A Novel Binocular Vision System for Surveillance Application

Zhigao Cui and Aihua Li
502 Faculty, Xi’an Institute of High-tech, Xi’an, China
Email: cuizg10@126.com, LCC@126.com

Abstract—Visual surveillance system has been well studied in past decades. Because of the conflict between surveillance range and resolution of single-camera system, visual surveillance using multiple cameras has attracted increasing interest in recent years. In multi-camera systems, dual-camera system which contains active camera such as PTZ (pan-tilt-zoom) camera is the simplest and typical one. The superiority of this system lies in that it can obtain multi-resolution information and depth information. But as far as we know, few of the dual systems make full use of the two superiorities. In this paper, we propose a new binocular vision system which is composed of two PTZ cameras. Different from other systems, stereo rectification is used to establish correspondence between two image coordinates. Then, both global and detailed image information can be obtained. Furthermore, we use the depth information to help solve the occlusion problem in tracking. The experimental results on indoor surveillance environments have demonstrated the effectiveness and robustness of our system.

Index Terms—Binocular Vision System, PTZ (Pan-Tilt-Zoom) Camera, Stereo Rectification, Occlusion

I. INTRODUCTION

With the rapid demands of security and safety, vision-based surveillance technologies have become an active subject in the field of computer vision [1-3]. Some large research projects have been conducted in many countries, e.g., the VSAM project supported by DARPA, the ESPRIT PASSWORDS project in Europe and Japan’s Cooperative Distributed Vision project. These systems have been well applied in commercial, law enforcement and military fields. The aim of vision-based surveillance system is to replace traditional passive surveillance system and accomplish the entire surveillance task as automatically as possible. There are several important factors in visual surveillance: global information, specific attention with high resolution (local information), stereo information and so on.

According to the number of cameras, the visual surveillance system can be classified into monocular and multi-camera system. Systems using single camera (monocular) have been widely applied in the past decades [4-6]. However, they are limited by their fixed view angles, fixed resolutions and limited depth information. Due to rapid progress in computer vision, camera control, network cooperation [7] technology, hardware development, and so on, multi-camera systems’ superiority become more and more obvious. Binocular vision system is a simple and typical one. In the past decades, lots of related works have been published. We only review some of them which we consider typical. Ref. [8] proposes a vision system using two omnidirectional cameras. This type of vision system is mainly used for robot vision, which can enlarge view field. Ref. [9] proposes a vision system with one omnidirectional camera and one PTZ camera. This kind of vision system is mostly applied in indoor environment. The omnidirectional camera is used for monitoring the whole surveillance scene, and PTZ camera is used for acquiring high resolution image of interested target. The two cameras work in a fixed master-slave mode. The main disadvantage of the above vision systems is the usage of the omnidirectional camera, whose image resolution is low and uneven. This drawback has great limitation in real application.

In addition to binocular vision system mentioned above, dual static, static plus active or dual active system are the mostly well applied vision systems in real application. Traditional dual static system could get depth information by stereo vision. Generally speaking, the larger the baseline, the more difficult the correspondence between two image points can be build. This kind of system can be integrated as a vision unit to improve its stability. All these systems work on a symmetrical mode. Ref. [10] proposes an object tracking method that combines color and depth information using dual static cameras. The works in [11] propose that use of plan-view maps to represent stereo information more efficiently. The main drawback of this kind of systems is that it could not obtain multi-resolution and multi-visual-angle information. Static plus active or dual active system always works on master-slave mode, which means that the two cameras are asymmetrical. In [12], Zhou et al. present a master-slave system for acquiring biometric imagery of interesting subject in an outdoor environment. The relationship between static image coordinates and active camera’s PTZ parameters is contained in a sample-interpolated mapping. Ref. [13] is similar to Ref. [12] in that it is also for outdoor scene, but it assumes that the ground in the area can be viewed as planar, and a linear transform associate with a mapping between object height in the master image and height in the slave image is applied to align two images. Bodor et al. present in Ref. [14] a system for both indoor and outdoor scene. Their
system disregards the base line between two cameras, and uses planar mapping from pixel in static image to pan and tilt parameters of PTZ camera. Li et al. [15] propose a master-slave surveillance system for indoor scene. The method uses a mosaic image created by images of slave camera to estimate the relationship between static master camera image plane and pan-tilt controls of slave camera. All these systems mentioned above use static and active cameras. Bimbo et al. in Ref. [16] propose a novel framework exploiting two PTZ cameras to achieve the task of relating the feet position of a person in the image of the master camera to head position in the image of the slave camera. The system also uses a global map or a mosaic image to realize master-slave configuration. The main advantage of these systems mentioned above is that it could obtain global information and local multi-scale information. But all these works did not use depth information.

