Computer Vision and Image Understanding 112 (2008) 184–194

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Stereo vision using two PTZ cameras ☆,☆☆

Dingrui Wan, Jie Zhou *

Department of Automation, Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 23 July 2007
Accepted 23 February 2008
Available online 7 March 2008

Keywords:
Dual-PTZ-camera
Stereo vision
Stereo rectification
Stereo matching
Disparity
Depth

ABSTRACT

The research of traditional stereo vision is mainly based on static cameras. As PTZ (Pan–Tilt–Zoom) cameras are able to obtain multi-view-angle and multi-resolution information, they have received more and more concern in both research and real application. Stereo vision using dual-PTZ-camera system, compared with using dual-static-camera system, is much more challenging. Dual-PTZ-camera system could have more extensive scope of application by combining the merits of PTZ-camera. However, few works about stereo vision with dual-PTZ-camera system were found in literature. In this paper, we propose a novel stereo rectification method for dual-PTZ-camera system, which is essential to greatly increase the efficiency of stereo matching. In dual-PTZ-camera system, the inconsistence of intensities in two camera images, which is caused by camera's self-adjustment of intensity under different illumination condition with different view fields, is also a challenge in stereo matching. In order to deal with this problem, we propose a two-step based stereo matching strategy. Experimental results show that our approach works well.

1. Introduction

Stereo vision is one of the most important embranchments of computer vision. It can be used in 3D reconstruction, scene analysis and other depth related usage [1]. In state-of-the-art research, most vision based approaches for extracting stereo information can be mainly classified into two categories. The first category is to use monocular camera with known scene information; and the second is to use traditional stereo vision by dual-camera system, which always has two cameras as one equipment for the convenience in stereo rectification and matching. The latter one, which is also called 'disparity' based approach, is more intuitive and general, this is because it is similar to human eyes system, and it does not need much pre-known scene information. Our study belongs to the second category.

Traditional stereo vision research usually uses static cameras for their low cost and relative simpleness in modeling. Brown et al. [2] summarized the computational stereo in these systems and its real-time implementations. Pan–Tilt–Zoom camera (we use the acronym PTZ for short) is a typical and the simplest active camera, whose pose can be fully controlled by pan, tilt and zoom parameters. As PTZ cameras are able to obtain multi-view-angle and multi-resolution information (i.e. both global and local image information), They have received more and more concern in research. On the other hand, as these cameras become cheaper, they are already taken into many real applications. Stereo vision using dual-PTZ-camera system (see Fig. 1), compared with using dual-static-camera system, is much more challenging as the intrinsic and external parameters of each camera can be changed in utility. The research on stereo vision with dual-PTZ-camera system is *significative*, because: (1) since PTZ-camera possesses multi-view-angle properties, we could get a much wider field of observation, which could be very useful for panoramic scene analysis or 3D reconstruction; (2) as PTZ-camera possesses multi-resolution properties, the precision of depth can be increased by improve the image resolution. Such system could have more extensive scope of application by combining the merits mentioned above.

Stereo vision with dual-PTZ-camera system can be regarded as an active stereo vision system, which is able to adjust its visual parameters to aid task-oriented behavior [3]. As an active stereo system may have many degrees of freedom (DOF) (for example the system in [4] has tens of kinematic parameters), a general way of dealing the stereo vision problem with this system is to simplify the system into traditional dual-static stereo vision system by delicate calibration. So, many literature focused on camera calibration of active vision system [4–6]. In this paper, we only review some head-eye systems which are similar to the proposed dual-PTZ-camera system.

CeDAR, the Cable-Drive Active-Vision Robot [7,8], which incorporates a common tilt axis and two pan axes, is regarded as one of

☆ Part of this work has been published in ICASSP 2007.

☆☆ This work was supported by Natural Science Foundation of China under grant 60673106, 60721003 and 60573062, and the Specialized Research Fund for the Doctoral Program of Higher Education.

* Corresponding author.
E-mail addresses: wandingrui00@mails.tsinghua.edu.cn (D. Wan), jzhou@tsinghua.edu.cn (J. Zhou).

1077-3142/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2008.02.005

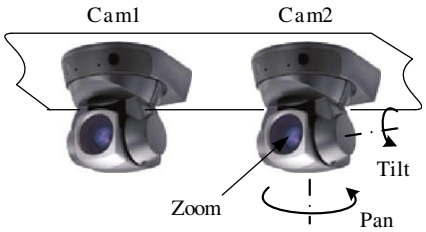

Fig. 1. The sketch map of dual-PTZ-camera system.

the simplest active stereo system. FOVEA, the FOveated VErgent Active Stereo System [9], is a similar system but using foveated cameras, which could obtain larger overlapped region between two views. There are also some other similar systems such as [10,11]. Compared with our proposed system, a common ground of these systems is that traditional stereo matching approach is in use. But none of them used a rectification model like the spherical rectification model proposed in this paper. Our rectification method is effective for non-translation dual active system, and it could preserve the inverse proportion between disparity and depth, which could facilitate the estimation of depth map directly from rectified image pairs.

In dual-PTZ-camera system, there exist at least two *difficulties*: (1) as the two PTZ cameras might have different orientations and zoom levels, the computation in direct stereo matching might be significant; and furthermore the robustness and accuracy can not be guaranteed; (2) because of camera's self-adjustment of intensity under different luminance condition (especially when zoom level changes or luminance in the field of view changes), the inconsistence in intensities of corresponding pixels in two images might become prominent. As this phenomenon hardly happens in symmetrical dual-camera system, very few traditional stereo matching algorithms concern about this problem. In our study, we will firstly propose a stereo rectification approach, and then, a luminance compensation is applied before stereo matching to standardize the problem so that many traditional stereo matching algorithms can be followed.

The paper is organized as follows: Section 2 describes the camera model and calibration; in Section 3, we propose a stereo rectification method which is called the spherical rectification; in Section 4, we describe the stereo matching approach used in our application; in Section 5, we give a simple analysis the precision of estimated depth and the relationship with disparity; experimental results are provided in Section 6; and in Section 7, we summarize this paper.

2. PTZ-camera model and calibration

Calibration of PTZ-camera is an important preparation for visual computing in our system. Ref. [12] reviewed some typical PTZ-camera calibration methods, including the ones using calibration targets and LEDs and requiring physical access to the cameras or the space in the field of view. Our method is similar to [12] in that they are both inherently feature-based. The difference is that we combine the parameters inquired from the PTZ-camera, which could simplify the problem a lot.

2.1. Camera model

The pin-hole camera model is used in our study, see Eq. (1). For simplicity, we do not consider either focal length variation or radial distortion.

$$\tilde{x} = \kappa K[R - Rt]\tilde{X}, \quad K = \begin{bmatrix} \alpha f & s & u_0 \\ & f & v_0 \\ & & 1 \end{bmatrix}, \tag{1}$$

where x and X are image coordinates and world coordinates, respectively; symbol '~' means homogeneous coordinates. α and s are the camera's pixel aspect ratio and skew, respectively; f is the equivalent focal length measured in pixel; (u_0, v_0) is the principal point in the image; κ is a scale factor. In order to simplify the PTZ-camera model, we make some assumptions as follows:

(1) The rotation axes of pan and tilt are orthogonal and intersect at one point, which is chosen as the origin of the world coordinate system, and hence we set $t = 0$;
(2) Aspect ratio $\alpha = 1$, and skew $s = 0$;
(3) Principal point (u_0, v_0) is replaced by the zoom center [13] approximately.

According to the assumptions, the camera model can be written as

$$\tilde{x} = \kappa KRX, \tag{2}$$

where the intrinsic matrix, K, could write in diagonal form by translating the origin of image to (u_0, v_0); and the rotation matrix, R, which is also the extrinsic matrix, which can be determined by the 'pan' and 'tilt' parameters.

Since the principal point is notoriously hard to compute accurately, and in contrast, the zoom center, which is also called 'center of expansion' [13], is easy to calculate from image data [14]. When camera changes zoom level, an image point will move radially along a line passing through an intersection point, and that is the zoom center. We use the zoom center to substitute the principal point (u_0, v_0) for two reasons. Firstly, the zoom center is relatively stable in many well developed commercial PTZ-cameras while zooming. Secondly, the estimation of zoom center is operable. This approximation has been used in many studies [12,15]. In our study, the SONY EVI D70 camera is in use, and experimental results show that this model works well.

