Dynamic Obstacle Detection through Cooperation of Purposive Visual Modules of Color, Stereo and Motion

Zhigang Zhu , Guangyou Xu , Shaoyun Chen and Xueyin Lin

Department of Computer Science
Tsinghua University, Beijing 100084, China

Abstract

In this paper we present a new framework for detection of dynamic obstacles in the unstructured outdoor road environment by purposively integrating binocular color image sequence. In our system, color image segmentation, stereo obstacle detection, visual egomotion estimation,and moving object analysis, are all built-in task-oriented modules and hence are efficient and robust. Most of these functions can be performed in realtime. They are activated and integrated adaptively in the manner of neural parallel distributed processing(PDP). Experimental results are given to validate our philosophy of so-called purposive vision , retinal mapping , adaptive integration and parallel distributed processing.

1. Introduction

Computer vision and image understanding have been a research field of intensive studies. Yet in spite of the availability of complex image analysis systems and very powerful processing capability, it is doubtful whether anyone would claim that the problems of image analysis had been nearly solved. There are even suggestions that the work in machine vision has to some extent failed [1]. Biological vision systems have been studied extensively both from the neurophysical and psychological viewpoints for a long time, and purposive , active and selective functions of the biological vision systems have been emphasized [2] . Currently purposive and active vision strategy is considered to be a promising direction for machine vision applications [3]. The critical point is that the method for machine vision should be developed and scheduled in a task-directed manner, so that an efficient and effective system can be built to meet the basic requirement of the underlying tasks.

Stereo, motion and color are the most important cues for human perception and it is believed that human perception behaves as a parallel distributed processing in the sense that these cues work parallelly and cooperatively[4,5]. We argue that by using and combining these cues purposively and adaptively, we can make a machine understand its environment incrementally. In this paper we try to solve some difficult problems for the visual navigation of a mobile robot moving in an unstructured and time-varying outdoor road environment, where detection of static obstacles and moving objects on the road is vital for the visual navigation. There are at least three contributions made in this paper. First, the purposive vision strategy is emphasized so that any task is processed at the most relevant level of precision. Second, a PDP neural model is used to organize and integrate different aspects of visual cues. Third, a novel image mapping, gaze transformation, which is suggested by the neural retinal mapping of biological vision system , is presented to deal with shape recovery and depth perception problems in the roadway environment. We have implemented the above ideas in an experimental visual navigation system .

2. Purposive visual modules

In this section three visual modules are described: stereo Obstacle_Detector, color Road_Segmenter and correlation-based Egomotion_Estimator. Each module is purposively designed to solve a particular problem in the visual navigation.

2.1. Stereo Obstacle-Detector

Correspondence problem is one of the main obstacles for stereo methods in 3D recovery. However, if we use stereo methods in a purposive and qualitative manner for robot navigation, the underlying tasks can be fulfilled satisfactorily. Based this idea, we have designed a novel algorithm for realtime obstacle detection whose efficiency is guaranteed by the realtime gaze transformation of the stereo pair.

(1). Gaze transformation and gaze images : We, human, turn our head and eyes to gaze at the points of our interest, which is called "eye movement" and focus

1050-4729/94 $03.00 © 1994 IEEE

of attention. Similarly, when we rotate the camera so that its optical axis aims directly at a specific plane, e.g. the ground surface where the robot moves, it will facilitate the visual process relative to the plane. Here we use the term "gaze transformation" to imply that the attention is focused on the specific plane. It is believed that gaze transformation is one of the important retinal mapping of human vision. Suppose the camera coordinates (X_c,Y_c,Z_c) is transformed to the gaze coordinates (X_g,Y_g,Z_g) by a software gaze transformation (Fig. 1), which is represented by a rotation matrix $\mathbf{R}=(r_{ij})_{3\times 3}$, then the Gaze Image coordinates (k,g) can be expressed by the original image coordinates (u,v) as

$$(k,g) = \left[\frac{r_{11}u+r_{12}v+r_{13}f}{r_{31}u+r_{32}v+r_{33}f} , \frac{r_{21}u+r_{22}v+r_{23}f}{r_{31}u+r_{32}v+r_{33}f} \right] \qquad (1)$$