In this paper, a novel binocular vision system using two PTZ cameras for indoor scene is presented. Different from other systems, the proposed system can work on multimode: asymmetrical mode and symmetrical mode. In asymmetrical mode or master-slave mode, we propose a rectification-disparity-based method to establish correspondence between two image coordinates, and it does not need any pre-configuration, such as image mosaic, mapping construction, etc. Then, the slave camera could capture high resolution image to assist master camera’ task. In symmetrical mode, we use images from the two cameras to estimate the depth of object located in common field of view (FOV). The calculated depth information is used to help solve the occlusion problem in tracking. The experimental results on indoor surveillance environments have demonstrated the effectiveness and robustness of our system. The paper is organized as follows: Section II gives an overall description of the PTZ camera calibration. Section III describes the stereo rectification method of dual PTZ cameras. Section IV introduces our proposed system, which consists of two modules, i.e. high resolution visual attention and object tracking using depth information. Section V gives the experimental results. Section VI summarizes the paper.

II. PTZ CAMERA CALIBRATION

Calibration of PTZ camera plays an important role in the visual system using PTZ cameras. In our system, PTZ camera calibration is an elementary preparation. This preparation can be done off-line. Our method is similar to [17] which is feature based combined with the parameters inquired from the PTZ camera. In order to calibration the PTZ camera, we first make some assumptions as follows:

1. The rotation axes of pan and tilt are orthogonal and intersect at the origin of the camera coordinate system. In this cases, the extrinsic matrix can be directly calculated from the pan and tilt parameters.
2. The camera’s pixel aspect ratio $\alpha = 1$, and skew $s = 0$.
3. The principal point $(u_0, v_0)$ is fixed, and we use zoom center to replace the principal point.

With the rapid improvement of camera manufacturing, these assumptions can be well satisfied. Based on the above assumptions, the simplified camera model can be written as:

$$x = \kappa K(Z) R(P,T) X$$

(1)

where $x$ and $X$ are image coordinates and world coordinates respectively, $\kappa$ is a scale factor. In (1), $K(Z)$ is the intrinsic matrix which can be determined by the zoom parameter:

$$K(Z) = \begin{bmatrix} f(Z) & 0 & u_0 \\ 0 & f(Z) & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

(2)

The extrinsic matrix $R(P,T)$ can be determined by the pan and tilt parameters which can be acquired from the camera:

$$R(P,T) = \begin{bmatrix} \cos(P) & 0 & \sin(P) \\ -\sin(P)\sin(T) & \cos(T) & \cos(P)\sin(T) \\ -\sin(P)\cos(T) & \cos(T) & \cos(P)\cos(T) \end{bmatrix}$$

(3)

A. Principal Point Estimation

As mentioned above in assumption (3), we use zoom center to replace the principal point. To estimate the zoom center, we capture successive images $I_0, \cdots , I_n$ from the camera at varying zoom level with fixed pan and tilt parameters, where $z$ is the maximum zoom level of the camera. Then we detect and match SIFT feature points [18] between $I_k (k = 1, \cdots , z)$ and $I_0$. For each of the matched points, we can obtain a straight line in the image plane. Ideally all the straight lines should converge at one point, i.e. the zoom center. So zoom center can be calculated by (4). Where $y = k_x x + b_i (i = 1, \cdots , n)$ is the straight line determined by each of the matched points, $n$ is the total number of the matched points.

$$\min_{x,y} \sum_{i=1}^{n} \frac{(y - k_i x - b_i)^2}{k_i^2 + 1}$$

(4)

Taking derivatives with respect to $x$ and $y$. Then we have

$$\begin{bmatrix} \sum_{i=1}^{n} \frac{k_i}{k_i^2 + 1} & \sum_{i=1}^{n} \frac{1}{k_i^2 + 1} \\ \sum_{i=1}^{n} \frac{-k_i}{k_i^2 + 1} & \sum_{i=1}^{n} \frac{1}{k_i^2 + 1} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \frac{-b_i}{k_i^2 + 1} \\ \sum_{i=1}^{n} \frac{-k_i b_i}{k_i^2 + 1} \end{bmatrix}$$

(5)

The solution of the above equation can be viewed as the estimation of the zoom center. Considering the existence of mismatching points, we use RANSAC method [19] to obtain a robust estimation of the zoom center finally.

