*MVA2002* IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara - ken New Public Hall, Nara, Japan

10—5

# Mosaics of Japanese Picture Scrolls

Naoki Chiba* and Hiroshi Kano
Digital Systems Development Center
SANYO Electric Co., Ltd.

## Abstract

We propose several new techniques for mosaicing images of Japanese picture scrolls. Previous methods cannot accurately stitch these images, because they are too wide. To solve this problem, we propose three techniques that can stitch without any seams. The first technique adjusts its height of each image in advance. The second technique uses planar projective transformation among three images, not two. These techniques enable us to stitch the images, considering perspective distortion. The third technique warps the image locally in addition to performing global geometric transformation. This technique can handle three-dimensional distortion caused by the roughness of the paper. We show the effectiveness of these techniques by experiments on real images.

## 1 Introduction

Digital archiving has recently been gathering considerable attention. Once we digitally archive historical art objects, such as sculptures [2, 4] or paintings, we can view the objects at anytime, anyplace, and from any viewing direction.

At the same time, image mosaicing has also been gathering attention because it is a useful technique for stitching images together in high resolution. This technique can be used to construct image mosaics of flat paintings or painted walls [5]. Methods for Japanese picture scrolls, however, have not been introduced so far.

Japanese picture scrolls are priceless historical art objects, because they were painted hundreds of years ago. Since they are very fragile, we cannot observe them thoroughly even in a museum. They are usually exhibited only in parts, because they are too fragile to show the entire scroll.

To digitize them, we need to capture several images because we cannot cover the entire scroll with a single image in high resolution. Therefore, it is very important to construct image mosaics of Japanese picture scrolls.

Previous methods cannot accurately stitch these images together, because the scrolls are too wide. Even with planar projective transformation, which is considered to be the most accurate, we cannot eliminate perspective distortion and three-dimensional distortion.

To solve these problems, we propose three techniques to accurately stitch images together for constructing image mosaics of Japanese picture scrolls. These techniques allow us to handle perspective distortion and three-dimensional distortion.

*Address: 1-18-13 Hashiridani, Hirakata, Osaka, 573-8534 Japan. E-mail: mail@naoki-chiba.com

## 2 Problems with Japanese Picture Scrolls

Image mosaicng methods using geometric transformation have already been proposed. Conventional geometric transformations include 2D rigid transformation and affine transformation. Recently, several researchers have proposed the use of planar projective transformation [6]. This transformation is considered to be the most accurate, because it can deal with perspective distortion. The accuracy of these methods, however, is not sufficient for mosaicing images of Japanese picture scrolls. The reason of this comes from the capturing environment, as we will see in the next section.

### 2.1 Capturing Environment

The best way to capture images of Japanese picture scrolls for image mosaicing is to use a special device to restrict the camera motion or projection model. If we can simplify the camera motion or the projection model, the process of image mosaicing becomes easy. For example, we can simplify camera motion by using a camera mounted on a cart on a rail track as in movie shooting. After flattening a picture scroll completely, we can capture a collection of images with this camera. This restricts the camera motion to one-dimensional movement. Another way is to use a flat bed scanner. This restricts the projection model to orthogonal projection, which does not contain any perspective distortion.

These ways of capturing images, however, are not practical in real situations, because Japanese picture scrolls are too fragile and priceless to make

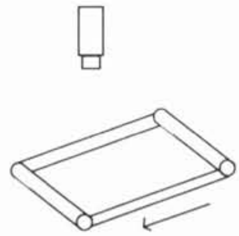

Figure 1: Shooting setting

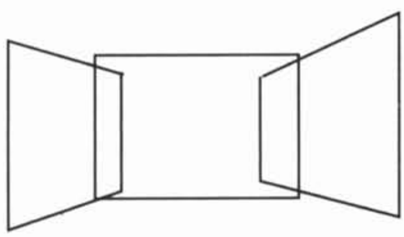

Figure 2: An example using planar projective transform

them completely flat. The typical way of capturing images is shown in Figure 1. The scroll is shown part by part. The camera is mounted on a tripod from a limited distance. The geometric relationship between the scroll and the camera might change, because we need to replace the scroll each time after scrolling it to show another part. Due to this capturing environment, the captured images contain the following distortions.

- rotation around the optical center
- scale change
- perspective distortion

The first two distortions are caused by the replacement of the scroll. The last distortion is caused by perspective projection due to the limited distance between the scroll and the camera.

The worst thing we face is that we are not allowed to capture the images for ourselves, because they are too fragile. We need to use images taken previously, just for recording purposes, not for mosaicing.

