A 3D Dynamic Visualization Surveillance System

Yuguan Xing[1], Hiroshi Nagahashi[2] and Xiaolin Zhang[3]

[1,2] Department of Information Processing, Tokyo Institute of Technology
Yokohama, Kanagawa, 226-8503, Japan

[3] Shanghai Institute of Microsystem And Information Technology
Shanghai, 200050, China

Abstract

We present a 3D dynamic visualization surveillance system based on wide-area PTZ surveillance cameras. In our system, PTZ cameras are used to generate 3D site model automatically. By using the 3D spatial information, pedestrians can be identified their spatial positions by a single PTZ camera when they come into its view. Furthermore, based on the spatial position coordinates, the pedestrians detected from real-time surveillance video can be not only projected onto 3D site model, but also tracked by multiple cameras from different directions with different magnifications.

Keywords: *Visualization surveillance system, PTZ camera, 3D site model, Multi-view tracking.*

1. Introduction

With the advance of smart city construction, various kinds of surveillance camera networks have been playing an important role. As a smart city's main sensor, the surveillance camera can acquire and record a lot of data for information analysis. When the traditional 2D surveillance systems are used for security, human operators have to watch multi-view 2D videos captured by dozens of cameras installed at different positions, and they must understand the relationship between 2D videos and 3D real world. However, when the number of available cameras becomes large, it is difficult to integrate the information in separate 2D images. If the scenes monitored are similar like residential or supermarket scenes, the operators will be sometimes confused. Even if the system can automatically monitor and detect suspicious events, this kind of problem exists so long as human beings should deal with the events. Therefore, 3D visualization surveillance systems, which can provide a more intuitive feeling and reduce the operator's cognitive load, become a significant issue.

3D visualization in surveillance video has been widely studied. For example, Jvckenwood Corporation developed a 3D Surveillance Camera System [1] by simply displaying 2D videos taken by surveillance cameras into a 3D site model. The HouseFly system developed by P. DeCamp et al. [2] projected multiple sensors' videos onto an indoor 3D model without separating persons. In addition, some systems address the issue of showing moving targets onto a reconstructed 3D environment. Y. Yang et al. [3] named it a “seeing as it happens” system, and their system can generate a steerable first-person immersive replay. However, when their system was used in outdoor sites, they just projected a virtual avatar onto 3D environment instead of real pedestrian images. The early VSAM (Video Surveillance and Monitoring) system [4] can detect people, vehicles, and insert selected images of interest into an environment map. This system cannot show targets and virtual environment so clearly. The USC AVE (Augmented Virtual Environment) [5] system projected moving objects within the video sequence onto 3D background. However, because the models were rendered by the real-time video images, the rendering of 3D background was affected by moving objects.

On the other hand, 3D site model can be obtained in several ways. P. DeCamp [2] used CAD models which were built manually. By using some image processing methods such as Structure-from-Motion [6-9] and optical flow method [10], many scene images taken from different directions are used to compute a geometrical model. Moreover, some laser scanners or other advanced 3D sensors such as LiDAR [11, 12], Kinect [13, 14] and ZED [15], can generate 3D models of scenes as well. However, no matter how 3D site models are generated, these 3D models in the existing systems were only used as display platforms.

In order to make good use of 3D model and expand the function of 3D surveillance system, we present a 3D visualization system that can reconstruct 3D site models by using surveillance PTZ cameras and display pedestrians in a 3D representation dynamically. In addition, because 3D site models contain the spatial information of scenes, we can calculate pedestrians' coordinate values and achieve the multi-view tracking.

The advantages of this system can be summarized as follows:

IJCSI International Journal of Computer Science Issues, Volume 13, Issue 5, September 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.20943/01201605.3644

- It is different from other 3D surveillance systems that 3D models in our proposed system are generated by system's own PTZ cameras. Hence, the 3D site models are described in the coordinate system of the camera group, and the coordinate system of the camera group can be seen as the world coordinate system. However, in other methods, they have to estimate the transformation (scaling, translation, and rotation) between the coordinate system of 3D model and the world coordinate system.
- The 3D mapping reconstruction can be operated automatically by controlling surveillance cameras. Therefore, even if there are some changes in the scene, the 3D model can be updated easily.
- Generally, when we want to measure the spatial position of a person or something else with surveillance cameras, at least two cameras are necessary. By contrast, utilizing the 3D site model generated by the system's own surveillance cameras, our system enables to obtain the position by only one camera. So that, if one camera can provide pedestrians' spatial coordinate values consecutively, pedestrians can be not only projected onto 3D site model in real time, but also tracked by other cameras in telephoto state.

2. Method

2.1 Camera coordinate system

The spherical PTZ camera shown in Fig.1 was used in our proposed system. This kind of camera can meet two conditions: 1) the optical axis of a camera passes through the rotation center which is the intersection point of axes of its pan motion and tilt motion, 2) rotation angles of the camera can be fetched and read out at any time.

