Depth-Based Target Segmentation for Intelligent Vehicles: Fusion of Radar and Binocular Stereo

Yajun Fang, Ichiro Masaki, *Senior Member, IEEE*, and Berthold Horn

***Abstract*—Dynamic environment interpretation is of special interest for intelligent vehicle systems. It is expected to provide lane information, target depth, and the image positions of targets within given depth ranges. Typical segmentation algorithms cannot solve the problems satisfactorily, especially under the high-speed requirements of a real-time environment. Furthermore, the variation of image positions and sizes of targets creates difficulties for tracking. In this paper, we propose a sensor-fusion method that can make use of coarse target depth information to segment target locations in video images. Coarse depth ranges can be provided by radar systems or by a vision-based algorithm introduced in the paper. The new segmentation method offers more accuracy and robustness while decreasing the computational load.**

***Index Terms*—Depth detection, image segmentation, motion stereo, obstacle detection, sensor fusion.**

I. Introduction

FOR intelligent vehicle (IV) applications, we expect a real-time highway environment interpretation system to provide complete three-dimensional (3-D) information for the targets in the driving environment, i.e., the target size, location, and depth (distance). For example, this information is essential for an intelligent cruise control system to adjust speed based on the distance between the intelligent vehicle and the nearest preceding vehicles, and on the types of these vehicles.

A single sensor is not adequate to provide reliable information for autonomous driving guidance in real time because of weather, ambient lighting, and other limitations. Vision systems are not reliable in depth detection because of miscorrespondence problems and failure to function well in bad weather. Radar systems offer relatively accurate depth information and robustness in bad weather, but their performance in spatial resolution is poor. Usually, there are three types of radar systems: broad single beam systems not having much horizontal resolution (type 1); multibeam systems having limited horizontal resolution (type 2); and scanning radar systems including electronic scanning and mechanical scanning (type 3). While the horizontal resolution improves, the cost of radar systems goes up significantly. Even for expensive specialized image radar that provides both depth and spatial resolution, the horizontal resolution remains worse than typical vision systems. Furthermore, radar systems used in IV are designed to detect moving targets to avoid false alarms from static objects *along* a road. Such systems might also ignore dead (static) vehicles *on* the road. Finally, distance resolution for lower performance radar systems is not satisfactory because of reflection.

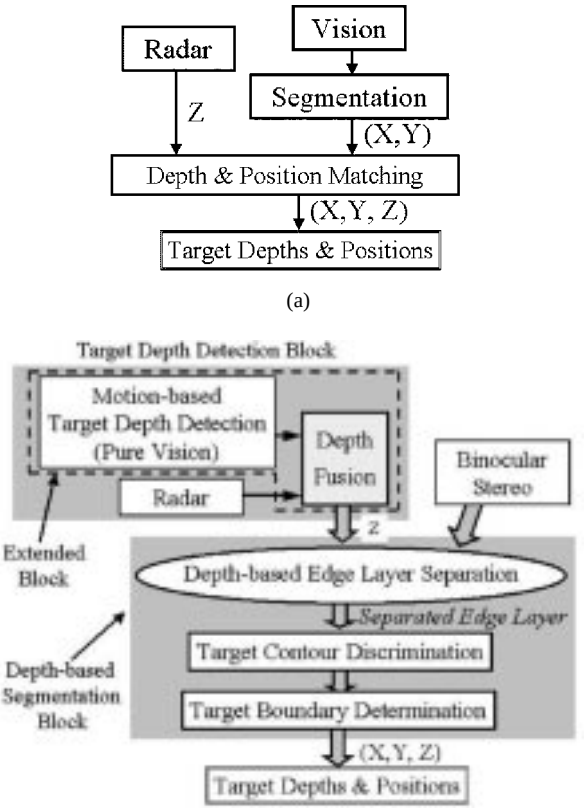


Fig. 1. (a) Traditional method of target detection. (b) Depth-based target detection: fusing radar and stereo.

Consequently, since the advantages and disadvantages of radar and vision systems are complementary, typical IV systems detect depth and target segmentation information separately from low-cost radar systems and video sequences respectively, and match the complementary information afterwards, as in Fig. 1(a).

However, these systems do not take advantage of the close relationship between depth and stereo images. In this paper, we present a data fusion strategy that actively introduces the depth information into the image segmentation process and propose

Manuscript received August 31, 2001; revised March 15, 2002. The Guest Editor for this paper was K. Ikeuchi.

Y. Fang and B. Horn are with the Artificial Intelligence Laboratories, Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: yjfang@mit.edu; bkph@ai.mit.edu).

