

Vision-based Real-time Obstacles Detection and Tracking for Autonomous Vehicle Guidance

Ming Yang*, Qian Yu, Hong Wang, Bo Zhang

State Key Laboratory of Intelligent Technology and Systems, Tsinghua University, CHINA

ABSTRACT

The ability of obstacles detection and tracking is essential for the safe visual guidance of autonomous vehicles, especially in urban environments. In this paper, we first overview different plane projective transformation (PPT) based obstacle detection approaches under the planar ground assumption. Then, we give a simple proof of this approach with relative affine, a unified framework that includes the Euclidean, projective and affine frameworks by generalization and specialization. Next, we present a real-time hybrid obstacle detection method, which combined the PPT based method with the region segmentation based method to provide more accurate locations of obstacles. At last, with the vehicle's position information, a Kalman Filter is applied to track obstacles from frame to frame. This method has been tested on THMR-V (Tsinghua Mobile Robot V). Through various experiments we successfully demonstrate its real-time performance, high accuracy, and high robustness.

Keywords: Obstacle detection, stereo vision, plane projective transformation, relative affine, watershed, Kalman filter

1. INTRODUCTION

The ability to obstacles detection and tracking is essential for the safe vehicle guidance of autonomous vehicles, especially in urban environments. In the early research of outdoor mobile robot, people used 3D laser radars to solve this problem [1,2]. Although 3D laser radars have very high accuracy, they're too expensive and the speed for data acquisition is too slow to satisfy the real time requirements for outdoor mobile robots. Besides that, the computation for detecting obstacles in 3D range data is also very high. In recent years, 2D Time-Of-Flight (TOF) laser radars are widely used instead of 3D laser radars [3]. 2D laser radars have faster acquisition speed, lower cost, and lower computation than 3D laser radars while the accuracy remains the same. However, the price of 2D laser radars is still too high and 2D information is not enough in many cases. Furthermore, laser radars are sensitive to the reflection of the object's surface, especially in outdoor environment.

Since camera is much cheaper and capable to provide much richer visual information than laser radars, numerous vision based obstacle detection approaches have been proposed in recent decades, and some of them have been successfully demonstrated [4,5]. Most monocular vision based methods usually detect specific obstacles according to the characters of obstacles, such as intensities, shape, etc. Therefore, they are sensitive to the road texture on the ground since no depth information is available unless the time-consuming ego-motion estimation is taken when the vehicle is moving. Compared with the monocular vision based method, the stereo vision based method is able to utilize the constant geometric relationship between a pair of images from the priori camera calibration. Traditional stereo vision is a reliable technique to extract 3D information from 2D images and is applicable to visual guidance. However, the 3D world reconstruction is based on correspondences searching between two stereo images, which require extremely powerful computational engines and cannot satisfy the real-time requirement, which is the most important to vehicle guidance. Furthermore, this method requires metric camera calibration to determine the camera intrinsic parameters (such as focal length, aspect ratio, lens effects, and nonlinearities in the optics or sensor) and extrinsic parameters (such as position, orientation) using calibration target whose 3D shape is strictly known, which demands considerable time and labor.

Assuming the ground is planar, if an obstacle is defined as object rising out from the ground plane, the problem of obstacle detection can be reduced to the problem of free-space detection, which can be simply detected by *Plane Projective Transformation*, and the complete 3D world reconstruction is no longer required. This paper presents a real-

* yangming97@mails.tsinghua.edu.cn; phone (86-10)62774276; <http://www.lits.tsinghua.edu.cn/ym/>; 30#416, Tsinghua University, Beijing 100084, CHINA

time hybrid obstacle detection method with a pair of images acquired by stereo cameras installed on a vehicle. In this method, a special plane projective transformation – *Homography Transformation* is first applied to warp left image into right image. Comparing the right image with the warped left image, points in non-free-space (i.e. obstacles area) will show disparities due to the obstacles' heights, while those on the free-space will not. Therefore, thresholding the differential image between right image and warped left image, we get the free-space image. After projecting this image onto the ground plane by an *Inverse Perspective Mapping*, we use two histograms to locate the obstacles. Taking this result as initial localization, we segment the original image with a Watershed transformation based algorithm to refine the localization. At last, with vehicle's position information from two optical encoders and a magnetic compass, a Kalman filter tracks obstacles from frame to frame. This method has been tested on THMR-V (Tsinghua Mobile Robot V), an autonomous vehicle developed by Tsinghua University, and demonstrates its real-time performance, high accuracy, and high robustness.

The rest of this paper is organized as the follows: Section 2 overviews the plane projective transformation based obstacle detection approach, and gives a simple proof based on *Relative Affine*; Section 3 describes the obstacle's rough and refined localization, and obstacle tracking in our method in detail. Section 4 gives experimental results and section 5 ends the paper with conclusions and description of future work.

2. PLANE PROJECTIVE TRANSFORMATION BASED OBSTACLE DETECTION

2.1 Plane Projective Transformation

Plane projective transformation (PPT) is the most general perspective transformation between two planes as shown in Fig. 1. A plane projective matrix $\mathbf{H}_{3 \times 3}$ can be used to present this transformation as the follows,

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \cong \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

where, $[x, y, 1]^T$ and $[x', y', 1]^T$ are the homogeneous coordinates of two correspondence point in two planes, respectively; $\cong$ is the equality up to scale. In order to eliminate the unknown scale factor, let one of the elements in $\mathbf{H}_{3 \times 3}$ to have a certain value, for example, $h_{33}=1$. From Eq. (1) we can find each point-correspondence can provide two linear equations. Since there are 8 unknowns in $\mathbf{H}_{3 \times 3}$, at least 4 non-collinear correspondences are needed to compute $\mathbf{H}_{3 \times 3}$. Although 4 correspondences are enough for solution, in practice tens of correspondences are used to reduce the impact of noise with a least-square method.

