scale

Note that we 'expand' (add step size to the feature points) two vertices at a time whereas only 'contract' (subtract step size from the current radius) one vertex in the same iteration. The reason is that as we are now performing the model estimation in a radial fashion, any expansion of vertex away from the center will induce lesser changes in the resultant *SDA* than that of contraction counterpart.

3.2 Local Reconstruction

Various established techniques are available for estimating the structures from a set of feature points tracked throughout a number of images of the same object. In this project, we used the factorization method [12] developed by Tomasi and Kanade due to its robustness and non-iterative nature. As the factorization method is now a well established technique, we suggest the interested readers to refer to [12] for the implementation details.

4 Results & Discussion

Both synthetic and real image data are used to confirm the validity of our algorithm. The first data set consists of a gazebo placed at the one of the ends of a cross shaped corridor. The gazebo is a public domain model of 750 vertices with 1246 polygons. A perspective view of the model is shown in fig 2). The panoramic view of the scene is also shown in the same figure. To handle the reconstruction, firstly the background i.e. the cross shaped corridor is estimated by applying the global scene recovery process. The 2D panoramic coordinates of the vertices as well as the connectivity are passed to the global recovery algorithm. The recovered model is shown in fig. 2. In local structure recovery, a set of tracked feature points is required. This is handled by first obtaining a panning sequence about the object, which is the gazebo in this data set. We created a panning sequence of horizontal motion about the gazebo and use the projected vertices coordinates as the input to the factorization algorithm. The reconstructed gazebo – only the face towards the camera, is shown in fig. 2.

Real image data presents a challenge to the application of our algorithm. Our algorithm is suitable for estimating the structure of a building interior and we test our algorithm with real images of an existing building. A laboratory scene is also used to test our algorithm for real image data. A scene showing a miniaturized corridor of L-shape is constructed with a protruding feature placed on one of the sides. The protruding feature is actually a box stacked on another of bigger size. We use a simple web-cam for images capture and mount the camera on a turn table to take a series of 25 images at the center of the scene. The perspective images are then input to a commercial panorama stitching program and the resulting panorama is shown in fig. 3.

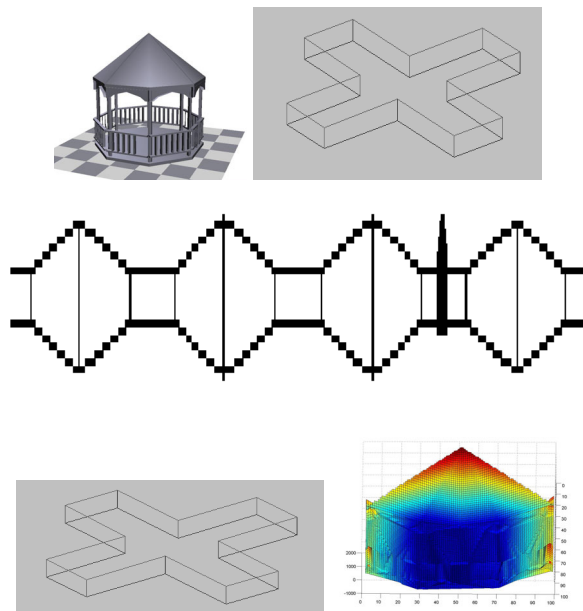


Figure 2: A synthetic scene of a corridor in a cross shape with a gazebo placed at one of the ends of the corridor and estimation result. Top left : A perspective view of the gazebo, top right: The original corridor model, middle : The panorama of the scene. The observation point i.e. the position of the camera is at the middle of the cross, i.e. center. bottom left : Estimated corridor, and bottom right: Recovered gazebo surface from the panning sequence.

For global structure estimation, the vertices together with their interconnection need be identified first. For real images data, a number of image processing techniques can be applied. For example, by applying some corner detectors to a blurred version of the panorama together with edge detection can somewhat automate the process. In our experiment, we manually locate the corners as well as connectivity due to the relatively small number of vertices, which is only 12.

The panning angles θ and the height of the feature point in 3D space are calculated from the pixel coordinates of the points (x_f, y_f) and image size $(imgwidth, imgheight)$ as $\theta = \frac{x_f \cdot 2\pi}{imgwidth}$ and $h = \frac{(imgheight - y_f)}{imgheight}$, thus the initial 3D Cartesian coordinates of the features are given by $x = \cos(\theta)$, $y = \sin(\theta)$, $z = h$.

The coordinates data are then passed to our modified Marill algorithm to perform the model recovery.

In local structure i.e. the protruding feature, estimation, we took a simple panning sequence of perspective images about the boxes at the same view point as the panorama. We used the KLT [1] tracker to obtain a sequence of tracked features. The performance of KLT tracker is good enough to supply quality feature points for the factorization algorithm. A view of the reconstructed mesh of the boxes and texture mapped view is shown in fig. 3. As can be seen from the pictures, the algorithm is correct in estimating the different depth of the two boxes from the camera.

