MULTIPLE-VIEWPOINT IMAGE STITCHING

Kai-Chi Chan and Yiu-Sang Moon
Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Keywords: Image stitching, Feature matching.

Abstract: A wide view image can be generated from a collection of images. Its field of view can be expanded as much as to capture a 360° scene. Common approaches, like panorama, mosaic, assume all source images are taken at the same camera center by pure rotation. However, this assumption limits the quality and feasibility of the generated images. In this paper, the problem of generating a wide view image from multiple viewpoint images is formulated. A simple and novel way is proposed to loosen the single viewpoint constraint. A wide view image is generated by 1) transforming images from different viewpoints into a unified viewpoint using SIFT feature matching, etc; 2) stitching the transformed images together by overlapping. Test results demonstrate that the proposed method is an efficient way for stitching images from different viewpoints.

1 INTRODUCTION

To synthesize a wide view image, a common approach is to create a panorama (Brown and Lowe, 2003). However, images have to be taken at the same camera center to create a realistic view. There are two common ways to generate a panorama. In the first way, the images are taken by a panning camera. However, the scene captured has to be static so that images taken by the panning camera are consistent. In the second way, multiple cameras with different camera centers are used to create a wide view dynamic scene. In this case, the scene captured can be dynamic but the multiple-viewpoint panoramic image will be blurred or distorted because the image transformation (rotation and translation) in the panorama generation is assumed to be constant among pixels. This assumption is invalid on different camera centers.

In this paper, a novel method is proposed to stitch images from different viewpoints. First, images from different camera centers are transformed to a reference camera center. Then, a wide view image is generated by overlapping them together. Also, we improve the quality of panorama by inputting the transformed images sharing the same viewpoint. We essentially 1) formulate the relationships of image pairs from different viewpoints; 2) derive a self-verification method for SIFT feature matching; 3) derive a novel method to stitch images from different viewpoints together.

2 RELATED WORKS

At Stanford University, large camera arrays (Wilburn et al., 2005) created photographs from combinations of images taken by a number of cameras. The cameras were placed closely to each other so that the images taken by each camera were assumed to share a single center of projection. The parallax was compensated using software. Although the computation of depth information can be saved, the configuration of cameras is limited and there is a lot of redundant information among their images. Peleg, et al. (Peleg and Herman, 1997) proposed a manifold projection method to overcome the difficulties caused by arbitrary camera motion. In the manifold projection, only 1-D strips in the center of images were used for image alignment and panorama creation. In this approach, the transformations among 1-D strips were still fixed when the camera motion involved translation and rotation. However, other strips not in the center of images were ignored and wasted. Also, the scene being captured should be static and the camera motion should be continuous.

Chan K. and Moon Y. (2010).
MULTIPLE-VIEWPOINT IMAGE STITCHING.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 190-193
DOI: 10.5220/0002893401900193
Copyright © SciTePress

3 MULTI-VIEW PERSPECTIVE PROJECTION

The mapping between a point I (I_x, I_y) on the image and a point W (W_X, W_Y, W_Z) in the 3D world is given by

$$s\begin{bmatrix} I_x \\ I_y \\ 1 \end{bmatrix} = \begin{bmatrix} -f_u & 0 & u_0 \\ 0 & -f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} W_X \\ W_Y \\ W_Z \end{bmatrix}, \quad (1)$$

where (f_u, f_v) is the focal length, (u_0, v_0) is the camera center and s is a scaling factor.
Let $\mathbf{P}$ be the projection matrix. Equation (1) can be rewritten as

$$s\vec{I} = \mathbf{P}\vec{\mathcal{W}}$$

Subscripts are now added to the above equation to indicate the viewpoint of an image. If the relationship between two different viewpoints are given by

$$\vec{\mathcal{W}}_1 = \mathbf{R}\vec{\mathcal{W}}_0 + \vec{\mathcal{T}}, \quad (2)$$

where $\mathbf{R}$ is a rotation matrix and $\vec{\mathcal{T}}$ is a translation vector. After simplifying the equations, images can be transformed to another viewpoint by transforming every pixel $\vec{I}_0^i$ in images by

$$s_1^i\vec{I}_1^i = s_0^i\mathbf{P}_1\mathbf{R}\mathbf{P}_0^{-1}\vec{I}_0^i + \mathbf{P}_1\vec{\mathcal{T}} \quad (3)$$