2.2. Zoom center estimation

To determine the zoom center, firstly, we capture a successive image sequence from each PTZ-camera at varying discrete zoom level (from $z_{\min}$ to $z_{\max}$) with fixed pan and tilt parameters. Secondly, we detect and match feature points by using Harris corner detector, between neighboring images. For ideal case, the tracks of matched points should belong to a ray bundle converging at one point, i.e. the zoom center. Considering the existence of errors, we use the Least Squares Fitting method to obtain a robust estimation of the zoom center finally. In our experiments, the zoom centers of two PTZ cameras are (150.1,127.2) and (159.4,123.3), respectively, while the image size is 320×240.

In our experiment, we found that all above assumptions hold well. After translating the origin of images coordinate system to the zoom center, (u_0, v_0), the intrinsic K can be written in diagonal form as $diag\{k, k, 1\}$. The calibration of parameter k, which is depended on zoom value, will be discussed in the following.

2.3. R and K estimation

The rotation matrix, R, can be directly calculated from pan angle (θ_p) and tilt angle (θ_t) which can be inquired from the camera:

$$R = \begin{bmatrix} \cos(\theta_p) & 0 & \sin(\theta_p) \\ -\sin(\theta_p)\sin(\theta_t) & \cos(\theta_t) & \cos(\theta_p)\sin(\theta_t) \\ -\sin(\theta_p)\cos(\theta_t) & -\sin(\theta_t) & \cos(\theta_p)\cos(\theta_t) \end{bmatrix}. \tag{3}$$

This formula contain two assumptions. The first one is that the rotation axes of pan and tilt are orthogonal; and the second is the two axes intersect at one point.

As the intrinsic $K(z) = diag\{k(z), k(z), 1\}$ has only one degree of freedom which associated to the zoom value, we use the approach in [12] to estimate K at several discrete zoom levels, and then choose a proper model to fit these samples. In our study, we use the model in Eq. (4) for approximation.

$$k(z) = a\exp(bz) + c\exp(dz), \tag{4}$$

where the four unknown parameters a, b, c, d can be solved by using curve fitting tools. This model works well in our experiment.

Given a specified zoom value z, we capture two images with overlapped view under the same zoom settings z but different pan and tilt angles. As the pan and tilt parameters are known by acquiring from cameras, the relative rotation matrix $R_{12} = R_2R_1^{-1}$ can be calculated by Eq. (3). In order to calculate $k(z)$, we solve the following optimization problem as:

$$\min_{k(z)} \sum_{x\in E} \|I_1(x) - I_2(W(x, k(z)))\|, \tag{5}$$

where E is the overlapped region between the two images, and $I_i(x)$ is the gray level at pixel $x \in E$ in image $I_i, i = 1, 2$.

The coordinates conversion can be implemented by Eq. (6):

$$W(x, k(z)) = [v_1/v_3,\ v_2/v_3]^T, \tag{6}$$

where $[v_1,\ v_2,\ v_3]^T = K(z)R_{12}K(z)^{-1}\tilde{x}$, and $\tilde{x}$ is the image homogeneous coordinates at x after the zoom center translation.

In order to minimize the effect of parameters errors inquired from the camera (for example, the PTZ values), we take several experiments to calculate the mean value as the final result of $k(z)$.

2.4. Classical stereo model

In the ideal dual-camera stereo model, two cameras have the same focal length, and the optical axes are parallel and perpendicular to the baseline. See Fig. 2. We call this model the classical stereo model. In this model, the disparity, which is the displacement of a projected point in one image with respect to the other, can be calculated if the two projective points are both visible. Depth (Z) and disparity ($d = x - x'$) have the inverse proportion relation, i.e. $Z = b \cdot f/d$, where f is the focal length and b is the baseline width.

As the two optical axes are parallel, this model is also called the nonverged stereo model [2], but in practice, this model does not always hold well for real applications, especially when two cameras have different PTZ parameters. Image rectification is a way to make the verged geometry to be nonverged. In our study, we propose a so-called spherical rectification method to deal with this problem, and so that the relationship between two camera coordinate systems can be built.

3. Spherical rectification

Although stereo matching is an important technique of stereo vision, many traditional stereo matching algorithms can not be directly applied on image pairs. For ideal dual stereo vision, the system satisfies nonverged geometry [2], i.e. the epipolar lines are parallel to each other. However, this assumption does not hold well for many realistic systems. Stereo rectification is a method to make arbitrary stereo image pairs (i.e., with verged geometry) become nonverged geometry [2]. Most stereo rectification algorithms are based on epipolar constraint to map the epipolar lines to image scan lines and ensure the same scan lines in two images correspond to a specific epipolar line pair. Stereo rectification is very useful in that it makes the searching computation of stereo matching to be confined to one dimension, and, hence, make the problem simplified [16].

Many stereo rectification approaches have been proposed in the past years [17]. Most of them are homography based (also called planar rectification) [18,19]. Simplicity is a typical merit for this kind of approaches, while one of the shortcomings is that it does not preserve distance along epipolar line. Ref. [20,?] use more general warping functions to solve this problem. Roy et al. [20] proposed a cylinder rectification approach instead of the planar one, and Polleyfeys et al. [21] proposed a polar rectification method, which only required the fundamental matrix. These approaches could preserve distance along epipolar line but always be computationally expensive. In dual-PTZ-camera system, since the orientation and zoom level of the camera may change at any moment, the veracity, robustness and computation complexity are the key factors in rectification performance. To achieve stereo rectification in such system, we propose a spherical rectification method by using the longitude–latitude common coordinates. This rectification method could preserve the inverse proportion between disparity and depth, so that, disparity (depth) map can be directly and conventionally calculated from the rectified image pair.

3.1. Basic notation

Each pixel on the image plane can be mapped onto a unit spherical surface with the center coincide with the camera optical center, see Fig. 3. Actually, the mapped point is the intersection between the sphere and the line passing through the optic center and the given pixel on the image plane. We define some concepts as following:

3.1.1. Epipolar plane

$\Pi(O_1, O_2, P)$, where O_1 and O_2 are the two cameras' centers, and all epipolar planes pass the line O_1O_2. A given point outside the line O_1O_2 could determine an epipolar plane.

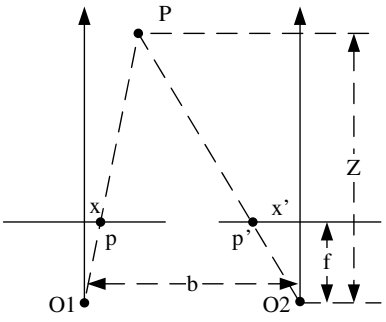

Fig. 2. The sketch map of classical stereo model.

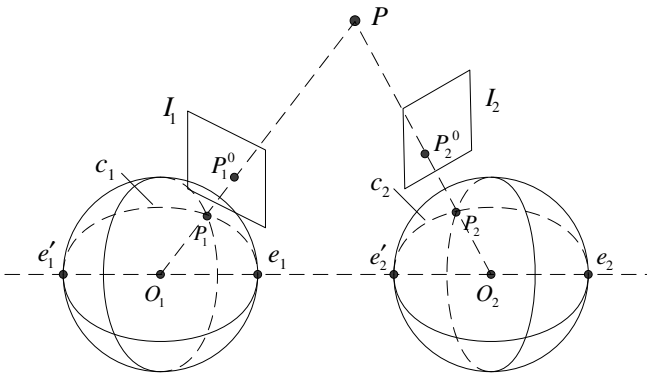

Fig. 3. The sketch map of spherical rectification.

3.1.2. Epipolar circle

c_1 and c_2, the intersection curves of epipolar plane $\Pi(O_1, O_2, P)$ and the unit spheres ϕ_1 and ϕ_2. A given point outside the line O_1O_2 could determine a pair of epipolar circles.

3.1.3. Epipole

The intersection point of the line O_1O_2 and the sphere. So there exit two epipoles, $e_i, e_i', i = 1, 2$, for each sphere. All epipolar circles pass through two epipoles on the same unit sphere.

