Chapter 4

Stereo Vision for Obstacle Detection

ABSTRACT

As the third module of the full-view visual navigation, this chapter presents a real-time visual obstacle detection system that is novel in both sensors and algorithms. Single-camera stereo vision is realized by using a set of handily placed mirrors, and optical and electrical identities of binocular image pair are guaranteed. A software gaze transformation method is presented to produce zero disparities for road surface figures, thus simplifies road obstacle detection. The proposed non-zero disparity filtering (NZDF) algorithm makes the process of obstacle detection simple and robust. Problems of feature detection and correspondence are avoided and the implemented system is robust, near real-time, and applicable to natural scene environments. Experimental results are given for outdoor real scene images.

Keywords: Mobile robot, Intelligent vehicle, Obstacle detection, Single-camera stereo, Gaze transformation

4.1. INTRODUCTION

A mobile robot must be capable of detecting obstacles in real time in order to carry out autonomous navigation tasks in an unstructured outdoor environment. Existing obstacle detection systems typically compute the position of an obstacle relative to the mobile robot by using range information [1,2], which may be obtained from a laser ranger finder or various vision-based techniques. Among them stereo/motion vision systems have advantages of low cost, low power consumption and a high degree of mechanical reliability. However the relatively slow processing speed of existing stereo vision methods limit their applications in real world problems, mainly due to the long computing time required for establishing stereo correspondences.

Usually, the reconstruction of a detailed 3D world map is noisy and computationally prohibitive, and may be unnecessary in some of the applications, such as obstacle detection. Furthermore, in a most general sense, distinguishing a visual target (e.g. an obstacle) is the notorious segmentation problem, and it is not clear how to extract this information from visual signals in all cases. In paper [3], camera movements are mechanically controlled to simplify the visual processing necessary for a target by keeping the camera locked on the target. The central idea is that localizing attention in 3D space makes the pre-categorical visual processing sufficient to hold gaze. Converged cameras produce a horopter (surface of zero stereo disparity) in the scene and hence binocular features with no disparity can be located with a simple filter, showing the target's location in the image. Instead of requiring a way to recognize the target, the system relies on active control of the camera movements and binocular fixation to locate the target.

Simple and fast stereo obstacle detection methods have been proposed based on the fact that obstacles mostly lie on a flat ground [4-7]. Badal et al [4] developed a practical obstacle detection and avoidance system for an outdoor robot. A fast *hierarchical stereo correspondence algorithm* produces sparse disparity information and compares it with that of the ground plane. The system can only detect obstacle protruding sufficiently high from the flat ground plane with the requirements of 1D epipolarity and identical cameras and digitizers. The system also requires a tedious calibration procedure. The system executes at 2 Hz on a Datacube MV-200 and SGI host.

Stojorham et al [5] proposed a computationally simple floor level detection method which can check the presence of obstacle above a ground plane. A geometrical image transformation, namely *inverse perspective mapping* was introduced to project both original stereo images onto the ground plane so that points on the ground plane are assigned zero disparities. Although this method is computationally efficient, resolutions of the mapped images could be greatly reduced duo to the severe distortion by the inverse perspective transformation in the case of forward looking cameras.

Zhu & Lin [6] proposed two fast obstacle detection algorithms based on the so-called *reprojection transformation*, which also projects the original images on the ground plane. The characteristic of the algorithms is that only one camera is used and motion parallax is utilized. The first algorithm differs from that of Stojorham's algorithm in that accurate motion parameters (e.g., baselines) between two images are not available so that a search for zero disparity match of the ground plane is explored. The second algorithm creates epipolar plane images from the re-projected image sequence and 3D coordinates of the obstacle are accurately estimated. These two algorithms have been implemented in PIPE, a pipelined image processing engine, at 5 Hz and 30 Hz respectively.

Cho et al [7] proposed a computationally efficient stereo vision method for obstacle detection by an indoor mobile robot. A pair of original stereo images is geometrically transformed into a common image plane located at the mid-point of the stereo image planes. In the transformed stereo images, image pixels of the floor plane have zero isodisparities, while obstacle features above the floor level correspond to non-zero isodisparities. The proposed correspondence scheme of the isodisparity prediction and verification facilitates frame-wise operations without an exhausted searching process for stereo matching, provided that the paired image edges have unique matches by using edge angles and intensity patterns. Camera calibration is needed for the geometrical transformation, which may introduce some system error. The isodisparity checking is carried out for every constant disparity value so that the required time for an entire detection period in a PC is 15 seconds.

Limitations of the aforementioned work are in twofold. First, the cameras should be carefully calibrated so that the exact zero disparity properties for a given plane can be guaranteed. Second, the assumption of photometric identity is difficult to keep for two physical separated cameras and digitizers. With an emphasis to overcome these limitations, we have developed a near real-time stereo vision system for obstacle detection using a single camera. The system has the following features:

1) A projective model is proposed for the so-called zero disparity *gaze transformation*. There is no need to calibrate the internal and external parameters of cameras in order to realize the gaze transformation. Therefore the transformation is straightforward and accurate. Meanwhile considering the geometric relation of the pair of original stereo image planes and the ground plane, an image re-projection transformation is selected that minimum the lose of image resolution.
2) Single-camera stereo vision systems have been designed so that the electrical and photometric identity can be satisfied for the stereo pair. In particular, a patented single camera binocular vision system will full horizontal field of view is designed and constructed where left and right views are projected to the up- and bottom-halves of a single image. The system is cheap because only a

single camera and a single digitizer are needed.

3) Our work explores the use of pre-categorical visual cues (i.e., prior to object recognition) in order to distinguish the visual target (obstacle) from the surrounding scene (road). The novelty of the (dynamic) gaze transformation method, which resembles gaze control of the human vision, lies in the fact that it brings the road surface zero disparity so that the feature extraction and matching procedures of stereo vision methods are completely avoided in the proposed obstacle detection algorithms. However no mechanical control of camera movement and vergence is needed.
4) Dynamic gaze transformation methods are developed so that obstacle detection algorithms can work on a rough road surface. The progressive processing strategy from yes/no verification, through focus of attention, to 3D measurement based on the gaze transformation make the hierarchical obstacle detection algorithms efficient, fast and robust.

This chapter is organized as follows. Section 2 describes the principle of the image gaze transformation. The designs of two types of single camera stereo systems are given in Section 3. Section 4 gives the basic obstacle detection algorithm based on the zero-disparity property of the ground plane. A few discussions and concluding remarks are given in the last section.

4. 2. PLANAR GAZE TRANSFORMATION

4.2.1. Principle

In this section, the principle of the planar gaze transformation is introduced by building the relation of Euclidean geometry and the projective geometry of binocular cameras. The introduction of the Euclidean geometry is for the calculation of the 3D location of an obstacle and the adaptation of the vision system to camera fluctuations due to bumping. Suppose the coordinate system of left and right cameras are $X_1Y_1Z_1$ and $X_2Y_2Z_2$ respectively, where Z_1, Z_2 are their respective optical axes, and the world (robot) coordinate system is XYZ (Fig. 4.2.1). We will use the right coordinate system as the *reference coordinate system* for the gaze transformation. A point $\boldsymbol{P} = (x, y, z)^T$ in the world coordinate system corresponds to a pair of points $\boldsymbol{P}_1 = (x_1, y_1, z_1)^T$, $\boldsymbol{P}_2 = (x_2, y_2, z_2)^T$ in the two camera coordinate systems. The relations between the three coordinates are

$$\boldsymbol{P} = \boldsymbol{R_1P_1} + \boldsymbol{T_1} = \boldsymbol{R_2P_2} + \boldsymbol{T_2}$$