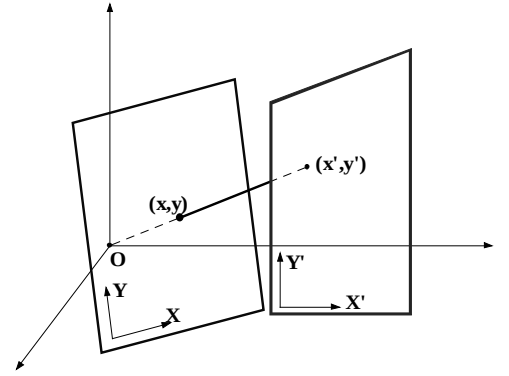


Figure 1. Plane projective transformation

2.2 PPT based Obstacle Detection

Assuming the ground is planar and defining obstacle as object rising out of the ground plane, we can make use of plane projective transformation to detect obstacles if they have different intensities again the ground. The Inverse Perspective Mapping (IPM) is a special case of plane projective transformation, which is the transformation between the ground plane and the image plane, and can be used to remove the perspective effect [6]. Let $P=[x, y, 1]^T$ be the homogeneous coordinates of a point on the ground plane, $p_l=[u_l, v_l, 1]^T$ and $p_r=[u_r, v_r, 1]^T$ be the homogeneous coordinates of its projections on the left and right original images, respectively. The IPMs between the ground plane and two image planes are described as,

$$P \cong H_l p_l \quad (2)$$

$$P \cong H_r p_r \quad (3)$$

where H_l and H_r are the projective matrix of these two IPMs. Applying these two IPMs to the original stereo images, we get two ground images representing the texture of the road surface as if they were observed from the top. Any difference in these two ground images represents a deviation from the assumption of ground plane, and thus indicates a non-free space, i.e. obstacle.

Let's consider a simple example as shown in Fig. 2. Given a point Q on the ground plane, by IPMs its corresponding pixels q_l and q_r in the left and right original images will be projected to the same location in the left and right ground images, respectively. Since the intensities of q_l and q_r are identical, the intensities of their projected pixels in both ground image are also identical, which shows zero in the differential image. On the other hand, for a point P above the ground plane, its corresponding pixel p_r in the right image is projected to P' in the right ground image. However, in the left ground image, P' is given rise by another pixel p'_l in left image, which corresponds to P' on the ground plane. If P and P' have different contents, then p_r and p'_l have different intensities, which leads to that the intensities of left and right ground images at P' are different. Thus, we can detect obstacles by finding those pixels that have different intensities in the left and right ground images with a threshold.

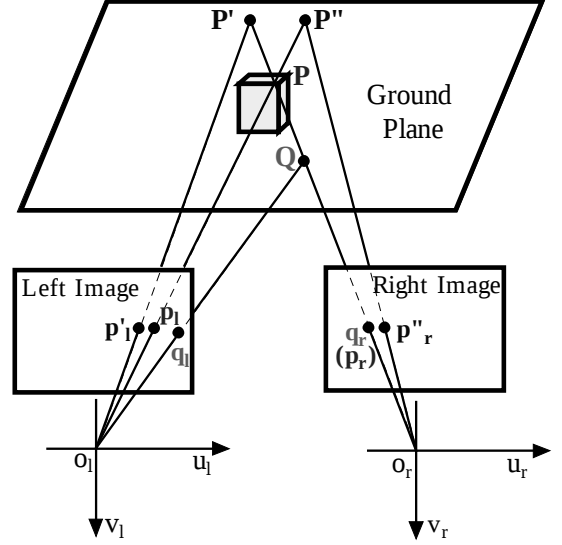


Figure 2. IPM based obstacle detection

IPM based obstacle detection has been successfully demonstrated on ARGO vehicle [7] and THMR-V vehicle [8]. In practice, the assumption of planar ground is often violated in an outdoor environment due to the un-flat ground and the vehicle's movement. Unfortunately, this method is sensitive to the error of the camera geometry with respect to the ground plane, which results in that some areas on the ground are wrongly detected as obstacles. In order to solve this problem, Onoguchi proposed a Planar Projection Stereopsis (PPS) method [9]. PPS examines the shapes of the obstacles and eliminates falsely detected obstacles by using the following properties: obstacles are projected to the shapes falling backward from the location where the obstacles touch the ground plane; and the length of shapes falling backward depends on the location of obstacles in relation to the stereo cameras and the height of obstacles in relation to the ground plane. Bertozzi presented an extension to the IPM [10], which made use of the features like lane markings on the ground plane to estimate the slope of the ground plane.

