PTZ CAMERA CALIBRATION FOR AUGMENTED VIRTUAL ENVIRONMENTS

Lu Wang, Suya You, Ulrich Neumann

Computer Science Department, University of Southern California

ABSTRACT

Augmented Virtual Environments(AVE) are very effective in the application of surveillance, in which multiple video streams are projected onto a 3D urban model for better visualization and comprehension of the dynamic scenes. One of the key issues in creating such systems is to estimate the parameters of each camera including the intrinsic parameters and its pose relative to the 3D model. Nowadays, PTZ cameras are popular in this kind of applications. How to rapidly calibrate them at an arbitrary PTZ setting is not clear in the literature. We propose an efficient approach with two steps. In the first step, panoramic images are generated at a set of zooms. The images composing these panoramas are calibrated and stored in a database. In the second step, an image is acquired at an arbitrary PTZ setting. Its best matching image in the database is found by using an efficient local feature recognition technique. Based on this image, the camera parameters at the new PTZ setting can be estimated.

***Index Terms*—** Augmented virtual environments, camera calibration, local feature, panorama

1. INTRODUCTION

In an Augmented Virtual Environment (AVE) [1], multiple live videos captured by surveillance cameras mounted on the top of buildings are projected onto a 3D urban model. Instead of watching these videos separately, the users can visualize them in a 3D context simultaneously, which helps the users better comprehend the dynamic scenes. To set up such a system, one of the main problems is to calibrate each camera so that the projection of the video can be accurately aligned with the 3D model.

Nowadays, PTZ cameras are popular in surveillance systems. Their calibration posts two challenges. First, due to limited field of view, the number of 2D to 3D feature correspondences contained in a single image is usually not enough to satisfy the requirement in classical DLT approach [2]. To solve this problem, some researchers used a set of landmarks put on the ground [3] but to measure the location of the landmarks is inconvenient. The second challenging problem is that the user can pan, tilt and zoom a camera freely. The recalibration of the camera at an arbitrary PTZ setting should be fast. The traditional approaches involving manual selection of 2D to 3D feature correspondences are not practical especially for a surveillance system requiring an instant respond.

The approach in [4] provides an effective way to calibrate the intrinsic parameters of a PTZ camera at different zoom settings. It has two phases. First, at the minimum zoom the camera is rotated and images are acquired in a spherical grid with certain discrete pan and tilt steps. They are stitched into a panorama, during which the intrinsic parameters of the camera at the minimum zoom can be estimated. In the second phase, an image sequence is captured in a fixed direction with the camera progressively zooming in. From it, the intrinsic parameters of the camera at each zoom step can be obtained. The advantage of this approach is that it does not need any information from 3D models. However, the extrinsic parameter estimation problem is not addressed.

In our system, the same approach as [4] is used to estimate the intrinsic camera parameters at each zoom step. However, instead of creating a panorama only at the minimum zoom we create panoramas for a set of zooms increased with a certain step. To estimate the extrinsic parameters, the user select the correspondences between the 2D features (point or line) in each panorama and the 3D features on the 3D model. Since a panorama has a much larger field of view, we can usually find enough such correspondences. An approach similar to [5] is used to estimate the camera pose of the images composing the panoramas. These images and their associated camera parameters are stored in a database.

The users can pan, tilt and zoom the camera to an arbitrary PTZ setting at which a new image is taken. Its best matching image in the database is found with an efficient local feature recognition technique called SURF [6]. From the homography between these two images, the intrinsic and extrinsic camera parameters at the new PTZ setting can be estimated.

We need to mention that the extrinsic camera parameters at the new PTZ setting have to be estimated online. Due to mechanical reasons, for not very expensive PTZ cameras the extrinsic parameters may have slight variation every time when the cameras are moved to the same pan and tilt settings. This variation can cause several pixels error of the alignment between the image projection and the 3D model. Therefore, the approach in which the extrinsic parameters at every pan and tilt step are estimated offline and saved in a lookup table cannot satisfy the accuracy requirement in an AVE system.

The paper is organized as follows: In Section 2 camera calibration based on panoramas is briefly introduced. Section 3 describes online re-calibration of the camera at an arbitrary PTZ setting. In section 4, some experimental results are presented. A conclusion is given in section 5.