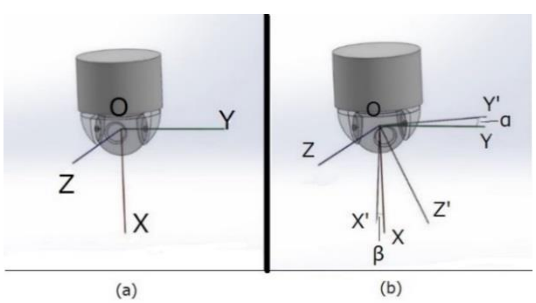

Figure 1 Coordinate system of a camera

Figure 1 (a) shows an initial coordinate system $o-xyz$ of the camera. The active surveillance camera performs pan motion (right and left movement) and tilt motion (up and down movement). However, once the camera rotates, the coordinate system will differ from its original position. As shown in Fig.1 (b), $o-xyz$ and $o-x'y'z'$ are assumed to be the initial coordinate system and a new one after rotating, respectively. The origins of the two coordinate systems are the same, and angle α between $y'-$ and $y-$axes gives the rotation angle of pan motion, angle β between $x'-$ and $x-$ axes also gives the rotation angle of tilt motion. In other words, the motion between $o-xyz$ and $o-x'y'z'$ can be considered as two rotations: rotation α in the pan direction, and rotation β in the tilt direction.

Assuming that there is a point P in a 3D space, the relation between the initial coordinate system and the new coordinate system after rotating can be calculated by the following equation.

$$\mathbb{P} = R\,\mathbb{P}' \tag{1}$$

where $\mathbb{P} = (X, Y, Z, 1)^T$ is a homogeneous coordinates of P in $o-xyz$, $\mathbb{P}' = (X', Y', Z', 1)^T$ is the coordinates of P in $o-x'y'z'$, and R is a transformation matrix. The transformation matrix R can be calculated with rotation angles α and β read from the camera encoder. We note that the rotation angles mentioned in this paper are relative to the initial coordinate system. If the optical axis aims at the point P after rotating α and β from the initial position, we stipulate that the point is located in the direction of α and β angles with respect to the initial camera position.

On the other hand, the relation between two cameras' initial coordinate systems is the same as that of initial coordinate systems before and after rotation of single camera. For example, the initial coordinate systems of two cameras are set to $l-xyz$ and$r-xyz$. $\mathbb{P}_l = (X_l, Y_l, Z_l, 1)^T$ and $\mathbb{P}_r = (X_r, Y_r, Z_r, 1)^T$ represent the homogeneous coordinate values of point P in the two coordinate systems. The point $\mathbb{P}_r$ can be transformed to $\mathbb{P}_l$ by a matrix T as shown in Eq. (2).

$$\mathbb{P}_l = T\mathbb{P}_r \tag{2}$$

In this paper, one camera's initial coordinate system was set to the world coordinate system as a reference. We consider multiple cameras more than two. So if the transformation matrices among initial coordinate systems of cameras are known, the status of each camera with respect to the world coordinate system can be calculated. Here, we used the L-shaped gauge method [16] to estimate the transformation matrices between each camera pair, which is suitable for wide-area PTZ camera's external parameter estimation. By the proposed new estimation method of external camera parameters, all cameras can be used to accurately measure distances and coordinates.

2.2 Spatial position calculation for pixels

In particular, when a camera rotates to make its optical axis pass through the point P, its new coordinate value can be written as $\mathbb{P}' = (0,0,L,1)^T$, where L is the distance from

IJCSI International Journal of Computer Science Issues, Volume 13, Issue 5, September 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.20943/01201605.3644

the point to the origin of coordinate system. By substituting it for $\mathbb{P}'$ in Eq. (1), the coordinate value of the point P in the initial coordinate system is calculated as follows.

$$\mathbb{P} = R\begin{pmatrix}0\\0\\L\\1\end{pmatrix} \tag{3}$$

Suppose that two cameras make their optical axes aim at a same point P, then the following equations can be given.

$$\mathbb{P}_l = R_l\begin{pmatrix}0\\0\\L_l\\1\end{pmatrix}, \quad \mathbb{P}_r = R_r\begin{pmatrix}0\\0\\L_r\\1\end{pmatrix} \tag{4}$$

By substituting them in Eq. (2), we can get Eq. (5). If the point's rotation angles with respect to two cameras are known, R_l and R_r can be calculated. Then, it is not difficult to solve Eq. (5) and calculate L_l and L_r, i.e., the distances from the point to each camera. Next, the coordinate value of this point can be obtained by using Eq. (3).

$$R_l\begin{pmatrix}0\\0\\L_l\\1\end{pmatrix} = TR_r\begin{pmatrix}0\\0\\L_r\\1\end{pmatrix} \tag{5}$$