Fig. 1 shows our experiment result, where the black lines are the straight lines determined by the
matched points, and the solid circle is the estimated zoom center. The size of image is $320 \times 240$.

![Figure 1. Zoom center estimation result.](image)

(b) (c)

(a)

![Figure 2. PTZ camera calibration verification.](image)

B. Focal Length Estimation at Different Zoom Level

As shown in (2), the focal length is determined by the zoom parameter. So we first estimate focal length at a specified zoom value $Z$. Two images under the same zoom value but different pan and tilt parameters are captured. We assume the parameters of the two images are $(P,T,Z)$ and $(P,T',Z)$ respectively. Then we use SIFT feature points to establish correspondence between the two images. For each of the matched points, we have

$$
\begin{bmatrix}
    f & 0 & u_0 \\
    0 & f & v_0 \\
    0 & 0 & 1
\end{bmatrix} R(P,T) X_j
$$

and

$$
\begin{bmatrix}
    f & 0 & u_0 \\
    0 & f & v_0 \\
    0 & 0 & 1
\end{bmatrix} R(P',T') X_j
$$

As the extrinsic matrix $R$ can be calculated by (3), and the principal point $(u_0, v_0)$ has been estimated in previous content. So solving (6), we can obtain the estimation of focal length for each of matched points. In order to improve the robustness, we use all the matched points to calculate the mean value as the final result of focal length.

After focal length has been estimated at several discrete zoom levels, we choose a proper model to fit these samples. In our study, we use (7) for approximation. The four unknown parameters $p_1, p_2, p_3$ and $p_4$ can be solved by using curve fitting tools.

$$
f(Z) = p_1 e^{p_2 Z} + p_3 e^{p_4 Z}
$$

To verify the effectiveness of our calibration procedure, we capture two images from the PTZ camera shown in Fig. 2 (a) and (b) at different parameters. We use the camera calibration result to warp image (b) to image (a) with the same image coordinates. The result is shown in Fig. 2 (c). The result shows that the warped image fits the origin image well.

III. STEREO RECTIFICATION USING DUAL PTZ CAMERAS

In traditional dual static system, the two cameras have the same intrinsic parameters then the acquired images are known as a rectified pair of stereo images. In these images, corresponding points lie on a same horizontal scan lines. So stereo matching algorithm can be directly applied to calculate the disparity or depth information. In dual PTZ camera system, it’s hard to directly estimate disparity because that the two cameras may have different extrinsic and intrinsic parameters especially the focal length. Stereo rectification is a way to reduce the searching scope in stereo matching from 2D to 1D. Many stereo rectification methods have been proposed in the past years [20-24]. In this paper, we use spherical rectification method [23] to rectify images captured from PTZ cameras. The reason for choosing this method lies in that it can rectify images under arbitrary PTZ settings. Meanwhile, disparity obtained by this method could reflect the depth of the scene.

The main idea of spherical rectification method is using two spherical surfaces as the bridge to achieve fast and robust stereo rectification. It includes three steps: (1) estimate a rotation matrix and establish spherical coordinate system for each of the PTZ camera, the goal of this step is to make sure the longitude components of corresponding points be the same after coordinate conversion from the camera coordinate to spherical coordinate; (2) use PTZ camera model to map the image plane to the unit spherical surface according to current PTZ parameters; (3) the rectification is applied from the sphere to rectified plane. This method is independent to specific PTZ parameters, which is very convenient in application. More details can be found in [23]. Fig. 3 shows an example of spherical rectification for two images with quite different PTZ settings.

In our system, we use spherical rectification method to establish correspondence between two image coordinates. The conversion between origin image coordinate and rectified image coordinate are frequently used in our
system. So we summarize the conversion procedure as follows.

![Example of spherical rectification](image)

**Figure 3.** Example of spherical rectification: (a) and (b) are the origin images captured from the two PTZ cameras with different parameters; (c) is the rectification result. The PTZ parameters: (a) $PTZ_a = [92.31, -10.65, 5.00]$ and (b) $PTZ_b = [93.28, -12.62, 6.31]$.

### C. From Origin Image Coordinate $(x, y)$ to Rectified Image Coordinate $(x', y')$

Step1. Get camera coordinate $X$ by the camera model $X = \kappa R X K^{-1} X$, where $\kappa$ is a scale factor and $x = [x, y, 1]^T$. $R$ and $K$ are the extrinsic matrix and intrinsic matrix respectively, and can be obtained from the calibration result (see (2) and (3)).