The spherical rectification mainly constitutes two steps: first, using the camera model to map the image plane to the unit spherical surface. Second, warping the valid sphere region to rectified image. The sphere is independent of PTZ parameters, so when PTZ parameters change, we only need to find the corresponding region on the spherical surface, and use the pre-calculated rectification parameters to warp the sphere to the goal plane. As this method could avoid the recomputation of rectification parameters, it is convenient for PTZ image pairs' stereo rectification. Following the idea of preserving the distance along epipolar line in [20,21], our method could also possess this advantage so that traditional stereo matching algorithm can be followed to estimate the depth of the scene directly. The two steps will be depicted in the following:

3.2. Image plane to sphere

Let $x = [u\ v]^T$ be the image coordinates, and X_w be the corresponding world coordinates, which can be calculated by camera model. We can restrict $|X_w| = 1$, so that X_w locates on the unit spherical surface. Let $X = [\alpha_x\ \beta_x]^T$ be the corresponding sphere coordinates. α_x and β_x, which are also called the longitude and latitude component, respectively, can be determined if two poles (e and e') and zero longitude are known. In Fig. 4, α_x is the angle from $\Pi(e, e', X)$ to OM, where $\alpha_x \in [-\pi, \pi)$, and β_x is the angle from Oe to OX, where $\beta_x \in [0, \pi]$. OM is the reference vector, which is used to represent the zero longitude, and it satisfies $OM \perp ee'$.

3.3. Construction of spherical coordinates

As we mentioned before, in order to construct the spherical coordinates, we have to know the epipoles (e_1 and e_2) and the zero longitude (represented by a reference vector O_1M_1 and O_2M_2). Here, we use several fundamental matrixes to solve this problem:

(1) Capture several image pairs from two cameras, in which each pair has similar field of view.
(2) For image pair j, use RANSAC method to estimate fundamental matrix F_j and matched point pairs $\{x_k^{1j},\ x_k^{2j}\}$. Assume camera parameters are (K_{1j}, K_{2j}) and (R_{1j}, R_{2j}) for two cameras, respectively, and then the following two steps can be used to estimate epipoles and zero longitude, respectively.
(3) Assume e_{1j} and e_{2j} are the traditional epipoles, which should be the image coordinates of the goal epipoles e_1 and e_2. According to epipolar constraint, $F_j e_{1j} = 0,\ e_{2j}^T F_j = 0$. Let

$$\begin{cases} F_j^1 = F_j K_{1j} R_{1j}, \\ F_j^2 = R_{2j}^T K_{2j}^T F_j, \end{cases} \tag{7}$$

so that $F_j^1 e_1 = 0$ and $e_2^T F_j^2 = 0$. Let

$$\begin{cases} A_1 = [F_1^{1T}, F_2^{1T}, \cdots, F_n^{1T}]^T, \\ A_2 = [F_1^2, F_2^2, \cdots, F_n^2], \end{cases} \tag{8}$$

and we can use SVD method to estimate e_1 and e_2 by solving $A_1 e_1 = 0$ and $e_2^T A_2 = 0$.

(4) We first convert $\{x_k^{1j},\ x_k^{2j}\}$ into normalized world coordinates $\{X_k^{1j},\ X_k^{2j}\}$ by using the camera model. Note that it is not unique to choose zero longitudes. After e_i is estimated in (3), we arbitrarily choose M_1 and M_2 located at the middle of the two longitudes circles, and then convert $\{X_k^{1j},\ X_k^{2j}\}$ into longitude–latitude coordinates. What we concern is the longitude component pairs $\{\alpha_s^1,\ \alpha_s^2\}$, where $s = 1, 2, \cdots, N$(where N is the total number of pairs). Finally, rotate M_2 around epipolar axis by $\Delta\alpha$ which is the average offset between α^1 and α^2.

After the poles (e_i) and reference vector (O_iM_i) are estimated, the spherical coordinates can be established. The above procedure can be implemented off line.

3.4. Sphere to rectified image

There are several ways to warp the sphere to a plane, for example, we can use the equation as $X_r = [u'\ v']^T = [f_\alpha(\alpha_x)\ f_\beta(\beta_x)]^T$. However, not all the warping methods can preserve constant relationship between disparity and real depth. We propose a method which is called the (α, γ) rectification to achieve this goal.

Let P be a point in the scene, and the spherical coordinates of two cameras are (α_1, β_1) and (α_2, β_2), respectively (see Fig. 5). Obviously, (α_1, β_1) and (α_2, β_2) are located on the epipolar plane. Assume we have a proper reference vector, O_2M_2, so that $\alpha_1 = \alpha_2$. Let y_P be the distance between P and the line O_1O_2, and $O_1P = x_1$, $O_2P = x_2$, $O_1O_2 = b$, where b is the baseline width. Then we have

$$\begin{cases} x_1 \sin\beta_1 = x_2 \sin\beta_2 = y_P, \\ x_1 \cos\beta_1 - x_2 \cos\beta_2 = b. \end{cases} \tag{9}$$

So, y_P can be computed as

$$y_P = \frac{b}{\cot\beta_1 - \cot\beta_2} = \frac{b}{\gamma_2 - \gamma_1}, \tag{10}$$

where $\gamma_i = -\cot\beta_i, i = 1, 2$. $d_\gamma = \gamma_2 - \gamma_1$ is the disparity. Eq. (10) is similar to the classical expression which shows the inverse relation between disparity and depth. As $|\cot\beta| \to \infty$(*when* $\beta \to 0$ *or* π), the distortion of rectified image might be obvious, so this rectification

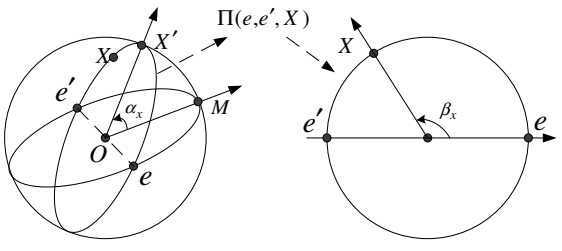

Fig. 4. The definition of α_x and β_x.

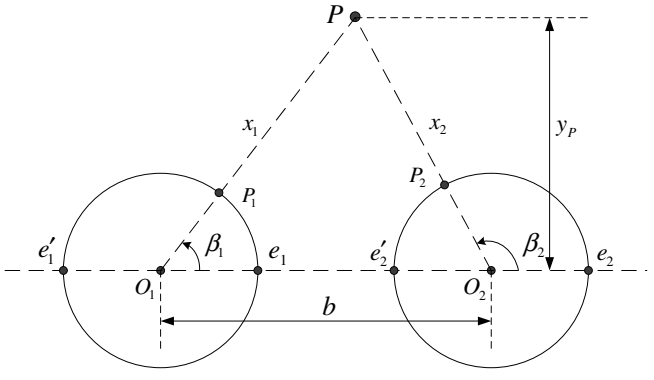

Fig. 5. The sketch map of depth in spherical rectification.

method could not deal with the case that one of the epipoles located in the image.

A linear transformation is used to map (α_x, γ_x) to rectified coordinates (u_r, v_r):

$$\begin{cases} u_r = k_\gamma(\gamma_x - \gamma_{min}), \\ v_r = k_\alpha(\alpha_x - \alpha_{min}). \end{cases} \tag{11}$$

Considering the valid region for α_x and γ_x, the following mapping parameters should be pre-calculated: $\alpha_{\min}$, $\gamma_{\min}$, $\Delta\alpha_u (= 1/k_\alpha)$ and $\Delta\gamma_u (= 1/k_\gamma)$. In order to reduce computation, we only check the four corners of the original image to calculate $\alpha_{\min}$, $\alpha_{\max}$, $\gamma_{\min}$ and $\gamma_{\max}$. For $\Delta\alpha_u$ and $\Delta\gamma_u$, we adopted the principle that to minimize the loss of pixel information [21]. Assume the point (α_x, γ_x) correspond to (u_r, v_r) in rectified image and (u_o, v_o) in original image, then, the displacement $[\Delta\alpha_u, 0]$ at (α_x, γ_x) indicates a vertical displacement $[0, 1]$ at (u_r, v_r) in rectified image and a displacement $[d_u, d_v]$ at (u_o, v_o) in original image, respectively. Minimizing the loss of pixel information indicates maximizing $\Delta\alpha_u$ under the constrain that $\|[d_u, d_v]\| \leqslant 1$. Similarly, as the displacement $[0, \Delta\gamma_u]$ indicates a horizontal displacement $[1, 0]$ in rectified image, $\Delta\gamma_u$ can be calculated in the same way. In order to save computation, we also only check the four corners to calculate $\Delta\alpha_u$ and $\Delta\gamma_u$.