2. CAMERA CALIBRATION BASED ON PANORAMAS

At the first step of our approach, panoramas are created at a set of zooms. At each zoom, the camera is rotated with certain pan and tilt steps. The captured images are overlapped with each other and form a hemispherical grid. The approach in [4] is used to stitch them together into a panorama and to estimate the intrinsic camera parameters. However, instead of using Harris corners we use SURF features [6] to match neighboring images because they are more robust and can be computed very fast.

To estimate the camera pose for each image in a panorama, the user selects the 2D features (point or line) in the panorama and the corresponding 3D features in the 3D model. In our system, the panorama is displayed with spherical projection. The 2D feature selection is directly in the panorama (not in the original images). However, for each 2D point (including the endpoints of lines) we know which image it is from and its coordinates in that image according to the spherical projection.

Take one of the images as the reference image. All the selected 2D features can be mapped to this image according to the homography (obtained from the panorama stitching) between the reference image and the other images. Therefore, the panorama can be treated as a very large single image and the approach in [5] can be used to estimate its camera pose relative to the 3D model. This camera pose is of the reference image. The camera pose of any other image differs from it by a rotation estimated during the panorama stitching. A final bundle adjustment [7] is applied to refine all the intrinsic and extrinsic parameters together.

The images composing the panorama are stored in a database with their associated camera parameters. The online recalibration for a new image is based on its best matching image in the database (see section 3). Because the ability of the image matching technique to handle scale variation is limited, panoramas are generated at a set of different zooms so that the images saved in the database cover a large range of scale change. In the example given in section 4, we created panoramas at 4 zooms. The numbers of images used to generate the 4 panoramas are 60, 120, 240 and 480 respectively. The camera field of view at the largest zoom is about $1/8$ of that at the minimum zoom.

We do not create panoramas for very large zooms because the camera field of view will become so small that the number of images needed to cover the whole hemisphere of the pan-tilt range will be too large. The automatic panorama stitching will be very slow and not robust because the image matching may fail due to many images containing only textureless area. To still be able to calibrate the camera at very large zooms, the approach in [4] is applied to estimate the intrinsic parameters of the camera at every zoom setting. It is based on an image sequence captured in a fixed direction with the camera progressively zooming in. The obtained intrinsic parameters at different zoom settings are stored in a lookup table. The online re-calibration of the camera at very large zooms will take two steps as will be discussed in section 3.

3. ONLINE RE-CALIBRATION FOR A NEW PTZ SETTING

After the first step, users can pan, tilt and zoom the camera to an arbitrary target state. An image i is taken at the new PTZ setting. Its matching images are found in the database by using SURF feature matching [6]. To improve speed, only the images with their PTZ settings close to the new PTZ setting are checked. The image with the largest overlapping with i is the best matching image denoted as j. Take the camera frame of j as the reference coordinate system. The projection of a 3D point in this image is computed with:

$$\mathbf{x} = K\mathbf{X}, \tag{1}$$

where $\mathbf{x}$ is the homogeneous coordinate of the 2D point, $\mathbf{X}$ is the coordinate of the 3D point and K is the calibration matrix (intrinsic camera parameters) of j obtained in panorama calibration. Assume the rotation from the camera frame of j to the camera frame of i is R and the calibration matrix of i is K'. Then the projection of a 3D point in i is computed with:

$$\mathbf{x}' = K'R\mathbf{X}, \tag{2}$$

where $\mathbf{x}'$ is the homogeneous coordinate of the 2D point in i. It is easy to see that:

$$\mathbf{x}' = K'RK^{-1}\mathbf{x}. \tag{3}$$

Let $H = K'RK^{-1}$. It is the homography [2] from image j to image i and can be computed linearly from the feature correspondences between the two images. Therefore, we have:

$$K'R = HK, \tag{4}$$

from which K' and R can be calculated with LQ decomposition of the matrix HK.

Denote the rotation and translation from the coordinate system of the 3D model to the camera frame of image j as R_0 and T_0, which are computed in step one. The rotation and translation from the coordinate system of the 3D model to the current camera frame at the new PTZ setting are RR_0 and RT_0 respectively.

