A Robust Method for Recovering Geometric Proxy from Multiple Panoramic Images

Ada S. K. Wan [1] *Angus M. K. Siu* [1] *Rynson W. H. Lau* [1,2] *C.W. Ngo* [1]

[1] Department of Computer Science, City University of Hong Kong, Hong Kong
[2] Department of CEIT, City University of Hong Kong, Hong Kong

ABSTRACT

The use of multiple panoramic images for walkthrough applications is attracting more attention in recent years. The geometric proxy, which can be recovered from wide-baseline images, can be used to reduce the sampling rate. However, correspondence matching across widely separated panoramic images requires searching a large disparity range, which can significantly increase the number of false matches. In addition, 3D point estimation along the antipodal direction of panoramic images is unreliable due to small vergence angle. Moreover, existing methods only recover 3D points, which do not provide information about object continuity in the scene.

In this paper, we propose a robust method for recovering geometric proxy of a scene from multiple wide-baseline panoramic images. Correspondences are matched under the epipolar constraint. A robust algorithm is developed to remove false matches and estimate a set of reliable sparse 3D points. We further recover information about object continuity with topology and pattern similarity checking.

1. INTRODUCTION

As panoramic images have the advantage of wide field-of-view (FOV), their application in image-based rendering (IBR) for walkthrough is becoming popular [11, 14]. However, a major practical problem in developing this kind of applications is the high manual effort in image acquisition and/or scene modeling. Image acquisition effort can be lowered by reducing the spatial sampling rate with the use of geometric information (geometric proxy) [3]. The most cost effective method to obtain the geometric proxy is recovering them automatically from the source images.

Geometric information recovery and 3D reconstruction from videos or image sequences has been extensively studied [5, 10]. Since consecutive views in a video sequence are closely-spaced, both the search range and the effect of occlusion are usually small. Hence, correspondences can be matched and dense 3D points can be obtained easily. However, if the recovered geometric proxy is to help reduce the sampling rate, the baseline between adjacent images is expected to be large. Then, existing 3D reconstruction methods cannot be used here as they assume that the input images are closely-spaced.

Recovering geometric information from panorama is in fact a challenging task. As shown in [12], 3D reconstruction with a pair of panoramic images yields very biased reconstruction, as the 3D points estimated from correspondences are distributed highly non-uniformly. Since the vergence angle α for 3D point estimation is small near the antipodal direction, the depth uncertainty d gets higher and the precision of the 3D point becomes low as shown in Figure 1(a). Meanwhile, for the region near the perpendicular direction to the image baseline, the disparity range β to be search becomes large, which may significantly increase the number of false matches as shown in Figure 1(b).

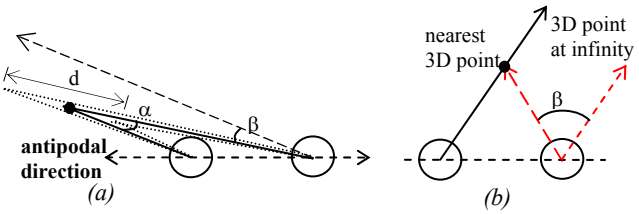


Fig. 1. *Two problems in correspondence matching of wide-baseline panoramic images: (a) Small vergence angle α for 3D point estimation near the antipodal direction; (b) Large search range β near the perpendicular direction to the image baseline.*

In order to improve the precision, multiple panoramic images can be used for 3D reconstruction. [9] recovers 3D points by minimizing an error function of correspondences in multiple images. Without the need to search for correspondences in the image domain, [2] samples several 3D positions along the viewing direction in spatial domain for each feature point. It projects the sampled 3D points to all images and finds one with maximal similarity. However, both methods aim at 3D point reconstruction only and neither identify individual 3D objects nor recover object discontinuity.

In this paper, we propose a robust method to recover geometric proxy for a scene from multiple wide-baseline panoramic images. Apart from the estimated 3D points, the geometric proxy also contains information about object continuity in the scene. The main contributions of this paper are:

0-7803-8554-3/04/$20.00 ©2004 IEEE.