I. Masaki is with the Microsystems Technology Laboratories, Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: masaki@mit.edu).

Publisher Item Identifier 10.1109/TITS.2002.802926.

1524-9050/02$17.00 © 2002 IEEE

a new depth-based segmentation algorithm [see Fig. 1(b)]. Basically, given a series of depth data for n targets from a radar system, say, $d_1, d_2, \ldots, d_n$, we expect to find their corresponding positions sequentially. The idea is to split an edge map of a binocular image into n edge layers corresponding to n given target depth information so that different layers contain the edge pixels of targets at different depth ranges. In this way, the original multiple-target segmentation task is decomposed into several simpler and easier single-target segmentation tasks on each depth-based target feature layer, thus, improving the segmentation performance.

The depth-based segmentation algorithm needs only coarse target depth information to separate depth-based feature layers and imposes lower performance requirements on depth detection sensors. A low-quality radar system is adequate and accurate depth information for every frame pixel is not needed. If radar is not available, we propose a pure vision-based target depth-detection algorithm to detect the number of targets and target depth information. For real applications, it is better to combine depth information from both radar and the additional "vision-based target depth detection" algorithm for better performance.

We expect the sensor fusion system to provide better target detection than separately detecting segmentation regions and depth information and matching them afterward. With the depth-based segmentation algorithm, the fusion system in Fig. 1(b) can provide better target segmentation performance than that in Fig. 1(a). On the other hand, segmented regions on stereo images can also improve the estimation of target depth. In Section II, we introduce the enhanced segmentation algorithm (by additional depth information). In Section III, we introduce how to detect target depth information required for Section II when radar information is not available.

II. Depth-Based Segmentation System

There are several traditional image segmentation approaches. Popular characteristic feature thresholding or clustering algorithms [1], [2], such as k-means [3] and ISODATA, do not exploit spatial information. Boundary detection methods [1], [2], such as edge detection algorithms, are not robust enough to be used alone. Region-growing-based algorithms [2] need seed points or seed regions to grow into larger regions. Motion-based segmentation methods [4], [5] segment images based on the target motion pattern (e.g., optical flow vectors). For real-time implementation, tracking-based segmentation methods [6] are the most popular, which assume the similarity of target features in consecutive frames and use it to remove initial segmentation errors. Symmetry-based segmentation algorithm [7] is another typical segmentation method for IV systems. The method assumes the symmetry of typical IV targets (such as vehicles, pedestrians) and asymmetry of background, and uses a symmetry finder to detect candidate target regions. The performance is invariant under nodding movements and under changes of target size. However, these assumptions are not necessarily held in real situations, which limit their applications.

This paper does not make assumptions on target shapes or the symmetry of targets. Our algorithms focus on static segmentation performance, since it is important to quickly detect new targets (cut-ins) for fast response in real-time driving. Better initial static segmentation always brings with it the performance improvement of an overall tracking system if the "tracking" feature is included. Furthermore, our algorithms do not segment all objects in the images (and provide accurate 3-D information for each pixel), because our attention is to those targets within the depth ranges of interest for the purpose of autonomous driving.

A. Depth-Based Segmentation Principle

In general, the edge map of a real scene image is composed of edge pixels that belong to different targets at different depth ranges. The target depth information imposes constraints on correspondence searching for the target. Given n depth data from radar, the relationship between target depth and its correspondence-disparity provides n corresponding groups of target feature points from the edge map of the original stereo images, leading to n edge-layers containing the edge pixels of targets within different depth ranges. (See detail in Section II-A.2.) In each depth-based feature layer, there are fewer targets or even only one target. Therefore, segmenting each target sequentially in separated edge layers is easier and more reliable than original segmentation task that segments all targets simultaneously. Thus, by providing binocular stereo systems with coarse target depth information from sensors (or algorithms in Section III), we become able to split an edge map into n depth-based edge layers, and to decompose the original task of multiple-target segmentation into n segmentation tasks on each depth-based target feature layer. In this way, more accurate and robust target segmentation can be achieved. The flow chart for our proposed depth-based segmentation algorithm is shown in Fig. 1(b) and we will next introduce each step.

1) Depth Determination: Based on the fusion strategy in Fig. 1(b), the first step is to detect targets' depth information. It will be shown that the detection algorithm does not impose high requirements on depth accuracy; radar data with low resolution is acceptable. We can also obtain target depth information from other sensors or other algorithms. (See detail in Section III).

