WIDE RANGE IMAGE SENSING USING A THROWN-UP CAMERA

Toshitaka Kuwa, Yoshihiro Watanabe, Takashi Komuro and Masatoshi Ishikawa

The University of Tokyo
Email: {Toshitaka_Kuwa, Yoshihiro_Watanabe, Takashi_Komuro, Masatoshi_Ishikawa}@ipc.i.u-tokyo.ac.jp

ABSTRACT
In this paper, we propose a wide-range image sensing method using a camera thrown up into the air. By using camera thrown up in this way, we can get images that are otherwise difficult to obtain, such as those taken from overhead. As an example of wide-range image sensing, we integrated video images captured by a thrown-up camera using an image mosaicing technique. When rotation about the optical axis of the camera can be ignored, we can integrate images by mosaicing using a translational approximation, which preferentially pastes pixels around the image center. To obtain the information about the camera direction, a rotational approximation using the angles of incident light rays is required. We also propose use of a high-frame-rate camera (HFR camera) in order to acquire a large amount of information. A seamless large image was obtained by synthesizing the images captured by a thrown-up HFR camera. We found that high frame rates of around 1000 fps were necessary.

Keywords— Image mosaicing, high frame rate camera

1. INTRODUCTION
Although cameras were originally developed to take pictures of people or scenes, they are also widely used in environmental sensing. However, if we try to take pictures while keeping the camera position fixed, only a narrow range of information is obtained. To get more information about the scene, moving cameras are often used instead of stationary cameras, and if we integrate the numerous captured images, more information can be obtained compared with a single image.

One way to integrate multiple images is a technique called 3-D shape reconstruction using structure from motion [1]. In this method, the 3-D shape of the scene and the motion of the camera are estimated at the same time from the captured images. However, if the camera moves with only slight translation, it is difficult to estimate the 3-D shape in principle. Another way to integrate images is image mosaicing [2], which generates a large image by combining a number of small images. Image mosaicing can be applied to video images [3], providing a large image formed by combining the individual frames that the camera captured. Conventionally, it is difficult to generate a seamless image from the video images captured by a fast moving camera since the image displacement between successive frames is large. Digital Route Panorama[4] is a method that can integrate fast moving images. In this method, line images that are orthogonal to the moving direction are obtained at a high frame rate, and a seamless image can be generated by combining the line images. However, this method requires the moving direction of the camera to be known and constant.

Use of a high-frame-rate camera (HFR camera) enables integration of video images that are captured by a camera attached to a high-speed vehicle, or a free-moving camera, which allows a wider range of applications.

In this paper, we propose a wide-range image sensing method using a free-moving HFR camera. The result of image mosaicing is shown as an example of the integration of video images.

2. IMAGE SENSING USING A THROWN-UP CAMERA
As an example of image sensing using a free-moving HFR camera, consider a camera that is thrown up into the air. Images that are otherwise difficult to get, such as those taken from overhead, may easily be obtained by using a camera thrown up in this way, as illustrated in Fig.1. Throwing the camera up while causing it to rapidly spin enables wide-range image sensing of the surroundings. The camera needs to be rotated about an axis other than the optical axis since rotation about the optical axis does not increase the amount of information captured.
With this technique, we can realize wide-range image sensing in a short time. However, if we use a standard camera with a typical frame rate of about 30 fps, the captured images will be blurred. Although we can solve this problem by shortening the exposure time, other problems remain. For instance, feature point tracking is often used to estimate the motion of cameras and to integrate video images, but the correspondence of feature points between frames cannot be estimated correctly if the image displacement is large. We therefore propose to use an HFR camera, which can capture more images in a given time than a standard camera. Using an HFR camera, the image displacement between frames becomes small, and the correspondence of feature points can thus be easily estimated. In addition, dense information may be obtained by integrating more images.

3. INTEGRATION OF IMAGES

As described above, although there are some methods for integrating a number of images, we choose image mosaicing here since the movement of a thrown-up camera is mainly rotational, making 3-D shape reconstruction difficult in principle.

