

Vision-Based Front Vehicle Detection and Its Distance Estimation

Jyh-Yeong Chang and Chien-Wen Cho

Abstract—Vision-based driver assistant systems are very promising in Intelligent Transportation System (ITS). This paper will propose a system that can detect front vehicles and estimate the nearest car distance from the host car. In a companion paper [1], we have developed a scene analysis module that deals with scene segmentation and natural object labeling of forward-looking images by the use of fuzzy Adaptive Resonance Theory (ART) and fuzzy inference techniques. Based on this technique, the proposed system can detect the front vehicles and then estimate the distance of the nearest car from us. The validity of our proposed scheme in car detection and the distance estimation was verified to be very successful by field-test experiments.

I. INTRODUCTION

It is the most important issue to resolve the safety problems encountered in the research of Intelligent Transportation System (ITS). Vehicle detection and its distance estimation play a vital role in driver safety assistance. Vision-based systems take several hints of vehicles as criteria to distinguish vehicles from other objects. Broggi *et al.* [2] used the characteristics of symmetry, bounding box shape, and road region constraints to distinguish vehicles from other road scene object. Kato and Ninomiya [3] reported their learning algorithm using a template matching method with modified quadratic discriminant functions. They exploited binocular vision images to remove the perspective effect from incoming images and to remap each pixel of these images toward new position. Sun *et al.* [4] utilized a Gabor filter bank for feature extraction in vehicle detection. To improve detection performance, they optimized these Gabor filters by genetic algorithms. Suzuki and Kanade [5] constructed a vehicle dynamics model and calculated the optical flow from sequential images. They further used extended Kalman filter to estimate the motion and orientation of vehicles. Tokoro discussed the advantages and disadvantages of laser and microwave radars in gauging the distance of automobile [6]. He also showed the application of radars in intelligent cruise control system. Recently, the combination of radar and vision in detecting obstacles was also implemented and reported. As

proposed by Kato *et al.* [7], the role of the vision sensor was to detect the boundaries of objects, whereas the distances to obstacles were measured by the radar. A stereo image based obstacle detection and distance measurement approach was adopted in Ruichek's study [8]. In stead of matrix cameras, the algorithm used two linear cameras to capture a pair of stereo images and then performed an edge matching.

In this paper, we will propose a framework to provide a practical solution to the detection of preceding vehicles and then the estimation of the distance. The reliability of vehicle detection is enhanced due to the use of road region segmentation. In addition, we evaluate the proposed distance estimation algorithm by comparing with the measurement from a radar sensor.

II. THE FUZZY ART BASED SCENE ANALYSIS

In a companion paper [1], we have developed a scene segmentation system capable of automatically classifying and labeling the objects in images via image pixel features. We first construct a fuzzy rule base to analyze the scene and then utilize this rule base to classify pixels of the given traffic scene images. Fuzzy approach is adopted so that the module can be better tolerant of uncertainty, ambiguity, irregularity, and noise existent in the image. To obtain suitable fuzzy sets of features, we use fuzzy Adaptive Resonance Theory (fuzzy ART) to summarize the grouping nature among pixel features. Unlike other clustering approach, such as fuzzy *c*-means, fuzzy ART can produce appropriate number of clusters subject to the specified setting of vigilance value.

A. Fuzzy ART Clustering to Construct Membership functions

Three inherent features, including hue, intensity, and height of image pixels, are extracted from a given $M \times N$ image in the proposed scene analysis module. For fuzzy ART requirement, we encode the feature vector of the i -th pixel using the complement coding form as $I_i = (a_i^1, a_i^2, a_i^3, a_i^{1c}, a_i^{2c}, a_i^{3c})$, where a_i^1 , a_i^2 , and a_i^3 are the pixel's hue, intensity, and height, respectively. Furthermore, a_i^{1c} , a_i^{2c} , and a_i^{3c} represent their corresponding complements. It is noted that a_i^1 , a_i^2 , and a_i^3 should be normalized in advance. We use fuzzy ART to cluster MN I_i 's. After the fuzzy clustering, the pixels of the image can be divided into several clusters.

We choose the Gaussian type membership function to represent the features because the Gaussian type membership

Manuscript received March 30, 2006. This research was supported in part by the MOE under grant EX-91-E-FA06-4-4, by the MOEA under grant 93-EC-17-A-02-S1-032, and by the NSC under Grant NSC 92-2213-E-009-112, Taiwan, R.O.C.

