

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A Wide-area Surveillance System Using Multiple Active Cameras
著者(和文)	ケイ玉冠
Author(English)	Yuguan Xing
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10548号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:長橋 宏,熊澤 逸夫,小池 康晴,山口 雅浩,小尾 高史
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10548号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Doctoral Dissertation

A Wide-area Surveillance System Using Multiple Active Cameras

Yuguan Xing

Department of information processing,
Tokyo Institute of Technology

Supervisor: Professor Hiroshi Nagahashi

March 2017

Abstract

Nowadays, in order to make the cities more comfortable and safe, many countries have been promoting the smart city construction. One of the primary objectives is to realize advanced information analysis by obtaining data from various sensors connected via network. As the main sensor of smart city, the surveillance active cameras are widely used for crime prevention or disaster prevention. However, separate control for each camera limits the entire system to play a more effective function. So it is still an important task to make cameras cooperate with others to improve the efficiency and convenience of the surveillance system.

This dissertation presented a wide-area surveillance system whose cameras are integrated organically by using a new external parameter estimation method. In this method, an L-shaped tool is used as a gauge, and the optical axis of the camera is adjusted to each endpoint of the L-shaped gauge. Then we can use only the information of the optical axis to calculate the external parameters, so that the accuracy does not depend on the lens distortion or the image processing error.

Based on the estimated external parameters, the cameras can be controlled to measure the spatial information of the surveillance area and generate the 3D site model. 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.

Acknowledgments

Special thanks are due to my supervisor Professor Hiroshi Nagahashi for his invaluable support and guidance through the hard moments of graduate school.

I am grateful to my parents for their love and encouragement, without whom I would never have an opportunity to study abroad.

I also thank the Rotary Yoneyama Memorial Foundation and Department of information processing, Tokyo Institute of Technology, for their financial support over the years.

Contents

Abstract	i
Acknowledgments	ii
List of figures	vi
List of tables	vii
1 Introduction	1
1.1 Research background	1
1.2 Research purpose	4
1.3 Related works	6
1.4 Outline	9
2 Proposed surveillance system	10
2.1 Whole image of the proposed surveillance system	10
2.2 External parameters of active PTZ camera	13
3 External parameter estimation	21
3.1 Existing external parameter estimation method	21
3.2 Proposed external parameter estimation method	24
3.2.1 The gauge used for distance estimation	24
3.2.2 Estimation of distances	26
3.2.3 Estimation of external parameter	28
3.3 Evaluation experiments	31
3.3.1 Synthetic experiment	31
3.3.2 Real data experiment	35
4 The generation of 3D site model and dynamic 3D display	43
4.1 Spatial position calculation for pixels	43
4.2 3D reconstruction	47
4.3 Position estimation by one camera	50
4.4 Experiment of dynamic 3D display	54
4.4.1 Experiment of 3D model generation	54
4.4.2 3D dynamic display experiment	55

5	Multi-view tracking	60
5.1	Extraction of pedestrians	60
5.2	Multi-view tracking experiment	64
6	Conclusion	70
	Publications	72
	References	73

List of Figures

1.1	Two types of surveillance camera	1
1.2	An operator watching screens	2
1.3	Similar sences	2
1.4	3D surveillance system developed by Jvckenwood Corporation . . .	6
1.5	Image of Y. Yang' system	7
1.6	Video surveillance and monitoring system	7
1.7	Augmented virtual environment system	8
2.1	Whole image of the proposed system	10
2.2	The composition of the system	11
2.3	VN-V686 camera	13
2.4	Initial coordinate system of a camera	14
2.5	Coordinate system after rotation	15
2.6	Pan rotation	16
2.7	Tilt rotation	17
2.8	Two Coordinate Systems	19
3.1	A checkerboard	21
3.2	DLT method	22
3.3	3 axis frame	23
3.4	L-shaped gauge	24
3.5	Flowchart of process	25
3.6	Coordinate system and L-shaped gauge	26
3.7	Configuration of L-shaped gauge	26
3.8	Step of angle measure	27
3.9	Simulation environment	31
3.10	Mean error for distance estimation experiment conducted without disturbance	32
3.11	Mean error for distance estimation experiment conducted with dis- turbance	33
3.12	Environment of actual experiment	35
3.13	L-shaped gauge taken by a camera	36
3.14	Result of distance measurement	37
3.15	Flowchart of multi-directional shooting	38
3.16	Multi-directional shooting experiment	39
3.17	Circle of radius 0.25m and upper body	42

4.1	Rotation angles corresponding to pixels	44
4.2	Relationship between the rotation angle and the image pixel	45
4.3	Rotation angle estimation	46
4.4	Block matching aided by epipolar line	47
4.5	Different segmentations for block matching	48
4.6	Generating 3D site model with surveillance camera pairs	49
4.7	Position estimation for a point	50
4.8	Position estimation for a pedestrian	51
4.9	Experiment of distance estimation	52
4.10	3D site model for a lawn field	54
4.11	3D site model for a road	55
4.12	Rendering a pedestrian onto 3D environment	56
4.13	Visualization of pedestrians with different orientation and magnifi- cation on a lawn field	57
4.14	Visualization of pedestrians with different orientation and magnifi- cation on a road	58
5.1	Example of MHI	61
5.2	Marking a moving person	61
5.3	MHI of a pedestrian	62
5.4	Adjacent pixels and corresponding vectors	62
5.5	Position prediction	63
5.6	Multi-view tracking	64
5.7	Tracking example 1	65
5.8	Tracking example 2	66
5.9	The site of the experiment	67
5.10	Master camera switching experiment	68

List of Tables

3.1	Data of target points	29
3.2	Statistical results	34
4.1	Results of distance measurement	52

Chapter 1

Introduction

1.1 Research background



Figure 1.1: Two types of surveillance camera

At present, as the main sensor of smart city construction, the use of surveillance cameras is increasing [1]. The surveillance camera can take and record a lot of video data for security prevention, disaster prevention, information analysis, and so on. Most of wide-area surveillance cameras can be classified into two types as shown in **Fig.1.1**: fixed camera and active PTZ camera. The fixed camera which can not move can cover a wide range, they can not move Fixed cameras are often used to observe large areas. Comparing with the fixed cameras, the active PTZ cameras have been used more and more, because they can be used to make up for blind spots of the

fixed camera and to pay more attention to some important targets.



Figure 1.2: An operator watching screens

When the traditional 2D surveillance systems in which the movement of each camera is individually controlled are used for security, people usually have to watch dozens of videos which are captured by cameras installed at different positions (**Fig.1.2**). Further, operators must remember where the cameras are installed.



Figure 1.3: Similar sences

However, if the number of available surveillance cameras is large, integrating information from separate 2D images becomes difficult. Further, if the monitored scenes are similar, such as residential, supermarket or hotel scenes shown in **Fig.1.3**, the operators will be more easily confused, and more likely to make mistakes.

Even if some of the latest surveillance systems can monitor and detect suspicious events automatically, as long as human beings have to be involved in dealing with the events, this kind of problem still exists. Therefore, the development of 3D visualization surveillance systems, which can improve monitoring efficiency and reduce the people's cognitive load, become a significant issue.

1.2 Research purpose

The purpose of this research is to improve availability and effectiveness of surveillance system by integrating the multiple cameras so that the system can complete some functions that are not available in traditional systems.

In the proposed surveillance system, there should be the following functions:

1. The cameras can be controlled to cooperate with each other. The camera group fused organically can do some things that can not be done with dispersive cameras.
2. The 3D site models of monitoring area can be reconstructed by using system's own cameras.
3. Moving objects such as pedestrians or vehicles can be displayed in 3D site model dynamically so that the important information can be concentrated. Needless to say, it is easier for the operators to concentrate their attention. However, this function also makes sense even for the unmanned surveillance systems. For example, when some events occurred and people want to view the video record, the 3D videos can visually and centrally show the information related to the events. Compared to watching 2D videos, people can more easily grasp the overall situation.
4. Specific objects can be tracked by multiple cameras. With this function, the system can continuously provide position information and enlarged images of the specific target. Needless to say, it is not necessary to track all moving objects. However, because there is always a master camera monitoring the entire region, the system does not miss any other information when multiple cameras are tracking a target.

The advantages of this system are summarized below:

- By controlling the system's own surveillance cameras, the 3D mapping reconstruction can be operated automatically. Based on this feature, even if some parts of the surveillance area have changed, 3D models can be updated timely and easily. The traditional methods can not do this.

- In our proposed surveillance system, the coordinate system of the camera group is set as the world coordinate system, and 3D models are generated by system's own PTZ cameras, so the 3D site models and the cameras are described in the same coordinate system. However, in other methods, 3D models are generated separately, so they have to calculate the transformation including scaling, translation and rotation between the world coordinate system and the coordinate system of 3D models.
- In traditional method, at least two cameras are necessary to measure the spatial coordinates of a target. By contrast, because the 3D site model in our system contains the spatial information of the surveillance area, it is possible to obtain the position by only one camera. Based on the spatial coordinates provided by one camera, moving objects can be projected onto 3D site model and tracked continuously by other cameras.

1.3 Related works

In this section, we will introduce the related research status about the application of 3D visualization in surveillance systems. As shown in **Fig.1.4**, a 3D Surveillance Camera System [2] developed by JVCKENWOOD CORPORATION simply displayed 2D videos of surveillance cameras into a 3D site model. Since they did not merge the 2D videos with the 3D model, the 2D videos were still dispersive. And the 2D videos became too small to see clearly.



Figure 1.4: 3D surveillance system developed by Jvckenwood Corporation

In the HouseFly system [3] developed by P. DeCamp et al., multiple sensors' videos were projected onto an indoor 3D model without separating persons.

Some systems addressed the issue of showing moving targets onto a reconstructed 3D site model. The system developed by Y. Yang et al. was called “seeing as it happens” system [4]. However, as you can see in **Fig.1.5**, they just projected virtual avatars onto 3D site model instead of real people images when their system was used in outdoor sites.

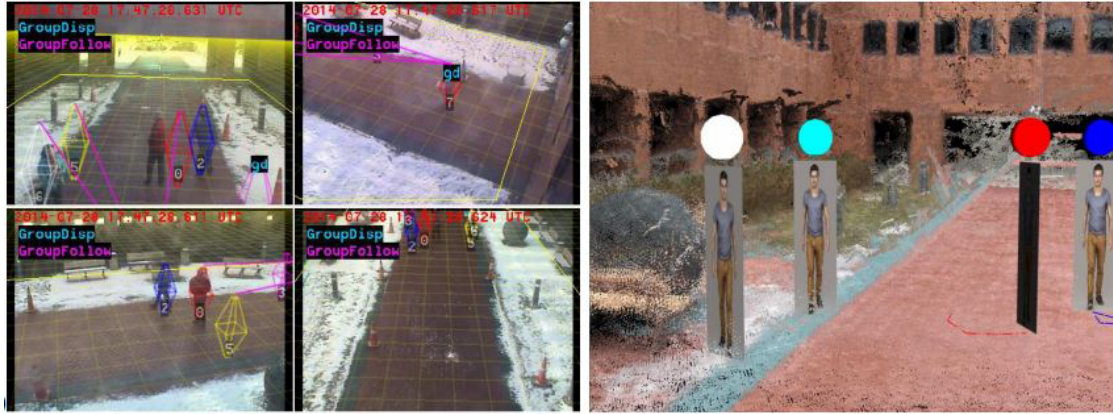


Figure 1.5: Image of Y. Yang's system

Figure.1.6 shows the early Video Surveillance and Monitoring (VSAM) system [5]. In this system, selected images of interest can be inserted into an environment 3D map. However, the virtual environment and the targets shown by this system were unclear.



Figure 1.6: Video surveillance and monitoring system

Figure.1.7 shows the Augmented Virtual Environment (AVE) [6] in which moving objects within the video sequence were projected onto 3D site model. However, as you can see, because the 3D site models were rendered by the real-time video images, there are some vacancies on the 3D background behind moving objects.



Figure 1.7: Augmented virtual environment system

On the other hand, 3D site models used as display platforms can be generated in several ways. For example, CAD models used by JVCKENWOOD CORPORATION [2] and P. DeCamp [3] were built manually. Some image processing methods, such as Structure-from-Motion [7, 8, 9, 10] and optical flow method [11], can be used to generate geometrical models. In addition, by using advanced 3D sensors such as laser scanner, LiDAR [12, 13], Kinect [14, 15] and ZED [16], 3D site models can be generated as well.

1.4 Outline

The dissertation contains six chapters, of which the first is the introduction.

Chapter 2 illustrates the whole image of the proposed surveillance system and its functions. Then, since obtaining external parameters of cameras is the basis for all the functions, we briefly discuss the meaning of external parameters.

Chapter 3 introduces the existing external parameter estimation method. After that, we illustrate a new external parameter estimation method and the evaluation experiments for this method.

Chapter 4 presents the process of 3D reconstruction and the dynamic 3D display function of the system.

Chapter 5 illustrates the theory of the multi-view tracking function and the experiments carried out.

Chapter 6 concludes the entire work with remark and future study.

Chapter 2

Proposed surveillance system

2.1 Whole image of the proposed surveillance system

As previously described, we proposed a new surveillance system which can achieve some functions that are not available in traditional systems. Firstly, we would like to introduce the composition of this system.

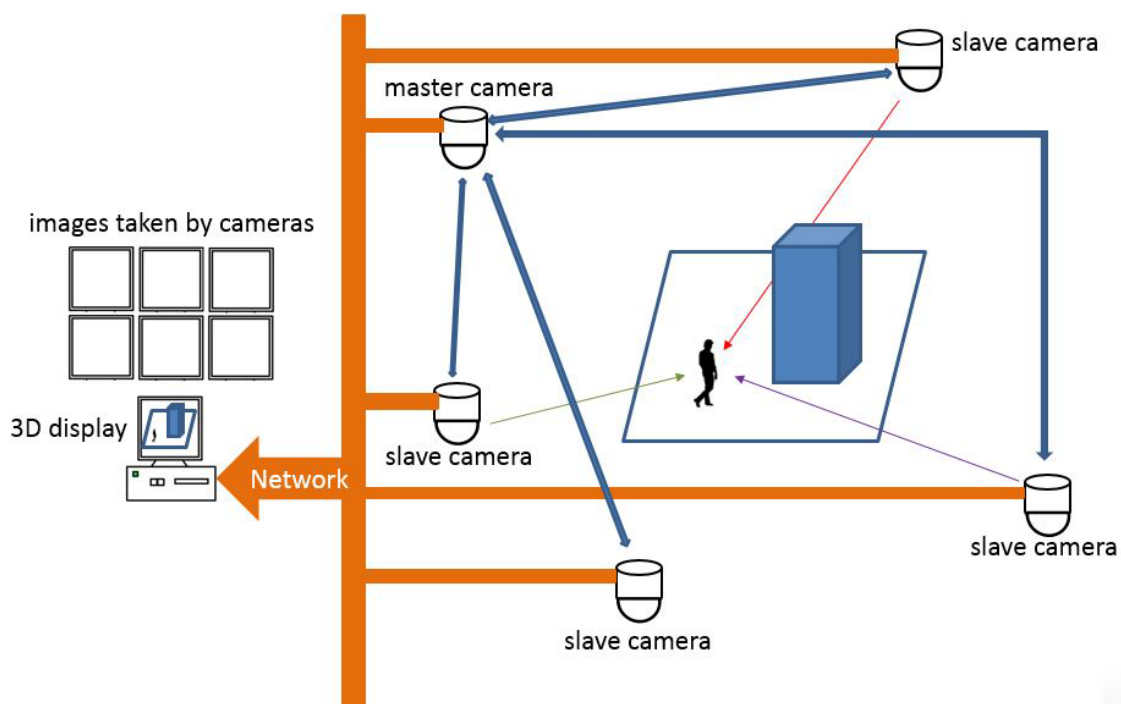


Figure 2.1: Whole image of the proposed system

As shown in the whole image of the proposed system(**Fig.2.1**), some active cameras are connected by a network. The operator can see not only the images taken from each camera, but also the overall 3D site model of the monitoring area in which the stereoscopic images of the pedestrians are shown in real time.

In addition, multiple cameras can do some combination movements. For example, if one camera as a master camera finds a target, the spatial coordinates can be calculated by the master camera, so other cameras can aim at and track the target from different directions automatically. Therefore, the multi-directional images of the target can be obtained. If the image of his/her face is necessary, the camera which can see his/her face can take the image immediately, furthermore, the enlarged image of the target can be acquired continuously. And even if the target leaves the field of view of master camera, another camera will switch to master camera automatically.

Next, we will introduce the function composition of the system.

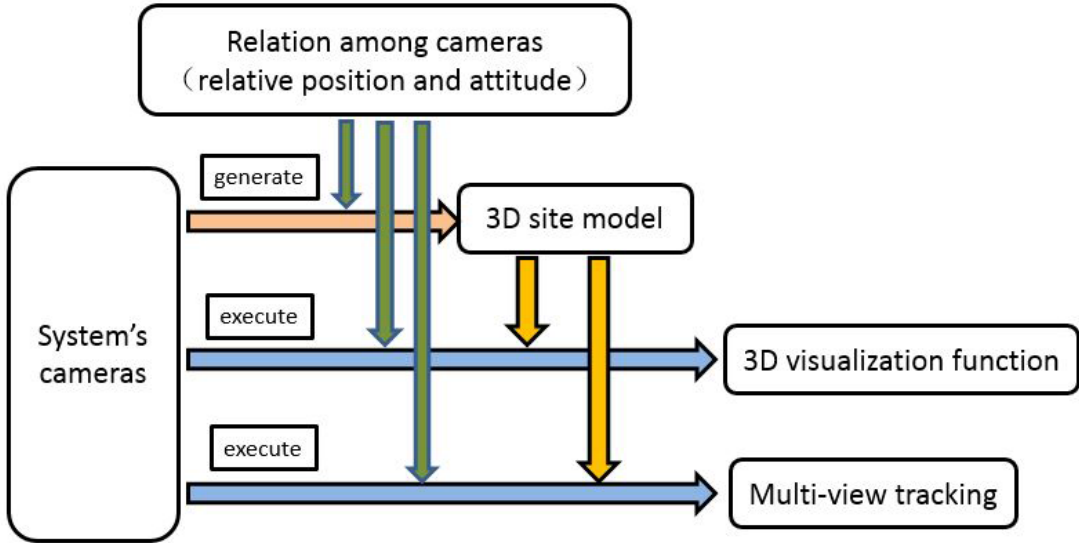


Figure 2.2: The composition of the system

Since all functions are performed by the cameras, the cameras are the core of the system. As shown in **Fig.2.2**, by using the system's cameras, three functions can be achieved. First, the 3D model of this system can be generated by the system's own cameras, unlike other 3D monitoring systems. It means the coordinate system of

the 3D model and the coordinate system of the camera group is the same one, so that it is not necessary to transform the coordinate value between their coordinate systems. Second, pedestrians can be rendered onto 3D site model in real time. So that the position of the pedestrian can be intuitively known by just looking at the dynamic 3D site model. Third, now that the position of the pedestrian is known, it is possible to let multiple cameras track the pedestrian from different directions. Thus we can see that the latter two functions are based on the 3D site model. In addition, in order to realize these three functions, it is necessary to know the relation among the cameras. So, as a basis for the system, the relation among cameras should be calculated firstly. The relation among cameras mentioned here refers to the relative position and the relative attitude, also known as external parameters. In the next section, we will present the knowledge of external parameters.

2.2 External parameters of active PTZ camera

Most of wide-area surveillance cameras can be classified into two types: fixed camera and active PTZ (pan, tilt, zoom) camera. Since the active PTZ camera can be used to make up for blind spots of the fixed camera and to pay more attention to an important target, this type of camera has been widely used.

