Producing Realistic Building Texture Using Video Mosaic

Fuan Tsai, Cheng-Hsuan Chen
Center for Space and Remote Sensing Research, National Central University
Chung-Li, Taiwan.
E-Mail: ftsai@csrsr.ncu.edu.tw, 93322086@cc.ncu.edu.tw

Jin-Kim Liu
Energy and Resources Laboratory
Industrial and Technology Research Institute
Hsin-Chu, Taiwan

Abstract:
For realistic building modeling, texture mapping via video sequences is an effective approach. The objective of this study is to semi-automatically produce photo-realistic texture images of building models using video sequences. The textures are generated using mosaic of frames extracted from video sequences and with 40% to 60% of overlaps. The overlapped video frames are merged together based on interest points identified from intersected line features of video images. These interest points are then matched to each other and used to register the images for mosaicing. The generated texture mosaic can then be mapped onto building models.
Keywords: video sequences, automatic mosaic, line-feature.

I. Introduction

Three-dimensional building model is one of the most important elements in a Cyber City implementation. Currently, 3D building models can be constructed effectively from high resolution aerial photographs, LIDAR images and other spatial data. However, most of building models lack adequate texture information. This may seriously limited the reality of building models and their applications. Although previous studies have demonstrated the viability of produce building texture from images acquired with digital camera (Tsai & Lin, 2004), it will need an abundance of digital images in order to generate realistic and complete texture for building models. On the other hand, digital video recorders can supply a large number of images efficiently. Consequently, it is valuable to investigate how to generate building texture images from video sequences.

One of the most challenging issues in video mosaic is to register individual video frames correctly. In other words, it is about how to match the corresponding features in video sequences and use them to merge video frames together. In addition to the geometric registration, it is also important to balance color and shading intensities among stitched video frames to create a seamless texture mosaic of a building facade.

This study developed a procedure to create texture mosaic semi-automatically from close-ranged video sequences for photo-realistic texture mapping of building models in Cyber City applications.

II. Video Mosaic

To merge or stitch video frames together into a texture mosaic, the frames need to be transformed to a common coordinate system. One way to do this is using interest points as the tie points to register video frames. A few algorithms have been developed for this purpose [1],[2]. However, if directly applied to building images, these algorithms will identify too many interest points on each frame of the video sequence. Using features instead is a better approach to avoid this drawback [3]. There are high-order characteristic features, like symbolic feature, block feature etc. However, because video sequences for building texture mapping are usually close-ranged images, the distortions are more severe. These high-order features in overlapped video frames may become too difficult to match. Therefore, using line-feature to construct building corners and other intersected points as interesting points seems to be a better approach.

When the interest features are extracted and interest points are identified, the next step is to stitch them in paired images. This is probably the most difficult and important step of the study. There are several automatic methods for this process. For example, LSM (least squares matching), which minimize the difference of search window [4] and NCC (normalized cross correlation) that finds the highest correlation coefficient within a searching window.

After interest points are matched correctly, they can be used as tie points to register the paired video frames. Afterward, texture mosaic can be generated by merging registered video frames. Besides geometry, the colors and shadings will also need to be adjusted in the registration in order to generate seamless texture mosaic.

III. Method and Procedure

The general procedure of creating texture mosaic of building models from video sequence can be separated into several steps as described below:

3.1. Finding interest points

After extracting frames from the video sequence, the first step is to identify interest points on individual image. Most regular building facades are retangular-shaped, so the corners will usually locate at the intersection of two straight lines. In this step, Hough Transform is applied on overlaped images to extract lines. Hough Transform is a method that transforms the feature points in image space into a γ-θ-space to extracted straight lines in an image. It can be expressed as a simple equation given by Eq. (1):

$$\gamma = x\cos\theta + y\sin\theta \qquad (1)$$

After the lines are extracted, intersection points of lines will then be calculated.
Let A,B and C,D be the ending nodes of two different lines. Then the directed line segments AB & CD are given by Eq. (2):

$$\begin{Bmatrix} AB = A + r(B - A) & r \in [0,1] \\ CD = C + s(D - C) & s \in [0,1] \end{Bmatrix} \qquad (2)$$

If AB & CD intersect, then (Eq. (3)):

$$A+r(B-A)=C+s(D-C) \;;\; r,s\in[0,1] \qquad (3).$$

The intersection point is give by Eq. (4):

$$P=A+r(B-A) \qquad (4)$$

And the intersection points are used as the interest points.