- ***Robust 3D point estimation from multiple panoramas:*** We have developed a method to estimate reliable 3D points from multiple wide-baseline panoramic images. Cylindrical epipolar constraint is established for guided matching, which reduces the search range and hence improves the correctness in correspondence matching. A robust method is included to remove mismatches while estimating 3D points from correspondences in different views.
- ***Object continuity information recovery:*** A method is proposed to construct globally consistent 3D unstructured meshes and to recover information about object continuity with topology and pattern similarity checking. We derive the plane homography for projective transform in similarity checking from the estimated 3D points. This checking helps further identifying and removing mismatches.

The rest of this paper is organized as follows. Section 2 describes our geometric proxy recovery method in detail. Section 3 presents some experimental results. Section 4 briefly summarizes the work presented in this paper.

2. GEOMETRIC PROXY RECOVERY

2.1 Overview

Geometric proxy is the geometric information that can be used to assist in the reconstruction of a desired view. In this paper, we recover an occlusion-adaptive geometric proxy, which is called ROAM [13]. Instead of a 2D standard mesh which enforces continuity of motion across the whole image, in ROAM, an image $\boldsymbol{I}_j$ is segmented into a set of 3D and 2D meshes. Matched patches $\mathbf{M}$ are considered as globally consistent 3D surfaces which can be visible in majority of images. Unmatched patches $\mathbf{U}_j$ may represent occluded regions. They are treated as 2D Graphical Objects (GO) [8] and defined adaptively in each image. Here, we define ROAM as:

$$\text{ROAM} = \{ \boldsymbol{I}_j (\mathbf{M}, \mathbf{U}_j, f) \mid j = 1...N\}$$
$$f : (\mathbf{M} \cup \mathbf{U}_j) \rightarrow \mathsf{C}$$

where $\mathbf{M} \subset \Re^3$ and $\mathbf{U}_j \subset \Re^2$. f is the attribute function mapping the patches to the color space C.

Our geometric proxy recovery method can be divided into 2 steps: 3D point estimation and mesh recovery. First, 3D positions of feature points are estimated robustly with multiple images. Second, common regions which are visible in several views are recognized and merged to obtain a set of globally consistent 3D meshes. For those unmatched regions, meshes are defined adaptively for each image.

2.2 3D Point Estimation

2.2.1 Cylindrical Epipolar Constraint for Matching

First, we estimate the camera pose with a method similar to [1]. Feature points are then extracted from images with the Harris operator [6] and matched with adjacent images. In order to improve the correctness in correspondence matching and reduce the computational time, we establish a cylindrical epipolar constraint between the image spaces of adjacent images for guided matching. By applying the constraint, the search range can be reduced from a 2D search window to a 1D epipolar curve *e*.

A feature point $\mathbf{p}^j=[u^j,v^j]^\mathsf{T}$ in the j^{th} image is first transformed to its projective ray $\mathbf{d}^j=[s,t,r]^\mathsf{T}$ by a function g^j, i.e., $g^j(\mathbf{p}^j) = \mathbf{d}^j$. This allows the linear mapping $\mathbf{E}^{kj}$ to be established between views j and k as follows:

$$g^k(\mathbf{p}^k)^\mathsf{T}\mathbf{E}^{kj}g^j(\mathbf{p}^j) = 0 \quad (1)$$

$$g\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \begin{bmatrix} f\, sin(u / f) \\ \tau v - h/2 \\ -f\, cos(u / f) \end{bmatrix} = \begin{bmatrix} s \\ t \\ r \end{bmatrix} \quad (2)$$

where $\mathbf{E}^{kj}$ is a 3x3 Essential matrix, f is the focal length, h is the image height, and τ is the skew factor. From (1), for a given $\mathbf{p}^j$, its corresponding point in image k should satisfy the following constraint:

$$g^k(\mathbf{p}^k)^\mathsf{T}\mathrm{M}^{kj} = 0 \quad (3)$$

where $\mathbf{M}^{kj} = \mathbf{E}^{kj}g^j(\mathbf{p}^j) = [\lambda_1, \lambda_2, \lambda_3]^\mathsf{T}$. By substituting (2) into (3), the corresponding epipolar curve e in the k^{th} image, induced by $\mathbf{p}^j$, can be expressed as follows:

$$v^k = \frac{h^k}{2\tau} + \frac{f^k}{\lambda_2 \tau}[\lambda_2 cos(\frac{u^k}{f^k}) - \lambda_1 sin(\frac{u^k}{f^k})] \quad (4)$$

Here,we use the Zero-mean Normalized Cross Correlation (ZNCC) as the similarity measurement function because it is unaffected by the overall change of luminance.

