Maneuvering Vehicle Tracking Based on Multi-sensor Fusion[1)]

CHEN Ying HAN Chong-Zhao

(*Institution of Synthetic Automation, Xi'an Jiaotong University, Xi'an* 710049)

(E-mail: chenying@mailst.xjtu.edu.cn)

Abstract Maneuvering targets tracking is a fundamental task in intelligent vehicle research. This paper focuses on the problem of fusion between radar and image sensors in targets tracking. In order to improve positioning accuracy and narrow down the image working area, a novel method that integrates radar filter with image intensity is proposed to establish an adaptive vision window. A weighted Hausdorff distance is introduced to define the functional relationship between image and model projection, and a modified simulated annealing algorithm is used to find optimum orientation parameter. Furthermore, the global state is estimated, which refers to the distributed data fusion algorithm. Experiment results show that our method is accurate.

Key words Information fusion, three-dimensional model, vision window, model matching, maneuvering target tracking

1 Introduction

The ability to track maneuvering targets is important to an intelligent vehicle navigation system. A vision sensor can provide wide-angle information and accurate angle measurements, but it is hard to produce accurate range measurements, which requires considerable computation costs. Radar can provide reliable longitudinal measurements, but it is hard to find lateral position, therefore it may easily lose the turning targets. Multi-sensor fusion can achieve synthetic environmental descriptions by integrating the redundant data from multiple sensors, which would significantly improve tracking system's reliability.

The previous related work overlooked the accurate angle measurements of image sensors[1,2], and the linear model adopted for vehicle moving description was hard to satisfy the complication of real-world vehicle motion. The paper presents a new tracking method based on radar and image fusion. An adaptive vision window, which improves accuracy of position and narrows down the working area of image processing, is established by fusing radar-based estimation and image intensity. Target's orientation is obtained via three-dimensional (3D) wire-frame model, improved Hausdorff distance and simulated annealing algorithm. Finally, data fusion technology is adopted to efficiently track strongly maneuvering targets by making full use of radar's accurate range and relative speed measurements, and image's accurate angle measurements.

2 System description

In order to satisfy the complicated motion property of the targets, the "bicycle" model is adopted, which describes the relationship between the rigid body motion of the vehicle and the steering and drive rates of wheels[3]. The relative kinetic equations denoting the relationship between the tracked vehicle and the host vehicle are

$$\begin{cases} \dot{x}_r = v \cdot \cos\theta - v_h \cdot \cos\theta_h \\ \dot{y}_r = v \cdot \sin\theta - v_h \cdot \sin\theta_h \\ \dot{v} = a \\ \dot{\theta} = \dfrac{v \cdot \tan\varphi}{B} \\ \dot{\varphi} = b \end{cases} \tag{1}$$

where B is the wheel-base length, a and b are the control parameters, x_r and y_r are the range coordinate data, v is the velocity of the tracked vehicle, θ is the orientation of the tracked vehicle, φ is the orientation of the front wheel; v_h is the velocity of the host vehicle measured by velocity sensor, and θ_h is the orientation of the host vehicle measured by angle sensor.

To make full use of the data from radar and image sensor, a fusion framework illustrated in Fig. 1 is proposed:

1) Supported by the Special Funds for Major State Basic Research Program of P. R. China (2001CB309403)
Received May 2, 2004; in revised form October 5, 2004

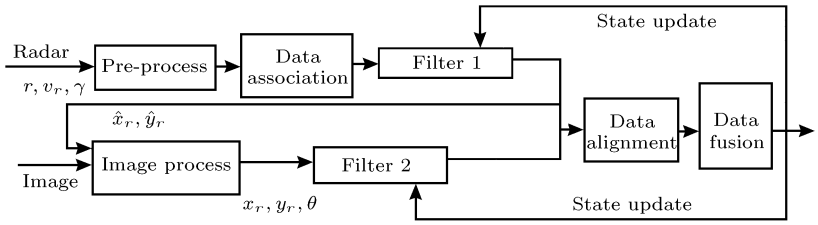


Fig. 1 Diagram of the fusion system

3 Vision window

The range information comes from a radar sensor. When the vehicle is not directly ahead, the radar performance declines. To solve the problem, the paper uses Kalman filter to post-process the radar sensor′s output, and meanwhile establishes a vision window to "capture" the obstacle by fusing filter estimation with image intensity.