where f is the focal length of the camera with

$$(u,v)=(f\ X_c/Z_c\ , f\ Y_c/Z_c)$$

and

$$(k,g) = (X_g/Z_g\ , Y_g/Z_g) = (X/Z\ , Y/Z) \qquad (2)$$

Using equation (1) the gaze image can be obtained without actually rotating the camera. In the application of the mobile robot navigation, we reproject the original image into the gaze image plane which is parallel to the road surface the robot moves on. It is obvious form (2) that the shapes of the figures on the ground plane remain unchanged in the gaze image, and there have more interesting properties which are very useful in motion analysis [6]. In this paper we use the gaze image to detect the obstacle on the road.

(2). Stereo obstacle detection algorithm : We fix a pair of color cameras (in parallel or in fixation) on either side of the mobile robot, with the same height H relative to the ground (Fig. 2). Both pictures taken by the binocular cameras are transformed into two gaze images respectively, which are parallel to the ground plane with the baseline b_x along the X axes of both gaze coordinates. One camera is defined as the main camera and the 3D coordinates of the objects are expressed in the gaze coordinates (X, Y, Z) of this camera with XY plane parallel to the ground surface and Y axis aim at the forward direction of the mobile robot. A offset $d_k=b_x/H$ is added to the gaze image of second camera so that every point on the road surface will have the same coordinates in both gaze image theoretically. In practice , we assume that the image preprocessing stages are sufficient to eliminate any problems of photometric variance arising from stereo projections. This means if the figure which appears in both gaze images is really on the ground, little difference of intensity can be detected from the corresponding regions, otherwise, significant difference may appear in the regions corresponding to obstacles high up above the ground.

Based on the difference detection , the question " Are there any obstacles" can be answered rapidly by calculating the Sum of Absolute Difference (SAD) of the binocular gaze images. The Stereo Difference Image (SDI) can be considered as a map of free path measurement. Based on this map the robot can move on the region where the SDI values are low and should avoid those regions where the SDI values are high for collision-free motion. The stereo obstacle detection algorithm needs no feature extraction and correspondence . And it has no restriction to the shape and texture of the object and the background. So it is robust and efficient for any kind of environment, both indoors and outdoors. The Obstacle_Detector has been implemented in PIPE image system hosted by a PC286 [6] . The processing time is less than 0. 2 seconds.

2.2. HSI color-based Road_Segmenter

Color image segmentation is one of the oldest and most important problems in image understanding and computer vision, but general method for image segmentation leads to few successful applications. By analyzing a large number of road images on campus, we noticed that HSI (Hue, Saturation and Intensity) system is a better base of feature space than RGB system commonly used. The image of road areas has higher intensity and lower saturation, but that of the adjacent no-road areas and objects on the pavement is usually on the contrary. By the analysis of K-L transformation in HSI space for many road images, we found that $S = I+\lambda$ is a proper discriminant function for most of the pictures with λ to be determined adaptively. The Road_Segmentor can be implemented in less than 0. 1 s in PIPE image system, and the detail of the algorithm can be found in Lin & Chen[7].

The Road_Segmentor is so designed that not only the pavement of the road can be separated from neighboring area, but also objects on the pavement can be detected without confusing with shadows, tarmac patches, and so on. However it should be noticed that Road_Segmentor algorithm refers to no 3D perspective, or it has no "height" sense for the objects.

2.3. Correlation-based Egomotion_Estimator

After obstacles(objects on the road) have been detected, the robot should decide whether each object is static or dynamic, and estimate the velocity and direction of the object's motion in order to avoid collision with this object. First of all the egomotion parameters of the mobile robot should be known at prior or estimated by visual motion. Egomotion parameters can be obtained from the optical encoders of the robot but it may be inaccurate due to the slipping and skipping of the robot. Therefore, a correlation_based visual method is designed for egomotion estimation.