Jyh-Yeong Chang* and Chien-Wen Cho are with the Department of Electrical and Control Engineering, National Chiao Tung University, Taiwan, R.O.C. (phone: +886-3-5712121 ext. 54336; e-mail: jychang@mail.nctu.edu.tw; address: 1001 Ta-Hsueh Road, Hsinchu, Taiwan 300, R.O.C.).

function can reflect the first order and second order statistics of clusters and is differentiable. As stated earlier, after the fuzzy ART clustering, the image pixels are assigned into different clusters. The corresponding mean and standard deviation of each cluster can be computed. Therefore, we can obtain associated Gaussian type membership functions and linguistic labels can be used to describe the features' characteristics.

B. Fuzzy Rule Base Extraction for Scene Analysis

Fuzzy rules can be generated by learning from examples. A training pattern with feature vector (a^1, a^2, a^3) is associated with the desired output of corresponding natural elements. Such image pixel constitutes an input-output pair. The rules generated are a series of associations of the form

"IF antecedent conditions hold, THEN consequent conditions hold."

The number of antecedent conditions equals the number of features. Note that antecedent conditions are connected by **"AND."** For illustrative purpose, assume now we have three linguistic labels, HIGH, MIDDLE, and LOW to describe the pixel's hue; three labels, BRIGHT, GRAY, and DARK to describe the pixel's intensity; and three labels, UP, MIDDLE, and DOWN to describe the pixel's height. For example, a pixel, its hue, intensity, height, and belonging categories being concatenated as vector format, is given by:

$$[a^1, a^2, a^3; D] = (0.80, 0.75, 0.93; SKY) \quad (1)$$

where a^1 , a^2 , and a^3 denote the normalized hue, intensity, and height of the pixel, respectively, and D is the corresponding belonging object category of the pixel. Let a^1 , a^2 , and a^3 map their maximal membership functions, HIGH, BRIGHT, and UP to values 0.90, 0.72, and 0.56, respectively. Hence, this pixel supports the rule of

"IF the pixel's hue is HIGH AND its intensity is BRIGHT AND its height is UP, THEN the pixel is SKY," (2)

with firing strength 0.56.

Due to a large number of training pixels, some conflicting rules may be generated. The conflicting rules have the same antecedent conditions but lead to different consequent conditions (for example, the pixel is ROAD or the pixel is TREE). For a set of antecedent conditions, we can have only one rule to reflect it. Therefore, we have to choose one from the two or more conflicting rules from each qualified cluster. To this end, we choose the rule that is supported by a maximum number of examples. Furthermore, to prune redundant or inefficient fuzzy rules, if the supporting pixels of a rule are less than a small fraction, 5% being used in our study, of the total pixels of the image, the rule is excluded from defining an **IF-THEN** rule. After the fuzzy rule base is

established, we use the max-min inference to classify the elements of images. Although fuzzy inference using fuzzy rule base is very promising in scene image segmentation, there are still erroneously classified pixels existing after our rule-base classification. These wrongly classified pixels occur most frequently across the boundary of two natural element categories. To reduce these mis-classified pixels, we adopt the modified fuzzy K -Nearest Neighbor (K -NN) algorithm [9] to correct the erroneous pixels.

III. CAR DETECTION AND DISTANCE ESTIMATION S

In the above section, we have summarized a scheme [1] to analyze and segment a forward-looking traffic scene. Accordingly, an important application of scene analysis to obstacle detection and collision avoidance for safety driving can be easily facilitated with a vision-based distance measurement scheme as described below.

A. Car Detection

Vehicles, which can have various colors, are difficult to detect using only color features, but they often can be better detected by some ground-truth cues described below.

- 1) Vehicles must be on the road.
- 2) Vehicles have shadows under the car bodies.
- 3) The height-width ratios of vehicles vary in a certain range, and are suitably represented by circumscribing rectangles.

For illustration purpose, as shown in Fig. 1(a), a forward-looking traffic scene image with some vehicles on the road. The road region in gray color is shown in Fig. 1(b). For car detection in a faster and reliable manner, we can limit the searching region of interest to a smaller area, i.e., the road, rather than the whole image. The shadow of a vehicle is shown in Fig. 1(c). Finally, as shown in Fig. 1(d), the preceding car is circumscribed with a rectangle. In the above, Cues 1) and 2) can be helpful to assure the car edges by some edge finding routines. Cue 3) is useful for car region refinement.