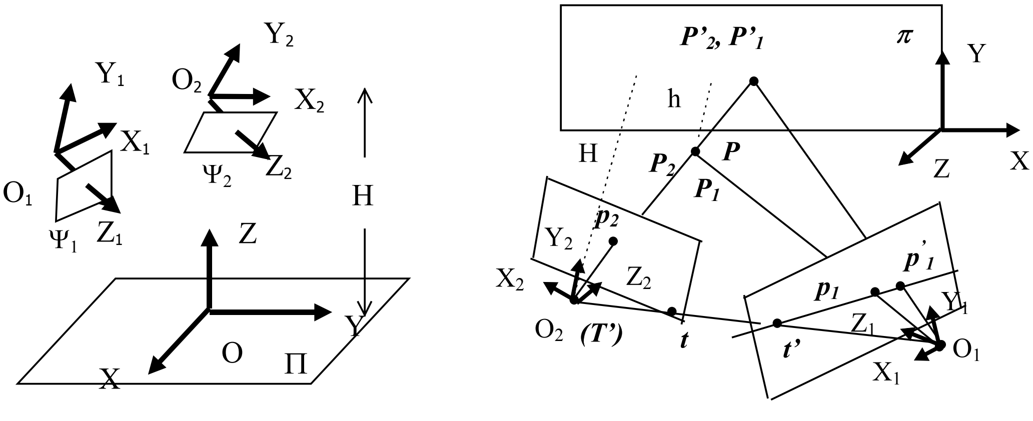
where $\boldsymbol{R_1}$, $\boldsymbol{R_2}$ are the rotation matrixes from coordinate systems $X_1Y_1Z_1$ and $X_2Y_2Z_2$ to the world coordinate system *XYZ* respectively, and T_1, T_2 are the corresponding transitional vectors. The binocular relation is

$$\boldsymbol{P_2} = \boldsymbol{RP_1} + \boldsymbol{T} = \boldsymbol{R(P_1} - \boldsymbol{T'\,)} \qquad (4.2\text{-}1)$$

where $\boldsymbol{R} = \boldsymbol{R_2^{-1}R_1}$ is a 3x3 rotation matrix, $\boldsymbol{T} = \boldsymbol{R_2^{-1}}\left(\boldsymbol{T_1} - \boldsymbol{T_2}\right)$ is the coordinates of the optical center of the right camera in the left coordinate system, and $\boldsymbol{T'} = -\boldsymbol{R}^{-1}\boldsymbol{T}$ is the coordinates of the optical center of the left camera in the right coordinate system. Under the pinhole camera model, the left and right image coordinates for the point $\boldsymbol{P}$ are

$$(u_i, v_i) = \left(f\frac{x_i}{z_i}, f\frac{y_i}{z_i} \right), or \;\; \boldsymbol{p}_{\mathrm{i}} = \begin{pmatrix} u_i \\ v_i \\ f \end{pmatrix} = \frac{f}{z_i}\boldsymbol{P}_i, \mathrm{i} = 1,2 \qquad (4.2\text{-}2)$$

where f is the equivalent focal length for both left and right cameras. In the following, the subscript "$_1$" represents the coordinates in the reference (right) coordinate system, the subscript "$_2$" represents the coordinates in the left coordinate system, and the prime " ' " represents the coordinates in the re-projected (transformed) coordinate system.



(a) Coordinate systems (2) Projective gaze transformation

Fig. 4.2.1. Coordinate systems and the gaze transformation

A plane π in the reference coordinate system can be represented as $\boldsymbol{N}^T \boldsymbol{P}_1 = d$ .The equation(4.2-1) for this plane can be written as

$$\boldsymbol{P}_2 = \left[\boldsymbol{R} + \boldsymbol{T}\boldsymbol{N}^T / d\right]\boldsymbol{P}_1 \qquad (4.2\text{-}3)$$

This 8-parameter planar projective transformation builds up the relation of any points in the plane π in the left and right coordinate systems. Assume the intersection point of the ray O$_2$P$_2$ and the plane π is $\boldsymbol{P_2'}$,and its vector of coordinates in the reference coordinate system is $\boldsymbol{P}_1'$, we can define a 3×3 matrix of a planar gaze transformation as

$$\boldsymbol{A} = \left[\boldsymbol{R} + \boldsymbol{T}\boldsymbol{N}^T / d\right]^{-1} = \left[\boldsymbol{I} - \boldsymbol{T}'\boldsymbol{N}^T d\right]^{-1} \boldsymbol{R}^{-1} \qquad (4.2\text{-}4)$$

Therefore the gaze transformation re-projects point $\boldsymbol{P_2}$ to $\boldsymbol{P}_1'$

$$\boldsymbol{P}_1' = \boldsymbol{A}\boldsymbol{P_2'} \cong \boldsymbol{A}\boldsymbol{P}_2 \qquad (4.2\text{-}5)$$

where “$\cong$” represents equality up to a scale factor(i.e.,$\boldsymbol{P_2'} \cong \boldsymbol{P}_2$).In other words, the gaze transformation ($\boldsymbol{A}$) about the plane π re-projects a point $\boldsymbol{p}_2$ in the left image intoa point $\boldsymbol{p'}_1$ in the reference image:

$$\boldsymbol{p'}_1 \cong \boldsymbol{A}\boldsymbol{p}_2 \qquad (4.2\text{-}6)$$

where $\boldsymbol{P}_1' = \left(x_1' \; y_1', z_1'\right)^T$, $\boldsymbol{p'}_1 = \dfrac{f}{z'_1}\boldsymbol{P}_1'$. For any point in the plane π, we have $\boldsymbol{P_2'} = \boldsymbol{P}_2$, hence we will have $\boldsymbol{P}_1' = \boldsymbol{P_1}$ and $\boldsymbol{p}'_1 = \boldsymbol{p}_1$ (i.e., the projection of pint P in the reference image is the same as the re-projection of its correspondence in the left image). However for points that are not in the plane, the projections are not the same.

From the matrix relation $\left[\boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^T\right]^{-1} = \left[\boldsymbol{I} - \gamma\, \boldsymbol{u}\boldsymbol{v}^T\right]$, where $\gamma = \dfrac{1}{1 + \boldsymbol{v}^T \boldsymbol{u}}, \left(\boldsymbol{v}^T \boldsymbol{u} \neq 0\right)$, we can re-write the gaze transformation matrix as

$$\boldsymbol{A} = \left[\boldsymbol{I} + \alpha \boldsymbol{T}'\boldsymbol{N}^T / d\right]\boldsymbol{R}^{-1} \qquad (4.2\text{-}7)$$

where $\alpha = d \big/ \left(d - \boldsymbol{N}^T \boldsymbol{T}'\right)$ with the condition $\boldsymbol{N}^T \boldsymbol{T}' \neq d$, i.e., the optical center of the left camera is not in the plane π. Substituting Eqs. (4.2-1) and(4.2-7) into (4.2-5) we have (For a prove see Appendix 4.1）

$$\boldsymbol{P}_1' \cong \boldsymbol{P}_1 - \frac{h}{H}\boldsymbol{T}' \qquad (4.2\text{-}8)$$

where h is the perpendicular distance (with sign) of the point$\boldsymbol{P}_1$ to the plane π, H is the perpendicular distance from the optical center of the left camera (i.e.,$\boldsymbol{T}'$) to the plane π(Fig. 4.2.1a).

Let $\boldsymbol{T} = (T_x T_y, T_z)^T$,$\boldsymbol{t}' = \frac{f}{T_z}\boldsymbol{T}'$ $(\text{if } T_Z \neq 0)$. By some tedious deduction (Appendix 4.2),the "disparity" of the left re-projected point $\boldsymbol{p}'_1$ and the right image point $\boldsymbol{p}_1$ is

$$\Delta \boldsymbol{p} = \boldsymbol{p}_1 - \boldsymbol{p}_1 = \frac{hT_z}{hT_z - Hz_1}(\boldsymbol{p}_1 - \boldsymbol{t}'), \quad (T_z \neq 0) \qquad (4.2\text{-}9)$$

and

$$\Delta \boldsymbol{p} = \boldsymbol{p}_1 - \boldsymbol{p}_1 = \frac{f\,h}{Hz_1}(T_x, T_y, 0)^T, \quad (T_z = 0) \qquad (4.2\text{-}10)$$

From Eq. (4.2-6) and Eq. (4.2-9) or Eq.(4.2-10), we have the widely used planar projective transformation (Kumar94; Sawhney94; Shashua96)

$$\boldsymbol{p}_2 \cong \boldsymbol{A}_\pi \boldsymbol{p}_1 + k\boldsymbol{t} \qquad (4.2\text{-}11)$$

where

$\boldsymbol{A}_\pi = \boldsymbol{A}^{-1}$ is the matrix of $2D$ planar projective transformation about the plane π

$\boldsymbol{t} = \boldsymbol{A}^{-1}T'$ is the epipole in the left image coordinate system

$k = -\frac{hf}{Hz_1}$ is the projective "depth" of the point $\boldsymbol{P}_1$ in the reference viewpoint.

It should be noted that $[\boldsymbol{p}_1, k]^T$ is the representation of a point P_1 fixed in the reference image coordinate system. This representation not only compensates the affect of camera rotation, but also provides a view-based scene description.

4.2.2 Properties

There are several interesting properties of this gaze transformation:

1. For a point $\boldsymbol{P}$ in the reference plane π(i.e.$h = 0$ in Eqs. (4.2-9) and (4.2-10)), the planar gaze transformation $\boldsymbol{A}$ re-projects the left image point$\boldsymbol{p}_2$ to its corrsponding point$\boldsymbol{p}_1$ in the reference image (i.e., $\boldsymbol{p}'_1 = \boldsymbol{p}_1$). Therefore the disparity of the point pair aftergaze transformation is 0, so is the

projective depth.