The PTZ camera can play three movements: pan (P), tilt (T) and zoom (Z). Although it is called an active camera, the position of the camera is fixed and only rotational movement is possible. Generally, roll movement around the optical axis is not necessary, so there is no roll rotation.



Figure 2.3: VN-V686 camera

The spherical camera is the mainstream PTZ camera. In this dissertation, we chose the VN-V686 PTZ network camera released by JVC to build a surveillance system. **Figure 2.3** shows the camera. It 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 $[0^\circ, 360^\circ]$ and its tilt range is $[-5^\circ, 185^\circ]$. After a long period of use (more than one week), the VN-V686 camera will automatically clear the cumulative

rotation error.

This type of PTZ camera has the following two characteristics. First, the optical axis of camera passes through the rotation center which is the intersection point of axes of pan motion and tilt motion. Second, this kind of camera is able to provide rotation angles.

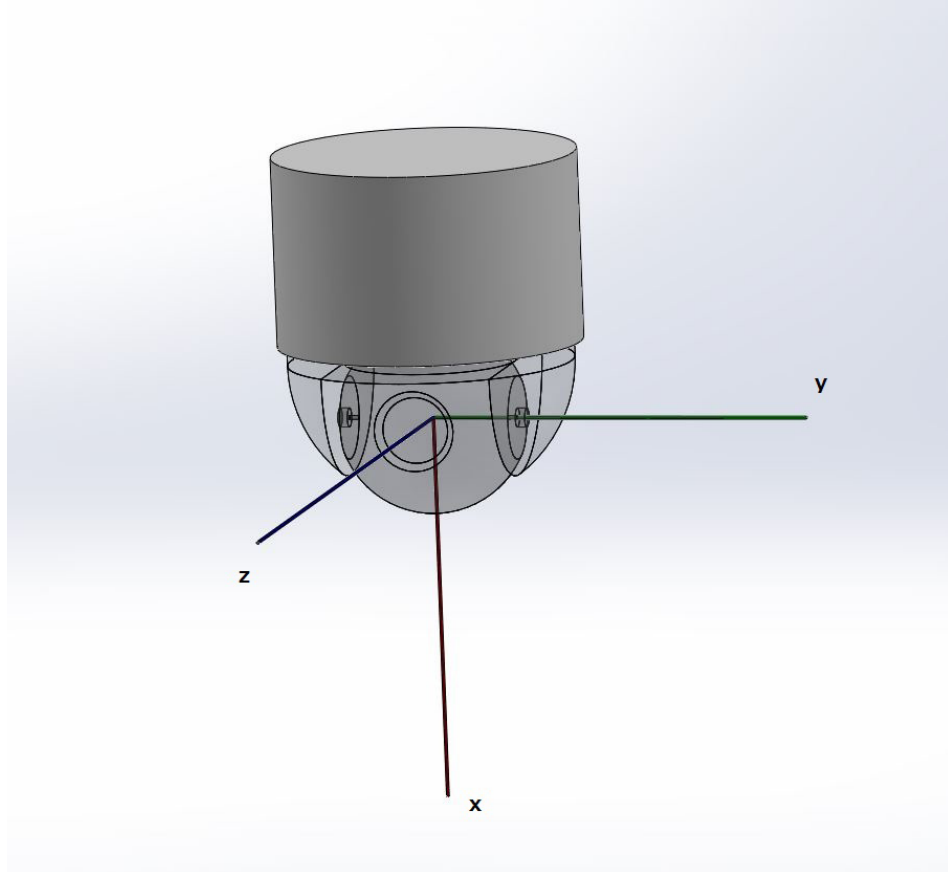


Figure 2.4: Initial coordinate system of a camera

Figure 2.4 shows the initial coordinate system $o - xyz$ of an active PTZ camera. The active PTZ camera performs two motions: pan motion (right and left rotation) and tilt motion (up and down rotation). In **Figure 2.4**, the optical axis of the camera is the z -axis of the coordinate system. The rotation around the x -axis represents the pan motion, and the rotation around the y -axis represents the tilt motion. However, once the camera rotates, the coordinate system will differ from its original position.

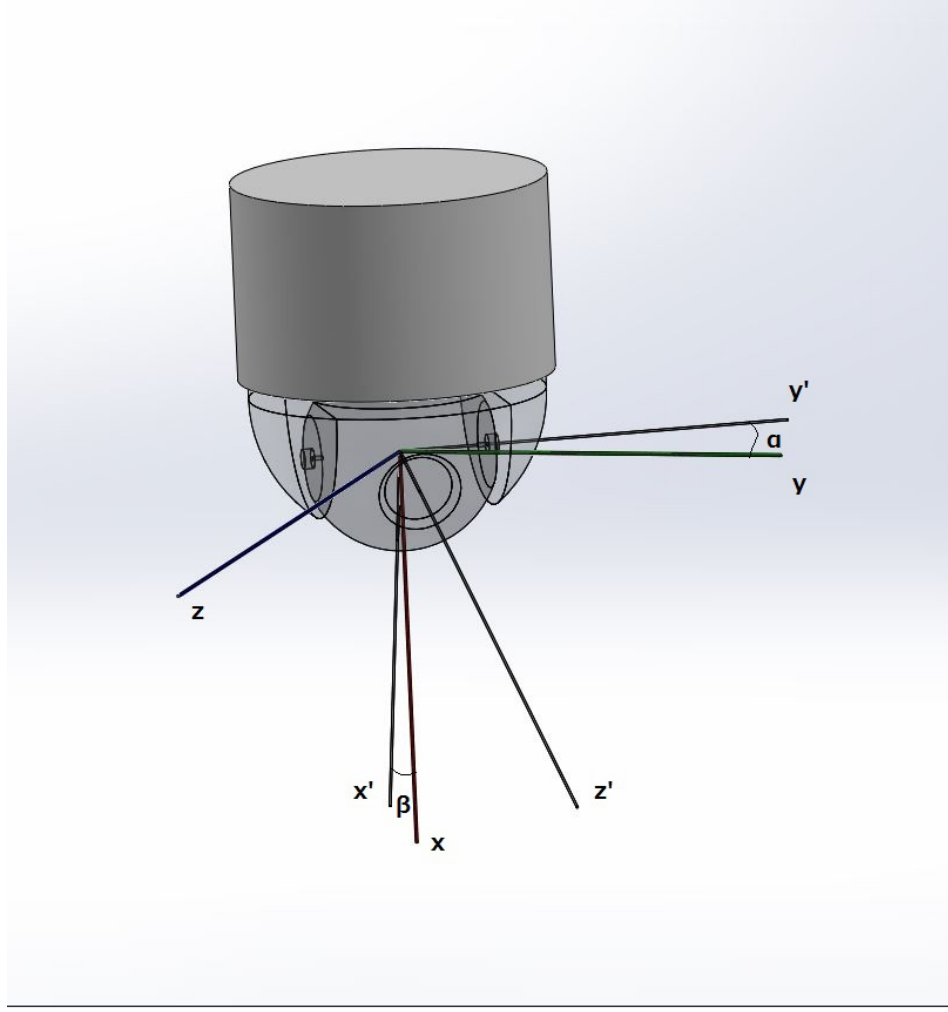


Figure 2.5: Coordinate system after rotation

As shown in **Fig.2.5**, $o - xyz$ and $o - x'y'z'$ are the initial coordinate system and a new one after rotating respectively. The origins of the two coordinate systems are the same. Angle α between y' -axis and y -axis is the rotation angle of pan motion. Angle β between x' -axis and x -axis is the rotation angle of tilt motion. Angle β can also be considered as the angle between z' -axis and yz -plane. In other words, the motion from $o - xyz$ to $o - x'y'z'$ can be considered as two rotations: angle α rotation on the pan direction, and angle β rotation on the tilt direction. In addition, according to the mechanism of the camera, we can know that y' -axis is always on the yz -plane.

Assuming that there is a point P in 3D space, the relation between the initial coordinate system and the new coordinate system can be

represented by the following relationship,

$$\mathbb{P} = {}^0T_t \mathbb{P}', \quad (2.1)$$

where $\mathbb{P} = (X, Y, Z, 1)^T$ is the coordinate value in $o - xyz$,

$\mathbb{P}' = (X', Y', Z', 1)^T$ is the coordinate value in $o - x'y'z'$ of the point P .

Due to the theory of the Euler angle, the relationship between before and after the rotation of the coordinate system can be calculated independently by each rotation step. It means the final result can be calculated by multiplying the result of each rotation step in sequence. Therefore, we calculated the results of pan rotation and tilt rotation respectively.

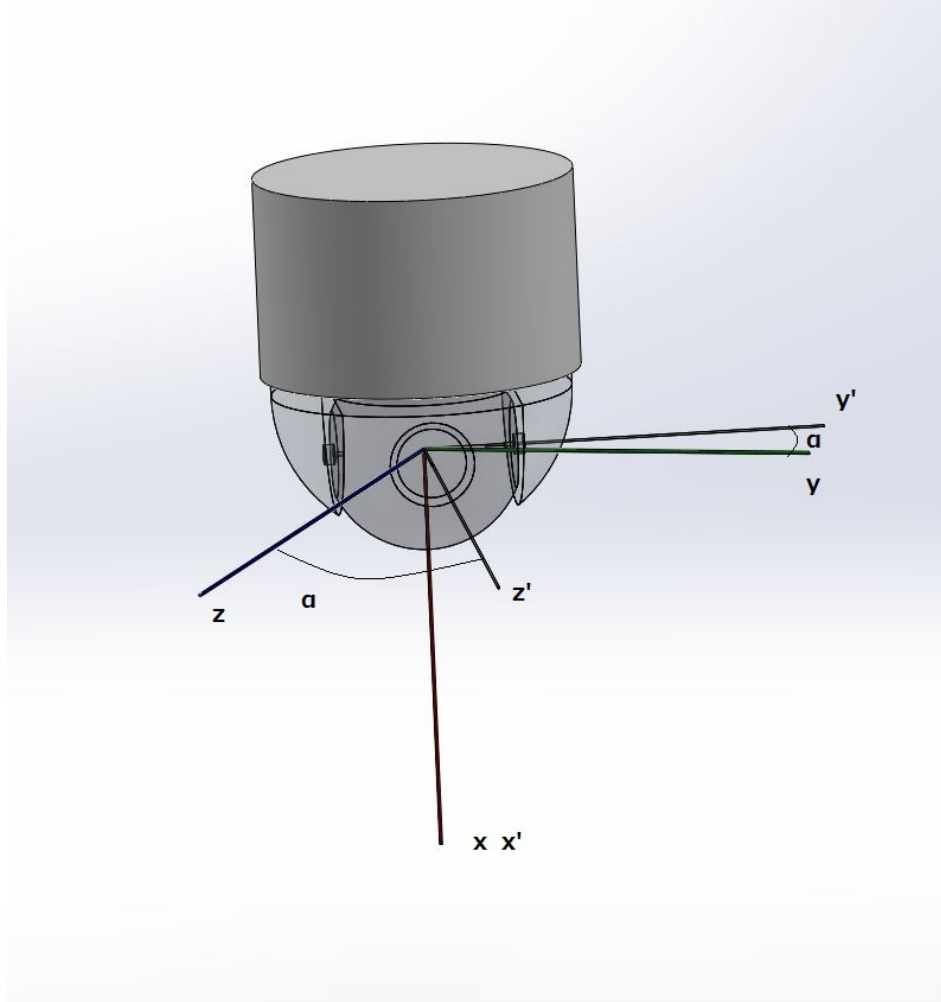


Figure 2.6: Pan rotation

When the coordinate system rotates by angle α in the pan direction as shown in **Fig.2.6**, the relationship can be written as $\mathbb{P} = A\mathbb{P}'$. A is represented by **Eq.2.2**.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.2)$$

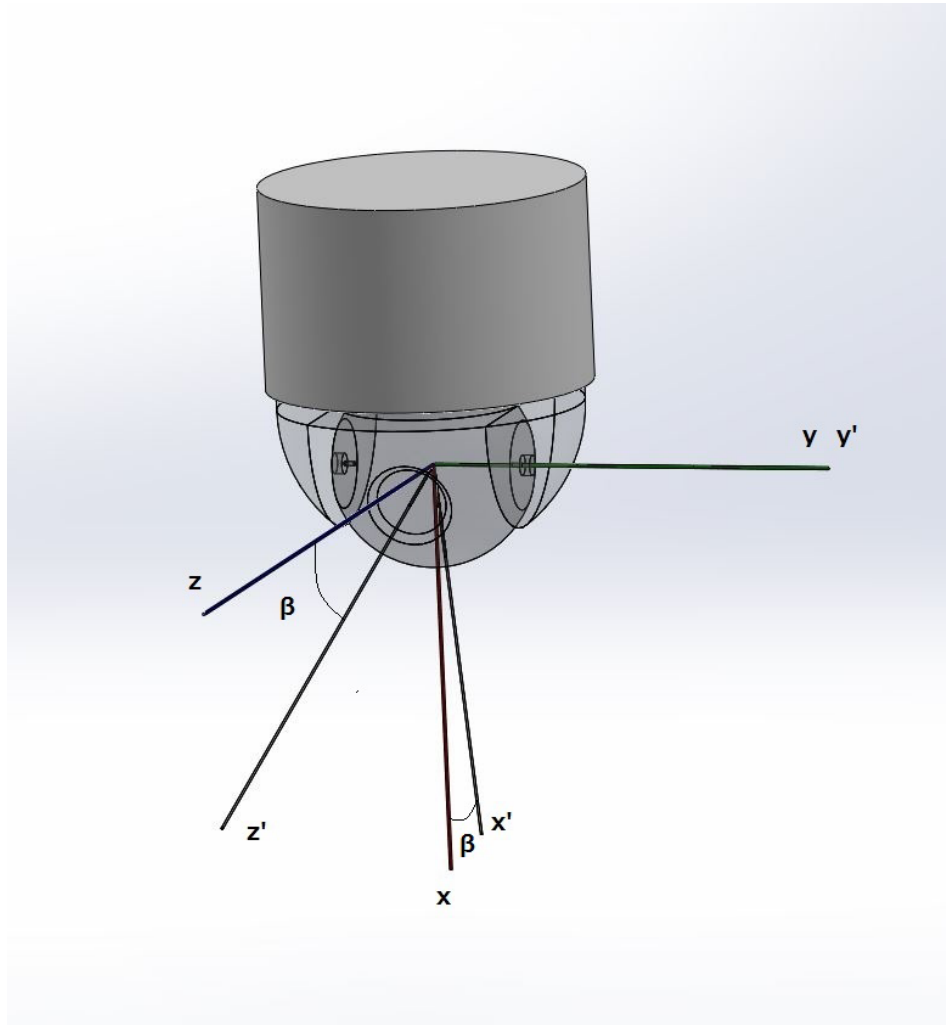


Figure 2.7: Tilt rotation

When the coordinate system rotates by angle β in the tilt direction as shown in **Fig.2.7**, the relationship can be written as $\mathbb{P} = B\mathbb{P}'$. B is represented by **Eq.2.3**.

$$B = \begin{pmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.3)$$

According to the theory of the Euler angle, we can know that 0T_t is multiplied by A and B as shown in **Eq.2.4**

$${}^0T_t = AB = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.4)$$

Further, the transformation matrix 0T_t can be calculated by the following equation.

$${}^0T_t = \begin{pmatrix} \cos \beta & 0 & \sin \beta & 0 \\ -\sin \beta \sin \alpha & \cos \alpha & \cos \beta \sin \alpha & 0 \\ -\sin \beta \cos \alpha & -\sin \alpha & \cos \beta \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.5)$$

Here, we would like to emphasize that rotation angles α and β mentioned in this dissertation are relative to the initial coordinate system.

In particular, when a camera rotates to make its optical axis pass through point P , point P 's coordinate value can be written as $\mathbb{P}' = (0, 0, L, 1)^T$, where L is the distance from point P to the origin of coordinate system. By substituting it for $\mathbb{P}'$ in **Eq.2.1**, the coordinate value of the point P in the initial coordinate system is calculated as follow:

$$\mathbb{P} = {}^0T_t \begin{pmatrix} 0 \\ 0 \\ L \\ 1 \end{pmatrix} = \begin{pmatrix} L \sin \beta \\ L \cos \beta \sin \alpha \\ L \cos \beta \cos \alpha \\ 1 \end{pmatrix}. \quad (2.6)$$

The relation between two cameras' initial coordinate systems (**Fig.2.8**) is similar. 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 coordinate values of point P in the two coordinate systems. $\mathbb{P}_l$ can be transformed to $\mathbb{P}_r$ by a matrix T as shown in **Eq.2.7**.

$$\mathbb{P}_l = T\mathbb{P}_r \quad (2.7)$$

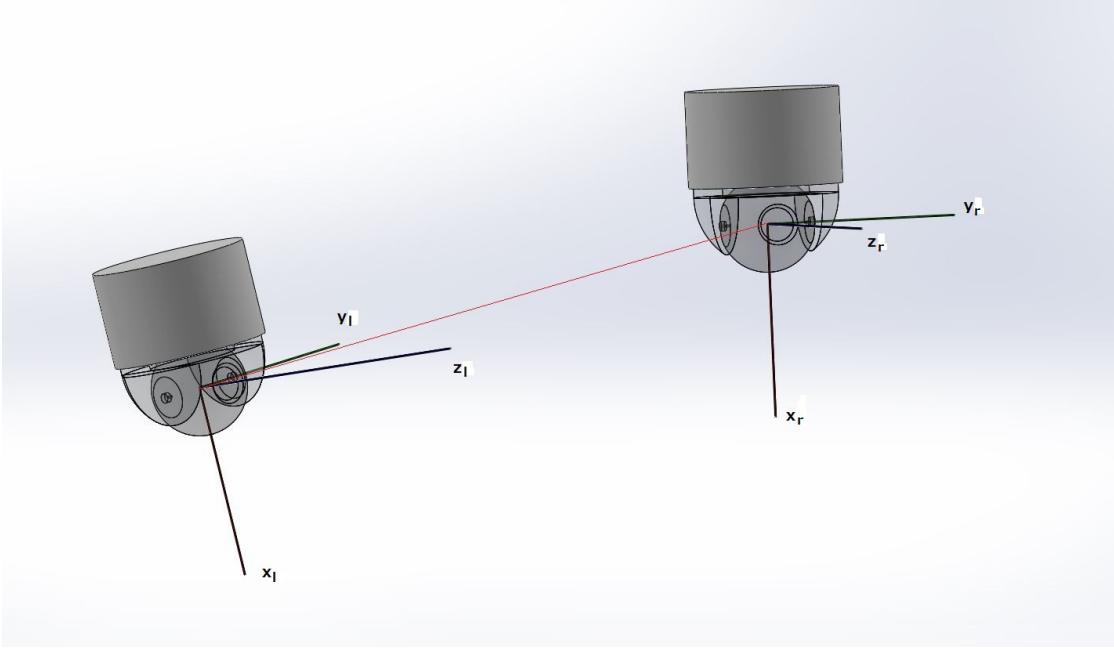


Figure 2.8: Two Coordinate Systems

T is a 4×4 matrix consisting of a rotating part R and a translating part t . R is a 3×3 matrix, and t is a 1×3 vector. It is expressed by **Eq.2.8**.

$$T = \begin{bmatrix} R & t^T \\ 0 & 1 \end{bmatrix} \quad (2.8)$$

The process of solving T can be regarded as the process of external parameter calibration between cameras. In general, the transformation matrix between two coordinate systems can be represented by three Euler angles and three components of a translation vector. Thus, in the minimum case, there are six unknowns in the transformation matrix. However, from algebraic calculation point of view, we calculate twelve unknowns contained in a 3×3 rotation matrix and a 3×1 vector. Hence, the external parameter estimation mentioned in this dissertation is the calculation of the twelve unknowns.

If there are several individual cameras, we can repeat the external parameter estimation for each camera pair.