Given an image, the main factors effecting the tracking accuracy come from three aspects[4]: intensity segmentation algorithm, window center position, and window size.

A common segmentation method is to set threshold value for the image. The system takes the most class variance criteria to identify the best threshold value[5].

The target centroid estimation error is mainly produced by the deviation between the window center and the target centroid. In order to make the former approach to the latter, we use the deviation as feedback to adjust the position of the window center. The original position of window center is the estimated position of radar-based filter, and the adjustment is

$$\boldsymbol{x}_0(k+1) = \boldsymbol{x}_0(k) + \mathrm{int}\{\hat{\boldsymbol{e}}(k) + (1-c)\mathrm{sign}[\hat{\boldsymbol{e}}(k)]\} \tag{2}$$

where c is the comparison threshold, int is the rounding operator, and sign is the symbolic operator. $\boldsymbol{x}_0(k) = (x_0(k), y_0(k))$ is the window center position of the k^{th} step, and $\hat{\boldsymbol{e}}(k) = [\hat{\varepsilon}_x(k) \quad \hat{\varepsilon}_y(k)]^{\mathrm{T}} = [\hat{x}_t(k) - x_0(k) \quad \hat{y}_t(k) - y_0(k)]^{\mathrm{T}}$ is the deviation of the step.

The window size is enlarged along with the change of window center position. The original size is set to be 10×10 pixels. Window size (w_x, w_y) is getting larger as the window center approaches to the target centroid, satisfying

$$w_x(k+1) = w_x(k) + \alpha \cdot \max|i - x_0(k)|, \quad w_y(k+1) = w_y(k) + \beta \cdot \max|j - y_0(k)| \tag{3}$$

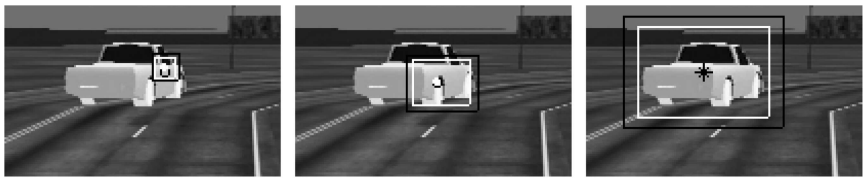
where (i, j) is the target pixel of the studied area. α and β are the weaken factors to limit speed increment of the window size, which are usually set to $0.15 \sim 0.35$.

To determine the final parameters (x_0, y_0, w_x, w_y), a function is established in window w

$$F(w) = \sum_{k=1}^{n} \max_{i,j \in w}(Grad_{i,j}, n) \tag{4}$$

where $Grad_{i,j}$ is the intensity gradient value of pixel (i, j), which can be obtained by gradient operators such as Sobel; n is the number of sample points.

The parameter set (x_0, y_0, w_x, w_y) that maximize $F(w)$ is the result. The processes and the result of the experiment are illustrated in Fig. 2, in which the two rectangles separately denote the background



(a) Original state (b) An adjusted state in-process (c) Final result

Fig. 2 Vision window

window (bigger one) and the target window, the hollow point denotes the current target centroid, the solid point denotes the current window center, and the asterisk denotes the final position of the window center that is almost coincident with the target centroid.

In general traffic scenes, vehicles are constrained to be in contact with the ground plane, and their height can be neglected, considering that their magnitude is far smaller than that of the distance between vehicles and cameras. Therefore, targets position can be determined by 3D reconstruction of window center (x_0, y_0).