2. For a point that is not in the plane π (i.e., $h \neq 0$), the gaze transformation eliminate the image motion introduced by camera rotation. The remaining planar motion parallax is introduced by camera translation ***T***. The motion parallax is the $\boldsymbol{p}_1 - \boldsymbol{p'}_1$, the difference vector of the real projection and the re-projection of the point ***P*** in the reference image.

(1) When $T_z \neq 0$, the parallax vector is

$$\boldsymbol{p}_1 - \boldsymbol{p}_1' = \frac{kT_z}{kT_z + f}(\boldsymbol{p}_1 - \boldsymbol{t}') = \frac{k'}{k'+1}(\boldsymbol{p}_1 - \boldsymbol{t}') \qquad (4.2\text{-}12)$$

This vector points to the epipole $\boldsymbol{t}' = f\frac{\boldsymbol{T}'}{T_z}$ in the reference image, where $k' = k\frac{T_z}{f}$.The projective depth ($k = -\frac{hf}{Hz_1}$) is proportional to the “height” of the point (h) and inversely proportional to the distance (“depth”) of the point (z_1)。

(2).When the line connecting the optical centers of the left and right cameras is parallel to the reference plane, i.e. $T_Z = 0$, the motion parallax filed will be a parallel field, and point to an epipole in infinity, i.e. $\left(T_X, T_y, 0\right)$. In this case we have

$$\boldsymbol{p}_1 - \boldsymbol{p}_1' = k\left(T_x, T_y, 0\right)^T$$

where the projective depth is $k = -\frac{hf}{Hz_1}$, which is still proportional to $\frac{h}{z_1}$.

3. The gaze transformation matrix ***A*** or the planar projective transformation matrix $\boldsymbol{A}_\pi$ includes rotation and part of translation, i.e.

$$\boldsymbol{A}_\pi \cong \boldsymbol{R} + \boldsymbol{T}\boldsymbol{N}^{\boldsymbol{T}}/d \qquad (4.2\text{-}13)$$

If the distance between the reference viewpoint and the reference plane is infinity, i.e., $d \to \infty$, $\boldsymbol{A}_\pi$ degrades to a rotation matrix, i.e., $\boldsymbol{A}_\pi = \boldsymbol{R}$ and

$$\boldsymbol{t} = \boldsymbol{A}^{-1}T' = -\boldsymbol{T}\text{ , } \quad k = -\frac{f}{z_1}\left(\because \frac{h}{H} \to 1\right)$$

Therefore Eq. (4.2-11) turns into a projective transformation that is equivalent to the combination of Eqs.(4.2-1) and (4.2-2)

$$\boldsymbol{p}_2 \cong \boldsymbol{R}\boldsymbol{p} + \frac{f}{z_1}\boldsymbol{T} \qquad (4.2\text{-}14)$$

4. The plnar gaze transformation is a planar projective transformation. From the 2D projective theorem, the gaze transformation can be determined by 4 point pairs (three of which are not colinear) on the ground plane. Since we do not need to measure any 3D coordinates in the world, the gaze transformation is based on uncalibrated image pairs.

4.2.3. Applications

Gaze control and image stabilization are considered to be the basic preprocessings in active vision (Ballard89,92; Aloinomos88; Taalebinezahaad92; Coombo94). By gaze control, the stereo cameras or a camera in motion can focus on a target (or a point) in interest, thus improve the computational efficiency and robustness. Many active vision system have achieved the goal of gaze transformation by mechanically control the motion of the cameras (Abbott88; Ballard 92; Coombo94). For "software" gaze control, Kanatani (Kanatani88) proposed a way to simplify the computation of 3D recovery from lines by using rotation transformation. Several algorithms have been proposed to simplify stereo correspondence by ground plane transformations (Zhu90; Cho94; Abbott95; Wong95).

The proposed gaze transformation here is a software gaze control. The gaze control is realized by a planar projective transformation, and therefore the gaze target is a plane. Gaze transformation can be used in the stabilization of an image sequence of the zero disparity horopter selection of a stereo vision system. The key point here is to define a reference plane π, find at least 4 pairs points on the plane and re-project the images. In the following, we will give three important applications that can make use of the gaze transformation.

1. Image stabilization

In Chapter 2 (Panoramic Scene Modeling), an image stabilization technique has been proposed to decompose the dominant motion with known properties and the unpredictable 6 DOF fluctuations, and thus simplify the 3D estimation of a scene. As a matter of fact, the image rectification for image stabilization is one kind of gaze transformation. When the dominant motion is a translational motion parallel to the image plane, the motion parallax filed of the stabilized image sequence will be a parallel field. The gaze transformation is realized in three steps in Chapter 2: image motion estimation and parametric modeling of the motion field; model-based motion filtering; and image rectification. If the fluctuation can be modeled by a 3 DOF rotation, then the gaze transformation matrix $\boldsymbol{A}$ for each frame will be a rotation matrix $\boldsymbol{R}$ which implies that the reference plane is in the infinity. In a real system, the fluctuation could not be a pure rotation, so the gaze transformation will be a projective transformation that includes both rotation and translation. In this case, the reference plane will be a virtual plane π. Through the alignments of points in the plane, we can approximately achieve a global translational motion across a long image sequence that satisfies Eq. (4.2-10).

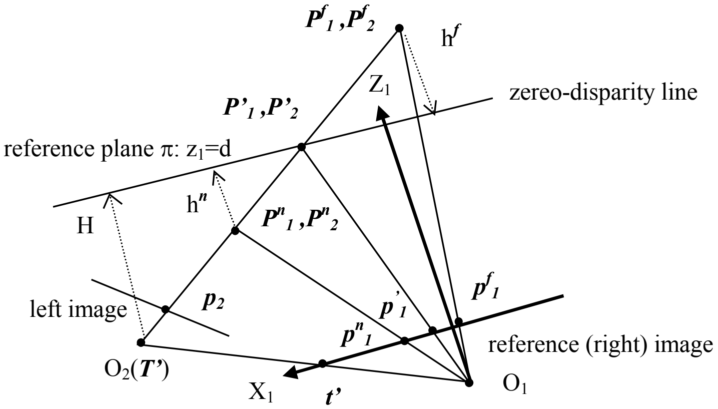


Fig. 4.2.2 Vertical zero-disparity gaze transformation (top view). In this figure the superscript “n” (near) denotes points whose distances are smaller than d, and the superscript “f” (far) denotes points whose distances are larger than d. The subscripts “1” and “2” represent coordinates in the right and left images respectively.

2. **Vertical zero-disparity gaze transformation**

In the application of road obstacle detection for a moving vehicle, the gaze transformation can be very effective. Since during the travel of the vehicle, obstacles are not allowed to appear within a certain distance d, the vision system specifically needs to check if there are obstacles whose distances 1) are greater than, 2) equal to or 3) smaller than distance d. Once an obstacle is detected within distance d, the vision system should send an alarm signal to the control system and take some necessary actions, such as reducing the speed or stepping on the brake. Without lose of generality, assume that the optical axis of the reference (right) camera is parallel to the road surface, and is along the motion direction of the vehicle. Define the reference plane π as

$$z_1 = d \qquad (4.2\text{-}15)$$

i.e., a vertical plane with the minimum safety distance *d*.. For this vertical plane we have $\boldsymbol{N} = (0, 0\ 1)^T$, therefore the gaze transformation matrix in Eq. (4.2-6) becomes

$$\boldsymbol{A}=\left(\boldsymbol{I}+\alpha\, \boldsymbol{T}'\, \boldsymbol{N}^T/d\right)\boldsymbol{R}^{-1} = \frac{1}{d-T_z}\begin{pmatrix} d-T_z & 0 & T_x \\ 0 & d-T_z & T_y \\ 0 & 0 & d \end{pmatrix}\boldsymbol{R}^{-1} \qquad (4.2\text{-}16)$$

After the reprojection of the left image, stereo disparities of any points in distance d will be zero. By alignments of features(edges or textures) in the stereo pair of the right image and the left reprojected image, we can check whether there are objects in distance d. It should be noted that we can make such judgements only if we can detect some kinds of salient visual features. Otherwise, in the intensity smooth areas, the zero-disparity conclusion cannot be made anyway. We have the following results: for any point (e.g. point P^f) with $z_1 > d$,disparities will be positive; otherwise for any points (e.g., point P^n)with $z_1 < d$, disparities will be negative(Fig. 4.2.2). If there is no obstacle on the road, there is only one scan line that has zero-disparities. Coombo and Brown (Coombo94) has achieved zero-disparity horopter by rotating the camera, and further computed the distance of the object with zero disparity. Here the zero-disparity gaze transformation avoids the mechanical control of the camera, instead a software gaze control is achieved. Furthermore, the zero-disparity property in distance d can be used to simplify stereo match for points near distance d since the search area is reduced.