In image mosaicing, it is important to generate a seamless image. In this section, we propose image mosaicing with a translational approximation and image mosaicing with a rotational approximation.

3.1. Image mosaicing with translational approximation

In conventional image mosaicing, the homography matrix between the previous image and the current image is estimated, and the current image is transformed based on the matrix to combine images. We propose a method of combining images on a plane with no transformation by using image mosaicing. A seamless image can be generated because the image change between frames captured by an HFR camera is small. The rotation about the optical axis is ignored, and the rotations about the other two orthogonal axes are approximated by translations. The algorithm is described below.

First, corners are extracted as feature points from the first image by using the Good Features To Track algorithm [5]. Correspondences are then estimated by tracking these points by using the Lucas-Kanade algorithm [6].

Next, the displacement of the camera is estimated from the correspondences of feature points that are tracked. Let \( \Delta x, \Delta y \) be the displacement of the centroid of the tracked feature points. The current image is pasted onto the mosaiced image after it is displaced based on these values. Images are combined on the \( xy \) plane whose origin is at the center of the first image.

\[
\begin{pmatrix}
\Delta x \\
\Delta y
\end{pmatrix} = \frac{1}{N} \sum_{j=1}^{N} \begin{pmatrix}
x_j \\
y_j
\end{pmatrix}
\]  

The generated image becomes distorted if the camera rotates, which is known as perspective projection distortion. Then we use the center pasting technique, which preferentially pastes pixels around the image center where the distortion is small.

Fig. 2. Synthesis of images and distance maps.

The algorithm is described below. First, we use an array whose size is the same as that of the captured images and to which the distance values from the image center are registered, as shown in Fig. 2(a). We call this array a distance map. In image mosaicing, each value of the distance map of the mosaiced image is compared with that of the distance map of the current image in the overlapped region. The pixel values of either the current image or the mosaiced image are used depending on the value of the distance map, as shown in Fig. 2(b). At the same time, the distance map is updated as shown in Fig. 2(c). Repeating this process, pixel values about the image center are preferentially used to generate an image.

If the number of feature points that are successfully tracked falls below a certain number, all feature points are re-extracted and are then tracked for motion estimation and image integration. In this update process, the displacement between the previously extracted corners and current ones is added to the displacement after this frame. Thus, images can be integrated on the \( xy \) plane whose origin is at the center of the first image.
The subject is far from the camera. The algorithm is described using rotations only. This technique works with a rotational approximation, which approximates the optical axis. To solve these problems, we propose image mosaicing when the camera rotates by a large amount about the origin of both the space. The coordinates of a feature point are obtained from the camera in the \( x\bar{y} \) coordinate space, which shows the direction of the feature point from the center of the camera. When the origins of both the \( x\bar{y} \) coordinate space and the \( \theta\phi \) coordinate space are placed at the center of the image, \( \theta \) is expressed as a function of \( x \), and \( \phi \) is expressed as a function of \( y \):

\[
\left( \frac{\theta(x)}{\phi(y)} \right) = \left( \tan^{-1} \left( \frac{x}{y} \right), \tan^{-1} \left( \frac{x}{y} \right) \right)
\]

(4)

\( \Delta \theta \) and \( \Delta \phi \) are given by:

\[
\left( \frac{\Delta \theta_i(\Delta \psi)}{\Delta \phi_i(\Delta \psi)} \right) = \left( \theta(x_i''(\Delta \psi)) - \theta(x_i), \phi(y_i''(\Delta \psi)) - \phi(y_i) \right)
\]

(5)