B. Camera Model

The camera model with front image plane [10] is illustrated in Fig. 2. The camera is represented by a lens together with an image plane lying a distance of f beyond the lens. The image v_p of a point v in the three-dimensional space is determined by the intersection of the image plane with the projecting ray defined by v and the lens center. Also we have aligned the Y axis with the optical axis of the camera, which is the ray from the lens perpendicular to the image plane. It is noted that the center of coordinates is the intersection of the principal ray with the image plane. Clearly, if we let $v = (x, y, z)$ and $v_p = (x_p, y_p, z_p)$, we have

$$x_p = \frac{f \cdot x}{f + y}$$

$$y_p = 0$$

and

$$z_p = \frac{f \cdot z}{f + y}. \quad (3)$$

C. Camera Alignment

The camera should be first calibrated before the proposed system starts to estimate the distance from the hosting vehicle to the preceding one. As inspired by Wang and Tsai [11], a camera can be calibrated by the concept of vanishing lines. In our application, we can further simplify the alignment process as described below. We consider three angular parameters of the camera, i.e., the pan, tilt, and swing angles; three shift parameters: the depth, horizontal, vertical offsets; and finally the focal length.

1) Alignment of the Pan, Tilt, and Swing Angles of the Camera

Consider a three-dimensional straight line L expressed as the intersection of two planes

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \end{aligned} \quad (4)$$

where $a_1, b_1, c_1, d_1, a_2, b_2, c_2,$ and d_2 are constants. From (3) and (4), it follows that, if y approaches infinite, the vanishing point will be

$$\lim_{y \rightarrow \infty} (x_p, y_p, z_p) = \left(-f \frac{b_1}{a_1}, 0, -f \frac{b_2}{a_2}\right). \quad (5)$$

In addition, for any straight line parallel to the optical axes, (4) can be reduced to

$$\begin{aligned} x &= d_1 \\ z &= d_2. \end{aligned} \quad (6)$$

And, hence, the vanishing point can be also reduced to

$$\begin{aligned} \lim_{y \rightarrow \infty} x_p &= 0 \\ \lim_{y \rightarrow \infty} y_p &= 0 \\ \lim_{y \rightarrow \infty} z_p &= 0. \end{aligned} \quad (7)$$

We should keep in mind that we intend to align the camera to reduce the affection of pan, tilt and swing angle of the camera. That is, we can adjust the pan, tilt, and swing angle of the camera so that if the car is in a position parallel to the lane lines, the optical axis will also parallel to these two lane lines. By (5), it is obvious that, we have to change the pan and tilt

angles of the camera until the projection on the image plane satisfies (7).

Finally, it follows from (3) that, without swing angle, the projection of two points with the same height will also have the same height in the image plane. We summarize the steps to align the pan, tilt, and swing angles of the camera as follows:

- i) Set the camera in a flat plane with two parallel lane markings.
- ii) Adjust the pan, tilt angles so that the lane markings crossing together in the origin in the image plane.
- iii) Place two markings with the same height and adjust the swing angle so that the height of the two markings are the same on the image plane.

2) Estimation of the Required Parameters

Once the pan, tilt, and swing angles of the camera are aligned, we then estimate the depth, horizontal, vertical offsets, and finally the focal length. We denote the horizontal, depth, and vertical offset parameters from the camera coordinate system to the real world coordinate system by Δx , Δy , and Δz . We then place several markings of different distance far from the camera and record their vertical projections in the image plane. Using (3) and replacing y with $y + \Delta y$ and z with Δz respectively, we can estimate, in the least square error sense, $f\Delta z$ and $f + \Delta y$ accordingly, where Δy and Δz are the depth and vertical offset parameters, respectively. In fact, we do not need to calculate the depth, horizontal, and vertical offsets as well as the focal length at all. Only $f\Delta z$ and $f + \Delta y$ is indeed needed to calculate the distance from the camera. To estimate these two parameters, we can collect our experimental data and solve for $f\Delta z$ and $f + \Delta y$ by the least square regression. Finally, the formulas to estimate the location of an object from the camera are

$$y = \frac{f\Delta z - z_p(f + \Delta y)}{z_p} \quad (8)$$

and

$$x = \frac{x_p(f + y + \Delta y)}{f} \quad (9)$$

where Δy and Δz are the depth and vertical offset parameters, respectively.