Chapter 3

External parameter estimation

3.1 Existing external parameter estimation method

The methods for estimating the external parameters of cameras have been extensively reported in the literature. However, because there are some shortcomings, most of state-of-the-art methods are not suitable for the scenario of wide-area surveillance system.

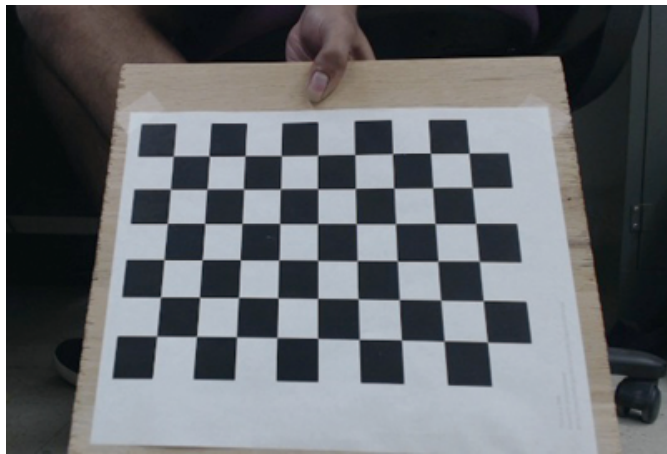


Figure 3.1: A checkerboard

Basically, in wide-area surveillance systems, cameras are commonly installed at outdoor high or far places, therefore objects in the observing area are far away from cameras. In this case, if certain kinds of checkerboards as shown in **Fig.3.1** are used for the camera parameter estimation based on traditional methods[18, 19, 20, 21], even if the checkerboards are enlarged, they still often cannot be seen clearly.



Figure 3.2: DLT method

Another common method is to set a world coordinate system in the observing area and use coordinate values of some target points for calculating the external parameters of cameras. Although points on an actual object can be regarded as target points directly, it may be sometimes difficult to measure their coordinate values. Therefore, some methods use structural frames and set target points on them to get the accurate coordinate values of points.

For example, as shown in **Fig.3.2**, Abdel-Aziz et al. [22] used a large frame which covers the overall space. Privanic et al. [23] used an orthogonally contacted 3-axis frame shown in **Fig.3.3**, and Ki-Young et al. [24] utilized the 3-axis frame and the movement of a wand to do the external parameter estimation. However, if these methods are used for outdoor surveillance cameras, it would become unpractical because the frames should be very huge.

On the other hand, most of the existing image processing based methods use points correspondences between image pairs to obtain the relation between each camera pair [25, 26, 27]. Generally, the accuracy may be strongly influenced by the distortion of lens [28, 29, 30]. And, the external parameters of a camera are affected by the estimation of the internal parameters.

However, once a camera zooms in or out, its internal parameters must change, so it is difficult to estimate the internal parameters with high precision [31, 32, 33, 34, 35]. Therefore, it is also difficult to achieve precise estimation of external parameters using the conventional methods in the case of wide-area surveillance system when

cameras should zoom in and zoom out frequently.

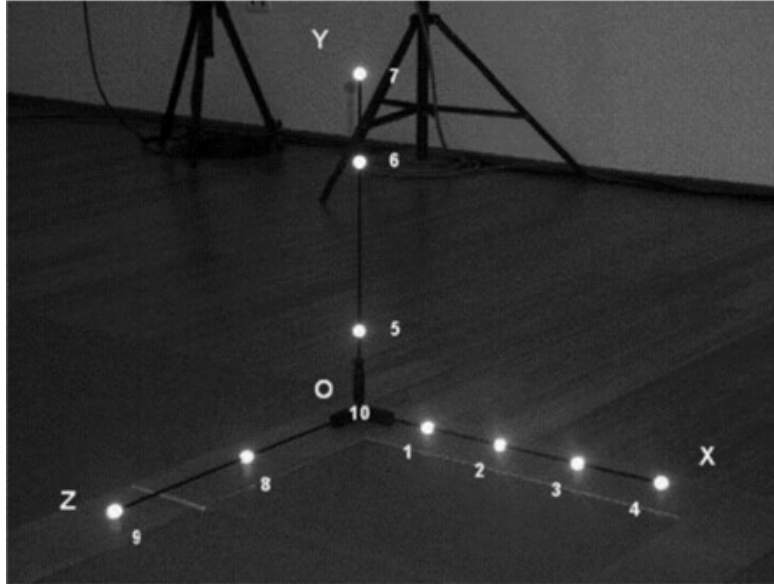


Figure 3.3: 3 axis frame

On account of above reasons, this dissertation proposes a simple external parameter estimation method for multiple PTZ cameras contained in a wide-area surveillance system. With this method, only the optical axis information is used, so that the accuracy can be not affected by the lens distortion. Moreover, since it is not necessary to calibrate the internal parameters, cameras can be operated to zoom in/out in accordance with a user's need during the estimation process.

3.2 Proposed external parameter estimation method

3.2.1 The gauge used for distance estimation



Figure 3.4: L-shaped gauge

As mentioned above, it is difficult to estimate the internal parameters of PTZ camera with high precision. Hence, we consider to estimate distances from target points to the camera directly, and obtain the coordinate values of target points to calculate the external parameters. It is challenging to measure the distance by only one camera without using the internal parameters. We propose a method [36] to calculate the distance by using an auxiliary instrument of L-shaped gauge as shown in **Fig. 3.4**. This method can utilize the movability of the PTZ camera and calculate the coordinate values of target points without internal parameters. Moreover, comparing with the existing methods, this gauge is simple and its placement is arbitrary.

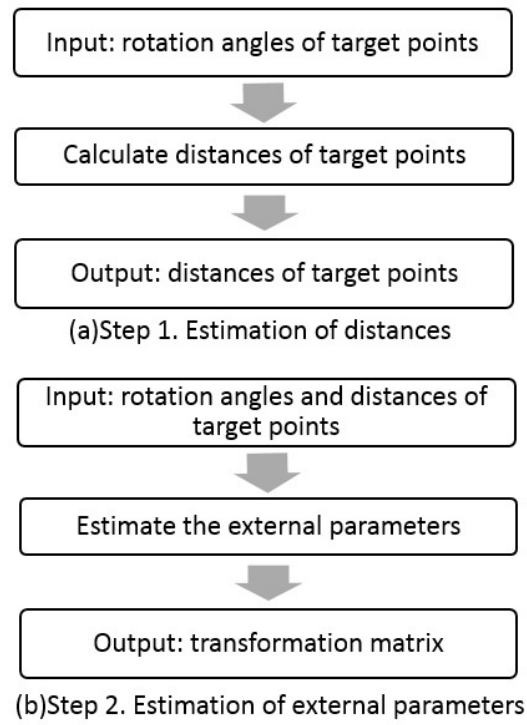


Figure 3.5: Flowchart of process

Figure 3.5 shows two main steps in our method, and these steps are described in following sections.

3.2.2 Estimation of distances

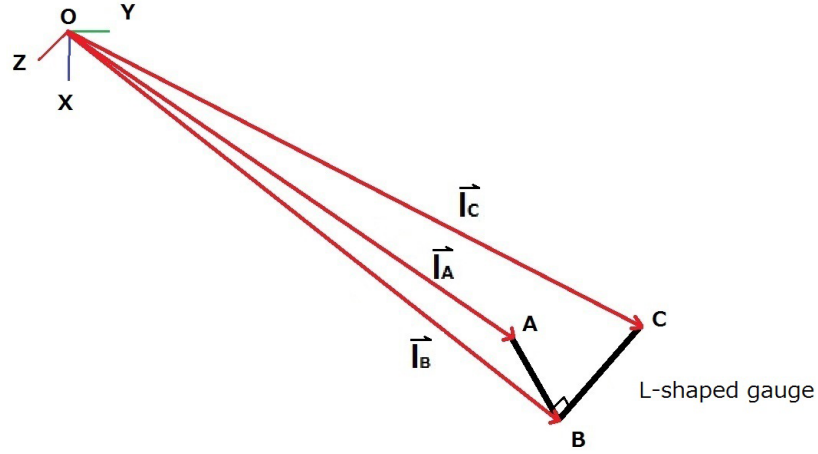


Figure 3.6: Coordinate system and L-shaped gauge

In **Fig.3.6**, points A, B, and C represent a L-shaped gauge used in our proposed method. Line BA and BC have a same length and are perpendicular. Similar to **Fig.2.4**, $o - xyz$ in **Fig.3.6** represents an initial coordinate system of a PTZ camera.

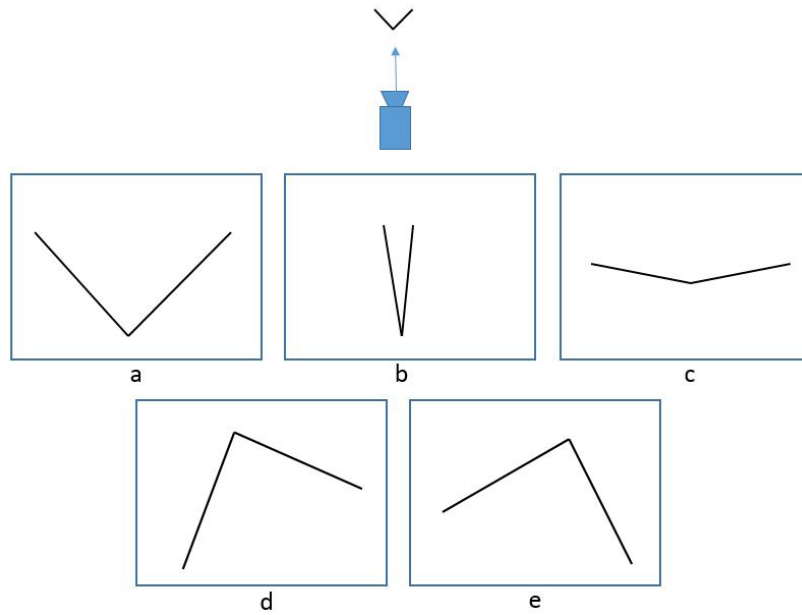


Figure 3.7: Configuration of L-shaped gauge

First, we install the L-shaped gauge in a common field of the views of all cameras. The configuration of L-shaped gauge is arbitrary as long as the three endpoints A, B and C can be seen very well in each camera.

Although it is not necessary to stipulate the posture of L-shaped gauge quantitatively, it is better to avoid extreme conditions. For example, the L-shaped gauge should be adjusted to make the image like **Fig.3.7** (a), not like **Fig.3.7** (b) (c). The rotations like **Fig.3.7** (d) (e) do not matter.

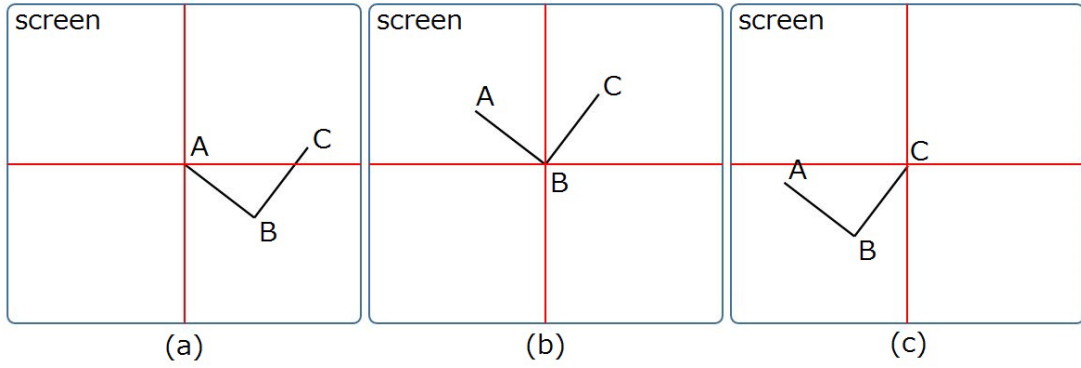


Figure 3.8: Step of angle measure

After it, each camera is operated to match its optical axis to the endpoints A, B and C one by one manually. In the figure, optical axis is assumed to pass through the cross section of horizontal and vertical center lines. In other words, the endpoints A, B and C are transferred to the center of the image by rotating each camera as shown in **Fig.3.8**.

$\vec{l}_A$, $\vec{l}_B$ and $\vec{l}_C$ in **Fig.3.6** represent the optical axis after rotating. When the optical axis passes through the endpoint, the rotation angle (values of pan rotation and tilt rotation) is read from the camera encoders. This operation is called *angle measure*. After this step, the values of pan ($\alpha_A, \alpha_B, \alpha_C$) and tilt ($\beta_A, \beta_B, \beta_C$) for three endpoints A, B, and C can be obtained.

In the camera initial coordinate system, vectors $\vec{l}_A$, $\vec{l}_B$ and $\vec{l}_C$ can be expressed as follows.

$$\vec{l}_A = (L_A \sin \beta_A, L_A \cos \beta_A \sin \alpha_A, L_A \cos \beta_A \cos \alpha_A), \quad (3.1)$$

$$\vec{l}_B = (L_B \sin \beta_B, L_B \cos \beta_B \sin \alpha_B, L_B \cos \beta_B \cos \alpha_B), \quad (3.2)$$

$$\vec{l}_C = (L_C \sin \beta_C, L_C \cos \beta_C \sin \alpha_C, L_C \cos \beta_C \cos \alpha_C), \quad (3.3)$$

where L_A , L_B and L_C in the above equations are the lengths of the vectors $\vec{l}_A$, $\vec{l}_B$ and $\vec{l}_C$.

According to the geometric relations shown in **Fig.3.6**, it is possible to derive the following equations.

$$f_1 : |\vec{BA}|^2 = |\vec{l}_A - \vec{l}_B|^2 = L^2 \quad (3.4)$$

$$f_2 : |\vec{BC}|^2 = |\vec{l}_C - \vec{l}_B|^2 = L^2 \quad (3.5)$$

$$f_3 : |\vec{AC}|^2 = |\vec{l}_C - \vec{l}_A|^2 = 2L^2 \quad (3.6)$$

$$f_4 : \vec{BA} \cdot \vec{BC} = (\vec{l}_A - \vec{l}_B) \cdot (\vec{l}_C - \vec{l}_B) = 0 \quad (3.7)$$

where L is the length of $\vec{BA}$ and $\vec{BC}$. The unknowns of these equation system are the distances L_A , L_B and L_C .

In this dissertation, the Levenberg-Marquardt [37] method is used to find the solution (L_A, L_B, L_C) of this non-linear problem. If more than one group of solutions are given, the one which makes the total error $(f_1 \sim f_4)$ smallest should be chosen as the solution of this equation system.

As indicated in **Fig.3.5(a)**, at the end of Step 1, the distances from each endpoint of the L-shaped gauge to the origin of the coordinate system are obtained.

3.2.3 Estimation of external parameter

According to the method described previously, if the L-shaped gauge is placed in the common field of views of cameras, it is possible to measure the rotation angles of the three endpoints A, B and C, and then the distances from the three endpoints to the camera can be calculated by using those rotation angles. So, if the gauge is installed

at m different locations, we obtain $n(n = m \times 3)$ points' rotation angles and distances for all cameras.

Table 3.1: Data of target points

NO.	camera- l			camera- r		
	pan	tilt	distance	pan	tilt	distance
1	${}^l\alpha_1$	${}^l\beta_1$	lL_1	${}^r\alpha_1$	${}^r\beta_1$	rL_1
2	${}^l\alpha_2$	${}^l\beta_2$	lL_2	${}^r\alpha_2$	${}^r\beta_2$	rL_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	${}^l\alpha_n$	${}^l\beta_n$	lL_n	${}^r\alpha_n$	${}^r\beta_n$	rL_n

The external parameters are estimated for every pair of cameras. Here, we simply consider one camera pair. **Table 3.1** is assumed to give the data of target points for one camera pair. Two cameras in each pair are called camera l and camera r respectively.

For example, the coordinates of target point $P_n(x_n, y_n, z_n)$ in two cameras' coordinate systems ($l-xyz$ and $r-xyz$) can be determined by the following equations because pan and tilt angles of left and right cameras are given from their encoders, and distances lL_n and rL_n from the both cameras to the point P_n are calculated by the non-linear problem solver.

$$\begin{bmatrix} {}^lx_n \\ {}^ly_n \\ {}^lz_n \\ 1 \end{bmatrix} = \begin{bmatrix} {}^lL_n \sin {}^l\beta_n \\ {}^lL_n \cos {}^l\beta_n \sin {}^l\alpha_n \\ {}^lL_n \cos {}^l\beta_n \cos {}^l\alpha_n \\ 1 \end{bmatrix} \quad (3.8)$$

$$\begin{bmatrix} {}^rx_n \\ {}^ry_n \\ {}^rz_n \\ 1 \end{bmatrix} = \begin{bmatrix} {}^rL_n \sin {}^r\beta_n \\ {}^rL_n \cos {}^r\beta_n \sin {}^r\alpha_n \\ {}^rL_n \cos {}^r\beta_n \cos {}^r\alpha_n \\ 1 \end{bmatrix} \quad (3.9)$$

By **Eq.2.7**, the following equation can be given.

$$\begin{bmatrix} {}^lx_n \\ {}^ly_n \\ {}^lz_n \\ 1 \end{bmatrix} = \begin{pmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} {}^rx_n \\ {}^ry_n \\ {}^rz_n \\ 1 \end{bmatrix} \quad (3.10)$$

After expanding **Eq.3.10**, three linear equations for each target point can be given. Since the transformation matrix T contains 12 unknowns, at least four target points are necessary. Generally, the more target points are used for calculation, the higher accuracy is achieved in the sense of least squares method. By using coordinate values of $n(n \geq 4)$ target points, we can get the following equations.

$$\left\{ \begin{array}{l} {}^l x_1 = {}^r x_1 T_{11} + {}^r y_1 T_{12} + {}^r z_1 T_{13} + T_{14} \\ {}^l y_1 = {}^r x_1 T_{21} + {}^r y_1 T_{22} + {}^r z_1 T_{23} + T_{24} \\ {}^l z_1 = {}^r x_1 T_{31} + {}^r y_1 T_{32} + {}^r z_1 T_{33} + T_{34} \\ \dots\dots\dots \\ {}^l x_n = {}^r x_n T_{11} + {}^r y_n T_{12} + {}^r z_n T_{13} + T_{14} \\ {}^l y_n = {}^r x_n T_{21} + {}^r y_n T_{22} + {}^r z_n T_{23} + T_{24} \\ {}^l z_n = {}^r x_n T_{31} + {}^r y_n T_{32} + {}^r z_n T_{33} + T_{34} \end{array} \right. \quad (3.11)$$

In this dissertation, the SVD (singular value decomposition) [38] was used to solve this equation system. So far, the transformation matrix between the initial coordinate systems of two cameras can be known.

3.3 Evaluation experiments

3.3.1 Synthetic experiment

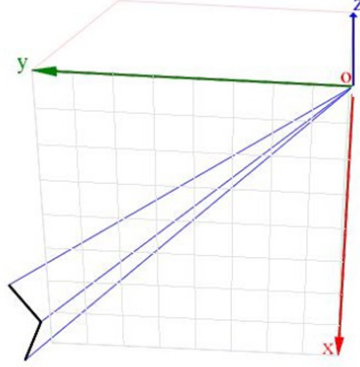


Figure 3.9: Simulation environment

As shown in **Fig. 3.5**, the estimation of the external parameters is based on the distances, so that in the following experiments, the accuracy of this distance estimation is investigated.

Figure 3.9 shows the L-shaped gauge and the coordinate system of a camera in our simulation environment. The size of the L-shaped gauge is set to $1\text{ m} \times 1\text{ m}$. It is placed at various positions. At each position, the rotation angles of the endpoints and their true distances from the origin can be obtained by using the coordinate values. The rotation angles were substituted to Eqs. (8),(9),(10) and (11), to calculate the distances. The results were compared with the true ones.

