MVA'98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

1—2 Feature-Based Image Mosaicing

Naoki Chiba, Hiroshi Kano, Minoru Higashihara, Masashi Yasuda, and Masato Osumi
Mechatronics Research Center*
SANYO Electric Co., Ltd.

Abstract

We propose an automatic image mosaicing method that can construct a panoramic image from digital still images. Our method is fast and robust enough to process non-planar scenes with free camera motion.

The method includes the following two techniques. First, we use a multi-resolution patch-based optical flow estimation for making feature correspondences to automatically obtain a homography. Second, we developed a technique to obtain a homography from only three points instead of four, in order to divide a scene into triangles.

Experiments using real images confirm the effectiveness of our method.

1 Introduction

Image mosaicing has become an active area of research in the fields of photogrammetry, computer vision, image processing, and computer graphics. The applications include construction of aerial and satellite photographs, photo editing, creation of virtual environments and image compression.

One conventional method is a cylindrical panorama that covers a horizontal view for creating virtual environments [3]. However, this method limits the camera to horizontal motion around the optical center, forcing the user to carry a tripod.

Several other methods try to avoid this limitation [8, 10, 13, 2], by using a planar projective transformation (homography). Given two images taken from the same viewpoint, or images of a planar scene taken from different viewpoints, the relationship between the images can be described by a linear projective transformation called a homography. There are two types of conventional methods for obtaining homographies. One uses a non-linear minimization framework without any features [8, 10] and the other uses image features such as corners [13, 2]. The problem of minimization-based methods is that they are slow to converge and sometimes require good initial manual estimation. Conventional feature-based methods require a lot of time to compute homographies, because they do not have feature correspondences.

*Address: 1-18-13 Hashiridani, Hirakata, Osaka 573-8534 Japan. E-mail: chiba@mech.rd.sanyo.co.jp

There has only been limited research [9, 6] fo a non-planar scene with unrestricted camera movement. However, these methods are numerically unstable and computationally expensive because they incorporate many variables in a non-linear minimization framework.

We propose a feature-based method which is fast and can deal with non-planar scenes with unrestricted camera motion. Our paper makes two contributions. First, our method is faster at computing homographies than previous methods, because we solve homographies linearly after making feature correspondences. Second, we describe a method for obtaining a homography from only three points, instead of four points in general, by using the epipolar constraint between images. This is for dividing the scene into several planes with triangles when we cannot regard the scene as a planar surface.

The remainder of our paper is structured as follows. After reviewing the Lucas-Kanade optical flow estimation in section 2, we show how to apply it to image mosaicing. Section 3 describes our novel technique for obtaining a homography from three sets of corresponding features. Section 4 presents our experimental results using real scene images. We close with a discussion and ideas for future work.

2 Feature-Based Image Registration

We use small rectangular regions such as corners, what we call point features, to obtain a homography. Here we show how to make correspondences by using optical flow estimation.

2.1 Lucas-Kanade Method

The Lucas-Kanade method is one of the best methods of optical flow estimation, because it is fast to compute, easy to implement, and controllable because of the availability of tracking confidence [1]. When the image brightness of an object is constant during the time periods t and $t + \delta t$, the intensity I at a point $u = (x, y)$ has the following constraint.

$$\frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t}\delta t = 0 \quad (1)$$

The Lucas-Kanade method assumes that neighboring pixels in a small window have the same flow vectors. Then we can choose the displacement vector

$\mathbf{d}$ so as to minimize the residue error E in a small region R, defined by the following equation [7].

$$E = \sum_{\mathbf{u} \in R} [I(\mathbf{u}+\mathbf{d}) - I(\mathbf{u})]^2 \quad (2)$$

Making the linear approximation that $I(\mathbf{u}+\mathbf{d}) \approx I(\mathbf{u}) + \mathbf{d}\frac{\partial}{\partial \mathbf{u}} I(\mathbf{u})$, we can solve the displacement $\mathbf{d}$ with the following equation.

$$\mathbf{d} = \frac{\sum_w g(\mathbf{u})[J(\mathbf{u}) - I(\mathbf{u})]}{\sum_w g(\mathbf{u})^2} \quad (3)$$

where $I(\mathbf{u}) = I(x, y, t)$, $J(\mathbf{u}) = I(x, y, t+\tau)$ and $g(\mathbf{u})$ is the derivative of $I(\mathbf{u})$.

The major drawback of the gradient methods, including the Lucas-Kanade method, is that they cannot deal with large motion. Coarse-to-fine multi resolution strategy can be applied to overcome this problem. A big problem, however, still remains. Low-textured regions have unreliable results. We solved this problem in [4], with a dilation-based filling technique after thresholding unreliable estimates at each pyramid level.