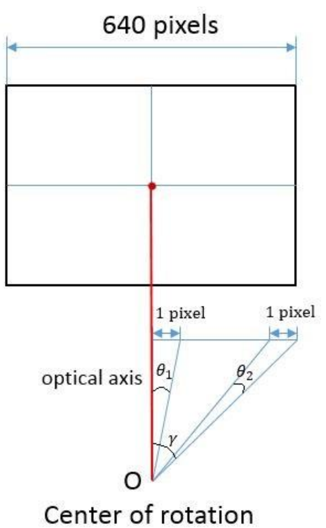

Figure 2 Rotation angles corresponding to pixels

In order to calculate the rotation angle of a pixel which is not in the center of image, we measured the angle of rotation between two adjacent pixels. As shown in Fig.2, γ is the rotation angle in pan direction from the center to the edge of the image, θ_1 is the angle corresponding to one pixel located in the center, and θ_2 is the angle corresponding to one pixel located in a peripheral area. The resolution of the camera is 640 × 480 pixels. When γ is 15 °, we can calculate $\theta_1 = 0.0048$ ° and $\theta_2 = 0.0045$ °. Their difference 0.003 ° is much smaller than the camera's accuracy of 0.03 °, so we considered it can be neglected. In our experiments, γ is kept to be smaller than 15 °, so the difference between θ_1 and θ_2 is always smaller than 0.003 °. Further, since the pixel's sizes in both horizontal and vertical directions are the same, we assumed that each pixel has approximately equal rotation angle θ in both horizontal and vertical direction whether it is in the center or in the peripheral area of an image.

Based on the above assumptions, we can measure θ for different zoom values. First, when zoom is a specific value, we make the center of camera aim at a spatial point P as shown in Fig. 3(a), and record the pan value α_P. Then, rotate the camera horizontally to move P's projection to the edge of the image, and record the current pan value α_1. Next, γ can be calculated as $\gamma = \alpha_P - \alpha_1$, and θ can be calculated as $\theta = \gamma/320$. We changed the zoom value and measured θ ten times, during which γ changed from 16.32 ° to 3.84 °. If the spatial point P becomes larger than one pixel as the magnification increases, we can choose different spatial points for this measurement.

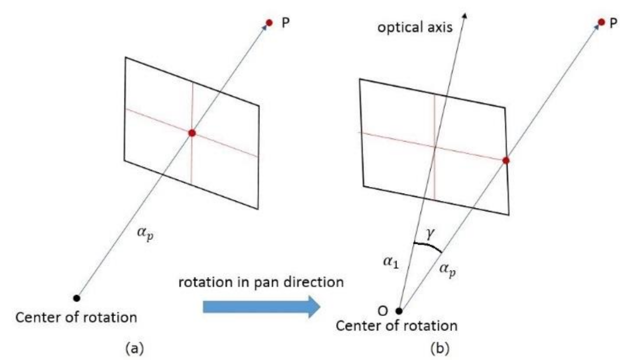

Figure 3 Relationship between the rotation angle and the image pixel

Similarly, a point's rotation angle with respect to camera's initial position can calculated by using its coordinates on the image plane. For example, as shown in Fig. 4, P_1 is the projection of a spatial point P on the image plane of Camera 1, so the rotation angle of P can be written as:

$$\begin{cases}\alpha_{P1} = \alpha_1 + \theta \times i\\ \beta_{P1} = \beta_1 + \theta \times j\end{cases} \tag{6}$$

where α_{P1} and β_{P1} are the point P's rotation angle with respect to the initial position of Camera 1, α_1 and β_1 are the rotation angles of Camera 1 with respect to its initial position, i and j are the number of pixels from the image center to P_1 in horizontal and vertical directions, and θ is the angle corresponding to one pixel. By using this method, P's rotation angle with respect to Camera 2 can be calculated too.

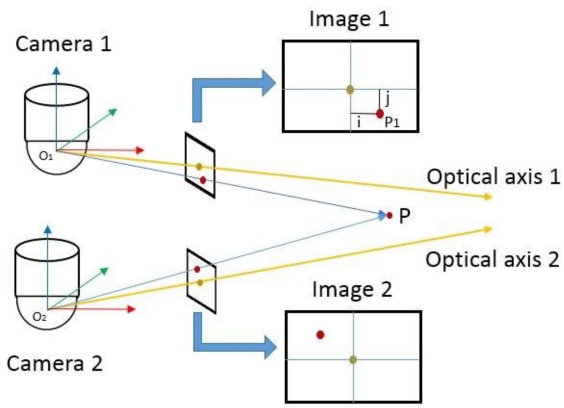

Figure 4 Rotation angle estimation

So far, rotation angles of any pixels can be computed by using their coordinate on the image plane. In summary, if the point's coordinates in two camera images are known, the point's rotation angles with respect to two cameras can be calculated, and then its coordinate value in world coordinate system can be obtained.