3.2. Matching interest points

In this step, the main objective is to identify one-to-one relationship among points in two overlapped images. This process is divided into two parts--the distance condition and the correlation coefficient condition. At the distance condition, it considers the translation between overlapped images. In this study, the video was shot horizontally. The threshold of the horizontal distance range is set to 0.4~0.6 of image width (the overlap of two frames), and the vertical threshold is set to -15~15 pixels. If the distance of a point pair is greater than the thresholds, they are excluded. Next, the NCC (normalized cross correlation) is used for precise matching. NCC is a method that searches the highest correlation coefficient between the target and search windows. The concept is illustrated in Fig. 1.

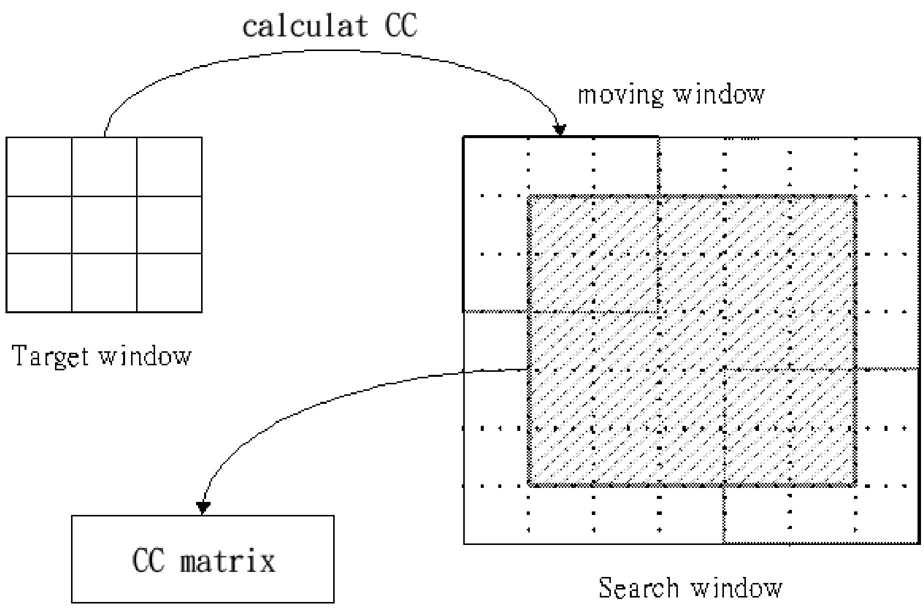

Fig. 1: NCC diagram (CC: correlation coefficient)

3.3. Registration and merging

The matched interest points now can be used as tie points to register the overlapped images. The registration can be done using affine (linear) transformation. To complete the registration, least squares fitting is used to solve eight-parameter transformation. After registration, the overlapped frames can then be merged together correctly using the polygon-based approach described in [5]. Applying this procedure iteratively for each extracted video frame, a complete texture mosaic is produced accordingly and ready to map onto corresponding building model facets.

IV. Result and Discussion

Fig.2. shows two overlapped frames extracted from a video sequence targeting at a facade of a long building. The overlap region is about 50% between the two extracted images. Because there are random noise in the images, they were teated with a mean filter processing.

Fig. 2: Two extracted overlapped video frames.