We use correlation of two successive original images to calculate the parameters of egomotion. By assuming that the ground, where the robot is running, is planar, the robot is restricted to a planar motion. There may be different motion (T_x, T_y, θ) between these two images while the robot is moving. By setting different values of (T_x, T_y, θ) incrementally and calculating the correlation, the one with maximal correlation value is selected as the best matching, and its corresponding (T_x, T_y, θ) as the estimated motion. With motion compensation, the two images will coincide rather well in the ground area, while there are differences in pixels of the heighted area. To improve the reliability of this estimation, we introduce weighted function to depress the noisy effect of heighted areas. The Egomotion_Estimator is a modified version of the Weighted Correlation algorithm[8].

However , the algorithm may offer inaccurate motion estimation if the heighted areas are overwhelming, e.g., obstacles are large and near the robot. Therefore, we use the difference image of Obstacle_Detector module as the negative weight for the correlation process so that the higher the SDI value, the less important the corresponding pixels for the egomotion estimation. The correlations under different motion settings are processed parallelly in PIPE system, therefore, the modified Egomotion_Estimator offers good motion estimation in every 3/60 second. The algorithms of Egomotion_Estimator and Obstacle_Detector are processed in parallel and pixel level integration of them is realized in realtime.

3. Purposive fusion strategy

3.1. System overview

The diagram of the purposive integration framework of stereo, color an motion are shown in Fig. 3. It is based on the PDP neural network model[4].

The basic visual modules , i. e. Road_Segmentor, Obstacle_Detector and Egomotion_Estimator work synchronously beginning at the capture of two color images. The difference image (SDI) of Obstacle_Detector module is used as the negative weights for the correlation process of Egomotion_Estimator module in order to reduce the negative effect of large and near obstacles(heighted areas) to the egomotion estimation. The road region of the current view is predicted by fusion the last road description results and the egomotion parameters from Egomotion_Estimator . This knowledge is used to guide the selection of Region Of Interesting (ROI) in SDI map . SAD value is calculated in the ROI and is used to decide whether there are obvious obstacles on the road. If the answer is "no" , then the Region_Fusion module and Motion_Analysis module do not activated, and the road description is made by simply fusing data from Road_Segmentor and Egomotion_Estimator modules. Otherwise the Region_Fusion module is activated to extract and fuse those regions on Color Segmented Images (CSI) and stereo difference image (SDI), which would be possible obstacles on the road. Region-based 3D estimation procedure is used to roughly estimate the pose and size of the obstacle by using the corresponding regions in the binocular gaze images. If the distance between robot and the object is less than the safe distance , then action must be taken to avoid collision. Otherwise, Motion_Analysis module is activated to estimate the motion of the object using the knowledge of 3D estimation from Region_Fusion module and the egomotion estimation from Egomotion_Estimator module. Finally, the results of Road_Segmentor, Region_Fusion module, Motion_Analysis module and Egomotion_Estimator are integrated to make up the road description.

3.2. Region_Fusion module

The fusion of data from Road_Segmenter and Obstacle_Detector modules can eliminate the ground patches from candidates of obstacle regions and derive 3D estimation of heighted area. It is noticed that stereo and color are different kind of data source and they are basically complementary and cooperative, so we prefer to use qualitative integration of stereo and color to deal with the uncertainty problem.

(1). Fusion of SDI map and CSI map : First, a binary map is created by thresholding, eroding and dilating the color segmented image (CSI) and stereo difference image (SDI). Then, a region extraction and grouping process is employed on the binary CSI and SDI maps so that the no-road region set $\{ C_i \mid i=1, \cdots, m \}$ and possible obstacle region set $\{ S_j \mid j=1, \cdots, n \}$ are

obtained. Connected or nearby streaks or points are grouped into one region and is described by (P, A, M, T, W) , where P is the contour of the region, A is the area, M is the centroid, T is color description for CSI region and SAD value for SDI region. W is the weight representing the importance of the region, which is calculated from A, M and T. The region set $\{ C_i \}$ and $\{ S_j \}$ are sorting respectively in a descend order of W.

Each region C_i with its weight greater than a certain value is checked by integrating with regions in set $\{ S_j \}$. The fused regions are classified as ground regions which should be eliminated, the obstacle regions which have been verified and are ready for 3D estimation, and the suspect regions which need further verification by focusing the attention on the suspect area in the color gaze images. Each verified obstacle region in obstacle set $\{ O_k \}$ is generated from the matched regions in $\{ C_i \}$ and $\{ S_j \}$.