As previously described, the difference between our system and existing systems [1-5] is that we appropriately use system's own PTZ cameras to get spatial information of scene and generate 3D site model. Since the cameras and the scene are fused in one coordinate system, coordinates transformation is not required. Moreover, there are additional features in our system: 1) it is enough to do the external parameter estimation only once. It means that, in-between camera rotation does not need to do the parameter estimation again nor to do the self-calibration [17-19]. 2) 3D coordinate values of a point can be calculated only by using its rotation angles with respect to two cameras.

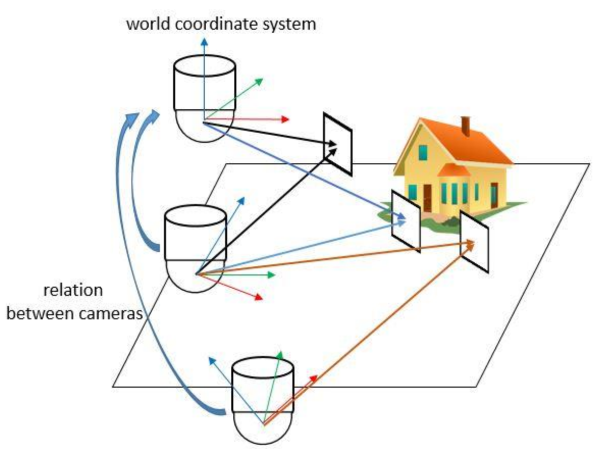

Figure 5 Generating 3D site model with surveillance camera pairs

Therefore, if two cameras aim at the same area, and the corresponding points in the two cameras' images can be extracted, the 3D site model of this area can be created. Furthermore, as indicated in Fig.5, two cameras can generate the 3D site model for a larger area though its scanning. In other words, if an area is seen by two cameras, its 3D site model can be obtained. The 3D site model mentioned in this paper refers to the 3D point cloud model. Each point in the 3D point model contains the spatial information (the 3D coordinates) and the color information.

2.3 Pedestrian location

In the conventional methods, it is necessary to use multiple cameras to calculate the spatial positions of targets. In this paper, because we have generated 3D site models which contain scenes' spatial information, pedestrian locations can be estimated by a single camera.

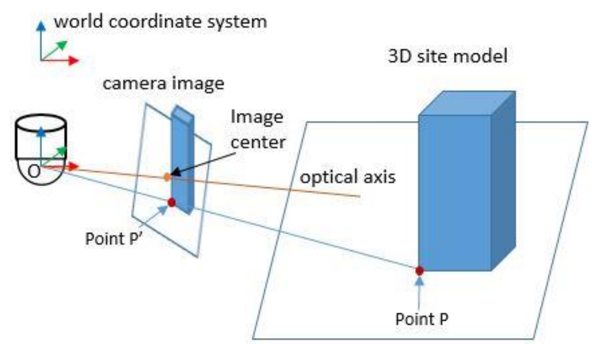

Figure 6 Position estimation for a point

Since the 3D site model, which contains the spatial information of real world, is generated by the surveillance cameras in advance, the locations and orientations of each camera, the 3D site models and the targets in the real world are all described in the same coordinate system. It means that it is not necessary to do the transformations among different coordinate systems. So that, if we want to know spatial positions of points in the image, we can just find its corresponding position on 3D site model. For example, as shown in Fig.6, in order to find the position of point P' in the 3D site model, camera's orientation information (represented by α_1 and β_1 in Eq.(6), the rotation angle corresponding to one pixel (represented by θ in Eq.(6)) and coordinates of P' on image plane are substituted into Eq.(6) to locate the direction of vector $\overrightarrow{OP'}$, and then its intersection (point P) with the 3D site model can be calculated.

When we want to estimate a pedestrian's spatial position, considering that persons usually walk upright and their feet contact the ground, we can obtain a person's position if the coordinate value of the contact point can be estimated from 3D site model. As shown in Fig.7, a walking person is

IJCSI International Journal of Computer Science Issues, Volume 13, Issue 5, September 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.20943/01201605.3644

detected in the camera image, and point A' is the midpoint of the bottom of the bounding box. According to previously described, the vector $\overrightarrow{OA'}$ can be calculated by using the rotation angles of cameras and the off-axis angle of line OA', furthermore, its intersection (point A) with the 3D site model can be calculated as well. Since every point in the 3D site model contains spatial information of real world, the 3D coordinate value of the point A can be considered the person's coordinate value.