2.2 Overlap Extraction

We need to extract the overlapping region before applying the optical flow estimation, when the overlapping region of images is less than about 50 %. This is because optical flow estimation may fail, even though our flow estimation incorporates a hierarchical multi-resolution technique.

We find a rough displacement by minimizing the sum of square difference (SSD), in the overlapping region, defined by the following equation.

$$E(\mathbf{d}) = \frac{\sum_w (I_1(\mathbf{x}) - I_2(\mathbf{x}+\mathbf{d}))^2}{M \times N} \quad (4)$$

where E is the squared difference of an overlapping region w (size:$M \times N$) between image I_1 and I_2 with the displacement $\mathbf{d}$. We use low-resolution images for faster computation. Figure 1 shows original images and the extracted overlapping region.

2.3 Feature Correspondence

After obtaining the overlapping region of two images, we first select good features in the first image from its image derivative and the Hessian [11]. Next, we estimate sparse optical flow vectors for small rectangular patches by using the improved Lucas-Kanade method described in 2.1. For making a point correspondence, we interpolate the nearest four patch results bi-linearly in sub-pixel order. Figure 2 shows the selected features and corresponding features.

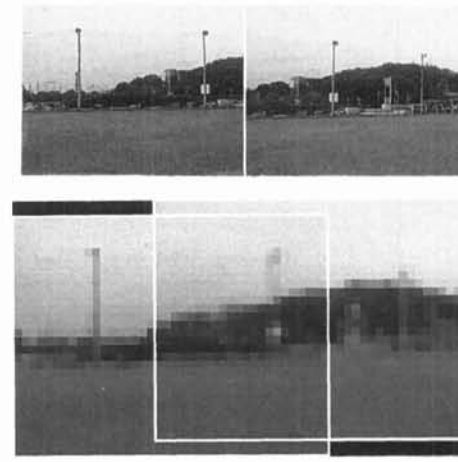


Figure 1: Extracted Overlapping Region

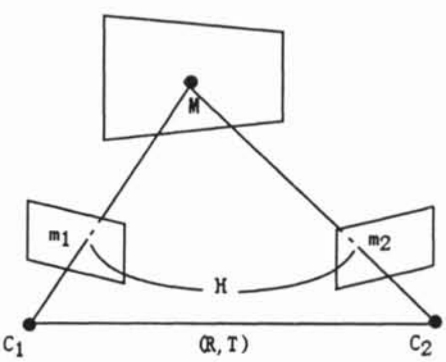


Figure 3: Planar Projective Transformation

2.4 Projective Planar Transformation

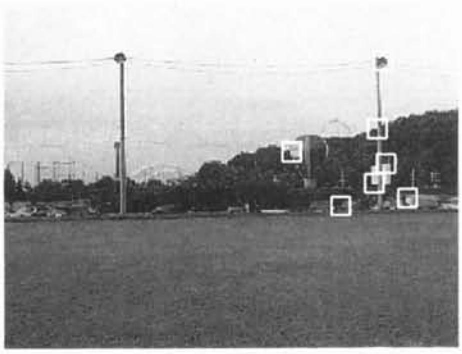
When the scene is a planar surface or when the images are taken from the same point of view, images are related by a linear projective transformation called a homography. Figure 3 illustrates the principal of the planar projective transformation. When we see a point M on a planar surface from two different viewpoints C_1 and C_2, we can transform the image coordinates $\mathbf{m_1}$ to $\mathbf{m_2}$ using the following planar projective transformation matrix $\mathbf{H}$ [5].

$$k\mathbf{m_2} = \mathbf{H}\mathbf{m_1} = \begin{pmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \quad (5)$$

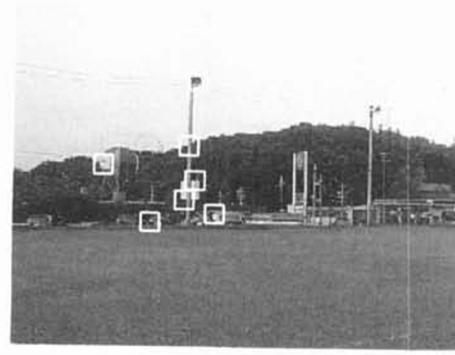
where k is an arbitrary scale factor. This relationship can be rewritten using the following equations.

$$\begin{cases} x_2 = \frac{h_0 x_1 + h_1 y_1 + h_2}{h_6 x_1 + h_7 y_1 + 1} \\ y_2 = \frac{h_3 x_1 + h_4 y_1 + h_5}{h_6 x_1 + h_7 y_1 + 1} \end{cases} \quad (6)$$

When a point on the planar surface is invisible from C_2 but visible from C_1, we can generate the corresponding point on image I_2 by this transformation.