4 IMAGE SYNTHESIS

In our approach to stitch images, images captured from different viewpoints are transformed as if they are captured from a unified viewpoint. This can be achieved by first computing the relationships (rotation and translation) among them. Then, the scaling factor is estimated in every pixel. Finally, pixels from one image are warped to another image captured from the unified viewpoint.

4.1 Viewpoint Correspondence

Viewpoint correspondence can be treated as transformations among 3D coordinate systems of cameras. Images from different viewpoints correspond to cameras located in different places and directions. In this way, the viewpoint relationship between image pair is the transformation between 3D coordinate systems of two cameras. As a result, the viewpoint correspondence problem can be solved by estimating the transformation of each camera pair in a 3D coordinate system. By taking each camera as the center of its 3D coordinate system, the rotation and translation between camera pair can be estimated using stereo camera calibration technique as in (Zhang, 1999) (Heikkila, 2000).

4.2 Scaling Factor Estimation

The scaling factor of a pixel in Equation (1) is the depth of the corresponding object in the 3D world. The depth of an image can be computed by single view modeling (Criminisi, 2002) or stereo triangulation (Davis et al., 2003).

Given a pair of pixels from two stereo images, the depth of the corresponding object can be estimated by stereo triangulation. An automatic depth estimation process can be achieved by automatically marking features from two different images and matching them. In this paper, SIFT (Lowe, 2004) features are used because they are invariant to rotation. They also provide robust matching to the change in 3D viewpoint and illumination. The matching process is implemented using K-d tree (Bentley, 1990) to reduce the searching time.

In the matching process, two features from an image pair are matched if the Euclidean distance between them is the shortest among other feature pairs. As mismatches among features may exist, to increase the robustness of the matching process, Feature Matching Verification, which will be explained later, is adopted in our work.

After the depth of every feature is computed, the depth of the remaining pixels can be computed through interpolation (Lee and Schachter, 1980) (Boissonnat and Cazals, 2002) (Lee and Lin, 1986). Then, they can be put together in a matrix (scale map) with the dimension equal to that of the image. Practically, the errors from depth estimation and interpolation can be reduced by smoothing using a Gaussian filter (Shapiro and Stockman, 2001), with the assumption that the change of depth between adjacent pixels is smooth or continuous.

4.2.1 Feature Matching Verification

To increase the accuracy of feature matching, two additional rules are added after the Euclidean distance matching.

1. The depth estimated ($\mathcal{D}$) is set to be bounded by predefined minimum ($\mathcal{M}_{min}$) and maximum ($\mathcal{M}_{max}$) values determined from the information of the scene.

$$\mathcal{M}_{min} < \mathcal{D} < \mathcal{M}_{max}$$

2. The estimated depth ($\mathcal{D}$) will be passed to Equation (3) to compute the the location of the trans-

formed pixel $(\mathcal{P}_x, \mathcal{P}_y)$. The transformed pixel location should be equal to the corresponding feature location $(\mathcal{F}_x, \mathcal{F}_y)$. Since the estimated depth is not exact, the distance between $(\mathcal{P}_x, \mathcal{P}_y)$ and $(\mathcal{F}_x, \mathcal{F}_y)$ is non-zero and is set to be smaller than a predefined threshold $(\mathcal{T}_x, \mathcal{T}_y)$ determined from the information of the scene.

$$|\mathcal{P}_x - \mathcal{F}_x| < \mathcal{T}_x \quad \& \quad |\mathcal{P}_y - \mathcal{F}_y| < \mathcal{T}_y$$

5 EXPERIMENTS

In our experiments, two cameras (QuickCam® Sphere AF) placed side by side were used. The resolution of each image was 800×600. The cameras were calibrated using Camera Calibration Toolbox for Matlab© (Bouguet).

