Institute of Advanced Engineering and Science

International Journal of Informatics and Communication Technology (IJ-ICT)
Vol.1, No.1, July 2012, pp. 20~25
ISSN: 2252-8776

Synthesizing the Whole View of a Large-Scale Object

MyintMyintSein*, Thin Lai LaiThein**

* Research and Development Department, University of computer Studies, Yangon
** Information System Department, University of computer Studies, Yangon

Article Info

Article history:

Received June 20th, 2012
Accepted July 28th, 2012

Keyword:

View synthesizing
Camera Calibration
Linear motion
Large Scale object
Image Registration
Foreground object
Geometric transformation

ABSTRACT

This paper presents the effective method for the whole view synthesizing of the large-scale object. This synthesizing method is developed based on the image registration approach. The foreground object is extracted from each image by color image processing. The information of the feature points of an object is obtained from a foreground image and establishing the correspondence among views. The translation and rotation parameters between two successive images are estimated by the computed camera calibration parameters. The second image is transformed by using these parameters and combine to first image. The integrated image is obtained by combining the overlapping parts of these images. A whole view of the large scale object is reconstructed image by applying the above process successively. Preliminary investigations with sampled images of a real scene have been done to confirm the effectiveness of the proposed method.

Corresponding Author:

First Author,
Research and Development Department,
University of computer Studies, Yangon.
Email: myintucsy@gmail.com

1. INTRODUCTION

The reconstruction of the complete view of a virtual environment has been paid substantial attention by the researchers. This paper presents the recent results of our research work related to construct the whole views of an object from the sampled viewpoint images of a scene. Especially, a full view image of the large-scale object is reconstructed. The whole view of a large-scale object is impossible to take the whole scene at once without using sufficient distance between camera and object. The system of the whole view synthesizing is developed base on the image registration approach. Image registration technique is very useful for computing the transformation and constructing a panoramic image from the sampled images of a scene. The multiple images are grabbed by a camera system from different camera viewpoints. The corresponding pairs among the successive image pairs are determined from the relation of the camera positions. The perspective projected geometry method is applied for detecting the relation between the camera and image coordinate system from the image pairs. In this approach, it is possible to produce a lot of intermediate views by using a few cameras. Then it will be provided to reduce the cost and number of camera in security system by using the synthesizing the arbitrary view of images.

Some reports have already been presented concerning the research work of the view synthesizing of an object and a scene[1-6]. N. Chiba [1] proposed the feature based image mosaicing technique for arbitrary depth scene. Yamamoto et al.[2] proposed the trinocular stereo system for searching the optimal correspondence among three images. S.E Chan and Williams [3] proposed the view interpolation method for image synthesizing by using image morphing technique. M.M.Sein et al., [4] presents an approach for reconstructing the arbitrary view of a large-scale object. This presents a new approach for synthesizing the arbitrary view based on the image morphing technique. T.Takahashi et al. [5] proposed a method for rendering views for large-scale scenes. An omni directional camera is used to capture the panoramic image running along a straight line. J. Mulligan et al [6] presented a technique to create virtual worlds using densely distributed stereo views. In their systems, the location of a third camera is at the center of a pair of stereo camera. The computation of dense trinocular disparity maps has been explored for non-planar camera configurations that arise when cameras are set surrounding the object, which is to be modeled. We developed a robust method for reconstructing the whole views of a large-scale object. Unlike the other virtualized reality system, we don't only synthesize the virtual stereo images but also reconstructed the arbitrary and virtual views of the unspecific configured object. It is also possible to create the new scene by merging the synthesized views of multiple objects.

2. Image projection and Camera Model

2.1 Projected Transformation

The principal of the projective transformation between two different viewpoints $V1$ and $V2$ will be considered. The transformation matrix **H** can be defined by the image coordinates m_1 and m_2 as:

$$k\, m_2 = \mathbf{H}\, m_1 \,, \tag{1}$$

where k is an arbitrary constant and $m_1=(x_1,y_1)$ and $m_2=(x_2,y_2)$ are the projected image points of the point M on the object. Equation (1) can be rewritten as

$$m_2 = k'\,\mathbf{H}\, m_1 \,. \tag{2}$$

where k' is an arbitrary nonzero scale factor. The linear solution of the transformation matrix **H** can be determined by using the four or more corresponding pairs of feature points. The aim of image projection is to find the correspondences between two images. We computed the transformation parameters by using mean square distance between two images. The correspondences are not easy to find due to incorrect feature detection and non-robust feature matching.