Step2. Use (8) to transform $X$ to spherical coordinate and obtain longitude and latitude coordinate $(\alpha, \beta)$. Where $R$ is rotation matrix which can transform the camera coordinate to spherical coordinate. $X'(m)$ is the $m$th component of vector $X'$.

$$
\begin{align*}
X' &= R X \\
\alpha &= \arctan(X'(3), X'(2)) \\
\beta &= \arccos(X'(1))
\end{align*}
$$

Step3. Adjust $(\alpha, \beta)$ to $(\alpha, \gamma)$ by $\gamma = -\cot \beta$ so that the disparity can be preserved.

Step4. Calculate linear transformation parameters and transform $(\alpha, \gamma)$ to $(x, y)$.

### D. From Rectified Image Coordinate $(x', y')$ to Origin Image Coordinate $(x, y)$

Step1. Linear transform $(x, y)$ to $(\alpha, \gamma)$ by using the calculated transformation parameters.

Step2. Calculate the longitude and latitude coordinates $(\alpha, \beta)$ by $\beta = \tan^{-1}(-1/\gamma)$.

Step3. Get corresponding camera coordinate $X$ from $(\alpha, \beta)$ by (9). Where $R^{-1}$ is the inverse matrix of $R$.

$$
\begin{align*}
X'(1) &= \cos \beta \\
X'(2) &= \sin \beta \cos \alpha \\
X'(3) &= \sin \beta \sin \alpha \\
X &= R^{-1} X'
\end{align*}
$$

Step4. Get origin image coordinate $(x, y)$ from the camera coordinate $X$ by camera model (see (1)).

## IV. THE PROPOSED SYSTEM

### A. System Overview

For our system, the hardware configuration is composed of two PTZ cameras. The flow chart of our system is shown in Fig. 4. Before main procedure starts, an initialization procedure is needed. We first choose one camera as the static view, and the other camera will be served as the active camera. Then, in order to reduce the computational complexity, some parameters can be pre-calculated, i.e. camera calibration, stereo rectification parameters, etc. The main procedure is composed of one main thread and some other assistant threads. The main thread is a common visual tracking loop in the static view, including foreground extraction, objects detection and tracking. The other assistant threads include high resolution visual attention part, two cameras work on symmetrical mode, and we estimate the disparity to build the relationship between two camera image coordinates. Therefore, visual attention of multi-resolution can be obtained. In object tracking using depth information part, two cameras work on asymmetrical or master-salve module, and we use the depth information on the rectified image fair. So that it can be used for object tracking under the situation of occlusion.

![System flow chart](image)

**Figure 4.** System flow chart.
advantage of both traditional dual static and static-active system.

B. High Resolution Visual Attention

In high resolution visual attention part, the two cameras work on asymmetrical mode or master-salve mode. The key technique of high resolution visual attention is to calculate the new PTZ parameters of the slave camera according to the master’s PTZ setting and the selected region in master camera’s image. So the slave camera could cover that region with a higher resolution. Traditional methods mainly have two kinds of control strategies: sampling-interpolation method and mosaic image-based method. The first method constructs a mapping from every pixel coordinates in static image to PT parameters (such as [12]). This method has two drawbacks. First, it assumes that the baseline is too small comparing to the depth of the scene. So if the scene has great change in depth, such as indoor environment, this method may have large error. Second, if the master camera’s view changes, the pre-configurations should be re-done. The second method uses a mosaic image created by snapshots of slave camera to estimate the relationship between static master camera image plane and pan-tilt controls of slave camera (such as [15, 16]). This method relies on feature matches of master camera image and slave camera mosaic image, and when the scene appearance changes, the mosaic image should be updated.

To overcome the drawbacks of these methods, we proposed another approach which directly uses rectified image and disparity to calculate new PTZ parameters. Rectified image pair and disparity map is a bridge between the two images coordinates, so that two image coordinates can be mapped from one to another by rectification mapping added a disparity translation. We denote \( I_1 \) and \( I_2 \) as the master camera image and slave camera image respectively, \( (P_1, T_1, Z_1) \) and \( (P_2, T_2, Z_2) \) as the parameters of the master and slave camera. We denote \( \mathcal{R} \) as the selected region in master camera’s image.

The procedure of our method is listed as follows:

Step1. Use spherical rectification method to rectify \( I_1 \) and \( I_2 \). We denote the rectified images as \( I'_1 \) and \( I'_2 \).

Step2. Find the center \( c_1 \) of the region \( \mathcal{R} \) in image \( I'_1 \).