5.1 Feature Matching Verification

In this experiment, the effectiveness of the Feature Matching Verification method was tested. Two tests were carried out. In these tests, the distance between each object and cameras along Z-axis was at least 60cm away. So, the lower bound of the depth was set to be 60cm (no upper bound). For each feature point on the left image, the distance between the pixel location calculated from Equation (3) and the corresponding matching feature on the right image was calculated. If the distance was larger than 5 pixels (which was found empirically) along X or Y axis on the right image, the matching pair would be eliminated.

The number of feature matching pairs before and after pruning is shown in Table 1. Under sampling, the accuracy of feature matching with confidence level 95% and confidence interval 10% is shown in Table 2.

Table 1: Number of matching pairs before and after verification.

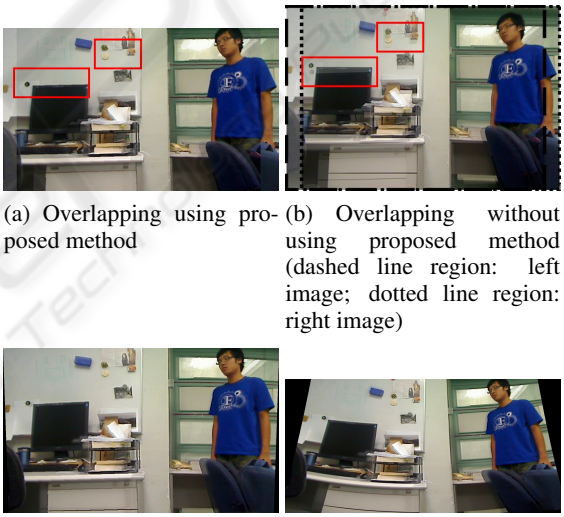
	Before	After
Test 1	1158	476
Test 2	1820	97

Table 2: Matching accuracy before and after verification (No. of sampling).

	Before	After
Test 1	57% (89)	95% (80)
Test 2	41% (91)	88% (49)

5.2 Viewpoint Synthesis

In this experiment, the performance of the proposed synthesis method is examined. The tests were done in two scenarios. In each scenario, the image captured by the left camera was transformed to the image from the right camera viewpoint using the parameter settings in Section 5.1. Then the transformed image and the right image were overlapped and cropped for comparison. Also, the images were used to generate a panorama. The results are shown in Figure 1 and Figure 2. Notice that the image misalignment effect (shown in rectangular boxes) is reduced using the proposed method compared with the overlapping of the source images. This indicates that, after viewpoint synthesis, the viewpoint of the left image is successfully transformed to that of the right image. This approach also reduces the distortion induced by different viewpoints in the panorama.



(a) Overlapping using proposed method

(b) Overlapping without using proposed method (dashed line region: left image; dotted line region: right image)

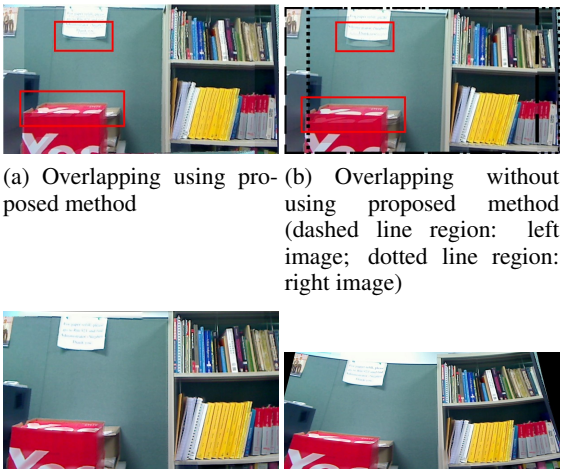
(c) Panorama using proposed method

(d) Panorama without using proposed method

Figure 1: Test 1 synthesis result.