Combining Eq. (2) and (3), we get

$$p_r \cong H p_l, \quad H = H_r^{-1} H_l \quad (4)$$

where H is the homography matrix between two image planes with respect to the ground plane. For a point P on the ground plane, homography relates its projection p_l in left image to its projection p_r in the right image. Compared with IPM based method, homography-based obstacle detection computes the homography matrix by weak calibration instead of metric calibration, which allows it to dynamically update homography matrix using the correspondences on the ground plane. Zhu Zhigang presented a reprojection transformation based obstacle detection method [11], in which several candidate homography matrices were computed according to different pitch angles in advance. The final homography matrix is selected as the matrix, by which the difference between right image and warped left image in ground plane area is minimal. Hattori introduced a pseudo-projection camera model that provided a good approximation to the projective camera model in road scenes [12]. Under this model, the homography matrix can be approximated to an affine transformation matrix, which can be computed easily from two or more lines. Besides that, in IPM based method, the differential image is sensitive to the noise in original images, especially at distant points, since the pixels at different distance have different weights due to the perspective effect. However, in homography-based method don't have this problem since the difference is taken on image plane and all the pixels have same weights.

2.3 A Simple Proof

In this subsection, we present a simple proof of homography-based obstacle detection by a unified relative affine framework proposed by Shashua [13], which includes by generalization and specialization the Euclidean, affine and projective framework in a natural and simple way. Under this framework, some vision related results are often reduced to a single-line proof. The affine invariant -- relative affine structure is a generalization of the affine structure under parallel projection and is a specialization of the projective structure, which can be described as a ratio of two relative affine structures. In relative affine framework, there are two important theorems:

THEOREM 1(Relative Affine Structure): Let π be some arbitrary plane and let $P_j \in \pi$, $j=1,2,3$ projecting onto p_j, p'_j in views ψ, ψ' , respectively. Let $p_0 \in \psi$ and $p'_0 \in \psi'$ be projections of $P_0 \notin \pi$. Let $v \in \psi$, $v' \in \psi'$ be the epipoles. Let H be a homography determined by the equations $Hp_j \equiv p'_j$, $j=1,2,3$, and $Hv \equiv v'$, scaled to satisfy the equation $p'_0 \equiv Hp_0 + v'$. Then, for any point P projecting onto $p \in \psi$ and $p' \in \psi'$, we have

$$p' \equiv Hp + kv' \quad (5)$$

The coefficient $k=k(p)$ is independent of ψ' , i.e., is invariant to the choice of the second view.

THEOREM 2: Let π denote a reference plane and $P_0 \notin \pi$ be some fixed point, and P be a point whose corresponding points p, p' in both view satisfy Eq. (5). Let d, d_0 denote the perpendicular distances of P and P_0 from π , and let z, z_0 denote the distances of P and P_0 from the first camera center O . Then the parallax term k is:

$$k = \frac{z_0}{z} \frac{d}{d_0} \quad (6)$$

Next, we give the description of homography-based obstacle detection, and prove it with above two theorems.

PROPOSITION: Let H be the homography matrix between left and right image planes (ψ_l, ψ_r) with respect to the ground plane. Let $I(p)$ be the intensity of a pixel p in the image. Let ψ'_l be the warped left image by the homography transformation. Let $p_r(x_r, y_r) \in \psi_r$ and $p'_l(x'_l, y'_l) \in \psi'_l$ be the pixels at the same location, and P be the correspondence point of p_r in the world. Suppose that different pixels have different intensities. The necessary and sufficient condition that P belongs to obstacles is $I(p_r) \neq I(p'_l)$.

PROOF: Let's consider the sufficiency first, i.e. if $I(p_r) \neq I(p'_l)$, then P belongs to obstacles. We can construct a proof by contradiction for the assertion above. Let $p_l(x_l, y_l)$ be the projection of P on the left image plane. We begin by assuming that P doesn't belong to obstacle, i.e. P belongs to the ground plane, then we have $p_r \equiv Hp_l$ and $I(p_r) = I(p_l)$ since both p_r and p_l are projections of P in the world. Let $p'_l(x'_l, y'_l) \in \psi'_l$ be the projection of p_l by the homography transformation, i.e. $p'_l \equiv Hp_l$ and $I(p'_l) = I(p_l)$. Then, from above we get $p'_l \equiv Hp_l \equiv p_r$ and $I(p'_l) = I(p_l) = I(p_r)$. For $p_r = p'_l$, $p'_l = p_r = p'_l$. Since both p'_l and p'_l are on the warped left image, $I(p'_l) = I(p'_l) = I(p_r)$, this yields a contradiction with the precondition $I(p_r) \neq I(p'_l)$. Therefore, we conclude that the assumption doesn't hold true.

Next, let's consider the necessity, i.e. if P belongs to obstacles, then $I(p_r) \neq I(p'_l)$. As mentioned above, obstacle is defined as object rising out from the ground plane, i.e. $d \neq 0$. Since z and d_0 are limited, the relative affine structure $k \neq 0$ (THEOREM 2). Let $v_r \neq 0$ be the epipole in right image, we get $p_r \equiv Hp_l + kv_r$ (THEOREM 1). Since $v_r \neq 0$ and $k \neq 0$, it follows that $kv_r \neq 0$ and $p_r \neq Hp_l$. Given $p'_l = Hp_l$, we have $p_r \neq p'_l$, and since $p'_l = p_r$, we get $p'_l \neq p'_l$. According to that different pixels have different intensities, $I(p'_l) \neq I(p_r)$. And given $I(p_r) = I(p_l)$, therefore, $I(p_r) \neq I(p'_l)$. $\square$