2.2.2 Robust 3D Point Estimation from Multiple Images

To avoid biased reconstruction, a set of N-view correspondences $\mathbf{S}_i = \{\mathbf{p}_i^j \mid j = 1..N\}$ is obtained through the correspondence matching. The corresponding 3D point $\mathbf{X}_i$ in the spatial domain can be estimated by obtaining the least square solution in the *linear triangulation* method [7] with $\mathbf{S}_i$, i.e., $\mathbf{X}_i = \boldsymbol{G}(\mathbf{S}_i)$. However, if there is any mismatch in $\mathbf{S}_i$, $\mathbf{X}_i$ would be incorrect. If more images are involved, mismatches are more likely to appear and the correctness is affected.

To improve the correctness, we apply RANdom SAmple Consensus (RANSAC) to robustly estimate the 3D points. A subset of correspondences is randomly selected in each trial and a candidate 3D point $\mathbf{X}_i$' is estimated from them. $\mathbf{X}_i$' is projected onto the images. $\delta^j = \|\mathbf{p}_i^j - \mathbf{p}_i^{j'}\|$, the distance between the original point $\mathbf{p}_i^j$ and the projected point $\mathbf{p}_i^{j'}$ for each view j is calculated. If δ^j is smaller than a threshold ɜ, the point is considered as inlier. After N trials, the largest inlier set $\mathbf{S}_i^k$ is used to re-estimated the final $\mathbf{X}_i$.

2.2.3 Re-matching

Outliers due to mismatch in some images are removed during 3D point estimation. After the final 3D point $\mathbf{X}_i$ is

computed, it is projected onto all images. Correspondence is searched again within a small window centered at the projected image point. Correspondence may not necessarily be found in this re-matching process due to occlusion.

2.3 Mesh Recovery

Dense 3D point reconstruction is difficult due to occlusion, insufficient texture and noise, while sparse 3D points do not represent object continuity. In this step, we use sparse 3D points together with triangulation and checking to identify globally consistent 3D meshes from occluded / non-rigid 2D Graphical Objects. With these operations, object continuity information is implicitly recovered.

2.3.1 Delaunay Triangulation and Patch Checking

We apply Delaunay triangulation on the correspondences of the reference images. A set of unstructured triangular patches are produced. Since some correspondences may be missing in some of the images, corresponding patches may not always be found in all images. To determine whether corresponding patches are matched or not, we apply two checking steps:

- ***Topology Checking:*** This is to ensure that patches in the same image do not overlap with each other and that each patch does not flip with respect to its corresponding patches in other images.
- ***Pattern Similarity Checking:*** Corresponding patches capture the same object surface from different views and hence they should contain similar pattern. Using the three 3D vertices $\mathbf{X}$ of a triangular patch and the camera pose, a plane homography $\mathbf{H}^{kj}$, which maps a pixel within a patch from view j to view k $g^k(\mathbf{p}^k)^\top=\mathbf{H}^{kj}g^j(\mathbf{p}^j)$. $\mathbf{H}^{kj}$ can be determined as :

$$\mathbf{H}^{kj} = \mathbf{R}^{k\top}(\mathbf{I} + \frac{\tilde{\mathbf{\Pi}}^\top(\mathbf{T}^k - \mathbf{T}^j)}{\Pi_4 + \tilde{\mathbf{\Pi}}^\top\mathbf{T}^j})\mathbf{R}^{j\top} \quad (5)$$

where $\mathbf{\Pi} = \begin{pmatrix} (\tilde{\mathbf{X}}_1 - \tilde{\mathbf{X}}_2)\times(\tilde{\mathbf{X}}_1 - \tilde{\mathbf{X}}_3) \\ -\tilde{\mathbf{X}}_3^\top(\tilde{\mathbf{X}}_1 \bullet \tilde{\mathbf{X}}_2) \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{\Pi}} \\ \Pi_4 \end{pmatrix}$ and $\mathbf{I}$ is 3x3 identity matrix. With $\mathbf{H}^{kj}$, pixels in corresponding patches are warped onto the same plane for similarity checking. We adopt a distortion function — Pel Difference Classification (PDC) [4] as measurement to determine whether two patches are matched.