(2). Region-based 3D estimation : For the obstacle region set $\{ O_l \}$ of left CSI image, we try to find the correspondence in the obstacle region set $\{ O_r \}$ of right CSI image using their region descriptions, eg. area centroid and the matched regions in $\{ S_k \}$. The typical situation is that corresponding regions are matched to (approximately) same SDI region(s).

Position and size of the obstacle are estimated by calculating the 3D coordinates (x_i, y_i, h_i) of corresponding contour points (k_{1i}, g_{1i}) in $\{ O_l \}$ and (k_{2i}, g_{2i}) in $\{ C_r \}$:

$$\begin{cases} h_i = (k_{1i} - k_{2i}) H / (d_k + (k_{1i} - k_{2i})) \\ x_i = k_{1i} (H - h_i) \\ y_i = g_{1i} (H - h_i) \end{cases} \tag{3}$$

where d_k is the offset relative to baseline b_x (ie. $d_k = b_x / H$), H is the height of the cameras , and h_i is the height of point $(x_i, y_i, z_i = H - h_i)$.

The 3D estimates are modified using the height and pose constraints: the visible surface of the obstacle on the road is often the front surface, so the height of an object is increasing from bottom to up , and the coordinate in Y axis do not scatter very much.

3.3. Motion_Analysis module

For each object detected, its motion should be estimated using image sequences in order to avoid collision with the robot. The 3D coordinates $\{(x_i, y_i, h_i)\}$ of the object are obtained from Region_Fusion module, and the egomotion parameters (T_x, T_y, θ) are continuously given by Egomotion_Estimator module. Next view of the main camera is planned adaptively by considering the 3D estimation of pose and size of object and the motion of the robot, so that the main part of the object can be in the sight of main camera . Active sensing planning is realized by controlling and monitoring the motion of the robot and select the instant of image capture. By assuming that the object is static, the projection of object in next view can be estimated as

$$(k_i^*, g_i^*) = (x_i^* / (H - h_i^*), y_i^* / (H - h_i^*)) \tag{4}$$

where

$$(x_i^*, y_i^*, h_i^*) = (x_i \cos\theta + y_i \sin\theta + T_y, -x_i \sin\theta + y_i \cos\theta + T_y, h_i)$$

In the gaze image of next view, attention is focused near the estimated region $\{Q_i^* = (k_i^*, g_i^*)\}$, and color segmentation is done in this particular region with the known color properties of the object from the last view. The observation of the object in the next view can be expressed as $\{Q_i = (k_i, g_i)\}$. The motion parameters of the object can be estimated using approximate translation motion (M_x, M_y):

$$\begin{cases} M_x = \dfrac{1}{n} \sum_{i=1}^{n} (H - h_i^*)(k_i - k_i^*) \\ M_y = \dfrac{1}{n} \sum_{i=1}^{n} (H - h_i^*)(g_i - g_i^*) \end{cases} \tag{5}$$

where (k_i, g_i) and (k_i^*, g_i^*) are the corresponding features of the observed and the estimated, such as centroids, portion of the contours and etc.

4. Experimental results

All the visual modules have been implemented in the PIPE, a multiple pipelined image processing system. The realtime execution of most procedures guarantees the purposive and adaptive integration of different visual modules. Fig. 4 to Fig. 8 show an example of one detection phase. Original binocular color images are shown in Fig. 4(a) and (b), and their color segmented images (CSIs) created by the Road_Segmentor in (c) and (d), where a man and some shadows were classified as non-road regions.

Fig. 5 (a) and (b) show the corresponding binocular gaze images in which the parallelism of the road edges and the shapes of the ground figures (e.g., shadows) were recovered. The stereo difference image (SDI) produced by stereo Obstacle_Detector is shown in Fig. 5(c), and the SAD value calculated inside the road region shows that there exist obvious obstacles on the road. Fig. 5(d) shows the possible obstacle regions

{Sj} extracted from SDI map.