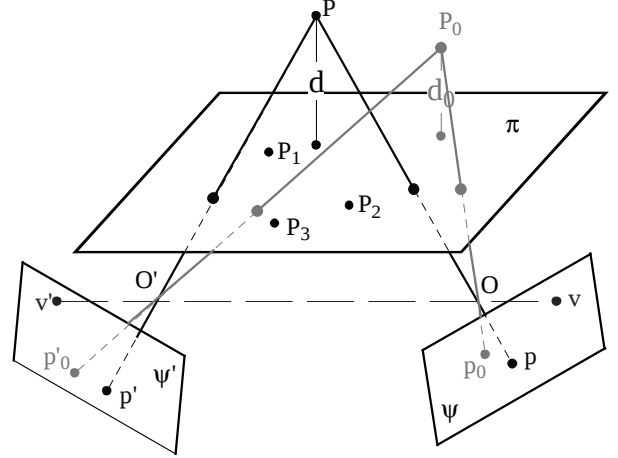


Figure 3. Relative affine structure

3. OBSTACLES LOCALIZATION AND TRACKING

Using PPT based obstacle detection, we could get a free-space image, in which each pixel is labeled as free-space or obstacles. For IPM based method, this free-space image is on the ground plane. For homography-based method, this image is on the right image plane and can be transformed to the ground plane by an IPM. In this section, we will discuss the problems of obstacles localization and tracking in ground free-space image.

3.1 Rough Localization of Obstacles

Bertozzi demonstrated that an ideal square homogeneous obstacle will produce two clusters of pixels with triangular shape in the ground free-space image due to the different coordinates of the projections C_l and C_r of the two cameras as shown in Fig. 4. Therefore, obstacle localization is reduced to the localization of these triangles, which can be solved by a polar histogram and a radial histogram [7].

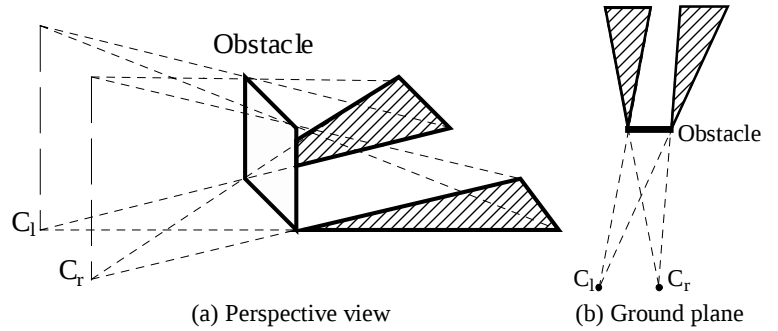


Figure 4. Relationships between obstacle and cameras

A polar histogram can be used to determine the orientation of an obstacle by detecting its triangles in the free-space image. Given a focus point on the ground plane, the polar histogram is computed by scanning the free-space image and counting the number of over-threshold pixels for every straight line originating from the focus. Since the polar histogram presents an appreciable peak corresponding to each triangle, the orientations of obstacles can be determined by using the amplitude and width of the peaks, as well as the interval between joined peaks. It is obvious that the corners of the triangles represent the obstacles' contact points on the ground plane. Thus, a radial histogram can be used to determine the obstacle's distance. For each peak in the polar histogram, the radial histogram is computed by scanning a specific sector of the free-space image; whose width is determined as a function of the width of the peak in the polar histogram. The number of over-threshold pixels lying in the sector is computed for every discrete distance from the focus and the result is normalized. Applying a threshold to the radial histogram allows the determination of the position of the triangle's corner, thus obtaining the obstacle distance.

3.2 Refined Localization of Obstacles

However, in practice these triangles in the free-space image are not so clearly defined and often not distinctly disjointed because of the texture, irregular shape and homogeneous intensities of real obstacles. Although clusters of pixels having an approximate triangular shape are recognizable, the accuracy of localization is poor, especially when the obstacle is distant since the pixels at the top of the image correspond to larger area on the ground plane than those at the bottom.

In order to improve the accuracy of localization, we present a hybrid obstacle detection method, which combines the PPT-based method with region segmentation based method. Although the former has lower accuracy in localization, it is fast and able to distinguish ground regions with textures from obstacles, which are main problems with the latter. Compared with PPT-based method, the region segmentation based method has higher localization accuracy since it is working in image plane and not affected by the violation of ground plane assumption. Furthermore, region segmentation based method can provide information about obstacle's shape, which is unavailable in PPT-based method.

In this hybrid method, we first detect obstacles with the homography-based method, then with a polar histogram and a radial histograms, we obtain the rough locations of obstacles. Next, we use a region segmentation based method to refine the obstacles' localizations. In order to reduce the computation involved in region segmentation, a Region-Of-Interest (ROI) is first computed for each obstacle according to the rough locations. After segmenting the image in ROI, we selected the regions belong to obstacles according to shape and locations of regions, the rough locations, etc. In the end, the selected regions are merged and a bounding box is computed, which indicates the shapes and locations of obstacles.

In region segmentation, we adopt a watershed transformation based segmentation method developed in the framework of mathematical morphology [14]. In watershed transformation, any intensity image is considered as a topographic surface. If we flood this surface from its minima and, if we prevent the merging of the waters coming from different sources, we partition the image into

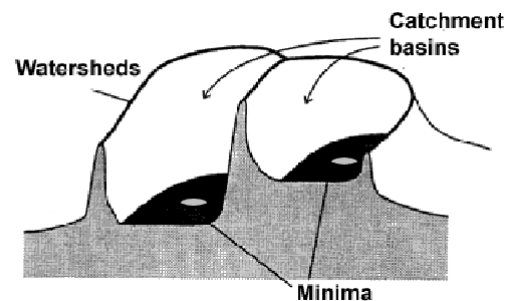


Figure 5. Watershed transformation

two different sets: the catchment basins and the watershed lines as shown in Fig. 5. If we apply this transformation to the gradient magnitude image, the catchment basins should theoretically correspond to the homogeneous intensity regions of this image. In practice, this transformation produces an over-segmentation due to noise or local irregularities in the gradient image. A marker-controlled watershed or a hierarchical segmentation with two successive watershed transformations can be used to solve this problem. Watershed transformation based segmentation has the advantages of high efficiency and high accuracy. Furthermore, it can always provide closed contours by construction and the segmentation is independent of shape and size.