(a) Selected Features



(b) Corresponding Features

Figure 2: Feature Correspondence

Our method solves this matrix faster than previous methods because it uses a least squares method with four or more corresponding points, instead of using a non-linear minimization framework.

3 Non-Planar Scene

We propose a method that can construct a panorama for non-planar scenes with unlimited camera motion. If the assuming condition is not satisfied, i.e., taking images of a non-planar scene from different viewpoints, there will be misregistration caused by motion parallax. Figure 5 (b) shows an example. Our method works on this situation with the following two steps. First, we divide a scene into multiple planes because the single plane assumption has significant errors. Since three points are minimum number of points to consist a plane, we triangulate a scene with corresponding feature points between images by using the Delaunay triangulation.

Second, we obtain the homography for each triangle. However, we need at least four points to obtain a homography since it has eight independent parameters. (It is a 3×3 matrix and is invariant to scaling.) We describe a novel technique to compute a homography from three points using the epipolar constraint.

The epipolar constraint between two images is described by the fundamental matrix $\mathbf{F}$, with a point $\mathbf{m}$ on an image I and the corresponding point $\mathbf{m}'$ on the other image I' as follows:

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0 \quad (7)$$

Recently a good method has been developed to obtain a fundamental matrix between uncalibrated cameras [12]. By using this, we first obtain the fundamental matrix between images from their corresponding points. By substituting equation (5) with (7), we have the following equation.

$$\mathbf{m}^T \mathbf{H}^T \mathbf{F} \mathbf{m} = 0 \quad (8)$$

Table 1: Processing Time

Function	Time (sec.)
Overlap Extraction	2
Feature Matching	3
Warping	3
Blending	2
Total	10

Since $\mathbf{H}^T\mathbf{F}$ means the outer product of vector $\mathbf{m}$, it should be skew symmetric:

$$\mathbf{H}^t\mathbf{F} = \begin{pmatrix} 0 & a1 & -a2 \\ -a1 & 0 & a3 \\ a2 & -a3 & 0 \end{pmatrix} \quad (9)$$

We obtain six equations from (9), because diagonal elements and the additions of skew symmetric elements should be zero. We have already six equations from three sets of corresponding points with (5). We can compute a homography by least-squares with three points and the fundamental matrix, because a total of 12 equations are available for eight unknown parameters.

4 Experiments

This section introduces the results of applying our feature-based techniques to image mosaicing. These images were taken from different viewpoints with a hand-held digital still camera without a tripod. We set a patch size of 13×13 pixels for optical flow estimation and feature correspondences.

4.1 Single Planar Scene

Figures 4, 6 and 7 show the results of a single planar scene. Figure 6 with a size of 640×480 pixels took ten seconds on a Pentium 300 MHz PC to process. Table 1 shows the detail of processing time.

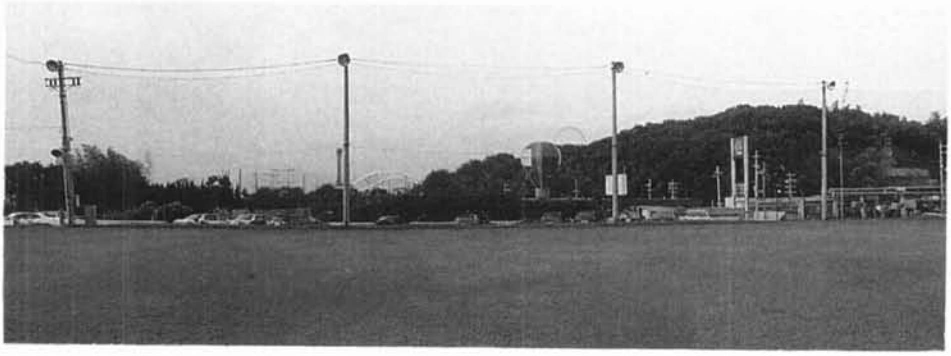


Figure 4: Landscape

The time of feature matching includes obtaining the homography. We blended intensities weighted by the distance from the boundaries of the overlapping region. Figure 7 shows an example of vertical panoramas.

4.2 Multiple Planar Scene

Figure 5 shows the result of a multiple planar scene. Since the scene has a certain amount of depth range, we cannot regard it as a planar surface. The overlapping part of the image is divided into 40 triangles using feature points. The average of absolute intensity differences in the overlapping region has been reduced to 24.0, from 28.5 that using a single planar model.

5 Conclusions

In this paper, we have illustrated two novel techniques for constructing image mosaics for any scene with any camera movement. First, in order to obtain a planar projective transformation for avoiding a limited cylindrical representation, we presented a method based on image features, obtained automatically by using an optical flow estimation. The computational cost is much lower than conventional methods that use a non-linear minimization framework or image features without any correspondences.