For the two images, after $\alpha^i_{\min}$, $\alpha^i_{\max}$, $\gamma^i_{\min}$, $\gamma^i_{\max}$, $\Delta\alpha^i_u$ and $\Delta\gamma^i_u$ $(i = 1, 2)$ are obtained, we take the following steps to get rectification parameters:

(1) Make the two rectified images comparable. Set $\Delta\alpha_u = \max(\Delta\alpha^1_u, \Delta\alpha^2_u)$ and $\Delta\gamma_u = \max(\Delta\gamma^1_u, \Delta\gamma^2_u)$.
(2) Make the same scan line in two rectified images correspond to the same epipolar plane. Let $\alpha_{\min} = \min(\alpha^1_{\min}, \alpha^2_{\min})$, and then set $\alpha^1_{\min} = \alpha^2_{\min} = \alpha_{\min}$.
(3) Make the two rectified images have the same size if the following procedures require.

In Fig. 6, we summarize the conversion between original image coordinates and rectified image coordinates.

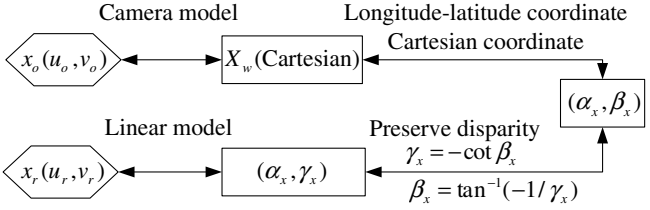

Fig. 6. the conversion between original image coordinates (u_o, v_o) and rectified image coordinates (u_r, v_r).

4. Stereo matching

Scharstein et al. [22] gave an intensive review about stereo matching algorithms. Generally speaking, there are mainly two subclasses in traditional stereo matching approaches, i.e. the feature-based and the 'direct' (i.e. pixel based) method. In traditional thought, after the spherical rectification, traditional stereo matching approaches can be used to achieve stereo vision. But in dual-PTZ-camera system, these might not work well because of the intensity inconsistence problem, which means the intensities in two rectified images are incomparable. This phenomenon hardly happens in symmetric binocular vision system, because the view fields of two cameras are nearly the same, and the characters of two cameras are almost the same. But in dual-PTZ-camera system, as the intensity (or color) in the image captured from cameras will be automatically adjust by the global luminance and contrast within the visual scene for difference field of view, especially when the two fields of view are different, which may increase the difficulty in stereo matching (see Fig. 7). In this paper, we combine the feature-based and pixel-based methods into a two-step approach to deal with stereo matching problem. Note that although some criterions such as luminance-invariant feature could avoid this problem, we intend to propose a general framework so that many traditional stereo matching algorithms with different criterions can be applied, That is why we put through a luminance compensation procedure.

4.1. Step 1

Finding the global intensity mapping between two images. In our application, we simply choose the linear model. In order to estimate the intensity mapping between rectified images, we first find reliable matched feature point pairs, and we intuitively assume that the intensities (gray levels) in the matched points' neighborhood are associated to the same object. Then, we build the correspondence between the sampled gray levels in the neighborhood of matched point pairs, i.e. $\{g^i_1\}$ and $\{g^i_2\}$. Finally, we use a linear model to fit this correspondence. Note that, this procedure is independent so that any more suitable luminance compensation model can replace the one we use. But in our system, the experimental results show that the linear model works well for the cameras in use.

As far as we know, SIFT is a good method to extract feature points with representative feature descriptor [23]. The merit that the feature descriptor is barely sensitive to global illuminance could be very helpful for points matching in our case. As the two images are rectified and they are nearly in the same scale

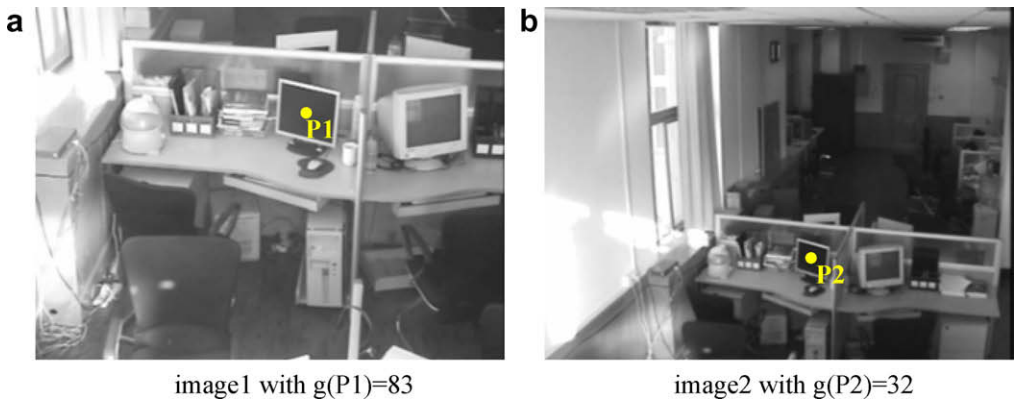

Fig. 7. An example for the discrepancy in intensity (gray level). (a) The image from camera 1, with P1's gray level is 83; (b) the image from camera 2 at the same time stamp, with P2's gray level is 32, which is the corresponding point of P1 in (a).

without rotation, we do not fully adopt SIFT method. Instead of that, we first choose Harris corners [24] as feature points because it can be fast computed, and then we imitate SIFT's idea to generate gradient feature descriptor in a size-fixed neighborhood: for each corner point, we segment the its neighborhood into 4 blocks (i.e. the left up, the right up, the left bottom and the right bottom), and for each block, we use an 8×1 dimensioned vector to express the gradient direction of the block. We match the corner pairs between the two images by comparing the $8 \times 4 = 32$ dimensional feature vector for each corner. One of the matching result is shown in Fig. 8(a). As the gradient direction feature has only relationship with the relative magnitude of the intensity, this approach could avoid the inconsistence of intensity between the two images affecting the stability of point feature. For each matched corner point pair, we calculate the median of intensities for the four blocks as samples. The linear model used to fit the relation between corresponding samples is shown in Fig. 8(b).

4.2. Step 2

After intensity adjustment by the estimated mapping model in step 1, the stereo matching problem has been converted to traditional classic one. So any existing well-performance matching algorithms can be applied in our case. Considering the computational cost, we use the 'direct' method (SSD criterion) with region-based aggregating.

The ordinary SSD method might induce many false matching, so the post-processing is necessary. As the ideal of region-based stereo matching has been inducted in many studies [25–27], we choose the region-based aggregation approach, so that many outliers can be eliminated. Without losing generality, we regard the image from camera 1 be the reference image, and the one from camera 2 be the compared image. The regions are segmented by intensity in reference image. If the variance of the primal calculated disparities in some region is too large, we set the disparities in this region zero, which indicates invalid.

Different matching algorithms have their own applicabilities for different situations. If more accurate matching result is needed, a global (iterative) optimization or hierarchical (coarse-to-fine) matching ideas can be considered into the second step. More information can be found in Ref. [22], which gives some detailed comparisons among typical stereo matching algorithms.

In traditional stereo matching approach, it always needs the maximum and minimum disparities to restrict the searching range (disparity and depth are in inverse proportion). In our studies, we compute the maximum and minimum disparities for each column (i.e. $d_{\max}(u_1)$ and $d_{\min}(u_1)$, where u_1 is the column index) of the rectified reference image by giving a rough depth range of the scene (i.e. $D_{\max}$ and $D_{\min}$). Compared with setting global maximum and minimum disparities, this method could give a smaller searching range, so our disposal could be more efficient.

