A NOVEL ALGORITHM TO STITCH MULTIPLE VIEWS IN IMAGE MOSAICS

Xiaoyong Sun, Eric Dubois

School of Information Technology and Engineering,
University of Ottawa,
Ottawa, Ontario, K1N 6N5,
xsun@site.uottawa.ca , edubois@site.uottawa.ca

ABSTRACT

This paper presents a new method for registration and stitching of adjacent views in an image mosaic, specifically for cylindrical panoramas. The possible registration errors for the overlap area of two pictures captured by a camera mounted on a tripod at different rotation angles have been analyzed based on a general camera rotation model. Then a novel algorithm is proposed based on affine and refocusing adjustments on the optimal stripe block where the stitching will be implemented. The registration errors can be largely reduced because a narrow stripe block is selected instead of the whole overlap area. The simulations on the overlap area registration algorithm show that our method can significantly reduce the stitching errors.

1. INTRODUCTION

A panorama can be generated by stitching together images taken from different directions with some overlap area [1]. In order to obtain a cylindrical panoramic view, images at different rotation angles have to be pre-captured by a camera mounted on a tripod and rotated around its optical center. The adjacent pre-captured images, that have some overlap area, are then stitched together. In this paper, we focus on the registration of two adjacent images to improve this stitching process.

Various methods to register the overlap area of two adjacent images have been proposed based on an 8-parameter perspective transformation [2], a polynomial transformation with more freedom [3], and other geometric corrections [4], etc. However, the registration is usually implemented on the whole overlap area, which is usually too large for one global transformation to get a good result. Thus local corrections have to be introduced which bring both discontinuities and distortions. The problem of the overlap registration is still open even for most of the commercial software to generate panoramic views.

A novel algorithm to register the overlap area of two adjacent images is proposed here for cylindrical panorama generation but the principle can also be applied to spherical panoramic view generation. Only global transformations are used in the proposed algorithm to avoid discontinuities and the registration is implemented only on a narrow stripe area instead of the whole overlap area. In addition, the camera focal lengths associated with different images are adjusted separately in the proposed algorithm to cope with the possible change of focal length.

In the following sections, we will see that registration errors are usually unavoidable based on a general camera rotation model. In section 3, a novel algorithm for the overlap area registration is proposed based on two basic techniques: affine adjustment and focal-length adjustment. The two techniques involve imaging-condition adjustment, both external and internal. An objective function is applied to find the minimal registration difference at the optimal stripe block within the overlap area. The stitching is then applied on this optimal stripe block. The simulation results on the proposed registration algorithm are given in section 4. Section 5 summarizes the conclusions of our work and the future work.

2. A GENERAL CAMERA ROTATION MODEL FOR ANALYSIS OF REGISTRATION ERRORS

In order to stitch two adjacent images, we must determine their spatial relationship, or the position of the overlap area in each image. Moreover, the difference between the overlap area must be minimized to obtain stitched images with good quality. This is the general idea on overlap area registration for a pair of adjacent images.

When a camera is mounted on a tripod and it captures images at different rotation angles, its motion usually deviates from a pure rotation. A general camera rotation model [4] is illustrated in Fig.1. It is not easy to determine the imaging center of a real camera. Thus the camera center O' usually moves along a circle, as shown in Fig.1, instead of the camera rotating around the camera center, or point O in the Fig.1.