Then calculate \( c_1 \)'s corresponding location \( c_2 \) in rectified image \( I'_2 \) (see detailed procedure in section III).

Step3. Use graph cuts based stereo matching algorithm [25] to estimate the mean disparity \( d \) at \( c_1 \). Denote \( c_2 \) as the corresponding point in rectified image \( I'_2 \) of the slave camera image \( I_2 \), where \( c_2 = c_1 + [d, 0]^{T} \).

Step4. Calculate \( c_2 \) in \( I_2 \) from corresponding \( c_2 \) in \( I'_2 \) (see detailed procedure in section III).

Step5. Calculate new PTZ parameters \( (P_2, T_2, Z_2) \) given point \( c_2 \) and the origin PTZ parameters \( (P_1, T_1, Z_1) \) of the slave camera. This step can be divided into two parts, the PT parameters and \( Z \) parameter calculation. Let \( c_2 = (x_2, y_2) \), and the principal point is \((u_0, v_0)\). According to (7) and zoom value of slave camera \( Z_2 \), we can calculate the focal length \( f \). Then the absolute angles of pan and tilt parameters are

\[
\begin{cases}
\Delta P = \arctan \frac{x_2 - u_0}{f} \\
\Delta T = \arctan \frac{y_2 - v_0}{f}
\end{cases}
\]

Based on the position of point \( c_2 \), we use (11) to compute the new pan and tilt parameters, \( P_2 \) and \( T_2 \). Fig. 5 is the sketch map for pan and tilt parameters estimation of slave camera.

\[
\begin{cases}
P_2 = P_2 + \Delta P, T_2 = T_2 - \Delta T & \text{if} \quad x_2 \geq u_0 \text{ and } y_2 \geq v_0 \\
T_2 = T_2 + \Delta T & \text{if} \quad x_2 = u_0 \text{ and } y_2 < v_0 \\
T_2 = T_2 - \Delta T & \text{if} \quad x_2 < u_0 \text{ and } y_2 \geq v_0 \\
T_2 = T_2 + \Delta T & \text{if} \quad x_2 < u_0 \text{ and } y_2 < v_0
\end{cases}
\]

Figure 5. Sketch map for pan and tilt parameters estimation of slave camera.

The new zoom parameter \( Z_2 \) of slave camera can be determined by the size of the bounding box of the given region. Firstly, we use a lookup table to get a zoom level \( Z_0 \) according to the size of given region bounding box. Then we define the reliability of disparity \( Y_0 \) as a function of the variance of the local disparity in the target region,
which satisfies $Y_s \in [0,1]$. We use $Y_s$ as a weight, so the final zoom parameter $Z_f = Z(Z_n, R_s)$ (in our system, $Z_f = Z_n(0.7 + 0.3R_s)$). That means if the reliability of disparity is low, we lower down the zoom level. Otherwise, we give a higher zoom level.

If the target region is not visible in the salve camera image, we can’t estimate the disparity translation on the two rectified images. In this situation, we first estimate rough PTZ parameters by the maximum $D_{max}$ and minimum $D_{min}$ depth, see Fig. 6. Then after the salve camera moves to the rough PTZ position, we use procedure described above to estimate precise PTZ parameters.

C. Object Tracking Using Depth Information

Depth information is very useful in visual surveillance. And the depth information can improve object segmentation and tracking in case of multiple occluding objects. All work before use dual static cameras. In this paper, the depth information is introduced to our dual PTZ camera system. In this part, the two cameras work on symmetrical mode. As we know, when occlusion happens, part of the occluded object is invisible. So the occluding and occluded objects have different depth. Although the depth information estimated by vision approach is always less credible because of the uncertainty of stereo matching, in our system, this coarse depth information is adequate in that we only need to give a depth order of occluded objects.

In our system, we implement a tracking loop in static camera procedure which works well with no occlusion. Under occlusion situation, the tracking procedure will send a message to active camera. Then two cameras can calculate depth information by spherical rectification and stereo matching. In order to reduce computational work, we only estimate the depth information in foreground area. The prior knowledge of segmentation and tracking comes from the tracking history which can be obtained from the tracking database, see Fig. 4. The prior knowledge includes the number of objects in the occlusion area and previous location of each object. Then we classify each foreground pixel into objects according to location and depth information of each pixel. To improve the robustness of our system, we also use a post process of validation. The validation considers the proportion of segmentation region. That a too small value suggests an invalid segmentation.

We summarize the procedure as follows:

Step1. Use spherical rectification method to rectify images of the two cameras.