In the case the current zoom is so large that it exceeds the zoom range covered by the database, the online re-calibration is conducted with two steps. First, the camera is zoomed out to the largest zoom setting in the database without changing the direction. An image is taken at this zoom based on which the extrinsic parameters are estimated with the aforesaid approach. The camera is then zoomed back to the target zoom setting. Its intrinsic parameters are read from the lookup table created in step one. Compared to the direct re-calibration, the two-step approach is slower since zooming takes time and it is less accurate because changing zoom may slightly change the extrinsic camera parameters.

4. EXPERIMENTAL RESULTS

We did the experiments with the 3D model of USC campus. Fig.1(a) is part of the 3D model. Because the model is created from airborne Lidar, it does not have any texture. A PTZ camera is installed on the top of a building. The red dot represents the approximate camera location. Fig.1(b) shows a panorama created at the minimum zoom. Fig.1(c)-(e) are three different views of its projection onto the 3D model according to the camera parameters estimated in step one. We did not implement color normalization and blending so the color looks uneven but this does not affect the accuracy of the calibration. We can see the registration between the panorama and the 3D model is very good. Obviously, the panorama projection can be used as a high-resolution texture map of the 3D model.

Fig.1(f) is an image captured at a new PTZ setting. Its best matching image in the database is Fig.1(g). After the online re-calibration, it is projected onto the 3D model as shown in Fig.1(h). In Fig.1(i), its projection is overlaid on the texture map of the 3D model. We can see the alignment is very accurate.

Fig.1(j) is an image acquired at another different PTZ setting. Fig.1(k)-(m) are its best matching image, its projection and its projection overlaid on the texture map of the 3D model respectively.

The online re-calibration usually takes about 3 seconds including the time spent on image capture. The two-step re-calibration in which the camera zoom is changed to an optimal value and then changed back takes longer time depending on the zooming speed of the camera.

5. CONCLUSION

In this paper, an efficient approach to calibrate the PTZ cameras in an AVE system is presented. First, the intrinsic camera parameters for each zoom step are calibrated from panoramas. The extrinsic parameters of the images composing the panoramas are estimated from a set of 2D to 3D feature correspondences. The images and their camera parameters are stored in a database. The users can then pan, tilt and zoom the camera freely to a target position. The camera parameters at the new PTZ setting is re-calibrated online based on the best matching image found in the database.

The proposed algorithm is validated by our experiments. The panoramas generated in the first step can also be used as a high-resolution texture map of the 3D model. The online re-calibration for the camera at an arbitrary PTZ setting usually takes about 3 seconds. The projection of the current image (or video) is accurately aligned with the texture map of the 3D model, which satisfies the requirements of an AVE system.

6. REFERENCES

[1] U. Neumann, S. You, J. Hu, B. Jiang, and J.W. Lee, "Augmented virtual environments (ave): dynamic fusion of imagery and 3d models," *IEEE Virtual Reality Conference*, pp. 61–67, 2003.

[2] R. Hartley and A. Zisserman, ," *Multiple View Geometry in Computer Vision, second edition*, 2003.

[3] R.T. Collins and Y. Tsin, "Calibration of an outdoor active camera system," *Computer Vision and Pattern Recognition Conference*, pp. 528–534, 1999.

[4] S.N. Sinha and M. Pollefeys, "Pan-tilt-zoom camera calibration and high-resolution mosaic generation," *Computer Vision and Image Understanding*, vol. 103, pp. 170–183, 2006.

[5] L. Wang, S. You, and U. Neumann, "Single view camera calibration for augmented virtual environments," *IEEE Virtual Reality Conference*, pp. 255–258, 2007.

[6] H. Bay, T. Tuytelaars, and L.V. Gool, "Surf: Speeded up robust features," *European Conference on Computer Vision*, pp. 404–417, 2006.

[7] B. Triggs, P. Mclauchlan, R. Hartley, and A. Fiztgibbon, "Bundle adjustment: a modern synthesis," *Vison Algorithms: Theory and Practice, LNCS*, vol. 1883, pp. 298–372, 2000.