Assume that a 3D point $\boldsymbol{X} = (x, y, z)^T$ in the camera coordinate system is projected to the position $\boldsymbol{u}_{[h]} = (u, v, 1)^T$ in image 1. Subscript h stands for homogeneous coordinates. It moves to $\boldsymbol{X}' = (x', y', z')^T$ due to the movement of camera, with image position $\boldsymbol{u}'_{[h]} = (u', v', 1)^T$ in image 2. Image 1 and Image 2 are two adjacent images with some overlap area. The relationship of imaging positions between adjacent images taken by the camera can be described as,

$$w' \begin{bmatrix} u' \\ v' \\ f' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} + T_x/z \\ r_{21} & r_{22} & r_{23} + T_y/z \\ r_{31} & r_{32} & r_{33} + T_z/z \end{bmatrix} \begin{bmatrix} u \\ v \\ f \end{bmatrix} \tag{1}$$

based on a pinhole camera model. Here, r_{ij} $(i = 1, 2, 3; j = 1, 2, 3$ are the components of the rotation matrix R and T_x, T_y, T_z are the translations along the x, y, z axis, respectively. Parameters f' and f are focal lengths associated with different camera positions. Parameter w' is a scale factor. Thus we see that the transformations to describe the

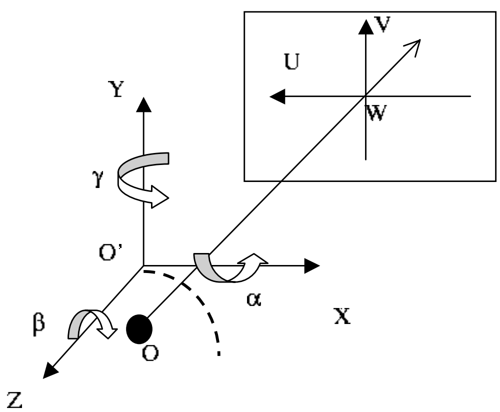

Figure 1 The general camera rotation model

correspondent points' positions in the overlap area of the two adjacent images is depth-dependent. Perfect registration can only be achieved after perfect 3D reconstruction of the depth distribution of the environment, which is too difficult to implement in practice.

3. THE OVERLAP AREA REGISTRATION ALGORITHMS

In the general camera rotation model, we find that the camera movement differing from a pure rotation is one reason that causes the registration error. The other source of registration error comes from camera calibration, which gives an inaccurate focal length for image warping, which is a necessary procedure to mapping two adjacent images onto a common cylindrical or spherical surface. We choose a cylindrical surface in this paper. The resulted mapped images are referred to here as *warped images*. The warping is implemented based on focal length; more detailed information can be found in [2]. In this section, we introduce our methods to reduce the registration errors. One technique that we use is the affine adjustment to reduce the registration errors caused by camera movement. The other one is a focal-length adjustment technique to reduce the estimation error of the focal length. Finally our objective function for optimization is defined on a stripe block within the overlap area whose position is also located with minimal registration error among all candidate stripe blocks, and the final stitching is carried out on this stripe block.

3.1. Camera pose adjustment through affine transformation

Affine transformation is defined on the continuous images. The affine transformation from a discrete image I to the discrete image I' is shown in Fig.2. Images I' and I have the same size, and they can be regarded as being sampled from I_c and I_c' respectively with the same sampling structure. Inversely, I_c can be interpolated from I.

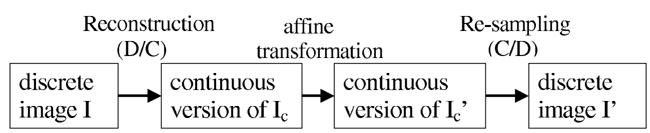

Figure 2 Affine transformation of a sampled image

The image I' is obtained by,

$$I'(\boldsymbol{u}) = I_c(\boldsymbol{u}') \tag{2}$$

The mapping relationship between $\boldsymbol{u'}$ and $\boldsymbol{u}$ is defined as

$$\boldsymbol{u'} = \boldsymbol{T} \cdot (\boldsymbol{u} - \boldsymbol{u_0}) + \boldsymbol{d} + \boldsymbol{u_0}, \tag{3}$$

where,

$$\begin{aligned} \boldsymbol{T} &= \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix}, \\ \boldsymbol{d} &= \begin{bmatrix} t_{31} & t_{32} \end{bmatrix}^T \end{aligned} \tag{4}$$

and $\boldsymbol{u}_0$ is the center of the image I'. Thus, there are six parameters to define an affine transformation; we use a vector to represent them, $\boldsymbol{t} = \begin{bmatrix} t_{11} & t_{12} & t_{21} & t_{22} & t_{31} & t_{32} \end{bmatrix}^T$.