2) Depth-Based Edge Layer Separation: The objective of this step is sequentially to locate image edge pixels corresponding to targets at given depth, say, $d_1, d_2, \ldots, d_n$. Assuming the resolution from the radar is Δd, the real depth of the ith target will be between $[d_i - \Delta d, d_i + \Delta d]$. To obtain the corresponding depth-based edge layer for each depth range, we locate target edge pixels that satisfy both depth constraints and binocular stereo correspondence constraints. The detailed definition of the two constraints is as follows.

- *Binocular Stereo Epipolar Line Constraints*

  For binocular stereo image pairs, there exists a strong positional constraint, i.e., matching feature pixels from stereo images should be on corresponding epipolar lines. For aligned cameras, the epipolar lines are frame rows. Usually the epipolar line constraint is not strong enough to locate corresponding feature points, and traditional binocular stereo methods need other complex feature constraints to further limit correspondence searching.

- *Depth Constraints*

  For a binocular stereo system with baseline distance b and focal length f, target depth z imposes constraints on its correspondence disparity as follows:

$$x_l - x_r = b\frac{f}{z} \tag{1}$$

where x_l, x_r are the x coordinates of target pixels in images from the left and right cameras, respectively, and $x_l - x_r$ is the disparity. Since the real depth of the ith target in the given radar data series is between $[d_i - \Delta d, d_i + \Delta d]$, in a stereo system with baseline distance b and focal length f, the corresponding disparity for target i will be in the following range:

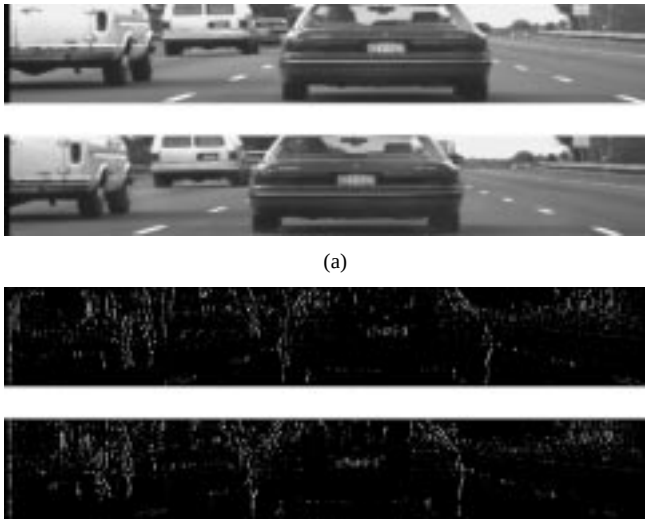
$$\left[\frac{bf}{d_i + \Delta d}, \frac{bf}{d_i - \Delta d}\right]. \tag{2}$$

We choose candidate feature pixels satisfying both constraints. For each feature pixel in the left-camera images, we check whether there exist corresponding feature pixels (at the epipolar lines in the right-camera images) whose disparities fall in the given disparity range, as in (2). All the feature pixels in the left-camera image with corresponding pixels make up a depth-based feature layer. At different disparity ranges, we obtain different edge layers corresponding to these ranges, for which we further detect whether (and where) there is any target. High-quality edge detection and strict edge layer separation are not necessary, since depth-range-based filtering later will easily remove edge layer separation noise, e.g., the residual background noise. Therefore, the simple and fast Sobel edge detection algorithm is chosen and it is tolerable for edge pixels to overlap among different depth-based edge layers. Low edge-detection thresholds are used to provide sufficient pixels for reliable target discrimination in the next step.

3) Depth-Based Target Contour Discrimination: The separated edge layer contains target pixels that are close together, while noisy pixels from other depth ranges scatter randomly throughout the layer. We delineate the candidate targets in the edge layer through a morphological "closing" operation. We first dilate edge lines in an edge layer to increase edge length. The aim is to connect all the boundaries of targets into long boundaries with lengths longer than those of noisy edges. Second, an erosion operation is applied to remove short edges (assumed to result from noise). After these operations, targets should stand out. The assumption is that the boundary of our target is much longer than the noise coming from other depth layers.

Other segmentation algorithms also adopt morphological operation for robustness. However, they try to identify the target contour in the original edge map that involves much more background noise. In our algorithm, separated layers have cleaner backgrounds and less noise from targets at different depth ranges. Furthermore, targets in each layer are a portion of total targets in the original images. Thus, segmentation for each layer is simpler and easier than for the original edge image.