### 2.2 Previous Methods

Among the various types of geometric transformations, planar projective transformation is the most accurate method for handling perspective distortion, including image rotation and scale changes. This method, however, has a problem in stitching together Japanese picture scrolls. Since it warps the image, the edges of the mosaics have severe distortion, as shown in Figure 2. Although the horizontal edges of the mosaics should be parallel, they are not very parallel.

This method cannot handle three-dimensional distortion either. In fact, Japanese picture scrolls are not very flat due to the roughness of the paper surface, which is naturally distorted over hundreds of years. Since planar projective transformation assumes that the object is planar, the roughness leads to misregistration

## 3 Stitching Japanese Picture Scrolls Together

To overcome these problems, we propose the following three techniques.

- image height normalization
- planar projective transformation among three images
- local registration

To handle perspective distortion, we adjust the image heights in the preprocess stage. In addition, we use planar projective transformation not with two images, but with three. To deal with three-dimensional distortion, we propose local image registration in addition to global geometric transformation.

### 3.1 Image Height Normalization

To avoid the problems of image rotation, scale changes and perspective distortion, we use the knowledge that the heights of the transformed images should be equal to each other. Based on this knowledge, we normalize each image in the preprocessing stage by performing geometric transformation.

Figure 3 illustrates the general idea of this transformation. Japanese picture scrolls have two horizontal lines that are parallel. In the images, these lines are not very horizontal or parallel because they are distorted by perspective distortion and camera rotation. After selecting these lines, we can compute a geometric transformation to compensate horizontal perspective distortion. In addition, the scales of the images are varied during the capturing process. We compute the geometric transformation to make them equal. These transforms can be written in a single form as follows:

$$I' = NI \tag{1}$$

where $I$ is the original image, $I'$ is the transformed image, and $N$ is geometric transformation. Next, we show how to compute this transform $N$ in detail.

First, we extract the two horizontal lines at the top and bottom boundaries from the images. We

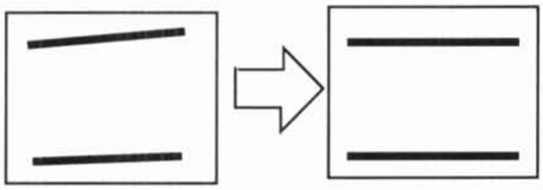

Figure 3: Image height normalization

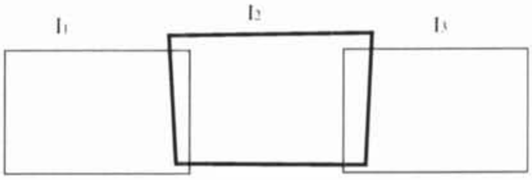


Figure 4: Planar projective transform among three images

manually extract these lines by tracing them with a mouse, because automatic line extraction using image processing is not perfect. Second, we compute the geometric transformation for normalizing images. The transformation has the following three requirements.

- rotation compensation
- perspective distortion elimination
- scale change equalization

All of the requirements can be done by linear geometric transformations. First, we can easily estimate a rotational matrix $R$ to compensate the rotation by measuring the angle of the extracted lines. This transformation can be written as follows:

$$I' = RI \tag{2}$$

where $I$ is the original image and $I'$ is the rotated image.

Second, we compute the geometric transformation for making the two lines parallel. We use the four end-points of the two lines. If we know four pairs of point correspondences between the original image and the transformed image, the geometric transformation can be described by planar projective transformation. When we set the horizontal coordinate of each line equal, the two lines will be parallel. To do this, the transformed coordinates $(u, v)$ of an end-point $p$ can be set by using the horizontal coordinates $v_1$ and $v_2$ of the original end-points as follows:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u \\ (v_1 + v_2)/2.0 \end{pmatrix} \tag{3}$$

By setting all of the four end-points, we can compute the transformation matrix $P$ as follows:

$$I'' = PI' \tag{4}$$

where $I'$ is the rotated image and $I''$ is the transformed image.

Finally, we adjust the scaling of each image. We can do this by setting the distance between the two lines. In this process, we aim at defining the actual image size in addition to scale normalization. If we know the resolution of the devices for image input (camera) and output (printer), we can create a perfect digital copy of Japanese picture scrolls with the actual size. We compute the scaling parameter $r$ by setting the distance $d'$ between the two lines in the transformed image while the original distance is $d$. By using the scaling matrix $S$ having the ratio $r = d'/d$, we can describe the scaling transformation as follows:

$$I''' = SI'' \tag{5}$$

where $I'''$ is the scaled image and $I''$ is the image after making the two lines parallel.