The affine adjustment is applied on the warped images. Assume two images $\widetilde{I}_{10}$ and $\widetilde{I}_{20}$ are warped, using the focal length f_0, from I_1 and I_2, respectively. Here, f_0 is the estimated focal length of the camera. The initial discrete overlap support area S_0 can be determined by the displacement $\Delta\hat{\boldsymbol{u}}$ of the support area $\widetilde{S_1}$ and $\widetilde{S_2}$ of image I_1 and I_2, which is defined as,

$$\Delta\hat{\boldsymbol{u}} = \arg\min_{\Delta\boldsymbol{u}} \sum_{\boldsymbol{u} \in \widetilde{S_2}} |\widetilde{I}_{10}(\boldsymbol{u} + \Delta\boldsymbol{u}) - \widetilde{I}_{20}(\boldsymbol{u})| \tag{5}$$

and

$$\widetilde{S_0} = (\widetilde{S_1} + \Delta\hat{\boldsymbol{u}}) \cap \widetilde{S_2} \tag{6}$$

Here, the plus sign indicates the shift of the support area. Assume $\widetilde{I}_1$ and $\widetilde{I}_2$ are two adjacent images after warping and their initial overlap area is $\widetilde{S_0}$. The affine adjustment is then defined by searching for parameter $\hat{\boldsymbol{t}}$,

$$\hat{\boldsymbol{t}} = \arg\min_{\boldsymbol{t}} \sum_{\boldsymbol{u}\in\widetilde{S}_0} |\widetilde{I}_1(\boldsymbol{u}) - \widetilde{I}_{\text{affine}}(\widetilde{I}_2, \boldsymbol{t})(\boldsymbol{u})| \tag{7}$$

where, $\widetilde{I}_{\text{affine}}(\widetilde{I}_2, \boldsymbol{t})$ denotes a discrete image which is transformed from $\widetilde{I}_2$ through the affine transformation with parameter vector $\boldsymbol{t}$.

In the proposed algorithm, an optimal parameter vector $\boldsymbol{t}$ is selected to update current $\widetilde{I}_2$. Similarly, we can apply affine adjustment on image $\widetilde{I}_1$ to match $\widetilde{I}_{\text{affine}}(\widetilde{I}_2, \boldsymbol{t})$.

3.2. Focal-length adjustment method

The focal-length adjustment is used to compensate the overlap registration error caused by the internal camera parameter, or focal length. The optimal focal length associated with each image for warping processing is found to minimize the registration error. In the proposed algorithm, the focal-length adjustment is applied on unwarped images. Assume that we want to apply focal-length adjustment on image I_2, or finding the optimal focal length associated with I_2 that can minimize the overlap registration error.

First, we warp I_1 with some fixed focal length f_1. It can be the focal length initially estimated from camera calibration, or it can be the focal length obtained from a previous optimization of focal-length adjustment applied on image I_1. Then, the search for an optimal focal length $\hat{f}_2$ for I_2 yield

$$\hat{f}_2 = \arg\min_{f_2} \sum_{\boldsymbol{u}\in\widetilde{S_0}} |\widetilde{I}_1(\boldsymbol{u}) - \widetilde{I}_{\text{warp}}(I_2, f_2)(\boldsymbol{u})|. \tag{8}$$

The initial guess for f_2 is f_{20}, which also can be the focal length initially estimated from camera calibration, or can be the focal length from a previous optimization of focal-length adjustment applied on image I_2 in an iterative fashion. $\widetilde{I}_{\text{warp}}(I_2, f_2)$ denotes the image warped from I_2 under the focal length f_2. Similarly, we can represent $\widetilde{I}_1 = \widetilde{I}_{\text{warp}}(I_1, f_1)$. $\widetilde{S}_0$ is the overlapped support area of image $\widetilde{I}_{10}$ and $\widetilde{I}_{20}$ as above.