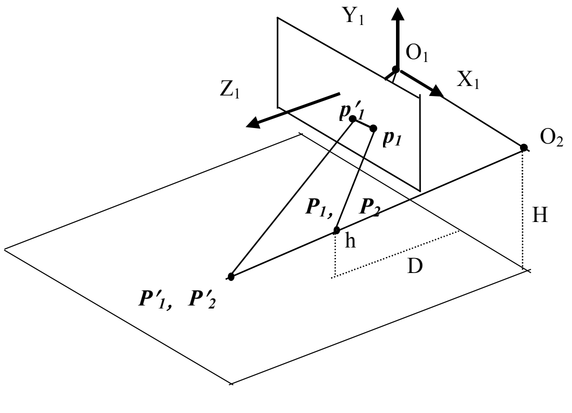


Fig. 4.2.3 Ground zero-disparity gaze transformation

3. **Ground zero-disparity gaze transformation**

In an environment with a large number of horizontal structures, such as an indoor environment, the ground surface gaze transformation can be used to greatly simplify stereo matching(Zhu90; Lin96). Suppose that the baseline of the binocular cameras is parallel to the floor, or a single camera is mounted on a mobile vehicle moving on the floor. Then the gaze transformation about the floor re-projects the left image in the plane of the right image, and disparities of points on the floor becomes zero. For Any figure on a plane parallel to the floor, the shape in the reference image $\{\boldsymbol{p}_1\}$ will be similar to that of the left re-projected image$\{\boldsymbol{p}'_1\}$. Assume that the XOY plane of the world coordinate system is the floor surface. If we perform the same image rotation transformation to mapping the left reprojected image and the right image to the floor surface, we have a pair of ground re-projected images

$$\begin{cases} \boldsymbol{p}' \cong \boldsymbol{R}_1 \boldsymbol{p}_1 \hat{=} \boldsymbol{A}_1 \boldsymbol{p}_1, \ (\boldsymbol{A}_1 = \boldsymbol{R}_1) \\ \boldsymbol{p}'' \cong \boldsymbol{R}_1 \boldsymbol{p}'_1 = \boldsymbol{R}_1 \boldsymbol{A} \boldsymbol{p}_2 \hat{=} \boldsymbol{A}_2 \boldsymbol{p}_2, \ (\boldsymbol{A}_2 = \boldsymbol{R}_1 \boldsymbol{A}) \end{cases} \qquad (4.2\text{-}17)$$

where “$\hat{=}$ represents “by definition”. This time, for any horizontal planar figure of any height, the same in the ground re-projected images $\{\boldsymbol{p}'\}$ and $\{\boldsymbol{p}''\}$ will be exactly the same. This property can greatly facility stereo match. However for doing this, we need to calibrate the cameras and find the rotation matrix $\boldsymbol{R}_1$

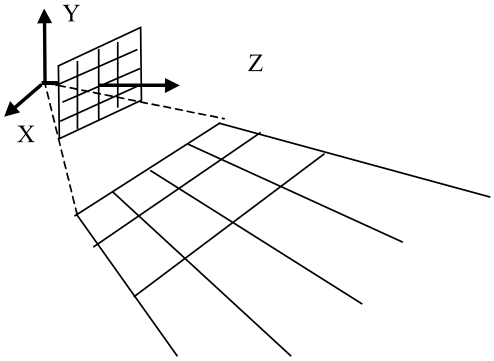


Fig. 4.2.4 Distortion in ground Gaze transformation

Another problem for ground reprojected images is the degradation of image resoltuion due to large perspective distortion. In the application of an outdoor obstacle detection, the optical axis could be

almost parallel to the ground plane since the height of the camera H（about 1- 2 m）will be much smaller than the distance that obstacles are needed to be detected (typically the average distance could be about 30 m). The image resolution will be greatly distorted if we apply the ground re-projection using Eq.（4.2-17）, see Fig. 4.2.4. Therefore obstacle detection will be directly performed on the left re-projected image { $\boldsymbol{p'}_1$ } and the original right image { p_1 }. In this case we do not need to calibrate the extrinsic camera parameters ***Ri*** and ***Ti***，(i=1,2), niehter do we need to measure the 3D coordinates of any ground points. What we need to do is to provide image correspondences of at least four pair of points in the left and right images, and find the gaze transformation matrix A.
In summary, iIf the left image re-projects to the right image using the gaze transformation of Eq. (4.2-6), the point in the ground plane will have zero disparities between the reprojected left image and the original right image. This property is very useful for the image-level ground obstacle detection. Comparing to the previous work, only one image, say left image, is reprojected. Furthermore the image resolution is preserved since no severe distortion is introduced by the geometrical transformation even the optical axis of the camera is almost parallel to the ground plane in order to detect distant obstacles.

In our obstacle detection algorithm, there is no requirement that the baseline of the binocular cameras is parallel to the ground plane (i.e., $T_Z= 0$). Instead in a genral setting where $T_z \neq 0$,the disparity Δu of the left re-projected image{ $\boldsymbol{p'}_1$ }and the original right image { $\boldsymbol{p}_1$ } will be

$$\Delta u = \frac{hT_z}{hT_z - HD}\Delta u_F \qquad (4.2\text{-}18)$$

where $\Delta u = \sqrt{(u_1 - u_1')^2 + (v_1 - v_1')^2}$ ， $\Delta u_F = \sqrt{(u_1 - u_F)^2 + (v_1 - v_F)^2}$ ， $\boldsymbol{t}' = (u_F, v_F, f)^T$ is the knowm epipole in the reference (right) image , $\boldsymbol{p}'_1 = (u'_1, v'_1, f)^T$ is the reprojection of the left image point, $\boldsymbol{p}_1 = (u_1, v_1, f)^T$ is the point in the right image, *H* is the height of the optical center of the right camera from the ground plane, *D* is the distance of the point P_1 in the Z_1 direction, and h is the height of the point P_1 to the ground plane.

In a real system, we can setup the binocular cameras such that $T_z << T_x$, $T_y << T_x$, and the optical axis is roughly parallel to the ground plane. Otherwise the accurate relation will be given by $\boldsymbol{R_1}, \boldsymbol{T_1}$, and the condition of horizontal optical axis can be achieved by a rotation transformation similar to that of Eq.(4.2-17). In the following discussion , we assume that the optical axis of the reference camera OZ_1 is parallel to the ground plane. Using Eq.(4.2-10) instead of Eq. (4.2-10) we can have a concise disparity equation between disparity Δu, height *h* and distance *D* (=z_1)

$$\Delta u = \frac{fT_x}{HD} h \quad \Delta v \approx 0 \tag{4.2-19}$$

where $\Delta u = u_1 - u_1', \Delta v = v_1 - v'_1$. From Eq. (4.2-19) we can easily find the relation among disparity, height and distance of a point in the space: Disparity Δu is proportional to the height of the point and inversely proportional to the distance of that point. Correspondence points are constraint to a scanline. (Fig. 4.2.3). It should be noted that Eq. (4.2-19) also holds for points with h <0（lower than the ground plane）and points with h>H（higher than the camera. Notice that points that are higher than the camera cannot be represented in a ground re-projection transformation.

4. 3. SINGLE CAMERA STEREO

The patented design [8] of an optical based binocular vision system using a single camera stems from the requirement to record the outdoor binocular image sequences by a camcorder frequently and portably. However advantages of a single-camera binocular vision system are far beyond this intuitive fact. First, the system is cheap and effective. A stereo image pair can be captured in a snap (simultaneously at the frame rate or field rate). Only a single camera and a single frame grabber are required, and no electrical synchronization is needed. Second, the stereo image pair is identical. Since a single camera and a single frame grabber are used to capture the stereo pair, no differences of electrical, optical and photometric properties exist, such as iris, focal length, noise level, sensitivity, contrast and etc. The image identity is favorable to the obstacle detection in gray-level image. Third, it can be constructed as a concise binocular vision system for various application, such as active vision and tele-reality. The fixation point, baseline length, pan/tilt angles can be easily controlled by rotating and translating the auxiliary mirrors.

In face, a real-time and compact binocular vision system has wide applications in photometry, medical imaging, remote operation and robotics. There are many existing designs such as rotating refection mirror method (Teoh84; Nishimoto87), dual back reflection mirror method (Goshtasby93) and light splitting prism method (Ariyaeeinia94) and tri-splitting lens method (Kurada95). In this section we will give a complete analysis of the geometry and the system aspects of our patented design(Shi & Zhu 95). Furthermore, an improved design will be given in which left and right views are projected to the up- and bottom-halves of a single image so that both left and right images will have full horizontal field of views.