4) Depth-Based Target Boundary Determination: Because target contours are manifest after the previous steps, the final step can create rectangular bounding boxes for targets through pixel aggregation (a region-growing method) with seed points being chosen from accentuated (i.e., processed by morphological closing) candidate points. Repeating this process at different disparity ranges, we can detect the locations and the sizes of targets within different depth ranges sequentially.



(a)

(b)

Fig. 2. Binocular stereo images and their corresponding edge maps. (a) Images from left (top) and right (bottom) cameras at frame 57. (b) Edge maps for (a).

B. Segmentation Example and Results

This section presents target detection results for the binocular stereo images in Fig. 2 (frame 57 in our video sequence) step by step, applying the steps from Section II-A. The three vehicles in Fig. 2(a) are called "left car," "middle car," and "right car." We need to detect three depth ranges for the three vehicles and derive vehicles' locations from their ranges. Without losing generality, we show only segmentation results for the left-cameras images.

1) *Depth Range Determination*

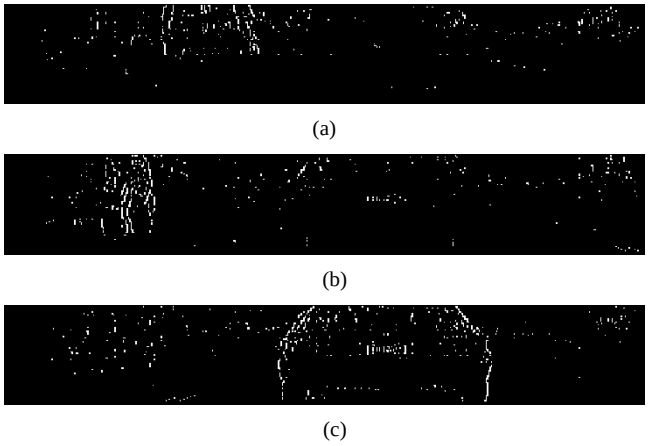
   First, we determine that the three nearest vehicles in the images separately correspond to disparity ranges 22–24, 18–20, and 7–9. (See detail in Section III.) The larger the disparity range, the closer the vehicle is to the observer.

2) *Depth-Based Edge Layer Separation*

   Second, the edge map at the top of Fig. 2(b) (for left-camera image) is decomposed into three edge layers corresponding to the three disparity ranges 7–9, 18–20, and 22–24, as in Fig. 3. Though there are edge pixel overlaps among the three edge layers, the three separate vehicles are distinct. The depth-range-based filtering later will remove the error resulting from simple edge layer separation algorithms.

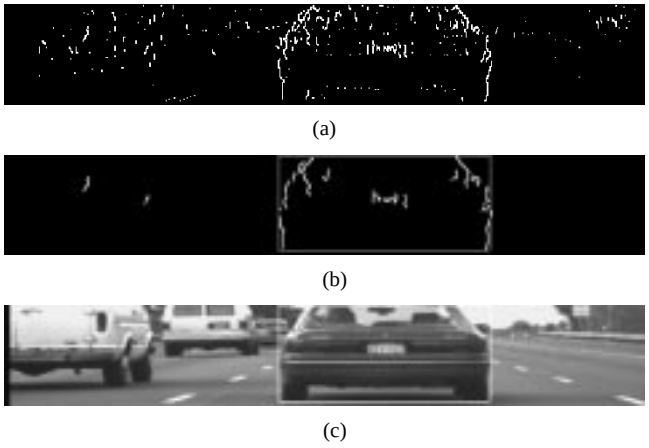
3) *Depth-Based Target Contour Discrimination*

   For the edge layer in Fig. 3(c), the dilation operation result is shown in Fig. 4(a) and the closing result (from applying both dilation and erosion operations) is shown in Fig. 4(b), in which our target vehicle stands out. Thus, the morphological operation successfully removes the error edge lines with short lengths. Without losing generality, we show only intermediate processing results for the right car with disparity range 22–24.



(a) (b) (c)

Fig. 3. Edge layers corresponding to different disparity ranges for frame 57 (Fig. 2). (a) 7–9. (b) 18–20. (c) 22–24.