The rotation error of VN-V686 camera is 0.03° , so we supposed that rotation errors of each camera used in actual experiment were within $\pm 0.03^\circ$, and their resolving power were set to 0.01° .

If a target point is very far, the distance error due to angle misalignment is much bigger than the error caused by human operation in pixel unit. Therefore, in this simulation experiment, the effect of human operation is neglected, and the error is set in the range of $[-0.03^\circ, +0.03^\circ]$.

Five ranges of distances were set, which were $[10\text{m}, 20\text{m}]$, $[20\text{m}, 30\text{m}]$, $[30\text{m}, 40\text{m}]$, $[40\text{m}, 50\text{m}]$ and $[50\text{m}, 60\text{m}]$. In each range, the

distance estimation experiment was repeated 1,000 times for different positions.

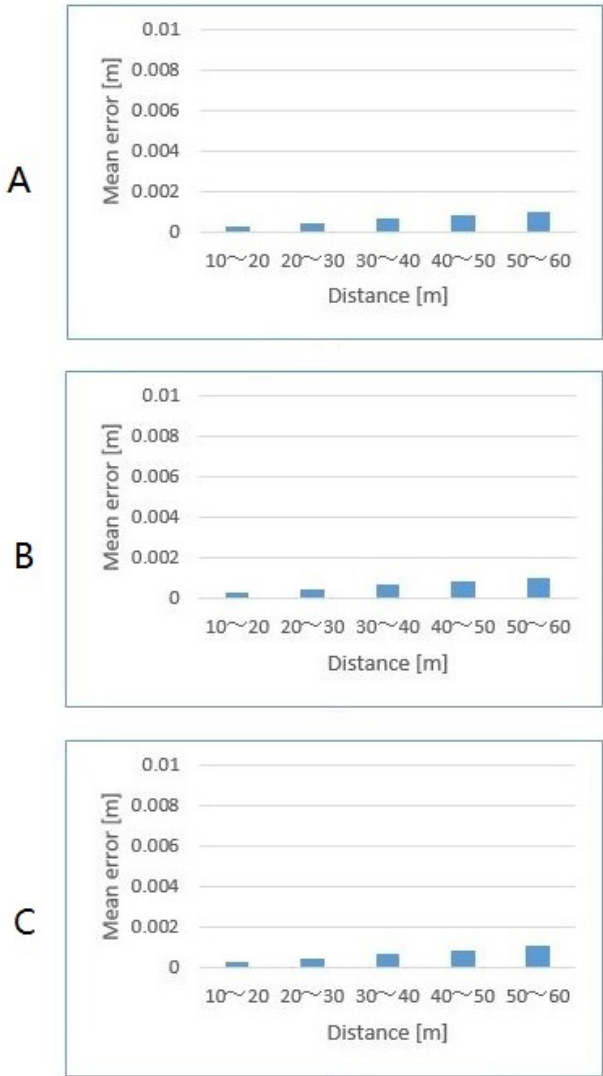


Figure 3.10: Mean error for distance estimation experiment conducted without disturbance

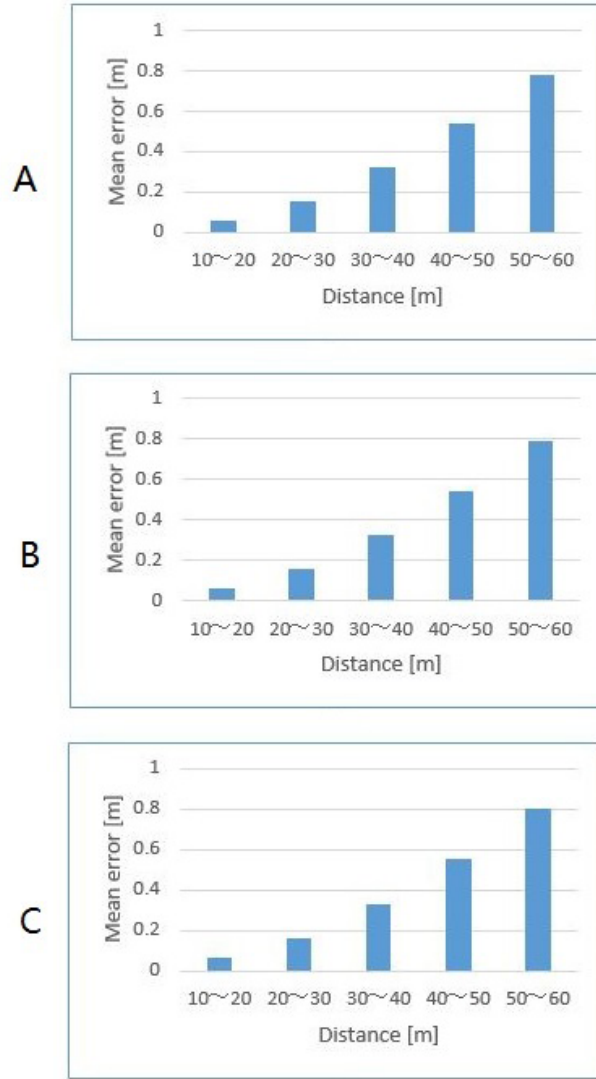


Figure 3.11: Mean error for distance estimation experiment conducted with disturbance

In order to clarify the effect of the rotation angle error, the experiment using rotation angles without disturbance was carried out. **Figure 3.10** shows absolute mean errors for endpoints A, B and C respectively.

The experiment using rotation angles added with random disturbance varying from -0.03° to $+0.03^\circ$ was also carried out. The result is shown in **Fig. 3.11**. As can be seen from **Fig. 3.10** and **Fig. 3.11**, the error of the former is much smaller than that of the latter. According to **Fig. 3.10**, we can conclude that it is possible

to estimate the distance of endpoint by this method when the rotation angles are provided properly. However, once the disturbance is added, the error becomes large. Of course, there is also a tendency that the farther the distance is, the bigger the error is.

Table 3.2: Statistical results

Range of distance	A			B			C		
	Mean error	Standard deviation	Max value	Mean error	Standard deviation	Max value	Mean error	Standard deviation	Max value
10 ~ 20	0.0564	0.0514	0.3039	0.0573	0.0521	0.3079	0.0623	0.0560	0.3241
20 ~ 30	0.1547	0.1183	0.6261	0.1567	0.1199	0.6312	0.1639	0.1253	0.6531
30 ~ 40	0.3180	0.2411	1.4165	0.3203	0.2445	1.4326	0.3301	0.2529	1.4677
40 ~ 50	0.5405	0.3835	2.1363	0.5427	0.3876	2.1526	0.5543	0.3961	2.1894
50 ~ 60	0.7810	0.5603	3.0339	0.7848	0.5631	3.0364	0.7998	0.5748	3.1016

Table 3.3.1 shows the mean error, the standard deviation and the max error for endpoints A, B and C. As you can see, when the distance is under 30m, the mean error is rather low. As the distance increases to range of 30m~50m, the mean error becomes about 1%. When the distance is from 50m to 60m, the mean error is almost 0.8 meter. According to this result, we can see that it is better not to place the L-shaped gauge too far away from the cameras when the estimation is operated. However, in many wide-area surveillance system, since cameras are commonly installed in high places, it is usually difficult to put the gauge close to cameras. In this dissertation, we tried to put the L-shaped gauge within 50m when the real experiment was done.

3.3.2 Real data experiment

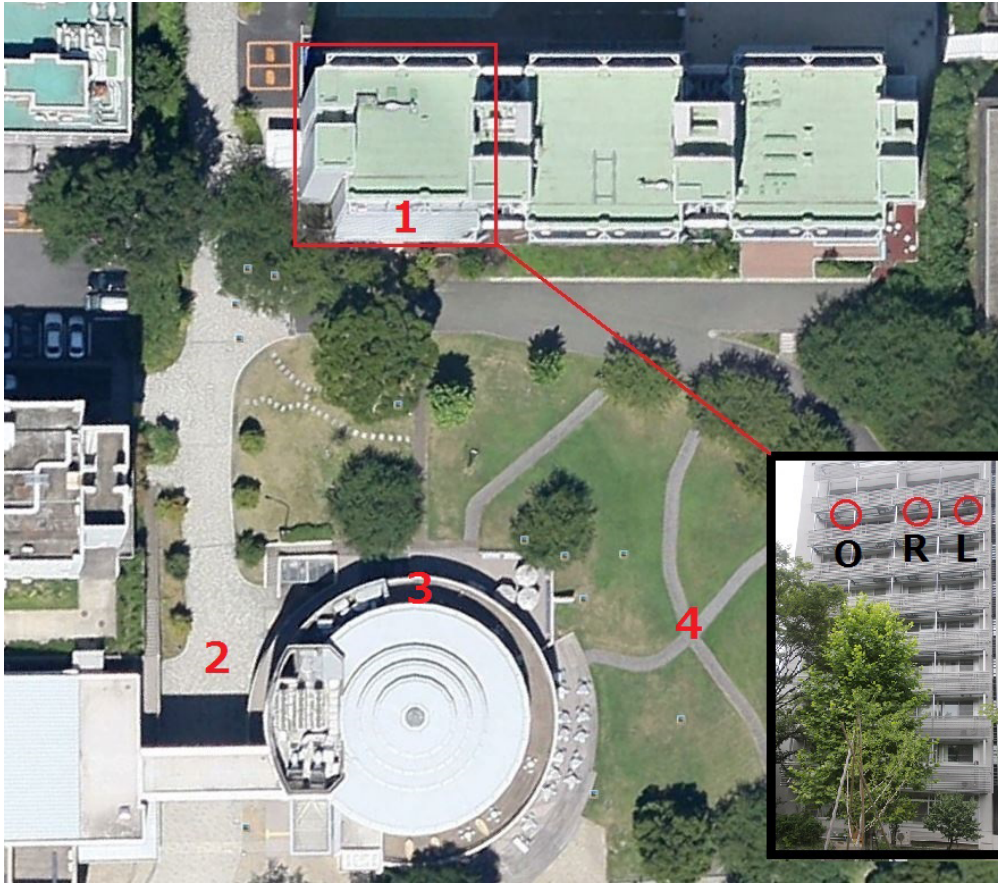


Figure 3.12: Environment of actual experiment

We also did actual experiments to inspect the proposed external parameter estimation method. For example, an experiment was carried out in the environment shown in **Figure 3.12**. Three cameras (red circles in **Figure 3.12**) are installed on the eighth floor of the building. Most of the objects in the field of camera view are far from the cameras more than 40 meters. The letters L, R and O in **Fig. 3.12** are code names of three cameras.

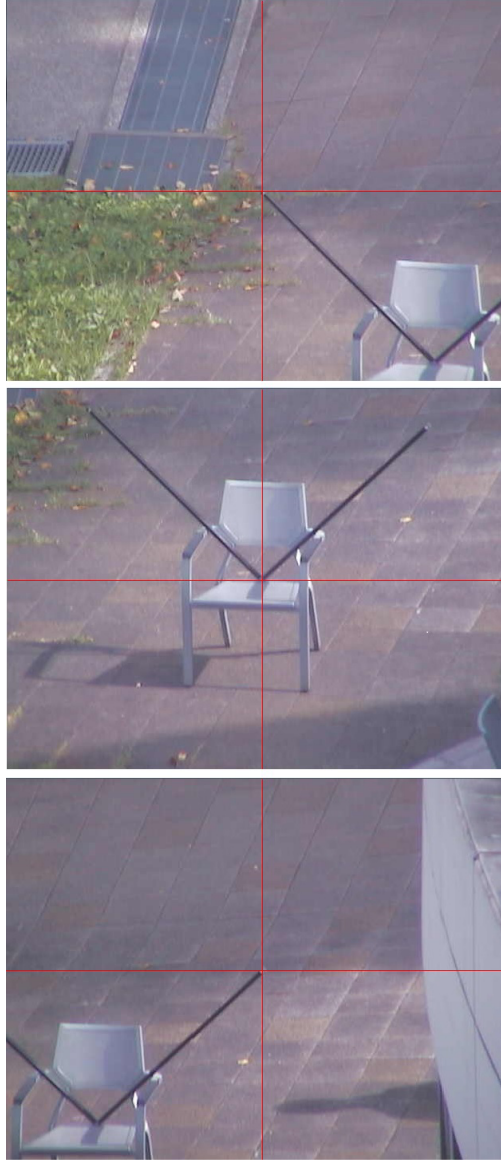


Figure 3.13: L-shaped gauge taken by a camera

As mentioned in section 3.2, at least four target points are necessary to calculate the transformation matrix T . Although, in theory, it is believed that more points are used, the higher accuracy can be achieved, we empirically found that if there are no great differences among the points in terms of distance and orientation, four quiet different typical positions for putting the L-shaped gauge is enough, so that we had $12(4 \times 3)$ points' data to do the estimation.

The numbers in red color in **Fig. 3.12** give the four locations chosen in this experiment. For example, 1: the balcony of eighth floor; 2: the road; 3: the third floor of the rotunda; 4: the lawn.

Figure 3.13 shows an example that the optical axis of a camera is matched to three endpoints of L-shaped gauge located at the road.

According to the method described in section 3.2, we estimated three transformation matrices for three camera pairs(L and R, L and O, R and O). When the two cameras aim at a same point, the distance from this point to each camera can be calculated by using the rotation angles and the transformation matrix. In other word, the distance to any point in the common field of views can be measured.

In this dissertation, we used the camera L and R as the main camera pair to do the distance measurement experiment. Similar to **Eqs.3.8**, **3.8** and **3.10**, when two cameras rotate to aim at a target point, **Eq.3.12** can be given.

$$\begin{pmatrix} D_L \sin \beta_L \\ D_L \cos \beta_L \sin \alpha_L \\ D_L \cos \beta_L \cos \alpha_L \\ 1 \end{pmatrix} = T_{R \rightarrow L} \begin{pmatrix} D_R \sin \beta_R \\ D_R \cos \beta_R \sin \alpha_R \\ D_R \cos \beta_R \cos \alpha_R \\ 1 \end{pmatrix}, \quad (3.12)$$

where $\alpha_L, \beta_L, \alpha_R, \beta_R$ represent their rotation angles, and D_L and D_R represent the distances from the target point to the Cameras L and R. The $T_{R \rightarrow L}$ is the transformation matrix between Camera L and Camera R. Then, we can use the SVD method to solve this equation to obtain the unknowns D_L and D_R .

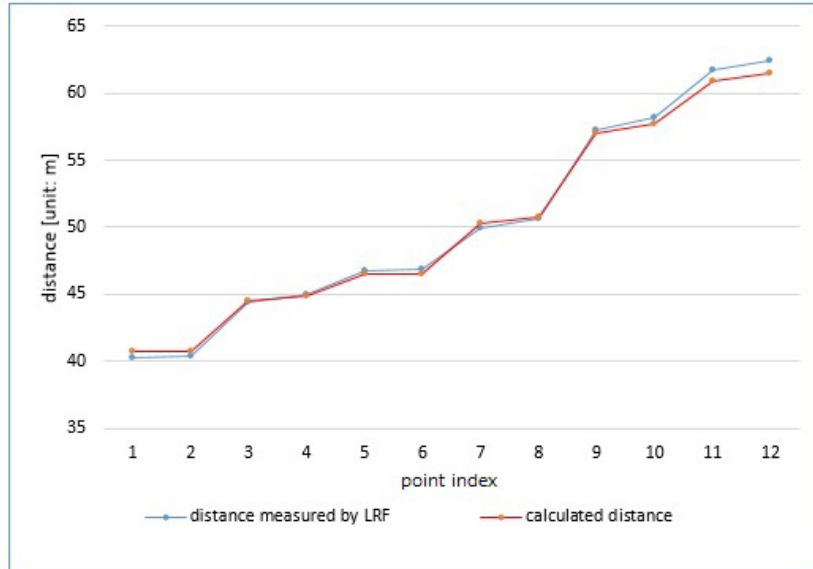


Figure 3.14: Result of distance measurement

In order to evaluate the accuracy, the obtained distances of multiple points were compared with the ground truth values provided by a laser rangefinder (LRF) as shown in **Fig.3.14**. From the results, we can know that the relative displacement can be hold down to 1% if the distance is within 60 meters.

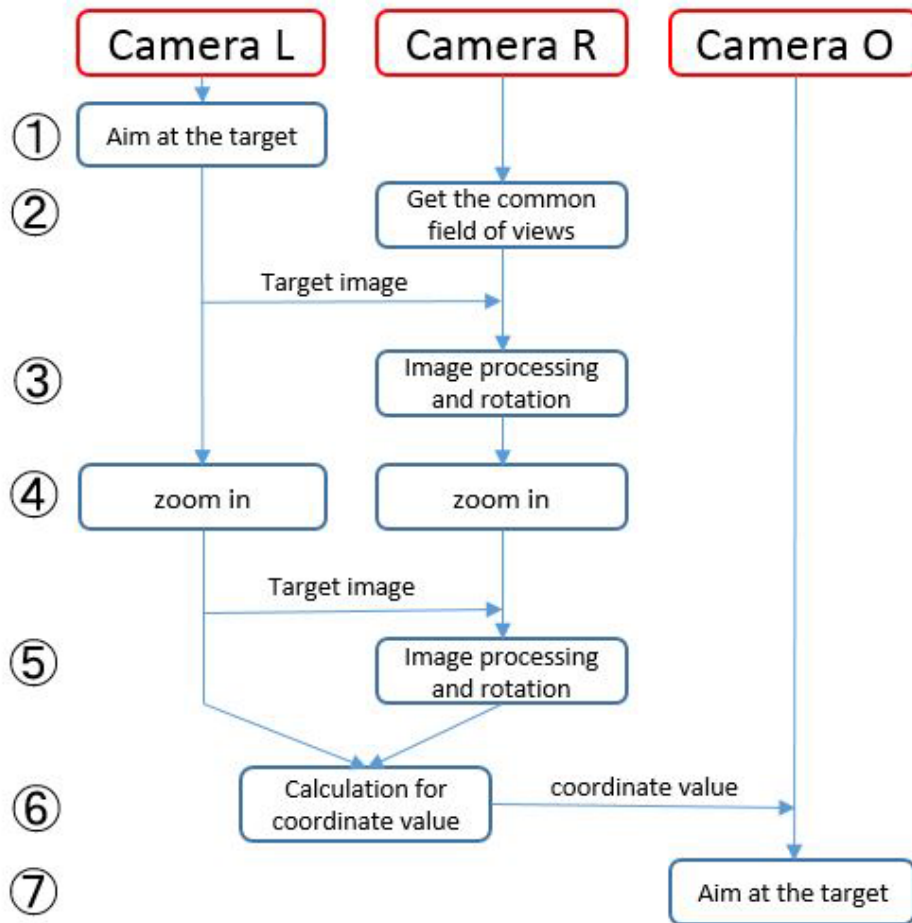


Figure 3.15: Flowchart of multi-directional shooting

When the distance of a point is obtained by two cameras, the coordinate of this point in each camera coordinate system should be calculated precisely as well. Therefore, in order to further evaluate the obtained transformation matrices, we implemented the function of multi-directional shooting based on the distance measurement. If the coordinate value can be calculated precisely, other cameras can be controlled to aim at the same point from different directions automatically. **Figure.3.15** shows the process of multi-directional

shooting experiment.