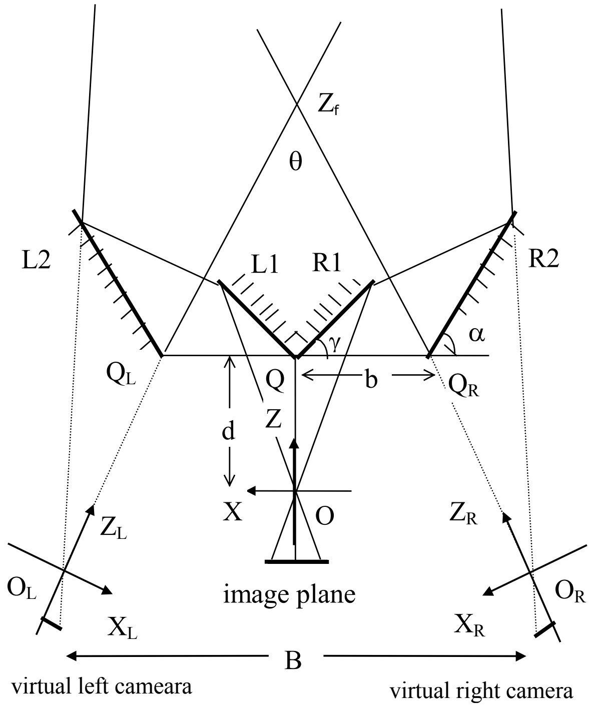


Fig. 4.3.1. Basic single-camera binocular vision system: left-right imaging (top view)

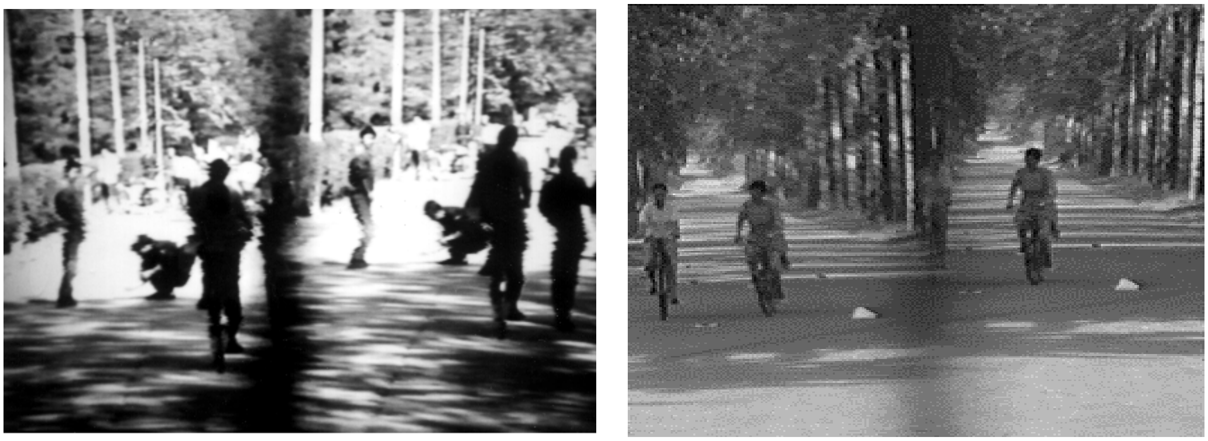


Fig. 4.3.2 Real images captured by basic single-lens binocular vision system

4.3.1. Single-lens binocular vision system I: right-left imaging

The basic single-lens binocular vision system is shown in Fig. 4.3.1. The design goal of the system is that images viewed from left and right viewpoints with certain baseline are projected into the left and right halves of a single image sensor respectively. The system consists of two primary mirrors, two auxiliary mirrors and a standard video camera. The coordinate system of the real camera is XYZ, with O the optical center, OZ the optical axis and XOY the plane parallel to the image plane. The intersection line of the two primary mirrors lies along the Y direction in the plane YOZ. Two auxiliary planar mirrors are parallel to this intersection line. In the standard setting, the angle between the left (right) primary mirror $L_1(R_1)$ and the image plane is $\gamma = 45^\circ$, and the angle between left (right) auxiliary mirror $L_2(R_2)$ and the image plane is $\alpha \in \left(45^\circ, 90^\circ\right)$. The perpendicular distance from the optical center (O) of the real camera to the intersection line of L_1 and R_1 is d . This intersection line passes through the two auxiliary mirrors and distance between left (or right) auxiliary mirror and the intersecting line is *b* along the X direction. Half viewing-angle of the camera is β . The single real camera, left/right primary and auxiliary mirrors construct a system of equivalent left and right virtual "cameras" with viewpoints O_L and O_R . The left and right images are projected into the left and right halves of a single sensing plane.

Given d, b, α , the coordinates of left and right optical centers (X_L, Y_L, Z_L) and (X_R, Y_R, Z_R) in the camera coordinate system XYZ are

$$\begin{cases} X_L = b - (b+d)\cos 2\alpha \\ Y_L = 0 \\ Z_L = d - (b+d)\sin 2\alpha \end{cases} \qquad (4.3\text{-}1)$$

$$\begin{cases} X_R = -b + (b+d)\cos 2\alpha \\ Y_R = 0 \\ Z_R = d - (b+d)\sin 2\alpha \end{cases} \qquad (4.3\text{-}2)$$

respectively. If the left and right images are rectified through the rotation transformation (similar to gaze transformation) so that the "new" optical axes are parallel, then the baseline length is

$$B = X_L - X_R = 2b + 2(b+d)\cos(\pi - 2\alpha) \qquad (13)$$

The fixation angle between the original left and right virtual optical axes is

$$\theta = 2\left(90^\circ - 180^\circ + 2\alpha\right) = 4\alpha - 180^\circ \qquad (4.3\text{-}3)$$

and the Z coordinate of the fixation point is

$$Z_f = bctg\frac{\theta}{2} + d = b\left(-tg2\alpha\right) + d \qquad (4.3\text{-}4)$$

Two real images captured by the basic system are shown in Fig. 4.3.2.

4.3.2 Single-lens binocular vision system: up-down imaging

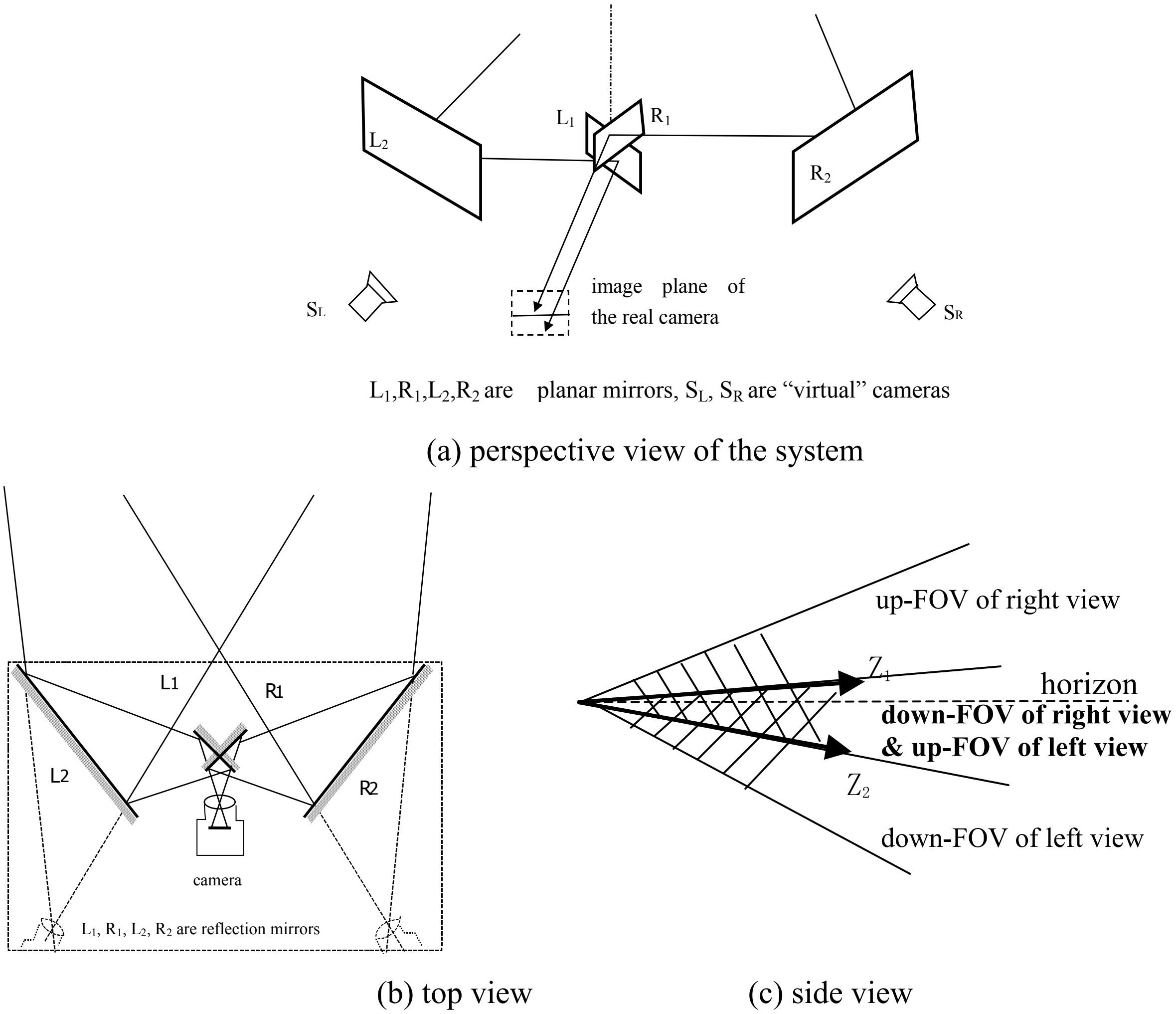
The modified single camera binocular imaging system is designed to satisfy the following requirements of the obstacle detection system.