As described in the previous section, Hough Transform are applied to the test images and the intersection points are calculated as candidates of interest points. The result is displayed in Fig.3

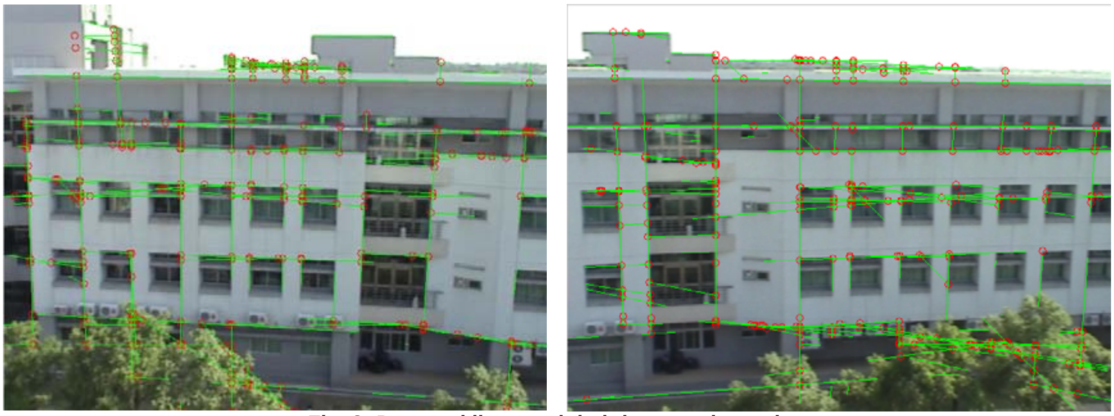

Fig. 3: Detected lines and their intersection points.

In Fig. 3, the red cycles are the intersection points. As can be seen in the figure, most of the corners are correctly detected.

The next step is to match these points in the two images and the matching result is displayed in Fig.4. These points are the interest points of the images and will be used as tie points to register the frames.

Fig. 4: Matching result

Fig. 5 shows the result of registration and mosaicing, which uses the left frame as the base, registers and appends the right frame onto it. There seems to be a systematic error in the affine transformation that causes the small dot lines on the right-hand part of the mosaic image. The cause of this artifact is still under investigation when this paper is being written up. For now, a median filter is used to clean up the mosaic image. The final result is displayed in Fig.6. From Fig. 5 and Fig. 6, it can be seen that the developed procedure has produced a texture mosaic of the building. However, there seems to be a error propagation that causes the discontinuity of the building outline at the top and bottom of the building along the stitching line. To reduce this imperfection will require a revision of the registration and merging algorithms. This issue is now under evaluation and will be addressed in the future.

Fig. 5: Registration and preliminary mosaic result.

Fig. 6: Median filter corrected result

V. Conclusion and Future Work

This paper presents a procedure for semi-automatically producing texture mosaic from video sequence. The preliminary result presented here shows that the developed method is a valid approach for dealing with photo-realistic texture mapping of 3D building models. However, some issues still need to be addressed and there definitely are still rooms for improvements. Among them, the error propagation in registration and merging that causes the discontinuity in geometry of the mosaic result will be the top priority of the research team to correct. The next is to implement the developed procedure into a cyber city system. Also, a more complete test in a city model will also be scheduled to conduct in order to understand the performance of the developed system in the real operation.

Acknowledgement

This study was support in part by the NCU-ITRI Joint Research Center (NCU-ITRI-930304) and the National Science Council of Taiwan (NSC-94-2211-E-008-031).

References

[1] Gui, T., D. Gledhill, D. Taylor, 2003, Comprehensive interest points based imaging mosaic. Pattern Recognition Letters, 24(9-10), pp. 1171-1179.

[2] Szeliski, R. 1996, Video mosaics for virtual environments. Proc. IEEE Computer Graphics and Application, pp. 22-30.

[3] Zhang, Z. and Kang, Z, 2004, The rendering of building texture from land-based video. IAPRS, Vol. XXXV, part B3, pp. 732-737.
[4] Ackermann, F., 1984, Digital Image Correlation : Performance and Potential Application in Photogrammetry. Photogrammetric Record, 11(64), pp.429-439.
[5] Tsai, F. and H-C Lin, 2004, Realistic Texture Mapping On 3D Building Models , Proc. 25th Asian Conference on Remote Sensing (ACRS2004), Nov. 2004, Chiang Mai, Thailand, pp. 204-208.