4 Orientation

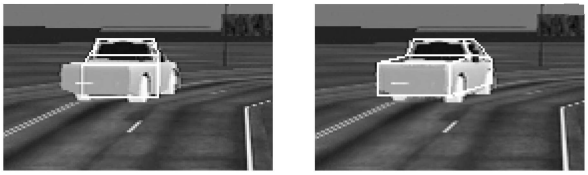
Orientation is obtained *via* 3D method. Traditional methods usually involve an initial stage of obtaining a symbolic description of the image in terms of predefined 2D features, and recognition is then achieved by making a match between sets of 2D image features and sets of similar 3D features[6,7]. Since the feature extraction and matching process are error-prone and time-consuming, they are hardly to satisfy practical applications[8]. In the paper, a new modified Hausdorff distance (HD), which uses dominant points instead of edge maps as features for measuring similarity between image and model projection, is employed to establish the matching function. The orientation is then achieved by optimizing the function.

Because it is not necessary for HD to build point-to-point correspondence between the model and image, HD is more tolerant to perturbations in point location than other techniques[9]. The traditional HD is sensitive to outlier points. Dubuisson *et al.* investigated 24 forms of different Hausdorff distances and indicated the modified Hausdorff distance (MHD)measure had the best performance[10]. In the paper, MHD is improved and employed to match dominant points rather than binary pixels[11]. The new modified Hausdorff distance (M^2HD) is defined as:

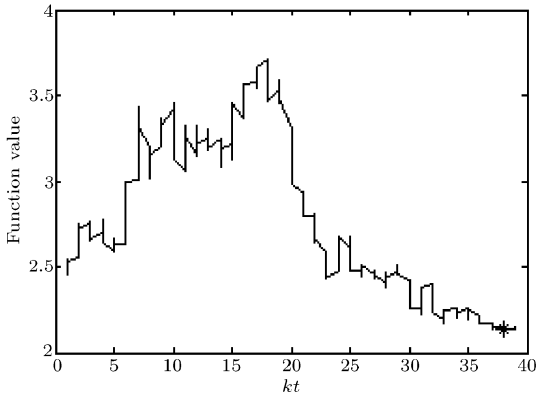
$$h(T, M) = \frac{1}{\sum\limits_{t_j \in T} w_{t_j}} \sum_{t_j \in T} w_{t_j} \cdot \min_{m_i \in M} \|t_j - m_i\| \tag{5}$$

where w_{t_j} denotes merit of point t_j, and $\| \bullet \|$ denotes the distance from a point to a line segment.

Since the function is complicatedly nonlinear and multi-variable, the paper uses simulated annealing algorithm with a mnemon and variant frequency to get the optimum solution[12]. The results are illustrated in Fig. 3 where "$*$" denotes the optimum.



(a) Original orientation (b) Adjusted orientation



(c) SA optimization

Fig. 3 Orientation searching

5 Data fusion

Convert (1) into a discrete-time state transition with T as the sampling interval. The state vector at a time k is defined as $\boldsymbol{x}(k) = (x_r(k), y_r(k), v(k), \theta(k), \varphi(k))^{\mathrm{T}}$, and the control vector as $\boldsymbol{u}(k) = (a(k), b(k))^{\mathrm{T}}$ with the model

$$a(k) = \bar{a}(k)(1+\delta_q) + \delta_a, \quad b(k) = \bar{b}(k)(1+\delta_s) + \delta_b \tag{6}$$

where δ_a, δ_b are disturbance errors and δ_q, δ_s are multiplicative errors that separately reflect the increased uncertainty in vehicles motion as speed and steer angles change. $\delta_a, \delta_b, \delta_q, \delta_s$ are modeled as constant, zero mean, uncorrelated white sequences with constant variances, respectively.

The state equation is defined as

$$\boldsymbol{x}(k+1) = f(k, \boldsymbol{x}(k), \boldsymbol{u}(k)) + \boldsymbol{v}(k) \tag{7}$$

where the process noise $\boldsymbol{v}(k)$ is a Gaussian noise vector with zero mean and variance matrix $\boldsymbol{Q}(k)$.