(1) The equivalent image pair from the left and right viewpoints is projected onto a single image sensing plane.

(2) Viewing angle in the horizontal direction is not divided between two images so that the system can capture both images with full horizontal field of views.

(3) Binocular images are projected onto the up and down halves of a single image plane so that the entire image is fully used. This is because the up-half of an original monocular image of a scene above the road, is not important to obstacle detection.

Modifying the design in Fig. 4.3.1 by extending the up-half of the left primary mirror and the down-half of the right primary mirror, the left and right virtual cameras can see the entire field of view (FOV) in the horizontal direction (Fig. 4.3.3). In order that the up- and down-half of the FOV form the common FOV, the left and /or right auxiliary mirrors should tilt slightly down and/or up respectively. The properties of binocular imaging geometry, baseline length and the fixation point are similar to those of the basic binocular vision system I. The difference is that in the modified system II, the optical centers O_L, O_R of the two virtual cameras are not located in the same height, i.e. $Y_R < 0, Y_L > 0$. Two real images captured by the modified system are shown in Fig. 4.3.4.



L_1,R_1,L_2,R_2 are planar mirrors, S_L, S_R are “virtual” cameras

(a) perspective view of the system

(b) top view (c) side view

Fig. 4.3.3. Modified single-camera binocular vision system: up-down imaging

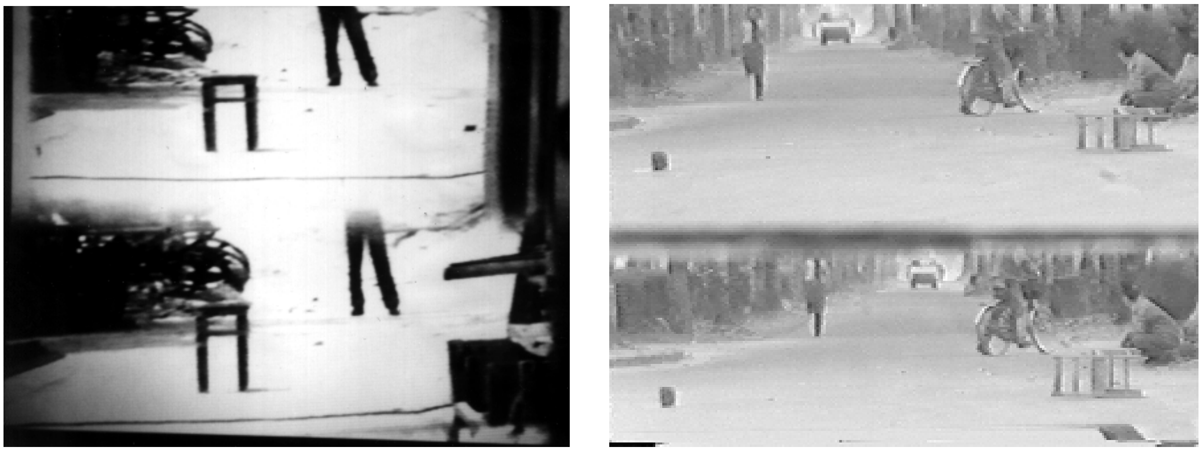


Fig. 4.3.2 Real images captured by modified single-lens binocular vision system

4.4. REAL-TIME OBSTACLE DETECTION

The basic real-time obstacle detection (RTSOD) system consists of three parts: the sensor, the gaze transformation and the non-zero disparity filtering algorithm. The sensor is the single-camera binocular imaging system given in Section 2. The gaze transformation has been described in Section 1. In this section a non-zero disparity filtering algorithm is given and an RTSOD implementation and its performance are discussed.

4.4.1. Basic algorithm

The non-zero disparity filter (NZD filter) is the heart of the RTSOD algorithm. In Coombs & Brown's binocular active vision system [3], the 3D shape of the horopter (surface of zero stereo disparity) is rather complicated (and may not be well defined), since it involves vertical, as well as horizontal, disparities. In the RTSOD system, the gaze transformation creates a zero disparity plane for all the features on the ground plane. In this way obstacles can be detected easily by the NZD filter. The image difference operation on a stereo pair after gaze transformation separates points with non-zero disparities from those with zero disparities by examining difference values. Theoretically, all the pixels of ground plane features will have zero values in the difference image. However, in practice, this property can not be kept strictly due to the photo-metric and geometric bias. So the NZD Filter is applied to throw away points whose difference values are lower than a certain threshold, and then to filter out small areas which imply small disparities. A grouping procedure gathers non-disparity points into each possible obstacle regions.

The basic algorithm consists of six steps:

Step 1. *Binocular imaging*: An image pair of left and right images L_i and R_i is captured simultaneously by the single camera binocular imaging system (I or II) and a frame grabber.

Step 2. *Gaze transformation*: A reprojected left image L_{pi} is created by re-projecting the left image L_i to the plane of the right image R_i.

Step 3. *Image difference*: Difference image D_i is produced by calculate the absolute difference of each corresponding points in L_{pi} and R_i

Step 4. *Non-zero disparity filtering*: Non-zero disparity image B_i is created by first throwing away the points in the difference image D_i whose value is smaller than a certain threshold T , then filtering out small areas which indicate small disparities.

Step 5, *Obstacle grouping*: A set of obstacle regions $\{O_i, i=1,\cdots N\}$ is obtained by grouping connected or nearby points into each obstacle region O_i. Sizes of obstacles and distances between obstacles are considered in the grouping procedure.

Step 6: *Parameter Computing*. For each obstacle (region), location (x, y)、width w and height h are calculated and an uncertainty value is estimated (see Section 4.4.2).

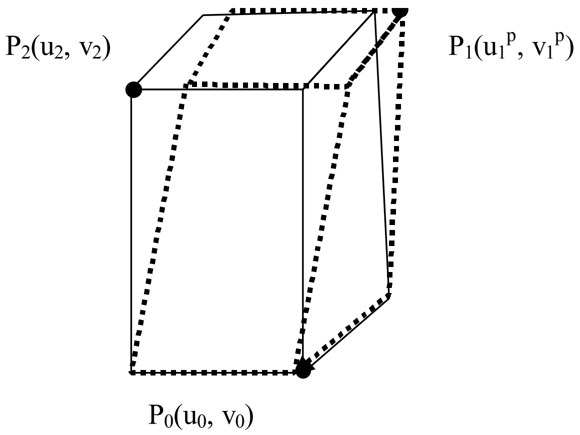


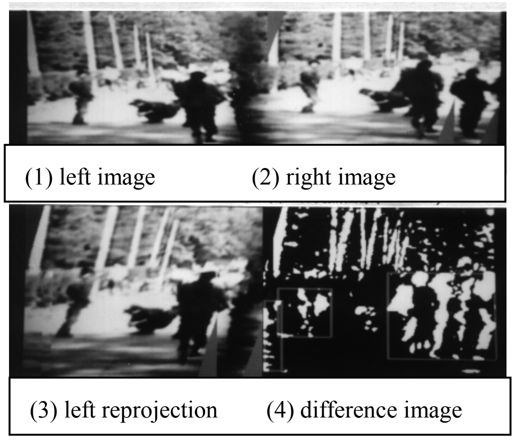
Fig. 4. 3D interpretation for the obstacle region

4.4.2. 3D measurements using representative points

Each obstacle region O is the result of binocular difference. In order to avoid the difficult correspondence problem, three representative points are selected to roughly estimate the 3D parameters (location, width and height) of the obstacle. Fig. 4 shows the 3D interpretation for the obstacle region in the difference image. The lowest point $P_0(u_0, v_0)$ of the region O is considered as the ground point. In practice it may be slightly high up from the ground due to the zero-disparity filtering process. This factor can be included to estimate the 3D parameters of the obstacle more accurate. The up-left point $P_2(u_2, v_2)$ is the up-left contour point of the obstacle in the image of the right camera, while the up-right point $P_1\left(u_1^p, v_1^p\right)$ is reprojected from the up-right contour point of the obstacle in the image of the left camera. *P₁* should be projected back to the original left image coordinates (u_1, v_1) by an inverse gaze transformation before the 3D computing. The 3D coordinates of these three points (x_i, y_i, z_i)(i=0,1,2) can be easily calculated by using the relations between the cameras' and the robot's coordinate system. The location, height and width can be derived then.