2.2 Perspective Camera Model

Nikon COOLPIX camera is used for acquisition of images. Let $P(X, Y, Z)$ denote the Cartesian coordinate of a 3D scene point with respect to the camera, and let $p(x, y)$ denote the corresponding coordinates in the image plane and 'C' is camera position as shown in figure 1. For the camera center coordinate system, the image plane is located at the focal length f and the relation between the 3-D point and 2-D point is expressed by:

$$x = f\,\frac{X}{Z}, \quad y = f\,\frac{Y}{Z}. \tag{3}$$

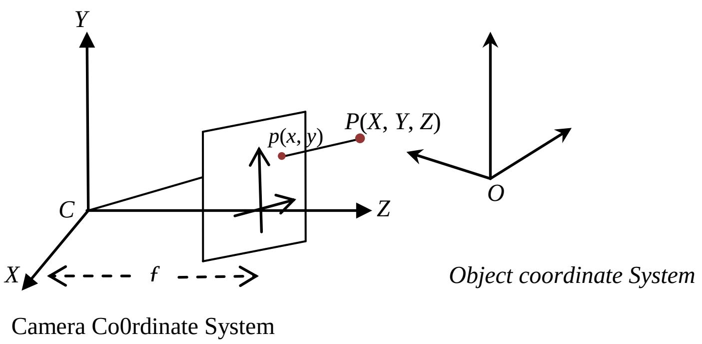

Figure 1. Illustration of the Perspective Geometry.

Equation (3) can be represented in the matrix form as:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{4}$$

The transformation matrix **H** is considered translation T and rotation matrix R, separately.

$$H = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{pmatrix} \quad (R \mid T) \tag{5}$$

The camera calibration can be computed from four planar points from an image by using equation (2) and equation (5).

3. Image Registration

Image registration is one of the fundamental tasks in image matching. The accurate transformation matrix is the key of the registration approach. It is the process of matching two images which are reference image and operated image. At least three matching pairs are needed to guess the initial transformation matrix. Let us consider the registration between two parts of an object. The first part is supposed to be a main part or destination part, and the second part is supposed to be a current part or transformed part. Let ***P'*** and ***P*** denote the points on the main part and current part of an object and their relation can be expressed as:

$$\boldsymbol{P'} = T\boldsymbol{P} + \boldsymbol{D}, \tag{6}$$

where ***D*** is the translation parameter and T contains the scaling and rotation parameters. In 2D Case, T and ***D*** can be defined as:

$$T = a\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and $$\boldsymbol{D} = [\, d_1 \; d_2]^t, \tag{7}$$

where a , θ and d_1 ,d_2 are the scaling, rotation and translation parameters, respectively. The initial transformation matrix can be detected the above relation. The accurate transformation matrix is computed iteratively by minimizing the distances between the control points of the two parts and that can express in the following equation.

$$D_i^{\,2} = \sum distance\,(T_i P_i, P_i') \tag{8}$$

where $T_i = T \circ T_{i-1}$. The difference δ_i from the $(k\text{-}1)^{th}$ to the k^{th} iteration is defined as

$$\delta_i = \frac{1}{N}(D^k - D^{k-1}) \tag{9}$$

where $|\delta_i| \leq \varepsilon_i$ and N is the number of control points on the curve. This process is continued until the difference $|\delta_i|$ become less than a threshold value $e_i (> 0)$.

4. Generating the Whole View of the Large Scale Object

An algorithm is developed for generating the virtual view and arbitrary viewpoint image of a scene from the images grabbed by moving a camera. The general process for generating system is illustrated in figure 2. The basic concept of using the planar projective transform is that it assumes scenes to be planar. By dividing a non-planar scene into multiple triangular patches, the planar projective transform can be applied for non-planar scene. Foreground object extraction is performed for each image based on the color segmentation. The camera calibration is calculated to obtain the precise transformation between two images. It will provide for image registration. The coordinate transformations between all pairs of an image sequences are computed.

The first input image is set to reference image and second input image is defined as operated image. After that we cut off the image which is not overlapping region from operated image and then merged it to the reference image. The feature points in each image can be transformed the coordinate frame. The transformations accompanied with the above correspondences are applied to each view and all the views are registered in the coordinate basis of a reference view. After all the views are registered, correspondences are established between all sets of overlapping views on the basis of nearest neighbor that are within a distance. Views that have correspondences less than a threshold are considered as non overlapping views and their correspondences are rejected.

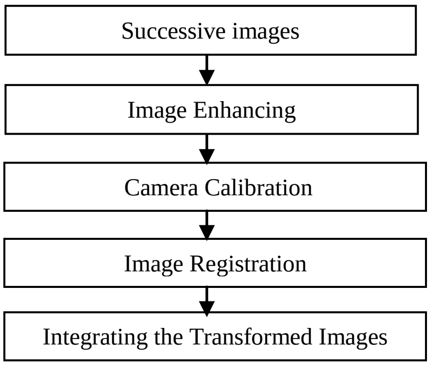

Figure2. Processing steps of the proposed system.

The overlapping region between two images is extracted and computed a maximum displacement of the overlapping region. After transform the operated image, it will be integrated to reference image. Then, this integrated image is defined as a reference image and pick up a new image as operate image. Finally, the whole view of a large scale object will be reconstructed by integrating the transformed object parts.