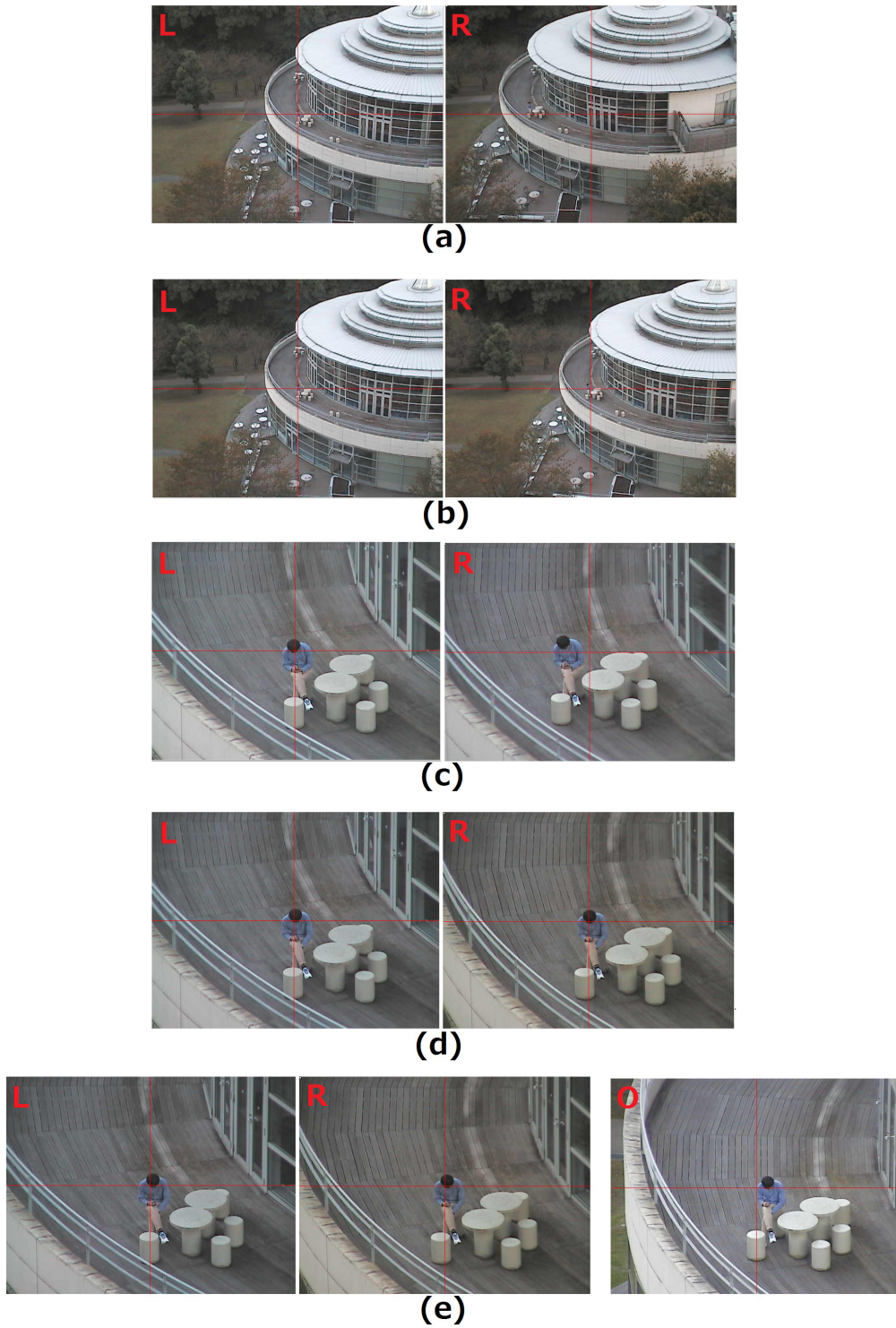


Figure 3.16: Multi-directional shooting experiment

An example of multi-directional shooting experiment is shown in

Figure 3.16. The details of the experiment are provided below.

1. In this experiment, it is assumed that a target which needs to be observed from different directions appears in the image of Camera L at first. We click the target to let the system know which one should be watched. After this step, the entire process is done automatically. First of all, Camera L rotates to make its optical axis aim at the target automatically. As shown in **Fig. 3.16(a)-L**, Camera L aims at a person sitting on the balcony who is chosen as the target.
2. Camera L and Camera R are set to wide-angle state. And, the Camera R rotates to make its optical axis parallel to the optical axis of Camera L to make sure that there are large common parts of two images. **Fig. 3.16(a)-R** shows the state of Camera R after this step.
3. The position of the target in the image of Camera R is found, and then Camera R rotates to move it to the center of the image as shown in **Fig. 3.16(b)**. Here, we take a 60×60 pixel block from the center of the image of Camera L, and search for a similar part in the image of Camera R by using a simple template matching method.
4. Because it is assumed that the zoom motion does not affect the position of optical axis in this dissertation, in order to improve the accuracy, two cameras are controlled to zoom in to make the target larger in two images. **Fig. 3.16(c)** shows the images after zooming. The horizontal angles of view are about 5° .
5. Repeat the step 3 until two cameras more precisely aim at the same place as shown in **Fig. 3.16(d)**. In this state, according to Eq.(16), we calculate the coordinate values in the initial coordinate systems of two cameras. They can be written as $(x_L, y_L, z_L, 1)^T$ and $(x_R, y_R, z_R, 1)^T$.
6. Calculate the target's spatial coordinate value in the coordinate system of Camera O by **Eq.3.13**.

$$\begin{pmatrix} x_O \\ y_O \\ z_O \\ 1 \end{pmatrix} = T_{L \rightarrow O} \begin{pmatrix} x_L \\ y_L \\ z_L \\ 1 \end{pmatrix} \text{ or } \begin{pmatrix} x_O \\ y_O \\ z_O \\ 1 \end{pmatrix} = T_{R \rightarrow O} \begin{pmatrix} x_R \\ y_R \\ z_R \\ 1 \end{pmatrix} \quad (3.13)$$

7. Camera O rotates to aim at the target. Rotation angles are calculated by **Eq.3.14**.

$$\alpha = \arctan\left(\frac{y_O}{z_O}\right), \quad \beta = \arctan\left(\frac{x_O}{\sqrt{y_O^2 + z_O^2}}\right) \quad (3.14)$$

In this experiment, zooming takes about 1.5s. The image template matching takes about 0.35s. The rotation speed of the cameras is about 200°/s. The time spent by other calculation is less than 1ms.

Figure 3.16 shows the result of multi-directional shooting experiment: first Camera L aims at a certain point, then Camera R and Camera O can aim at the same point automatically. In **Fig. 3.16(e)**, it can be seen that there is a small deviation from the optical axis of Camera O to the target point. When the Camera O aims at the point exactly, the pan and tilt values at this position are regarded as the true values. For the true position and the result position, the vectors of optical axis can be expressed with $\vec{v} = (\sin \beta, \cos \beta \sin \alpha, \cos \beta \cos \alpha)$ and $\vec{v}' = (\sin \beta', \cos \beta' \sin \alpha', \cos \beta' \cos \alpha')$, where α and β are the true values of pan and tilt angle, α' and β' are the result values. According to the formula for scalar product of two vectors, the error of optical axis in angle can be derived as **Eq.3.15**.

$$\arccos\left(\frac{\vec{v} \cdot \vec{v}'}{|\vec{v}| |\vec{v}'|}\right) \quad (3.15)$$

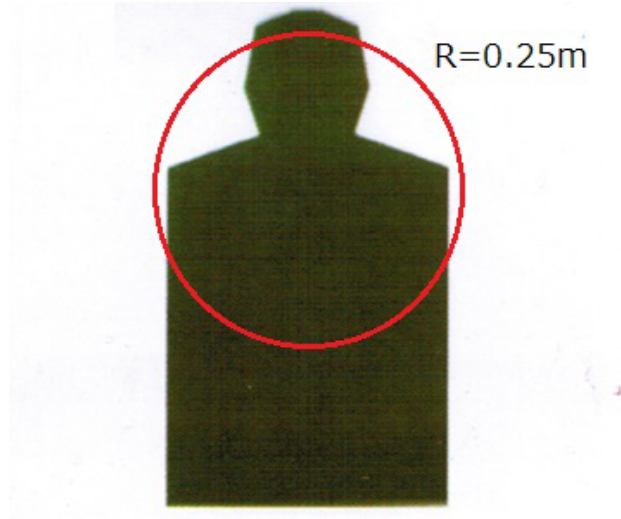


Figure 3.17: Circle of radius 0.25m and upper body

This multi-directional shooting experiment was carried out twenty times for twenty points chosen randomly. The experimental results indicate that the mean error of optical axis declination is 0.2263° , and the standard deviation is 0.0323° . When a target is 60 meters far from the camera, the error in distance caused by the optical axis declination is about 0.25m. By considering a circle of radius 0.25m with an upper body of person as shown in **Fig.3.17**, we can conclude that if the target is a person, the optical axis of the third camera can aim at the range of his/her upper body automatically.

Chapter 4

The generation of 3D site model and dynamic 3D display

4.1 Spatial position calculation for pixels

As described in **Section 3.3**, when two cameras make their optical axes aim at a same point P , then **Eq.3.12** can be given. If the point's rotation angles with respect to two cameras are known, it is not difficult to calculate D_l and D_r , i.e., the distances from the point to each camera. Next, the coordinate value of this point can be obtained by using **Eq.2.6**.

In order to calculate the rotation angle of any pixel on the image, the rotation angle between two adjacent pixels should be measured. γ shown in **Fig.4.1** is the rotation angle on 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 we used is 640×480 pixels. Assuming that γ is 15° , we can calculate $\theta_1 = 0.0048^\circ$ and $\theta_2 = 0.0045^\circ$. Their difference is 0.003° which is much smaller than the camera's accuracy of 0.03° . In the case of 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° . Therefore, it is considered that the difference can be neglected. 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 the

camera image.

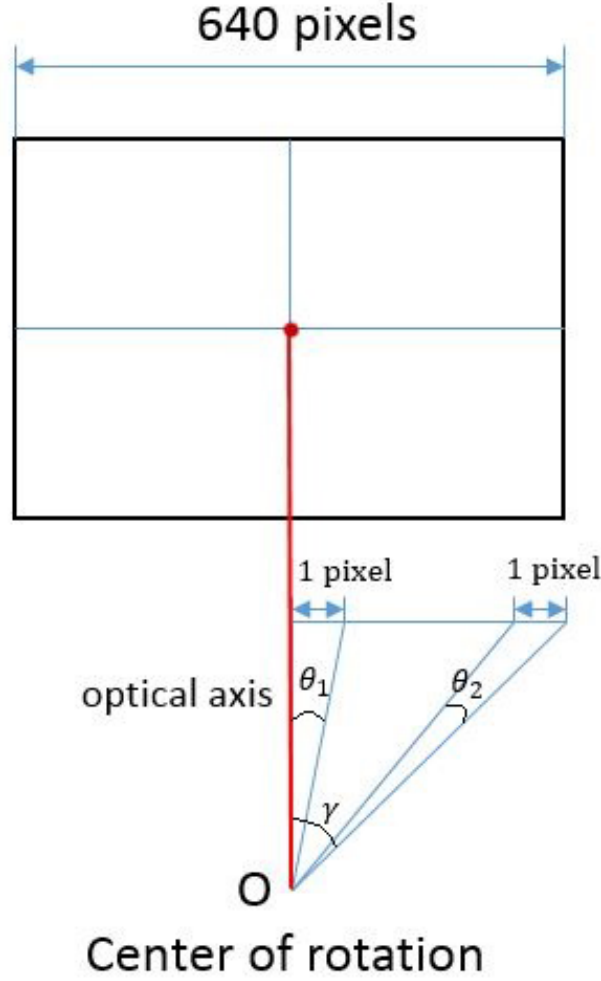


Figure 4.1: Rotation angles corresponding to pixels

Based on the above assumptions, θ can be measured for different zoom values. When zoom is a specific value, we make the camera aim at a spatial point P as shown in **Fig.4.2(a)**, and record the its pan value α_P . Next, rotate the camera horizontally to move P 's projection to the edge of the image, and record the current pan value as α_1 . Then, γ can be calculated as $\gamma = \alpha_P - \alpha_1$, and θ can be calculated as $\theta = \gamma/320$. We 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.

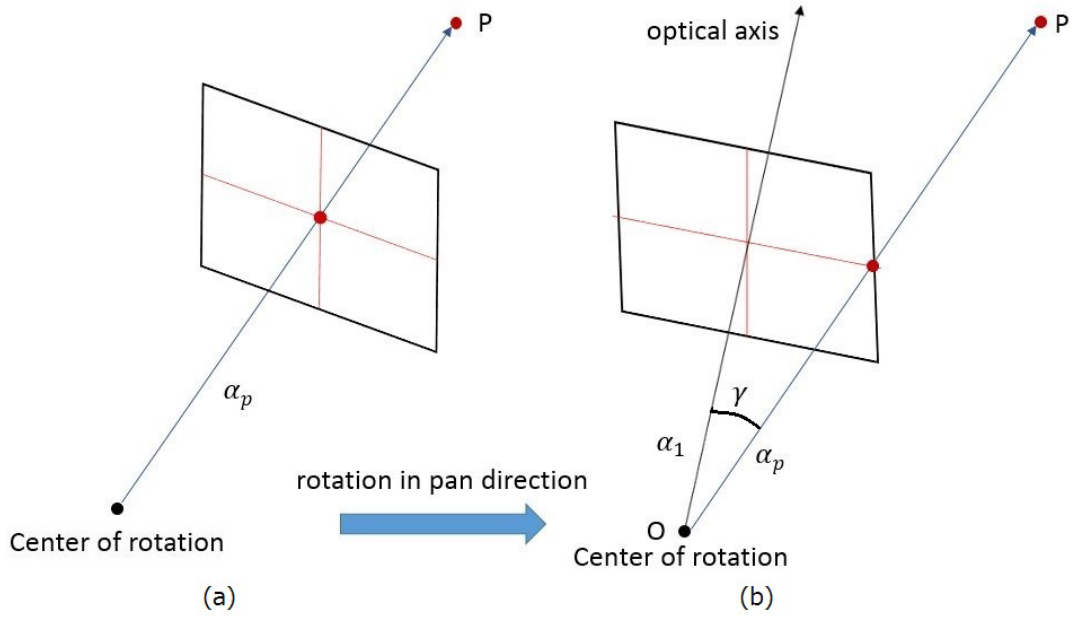


Figure 4.2: Relationship between the rotation angle and the image pixel

Similarly, any point's rotation angle with respect to camera's initial coordinate system can be calculated by using its coordinates on the image plane. For example, as shown in **Fig.4.3**, 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:

$$\begin{cases} \alpha_{P1} = \alpha_1 + \theta \times i \\ \beta_{P1} = \beta_1 + \theta \times j \end{cases} \quad (4.1)$$

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

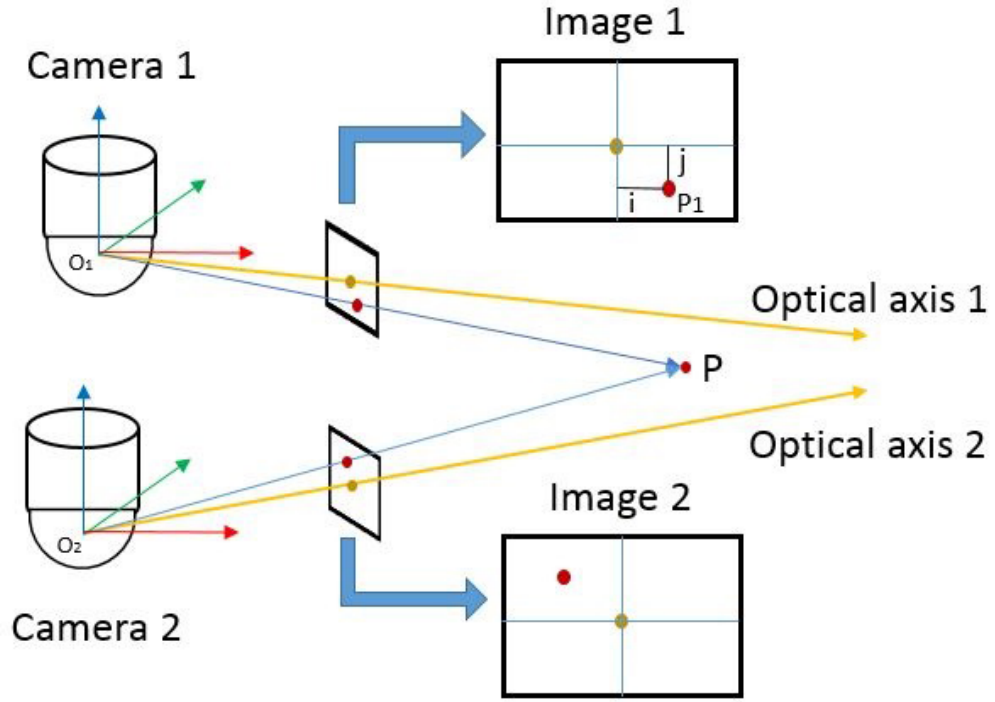


Figure 4.3: Rotation angle estimation

Thus we can see that 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. Further, by the previous explanation, we can conclude that its coordinate value in world coordinate system can be obtained.

4.2 3D reconstruction

As previously described, the difference between proposed system and existing systems [2, 3, 4, 5, 6] is that we appropriately use system's own PTZ cameras to calculate spatial information of scene and generate 3D site model. Since the cameras and the scene are fused in one coordinate system, it is no need to do the coordinates transformation among different coordinate systems. Moreover, our system has additional features: 1) it is enough to do the external parameter estimation only once. It means that, when camera rotates, it is not necessary to do the parameter estimation again nor to do the self-calibration [39, 40, 41]. 2) 3D coordinate values of a point can be calculated by only using its rotation angles with respect to two cameras.

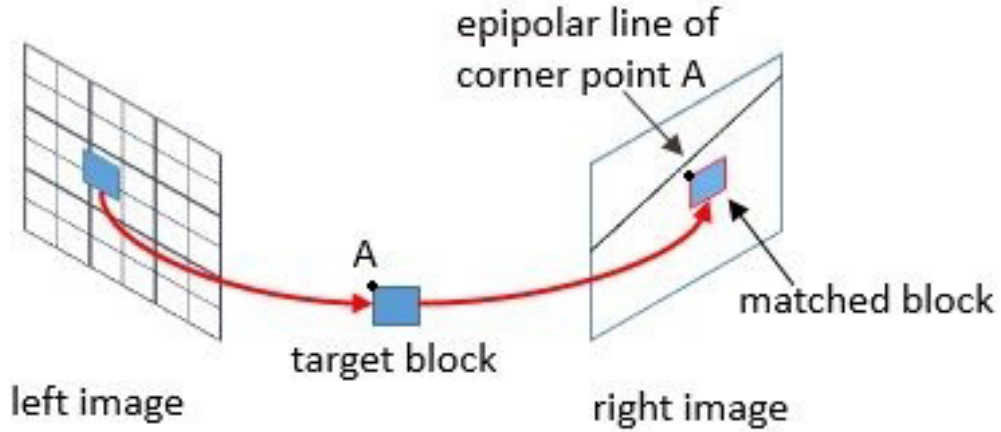


Figure 4.4: Block matching aided by epipolar line

As described above, two cameras in close proximity to each other in the surveillance system can be used to calculate the spatial information of their common field of views. In order to make a dense point cloud model, we should find corresponding points of two images as many as possible. In this dissertation, we divided one image (left image in **Fig.4.4**) into a number of blocks, and search for the similar part for each block in the other image (right image in **Fig.4.4**) by using a template matching method. Then, in order to screen the

matched blocks, we used the epipolar line to remove false matched blocks. For example, as shown in **Fig.4.4**, we calculated the equation of epipolar line which corresponds to the upper-left corner point of the target block. If the distance between the line and the upper-left corner point of matched block is greater than a threshold value, it is considered a miss matching. Otherwise, we considered that the matched block is true corresponding block, and all points of this block are corresponding points. The block matching step was executed many times with different segmentations of left image, so that we can find more corresponding points.