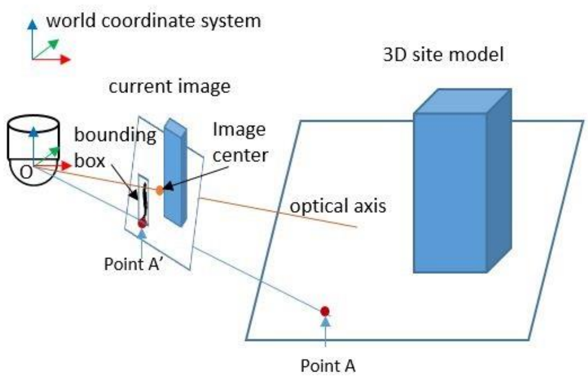

Figure 7 Position estimation for a pedestrian

In short, if a pedestrian is captured by one camera of the surveillance system, his/her coordinates can be obtained. Furthermore, by using the 3D coordinates, other cameras in this system can track him/her from different directions.

3. Experiment

3.1 Environment of experiment and 3D model generation

Figure 8 VN-V686 camera

We did some experiments to evaluate this approach. The VN-V686 PTZ network camera shown in Fig.8 released by JVC was used in the experiments. The camera has a CCD sensor with 380,000 pixels and a 36x zoom lens. With an accuracy of 0.03 ˚, it can precisely maintain camera positions even after frequent panning and tilting. Its pan range is [0 ˚, 360 ˚] and its tilt range is [-5 ˚, 185 ˚].

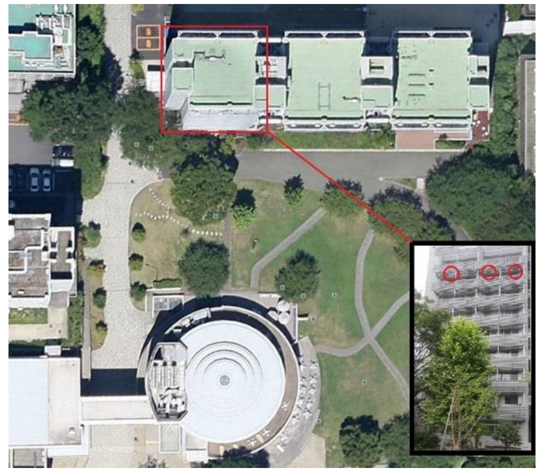

Figure 9 Environment of experiment

Figure 9 shows the environment where the experiments were carried out. Three cameras (shown by red circles) are installed on the eighth floor of the building. The distances between the cameras are greater than 4.5 meters. The ground is far from the cameras more than 30 meters.

According to the method described above, if surveillance cameras can be used to accurately measure the spatial information of scenes and generate 3D site models, the single-camera geolocation and multi-view tracking can be achieved based on the 3D site models. So we defined the world coordinate system, and estimated the external parameters of cameras. It is easy to make two cameras aim at the same point to obtain the maximum common field of views. Next, by using a template matching method to find dense corresponding points, we generated the 3D site model. Then, we removed their noise points and smoothed the rest 3D points with a Gaussian filter of σ = 5.0.

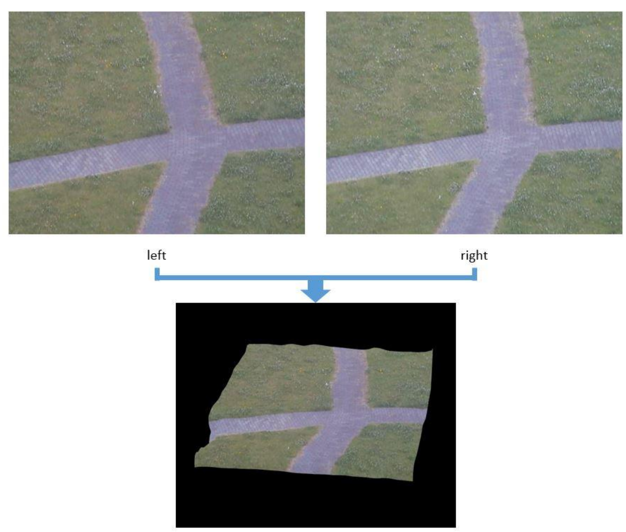

Figure 10 3D site model for a lawn field

IJCSI International Journal of Computer Science Issues, Volume 13, Issue 5, September 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.20943/01201605.3644