\((x_i, y_i)\) is the \( xy \) coordinate of a feature point of the previous image and \((x_i'', y_i'')\) is the \( xy \) coordinate, given by:

\[
\left( \begin{array}{c}
\theta(x_i''(\Delta \psi)) \\
\phi(y_i''(\Delta \psi)) \\
\end{array} \right) = \left( \begin{array}{cc}
\cos(\Delta \psi) & -\sin(\Delta \psi) \\
\sin(\Delta \psi) & \cos(\Delta \psi) \\
\end{array} \right) \left( \begin{array}{c}
x_i' \\
y_i' \\
\end{array} \right)
\]

\[(x_i', y_i')\) is the \( xy \) coordinate of a feature point of the current image, and \((x_i'', y_i'')\) is the \( xy \) coordinate of the point after rotation about the light axis \( \Delta \psi \). The horizontal rotation angle \( \Delta \theta \) and the vertical rotation angle \( \Delta \phi \) are given by:

\[
\left( \begin{array}{c}
\Delta \theta \\
\Delta \phi \\
\end{array} \right) = \frac{1}{N} \sum_{j=1}^{N} \left( \begin{array}{c}
\Delta \theta_j(\Delta \psi) \\
\Delta \phi_j(\Delta \psi) \\
\end{array} \right)
\]

(7)

Images are combined by using the rotation angles of the camera, \( \Delta \theta \), \( \Delta \phi \), and \( \Delta \psi \), on the \( \theta \phi \) plane whose origin is at the center of the first image. The current image is pasted onto the mosaiced image after it is rotated by \( \Delta \psi \) and displaced by \((\Delta \theta, \Delta \phi)\) on the \( \theta \phi \) plane.

In the same way as in image mosaicing with the translational approximation, the feature points are updated when the number of feature points that are successfully tracked falls below a certain number.

4. EXPERIMENT

We threw a digital camera (CASIO EX-FC100) up into the air to capture video images for image mosaicing. This camera can capture \( 224 \times 64 \) pixel images at 1000 fps. Figure 4 shows the images captured by the camera. We used 450 video images (0.45 sec) for the experiment. The captured images are so small that we cannot get an impression of the overall scene. We integrated these images using the algorithm described above.

We used OpenCV to implement the algorithm. We used the function \texttt{cvGoodFeaturesToTrack} for corner extraction, and the function \texttt{cvCalcOpticalFlowPyrLK} for feature point tracking. The maximum number of extracted corners was set to 100. The threshold number of feature points for re-extraction was set to 40.
First, the result of image mosaicing with the translational approximation is shown in Fig. 5. Although we could not get an impression of the whole scene from the individual video images, now we can see the whole scene from the combined image. The thrown-up camera successfully generated an image taken from overhead.

Next, the result of image mosaicing with the rotational approximation is shown in Fig. 6. The combined image generated in the $\theta\phi$ plane was projected onto a virtual sphere. The figure shows views of the sphere seen in four directions from the sphere center, at a viewing angle of 90 degree.

We explained that the reason for using an HFR camera is to make feature point tracking stable, as well as to obtain dense information. We conducted an experiment to examine whether a high camera frame rate is necessary.

We compared the combined image generated from 1000 fps video images with that generated from lower frame rate video images. Lower frame rate video images were generated by sampling images at a certain interval from the 1000 fps video images and equalizing the conditions, such as the exposure time. Feature point tracking failed using the images sampled at frame rates of 250 fps or lower. Next, we compared the image generated from the 1000 fps video images with that generated from 333 fps video images. The images generated from both video images are shown in Fig. 7.

As a result, we confirmed that the image generated from high-frame-rate video images had smoother edges. The reason for this result may be that feature points were tracked with high accuracy because of the small image changes between frames when using the HFR camera.

5. CONCLUSION

In this paper, we proposed a method of image integration for wide-range image sensing by using a free-moving camera. Video images captured by an HFR camera thrown up into the air were integrated by image mosaicing with a translational approximation and image mosaicing with a rotational approximation. We also confirmed the necessity for high frame rates by experiment.

In future work, we plan to introduce a motion model of a free-moving camera to enable stable tracking even with images having little texture.

6. REFERENCES

Fig. 6. The sphere on which the mosaiced images were projected.


Fig. 7. A comparison of (a) the image generated from 1000 fps video images and (b) that generated from 333 fps video images. The former has smoother edges than the latter.