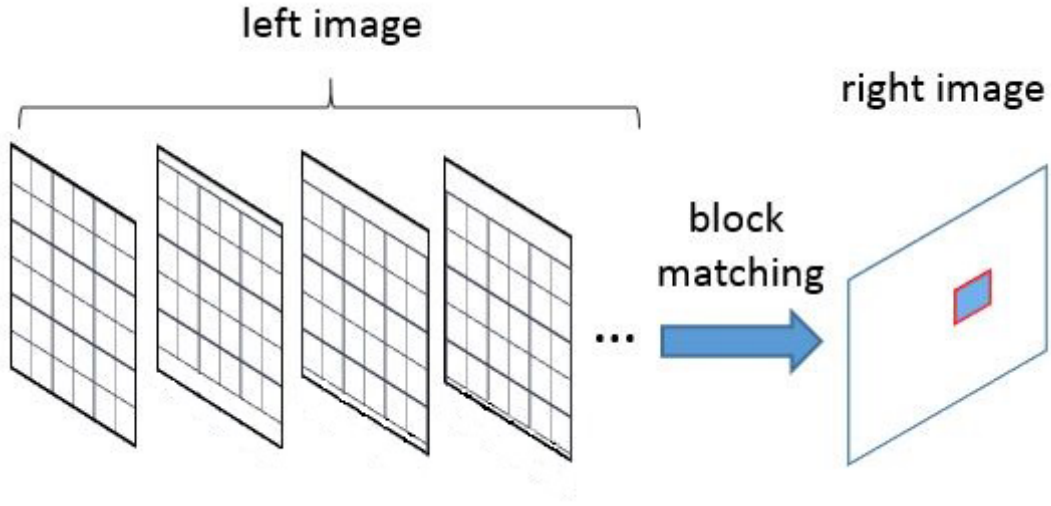


Figure 4.5: Different segmentations for block matching

In this dissertation, the camera image resolution is 640×480 pixel, and 40×40 pixel blocks were used for template matching. **Figure 4.5** gives a segmentation approach: the left image was divided into 40×40 pixel blocks 40 times with different starting points followed by $(0,0)$, $(1,0) \cdots (39,0)$. The blocks smaller than 40×40 will be discarded, and the point without matching point will be discarded as well. If a point has multiple matching points, we took the means of their coordinate values in the horizontal and vertical directions as the matched location.

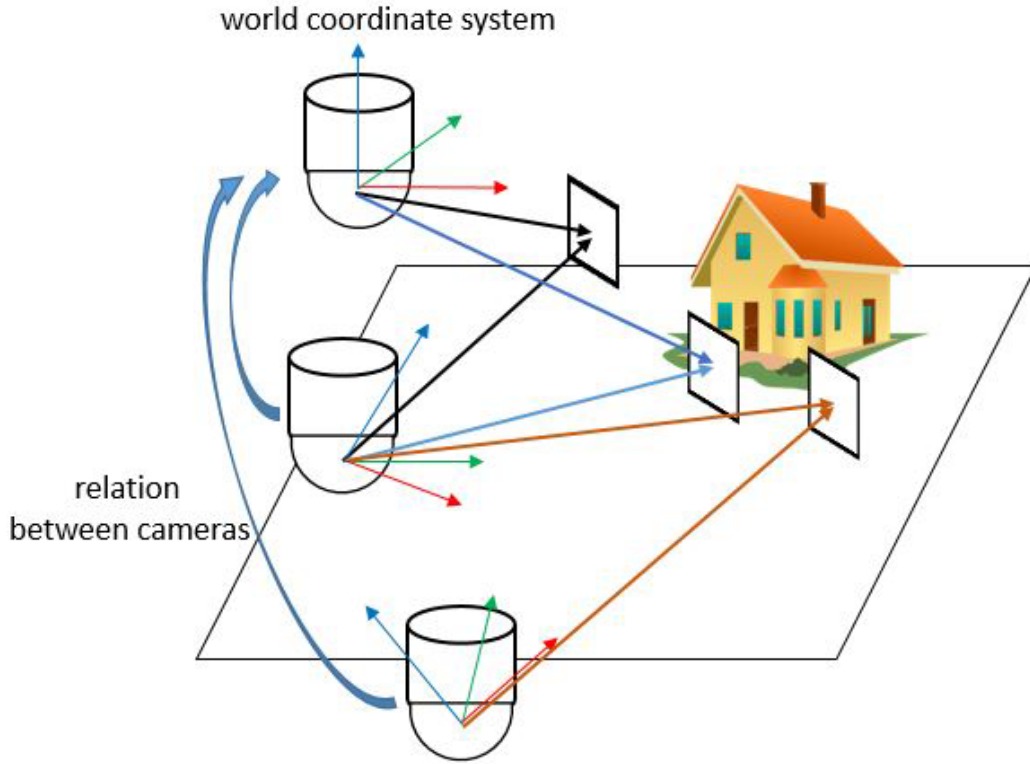


Figure 4.6: Generating 3D site model with surveillance camera pairs

Therefore, if two cameras aim at the same area, and the corresponding points on the two cameras' images can be extracted, the 3D site model of this area can be generated. Furthermore, as indicated in **Fig.4.6**, two cameras can generate the 3D site model for a larger area through its scanning. In other words, if an area is seen by two cameras, its 3D site model can be obtained. In this dissertation, the 3D site model 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.

4.3 Position estimation by one camera

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

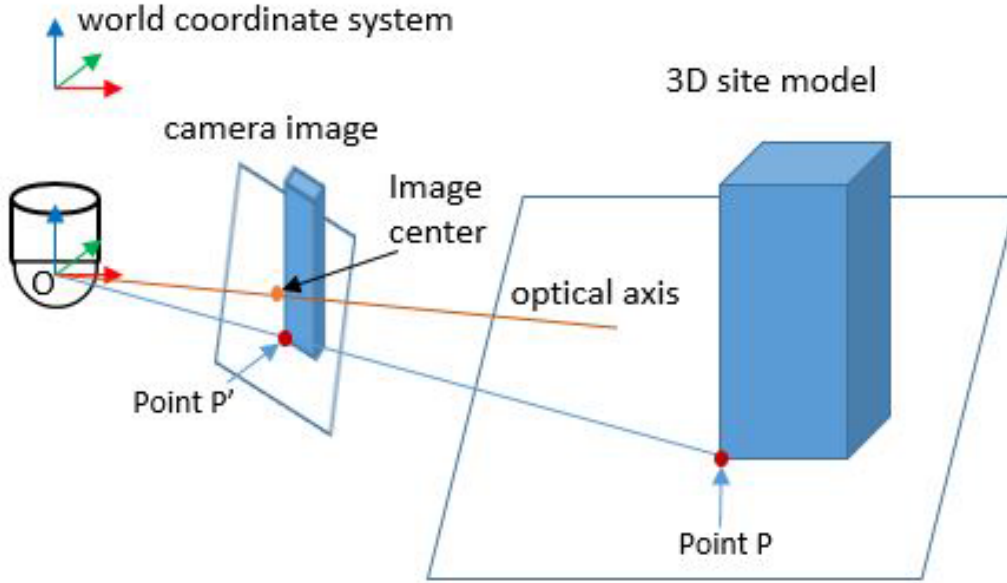


Figure 4.7: Position estimation for a point

Since the 3D site model 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. As emphasized in previous section, it is not necessary to do the transformations among different coordinate systems. So that, if we want to know spatial positions of points on the image, we can just find its corresponding position on 3D site model.

For example, as shown in **Fig.4.7**, in order to find the point P' position in the 3D site model, substitute camera's orientation information (represented by α_1 and β_1 in **Eq.(4.1)**), the rotation angle corresponding to one pixel (represented by θ in **Eq.(4.1)**) and coordinates of P' on image plane into **Eq.(4.1)** to locate the direction

of vector $\overrightarrow{OP'}$, and then its intersection (point P) with the 3D site model can be obtained easily.

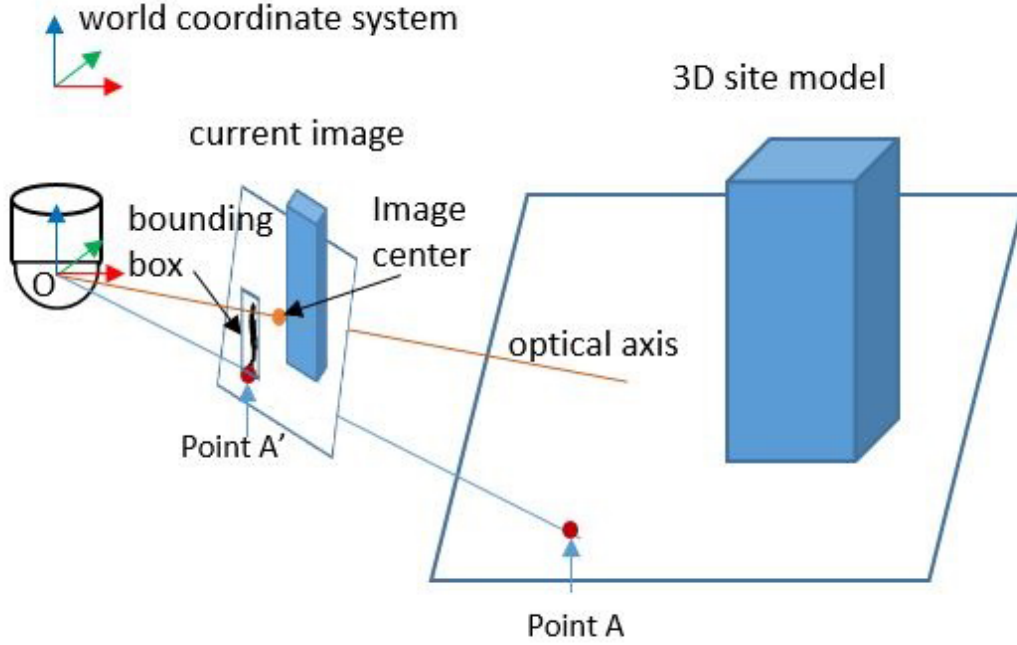


Figure 4.8: Position estimation for a pedestrian

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 with 3D site model. As shown in **Fig.4.8**, a walking person detected from the camera image is marked by a box, 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 each point of the 3D site model contains spatial information of real world, the 3D coordinate value of the point A can be regarded as the person's coordinate value.

In summary, if a pedestrian is shot by one camera of the surveillance system, his/her coordinates can be calculated. Not only that, by using the 3D coordinates, other cameras in this system can be controlled to track him/her from different directions.

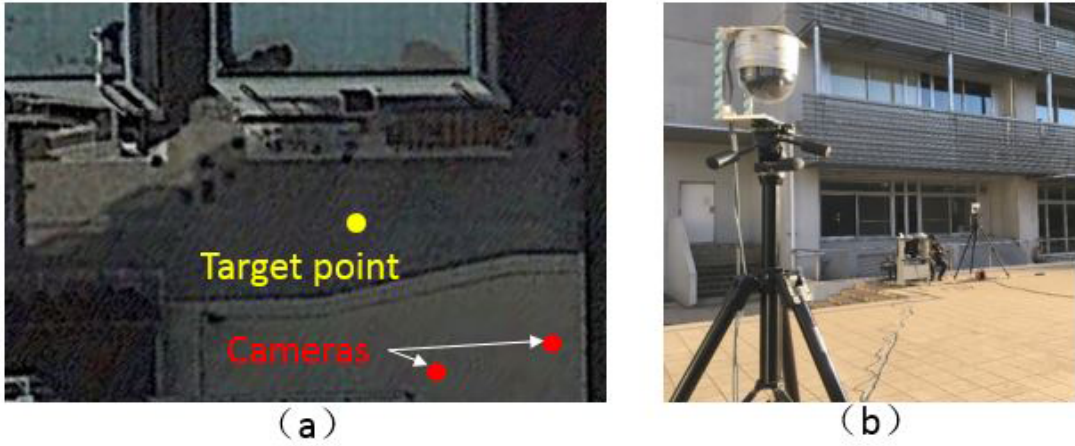


Figure 4.9: Experiment of distance estimation

In order to test the accuracy of this method, we did a distance measurement experiment. Here, we just compared the distances measured by different methods, because the accuracy of distances can represent the accuracy of spatial positions. **Figure.4.9(a)** shows the experimental site, the approximate position of two cameras and one target point. Ten similar target points are set on the ground. **Figure.4.9(b)** shows the layout of two cameras used in this experiment, which were mounted on tripods. The distance between two cameras was about 10 meters.

Table 4.1: Results of distance measurement

No.	Truth values	Distances measured by two cameras (error)	Distances measured by one camera (error)
1	16.54	16.7355 (0.1955)	16.8705 (0.3305)
2	17.66	17.5166 (0.1434)	17.3008 (0.3592)
3	17.58	17.3268 (0.2532)	17.3460 (0.2340)
4	16.83	16.6666 (0.1634)	16.4588 (0.3712)
5	17.05	16.9274 (0.1226)	16.6770 (0.3730)
6	21.00	20.9778 (0.0222)	21.4445 (0.4445)
7	18.28	18.5629 (0.2829)	18.5730 (0.2930)
8	21.75	21.6736 (0.0764)	21.4789 (0.2711)
9	23.96	23.8794 (0.0806)	24.3582 (0.3982)
10	25.97	25.7810 (0.1890)	26.3904 (0.4204)

One camera's coordinate system was set as the world coordinate system. Then we measured the distances from target points to the origin of the world coordinate system with different methods. Firstly,

we measured the distances with a laser range finder. The data measured by the laser range finder was used as true value. Next, according to the methods described above, we measured the distances with the camera pair and only one camera separately. The results are shown in the **Table.4.1**. As we can see, the accuracy of two cameras measurement is higher than the accuracy of one camera measurement, and the error of two cameras measurement does not change so much as the distance changes. On the other hand, the error of one camera measurement is relatively large, the largest one is almost equal to the width of an adult upper body. The low camera height is considered to be a major cause of the large error. Because the low height causes the viewing angle to be small, so that the target points was not easily observed on the camera image.

4.4 Experiment of dynamic 3D display

4.4.1 Experiment of 3D model generation

According to the method described above, the single-camera geolocation and multi-view tracking can be achieved based on the 3D site models, so we tried to generate the 3D site models on scences. First of all, we made two cameras aim at the same point to obtain the maximum common field of views. Next, we generated the 3D site model by using a template matching method to find dense corresponding points. Finally, we removed the noise points and smoothed the rest 3D points with a Gaussian filter of $\sigma = 5.0$.

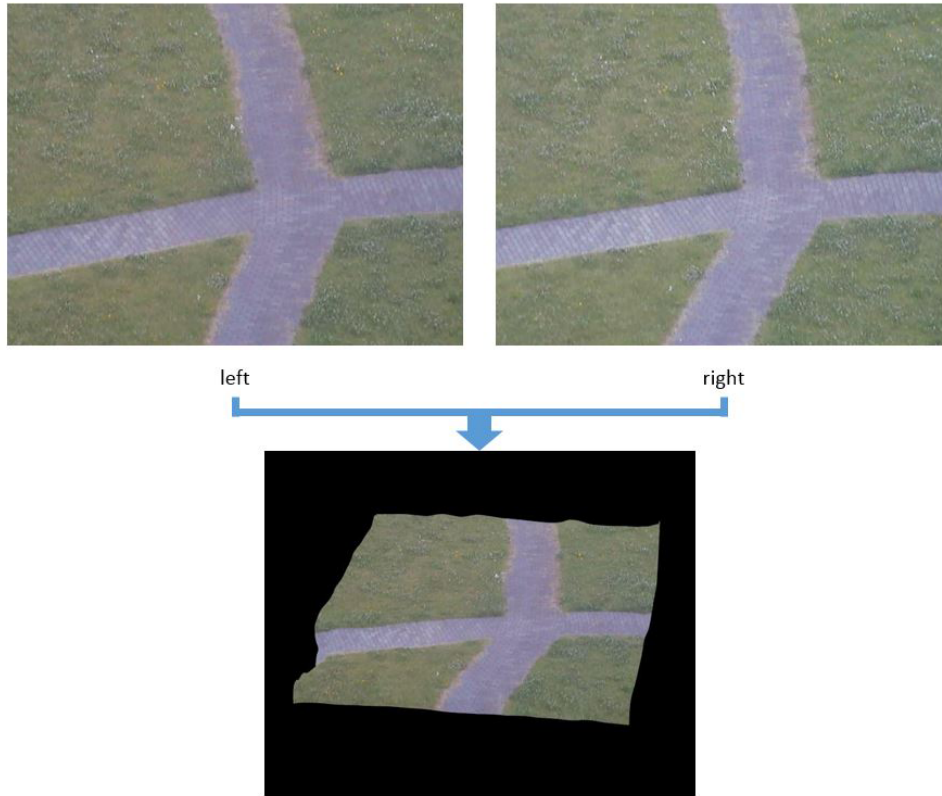


Figure 4.10: 3D site model for a lawn field

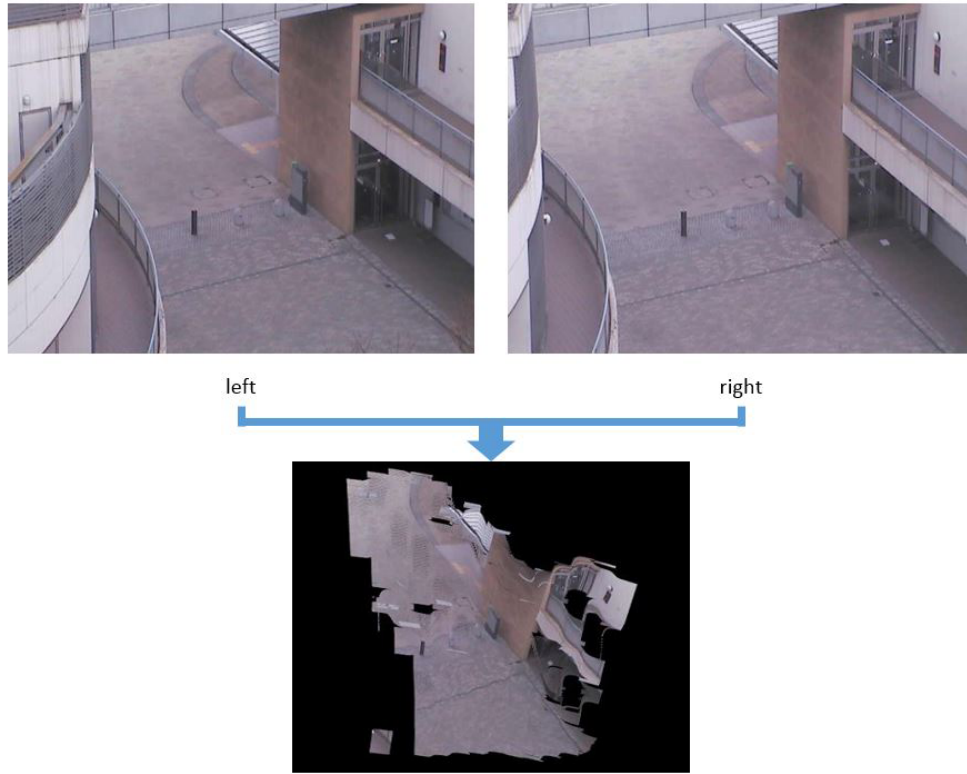


Figure 4.11: 3D site model for a road

Figure 4.10 and **Fig.4.11** show two experiments of 3D site model generation with different zoom values. The two upper images in **Fig.4.10** 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.4.10**, the scene in **Fig.4.11** is a little complicated. As shown in **Fig.4.11**, some of left and upper parts cannot be matched. And, 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 by two cameras due to the limitation of camera orientations. Even so, the ground and the brown wall, as the major part of this picture, are almost reconstructed.

4.4.2 3D dynamic display experiment

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

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.