The reconstructed scene together with the local model estimated in factorization algorithm in texture mapped format is shown in fig. 3. We converted the result into VRML format so that we can examine and compare more closely

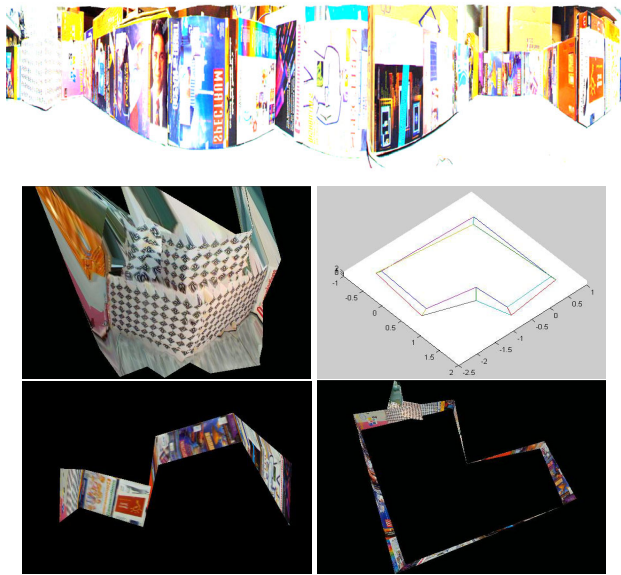


Figure 3: Real image data to simulating a L-corridor with a protruding feature at one of the end of the corridor (at left end of the image). Top: The panorama of the scene. The observation point i.e. the position of the camera is at the interior of a L corridor. The recovered scene (from middle to bottom). First: Textured view of the reconstructed protruding boxes from the corridor, second: Wireframe view of the reconstructed corridor from our algorithm. Bottom two images: perspective views of the final scene. The final result is shown in VRML model format.

the reconstructed result with the actual one. The estimated model of the corridor showed that the model is a bit slanted, which is possibly due to the mounting of the camera not at the same axis with the ground. We have also measured the ratio of the length of sides of the corridor and found that the ratio in the estimated model agreed with the actual values – average error within 5% of the original.

Although it is found that the reconstructed model not exactly agree with the actual one, we argue that it has already achieved the goal of model recovery. The reason is that using the estimated result, one can edit the model using commercial modeling packages such as *3D studio*. One of the advantages is that our system already provide the textures and the 3D model required in virtual environment authoring.

During the reconstruction of the scene above, we found that the incorporation of local details into the global structure presented some challenges. The problem is that the estimated model from the factorization algorithm is having a different scale with that of global recovered data. The integration of the two models thus required careful alignment and scaling of the two models. Automatic techniques to estimate the pose and location of the local structure inside the global one is another direction which we are now investigating. In addition, it would be more convenient for the system to identify the corners and their interconnections from the panorama so that the resulting system can suggest or even automatically carry out the recovery process for the user.

5 Conclusions

We propose a method to perform scene recovery with fine details. By using a global estimation algorithm derived from Marill's algorithm, we can build a 3D model of a scene from its panorama. Structure from motion algorithm can thus extract the fine details of selected portions of the scene. We demonstrated the effectiveness of our algorithm by applying it successfully in both synthetic and real image data. The propose algorithm can be used in various virtual environment authoring application such as computer game production.

Acknowledgments

The work described in this work was supported by a grant from the Research Grant Council of Hong Kong Special Administrative Region. (Project Number. CUHK4389/99E).

References

- [1] S. Birchfield. Klt: An implementation of the kanade-lucas-tomasi feature tracker. <http://vision.stanford.edu/birch/klt/>.
- [2] S. E. Chen. Quicktime vr - an image-based approach to virtual environment navigation. *Proceedings of SIGGRAPH 95*, pages 29–38, Aug. 1995.
- [3] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems. *Proc. Sixth European Conference on Computer Vision, Dublin, Ireland., 2000*.
- [4] S. B. Kang and R. Szeliski. 3-d scene data recovery using omnidirectional multibaseline stereo. *Intl J. Computer Vision*, 25(2), 1997.
- [5] Y. Leclerc and M. Fischler. An optimization-based approach to the interpretation of single line drawings as 3d wire frames. *IJCV*, 9(2):113–136, November 1992.
- [6] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. *Proc. of Eurographics*, 1999.
- [7] T. Marill. Emulating the human interpretation of line-drawings as three-d imensional objects. *IJCV*, 6:147–161, 1991.
- [8] S. Rusinkiewicz, O. H. Holt, and M. Levoy. Real-time 3d model acquisition. *Proc. SIGGRAPH'2002*, pages 433–442, Aug. 2002.
- [9] H.-Y. Shum, M. Han, and R. Szeliski. Interactive construction of 3d models from panoramic mosaics. *IEEE Proc. Computer Vision and Pattern Recognition (CVPR98)*, pages 427–433, June 1998.
- [10] P. Sturm. A method for 3d reconstruction of piecewise planar objects from single panoramic images. *Proc. CVPR Workshop on Omnidirectional Vision (OMNIVIS) Hilton Head Island, SC*, pages 119–126, June 2000.
- [11] R. Szeliski and H.Y.Shum. Creating full view panoramic image mosaics and texture-mapped models. *Proc. SIGGRAPH'97(Los Angeles)*, pages 251–258, Aug. 1997.
- [12] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Intl J. Computer Vision*, 9(2):137–154, 1992.