The results of region fusion and 3D estimation is shown in Fig. 6. Fusion of regions in two CSIs and the SDI produces a pair of corresponding obstacle regions which is represented in the gaze image plane in Fig. 6(a) and (b) respectively. 3D estimation is obtained by matching the left and right contours of the region pair (Fig. 6(c)). The initial heights of the points in the left and right edges are shown in dash lines in Fig. 6(d), where the horizontal axis represents g coordinates which is from up to bottom in gaze image, and vertical axis the heights of the obstacle points. It can be seen that the height values are not accurate due to the noise of region extraction. Statistics of the y coordinates shows that they do no scatter too much (y=4. 75 m, ρ_y =0. 15 m). Using this constraint the heights of two edges are modified to be smoothly decreasing top-down (shown in the bond lines in Fig. 6(d)). The height of visible part of the man is about 1. 65 m and it is located around (x= -0. 15m to 0. 10 m, y = 4. 75 m).

The motion parameters were calculated continuously by the Egomotion_Estimator while the robot was moving forward. The next view for Motion_Analysis module is planned adaptively according to the 3D estimation of the object (man) and the egomotion parameters. The image sequence between current views and next view is shown in Fig. 7. The egomotion between this time interval is shown in Fig. 8 (a), in which the accumulate egomotion is (T_x = 0. 351 m, T_y = 2. 183 m, $\theta=4^o$). If we assume that the object (a man in this example) is static, then its projection in the next view is shown in black contour in Fig. 8(c). But comparing the actual region and the estimated one of the man in the next view (Fig. 8(d)), the approximate motion is estimated as translation (M_x=0.486m ,M_y=1.551m), showing that the man went away from the robot.

5. Summary and discussion

In this paper, an active fusion framework is presented for the visual detection of dynamic obstacles in unstructured outdoor road environment. Purposive visual modules, which avoid the difficult problems of traditional vision methods and aim at the particular problems, are proved to be efficient and effective. Adaptive integration and cooperation of different visual modules shows advantages to fulfill difficult tasks with low computation cost. Experimental results have been given to demonstrate the robustness and effectiveness of the methods.

In spite of the advantage of the gaze transformation and gaze image, the resolution of the gaze image is reduced . We are planning to modify the reprojection process so that the visual operations are employed to the original image of main camera and the rectification image of the second one. The difference measurement in the current Obstacle_Detector module is intensity of the images. Further work is to be done to use color information as the difference measurements in order to cope with the color segmentation. Further work will also include road description and free path planning .

Acknowledgements

This work was supported in part by the 863 High Technology Program of China. Support was also given by the National Laboratory of Intelligence Technology and System.

References

[1] Jain R C and Binford T O, Ignorance, myopia and naivete in computer vision systems, CVGIP:Image Understanding,53, No 1,pp112-117, Jan 1991

[2] Arnheim R, Visual thinking, University of California Press, 1969

[3] Aloimonos J, "Purposive and qualitative active vision," Proc. 10th ICPR, pp 346-360, 1990

[4] McClelland J L and Rumelhart D E, Parallel distributed processing (PDP), MIT Press, 1985

[5] Blackbum M R and Nhuhen H G, Biological model of vision for an artificial system that learns to perceive its environment, IJCNN, pp II 219 - II 226, June 1989

[6] Zhu Z and Lin X., "Realtime Algorithms for Obstacle Avoidance by Using Reprojection Transformation", Proc. IAPR Workshop on MVA, pp 393-396, 1990.

[7] Lin X and Chen S, "Color Image Segmentation Using Modified HSI System for Road Following", Proc. ICRA, pp 1998-2003, 1991

[8] Chen S, Lin X and Zhu Z," Qualitative Visual Navigation Using Weighted Correlation ", Proc IEEE CVPR,1993.