2.3.2 Merging of Globally Matched Patches

If a triangular patch is matched, it will be merged with the matched meshes. Then, checking will be performed on its neighboring patches, until all patches are checked. This operation produces globally matched meshes.

2.3.3 2D Mesh Recovery for Occluded Regions

After recovering the matched meshes, there may still be areas that cannot be matched among the images. They may represent occluded regions and are treated as 2D graphical objects. We construct meshes adaptively for these regions in each image. First, the globally matched meshes are projected to every image to get the locally matched meshes. Unmatched meshes are built by using the bounding edges via Edge-constraint Delaunay triangulation.

3. EXPERIMENTAL RESULTS

We have implemented our method and tested it with real scenes on a Pentium 4 2.4GHz PC. Table 1 shows the computational cost of the method when applied to seven panoramic images, each at a resolution of 4500x450. We can see that the total processing time of the method is around one hour, which is more efficient than the dense matching methods. Table 2 further compares the performance in correspondence matching with and without epipolar constraint. We can see that the matching time with epipolar constraint is significantly faster, because the search range is reduced from 2D to 1D.

Table 1. Performance of our method.

Process	**Time (mins)**
Correspondence Matching	14.41
3D Point Estimation	1.40
Global Mesh Recovery	34.65
2D Mesh Recovery	14.77
Total	65.23

Table 2. Performance of correspondence matching.

Process	**Time (mins)**
Without epipolar constraint	52.33
With epipolar constraint	14.41

Figure 2 shows an experimental image of a circular space bounded by a wall. Majority of the feature points are expected to lie within the wall region and a small amount may lie within the entrance region. Figure 3 shows the results obtained with different 3D point estimation methods. In Figure 3(a), the 3D points are

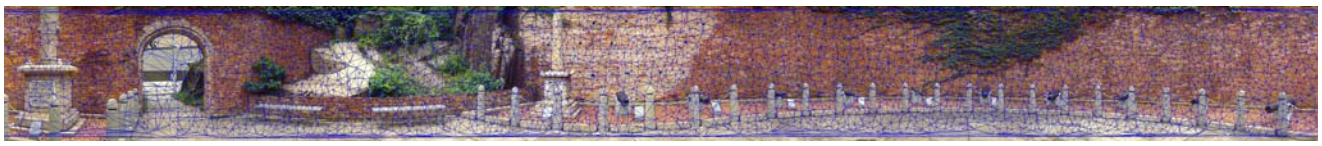
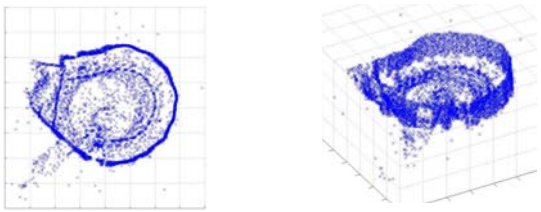
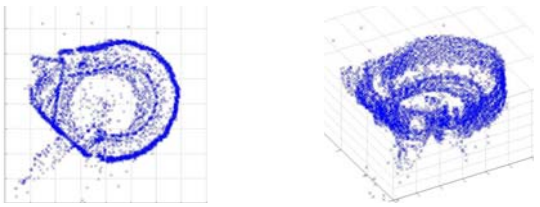