3.3 Obstacles Tracking

Although above proposed method can refine localization of obstacles, in practice some obstacle may be lost in the detection, or one obstacle may be detected as two or more obstacles close to each other. Considering the common constant acceleration motion, we apply a Kalman filter to solve the reliability problem and further improve the accuracy.

In the above method, the positions of obstacles are represented in vehicle coordinates. Since the tracking is carried out in the world coordinates, we must transform them to the world coordinates. This requires the vehicle's position information. Given the vehicle's position, the transformation between vehicle coordinates and world coordinates can be described as the follows,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta_k & -\sin \theta_k \\ \sin \theta_k & \cos \theta_k \end{bmatrix} \begin{bmatrix} x_v \\ y_v \end{bmatrix} + \begin{bmatrix} x_k \\ y_k \end{bmatrix} \quad (7)$$

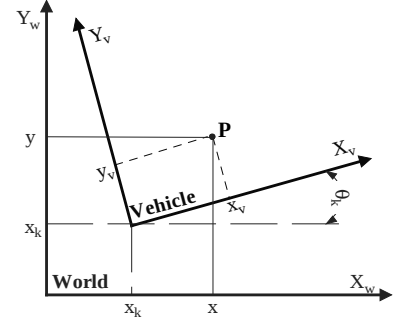


Figure 6. Relationship between world coordinates and vehicle coordinates

where, $[x_v, y_v]^T$ is the obstacle's position in vehicle coordinates, $[x, y]^T$ is the obstacle's position in world coordinates, (x_k, y_k, θ_k) is the vehicle's position in world coordinates at time instance k .

Let the measurement vector be the obstacle's position $z_k = [x, y]^T$. The state of the Kalman filter can be described by a six dimensional vector $x_k = [x', y', v_x, v_y, a_x, a_y]^T$, where $(x', y')^T$, $(v_x, v_y)^T$, and $(a_x, a_y)^T$ are the position, velocity and acceleration of the obstacle, respectively. All above variable are defined with respect to the world coordinates. The state equation and measurement equation of a Kalman filter can be described as

$$\begin{cases} x_k = AX_k + w \\ z_k = HX_k + v \end{cases} \quad (8)$$

where, the subscript k denotes the time instance; A and H are state transition matrix and measurement matrix, respectively. For a constant acceleration motion, they can be defined as the follows,

$$A = \begin{bmatrix} 1 & 0 & t_k & 0 & 0.5t_k^2 & 0 \\ 0 & 1 & 0 & t_k & 0 & 0.5t_k^2 \\ 0 & 0 & 1 & 0 & t_k & 0 \\ 0 & 0 & 0 & 1 & 0 & t_k \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (9)$$

where t_k is the time interval between time instance $k-1$ and k . In Eq. (8), the random variable w and v represent the state error and the measurement error, respectively, which are assumed to be independent of each other, white, and with normal probability distribution: $p(w) \sim N(0, Q)$, $p(v) \sim N(0, R)$, where Q and R are the status noise covariance and the measure noise covariance matrices, respectively. In practice, R can be measured prior to the operation of Kalman filter with some off-line sample measurements.

By using Kalman filter, we get an estimation of the obstacle's position according to the past states, which enables us to check the current detection result if any obstacle is lost. This could greatly increase the reliability of the detection and improves the accuracy at the same time.

4. EXPERIMENTS

The method proposed in this paper has been implemented in C++ on a commercial AMD 1.33Ghz PC with two cameras and a Matrox's Meter RGB/PPB digitizer. We have successfully run this system on THMR-V (Tsinghua Mobile Robot), an intelligent vehicle developed at Tsinghua University, and tested it within all kinds of environments. To compute the homography matrix and IPM matrices, a semi-hand calibration is performed using the feature points on the ground plane before experiments.

Fig. 7.a&b show the stereo images captured in a typical scene. By using the homography transformation, we warp left image to right image (Fig. 7.c). After thresholding the differential image between right image and warped left image, we get the free-space image (Fig. 7.d), in which the white pixels indicates the free-space while black pixels corresponds to the obstacles. Fig. 7.e shows the ground free-space image projected by an IPM. Then, a polar histogram and a radial histogram are applied to the ground free-space image to search the obstacle's location as the gray circle shown in Fig.