The radar measurement is:

$$\begin{aligned} &\boldsymbol{z}_1(k) = \boldsymbol{h}_1(k, x(k)) + \boldsymbol{w}_1(k) \\ &\boldsymbol{h}_1(k, \boldsymbol{x}(k)) = \begin{bmatrix} r(k) = (x_r^2(k) + y_r^2(k))^{1/2} \\ v_r(k) = v(k) - v_h(k) \\ \gamma(k) = \tan^{-1}\left(\dfrac{y_r(k)}{x_r(k)}\right) \end{bmatrix} \\ &\boldsymbol{w}_1(k) = (w_r, w_{v_r}, w_\gamma) \end{aligned} \tag{8}$$

where measurement noise w_r, w_{v_r}, w_γ are Gaussian white noise whose statistics are $(0, \sigma_r^2), (0, \sigma_{v_r}^2), (0, \sigma_\gamma^2)$.

The image measurement is:

$$\begin{aligned} &\boldsymbol{z}_2(k) = H_2(k)\boldsymbol{x}(k) + \boldsymbol{w}_2(k) \\ &H_2(k) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \\ &\boldsymbol{w}_2(k) = (w_{x_r}, w_{y_r}, w_\theta) \end{aligned} \tag{9}$$

where measurement noise $w_{x_r}, w_{y_r}, w_\theta$ are Gaussian white noise whose statistics are $(0, \sigma_{x_r}^2), (0, \sigma_{y_r}^2), (0, \sigma_\theta^2)$.

As for the system described by (7), (8), (9), a distributed data fusion algorithm[13] is adopted to get the fusion state estimation $\hat{\boldsymbol{x}}_F(k)$ and its covariance $P_F(k)$ at time k based on global information

$$\hat{\boldsymbol{x}}_F(k) = \hat{\boldsymbol{x}}_F(k|k-1) + P_F(k) \cdot \sum_{i=1}^{2} \{P_i^{-1}(k)[\hat{\boldsymbol{x}}_i(k) - \hat{\boldsymbol{x}}_F(k|k-1)]\} \tag{10}$$

$$P_F^{-1}(k) = \sum_{i=1}^{2} P_i^{-1}(k) - P_F^{-1}(k|k-1) \tag{11}$$

where $\hat{\boldsymbol{x}}_F(k|k-1), P_F(k|k-1)$ are respectively one-step prediction estimation and error covariance based on global information, and $\hat{\boldsymbol{x}}_i(k), P_i(k)$ are respectively the state estimation and error covariance for ith $(i = 1, 2)$ sensor based on global information

$$\hat{\boldsymbol{x}}_F(k|k-1) = f(\hat{\boldsymbol{x}}_F(k)) \tag{12}$$

$$P_F(k|k-1) = F(k-1)P_F(k-1) \cdot F^{\mathrm{T}}(k-1) + Q(k) \tag{13}$$

$$\hat{\boldsymbol{x}}_i(k) = \hat{\boldsymbol{x}}_F(k|k-1) + K_i(k)[\boldsymbol{z}_i(k) - H_i(k-1) \cdot \hat{\boldsymbol{x}}_F(k|k-1)] \tag{14}$$

$$P_i(k) = (I - K_i(k)H_i(k-1)) \cdot P_F(k|k-1) \tag{15}$$

$K_i(k)$ in (15) is the gain matrix of filter i, $i = 1, 2$.