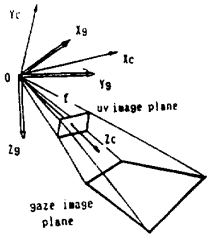

Fig.1. Gaze Transformation

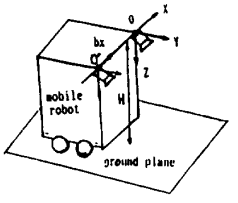

Fig.2. Coordinate system

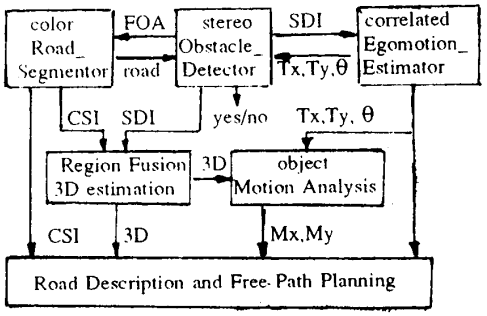

Fig.3 PDP-based Purposive Fusion Diagram

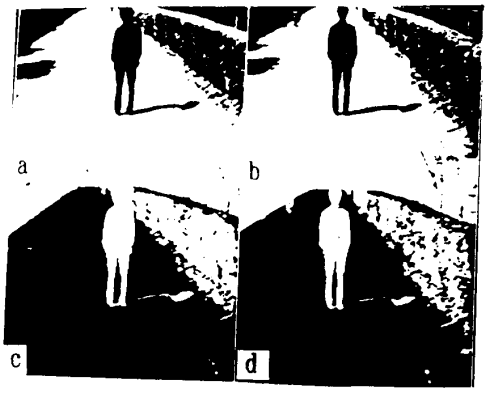

Fig. 4. Color road segmentation

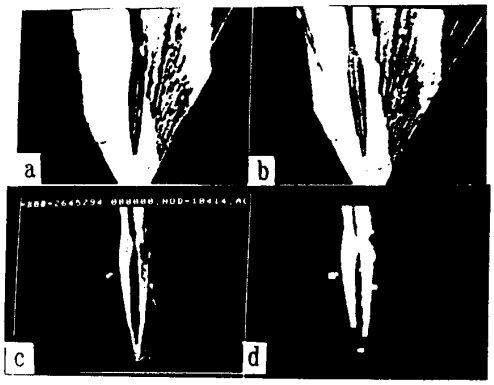

Fig. 5. Stereo obstacle detection in gaze images

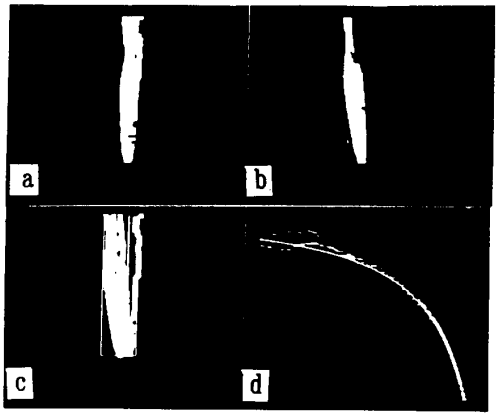

Fig. 6. Region fusion and 3D estimation

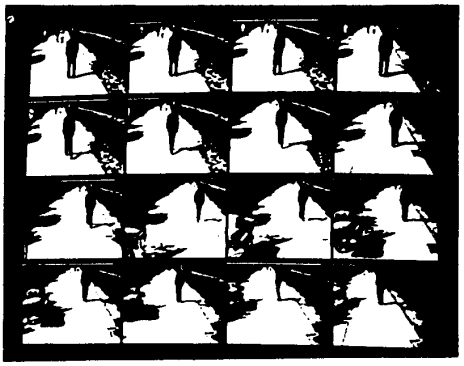

Fig. 7. Image sequences for Egomotion estimation

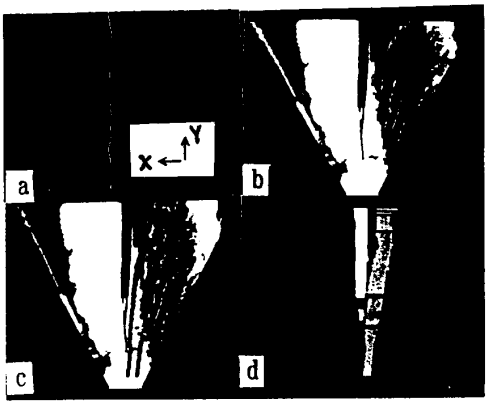

Fig. 8. Dynamic object analysis