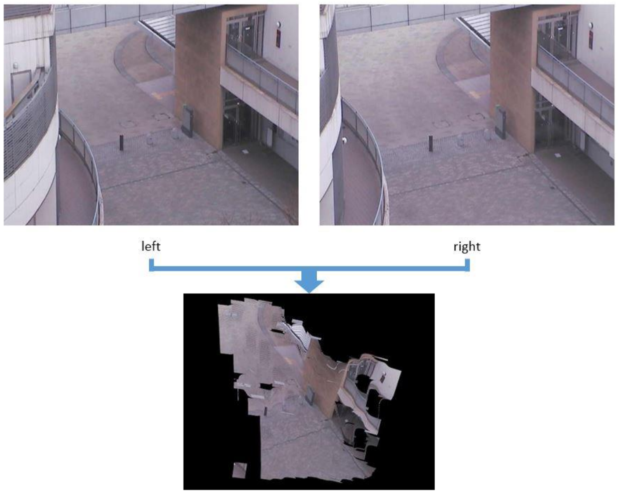

Figure 11 3D site model for a road

Figures 10 and 11 show two experiments of 3D site model generation with different zoom values. The two upper images in Fig.10 are lawn images taken from a camera pair, and lower one is the 3D site model generated by these two images. As you can see, except for the edge portion, the reconstructed surface of this lawn is relatively flat.

Comparing with Fig.10, the scene in Fig.11 is a little bit complicated. Some of left and upper parts cannot be matched. In addition, as for the first floor wall and the second floor aisle in the right side of images, there are some empty parts because they are not seen due to the limitation of camera orientations. Even so, the ground and the brown wall, the major part of this picture, are relatively reconstructed.

Based on these 3D site models and the spatial information they contain, we carried out 3D dynamic display experiment and multi-view tracking experiment.

3.2 3D dynamic display experiment

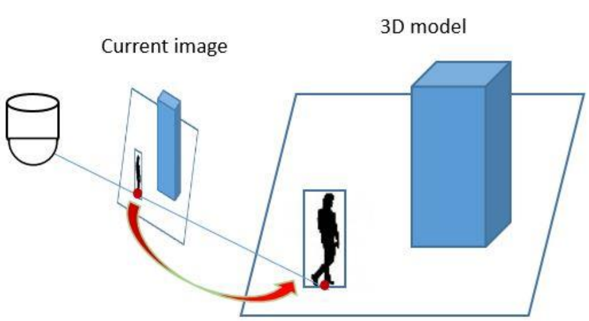

Figure 12 Rendering a pedestrian onto 3D environment

The purpose of 3D dynamic visualization is not only to provide an intuitive way of monitoring areas, but also to test whether the spatial coordinate values of targets can be calculated accurately by using the method described above.

As discussed in Section 2.3, by using the 3D site model, we can calculate the spatial positions of walking persons even if they are captured by only one camera. Figure 12 illustrates that a moving person is detected from every frame of original video and rendered onto a 3D site model. In this paper, we rendered the detected target image as a planar box at appropriate positions in the 3D site model. Here, we used MHI (motion history image) [20] method to detect moving persons, and the orientation and velocity of the target were estimated by the method.

We carried out the experiments in two different scenes. Walking persons were detected and their bounding boxes with textured images (from camera video frames) were projected onto the 3D site models. In order to show that the pedestrian location can be estimated at different camera orientation and magnification, we constantly changed the orientation and the magnification during the experiments. In both Fig.13 and 14, the left columns show the surveillance video images with different rotation angles and zoom values, and the right ones show the real-time dynamic 3D site models in which pedestrians' images are projected on. From these results, we can see that even if the camera's orientation and zoom value changes, the pedestrians can be put at correct positions in the 3D site model.

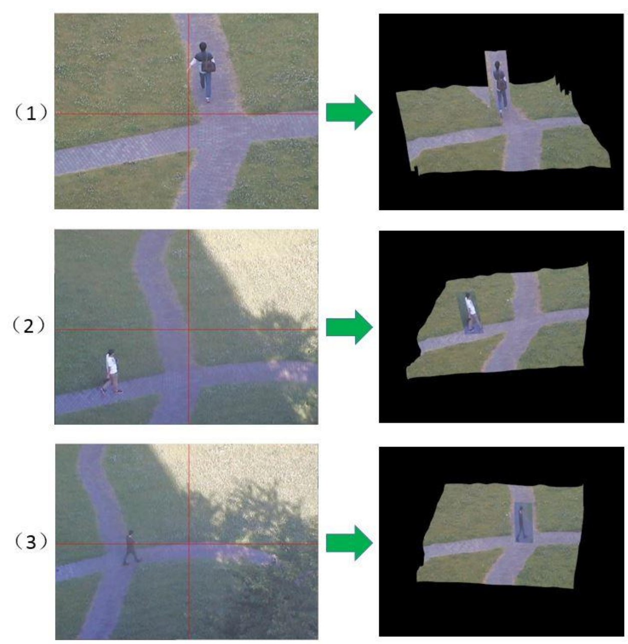

Figure 13 Visualization of pedestrians with different orientation and magnification on a lawn field

IJCSI International Journal of Computer Science Issues, Volume 13, Issue 5, September 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.20943/01201605.3644