IV. EXPERIMENTAL RESULTS

A. System Installation and Calibration

In Fig. 3 is shown the experimental host car, a Mitsubishi Savrin equipped with a Sick LMS 291-S05 laser range finder, which has measurement resolution of 10 mm, measurement accuracy of ± 35 mm and operating range of 30 m, in the front end of the car and a CCD camera mounted inside the car. The proposed algorithm, including scene analysis and distance estimation, was implemented on a Centrino 1.5 GHz

Notebook. The host vehicle was first parked in a flat road with lane markings, and the image of the road is taken. Next, the Sobel filter was applied to the image to find out the edges of the path lines, and then Hough transform was used to determine the path lines and their interception. By the camera alignment procedures mentioned earlier, the camera was aligned with its pan and tilt angle so that the interception of path lines coincided the center of the image. Afterwards, the alignment of the camera's swing angle was done by tuning and checking if two markings of the same height projected to the same height in the image. Finally, a preceding car was stopped at 20 different places, and the distance and projection in the image were recorded so that the parameters $f\Delta z$ and $f + \Delta y$ were estimated by the least square regression. Note that in our experiment, the unit of length was chosen to be cm. Thus we had $f + \Delta y = -129.12$ and $f\Delta z = -43323$.

B. Fuzzy ART Based Traffic Scene Analysis

Twenty 256×192 color scene images taken on a highway were used to train the fuzzy rule base for scene analysis. In the first step, we construct the membership functions for the fuzzy rule base by fuzzy ART clustering. The means, standard deviations, and resultant membership functions of three features defined for each category were computed accordingly. For brevity, we only show one examples in Fig. 4. Fig. 4(a) is one of the traffic scene image. Fig. 4(b) is the segmented output image by the fuzzy rule base learned. In Fig. 4(c), to further correct the erroneous pixels, the image is processed by the modified fuzzy K -NN algorithm. In Fig. 4(d), the detected vehicles are circumscribed by rectangles. As shown in Fig. 4(e), the pixels inside the rectangles are all assigned to VEHICLE class, whereas each pixel outside those rectangles is assigned to the class having the second high score if it is mis-classified to VEHICLE class through fuzzy rule inference.

After vehicles were detected, the distance from the vehicle's bottom edges to the center of the image could be computed. Since we had estimated the required two parameters $f\Delta z$ and $f + \Delta y$ during calibration procedure, we then used (8) and (9) to estimate the distances of the preceding cars from the host car. In addition, the radar also returned measurement values per second in different angles from 0° to 180° . The corresponding distances of the front car from the host car were calculated accordingly. An experimental set-up showing the results of our distance estimation and the measurement values of the radar is shown in Fig. 5. An image from image sequences is shown in the upper left part of the figure. The bottom edge of the front car was detected in the neighborhood of the segmented car region. To reduce processing time of the distance estimation, we did not present the segmentation result when running distance estimation program. The detected bottom edge of the front car is marked in red, and the estimated distance was shown in the lower left corner. For comparison, the radar read out plot was drawn in the upper right part of the figure. Also

as shown in Table I, we verified our vision based front vehicle distance estimation algorithm by comparing with the read out values of the radar equipment. The experiment was done in a scenario that the host car followed the front car, driving below 60 km/hr. We ignored the distance values that change suddenly either from our algorithm or the radar. In this experiment, when the distance exceeds approximate 27 m, according to Table 1, the error became larger since the projection of a farther point tends to converge to the center of the image plane; and such small deviation from the image center could induce great error in the distance estimation. Thus, for car's bottom edge close to the origin of the image, invisibility caused by image capturing or vibration noise of the camera would cause more serious error. For front vehicles whose distances were less than about 27 m from the host car, the error can be less than 5.46%. Moreover, we have observed that, when the host car is close (less than about 8 m) the front vehicle, the error is also larger. This is because that the least square regression puts more emphasis on larger error terms, caused from larger distance values, and thus results in a relatively better estimation in farther distance cases. In consequence, the proposed vision based distance estimation has a better accuracy when the front vehicle falls between 8 and 27 m ahead.