6 CONCLUSIONS

We have presented a novel method to generate a wide view image by stitching individual images from different viewpoints together. The method involves three steps (viewpoint correspondence, scaling factor estimation and image warping). In viewpoint correspondence, images taken from different viewpoints are related to each other by rotation and translation. Then, the scaling factors are estimated by matching the SIFT features in each image pair. Finally, images are synthesized as if they are captured at a single viewpoint.



(a) Overlapping using proposed method

(b) Overlapping without using proposed method (dashed line region: left image; dotted line region: right image)

(c) Panorama using proposed method

(d) Panorama without using proposed method

Figure 2: Test 2 synthesis result.

The feature matching verification which involves depth and pixel locality verifications is derived. It prunes mismatching pairs and is essential in depth estimation.

Although this method eliminates the single viewpoint constraint and can be used not only in static but dynamic scene, there are limitations on our work. The quality of the wide view image depends on the numbers of features in the overlapping regions of image pairs. Also, this method takes longer execution time compared with the panoramic approach because it takes one more step to estimate the depth of every pixel.

The preliminary test results demonstrate that it is a feasible method for wide view image creation. It also improves the feasibility and quality of panorama using multiple-viewpoint images.

ACKNOWLEDGEMENTS

The work described in this paper was substantially supported by a grant from the MPECENG(SEEM).

REFERENCES

Brown, M. and Lowe., D.G. (2003). Recognising Panoramas. In *ICCV'03, 9th IEEE International Conference on Computer Vision*. IEEE Computer Society Press.

Wilburn, B., Joshi, N., Vaish, V., Talvala, E.V., Antunez, E., Barth, A., Adams, A., Horowitz, M., Levoy, M. (2005) High performance imaging using large camera arrays. In *ACM SIGGRAPH'05, ACM Transactions on Graphics (TOG)*. ACM Press.

Peleg, S. and Herman, J. (1997) Panoramic mosaics by manifold projection. In *CVPR'97, Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Press.

Zhang, Z. (1999) Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In *ICCV'99, 7th IEEE International Conference on Computer Vision*. IEEE Computer Society Press.

Heikkila, J. (2000) Geometric Camera Calibration Using Circular Control Points. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE Computer Society Press.

Criminisi, A. (2002) Single-View Metrology: Algorithms and Applications. In *Lecture Notes in Computer Science, 24th DAGM Symposium on Pattern Recognition*. Springer-Verlag Press.

Davis, J., Nehab, D., Ramamoorthi, R., Rusinkiewicz, S. (2003) Spacetime Stereo: A Unifying Framework for Depth from Triangulation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE Computer Society Press.

Lowe, D. G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. In *International Journal of Computer Vision*. Kluwer Academic Press.

Bentley, J. L. (1990) K-d trees for semidynamic point sets. In *Annual Symposium on Computational Geometry, 6th annual symposium on Computational geometry*. ACM Press.

Lee, D. T. and Schachter, B. J. (1980) Two algorithms for constructing a Delaunay triangulation. In *International Journal of Parallel Programming*. Springer Netherlands Press.

Boissonnat, J. D. and Cazals, F. (2002) Smooth surface reconstruction via natural neighbour interpolation of distance functions. In *Annual Symposium on Computational Geometry, 16th annual symposium on Computational geometry*. ACM Press.

Lee, D. T. and Lin, A. K. (1986) Generalized delaunay triangulation for planar graphs. In *Discrete and Computational Geometry*. Springer New York Press.

Shapiro, L. G. and Stockman, G. C. (2001) *Computer Vision*. Prentice Hall.

Bouguet, J. Y. Complete Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib doc/index.html