After the optimal focal length $\hat{f}_2$ is found, warping with this focal length is applied on I_2 to update $\widetilde{I}_2$ for further processing. As we mentioned, focal-length adjustment is only applied on un-warped images in the proposed algorithm. If $\widetilde{I}_2$ has been updated, we apply un-warped (inverse transformation of warping) transformation on $\widetilde{I}_2$ to update I_2. Also, we can apply a similar focal-length adjustment procedure on I_1.

3.3. Choice of a narrow stripe block for optimization

The fact that the final stitching of the two images can be implemented just on a narrow stripe area is important to the proposed algorithm. Thus we perform the optimization on all possible stripe blocks with fixed size in the overlap area, instead of the whole overlap area, and thus a better registration result can be expected. The stripe block $\widetilde{S}_i$ is defined as a sub-area of the overlap area, or $\widetilde{S}_i \subset \widetilde{S_0}$. The subscript i denotes the location where the stripe block starts. Then our optimization models have been changed to

$$[\hat{f}_2, i] = \arg\min_{f_2, i} \sum_{\boldsymbol{u}\in\widetilde{S}_i, \widetilde{S}_i\subset\widetilde{S}_0} |\widetilde{I}_1(\boldsymbol{u}) - \widetilde{I}_{\text{warp}}(I_2, f_2)(\boldsymbol{u})| \tag{9}$$

for focal-length adjustment applied on image I_2, and

$$[\hat{\boldsymbol{t}}, i] = \arg\min_{\boldsymbol{t}, i} \sum_{\boldsymbol{u}\in\widetilde{S}_i, \widetilde{S}_i\subset\widetilde{S_0}} |\widetilde{I}_1(\boldsymbol{u}) - \widetilde{I}_{\text{affine}}(\widetilde{I}_2, \boldsymbol{t})(\boldsymbol{u})| \tag{10}$$

for affine adjustment of image $\widetilde{I}_2$.

3.4. The algorithm for registration based on focal-length adjustment and affine adjustment

The overall proposed algorithm for overlap area registration can be described as the following procedure for the adjacent images I_1 and I_2 captured by a rotating camera.

- The images I_1 and I_2 are first warped using an estimated focal length, approximated from camera calibration. The estimated focal length does not need to be very accurate, which makes the calibration inexpensive. The warped versions of images I_1 and I_2 are $\widetilde{I}_1$ and $\widetilde{I}_2$ respectively.

- A coarse registration is first applied based on minimal absolute difference error searching on the overlap area to find the horizontal and vertical displacement of image $\widetilde{I}_1$ and $\widetilde{I}_2$. This also defines the overlapped support area $\widetilde{S}_0$.

- Apply affine adjustment on image $\widetilde{I}_1$ and $\widetilde{I}_2$ separately, and update $\widetilde{I}_1$ and $\widetilde{I}_2$ based on the determined optimal affine transformation parameters.

- Un-warp $\widetilde{I}_1$ to obtain I_1 and implement focal-length adjustment on I_1; similarly, apply focal-length adjustment on I_2.

- apply affine adjustment again to the updated $\widetilde{I}_1$ and $\widetilde{I}_2$ which are warped with the determined optimal focal lengths respectively; update $\widetilde{I}_1$ and $\widetilde{I}_2$ based on the determined optimal affine transformation parameters.

- record the stripe block position i where the final minimal registration error has been obtained.