Fig. 2*. The triangular patches on the panoramic images.*

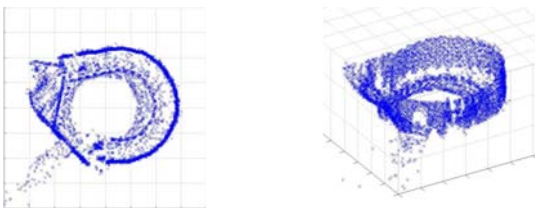
estimated from correspondences which are matched through 2D searching. We observe that the shape of the wall is roughly recovered (the outer circle shown in figure 3(a)). However, there are also some unexpected 3D points appearing outside the wall and around the center. These are obviously mismatches. On the other hand, the cluster of points outside the wall lying at the bottom left hand corner are not mismatches. They are the points which can be seen through the entrance. In Figure 3(b), the 3D points are obtained from correspondences matched under epipolar constraint. We note that the amount of mismatches is reduced. For example, there are fewer 3D points appeared around the center. In Figure 3(c), the 3D points are obtained from our method with epipolar constraint. We can see that most of the 3D point can be estimated reliably. There are no 3D points located around the center nor outside the wall. In addition, the shape of the wall is well recovered and appears very smooth as shown in the right diagram of Figure 3(c). In Figure 2, we superimpose the recovered triangular meshes on the input panoramic image. Although the scene contains a lot of repetitive patterns, which increases the chance of mismatch, our method addresses this problems as well as compared with existing methods.



(a) minimizing all correspondences without applying epipolar constraint



(b) minimizing all correspondences with epipolar constraint



(c) applying our method for removing mismatches and estimating 3D points with epipolar constraint

Fig 3. *3D point positions of the scene.*

4. DISCUSSIONS

In this paper, we have proposed a robust method for recovering geometric proxy from wide baseline panoramic images. We have established the cylindrical epipolar constraint for correspondence matching, which improves accuracy while reduces computation time. We use RANSAC to estimate 3D points from correspondences robustly. It successfully removes most false matches and obtains reliable 3D points. Information about object continuity is also recovered through topology and pattern similarity checkings.

ACKNOWLEDGEMENTS

The work described in this paper was partially supported by a CERG grant from the Research Grants Council of Hong Kong (RGC Reference Number: CityU 1308/03E).

REFERENCES

[1] M. Antone and S. Teller, "Scalable Extrinsic Calibration of Omni-Directional Image Networks," *IJCV*, **49**(2/3):143-174, Sept./Oct. 2002.

[2] R. Bunschoten and B. Kröse, "3-D Scene Reconstruction from Cylindrical Panoramic Images," *SIRS*, pp. 199-205, 2001.

[3] J. Chai, X. Tong, S. Chan, and H. Shum. "Plenoptic Sampling," *Proc. ACM SIGGRAPH*, pp. 307-318, 2000.

[4] E. Chan, A. Rodriguez, R. Ghandi, and S. Panchanathan, "Experiments on Block-matching Techniques for Video Coding," *Multimedia Systems*, **2**(3):228-241, Dec. 1994.

[5] O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin, and I. Zoghlami, "3-D Reconstruction of Urban Scenes from Image Sequences," *CVIU*, **69**(3):292-309, Mar. 1998.

[6] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Conf. on Alvey Vision*, pp. 147-151, 1988.

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

[8] G. Joans, V. Luiz, C. Bruno, and D. Lucia, *Warping and Morphing of Graphical Objects*, Morgan Kaufmann, 1998.

[9] S. Kang and R. Szeliski, "3-D Scene Data Recovery using Omnidirectional Multibaseline Stereo," *Proc. IEEE CVPR*, pp. 364-370, 1996.

[10] M. Pollefeys, R. Koch, M. Vergauwen and L. Gool, "Automated Reconstruction of 3D Scenes from Sequences of Images," *ISPRS*, **55(**4):251-267, 2000.

[11] H. Shum, M. Han, and R. Szeliski, "Interactive Construction of 3D Models from Panoramic Mosaics," *Proc. CVPR*, pp. 427-433, 1998.

[12] H. Shum, A. Kalai and S. Seitz, "Omnivergent Stereo," *ICCV*, pp.22-29, 1999.

[13] A. Siu and R.W.H. Lau, "Relief Occlusion-Adaptive Meshes for 3D Imaging," *Proc. ICME*, 2003.

[14] S. Teller, "Automated Urban Model Acquisition: Project Rationale and Status," *IUW*, pp. 455-462, 1998.