(a) (b) (c)

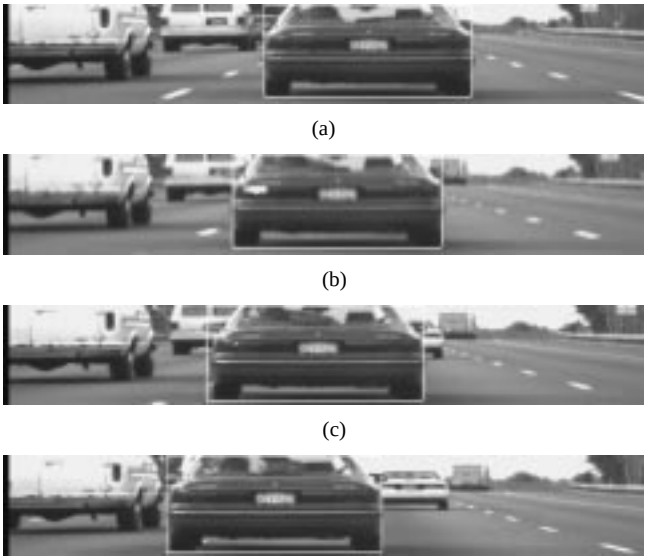
Fig. 4. Procedure to locate targets within a depth range (which corresponds to disparity range 22–24) for frame 57 (Fig. 2). (a) Edge dilation result. (b) Edge closing result. (c) Segmentation result.

4) *Depth-Based Target Boundary Determination*

The segmented region for the right vehicle (at disparity range 22–24) is drawn in a rectangular box, as in Fig. 4(b), (c) for frame 57 and in Fig. 5 for frames 62, 69, 73, and 80. Segmentation results at disparity ranges 7–9 for frame 57 and 5–6 for frame 31 are shown in Fig. 6.

All the results detect the targets with accuracy and robustness. The algorithm uses the same initial parameters to segment at different depth ranges or in different video frames, as in Figs. 4, 5, 6, and 8, showing its robustness. Furthermore, the algorithm does not impose high requirements on the performance of target depth determination and edge detection. We can use several different disparity ranges, 22–24, 23–25, 22–25, etc., and use different Sobel edge-detection thresholds to yield the same segmentation results in Fig. 4. The method does not assume target shapes or symmetry. With some modification, the algorithm has detected pedestrians accurately, showing its ability to detect targets of various shapes (as in Fig. 7; see detail in [8]). On the other hand, the segmented regions from stereo images can also improve depth estimation if the resolution of given depth is not satisfactory.

Unlike tracking-based segmentation, our segmentation algorithm does not need initialization time. Among the steps in Section II-A, most computation time is spent in correspondence searching at the step of "depth-based edge layer separation." For each edge pixel, the correspondence searching time is in direct proportion to the width of the disparity range (instead of, typically, the image width). For the whole algorithm, the computational load is in direct proportion to the product of the number of edge pixels and the disparity range width (if edge maps are given). Furthermore, the correspondence searching process and several other steps can be implemented simultaneously at different rows or even at different pixels, which makes parallel hardware implementation possible and further increases calculation speed. Thus, the algorithm shows the potential to detect quick cut-ins in real time when environmental scenes change rapidly.



(a) (b) (c) (d)

Fig. 5. Example of detecting right vehicle at frames (a) 62; (b) 69; (c) 73; and (d) 80.

The algorithm is limited in situations where only single sides of targets are in images, as in Fig. 8. Depth-based segmentation needs to capture target boundaries. If there are only single sides of targets in an image, we capture only partial target regions.

III. Pure Vision Scenario: Motion-Based Target Depth Detection System

Up to now, we assume that the depth information required in Section II is provided by radar. When radar information is not available, we propose a vision-based system to provide target depth information. Combining the proposed "motion-based target depth detection system" with "depth-based segmentation system" (Section II), we present a "pure vision" system to detect target depth and locations without introducing too much complexity.

Generally, it is hard to obtain accurate 3-D reconstruction from binocular stereo images using simple correspondence matching algorithms. However, our proposed segmentation algorithm expects only coarse target depth, which simple algorithms can provide. Furthermore, binocular stereo video (multiple frames) can provide more information than static images to increase reliability. Thus, proposed algorithm structure

Fig. 6. Example of locating targets at different disparity ranges. (a) and (b): Disparity range 7–9 at frame 57 (Fig. 2). (c) and (d): Disparity range 7–9 at frame 31. (e) and (f): Disparity range 5–6 at frame 31. (a), (c), and (e): Edge closing operation result. (b), (d), and (f): Segmentation result.

Fig. 7. Example of detecting both ball and pedestrian using the same initial parameters.

Fig. 8. Target segmentation results within disparity range 18–20 for frame 57 (Fig. 2). An example of "single-sided limitation" where only part of the target is captured, which leads to target size detection error. It happens when only part of target is in the image.