V. CONCLUSION

In this paper, we have proposed a framework which can provide a practical solution to the detection of preceding vehicles and then the estimation of the distance. In a companion paper, we have developed a scene analysis module that deals with scene segmentation and natural object labeling of forward-looking images by the use of fuzzy Adaptive Resonance Theory and fuzzy inference techniques. Based on this technique, the proposed system can detect the front vehicles and then estimate the distance of the nearest car from us. We demonstrate our system using road scene images as a test-bed. We detect cars head and estimate the distances of the preceding cars. The car detection and the distance estimation error are tested to be very reliable and promising in our experiment.

REFERENCES

- [1] J. Y. Chang and C. W. Cho, "Vision-Based Forward-Looking Traffic Scene Analysis Scheme," submitted to *IEEE Int. Conf. Intell. Transport. Syst.*, 2006.
- [2] A. Broggi *et al.*, "Visual perception of obstacles and vehicles for platooning," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, no. 3. pp. 164–176, Sept. 2000.
- [3] T. Kato and Y. Ninomiya, "An approach to vehicle recognition using supervised learning," *IEICE Trans. Inform. Syst.*, vol. E83-D, no. 7, pp.1475–1479, 2000.
- [4] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Trans. Intell. Transport. Syst.*, vol. 6, no. 2, pp. 125–137, 2005.

[5] T. Suzuki and T. Kanade, "Measurement of vehicle motion and orientation using optical flow," in *Proc. IEEE Int. Conf. Intell. Transport. Syst.*, 1999, pp. 25–30.

[6] S. Tokoro, "Automotive application systems of a millimeter-wave radar," in *Proc. IEEE Intell. Vehic. Symp.*, Tokyo, Japan, Sept. 1996, pp. 260–265.

[7] T. Kato, Y. Ninomiya, and I. Masaki, "An obstacle detection method by fusion of radar and motion stereo," *IEEE Trans. Intell. Transport. Syst.*, vol. 3, no. 3, pp. 182–188, 2002.

[8] Y. Ruichek, "Multilevel- and neural-network-based stereo matching method for real-time obstacle detection using linear cameras," *IEEE Trans. Intell. Transport. Syst.*, vol. 6, no. 1, pp. 54–62, 2005.

[9] J. Y. Chang and C. W. Cho, "Scene analysis system using a combined fuzzy logic-based technique," *J. Chinese Instit. Eng.*, vol. 25, no. 3, pp. 297–307, 2002.

[10] R.O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: John Wiley & Sons, pp. 379–382, 1973.

[11] L. L. Wang and W. H. Tsai, "Camera calibration by vanishing lines for 3-D Computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 4, pp. 370–376, 1991.

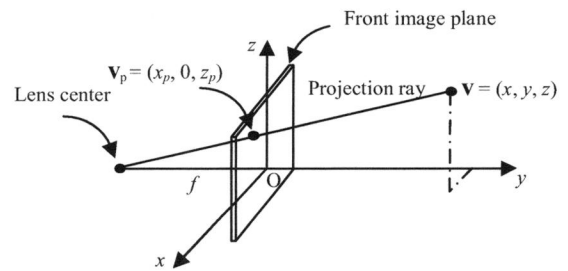


Fig. 2. The camera model.

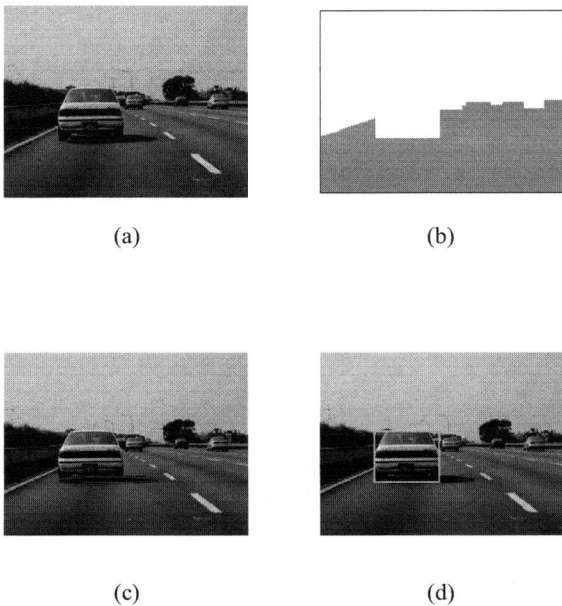


Fig. 1. (a) A forward-looking traffic scene image. (b) The road region is in gray color. (c) The shadow of the preceding car is marked. (d) The preceding car is circumscribed with a rectangle.