Second, for a scene that we cannot assume to be a single plane and with images taken from different viewpoints, we described a method of dividing images into triangles with corresponding features. Our method provides the homography for each triangle from three points using the epipolar constraint, although conventional methods require four points to compute. The technique is fast and robust because it is linear.

We showed that our feature-based method is faster and more robust than previous featureless methods, because it is based on linear techniques. In future work, we plan to develop a display method with which we can feel 3D-depth by stepping out of an optical center.

References

[1] J. Baron, D. Fleet, and S. Beauchemin "Performance of Optical Flow Techniques," Int. J. Computer Vision, Vol. 12, No. 1, pp. 43-77, 1994.

[2] D. Capel and A. Zisserman,"Automated Mosaicing with Super-resolution Zoom," Proc. IEEE CVPR, pp. 885-891, 1998.

[3] S. E. Chen, "QuickTime VR - an image-based approach to virtual environment navigation," SIGGRAPH'95, pp. 29-38, 1995.

[4] N. Chiba and T. Kanade, "A Tracker for Broken and Closely-Spaced Lines," Proceedings of ISPRS International Symposium on Real-Time Imaging and Dynamic Analysis, pp. 676-683, 1998.

[5] O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint," MIT Press, 1993.

[6] R. Kumar, P. Anandan and K. Hanna, "Direct recovery of shape from multiple views: a parallax based approach," Proc. ICPR, pp. 685-688, 1994.

[7] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," In Seventh International Joint Conference on Artificial Intelligence (IJCAI-81), pp. 674-679, 1981.

[8] S. Mann and R. W. Picard, "Virtual Bellows: Constructing High Quality Stills From Video," Proc. IEEE ICIP, pp. 363-367, 1994.

[9] R. Szeliski, "Image Mosaicing for Tele-Reality Applications," CRL Technical Report 94/2, Digital Equipment Corporation, 1994.

[10] R. Szeliski and H. Y. Shum, "Creating Full View Panoramic Image Mosaics and Environment Maps," SIGGRAPH'97, pp.251-258, 1997.

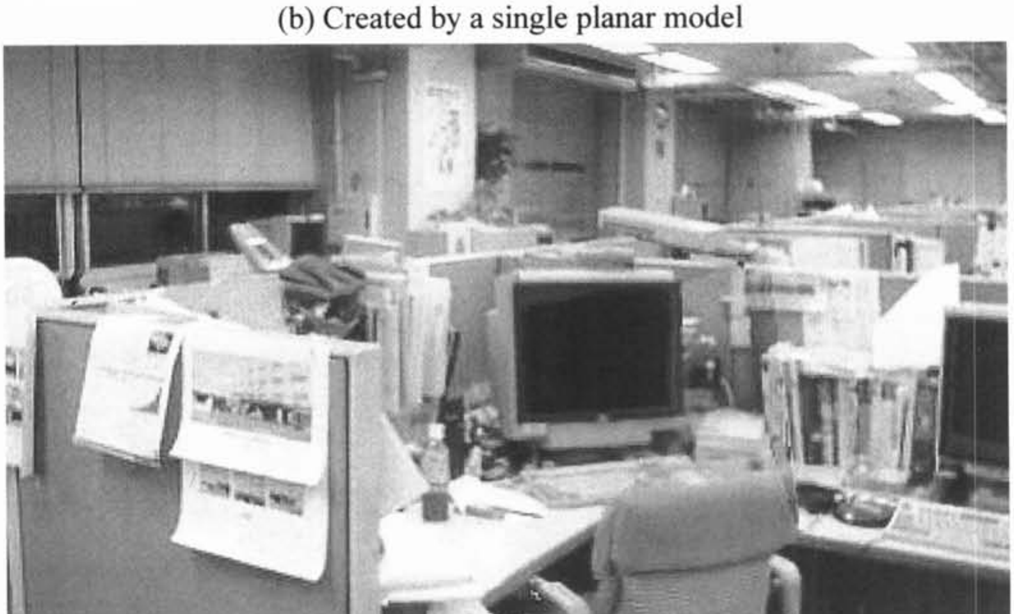
[11] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams: a Factorization method - Part 3 Detection and Tracking of Point Features, " CMU-CS-91-132, Carnegie Mellon University, 1991.

[12] Z. Zhang, "Determining the Epipolar Geometry and its Uncertainty: A Review," Int. Journal of Computer Vision, Vol. 27, No. 2, pp. 161-195., 1998.

[13] I. Zoghlami, O. Faugeras and R. Deriche, "Using geometric corners to build a 2D mosaic from a set of images," CVPR 97, pp. 420-425.

(a) Original images

(b) Created by a single planar model



(c) Created by a multiple planar model

Figure 5: Multi Planar Mosaics

Figure 6: Planar Mosaics (Clock)

Figure 7: Planar Mosaics (Tower)