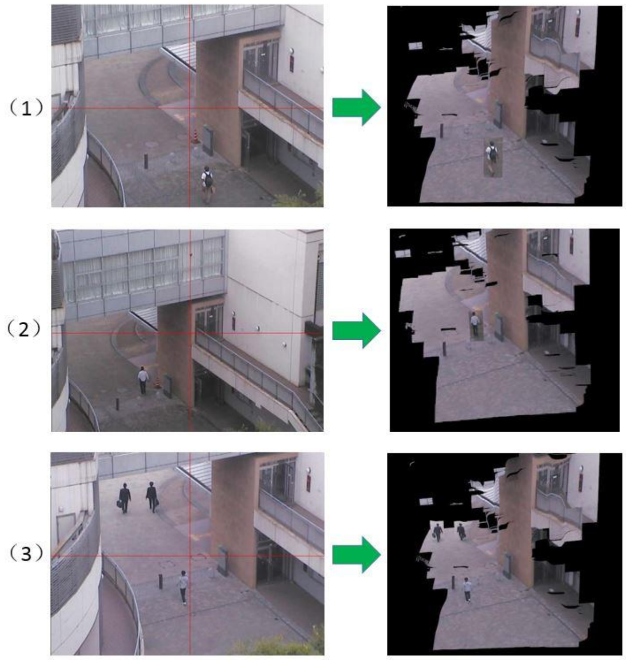

Figure 14 Visualization of pedestrians with different orientation and magnification on a road

3.3 Multi-view tracking experiment

Based on the above two experiments, we realized a multi-view tracking function. As shown in Fig.15, we used the master camera to calculate a pedestrian's position projected onto the ground (shown as point A). Then, a length of t (t is set to 1.2m) is added to the point A in the vertical direction to get the approximate location of pedestrian's body (shown as point B). Next, the slave camera is controlled to aim at the point B. By repeating the above steps for each frame of the master camera, we can achieve a continuous new tracking. As shown in Fig.16 and 17, the master camera is set to wide-angle state to get a wide field of view and the slave camera is set to telephoto state to get enlarged images which can show details more clearly.

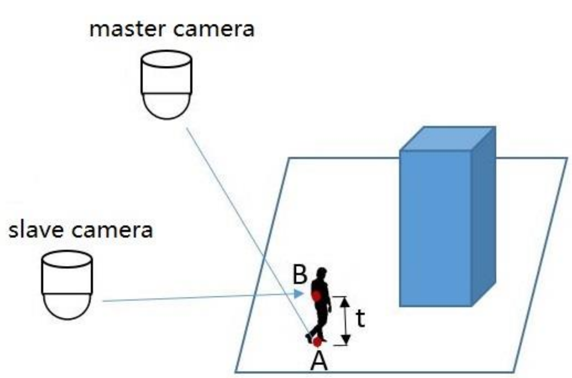

Figure 15 Multi-view tracking

Figure 16 Tracking example 1

Figure 17 Tracking example 2

In the actual experiments, we need to predict the position of pedestrians to offset the delay of camera motion. Further, since the targets are far from the cameras more than 50 meters, camera rotation error also causes a relatively large effect. For these reasons, the optical axis could not aim at pedestrian's waist stably, but floated in the range from chest to thigh.

Comparing with other long distance tracking methods [21-25], our method has two advantages: 1) even if the line of sight of the slave camera is blocked, the tracking can be continued so long as the master camera can see the target; 2) the slave camera' deviation angles do not affect the tracking. It means that the slave cameras can track the target from quite different directions. These features are derived from the reason that the slave camera controls its tracking based on 3D locations rather than images.

4. Conclusions

This paper presented a 3D visualization surveillance system based on wide-area PTZ surveillance cameras. Because cameras' motions can be precisely controlled, they were used not only to generate 3D site model automatically, but also used to calculate the detected pedestrian's spatial position. In this method, the pedestrians detected from real-time surveillance video can be projected onto 3D site model. Furthermore, based on the position coordinates, targets can be tracked by multiple cameras from different directions with different magnifications. Future work of this research includes the improvements on 3D site model, such as making models of wider scenarios and improving the integrity of the models. Moreover, some recognition techniques will be adopted to achieve 3D trajectory recording for specific persons.