Since the entire transformation is the linear combination of these transformations, we can represent it as follows:

$$I''' = SPRI \tag{6}$$

where $I'''$ is the normalized image and $I$ is the original image.

### 3.2 Planar Projective Transform among Three Images

After normalizing each image, we compute the geometric transformation for stitching the images together. Although the images are normalized, the accuracy of a simple geometric transformation such as 2D translation or affine transformation is not sufficient. This is because the transformed images contain vertical distortion. For example, when the optical axis of the camera is towards the upper direction, the image content of the bottom will be larger than that of the top, because it is closer to the camera. Pair-wise planar projective transformation is not good enough either. Although this transformation considers perspective distortion when mosaicing, it does not eliminate the distortion.

We propose a new technique to use planar projective transformation among three images to maintain the image heights while eliminating vertical perspective distortion. Figure 4 illustrates the general idea. To compute planar projective transformation, we need at least four pairs of

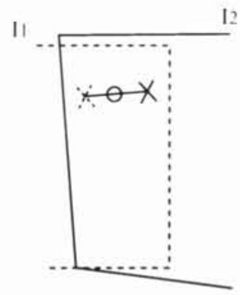


Figure 5: Feature correspondence for planar projective transform using three images

point correspondences in the overlapping region between images. Usually when mosaicing, we find these correspondences between two images. For Japanese picture scrolls, we find them in the two overlapping regions among three images. If we consider the geometric transformation for image $I_2$, we use the overlapping region between $I_1$ and $I_2$ and the overlap between $I_3$ and $I_2$.

To establish point correspondence, we use both an automatic method and a manual method. First, we automatically select prominent point features by computing image derivatives. After that, we automatically find the correspondences using an optical flow technique [3]. Some parts of Japanese picture scrolls have regions with low texture. For these parts, we use manual operation by mouse. After selecting a point with a mouse, we roughly find the corresponding pixel on the other image by hand. The system automatically finds the correspondence in sub-pixel precision by using area-based template matching which is based on normalized correlation.

After establishing point correspondence, we compute the planar projective transformation. We repeat this process for all of the images. At this point, we have some small problems to solve. If we transform an image, this transform will affect the rest of the images. Conventional image mosaicing methods allow this operation to process sequentially. This operation, however, becomes a problem for our purpose because we need to maintain the image heights.

To solve this problem, we describe a new way of setting point correspondences for planar projective transformation among three images. Figure 5 illustrates the idea. We use the intermediate point coordinates as target coordinates, instead of using the point coordinates in the other image. If a point on image $I_1$ is described by a dotted X and its corresponding point on image $I_2$ is X, we set the target point as the middle-point between these points, indicated by the circle, when estimating the transformation matrix both for image $I_1$ and $I_2$. This process will be applied between image $I_2$ and image $I_3$ as well. In this way, each image will be transformed to meet the other images.

One more problem that we need to solve is the operation for the first and the last images of the image mosaics. Since these two images have only one overlapping region with their neighboring images, we need a special treatment. For the side with no overlapping region, we use the two corners of the images themselves as fixed feature points. By adding these feature correspondences, we can compute the transformation matrix for these images as well as the other images.

### 3.3 Local Registration

Even with the techniques described above, some three-dimensional distortions still remain, caused by the roughness of the paper. Although planar projective transformation assumes that the surface of the object is flat, we cannot make picture scrolls very flat. This is because they are too fragile. To solve this problem, we use local registration in addition to the global geometric transformations described above. To do this, we compute the dense optical flow vectors for each pixel in the overlapped regions.

After using the transformed coordinates as the initial position, we perform local registration. We use the Lucas-Kanade optical flow method since it is faster to computer than area-based template matching methods [3]. The pixel transformation consisting of geometric transformation $H$ and its flow vector $d_i$ for pixel $p_i$ can be written as follows:

$$p_i' = Hp_i + d_i \tag{7}$$

where $p_i'$ is the transformed pixel.

Here, we need to solve a problem in optical flow estimation, called the aperture problem. When the image content has no texture or linear, unlike corners, the accuracy of the optical flow becomes poor. We solve this problem by interpolating reliable flow results [1].

### 3.4 Color Adjustment

Since the colors of images are different from each other, we have to adjust for these color differences. We use the following two types of methods.