For each pixel (u_1, v_1) in the reference image, which is located on a ray though O_1 with angle β_1, where β_1 can be obtained from u_1 by rectification coordinates conversion (see Fig. 9). P_1 and P_2 are the points on this ray at minimum and maximum depth, respectively, and the corresponding angles in coordinate system of camera 2, β_2^1 and β_2^2, can be obtained by using trigonometry. Analogously, the rectified image coordinates u_2^1 and u_2^2 can be calculated from β_2^1 and β_2^2 by the rectification coordinates conversion. Then, the maximum and minimum disparities are

$$\begin{cases} d_{\max}(u_1) = u_2^2 - u_1, \\ d_{\min}(u_1) = u_2^1 - u_1. \end{cases} \tag{12}$$

Fig. 8(c) shows the estimated disparity map of the image pair in Fig. 7. The gray level in disparity map shows the relative magnitude of disparities respect to a certain minimum disparity. If the gray level is 0, it means that the disparity is unreliable. In order

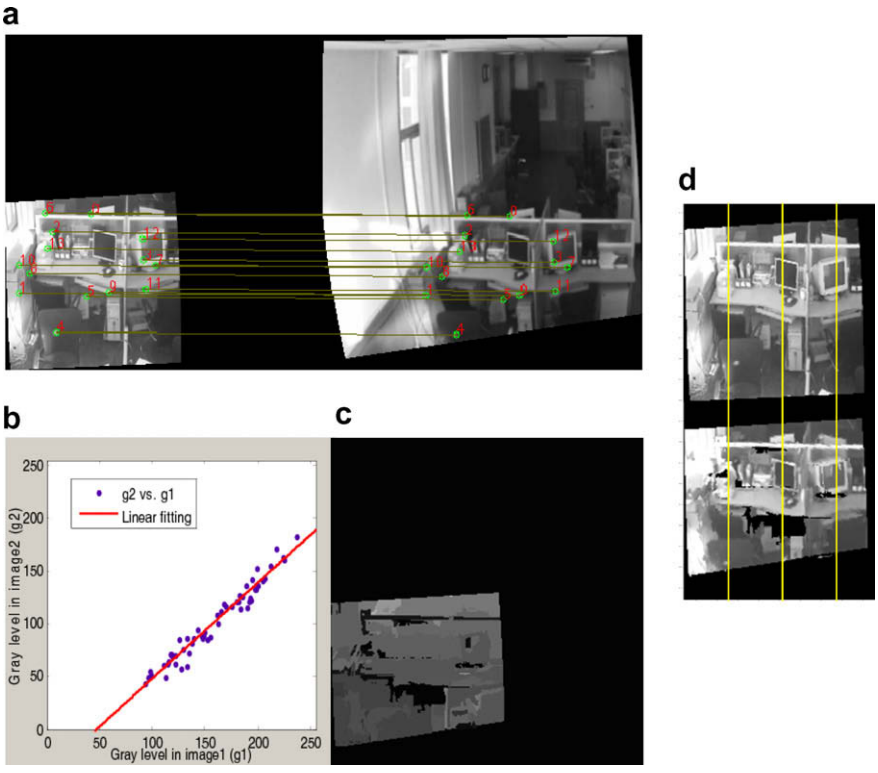

Fig. 8. Intensity (gray level) mapping for the images in Fig. 7. (a) Feature points matching result; (b) Linear fitting of the corresponding gray level pairs; (c) estimated disparity map, where the intensity indicates the relative magnitude of disparity, and 0 intensity indicates uncertain disparity; (d) a qualitative comparison between reference image (upper one) and the virtual image (lower one) which is generated by the compared image and the estimated disparity map.

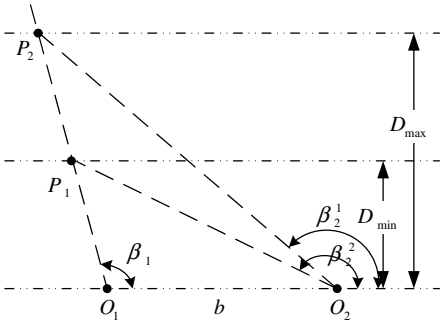

Fig. 9. Sketch map for maximum and minimum disparities.

to show the validity of the estimated disparity map, a virtual image is reconstructed by the compared image and the estimated disparity map. See Fig. 8(d).

5. Depth calculation from disparity

In this section, we will describe the relationship between disparity and depth. As the precision of disparity is in pixel level, it has to be aware of the uncertainty of depth caused by the precision of disparity and the distinguishability of depth, i.e. the minimal difference in depth that can be distinguished by disparity.

In Eq. (10), we know that the real depth D and disparity are in inverse proportion, i.e. $D = \frac{b}{\gamma_2 - \gamma_1}$, where b is baseline width, and we call $d_\gamma = \gamma_2 - \gamma_1$ as the real disparity. Denote $d_u = u_2 - u_1$ as the image disparity, we have

$$d_\gamma = d_u \cdot \Delta\gamma_u + \gamma^2_{\min} - \gamma^1_{\min}, \tag{13}$$

where $\Delta\gamma_u$ and $\gamma^i_{\min}$, $(i = 1, 2)$ are the rectification parameters (see Eq. (11)) that already calculated in stereo rectification. $\Delta\gamma_u$ is directly related to the resolution of the reference image or the zoom level of the camera corresponding to the reference image. When pan and tilt values are fixed, if the zoom level becomes smaller, the resolution will be lower and $\Delta\gamma_u$ will be greater.

5.1. The uncertainty of real depth

Denote the uncertainty of d_γ, d_u and $\gamma^i_{\min}$ as $\varepsilon(d_\gamma), \varepsilon(d_u)$, and $\varepsilon(\gamma^i_{\min})$, respectively. According to Taylor expansion, the uncertainty of real depth $\varepsilon(D)$ satisfies

$$\varepsilon(D) = \frac{D^2}{b}\varepsilon(d_\gamma) + o(\varepsilon(d_\gamma)), \tag{14}$$

where D is the real depth value. According to Eq. (13), $\varepsilon(\gamma^i_{min}) < \Delta\gamma_u$ $(i = 1, 2)$, and $\varepsilon(d_u) < 1$, so $\varepsilon(d_\gamma) < 3\Delta\gamma_u$. Then the upper limit of $\varepsilon(D)$ can be estimated:

$$\varepsilon(D) < \frac{3D^2}{b}\Delta\gamma_u. \tag{15}$$

Eq. (15) shows that if D is given, the uncertainty $\varepsilon(D)$ is in direct proportion to $\Delta\gamma_u$. For example, if we use two image pairs to estimate the depth of some given object, the pair with smaller $\Delta\gamma_u$, which always indicates a higher resolution, will have a smaller uncertainty of estimated depth. On the other hand, if $\Delta\gamma_u$ is given, the uncertainty $\varepsilon(D)$ is proportional to D^2. This case could happen when we estimate two objects' depth in the same image pair, $\Delta\gamma_u$ is the same for two objects. Then, the object with smaller D, will have smaller uncertainty.

5.2. The distinguishability of depths

If two objects in two images are both visible, we consider the minimal depth discrepancy so that we can possibly distinguish two objects by using the calculated disparities. In other words, let D_0 be the known depth, we are going to find the minimum δD which corresponds to more than 1 pixel in stereo image pair. We call δD the distinguishability of D_0 by using stereo vision approach.

From Eq. (13), as $\gamma^1_{\min}$ and $\gamma^2_{\min}$ are constant in given image pair and the minimal distinguishable disparity $\delta(d_u) = 1$ (in pixel), we have $\delta d_\gamma = \Delta\gamma_u$, and from Eq. (15), we can get

$$\delta D \doteq \frac{D_0^2}{b}\Delta\gamma_u. \tag{16}$$

For example, if $b = 0.75$ m and let $D_0 = 15$ m, the distinguishabilities of depth, δD, under different zoom settings of the camera corresponding to the reference image are as in Table 1. From the result, we can conclude that the greater D_0, the greater δD, and the weaker the distinguishability of depth. On the other hand, the smaller the zoom level, the lower the resolution, and the weaker the distinguishability of depth.