5 Experiments and Results

In this section, some experimental results and analyzing of proposed technique are presented. FUGIFILM Fine-Pix J12 digital camera is used for taking the images from arbitrary view positions. The first experiment is synthesizing the panorama view of a scene. The successive image sequence is grabbed by a camera with horizontal motion. It will effect only the translation and the rotation parameter is small. There are 20 successive images of a scene contains a building. Figure 3(a) is the some grabbed images. Feature point extraction and edge detection are performed before computing the transformation parameters. Some transformation errors are due to the matching errors. The accuracy rate percentage is the ratio of the number of corresponding pairs and maximum number of feature points in images.

$$\text{Error rate} = 100 - \frac{\text{number of corresponding pairs x } 100\%}{\text{number of all feature points pairs}}$$

The corresponding feature points on the two input images are detected and the distance in the overlapping region is calculated. The non overlapping region of the operated image is cut off and then merged it to the reference image. Figure 3(b) shows the synthesis panorama view of a scene.

(a) Some successive images

(b) Synthesis image

Figure 3. Synthesize panorama view of a scene.

Another experiment is the whole view reconstruction of a large scale object. The successive image sequence is grabbed by a camera moves with horizontal and vertical displacement. Both of the translation and the rotation parameters will effect in this experiment. There are 45 successive images of a large object. Some images are shown in figure 4.

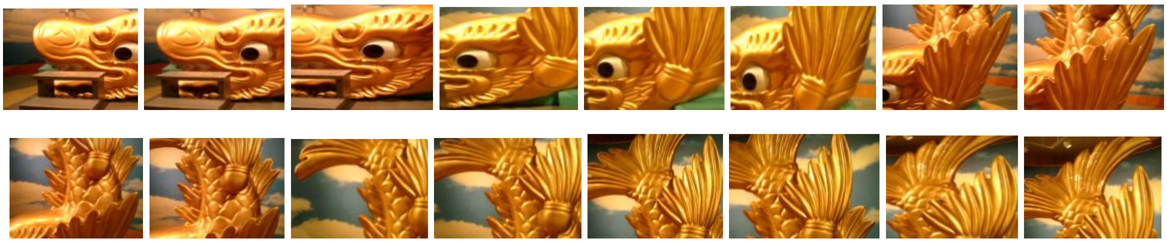

Figure 4. Some images of a large-scale object

Foreground extraction is done by eliminating the background color. Some extracted foreground images are illustrated in figure 5(a). The feature point extraction and computing the camera calibration process are required to perform before image registration process. The integrated part of the first 20 successive images is shown in figure 5(b) and the whole view reconstruction is shown in figure 5(c), respectively. Nearly 5% matching error is occurred due to the error of camera calibration.

Some images

Foreground images

(a) Background elimination

(b) Integrated part

(c) The whole view

Figure 5. The whole view reconstruction of a large scale object

6. CONCLUSION

The proposed system reconstructed the full-view of large-scale object from the successive images. The camera calibration and the absolute coordinate transform are computed for each image. The transformation parameters between each successive image are calculated and then used in blending operation. For each image pair, we could get very good results but when we use these parameter in blending, it could cause some artifacts. The perfect and good image has been received when blending the multiple images into one. The advantage of this process is useful for showing irremovable objects from one place to another at museum.

REFERENCES

[1] N. Chiba,“ Image mosaicing for arbitrary depth”, Image Labo, 11(8)220-230, August 2000.

[2] T. Yamamoto et al., “Correspondence Search of Trinocular Stereo Using Dynamic Programming”, IPSJ SIGNotes Computer Vision, Vol. 046, 1986.

[3] S.E.Chan and Williams, “View Interpolation for Image Synthesis”, Proc.SIGGARPH 93, pp. 279-299, 1993.

[4] M.M.Sein et.al, “ Reconstruction the Arbitrary View of an Object Using the Multiple Camera System”, IEEE International Symposiun on Micromechatronic and Human Science(MHS 2003), Nagoya, Japan,pp.83-88, Oct. 19-22,2003.

[5] T.Takahashi et al., “Arbitrary View Position and Direction Rendering for Large-Scale Scenes”, IEEE International Conference on Computer Vision and Pattern Recognition, pp.1063-6919, vol2.2000.

[6] Y.Chen and G.Medioni, “Object modeling by registration of multiple range images”, Image and Vision Computing, 10(3):145-155, Apr 1992.

BIOGRAPHY OF AUTHORS

Myint Myint Sein received the Ph.D in Electrical Engineering from the Graduate School of Engineering, Osaka City University, Osaka, Japan in 2001. She is presently serving as a professor in the Research and Development Department, University of Computer Studies, Yangon, Myanmar from 2005. She is a member of Program Committees and Organizing Committees of many National/ International Conferences. She has worked/ presently works as referees for many international journals and is a member of the technical committee for a couple of international conferences. She has published more than 70 research articles in peer reviews journals, book chapters and international conferences. Her research interests are Pattern Recognition, Image Processing, Soft computing, 3D reconstruction and 3D Image Retrieval.

Thin Lai Lai Thein received the Ph.D (IT) in 2007 from University of Computer Studies, Yangon, Myanmar. She is currently works at University of Computer Studies, Yangon, Myanmar as Associate Professor. She is interesting in pattern Recognition, Image Processing and 3D virtual view reconstruction, etc.