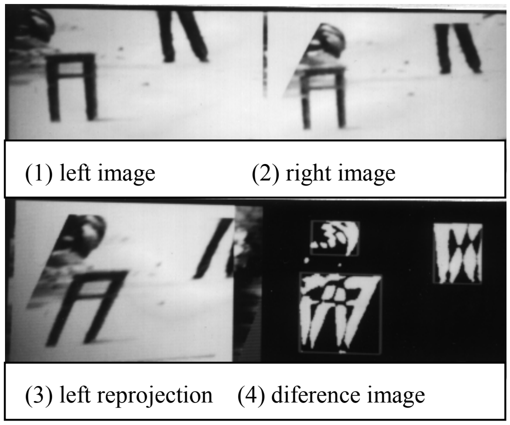
4.4.3. Real-time implementation

The RTSOD algorithm is implemented in a PC 486/66MHz with a frame grabber PC-Video 100. For a pair of 256*256*8 gray level image, the total time for a detection period is about 0.5-0.6 seconds, among which 250 ms is for 512*512 image capture and data transfer, 110ms for the gaze transformation, image difference and non-zero disparity filtering, 130-230ms (which varies with number of regions) for region grouping, and 10ms for parameter calculation and image display. The total processing time is reduced under 0.4 seconds in a Pentium/75MHz PC with the same frame grabber. The gaze transformation is carried out by using look-up tables of two-value functions $u_2 = U(u_1, v_1)$ and $v_2 = V(u_1, v_1)$ in case that the vehicle moves smoothly and the normal vector (p, q, r) of the ground plane keeps unchanged. The system still works properly if the normal vector (p, q, r) does not change severely. The work condition is analyzed in the following subsection and the algorithm has been developed for the case when the bump of the vehicle cannot be neglected [9].


(1) left image (2) right image

(3) left reprojection (4) difference image

Fig. 5. Example 1: heavy shadows



(1) left image (2) right image

(3) left reprojection (4) diference image

Fig. 6. Example 2: multiple obstacles

Fig.5-7 give some of the examples for the obstacle detection in the outdoor natural scene.. Fig. 5 shows the experimental results of the image in Fig. 2(b) where heavy shadows cover the road. Fig. 6 shows the results of the image in Fig. 3(b) where three obstacles are detected correctly. Fig. 7 shows one of the examples during the real-time operation of the RTSOD system.

4.4.4. Performance analysis

Detectable height: It is interesting to know the minimum height of an obstacle that the system can detect. Suppose at distance D, the detectable height of an obstacle is h. We will analyze the manifest of the point $(x, y, z) = (0, D, h)$ in the left, right and the reprojected images. Explicit results cannot be easily derived under the general case. Without the loss of generality, we suppose the optical axes of both left and right camera are parallel to the ground, and their image planes lie in the same plane ($T_z = 0$), then we have $R = I$, $T = (T_x, 0, 0)^t$. Assume the distance from each of the optical center to the ground is H. It can be proved that the "disparity" of the point $(0, D, h)$ in the right image and the reprojected left image is

$$\Delta u = du - du^0 = \frac{FT_x}{DH} h,$$

where du is the actual disparity of $(0, D, h)$ before the gaze transformation and du^0 is the disparity of the ground point $(0, D, 0)$. Hence we build up the relationship between the obstacle's detectable height h with the distance D, focal length F and camera's height H:

$$h = \frac{DH}{FT_x}\Delta u \qquad (16)$$

It should be noticed that the magnitude of "disparity" Δu is the measure of the distinctness of an obstacle. For example, given

$$D = 30m, H = 2.5m, T_x = 1m, F = \frac{256 pixel}{8.8mm} \times 25mm$$

if the minimum distinguished "disparity" is $\Delta u = 2\, pixel$, then the minimum height is about $0.20\, m$.

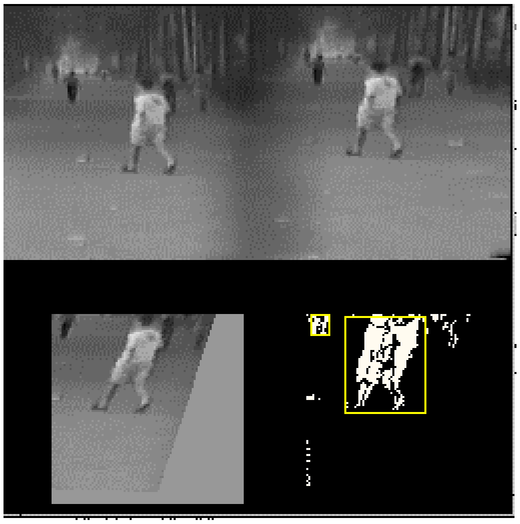
We have the following conclusions.

(1) The obstacle's detectable height is proportion to the image resolution of the discrete image. The smaller is the region in the difference image that can be distinguished, the lower is the detectable height of the obstacle.

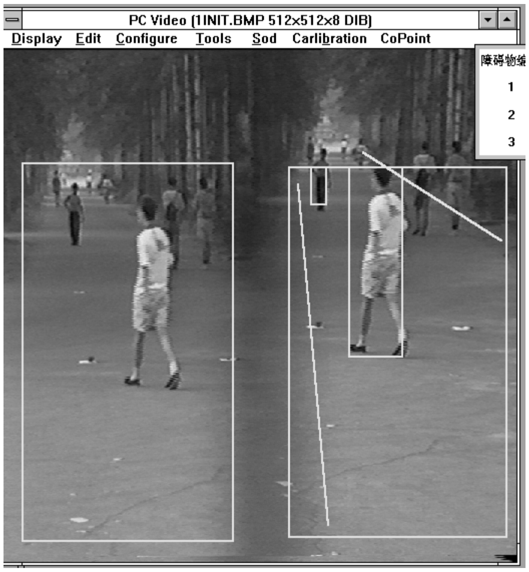
(2) For the same height h, the farther be the distance D, the smaller is the "disparity" Δu. Therefore small obstacle is easier to detect in near-distance.

(3) For the same height h, the disparity Δu increases by the increasing of the focal length and /or the baseline length T_x, however the (common) field of view of the two images is reduced.

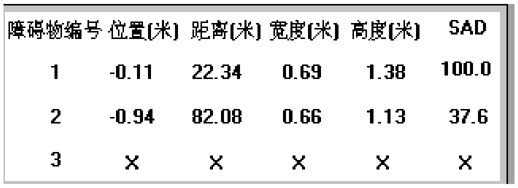
(4) It is interesting to find that the obstacle's minimum detectable height is also proportional to the camera's height H. The higher is the height H, the higher is the minimum detectable height. For this reason the camera should be installed as low as possible.



(a)Left image, right image, left reprojected image and obstacle regions in the difference image (two obstacle are detected)


PC Video (1INIT.BMP 512x512x8 DIB)
Display Edit Configure Tools Sod Carlibration CoPoint
障碍物编
1
2
3

(b)The results (two small rectangles) overlay in the original right image. The large rectangles in both the left and right images indicate the areas of interests (AOIs). Obstacles are detected in the intersection region of the two AOIs and within the road region indicated by two white lines in the right image



障碍物编号	位置(米)	距离(米)	宽度(米)	高度(米)	SAD
1	-0.11	22.34	0.69	1.38	100.0
2	-0.94	82.08	0.66	1.13	37.6
3	×	×	×	×	×

(c)Obstacle parameter list: Number, position (m), distance (m), width (m), height (m) and the belief measure (%) estimated from the sum of absolute difference (SAD) in each obstacle region. The actual distance and height of the first obstacle (a pedestrian) are 23.0 m and 1.7m respectively. The baseline is 0.34 m.