References

[1] Jvckenwood Corporation, http://www.diginfo.tv/v/12-0041-d-en.php, 2012
[2] P. DeCamp, G. Shaw, R. Kubat, D. Roy, An immersive system for browsing and visualizing surveillance video, MM. ACM,2010, pp.371-380
[3] Y. Yang, MC Chang, P. Tu, S. Lyu, Seeing as it happens: Real time 3D video event visualization, ICIP, 2015
[4] T. Kanade, R. Collins, A. Lipton, P. Burt, L. Wixson: advances in cooperative multi-sensor video surveillance, Proc. of DARPA Image Understanding Workshop, 1998, Vol.1, pp.3-24
[5] I. O. Sebe, J. Hu, S. You, U. Neumann, 3D video surveillance with augmented virtual environments, in: First ACM SIGMM International Workshop on Video Surveillance. ACM, 2003, pp.107-112
[6] C. Tomasi, T. Kanade, Shape and motion from image streams under orthography: a factorization method, International Journal of Computer Vision, 1992, 9-2, 137-154
[7] M. Jancosek, T. Pajdla, Multi-view reconstruction preserving weakly-supported surfaces, CVPR.IEEE, 2011
[8] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, Visual modeling with a hand-held camera, IJCV, 2004, vol. 59, no. 3, pp.207-232
[9] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, R. Szeliski, Building Rome in a day, Commun. ACM, 2011, vol. 54, no. 10, pp.105-112
[10] Y. Diskin, Vijayan K. Asari: Dense 3D point-cloud model using optical flow for a monocular reconstruction system, IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2013
[11] A. Sampath, J. Shan, Segmentation and reconstruction of polyhedral building roofs from aerial Lidar point clouds, IEEE Transactions on Geoscience and Remote Sensing, 2010, VOL. 48, NO. 3
[12] S. Fleck, F. Busch, P. Biber, W. Straber, 3D surveillance a distributed network of smart cameras for real-time tracking and its visualization in 3D, CVPR Workshop. IEEE, 2006, pp. 118–118
[13] Y. Zou, W. Chen, X. Wu, Z. Liu, Indoor localization and 3D scene reconstruction for mobile robots using the Microsoft Kinect sensor, IEEE 10th International Conference on Industrial Informatics, 2012
[14] J. Zhang, Q. Huang, X. Peng, 3D reconstruction of indoor environment using the Kinect sensor, Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control, 2015
[15] S. Pillai, S. Ramalingam, J. J. Leonard, High-performance and tunable stereo reconstruction, International Conference on Robotics and Automation (ICRA), 2016
[16] Y. Xing, H. Nagahashi, X. Zhang, K. Aoki, An external parameter estimation method for wide-area active surveillance cameras, IIEEJ Transactions on Image Electronics and Visual Computing, 2016, Vol.4 No.1
[17] Z. Zhang, Q. Luong, O. Faugeras, Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction, IEEE Transactions on Robotics and Automation, 1996, VOL. 12, NO. 1
[18] P. Sturm, Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97), 1997
[19] M. Pollefeys, R. Koch, L. V. Gool, Self-calibration and metric reconstruction in spite of varying and unknown intrinsic camera parameters, International Journal of Computer Vision, 1999, 32(1), 7–25
[20] J.W. Davis, A.F. Bobick, The recognition of human movement using temporal templates, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23 (3): 257–267
[21] X. Zhou, R. T. Collins, T. Kanade, P.r Metes, A master-slave system to acquire biometric imagery of humans at a distance, ACM SIGMM 2003 Workshop on Video Surveillance, 2003, pp.113-120
[22] J.A.Fayman, O.Sudarsky, E.Rivlin, M.Rudzsky, Zoom tracking and its applications, Machine Vision and Applications I3 (I), 2001, pp.25-37
[23] X. Zhang, An object tracking system based on human neural pathways of binocular motor system, Proc. ICARCV, Singapore, 2006, pp.285-292
[24] Y. Gao, X. Zhang, Zoom control of wide area tracking system, IEEE International Conference on Intelligent System, London, 2010, pp.402-407

[25] Y. Gao, X. Zhang, Active zoom cameras control system using stereo parallax, IEEE International Carnahan Conference on Security Technology (ICCST), 2010

Yuguan Xing received his B.E. degree from Sichuan University, China, in 2008, and M.Eng. degree from Tokyo Institute of Technology in 2013. Since 2013, he has been with Department of Information Processing at Tokyo Institute of Technology, where he is studying for Ph.D. degree. His research interests include camera calibration, video surveillance, and image processing.

Hiroshi Nagahashi received his B.S. and Ph.D. degrees from Tokyo Institute of Technology in 1975 and 1980, respectively. Since 1990, he has been with Imaging Science and Engineering Lab oratory at Tokyo Institute of Technology, where he is currently a professor. His research interests include computer vision, computer graphics, and geometric modeling. He is a member of IEEE, the Institute of Electronics Information and Communication Engineers, the Information Processing Society of Japan, and the Institute of Image Electronics Engineers of Japan.

Xiaolin Zhang received his B.E. degree in Power Systems from Northeast China Institute of Power Engineering in 1985, and M.S.E. and Ph.D. degrees from Yokohama National University in 1989 and 1995, respectively. From 2003 to 2013, he served as an associated professor and later as a full professor in Precision Engineering Institute, Tokyo Institute of Technology. In 2012, he became a researcher and a Ph. D supervisor at Chinese Academy of Sciences. His research interests include robot vision, artificial intelligence, medical engineering, virtual reality, and image recognition.