This conclusion can be used in 3D scene analysis or 3D reconstruction. For some given targets in the field of view, if we know the rough depth, D_0, and we want the minimal distinguishable depth to be δD, we have to choose a proper resolution so that these targets can be distinguished by disparity with stereo vision approach. According to Eq. (16), the demanded upper limit of $\Delta\gamma_u$ is $\frac{D_0^2}{b\delta D}$. In order to choose minimal zoom level, we utilize the relationship between zoom level $z_i (i = 1, 2)$ and $\Delta\gamma_{ui}$. As $\frac{k(z_1)}{k(z_2)} = \frac{\Delta\gamma_{u1}}{\Delta\gamma_{u2}}$, where $k(z_i)$ $(i = 1, 2)$ is the equivalent focal length in camera model, so the zoom level can be chosen easily by Eq. (4).

6. Experimental results

In our experiment, we use two SONY EVI D70 cameras to compose the dual-PTZ-camera system, which is running on one computer with Intel 3.0G CPU and 1G memory. The size of captured images is 320×240. The two cameras are fixed on the top window frame with baseline $b = 0.75$ m, so the indoor and outdoor scene are both visible by changing their pan and tilt parameters.

6.1. Stereo rectification

The spherical coordinates are established off-line. For two images $I_1(PTZ_1)$ and $I_2(PTZ_2)$, let I_1^r and I_2^r be the rectified images. Firstly, we traverse all points $S_p(i,j) \in I_1^r$, and follow a series of coordinate transformation summarized in Fig. 6, we get the sphere coordinates (α_p, β_p), the camera coordinates X_p and the original image coordinates $x_p(u_p, v_p)$. Secondly, we choose a proper interpolation method (the bilinear is used in our system) to estimate the gray level at $x_p(u_p, v_p)$. This gray level is assigned to (i,j) in I_1^r. Thirdly, we perform the same procedure for I_2^r.

Two sets of results are shown in Fig. 10. Two rectified images are normalized to the same size. Although the PTZ parameters are different between (a) and (b), the rectification coordinates do not need to be recalculated, while only rectification parameters do. All the experiments in this paper are under the same rectification coordinates. From the results, we can see that as $\gamma = -\cot\beta$, if

Table 1
The distinguishability of depth under different zoom settings of the camera corresponding to the reference image

Zoom level	$\Delta\gamma_u$	D_0 (m)	δD (m)
0	0.0033	15	0.99
		5	0.11
5.5	0.0016	15	0.48
		5	0.053
9	0.0009	15	0.28
		5	0.031

Fig. 10. Two spherical rectification experimental results. The left two images are the original images captured from the two PTZ cameras. The right two are the rectified images, whose sizes are normalized to that of the original image. The PTZ parameters: (a) $PTZ_1 = [96.59\ -12.45\ 5.27], PTZ_2 = [93.96\ -18.08\ 1.34]$ and (b) $PTZ_1 = [127.48\ -16.80\ 2.35], PTZ_2 = [123.58\ -16.80\ 2.35]$.

β is close to π or 0, the difference of densities along horizontal axis becomes larger, and the rectified images become more contorted, like the ones in Fig. 10(b).

In our experiments, the uncertainties of PTZ parameters acquired from the camera might have small effect on the precision of stereo rectification. According to our experience, the lower the zoom level, the smaller the error. For the maximal zoom level, the error in vertical pixel is always smaller than 10 (pixels) with image size 320×240. If there is a feature point matching procedure followed, we could use another way to minimize this error: the matched feature points could be used to estimate the global vertical translation, and this translation will be used for error compensation. This adjustment is adopted in our stereo matching experiments.

The computational complexity of spherical rectification is about $O(WH)$, where W and H are the width and height of rectified image, respectively. We decompose this procedure into three parts: Firstly, calculate the rectification parameters by using PTZ values and examine the four corners of the original image, of which the computational complexity is $O(1)$. Secondly, generate the $\sin(\cdot)$ and $\cos(\cdot)$ tables for all rows (i.e. the discrete αs) and columns (i.e. the discrete βs) of the goal rectified image, of which the computational complexity is $O(W) + O(H)$. Thirdly, fill in the rectified image pixel by pixel, of which the computational complexity is $O(WH)$. So the rectification procedure could be executed very fast.

6.2. Stereo matching

We list five sets of stereo vision experimental results in this section. The first three are shown in Fig. 11, which is meant to explain that our proposed method can be used in dual-PTZ-camera system with different view-angles (for both indoor and outdoor scene) and different zoom levels. The next two are shown in Fig. 12, which is meant to explain the stability of depth estimation of the same scene with different image pair and the relationship between zoom level and resolution of disparities. The PTZ parameters and some intermediate results of these five experiments are shown in Table 3.

For each experiment in Fig. 11, the left two images in the first column are original images captured from the two PTZ cameras. We firstly apply the spherical rectification on the two original image pairs of two cameras with arbitrary PTZ settings. Secondly, we extract corner points on each rectified image and match them (see the right two images in the first row in Fig. 11). Thirdly, we estimate the intensity mapping between two rectified images with linear model (see the second image in the second row in Fig. 11). Finally, we calculate the disparity map between two rectified images with intensity adjustment by the estimated mapping model, and convert the disparity map from rectified image coordinates to original image coordinates (see the third image in the second row in Fig. 11). The disparity is represented by intensity. The intensity indicates the relative magnitude of disparity with respect to a referenced minimum disparity. Intensity 0 indicates the disparity is uncertain. For Fig. 11(a) and (b), the greater the intensity, the smaller the disparity, the greater the depth. For Fig. 11(c), the magnitude of intensity is opposite because for indoor scene, the reference camera is on the left; but for outdoor scene, as there is a nearly 180° changing in 'pan', the reference camera turns to be on the right, and consequently, the disparity reverses its sign. However, this does not affect our algorithm. The only difference in pre-settings for these two conditions is the maximum and minimum depth. For indoor scene, we set $D_{\max} = 20$ and $D_{\min} = 3$, while for outdoor scene, $D_{\max} = 2000$ and $D_{\min} = 20$ (m). Note that, a rough and large enough bound of depth would not affect stereo matching theoretically, however, the search range could be larger, and the possibility of being trapped in local minimum grows.

It should be noticed that, precise disparity map estimation is still an intractable problem, and there is no approach working well in any situations. Although the region-based SSD stereo matching approach has lower computation with respect to many other approaches, it has its intrinsic shortcomings. Region-based aggregation will likely fail when some object has an uniform intensity but different depth, such as walls and ground plane which are viewed from the side. Furthermore, some cases are difficult for almost all the stereo matching methods. For example, the sky in Fig. 11(c) and the floor in Fig. 12.

We also designed some experiments to illuminate: (1) The stability of depth estimation from different image pairs especially with different zoom values; (2) the precision of estimated depth, which revealed by the relationship between zoom level and resolution of disparities. All the experimental results have supported our conclusion, but for space limitation, we only choose one example shown in Fig. 12.

Fig. 12 shows two experimental results that we estimate the depth map of the same scene with different image pairs. For Fig. 12(a) and (b), the left two images are original images, and the third image is disparity map estimated by the same stereo matching algorithm within rectified image coordinates. The intensity indicates the relative magnitude of disparity with respect to a referenced minimum disparity. The last image is the depth map