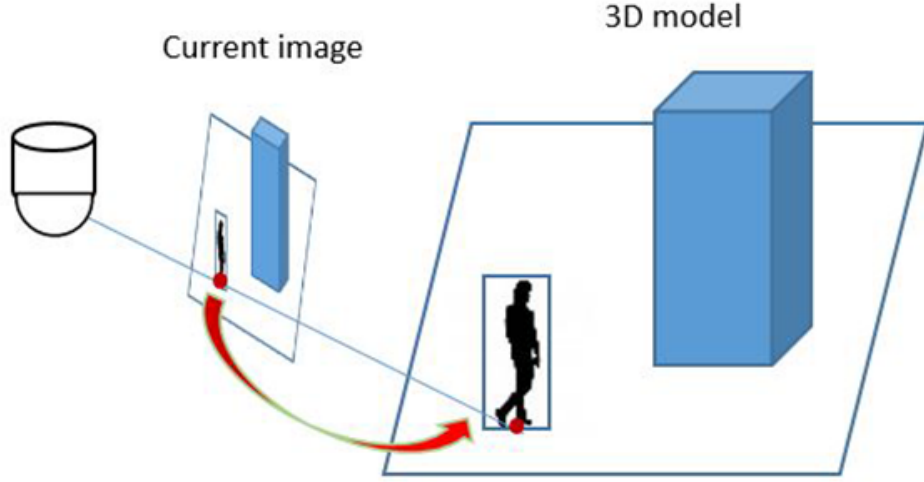


Figure 4.12: Rendering a pedestrian onto 3D environment

As discussed in **Section 4.3**, by using the 3D site model, we can calculate the spatial positions of walking persons even if they are shot by only one camera. **Figure 4.12** illustrates that a moving person is detected from every frame of original video and rendered onto a 3D site model. In this dissertation, 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) method to detect moving persons, and the orientation and velocity of the target can also be estimated by the method. We present MHI method in detail in the next chapter.

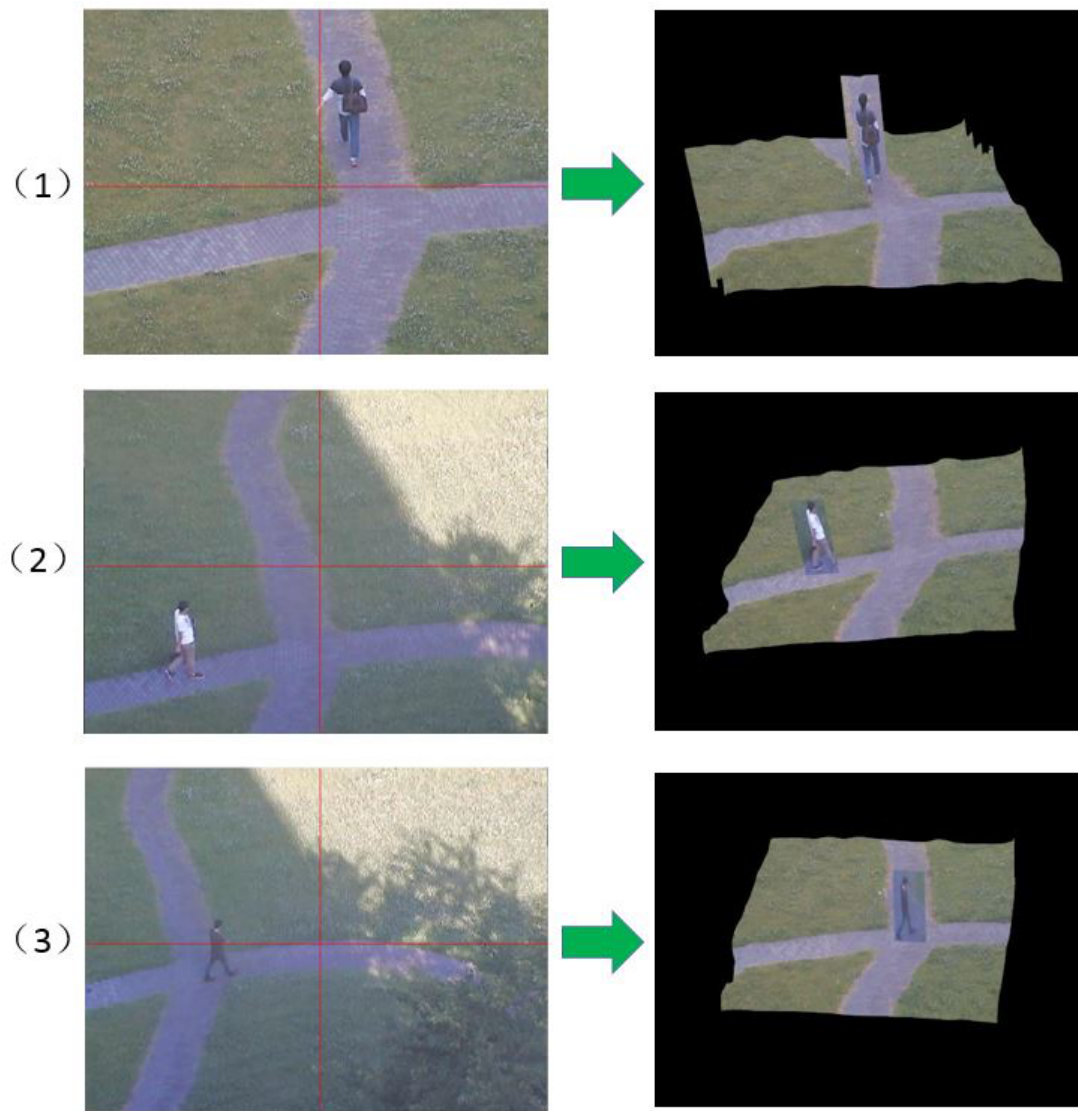


Figure 4.13: Visualization of pedestrians with different orientation and magnification on a lawn field

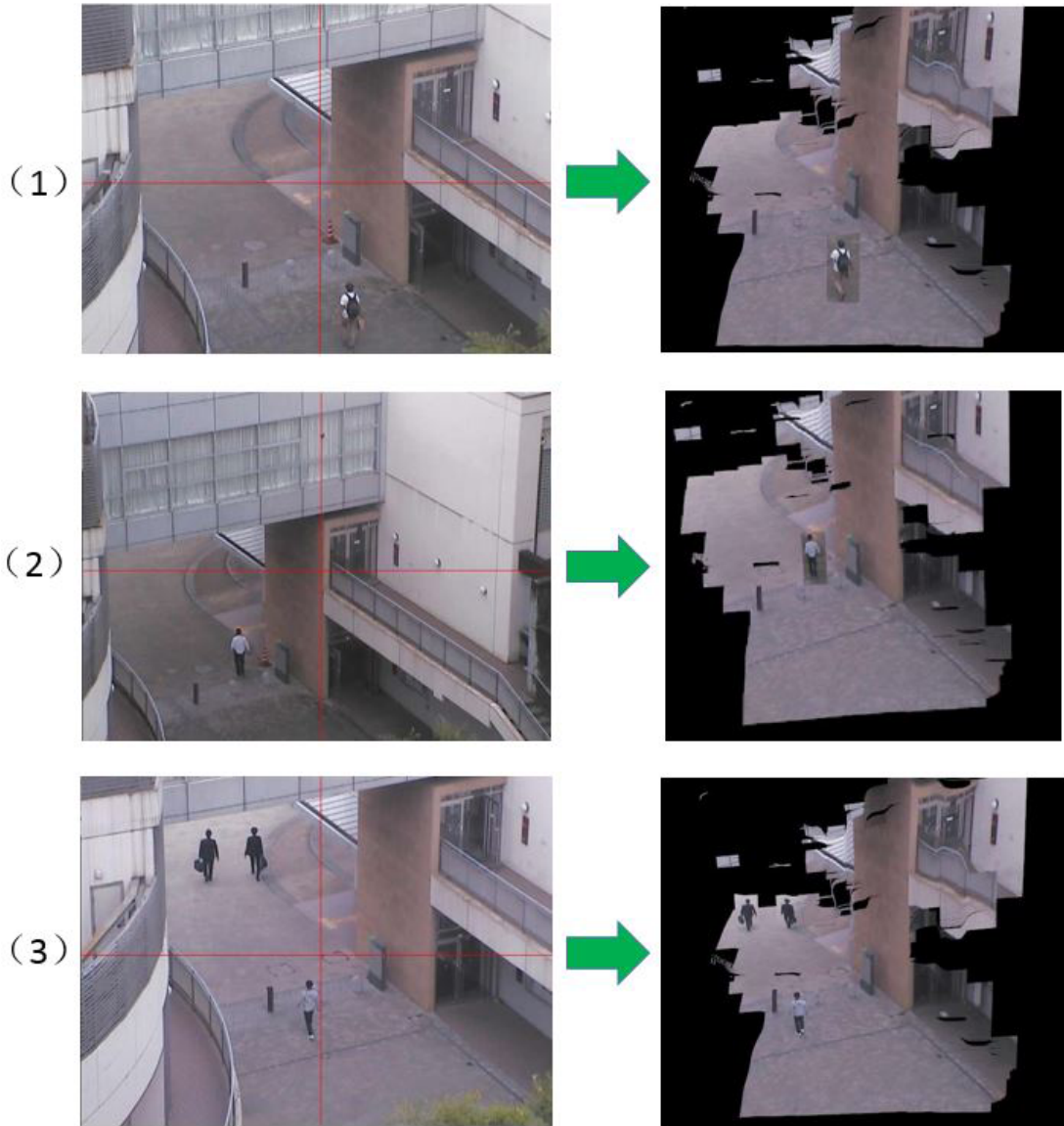


Figure 4.14: Visualization of pedestrians with different orientation and magnification on a road

We carried out the 3D dynamic display 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 test that the pedestrian location can be estimated with different camera orientation and magnification, we constantly changed camera orientation and the magnification during the experiments. In both **Fig.4.13** and **Fig.4.14**, 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 conclude that even if the camera's orientation and zoom value changes, the position of pedestrians can be calculated correctly, and the pedestrians can be rendered at correct locations in the 3D site model.

Chapter 5

Multi-view tracking

5.1 Extraction of pedestrians

As mentioned in Chapter 4, we used motion history image (MHI) method to extract pedestrians from the frames of camera video. By calculating the corresponding points in 3D site model, the 3D coordinate values of pedestrians can be obtained.

The motion history image (MHI) proposed by Bobick and Davis [42] utilizes motion shape information of a video to recognize actions. In MHI, pixel intensity is a function of the temporal history of motion at that point. Given an action video, the MHI H_τ can be obtained as follows. (1) Extract a binary image sequence $D(x, y, t)$ which indicates the regions of motion. $D(x, y, t)$ can be obtained by frame differentiating, where x, y show the position and t shows the index of frames. (2) Based on $D(x, y, t)$, the motion history image is defined as

$$H_\tau(x, y, t) = \begin{cases} \tau, & \text{if } D(x, y, t) = 1 \\ \max(0, H_\tau(x, y, t-1) - \delta), & \text{otherwise} \end{cases} \quad (5.1)$$

where τ is the maximum duration a motion is stored, δ is the decay parameter.

The result of MHI is a scalar-valued image where more recently moving pixels are brighter. **Fig.5.1** shows an example of the motion history image method. **Figure 5.1(a)** is a binary image obtained by frame differentiating. **Figure 5.1(b)** is a motion history image in which noise points are removed.



Figure 5.1: Example of MHI

By finding the contours in motion history image, the moving person can be marked with a box as shown in **Fig. 5.2**.

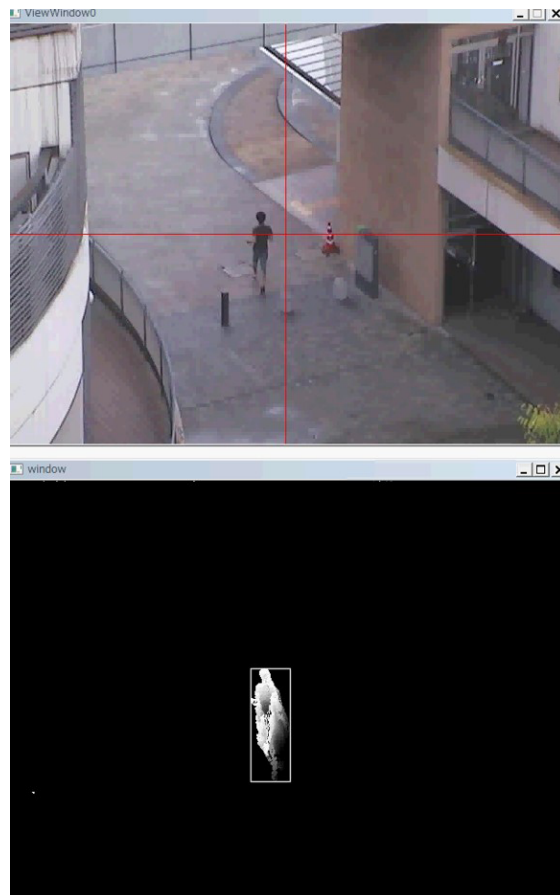


Figure 5.2: Marking a moving person

It can be seen in **Fig. 5.2** that the trajectory of the person in a period of time is all included in the box. Since the motion history

image contains the movement information in a period of time, it can be used to estimate the pedestrian's movement direction and speed.

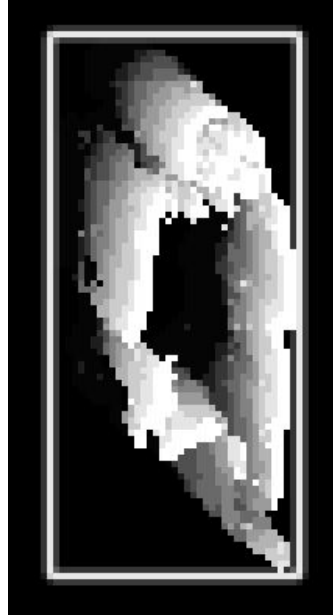


Figure 5.3: MHI of a pedestrian

Figure 5.3 shows a pedestrian's motion history image which is made by 10 frames of original video. In addition, the decay parameter δ is 25, so the brightest pixels are 255, the second bright pixels are 230.

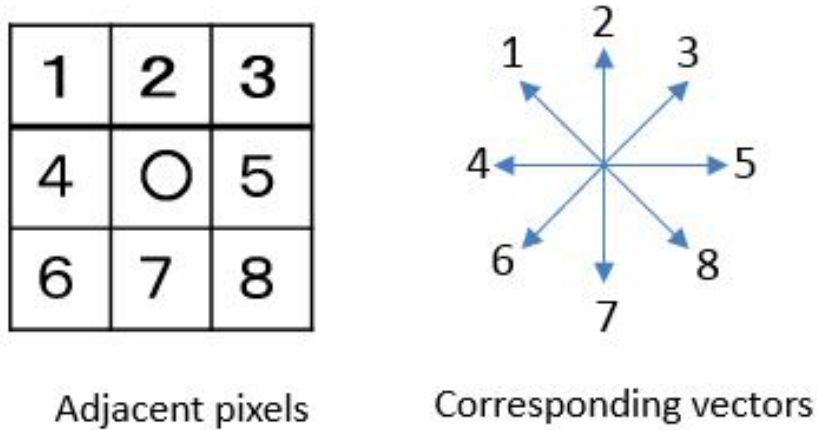


Figure 5.4: Adjacent pixels and corresponding vectors

In this research, we estimate the direction by counting the number of pixels whose value is 255 in the neighbors of each pixel whose value

is 230. Specifically, as shown in **Fig.5.4**, assume that each adjacent pixel corresponds to a vector represented by 1,2,3...8 respectively. The vectors are set to 1: $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, 2:(0,1), 3: $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, 4:(-1,0), 5: (1,0), 6: $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$, 7: (0,-1), 8: $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$.

The following example shows how to calculate the direction of motion. Assume the numbers of vector 5,7,and 8 are 20, the numbers of the rest vectors are 0, the number of pixel of value 230 is 25, so the motion vector of the target can calculate as $\vec{M} = [20 \times (1, 0) + 20 \times (0, -1) + 20 \times (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})]/25 \approx (1.71, -1.71)$. So the motion vector for 10 frames (about 1/3 second) can be calculated.

Furthermore, by using the motion vector, the location can of the pedestrian after a short time can be predicted. As shown in **Fig.5.5**, the white circles show the predicted positions of the feet after about one second.