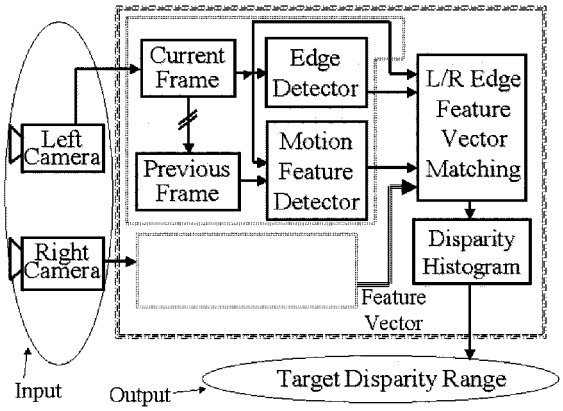


Fig. 9. Target depth detection system based on fusing binocular frames and motion. Detail for Fig. 1(b).

in Fig. 9 [detail for Fig. 1(b)] can detect target depth. We first introduce "motion-based correspondence matching criteria" in Section III-A, and then use them to detect target depth range in Section III-B.

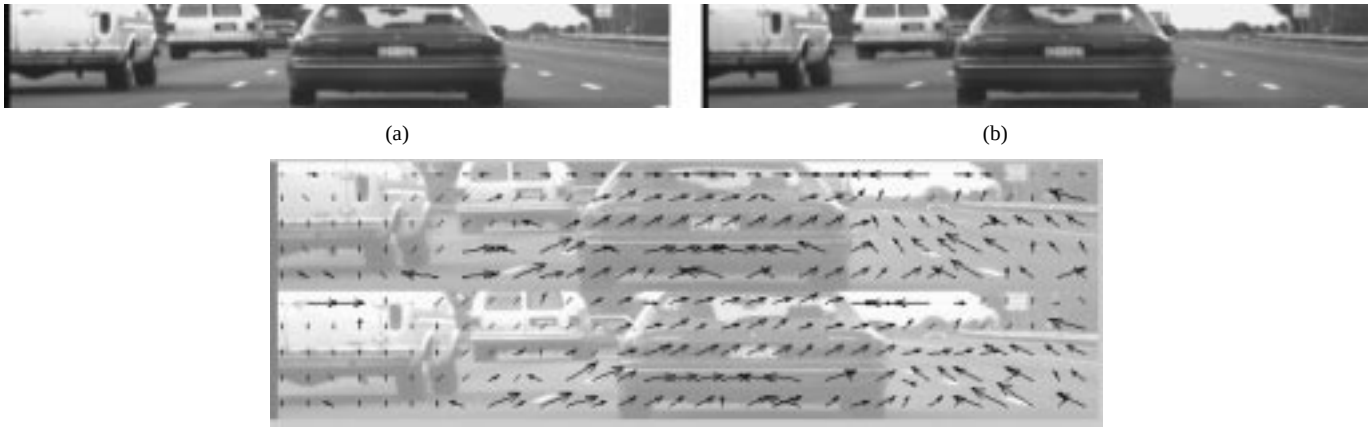
A. Motion-Based Correspondence Matching Criteria

Since the epipolar line constraint (Section II-A.2) alone is not enough to guarantee stereo correspondence matching, additional local content constraints (cornerness, edge, color, intensity, etc.) and global constraints (continuity) are needed. However, imposing global constraints is complex and time consuming, while algorithms based on local constraints are not reliable. Trinocular stereo [9] is one solution for both simplicity and effectiveness requirements. The algorithm imposes epipolar line constraints among three aligned camera images, i.e., the horizontal image position of a target point from the center camera must be the midpoint of image positions of the target point from the left and right cameras.

However, for a binocular stereo setting, there is no similar simple and effective method. Instead, we improve correspondence reliability by imposing extra motion constraints. Because of rigidity, motion information for the same targets calculated from the left and right camera video frames should be similar. As in Fig. 10, motion vectors for both stereo video frames show similar patterns even if detected motion vectors are noisy. To take advantage of such motion information similarity, we define multidimensional feature vectors including location constraints, motion constraints and feature similarity constraints. For each edge pixel at position (i,j), feature vector $\vec{V}(i,j)$ includes the following information:

$$\vec{V}(i,j) = [\mathrm{MV}_x(i,j)\mathrm{MV}_y(i,j)E_{\mathrm{cor}}(i,j)i_u(i,j)I(i,j)]' \tag{3}$$

where I is the intensity value. Motion vectors $(\mathrm{MV}_x, \mathrm{MV}_y)$ are calculated based on correlation methods. When searching for



(a) (b)

(c)

Fig. 10. Similarity of motion vectors for binocular stereo images. (a) and (b) Images from left and right cameras at frame 58. (c) Motion vector between frames 58 and 57 (Fig. 2).

the minimum correlation error E_{cor}, the uniform index i_u is defined as the function of the difference between the maximum correlation error and the minimum correlation error, which reflects the texture information of surrounding image patches. If there are significant changes within the image patch surrounding pixel (i,j), the uniform index $i_u(i,j)$ is large. Otherwise, i_u is small.