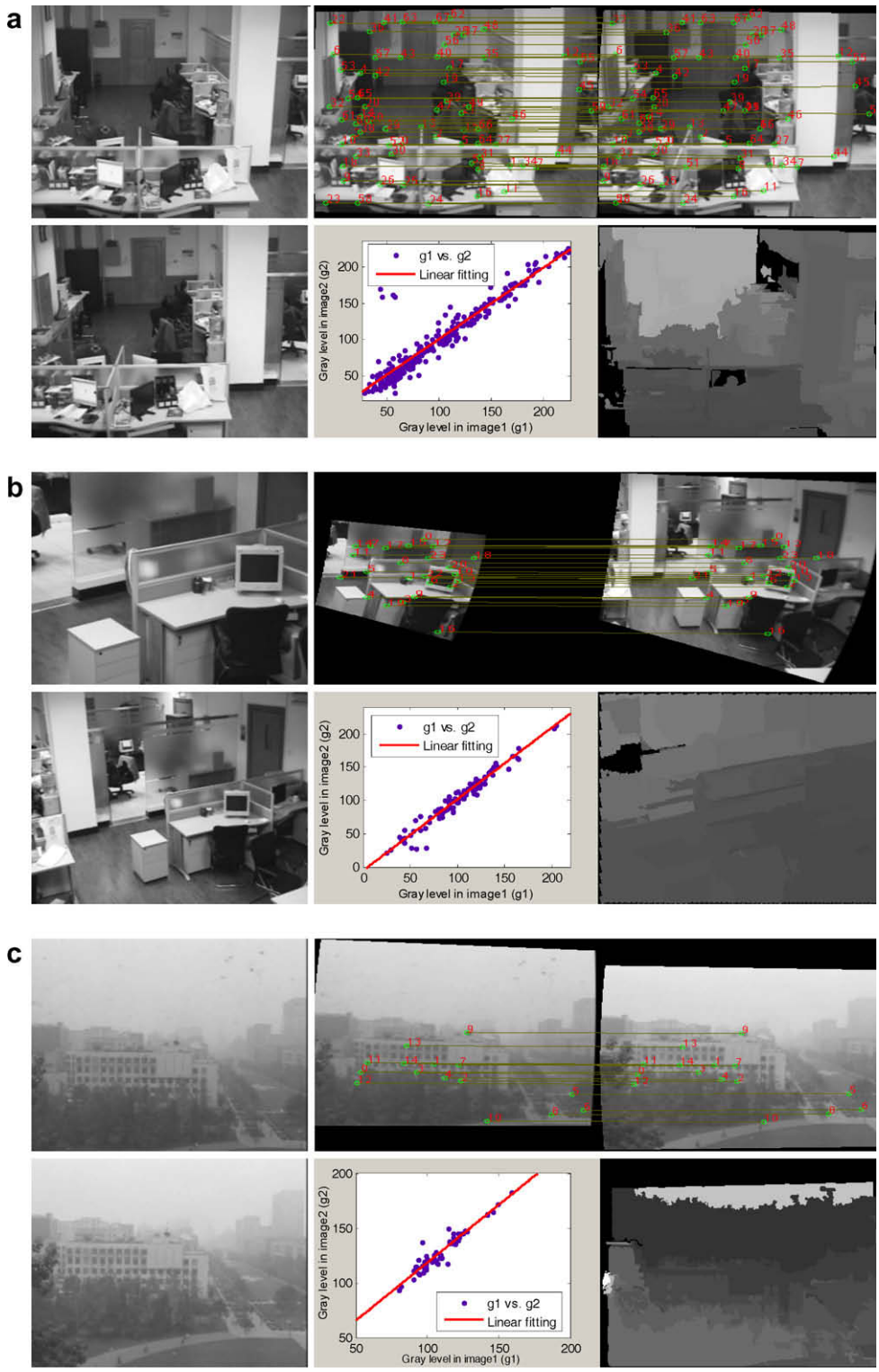

Fig. 11. Three experiments: the proposed stereo vision approach can be used for different PTZ settings (see Table 3). (a) and (b) are indoor, and (c) is outdoor.

within original image coordinates, while the magnitude of intensity has a linear relationship with depth. Zero intensity in both two images also indicates unreliable value.

In order to test the veracity of estimated disparity, we sample two points (A and B) in the scene. In Fig. 12(a), the disparities at A and B are 47 and 15 (groundtruth is 49 and 14), respectively; in Fig. 12(b), the disparities are 53 and 2 (groundtruth is 54 and 2), respectively. We deemed that the groundtruth of disparities can be obtained by manual stereo matching with pixel precision. The error of estimated disparity with respect to groundtruth can be resulted from false matching, aggregation, smooth operation, etc. In Table 2, we compared the disparity-based depth estimation in Fig. 12(a) with that in Fig. 12(b), which represent low zoom level and high zoom level, respectively. The groundtruth disparities will be used to calculate the depth's theoretical minimal uncertainty, $\varepsilon(D)$, which can be obtained by Eq. (15). $\varepsilon(D)$ is based on the

Fig. 12. Two experiments: depth map estimation of the same scene with different image pairs. The reference images of two experiments are identical, but the compared image of (a) has a lower zoom level than that of (b) (The specific parameters are in Table 3). A and B in images are testing points.

Table 2
Depth estimation at A and B in Fig. 12(a) and (b)

		A	B
Groundtruth D_0 (m)		10.05	4.95
Fig. 12(a)	D (ideal matching)	10.48	5.08
	D' (our method)	9.88	5.16
	$\Delta\gamma_u$	2.255×10^{-3}	
	ε (D)	0.99	0.23
Fig. 12(b)	D (ideal matching)	10.26	5.04
	D' (our method)	10.06	5.04
	$\Delta\gamma_u$	1.512×10^{-3}	
	ε (D)	0.64	0.15

Table 3
The parameters and some intermediate results are shown in Figs. 11 and 12

	PTZ_1	PTZ_2	Intensity mapping
Fig. 11(a)	[94.49, −16.80, 2.35]	[90.44, −16.80, 2.35]	$g2 = 0.99g1 + 0.95$
Fig. 11(b)	[124.26, −20.33, 6.57]	[116.08, −19.20, 1.24]	$g2 = 1.07g1 - 5.70$
Fig. 11(c)	[−94.71, −0.08, 3.00]	[−91.79, −1.13, 3.00]	$g2 = 1.06g1 + 12.55$
Fig. 12(a)	[81.51, −18.45, 6.62]	[79.71, −17.03, 2.56]	$g2 = 1.03g1 - 0.54$
Fig. 12(b)	[81.51, −18.45, 6.62]	[75.06, −16.93, 6.00]	$g2 = 1.02g1 - 3.56$

assumption that the disparity is exact in pixel, and it reveals the error in depth at location with depth D, caused by 1 pixel error in disparity.

The result in Table 2 shows: (1) although the compared images (the second image) in Fig. 12(a) and (b) have different zoom level, the estimated depths at A and B are both close to the groundtruth. This phenomenon might not prove that our matching approach can provide precise estimation of depth at any place, but it could sustain the stability of depth estimation of the same scene with different image pairs. (2) Fig. 12(b) has smaller uncertainty $\varepsilon(D)$ than that of Fig. 12(a) does. As there is no matching algorithm can guarantee precision of estimated disparity in pixel, we use the groundtruth disparities at specified location, so that we can compare the depth's uncertainties of two cases fairly. As we discussed in Section 5, depth's uncertainty is directly related to resolution of disparity, which is decided by the rectification parameter $\Delta\gamma_u$. Actually, in spherical rectification, the rule that minimizing loss of pixel information, makes the lower zoom level of the two original images have the primary effect to decide $\Delta\gamma_u$. Consequently, Fig. 12(b) has smaller $\Delta\gamma_u$, and higher resolution of disparity, and smaller uncertainty of depth.

6.3. Discussion

By observing a lot of experimental results, we summarize following three factors that might affect the performance of stereo vision result:

(1) *The veracity of stereo rectification*. The validity of proposed spherical rectification method is based on following assumptions: the PTZ values provided by camera are accurate enough, the pinhole camera model fits real model well, and the error in calibration of spherical rectification is acceptable. Here we only discuss the errors in acquired PTZ values (also called the uncertainty) caused by mechanical clearance. These errors are uncertain, so we only consider the worst case. As the uncertainties of pan and tilt are independent of zoom value, the worst case happens when zoom value reaches Z_{max}. According to our experiments, the uncertainties of pan and tilt are about 0.1° for SONY EVI D70 camera. For $Z < 2$, the equivalent error in pixel is about 1 pixel; and for $Z < 12$, it is less than 5 pixels (while image size is 320×240). As the error in depth estimation caused by these errors is also related to the 3D location of specific object, we do not quantitatively analyze this effect. In real application, this phenomenon could affect rectification result a little. In our system, a global vertical translation is calculated by the matched points in two rectified images for compensation, if the average error in y-direction is larger than 3 pixels. This trick is hardly used for small Zoom values. Although these PTZ uncertainties are unavoidable, by choosing better cameras with more accurate PTZ parameters, or re-estimating the PTZ values from image content might achieve smaller errors. This consideration is beyond the scope of the study in this paper.