(a)



(b)

Fig. 3. The experimental set-up including (a) the host vehicle with the radar mounted on the front end and (b) the camera installed inside the vehicle.

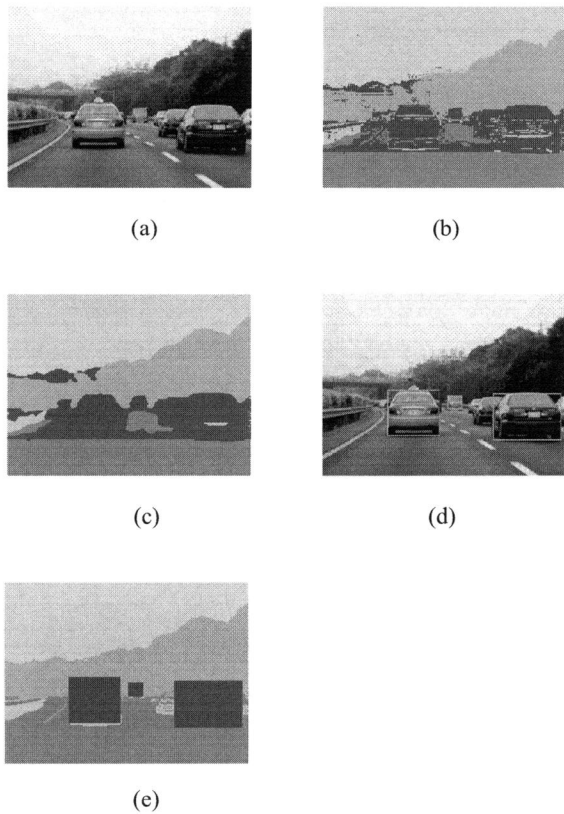


Fig. 4. An example selected from 20 test images and its segmentation results. (a) The original image. (b) The fuzzy ART based classification output image. (c) The segmentation result after fuzzy K -NN classification. (d) The detected vehicles. (e) The segmentation result combining vehicle detection.

TABLE I
COMPARISON OF THE DISTANCE ESTIMATION WITH RADAR
MEASUREMENT

Our estimate (m)	Radar value (m)	Error (m)	Error (%)
6.57	6.23	0.34	5.46
7.23	6.88	0.35	5.09
7.86	7.60	0.26	3.42
9.03	9.03	-0.00	0.00
9.17	9.35	-0.18	-1.93
10.13	10.21	-0.08	-0.78
10.92	11.07	-0.15	-1.36
11.61	11.95	-0.34	-2.85
11.61	12.04	-0.43	-3.57
12.69	12.79	-0.10	-0.78
13.33	13.34	-0.01	-0.08
14.03	14.24	-0.21	-1.48
14.83	15.54	-0.71	-4.57
15.73	16.15	-0.42	-2.60
17.95	17.67	0.28	1.585
20.13	19.85	0.28	1.411
21.92	21.86	0.06	0.274
28.37	27.24	1.13	4.148
34.62	31.17	3.45	11.07
37.39	34.48	2.91	8.44
AVG		0.58	3.04

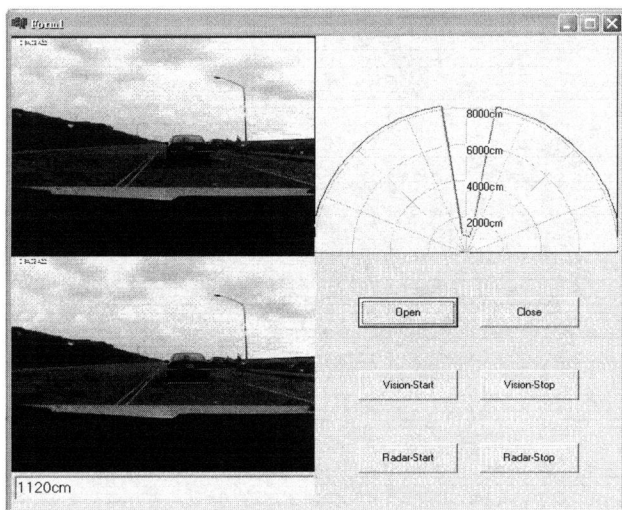


Fig. 5. The preceding car distance estimation program: the original image (upper left), the bottom edge of the preceding car and the estimate distance (lower left), and the radar read out plot (upper right).