Fig. 7. An operation example of RTSOD system

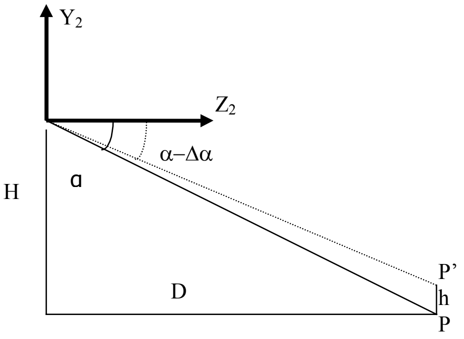


Fig.8. Tolerance of the vehicle's bump

Tolerance of the vehicle's bump: We also consider the simplified model, i.e. the optical axes of the cameras are parallel to the ground plane. For a ground point P at distance D, the angle between the projection ray O_2P and the optical axis Z_2 of the right camera is $\alpha = \tan^{-1}\frac{H}{D}$. Suppose the camera tilts down an angle $\Delta\alpha$ due to the bump of the vehicle, it is approximately equivalent to the situation that the point P moves up to point P' of height h, so that the angle between O_2P' and Z_z becomes $\alpha - \Delta\alpha$, therefore we have

$$h = H - D\tan(\alpha - \Delta\alpha) = D[\tan\alpha - \tan(\alpha - \Delta\alpha)]$$

The tolerance tilted angle range $\Delta\alpha$ can be estimated from $h \le h_{\min}$ (the minimum detectable height):

$$|\Delta\alpha| \approx |\sin\Delta\alpha| \le h_{\min}\cos^2\alpha = \frac{D}{D^2 + H^2}h_{\min} \qquad (17)$$

Given $H = 2.5m, D = 30m, h_{\min} = 0.40m$, we have $|\Delta\alpha| \le 0.8^\circ$. It means that if the tilted angle is larger than the tolerance, the figures (e.g. shadows on the road) on the road may be misdeemed as obstacles. However if orientation of the ground textures are perpendicular to the road direction (i.e. x axis of the camera), the negative effect of tilted bump of the vehicle will greatly reduced.

4.5. DISCUSSIONS AND CONCLUDING REMARKS

In the basic RTSOD algorithm, the operation condition is that the vehicle moves smoothly, i.e. the tilted angle due to the vehicle's bump is smaller than the tolerance value. So that the gaze transformation can be realized with look-up tables with known camera posture parameters. However mistakes may occur if the vehicle bumps severely on a textured road surface. In this case the tilted angle should be measured whenever necessary in order to modify the parameters of the gaze transformation. We have proposed the dynamic gaze transformation algorithm, which can be implemented in near real time [9]. The basic idea is to establish the relation between the tilted angle and the gaze parameters, and find the best parameters for zero-disparities of the ground plane.

The single-camera stereo equipment and the obstacle detection system can be used in smart cars, traffic monitoring, surveillance, virtual reality and medical imaging. The novel gaze transformation is a powerful pre-processing method for 3D scene modeling[10] and motion analysis. It can greatly simplify the correspondence problem of stereo vision and the tracking problem in visual motion.

7. REFERENCES

[1] Thorpe C, et al. Vision and Navigation for the CMU Navlab. *IEEE Trans. PAMI*, 1988, 10(3): 362 ~ 376

[2] Kayaalp A, Eckman J L, Near real-time range detection using a pipeline architecture, *IEEE Trans SMC*, vol 20, no 6, Nov/Dec 1990.

[3] Coombs D and Brown C, Real-time binocular smooth pursuit, *Int J Computer Vision*, vol. 11, no. 2, 1993: pp 147-164.

[4] Badal S, Ravela S, Draper B, Hanson A, A practical obstacle detection and avoidance system, In: *Proc. IEEE Workshop on Applications of Computer Vision*, 1994.

[5] Storjoham K, et al, Visual obstacle detection for automatically guided vehicles. In: *Proc IEEE ICRA*, pp761-766, 1990.

[6] Zhu Z G. Lin X Y, Realtime algorithms for obstacle avoidance by using reprojection transformation. In: *Proc IAPR Workshop on Machine Vision Application*, 1990. 393-396

[7] Cho Y C, Cho H S, A Stereo Vision-based Obstacle Detection method for mobile robot navigation, *Robotica*, 12,1994:203-216

[8] Shi D J, Zhu Z G, Chen Y K, Real-time stereo imaging equipment, *Chinese Novel Practical Patent* No ZL95211525.5, 1995.12.

[9] Zhu Z G, *Environment Modeling for Visual Navigation*, Ph.D. Dissertation , Tsinghua University, May 1997 (in Chinese)

[10] Lin X Y, Zhu Z G, Den W, "A Stereo Matching Algorithm Based on Shape Similarity for Indoor Environment Model Building", *Proc.* 13th *IEEE International Conference Robotics and Automation*, 1996: pp 765-770

Appendix 4.1. Gaze Transformation Geometry

[Proof]：

$$\begin{aligned} \boldsymbol{P'}_{\boldsymbol{2}} &\cong \boldsymbol{AP_2} = \boldsymbol{AR(P_1 - T')} \\ &= \left[\boldsymbol{I} + \alpha \boldsymbol{T'} \boldsymbol{N}^{\boldsymbol{T}}/d\right]\boldsymbol{R}^{-1}\boldsymbol{R}\left(\boldsymbol{P_1} - \boldsymbol{T'}\right) \\ &= \boldsymbol{P_1} - \boldsymbol{T'} + \alpha \boldsymbol{T'} \boldsymbol{N}^{\boldsymbol{T}} \boldsymbol{P_1}/d - \alpha \boldsymbol{T'} \boldsymbol{N}^{\boldsymbol{T}} \boldsymbol{T'}/d \\ &= \boldsymbol{P_1} + \left(-\boldsymbol{1} + \alpha\, \boldsymbol{N}^{\boldsymbol{T}} \boldsymbol{P_1}/d - \alpha\, \boldsymbol{N}^{\boldsymbol{T}} \boldsymbol{T'}/d\right)\boldsymbol{T'} \end{aligned} \qquad \text{(A4.1-1)}$$

where both $\mathbf{N}^{\mathrm{T}}\mathbf{P_1}$ and $\mathbf{N}^{\mathrm{T}}\mathbf{T'}$ are scalar values. Let

$$\beta = -1 + \alpha\, \boldsymbol{N}^{\boldsymbol{T}} \boldsymbol{P}_1/d - \alpha\, \boldsymbol{N}^{\boldsymbol{T}} \boldsymbol{T'}/d$$

Since $\alpha = d/\left(d - \boldsymbol{N}^T\boldsymbol{T'}\right)$， we have

$$\beta = \text{-}1 + \frac{\boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{P_1}}{d - \boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{T'}} - \frac{\boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{T'}}{d - \boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{T'}} = \frac{-\left(d - \boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{P_1}\right)}{d - \boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{T'}} \overset{\Delta}{=} -\frac{h}{H} \qquad \text{(A4.1-2)} \ \square$$

Appendix 4.2. Stereo Disparity in Gaze Transformation

(1) When $T_z \neq 0$, we have

$$\begin{aligned} \boldsymbol{p}_1 - \boldsymbol{p'}_1 &= \frac{f}{z_1}\boldsymbol{P}_1 - \frac{f}{z'_1}\boldsymbol{P}'_2 \\ &= \frac{f}{z_1}\boldsymbol{P}_1 - \frac{f}{z_1 + \beta T_z}\left(\boldsymbol{P}_1 + \beta \boldsymbol{T'}\right) \\ &= \frac{\beta T_z}{z_1 + \beta T_z}\frac{f}{z_1}\boldsymbol{P}_1 - \frac{fT_z}{z_1 + \beta T_z}\frac{f}{T_z}\boldsymbol{T'} \\ &= \frac{\beta T_z}{z_1 + \beta T_z}\left(\boldsymbol{p}_1 - \boldsymbol{t'}\right) = \frac{hT_z}{hT_z - Hz_1}\left(\boldsymbol{p}_1 - \boldsymbol{t'}\right) \end{aligned}$$

(2) When T_Z=0 we have z_1=z'_1 , therefore

$$\boldsymbol{p}_1 - \boldsymbol{p'}_1 = \frac{f}{z_1}\boldsymbol{P}_1 - \frac{f}{z'_1}\left(\boldsymbol{P}_1 + \beta \boldsymbol{T'}\right) = -\beta\frac{f}{z_1}\boldsymbol{T'} = \frac{fh}{Hz_1}\left(T_x, T_y, 0\right)^T \ \square$$