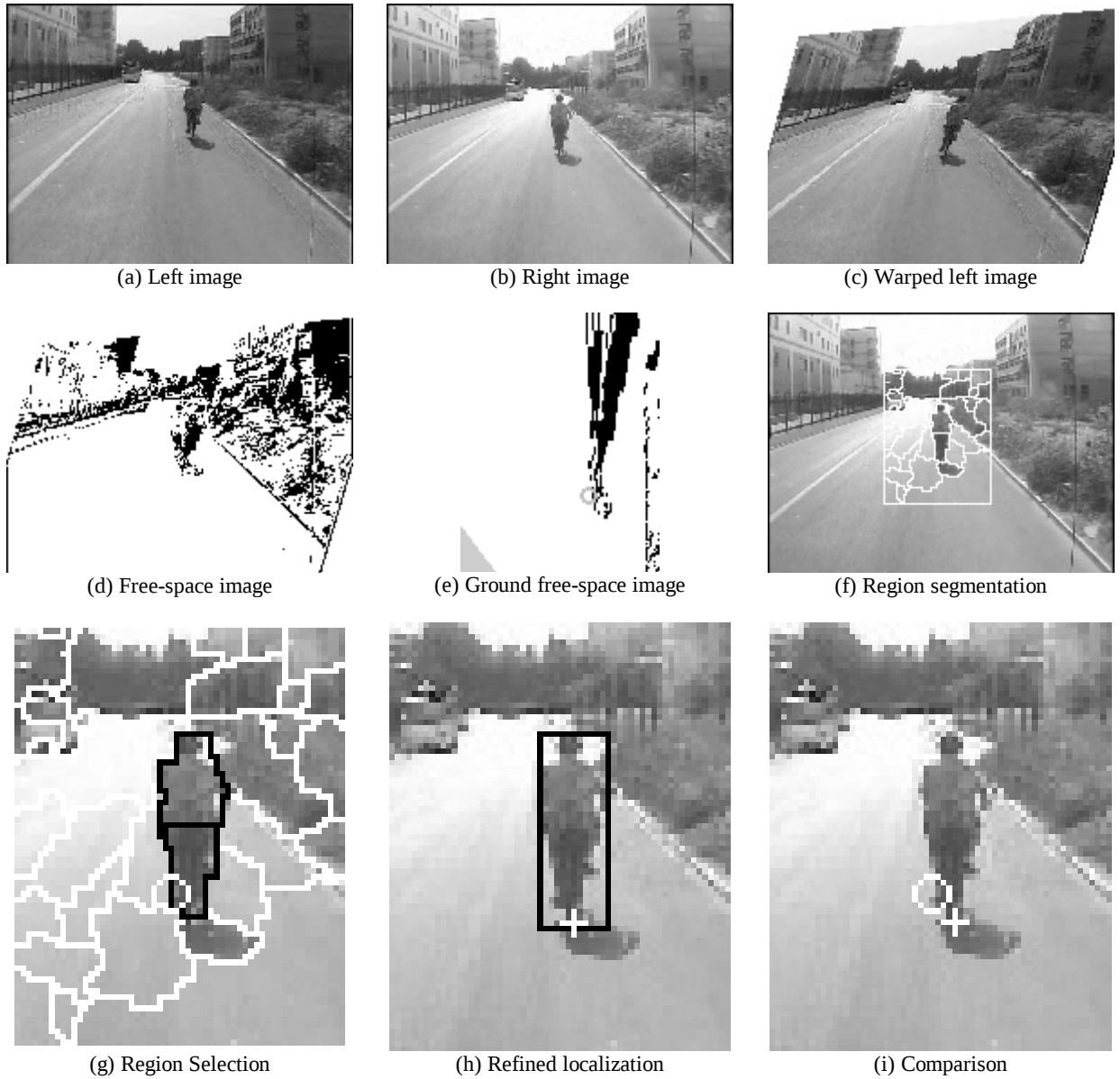


Figure 7 A typical example of obstacle detection

7.e. According to this rough location, a ROI is computed for the region segmentation to reduce the computation as the white bounding box shown in Fig. 7.f. In this ROI, watershed transformation based region segmentation is performed Fig. 7.f). A hierarchical segmentation with two successive watershed transformations is used here to solve the over-segmentation problem. Next, we select the regions belonging to the obstacle according to the rough location, region shape and location, etc. In Fig. 7.g., the edges of selected regions are marked in black. At last, by grouping the selected regions, we obtain a bounding box as the black box shown in Fig. 7.h, which indicates the shape and location (white cross) of the obstacle in image plane. Through the comparison in Fig. 7.i, we show that the combination of homography-based method and watershed region segmentation based method obviously improve the accuracy of localization, where the white circle and white cross represent the result from rough localization and refined localization, respectively. At the same time, this hybrid method doesn't have road texture problem with region segmentation based method. The road marking and shadow on the ground in this case are correctly detected as the free-space.

The obstacle tracking has also been tested with several stereo images sequences. Two typical tracking examples are presented here. A Kalman filter described as in section 3 is applied to track the obstacles from frame to frame. In order to test the performance of Kalman filter, we use the results of rough localization as the inputs. In the first example, the vehicle was still and two obstacles were moving straight in front of the vehicle. In the Fig. 8, we show the results of