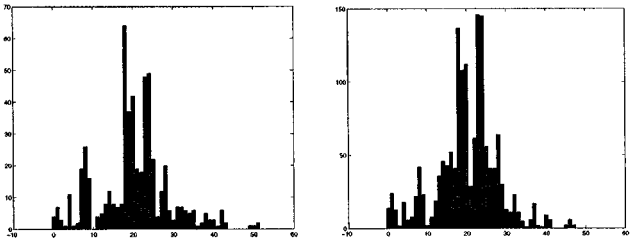
The *matching error index* $E_{\text{match}}(i_l, i_r, j)$ between pixel (i_l, j) in the left-camera image and pixel (i_r, j) in the right-camera image is defined as the weighted sum of their respective feature vector component differences

$$E_{\text{match}}(i_l, i_r, j) = \sum_k w_k * [\vec{V_l}(i_l, j) - \vec{V_r}(i_r, j)]_k \qquad (4)$$

where k represents the vector index. For edge pixel EL of left-camera images, the matching candidates, edge pixels ER_n, are on the same row as EL in right-camera images, among which the multidimensional correspondence pair EL and ER_m provides the minimum matching error and $x_l - x_{r_m}$ gives out the corresponding disparity. Thus, the corresponding disparity is obtained and recorded in a disparity histogram, as in Fig. 11(a).

B. Depth Range Detection Through Disparity Histogram

The concentration of edge points at peaks in the disparity histogram comes from targets in the field of view. Normally, peaks at large disparity location represent close targets and peaks at small disparity location represent targets far away. Since closer vehicles occupy larger image space (excluding the possibilities of toy cars on the road), leading to higher peaks, the higher peaks are normally located at the right side of the lower peaks in the disparity histogram. Because of the complexity of targets, it is also possible that farther vehicles have more edge pixels on images and have higher peaks than closer vehicles. Generally, to identify a peak as a target other than noise, the larger the peak's disparity, the higher the threshold should be. Thus, based on these common characteristics of the histogram peak distribution shape, target depth information can be derived. Note that the algorithm can be implemented in parallel and is, therefore, suitable for very large-scale integration hardware implementation.



(a) (b)

Fig. 11. Disparity histogram for edge pixels in Fig. 2(b). x coordinate: disparity value. y coordinate: the number of edge pixels within the disparity range. (a) Motion-based binocular stereo algorithm using frames 57 [Fig. 2(a)] and 58 [Fig. 10(a) and (b)]. (b) Trinocular algorithm.

TABLE I
DETECTED INFORMATION FOR VEHICLES IN FIG. 2

category	Disparity	Closeness	Edge Pixel #
Left Car	18-20	2nd nearest	most
Middle Car	7-9	3rd nearest	3rd most
Right Car	22-24	1st nearest	2nd most

C. Target Depth Detection Results

The disparity histogram in Fig. 11(a) shows the motion-based correspondence matching result for stereo frames 57 [Fig. 2(a)] and 58 [Fig. 10(a) and (b)]. Treating two close peaks at disparity 18 and 20 as one peak, the three highest peaks correspond to the three nearest vehicles in frame 57 with detailed information shown in Table I.

The peak for the left car is higher than for the right car because the depths of the left car and the right car in Fig. 2(a) (frame 57) are close and the left car has more complex projection in the edge map shown in Fig. 2(b). The lower peaks at the right side of the three highest peaks are noise. In contrast, the three highest peaks for trinocular algorithm [9] [Fig. 11(b)] are at disparity ranges 28, 18–20, 23–24 (from low to high) while missing the target at disparity range 7–9 and introducing a false target (peak) at disparity 28. The detection error is caused by miscorrespondence. We define the signal-to-noise ratio (SNR) as the ratio of the lowest peak corresponding to real vehicles over the highest peak corresponding to false alarms. For the trinocular stereo algorithm, the SNR is less than *one* while the SNR from

the proposed algorithm is around *four*, which offers better performance. The histogram comparison in Fig. 11 shows the performance improvement of motion-based target detection over trinocular algorithms. The obtained vehicle disparity ranges are those used in Section II-B to segment vehicles. The combination of two blocks, "vision-based target detection algorithm" and "depth-based segmentation algorithm (Section II-A)," as in Fig. 1 can offer complete depth and segmentation information for targets.