Step2. Estimate the disparity map in the foreground region by using graph cuts based stereo matching algorithm [25].

Step3. Collect prior knowledge from the tracking database, such as the objects number in occlusion area $n$ and the previous location of each object.

Step4. Classify all the foreground pixels into $n$ classes according its disparity and coordinates.

Step5. Calculate the property of each new class, such as area, centre and depth.

Step6. Validate the segmentation result. We consider the proportion of segmentation area, that a too small value suggests an invalid segmentation.

Step7. Update new class centre and depth into tracking database for valid segmentation result.

V. EXPERIMENTAL RESULTS

The validity of our system has been evaluated on an indoor environment. The system runs on one computer with Intel 3.0G CPU and 1.5G memory. Two SONY EVI D70 cameras were placed on the top window, at about 1.2 meters far from each other. Images from both cameras were taken at $320 \times 240$ pixels of resolution.

Figure 7. Results of high resolution visual attention. The top two images of (a) and (b) are the origin images of the master and salve camera respectively. The bottom images of (a) and (b) are salve camera images after moving to newly estimated values. (a) The selected object is visible in the salve camera image; (b) The selected object is invisible in the salve camera image.

Firstly, we present the experimental results of high resolution visual attention. The experiments tries to explain the robust result in two cases of salve camera’s state: the selected object is visible in the salve camera image and the selected object is invisible in the salve camera image. Fig. 7 shows the results under these two conditions. In the initialization, the target region is selected in master camera image highlighted by white box. The parameters
Figure 8. Experimental results of object tracking under occlusion. (a) Rectified image pair; (b) Disparity map of foreground region; (c) Tracking result. The parameters of the two cameras are \( PTZ_1 = [90.51, -14.55, 4.03] \) and \( PTZ_2 = [86.91, -13.73, 5.65] \) respectively.

of master camera are \( PTZ_1 = [90.44, -14.55, 4.03] \). In Fig. 7 (a), the initial parameters of the salve camera are \( PTZ_2 = [99.14, -15.68, 5.11] \), and the selected object is visible in the salve camera image. We calculate the new PTZ parameters of the salve camera using method illustrated in section IV. The estimated parameters of salve camera are \( PTZ_2' = [94.93, -11.18, 14.33] \). In Fig. 7 (b), the initial parameters of salve camera are \( PTZ_2 = [83.84, -10.88, 7.51] \), and the selected object is invisible in the salve camera’s image. So we first use the given rough depth range \([2, 15\, \text{mm}]\) to calculate the rough parameters of salve camera. After the camera moves to the new setting, we estimate the precise parameters finally. The estimated parameters of salve camera are \( PTZ_2' = [94.71, -11.48, 12.81] \).

When occlusion happens in the static camera view, we use two cameras to estimate depth information, and the depth is used in object tracking in our system. The designed algorithm is based on disparity map, and no appearance features are used. We take the previous location as the prediction for current frame, and the goal is to find the new center of each person in current frame. Fig. 8 gives four groups of results. For each group, the left two images are rectified images captured from two cameras at the same time stamp. The third image is the disparity map of the foreground region, the greater the intensity, the smaller the disparity, the greater the depth. The right image shows the tracking result. The two objects are presented by different intensity. The segmented region will be used to update the property of tracked object in tracking database. Note that the gray levels in different images are not comparable. This experiment is used to validate the depth information acquired from two PTZ cameras can be used in object tracking, especially when occlusion happens. And this application could handle those partly occlusion situation with not too small difference in depth.

VI. CONCLUSION

In this paper, we propose a novel binocular vision system for indoor scene visual surveillance. Our system can work on multimode: asymmetrical mode (master-salve mode) and symmetrical mode. In asymmetrical mode, we propose a rectification-disparity-based method to establish correspondence between two image coordinates. Then, both global and detailed image information can be obtained. In symmetrical mode, we use depth information to deal with occlusion problem in object tracking. The proposed system has the advantages of both dual-static system and static-active or dual active system. The experimental results show the merits of the proposed system. In our future work, we plan to combine multiple other features and depth information to deal with occlusion problem in object tracking.
REFERENCES


Zhiqiao Cui born in 1984, Hengshui, Hebei province, China. He is currently a doctor of Xi’an Institute of High-tech of China. His research interests include binocular vision surveillance and image processing.

Aihua Li was born in 1966. Professor and Ph.D. supervisor in Xi’an Institute of High-tech of China. His main research interests include pattern recognition, signal processing and computer vision.