- blending
- color transformation

Blending is a technique to smooth the color difference in the overlapping region between images. When rendering images, we mix the pixel values of the images according to the weights computed from the distance to the pixel. Color transformation is a pre-processing technique before rendering. If we know the color differences for each pixel,

Table 1: Comparison of pixel value differences

| | |
|---|---|
| Without local registration | 20.2 |
| With local registration | 16.9 |
| Color adjustment and local registration | 5.4 |

we can compute the linear transformation. By using this transformation, we can transform images to compensate the color differences as follows:

$$I' = aI + b \tag{8}$$

where $a$ and $b$ are coefficients, $I$ is the original image, and $I'$ is the transformed image. Since we know the pixel correspondences, we can compute these coefficients by the least squares method.

We thus perform image mosaicing of Japanese picture scrolls by using computed geometric transformations, flow vectors, and color transformation.

## 4 Experiments

We have conducted experiments on real images. The images were taken with a film camera. They were then digitized by a professional-use film scanner. The image size was 10,630 by 6,378 pixels. With this image size, we can print the result with a printer whose resolution is 300 DPI. Figure 6 shows the entire image mosaic. The image size is 114,608 by 5,085 pixels. We can see that the result is good. We cannot recognize any seams, even if we look at the printed result very carefully. The image height is maintained from start to end.

We compared the result by our method with that by a previous method, pair-wise planar projective transformation. Figure 10 shows the comparison. In the result by the previous method, the image height is not maintained and is distorted at the vertical edges of the mosaic. The accuracy in the part on the left side of the cow cart is extremely low. This is because it has low texture and the overlapping area is small. On the other hand, the result by our method is good all through the mosaic. The image height is maintained.

Figure 8 shows the effectiveness of our local registration. In figure 8(a), we can notice the doubled edges caused by registration error. On the other hand, the result with local registration shown in figure 8(b) is very good. Figure 9 shows another example. Figure 9(b) with local registration is much better than figure 9(a) without local registration.

We evaluated the proposed method quantitatively. We measured the average error $E$ of pixel value differences in the overlapped region $w$ by the following error definition.

$$E = \frac{\sum_w |I' - I|}{W} \tag{9}$$

where $W$ is the number of pixels in the overlapped region. The result is shown in Table 1. The error has been reduced by the proposed method from 20.2 to 5.4, which is almost one-fourth.

## 5 Discussion

This paper presents three techniques for stitching together images of Japanese picture scrolls to make a digital copy. We confirmed that we can accurately stitch images together with the proposed techniques. The techniques can eliminate perspective distortion and three-dimensional distortion caused by the roughness of the paper.

The most significant achievement of our method is that we do not have to capture images again for mosaicing purposes. We can use images taken previously for recording purposes. This benefit is very important for digital archives of historical art, because they are often too fragile to allow image capturing.

This time, we did not have to consider radial lens distortion. We might need to handle this distortion with a certain type of optical system. For this purpose, several methods have already been introduced. By using these techniques, we can easily avoid this problem.

## 6 Conclusion

This paper presents the following three techniques to accurately stitch together images of Japanese picture scrolls.

1. Adjustment of the height of each image using knowledge of the scrolls.
2. Planar projective transformation among three images
3. Local registration in addition to global registration

The first two techniques are used to compensate perspective distortion. The third technique is used to handle three-dimensional distortion caused by the roughness of the paper. Experiments showed that the quality of the image mosaic is very good. In the future, we would like to extend these techniques to other types of paintings.

**Acknowledgements**

We are thankful to the people of TOPPAN FORMS CO. LTD. and the Tokyo National Museum for providing the images.

Figure 6: Constructed image mosaic

(a)The constructed mosaic by planar projective transformation

(b)The constructed mosaic by our method

Figure 10: Comparison between a mosaic with planar projection and one with our method.

Figure 7: Raw image

(a) Without local registration

(b) With local registration

Figure 8: Comparison (image: left face)

(a) Without local registration (b) With local registration

Figure 9: Effectiveness of local registration (image: cap)

## References


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

[2] M. Levoy: The Digital Michelangelo Project, Proc. 2nd Internat. Conf. on 3-D Digital Imaging and Modeling, pp. 2-11 (1999).

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

[4] K. Nishino and K. Ikeuchi: Robust Simultaneous Registration of Multiple Range Images, Proc. Asian Conf. on Computer Vision, pp. 454-461 (2002).

[5] W. Puech and A. G. Bors: Mosaicing of Paintings on Curved Surfaces, Proc. Workshops on Applications of Computer Vision, pp. 44-49 (1996).

[6] R. Szeliski: Video mosaics for virtual environment, IEEE Computer Graphics and Applications, pp. 22-30, March (1996).