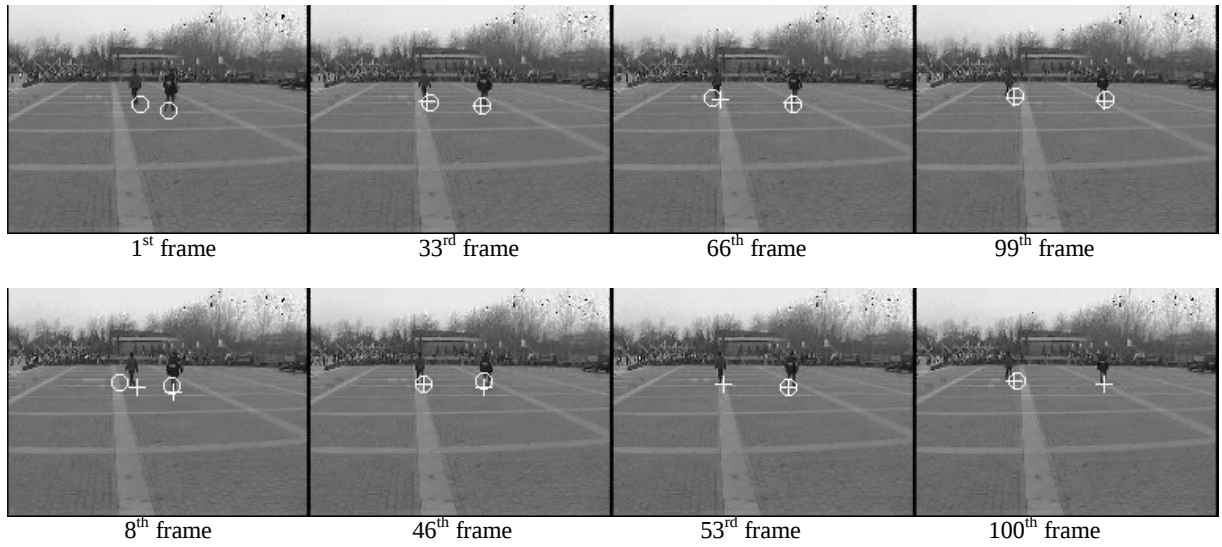


Figure 8 A sequence of obstacle detection and tracking

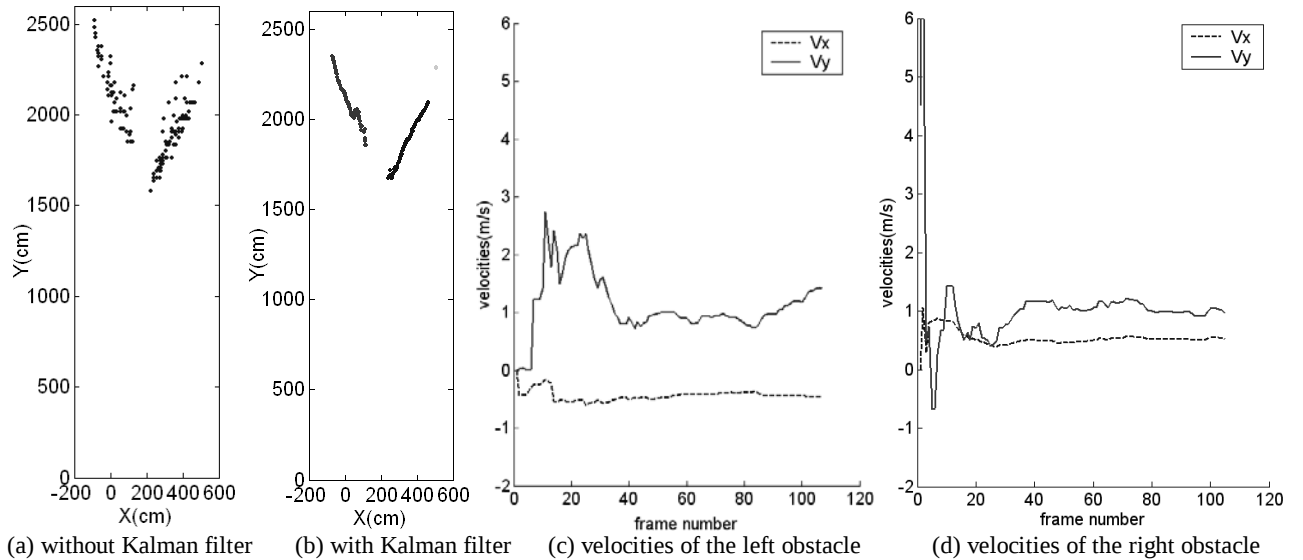


Figure 9 Trajectories of the 1st tracking example

tracking, where the white circle and white crossing represent the obstacles' locations without or with Kalman filter, respectively. Although sometimes the accuracy of obstacle's location is so poor, for example, the left obstacle in the 8th and 66th frame, and the right obstacle in the 8th and 46th frame, we still can successfully locate the obstacles accurately with Kalman filter, even if the obstacles are lost for a while, for example, the left obstacle in the 53rd frame and the right obstacle in the 100th frame. Similar results can also be found in Fig. 9.a&b, which are the trajectories of obstacles without or with Kalman filter, respectively. It is obvious that the trajectories with Kalman filter are much smoother than those without Kalman filter and exhibit clear straight trajectories. From Fig. 9.c&d, we find that the velocities of obstacles converge after about 20 frames.

In the second tracking example, the vehicle was not still any more. It first went straight for about 5 meters then turned left as shown in Fig. 10. At the same time, an obstacle was moving straight in front of the vehicle at similar direction. If without the vehicle's position information, the obstacle will be wrongly tracked as shown in Fig. 10.a. In our method, we use two optical encoders fixed on the rear wheels to provide translation information, and a magnetic compass to provide rotation information. Attaching the world coordinates to the initial position of the vehicle, we can obtain the vehicle's position in the world coordinates at each time instance by dead reckoning. With this vehicle's position, the obstacle is correctly tracked while the vehicle is moving as shown in Fig. 10.b.