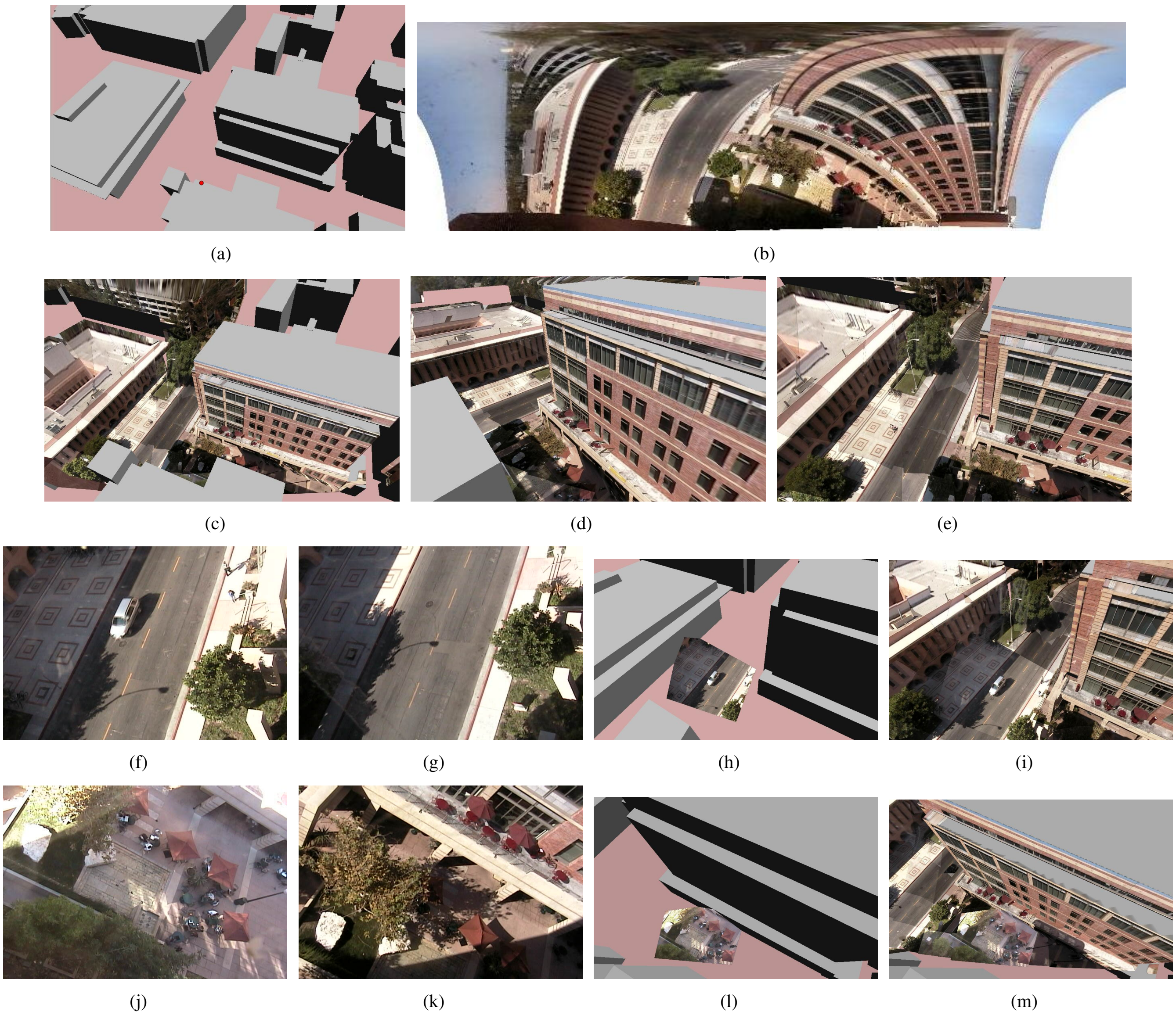

Fig. 1. (a):3D model. (b):The panorama at the minimum zoom. (c)-(e):Three views of the panorama projection onto the 3D model. (f):An image captured at a new PTZ setting. (g):Its best matching image in the database. (h):The projection of f onto the 3D model without showing the texture map of the 3D model. (i):The projection of f overlaid on the texture map. (j):An image captured at another new PTZ setting. (k):Its best matching image in the database. (l):The projection of j onto the 3D model without showing the texture map. (m):The projection of j overlaid on the texture map of the 3D model.