Specifically, the proposed target detection algorithm can be fused with radar information (as in Fig. 1). Target number from radar helps to differentiate target peaks from noise peaks in choosing peaks from disparity histograms and to give out more reliable target depth information. Fusing depth information from two sources can remove radar system error from reflection or vision system error from miscorrespondence, thus, improving accuracy and reliability.

IV. Conclusion

This paper proposes a new depth-based target detection method based on fusing target depth information and binocular stereo images. For general sensor fusion methods, all sensor data are of similar characteristics and are fused using standard mathematical algorithms, such as kalman filtering, fuzzy logic, Dampster–Shafer evidential reasoning, etc. The sensor set-up and fusion in this paper are tailored to the given task. Information from one sensor (coarse depth) is used to guide another sensor's processing (segmentation). The informed segmentation algorithm is easier and more reliable by decomposing the multitarget segmentation task into several single-target segmentation tasks on individual depth-based target feature layers. Usually radar can offer required target depth. Higher reliability will be achieved if radar information can be fused with our proposed target depth detection algorithm (based on the motion information similarity between binocular stereo images), which can be used separately if radar is not available.

Since the task does not need accurate 3-D information for every frame pixel, our design focuses computation power on the targets of our interest. The implementation results show that the accuracy of static image segmentation offers a good basis for further information tracking in video sequences. The algorithms can be implemented in parallel in hardware to increase the computational speed, and the light computational load makes real-time implementation possible in practical situations.

References

[1] J. S. Lim, *Two-Dimensional Signal & Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[2] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1993.

[3] D. Pelleg and A. Moore, "Accelerating exact k-means algorithms with geometric reasoning," in *Proc. 5th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Aug. 1999, pp. 277–281.

[4] T. Meier and K. N. Ngan, "Automatic segmentation of moving objects for video object plane generation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 525–538, Sept. 1998.

[5] D. Willersinn and W. Enkelmann, "Robust obstacle detection and tracking by motion analysis," in *Proc. IEEE Conf. Intelligent Transportation Systems*, 1997, pp. 717–722.

[6] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[7] M. Bertozzi, A. Broggi, A. Fascioli, and S. Nichele, "Stereo vision-based vehicle detection," in *Proc. IEEE Intelligent Vehicles Symp.*, 2000, pp. 39–44.

[8] Y. Fang, I. Masaki, and B. Horn, "Distance/motion based segmentation under heavy background noise," in *Proc. IEEE Intelligent Vehicles Symp.*, 2002.

[9] J. Bergendahl, "A computationally efficient stereo vision algorithm for adaptive cruise control," Master's thesis, Massachusetts Inst. Technol., Cambridge, 1997.

Yajun Fang received the B.E. degree in automatic control from Huazhong University of Science and Technology, Hubei, China, the M.S. degree in automation from Tsinghua University, Beijing, China, and the M.S. degree in electrical engineering from the University of Massachusetts at Amherst, in 1991, 1994, and 1997, respectively. Currently, she is working toward the Ph.D. degree in electrical engineering and computer science at the Massachusetts Institute of Technology, Cambridge.

Her research interests include computer vision, machine learning, image/video processing, intelligent transportation systems, and signal processing.

Ichiro Masaki (A'80–SM'92) was born in 1948 in Japan. He received the B.S. and M.S. degrees from Waseda University, Tokyo, Japan, in 1971 and the Ph.D. degree from Osaka University, Osaka, Japan, in 1973.

From 1973 through 1981, he worked on robotics and machine vision at Kawasaki Heavy Industries, Tokyo, Japan, Waseda University, and Unimation Inc., Danbury, CT. From 1981 through 1993, he worked at General Motors Research Laboratories, Grand Rapids, MI. During this time period, he started an intelligent vehicle project at GM for developing cars which are equipped with machine vision systems to assist drivers. In 1993, he joined the Massachusetts Institute of Technology (MIT), Cambridge, where he is currently the Director of the Intelligent Transportation Research Center at MIT's Microsystems Technology Laboratories.

Dr. Masaki started IEEE annual international symposia on intelligent vehicles.

Berthold Horn is currently a Professor in the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, where he is also a member of the Artificial Intelligence Laboratories.