(2) *The intrinsic limitation of the proposed spherical rectification*. As we analyzed before, when β is close to π or 0 (i.e. the orientations of cameras are closer to the baseline), the rectified images become more contorted, and the resolution of rectified images can be degraded for those contorted region because of the criterion that minimizing the loss of pixel information [21] (for example, Fig. 10(b)). So the precision of disparity estimation will be restricted. Actually, in this case, $\Delta\gamma_u$ is always larger than that in those situations with β close to 0.5π under the same zoom level. Some experimental results are listed in Table 4 (note that the exact value of $\Delta\gamma_u$ is related to specific PTZ parameters of both two cameras, so $\Delta\gamma_u$ values in Table 4 are only used for qualitative demonstration):

According to the analysis about depth uncertainty depicted in Section 5, we have similar conclusion, i.e. the greater $\Delta\gamma_u$, the greater the uncertainty of depth. On the other hand, intuitively thinking, as β is closer to π or 0, the equivalent baseline length be-

Table 4
The comparison of Δ_{γ_u} values under different β and zoom level (Z) settings

	$\beta \approx 90^\circ$	$\beta \approx 60^\circ$	$\beta \approx 30^\circ$
$Z = 0$	0.0027	0.0063	0.1306
$Z = 5$	0.0015	0.0027	0.0159
$Z = 10$	0.0007	0.0011	0.0041

We make two cameras have the same zoom level and similar FOV, and β is defined as $(\beta_1 + \beta_2)/2$, where β_1 and β_2 are the β-component of spherical coordinates at two image centers, respectively.

comes smaller, then the ability of depth perceptivity becomes weaker, just like the vision system of human beings.

(3) *The performance of stereo matching algorithm*. The goal of this study is to proposed an integrated stereo vision framework with dual-PTZ-camera system, so we chose a simple local SSD stereo matching method to validate this framework. As this module is independent of previous rectification procedures, if the computation is accepted, the performance of depth estimation can be improved by adopting some other techniques, such as belief propagation, dynamic programming, graph cut, etc.

In this section, we provide several experimental results to testify the proposed spherical rectification method and the two-step (feature-based intensity mapping estimation and region-based SSD matching) stereo matching framework. From the results, we can conclude that, (1) as depth map can be estimated under different PTZ settings of two cameras, so that a wide view of depth information can be obtained by changing PTZ parameters; (2) depth precision can be improved by raising the zoom levels of image pair.

7. Conclusion

PTZ cameras have been used widely in visual surveillance application, but the research with dual-PTZ-camera system is very few in literature. Stereo vision is an important brunch in computer vision, however, as far as we know, no research has been found to apply stereo vision in such system. The study in this paper might be one of the first attempts in this area. We aim to extend the application occasions of stereo vision and to discover the useful and realizable functions of dual-PTZ-camera system.

In this paper, we discussed stereo vision in dual-PTZ-camera system. Multi-view-angle property and multi-resolution property have been brought into stereo vision, so that we can get depth information in a varied and wide scene by changing cameras' PTZ settings. We first proposed a spherical rectification method to deal with the stereo vision in dual-PTZ-camera system. We then proposed a two-step stereo matching framework to deal with the gray level auto adjustment problem which exists in the dual-PTZ-camera system. Experimental results show that both rectification and stereo matching work well. Thus, depth information can be estimated under different PTZ settings for two cameras. Experimental results also validate the two advantages of stereo vision in dual-PTZ-camera system: (1) depth information can be obtained from two images with different PTZ settings; (2) depth precision can be improved by using higher zoomed image pair.

As in real applications, there could be some errors in stereo rectification, which might be resulted from two reasons, i.e. the simplification of camera model and the uncertainties of PTZ parameters. For the first reason, a more precise camera model can replace current model in use. For example, more factors can be taken into consideration, such as lens distortion, focus length variation, and the assumption that optical center coincides with motion center, etc. For the second reason, unless using a more accurate camera, it always needs more image information or scene information to refine the parameters. In depth map estimation, as our purpose is to propose a general framework for stereo vision with dual-PTZ-camera system, we choose a simple region-based SSD matching algorithm which might not reach the demand of real applications. The stereo matching approach with more convenience and better performance might be considered in our future research.

Acknowledgments

The authors acknowledge support from Natural Science Foundation of China, Natural Science Foundation of Beijing, National 863 Hi-Tech Development Program of China, and Basic Research Foundation of Tsinghua University.

References

[1] D. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall, 2003.
[2] M.Z. Brown, D. Burschka, G.D. Hager, Advances in computational stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (8) (2003) 993–1008.
[3] A.B.J. Alomoinos, I. Weiss, Active vision, International Journal of Computer Vision 2 (1988) 353–366.
[4] S.-W. Shih, Y.-P. Hung, W.-S. Lin, Calibration of an active binocular head, IEEE Transactions on Systems, Man, and Cybernetics, Part A 28 (4) (1998) 426–442.
[5] S. Ma, A self-calibration technique for active vision systems, IEEE Transactions on Robotics and Automation 12 (1996) 114–120.
[6] F. Du, M. Brady, Self-calibration of the intrinsic parameters of cameras for activevision systems, in: CVPR, 1993, pp. 477–482.
[7] H. Truong, S. Abdallah, S. Rougeaux, A. Zelinsky, A novel mechanism for stereo active vision, in: Proceedings of the Australian Conference on Robotics and Automation, 2000.
[8] A.Z.A. Dankers, N. Barnes, Active vision—rectification and depth mapping, in: Proceedings of the Australian Conference on Robotics and Automation, 2004.
[9] W.N. Klarquist, A.C. Bovik, Fovea: a foveated vergent active stereo system for dynamic three-dimensional scene recovery, in: ICRA, 1998, pp. 3259–3266.
[10] S.-C. Park, S.-W. Lee, Fast distance computation with a stereo head-eye system, in: Biologically Motivated Computer Vision, 2000, pp. 434–443.
[11] B. Scassellati, A binocular, foveated active vision system, Tech. rep., (1998).
[12] S. Sinha, M. Pollefeys, Towards calibrating a pan–tilt–zoom cameras network. Available from: URL <citeseer.ist.psu.edu/721823.html>.
[13] M. Li, J.-M. Lavest, Some aspects of zoom lens camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (11) (1996) 1105–1110.
[14] R. Collins, Y. Tsin, Calibration of an outdoor active camera system, in: CVPR, 1999, pp. 528 – 534.
[15] R. Willson, Modeling and calibration of automated zoom lenses, Tech. rep., CMU-RI-TR (1994).
[16] D. Papadimitriou, T. Dennis, Epipolar line estimation and rectification for stereo images pairs, IEEE Transactions on Image Processing 3 (4) (1996) 672–676.
[17] M. Pollefeys, S.N. Sinha, Iso-disparity surfaces for general stereo configurations, in: ECCV, issue 3, 2004, pp. 509–520.
[18] C.T. Loop, Z. Zhang, Computing rectifying homographies for stereo vision, in: CVPR, 1999, pp. 1125–1131.
[19] R.I. Hartley, Theory and practice of projective rectification, International Journal of Computer Vision 35 (2) (1999) 115–127.
[20] S. Roy, J. Meunier, I.J. Cox, Cylindrical rectification to minimize epipolar distortion, in: CVPR, 1997, pp. 393–399.
[21] M. Pollefeys, R. Koch, L.J.V. Gool, A simple and efficient rectification method for general motion, in: ICCV, 1999, pp. 496–501.
[22] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, International Journal of Computer Vision 47 (1-3) (2002) 7–42.
[23] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[24] C.G. Harris, M. Stephens, A combined corner and edge detector, in: Fourth Alvey Vision Conference, 1988, pp. 147–151.
[25] H. Tao, H.S. Sawhney, R. Kumar, A global matching framework for stereo computation, in: ICCV, 2001, pp. 532–539.
[26] F. Ernst, P. Wilinski, C.W.A.M. van Overveld, Dense structure-from-motion: An approach based on segment matching, in: ECCV, issue 2, 2002, pp. 217–231.
[27] Y. Wei, L. Quan, Region-based progressive stereo matching, in: CVPR, issue 1, 2004, pp. 106–113.