As mentioned before, luminance difference at the overlap area is unavoidable, and it will affect the registration accuracy. A luminance compensation factor is introduced to roughly compensate the luminance difference in the overlap area. The luminance compensation factor is defined as

$$\eta = \frac{\sum_{\boldsymbol{u}\in\widetilde{S}_0} \widetilde{I}_1(\boldsymbol{u})}{\sum_{\boldsymbol{u}\in\widetilde{S}_0} \widetilde{I}_2(\boldsymbol{u})}. \tag{11}$$

Thus, the modified optimization models are,

$$\begin{aligned} [\hat{f}_2, i] &= \arg\min_{f_2,i} \sum_{\boldsymbol{u}\in\widetilde{S}_i,\widetilde{S}_i\subset\widetilde{S}_0} |\widetilde{I}_1(\boldsymbol{u}) - \eta\cdot\widetilde{I}_{\text{warp}}(I_2, f_2)(\boldsymbol{u})|, \\ [\hat{\boldsymbol{t}}, i] &= \arg\min_{\boldsymbol{t},i} \sum_{\boldsymbol{u}\in\widetilde{S}_i,\widetilde{S}_i\subset\widetilde{S}_0} |\widetilde{I}_1(\boldsymbol{u}) - \eta\cdot\widetilde{I}_{\text{affine}}(\widetilde{I}_2, \boldsymbol{t})(\boldsymbol{u})|. \end{aligned} \tag{12}$$

Through the above registration algorithm, we not only obtain two images $\widetilde{I}_1$ and $\widetilde{I}_2$ which are ready to be stitched together, but also the position of the optimal stripe block where the stitching will be implemented.

4. SIMULATION RESULTS

The simulation was carried out on a pair of images, which were chosen intentionally with large depth variations. The camera pose was just roughly adjusted to be levelled by hand. Due to the large depth variations, these two adjacent images make it difficult to get a good registration result at the overlap area. The registration was implemented with the proposed algorithm and the stripe block that we selected was 100 pixels in width and the same height as the original image. The optimizations were implemented using the MATLAB optimization toolbox. The stitched result after the registration with the proposed method is shown in Fig.3(b). Fig.3(a) shows the stitching result of a commercial software for comparison, and we can see the stitching error (within the circular mark) is very striking.

(a) (b)

Figure 3 the stitched images with a commercial software (a) and the proposed algorithm (b)

5. CONCLUSION AND FUTURE WORK

In this paper, an algorithm has been proposed for the registration of the overlap area of two adjacent images in order to better stitch them together. The images are captured by a camera mounted on a tripod at different rotation angles. The application of the proposed algorithm includes obtaining large filed of view images, generating panoramas, etc.

The proposed algorithm consists of the external adjustment, or camera pose adjustment by affine transformation, and the internal adjustment, or focal-length adjustment. It is obvious that the affine transformation is not the only method for camera pose adjustment but it is a relatively simple one. More sophisticated transformations, such as global perspective transformation and polynomial transformation, can be applied to replace the role of affine adjustment in the framework of the proposed algorithm.

6. ACKNOWLEDGEMENT

This work was supported by a grant from the National Capital Institute of Telecommunications, Canada.

7. REFERENCES

[1] S. E. Chen, "Quicktime VR — an image-based approach to virtual environment navigation," *Computer Graphics (SIGGRAPH'95)*, pp. 29–38, August 1995.

[2] H.-Y. Shum and R. Szeliski, "Construction of panoramic image mosaics with global and local alignment," *International Journal of Computer Vision*, vol. 36, no. 2, pp. 101–130, 2000.

[3] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, "Mosaic representation of video sequences and their applications," *Signal Processing: Image Communication, special issue on Image and Video Semantics: Processing, Analysis, and Application*, vol. 8, no. 4, pp. 327–351, 1996.

[4] Z. Zhu, G. Xu, E. M. Riseman, and R. Hanson, "Fast generation of dynamic and multi-resolution 360-degree panorama from video sequence," *Proceedings of 1999 IEEE International Conference on Multimedia Computing and Systems*, vol. 1, pp. 400–406, 1999.