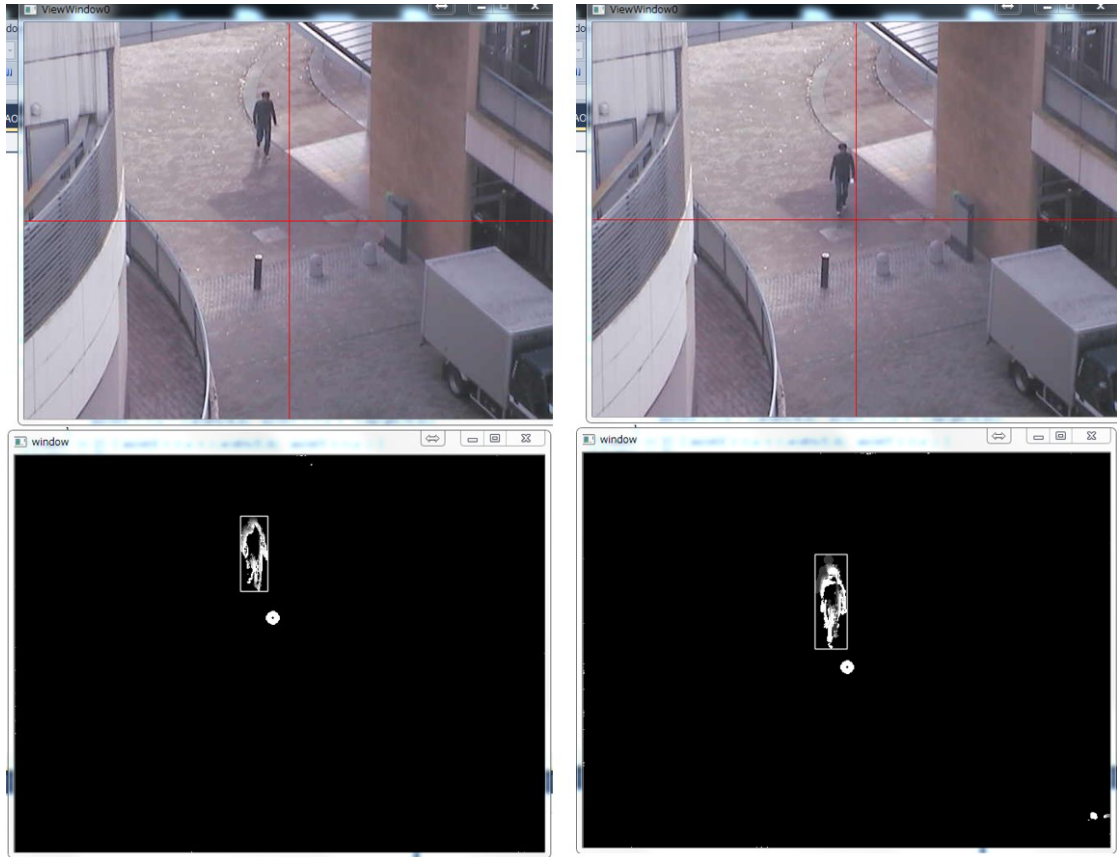


Figure 5.5: Position prediction

5.2 Multi-view tracking experiment

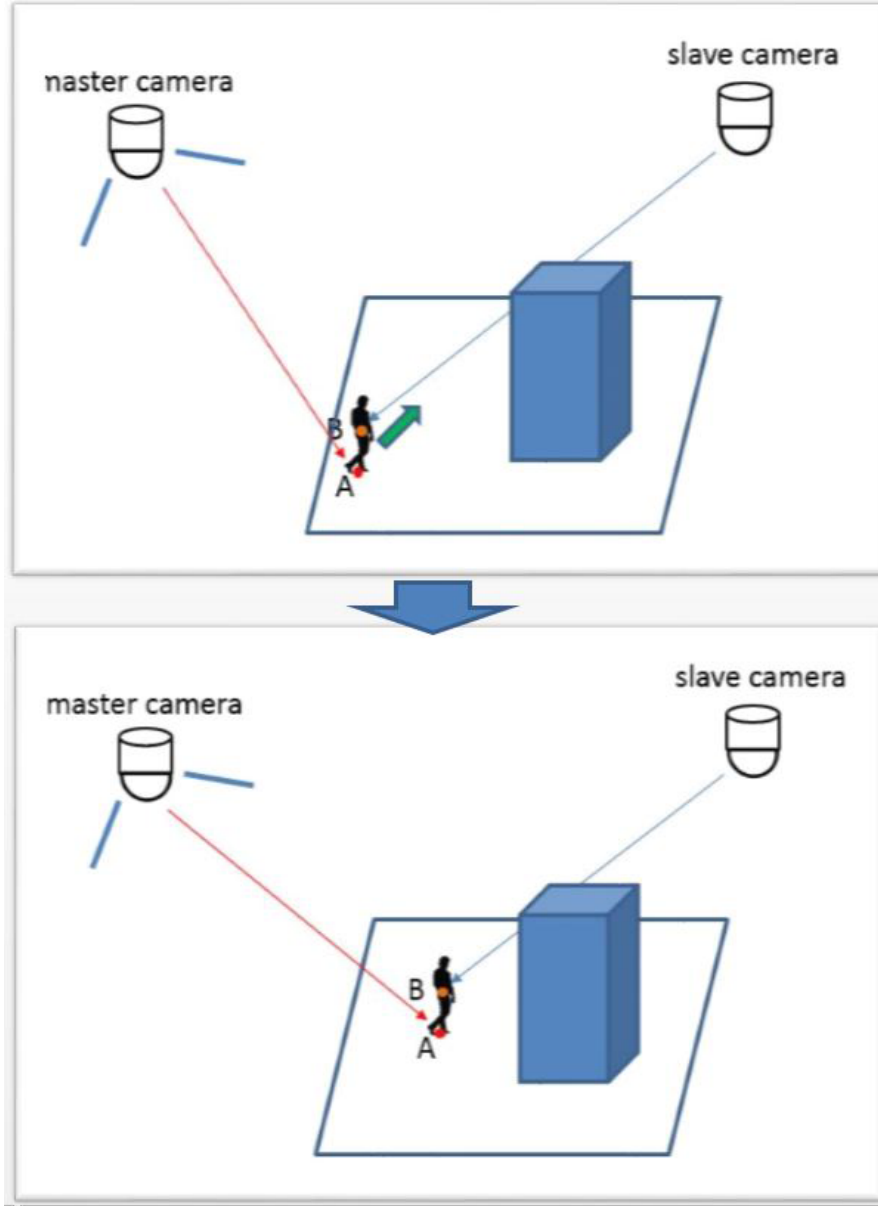


Figure 5.6: Multi-view tracking

Based on the reconstructed 3D site models and the 3D dynamic display experiment, we realized a multi-view tracking function. As shown in **Fig.5.6**, 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 1.5m) is added to the point A in the vertical direction to get the approximate location of pedestrian's upper body (shown as point B). Next, the slave camera is controlled to aim at

the point B. Then, we can repeat the above steps for each frame of the master camera to achieve a continuous tracking. As shown in **Fig.5.7** and **Fig.5.8**, in order to cover a large area, the master camera is set to wide-angle state to get a wide field of view and the slave camera with telephoto state can get enlarged images so that it can see more details.

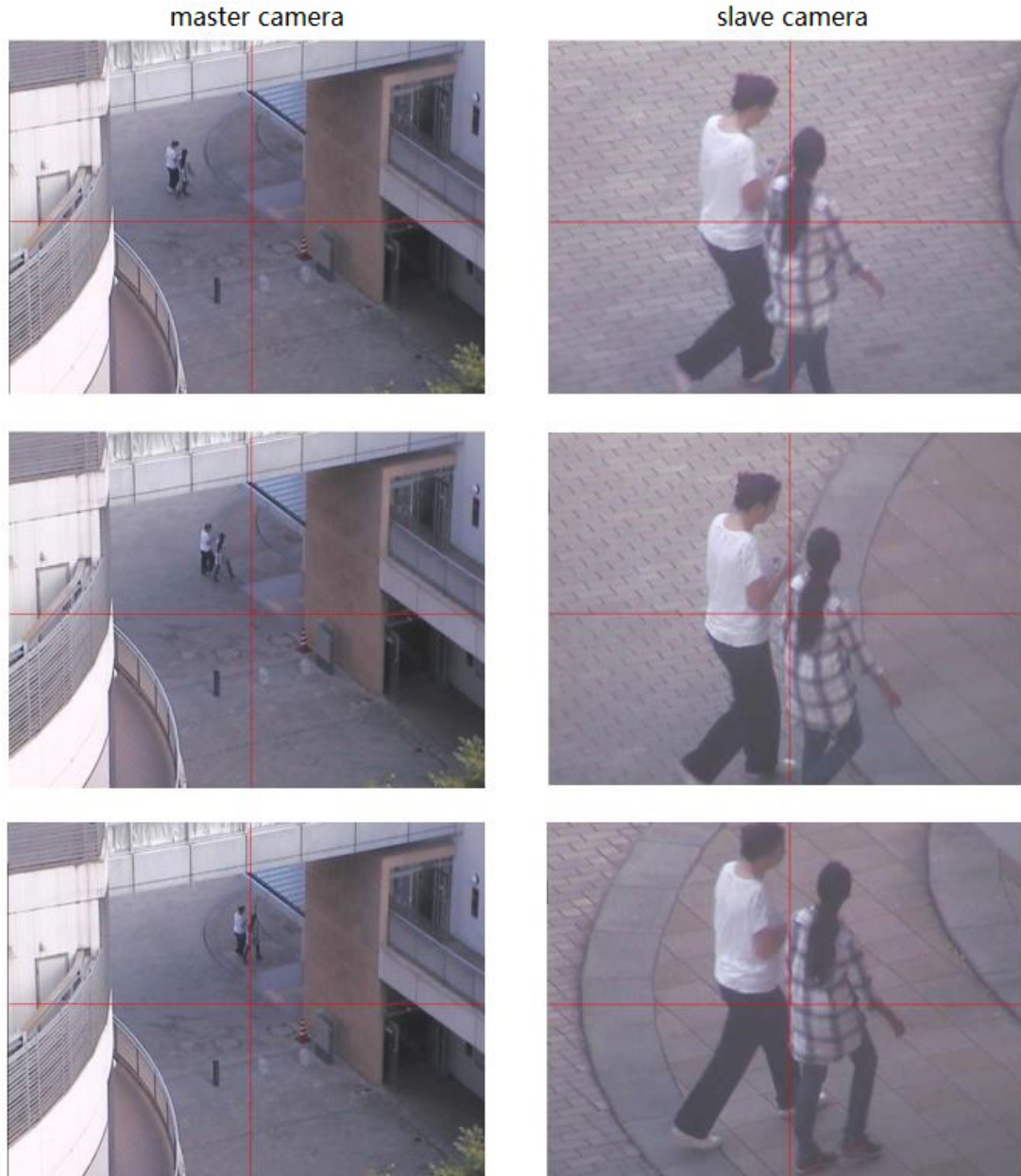


Figure 5.7: Tracking example 1



Figure 5.8: Tracking example 2

In the actual experiments, we can see that the optical axis could not aim at pedestrian's waist stably, but floated in the range from chest to thigh. That's because the targets are far from the cameras more than 50 meters, so the camera rotation error causes a relatively large effect.

Comparing with other long distance tracking methods [43, 44, 45, 46, 47], our method has two features:

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's 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. Furthermore, this method also has another advantage. For the traditional methods which use image information to track a target, the camera often zooms in to see more clearly. However, if the target is relatively fast, it is likely to run out of the camera's field of vision. In contrast to it, because the main camera which provides the target location maintains a large field of vision, it reduces the likelihood of missing targets

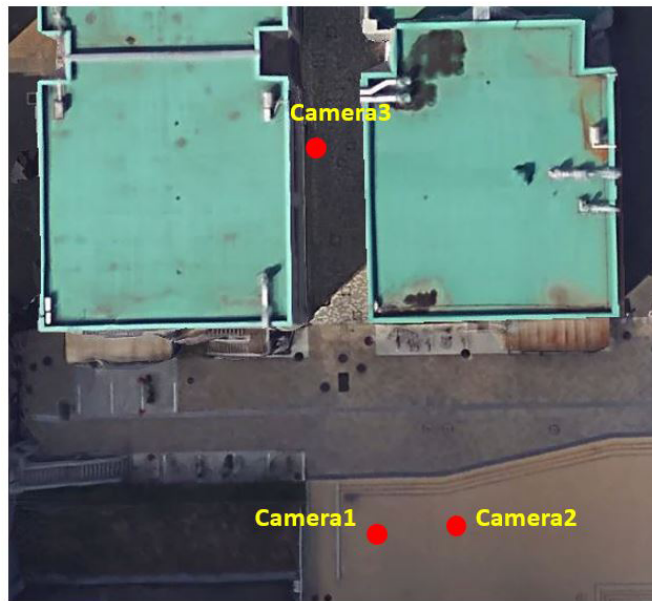


Figure 5.9: The site of the experiment

Although the coordinates of the pedestrian can be calculated by only master camera, the master camera must remain stationary. When the target moves out of the master camera's field of vision or is blocked, tracking has to stop even if the other cameras can still see the target. So in order to track the target continuously, the system needs to achieve the camera switching function. It means that

once the target disappears from the field of view of the main camera, the other camera which can see the target becomes the main camera, so that the coordinates of the target can continue to be obtained.

Figure 5.9 shows the experiment site and the locations of three cameras. As you can see, each camera has a limited field of view.



Figure 5.10: Master camera switching experiment

Figure 5.10 shows the process of continual tracking experiment. In **Figure 5.10(a)**, camera 2 is the master camera, camera 1 and camera 3 are aiming at the pedestrian. However, line of sight of the camera 3 is blocked by the wall, so that the target is not visible in

the image taken by camera 3. As shown in **Figure 5.10(b)**, when the pedestrian moves out of camera 2's field of view and camera 3 can see him, the camera 3 becomes the master camera, and camera 2 becomes slave camera. In **Figure 5.10(c)**, camera 1 and camera 2 are aiming at the target, but only camera 1 can see him. Then, when the pedestrian moves out of the wall as shown in **Figure 5.10(d)**, camera 2 can see him again, and camera 1 and 2 can track him from different directions.

As illustrated above, in this experiment, because the master camera can be changed automatically and the slave cameras can aim at the target continuously, as long as one camera can see the target, the coordinate value of this target will be not lost.

Chapter 6

Conclusion

This dissertation presented a wide-area surveillance system based on multiple PTZ surveillance cameras.

In the conventional surveillance systems, each camera is independently controlled, and the operators have to watch multi-view 2D videos captured by dozens of cameras, so that the burden on the operators is heavy. In order to improve the surveillance efficiency considerably, it is necessary to integrate cameras installed at different positions by estimating their external parameters which refer to the relative position and the relative attitude.

In this dissertation, we proposed a new external parameter estimation method which aims at the features of the active PTZ camera. In this method, by using the L-shaped gauge, it is possible to estimate the external parameters based on the information of the optical axis, and avoid distortion of the lens and the displacement of corresponding points in conventional image processing method. In addition, since it is not necessary to calibrate the internal parameters, cameras can be operated to zoom in/out in accordance with a user's need during the estimation process. Once the external parameters are determined by this novel method, it is then possible to coordinate the motion of cameras.

Based on the calibrated external parameters, the cameras can be controlled to measure the spatial information of the surveillance area and generate its 3D site model. Then, in order to make good use of 3D site model and expand the use of surveillance system, we implemented the following two functions: dynamic 3D display and multi-view tracking.

The dynamic 3D display function means that the pedestrians can be displayed in a 3D representation in real time. With this function, the operator can master the overall situation of the surveillance area by just watching the 3D site model. Therefore, it can provide a more intuitive feeling and reduce the operator's cognitive load.

With the multi-view tracking function, one camera as the master camera provides the coordinate value of the target, other cameras can aim at the target automatically. If the target moves out of the field of view of the master camera, the camera which can still see the target is converted to the master camera so that it can continue to provide the coordinate value, thereby the target can be tracked by relays of multiple cameras.

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.

List of related publications

Journal papers

- An External Parameter Estimation Method for Wide-area Active Surveillance Cameras, IIEEJ Transactions on Image Electronics and Visual Computing, Vol.4 No.1, pp.53-62, 2016
- A 3D Dynamic Visualization Surveillance System, IJCSI International Journal of Computer Science Issues, Volume 13, Issue 5, pp.36-44, 2016

International conferences

- Study on External Parameter Estimation for Wide-area Active Surveillance Cameras, IEVC 2014

Domestic conferences

- 広域アクティブ監視カメラの外部パラメータの推定に関する研究, 画像電子学会 2014 年次大会
- 広域アクティブカメラを用いた空間位置の計測, 画像電子学会 2015 年次大会

References

- [1] Y. Xing, H. Nagahashi, X. Zhang:“ A 3D Dynamic Visualization Surveillance System”, IJCSI International Journal of Computer Science Issues, Volume 13, Issue 5, pp.36-44, 2016
- [2] Jvckenwood Corporation, <http://www.diginfo.tv/v/12-0041-d-en.php>, 2012
- [3] P. DeCamp, G. Shaw, R. Kubat, D. Roy:“ An immersive system for browsing and visualizing surveillance video”, MM. ACM, pp.371-380, 2010
- [4] Y. Yang, MC Chang, P. Tu, S. Lyu:“ Seeing as it happens: Real time 3D video event visualization”, ICIP, 2015
- [5] T. Kanade, R. Collins, A. Lipton, P. Burt, L. Wixson:“ advances in cooperative multi-sensor video surveillance”, Proc. of DARPA Image Understanding Workshop, Vol.1, pp.3-24, 1998
- [6] 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, pp.107-112, 2003
- [7] C. Tomasi, T. Kanade:“ Shape and motion from image streams under orthography: a factorization method”, International Journal of Computer Vision, 9-2, 137-154, 1992
- [8] M. Jancosek, T. Pajdla:“ Multi-view reconstruction preserving weakly-supported surfaces”, CVPR.IEEE, 2011
- [9] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch:“ Visual modeling with a hand-held camera”, IJCV, vol. 59, no. 3, pp.207-232, 2004

- [10] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, R. Szeliski: “ Building Rome in a day”, Commun. ACM, vol. 54, no. 10, pp.105-112, 2011
- [11] 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
- [12] A. Sampath, J. Shan: “ Segmentation and reconstruction of polyhedral building roofs from aerial Lidar point clouds”, IEEE Transactions on Geoscience and Remote Sensing, VOL. 48, NO. 3 , 2010
- [13] 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, pp. 118118, 2006
- [14] 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
- [15] 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
- [16] S. Pillai, S. Ramalingam, J. J. Leonard: “ High-performance and tunable stereo reconstruction”, International Conference on Robotics and Automation (ICRA), 2016
- [17] G. Xu, T. Saburo: “Three-Dimensional Vision” (in Japanese), Kyoritsu Shuppan, p.7-11, 1998.
- [18] Z. Zhang: “A Flexible New Technique for Camera Calibration”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(22):13301334, 2000
- [19] D. Herrera C., J. Kannala, J. Heikkila: “Joint Depth and Color Camera Calibration with Distortion Correction”, IEEE Trans-

actions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 10, 2012

- [20] Y. Xu, Y. Zhao, F. Wu, K. Yang: “Error Analysis of Calibration Parameters Estimation for Binocular Stereo Vision System”, IEEE International Conference on Imaging Systems and Techniques, 2013
- [21] L. Wang, Z. Liu, Z. Zhang: “An On-line Calibration Algorithm for External Parameters of Visual System Based on Binocular Stereo Cameras”, International Symposium on Optoelectronic Technology and Application, 2014
- [22] Y. I. Abdel-Aziz, H. M. Karara: “Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-range Photogrammetry”. In Proc. of Symposium on Close-Range Photogrammetry, pages 118, 1971
- [23] T. Pribanic, P. Strum, S. Peharec: “Wand-based Calibration of 3D Kinematic System”, IET Computer Vision. Vis., Vol. 3, No. 3, pp.123-129, 2009
- [24] K. Shin, J. H. Mun: “A Multi-camera Calibration Method Using a 3Axis Frame and Wand”, International Journal of Precision Engineering and Manufacturing, Vol. 13, No. 2, PP. 283-289, 2012
- [25] L. Quan, Z. Lan: “Linear N-Point Camera Pose Determination”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 21, No. 8 (1999).
- [26] J. Davis, X. Chen: “Calibration Pan-tilt Cameras in Wide-area Surveillance Networks”, Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003
- [27] Z. Wu, R.J. Radke: “Using Scene Features to Improve Wide-Area Video Surveillance”, Proc. IEEE Workshop Camera Networks and Wide Area Scene Analysis, 2012
- [28] R. Steele, C. Jaynes, A. Leonardis, H. Bischof, and A. Pinz: “Over constrained Linear Estimation of Radial Distortion

and Multi-View Geometry”, Proc. of European Conference on Computer Vision, 2006

- [29] B. He, Y. Li: “Camera Calibration with Lens Distortion and from Vanishing Points”, Optical Eng., Vol. 48, No. 1, p. 013603, 2009
- [30] Z. Kukelova, T. Pajdla: “A Minimal Solution to Radial Distortion Auto-calibration”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 33, No. 12, pp. 2410-2422, 2011
- [31] L. de Agapito, R. Hartley, E. Hayman: “Linear Calibration of a Rotating and Zooming Camera”, Computer Vision and Pattern Recognition, 1999
- [32] L. de Agapito, E. Hayman, I. Reid: “Self-Calibration of Rotating and Zooming Cameras”, Int’l J. Computer Vision, Vol. 45, No. 2, pp. 107-127, 2001
- [33] E. Hayman, D. Murray: “The Effects of Translational Misalignment When Self-Calibrating Rotating and Zooming Cameras”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 8, pp. 1015-1020, 2003
- [34] M. Li, J.-M. Lavest: “Some Aspects of Zoom Lens Camera Calibration”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 18, No. 11, pp. 1105-1110, 1996
- [35] Z. Wu, Richard J. Radke: “Keeping a Pan-Tilt-Zoom Camera Calibrated”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 8, 2013
- [36] Y. Xing, H. Nagahashi, X. Zhang, K. Aoki: “An external parameter estimation method for wide-area active surveillance cameras”, IEEE Transactions on Image Electronics and Visual Computing, Vol.4 No.1, 2016
- [37] K. Madsen, H.B. Nielsen, O. Tingleff: “Methods for non-linear least squares problems”, Technical University of Denmark, p.24-29, 2004

- [38] M. J. Greenacre: "Theory and Applications of Correspondence Analysis", Academic Press. Appendix A: Singular Value Decomposition (SVD) and Multidimensional Analysis, 1984
- [39] 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
- [40] P. Sturm, Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97), 1997
- [41] 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), 725
- [42] J.W. Davis, A.F. Bobick, The recognition of human movement using temporal templates, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (3): 257267, 2001
- [43] 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, pp.113-120, 2003
- [44] J.A.Fayman, O.Sudarsky, E.Rivlin, M.Rudzsky, Zoom tracking and its applications, Machine Vision and Applications I3 (I), pp.25-37, 2001
- [45] X. Zhang, An object tracking system based on human neural pathways of binocular motor system, Proc. ICARCV, Singapore, pp.285-292, 2006
- [46] Y. Gao, X. Zhang, Zoom control of wide area tracking system, IEEE International Conference on Intelligent System, London, pp.402-407, 2010

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