3 DoF Image Stitching Using Inertia Sensor Data

Minwoo Kim and Sang-Kyun Kim*
Computer Engineering Department
Myongji University
Yongin, Korea
{minwoo1989 & goldmunt}@gmail.com

Abstract—In this paper, a method to generate panoramic images by combining conventional feature extraction algorithms (e.g., SIFT, SURF, MPEG-7 CDVS) with sensed data from an inertia sensor is proposed to enhance the image stitching results. The challenge of image stitching increases when the images are taken from two different mobile phones with no posture calibration. Using inertia sensor data obtained by the mobile phone, images with different yaw angles, pitch angles, roll angles are preprocessed and adjusted before performing stitching process. Performance of stitching (e.g., stitching accuracy) between conventional feature extraction algorithms is reported along with the stitching performance with/without using the inertia sensor data. Finally, the method of speeding up the stitching of videos (i.e., a sequence of images) is proposed using the same inertia sensor data.

Keywords—image stitching; video stitching; panoramic video; ultra wide viewing video; sensor-based image stitching

I. INTRODUCTION

This paper proposes a three DOF image stitching method using inertia sensor data obtained from image capturing devices. The stitching results between SIFT [1], SURF [2], and MPEG-7 CDVS (Compact Descriptor for Visual Search) [3][4] are compared by stitching time and accuracy.

The three DOF denotes degrees of panning, tilting, and rolling, which are the posture of mobile devices when capturing images or video clips. Fig. 1 shows postures of a capturing device with respect to the three DOF.

This paper is organized as follows. The proposed method is described in Section II. Section III explains experimental results with experimental settings and procedures. Finally, a conclusion is presented in Section IV.

Fig. 1. Capturing device postures in three DOF.

II. IMAGE STITCHING

A. Image stitching process

Fig. 2 shows the ordinary image stitching process. We used inertia sensor data to preprocess the input images before the feature point extraction algorithms are applied.

B. Preprocessing with inertia sensor data

Before applying the sensed data, the x, y, z-axis of inertia sensors are aligned with the x, y, z-axis of the image captured.

The discrepancies of posture angles between two corresponding images are sought as \( \Delta \text{pitch} \), \( \Delta \text{roll} \), and \( \Delta \text{azimuth} \).

The 2D image is transformed to 3D homogeneous coordinates to be combined with 3D rotational matrices. Average angles for each direction are calculated with the angle discrepancies (\( \Delta \text{pitch} \), \( \Delta \text{roll} \), and \( \Delta \text{azimuth} \)). Each image is rotated with the average angles for each direction. Then 3D homogeneous coordinates are projected back to the 2D plane. The composite matrix is shown in equation (1).

\[
M_{\text{composite}} = P_{1 \rightarrow 2} \cdot \alpha(\vec{r}_{uv} \cdot P_{1 \rightarrow 2}).
\] (1)

Two corresponding images are warped with the composite matrix sought in equation (1). Fig. 3 shows the two corresponding images to stitch before warping. The inertia sensor data are shown below. Fig. 4 shows the warped images after applying the composite matrix calculated by the inertia sensor data.
sensor data. The two images become aligned after applying the sensed data.

Fig. 3. Two images before warping.

Fig. 4. Two images after rotating with inertia sensor data.

C. Video stitching using sensed data

A video is a sequence of image frames. In order to stitch two corresponding videos, one should follow the process as shown in Fig. 5.

The stitching of every single corresponding image frames between two videos, however, consumes much time mainly because of the feature extraction time of the matching algorithms.

In order to save the video stitching time, the feature point extraction and the homography matrix calculation are only performed on every K frames. The homography matrices in-between K frames are calculated using linear interpolation of two successive homography matrices obtained from the feature point extraction (e.g., homography matrices from \( I^0 \) image frame and \( (I+K)^t \) image frame) as shown in Fig. 6. Also, the images are rotated with the inertia sensor data before stitching to minimize incorrect stitching results.

Fig. 5. Video stitching process.

Fig. 6. Speed up video stitching with linear homography matrix generation.

III. EXPERIMENTAL RESULTS

A. Experimental conditions

We performed the experiment on Intel Core i7-4770 3.4GHz CPU with 8GB RAM, and used libraries from OpenCV 2.4.11 and reference software from MPEG-7 CDVS evaluation framework.

B. Data acquisition for image stitching

The distance between two mobile phones is approximately 20 cm. Sensed data from the inertia sensor captured for every image frame by Android Apps programmed for the experiment.

The left mobile phone is aligned with 0° in yaw, pitch, and roll direction while the right mobile phone is rotated 0~12° for yaw, pitch, and roll direction, respectively, and captures the corresponding images.

C. Image stitching result comparison

Because there is no clear metric to evaluate the accuracy of stitching, we evaluate the stitching results with the naked eye. If the two corresponding images are stitched perfectly, give 3 points. If they are not aligned perfectly and create a little afterimage, give 2 points. If the images are misaligned noticeably, give 1 point. If the homography matrix is abnormal so that stitching is perfectly failed, give 0 point. SIFT produces the biggest stitching failure rates (7.94%), and its accuracy rate is 2.33. SURF generates no stitching failures and the stitching accuracy is 2.65. CDVS produces 1 point stitching results most, and its accuracy rate is 1.98, which is the lowest among three feature point extraction algorithms.

The failure stitching (e.g., 0 or 1 point) can be enhanced by using the proposed method with the sensed data preprocessing. The stitching accuracies are increased to 2.57, 2.83, and 2.20 for SHIF, SURF, and CDVS, respectively. Fig. 7 shows how the stitching is enhanced when the preprocessing with inertia sensor data was used. The upper image is the result of stitching using SIFT only, where the lower image is the result of stitching using the sensed data preprocessing with SIFT. Stitching result is clearly enhanced in the lower image.
D. Video stitching results

Video stitching was performed using the SURF algorithm. Table 1 is the stitching time of 10,000 pair of image frames. The whole frame stitching using SURF takes 2.6 seconds per an image pair. When the $K$ is set to 9, and generate homography matrices linearly in between, the stitching time becomes reduced to 1.05 seconds per an image pair. When the sensed data adjustment was added, the stitching time becomes 1.1 seconds per an image pair. The stitching accuracy becomes reduced from 2.36 to 2.03 and 2.16, respectively.

<table>
<thead>
<tr>
<th>Stitching time</th>
<th>Whole frame stitching</th>
<th>Linear homography matrix</th>
<th>Linear homography matrix along with sensed data adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000 pairs of image frames (sec)</td>
<td>25951.83</td>
<td>10463.64</td>
<td>11043.28</td>
</tr>
<tr>
<td>Stitching time per an image pair (sec)</td>
<td>2.60</td>
<td>1.05</td>
<td>1.10</td>
</tr>
<tr>
<td>Stitching accuracy</td>
<td>2.36</td>
<td>2.03</td>
<td>2.16</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

A method using inertia sensor data to enhance the image stitching accuracy was proposed. The image stitching results are enhanced with the proposed method for all of the feature point extraction methods (i.e., SIFT, SURF, CDVS). In order to speed up the stitching time, we propose the linear homography matrix method. When this method is combined with the image posture adjustment using inertia sensor data, the stitching time gets reduced by 57.7% while the stitching accuracy gets reduced only by 8.5% in comparison to the whole image frame stitching using SURF.

REFERENCES