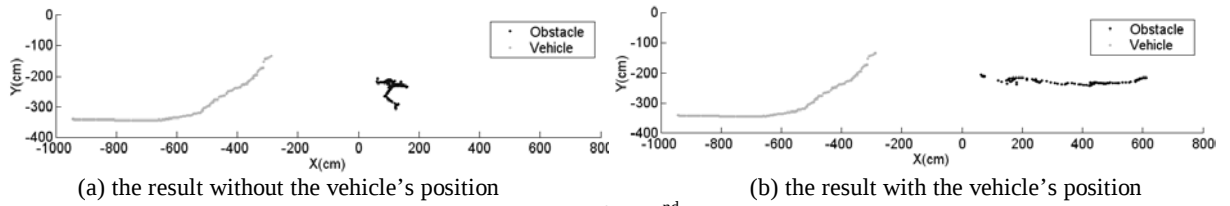


Figure 10 Trajectories of the 2nd tracking example

Although the proposed hybrid method consists of time-consuming region segmentation, it shows high real-time performance in all the experiments due to the use of ROI. For the above sequences, the whole processing takes no more than 30 ms per frame for 192x144 images with two obstacles and without any extra code optimization, in which near 4 ms for the rough localization including homography-based obstacle detection and histograms based localization, 13 ms for refined localization of each obstacle including watershed transformation based region segmentation, region selection and merging, and no more than 1 ms for tracking with Kalman filter. The real-time requirement is satisfied without any problem here.

5. CONCLUSIONS AND FUTURE WORK

This paper overviews the plane projective transformation based obstacle detection based on planar ground assumption, which includes IPM based method and homography (or reprojection) based method. A simple proof of sufficiency and necessity of this approach is given based on relative affine framework. Then we present a real-time hybrid obstacle detection method, which combines homography-based method with a Watershed region segmentation based method. Experimental results show that the combination of these two methods obviously improves the accuracy of obstacles' locations while the advantages of homography-based method are remained, such as low computation, robustness to road texture, etc. At last, with the vehicle's position information from optical encoders on the rear wheels and a magnetic compass, a Kalman filter is applied to track these obstacles from frame to frame, which improves the reliability of detection and accuracy of localization as shown in experiments on THMR-V.

Since the ground plane assumption is often violated in an outdoor environment in practice, we plan to dynamically compute the homography matrix using the features on the ground plane with a robust method. Since the vehicle's position has great influence on the results of tracking, vision based ego-motion estimation is considered to replace the optical encoders to improve the accuracy by using the features on the ground plane. Furthermore, a high accurate GPS is also considered in the future.

ACKNOWLEDGEMENTS

This research was supported in part by Chinese High Technology Development Program and Portugal-China Science and Technology Cooperation Project. We wish to thank Prof. Helder Araujo for his valuable comments and discussions, Yunpeng Cai, Jianye Lu for their implementation of Watershed algorithm, Changjiang Liang, Bin Dong for their collaboration in the video data acquisition on the THMR-V.

REFERENCES

1. C. Thorpe, M. Herbert, T. Kanade, and S. A. Shafer, "Vision and navigation for the Carnegie-Mellon Navlab", *IEEE Transaction on Pattern Analysis and Machine Intelligence* **10**, pp. 362-373, 1988.
2. Q. Zhang, W.K. Gu, "Real-time algorithms study and error analysis on obstacle detection using range image of laser range sensor in ALV", *Chinese Journal of Robot* **19**, pp. 122-128, 1997. (In Chinese)
3. M. Yang, H. Wang, K.Z. He, and B. Zhang, "Laser radar based environment modeling and obstacle avoidance of mobile robot", *Journal of Tsinghua University (Science and Technology)* **40**, pp. 112-116, 2000. (In Chinese)
4. M. Bertozzi, A. Broggi, and A. Fascioli, "Vision-based intelligent vehicles: State of the art and perspective", *Journal of Robotics and Autonomous Systems* **32**, pp. 1-16, 2000.
5. M. Yang, J.Y. Lu, H. Wang, and B. Zhang, "Vision based road following", *Chinese Journal of Pattern Recognition and Artificial Intelligence* **14**, 186-193, 2001. (In Chinese)
6. M. Bertozzi, A. Broggi, A. Fascioli, "Stereo inverse perspective mapping: theory and applications", *Image and Vision Computing* **16**, pp. 585-590, 1998
7. M. Bertozzi, A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection", *IEEE Transaction on Image Processing* **7**, pp. 62-81, 1998.
8. M. Yang, G. Li, H. Wang, and B. Zhang, "Vision-based real-time vehicle guidance on THMR-V, Part II: obstacle detection", In: *Proc. of the International Symposium on Test and Measurement* **2**, pp. 1371-1375, 2001
9. K. Onoguchi, N. Takeda, and M. Watanabe, "Planar projection stereopsis method for road extraction", *IEICE Transaction on Information and System* **E81-D**, pp. 1006-1018, 1998.
10. M. Bertozzi, A. Broggi, A. Fascioli, "An extension to the Inverse Perspective Mapping to handle non-flat roads", In: *Proceedings of the IEEE Intelligent Vehicles Symposium '98*, pp. 305-310, 1998.
11. Z.G. Zhu, X.Y. Lin, D.J. Shi, et al, "A real-time visual obstacle detection system based on reprojection transformation", *Computer Research and Development* **36**, pp. 77-84, 1999 (In Chinese)
12. H. Hattori, A. Maki, "Stereo without depth search and metric calibration", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 177-184, 2000.
13. A. Shashua, N. Navab, "Relative affine structure: canonical model for 3D from 2D geometry and applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**, pp. 873-883, 1996.
14. E.R. Dougherty, *Mathematical morphology in image processing*, Chapter 12. The morphological approach to segmentation: the watershed transformation, pp. 433-481, Marcel Dekker, Inc., New York, 1993.