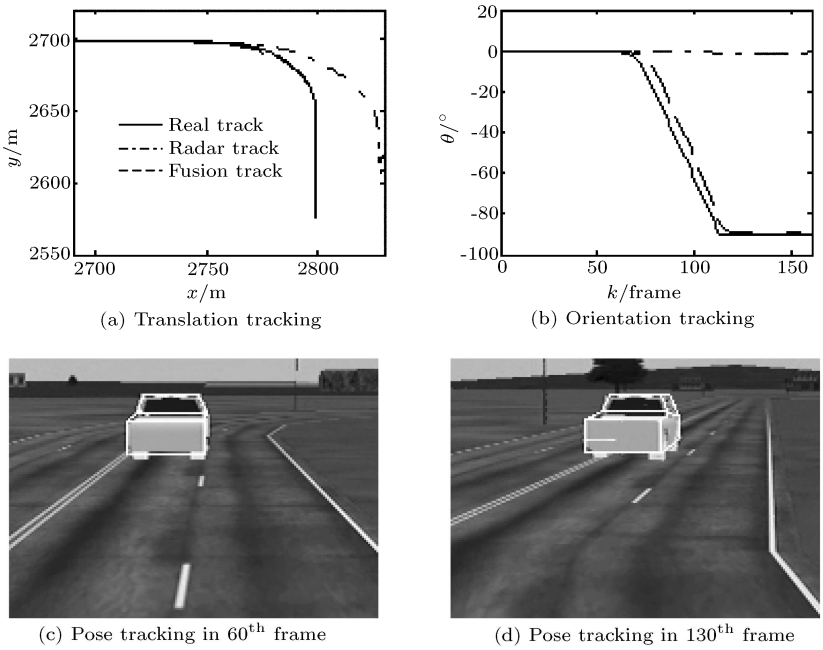
$$K_i(k) = P_i(k|k-1)H_i^{\mathrm{T}}(k-1) \cdot (H_i(k-1)P_i(k|k-1)H_i^{\mathrm{T}}(k-1) + R_i(k))^{-1} \tag{16}$$

In addition, $F(k-1) = \left.\dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_F(k-1)}$ is the Jacobians matrix of partial derivatives of $\boldsymbol{f}(\cdot)$, $H_1(k-1) = \left.\dfrac{\partial \boldsymbol{h}_1}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_F(k|k-1)}$ is the Jacobians matrix of partial derivatives of $h(\cdot)$, and $R_1 = \mathrm{diag}(\sigma_r^2, \sigma_{v_r}^2, \sigma_\gamma^2)$, $R_2 = \mathrm{diag}(\sigma_{x_r}^2, \sigma_{y_r}^2, \sigma_\theta^2)$ are respectively the measurement noise covariance matrices of sensor 1 and sensor 2.

6 Tracking experiments

To validate the approach proposed above, a 3D virtual reality is created using Vega as the scene arranger, and Visual V++6.0 as the development platform. The vehicle first moves with a constant velocity, and starts to turn after 60^{th} frame when its turning rate and acceleration change at all time. After 110^{th} frame, the vehicle again enters on the straight-line road.

With data and images from the simulated reality, experiments were taken. The results are shown in Fig. 4, which indicate that the moving vehicle can be accurately tracked even after a 90-degree turn.



(a) Translation tracking (b) Orientation tracking

(c) Pose tracking in 60^{th} frame (d) Pose tracking in 130^{th} frame

Fig. 4 Tracking results

7 Conclusion

The presented new maneuvering vehicle tracking scheme based on multi-sensor information fusion makes full use of the reliable longitudinal information from radar sensor and the accurate lateral information from image, and improves significantly the target tracking performance. The position measurement mainly comes from radar, and the lateral error is corrected by integrating image information. The angle measurement is achieved *via* improved 3D image matching which can reduce the image processing time and improve accuracy in measurement. Data fusion algorithm efficiently synthesizes the measurements from the two kinds of sensors, and achieves the accurate state estimation of preceding mobile target. The experiments confirm that the proposed fusion scheme has strong robustness for the target motion uncertainty.

References

1 Caveney D, Feldman B, Hedrick J K. Comprehensive framework for multisensor multitarget tracking in the adaptive cruise control environment. In: Proceedings of International Symposium on Advanced Vehicle Control, Tokyo: Society of Automotive Engineers of Japan, 2002. 697∼702

2 Becker J C, Simon A. Sensor and navigation data fusion for autonomous vehicle. In: Proceedings of the IEEE Intelligent Vehicles Symposium, Piscataway: IEEE, 2000. 156∼161

3 Alexander J C, Maddocks J H. On the kinematics of wheeled mobile robots. *International Journal of Robotics Research*, 1989, **8**(5): 15∼27

4 Ling Z Y. Dynamic Image Analysis. Beijing: National Defence Industry Press, 1999

5 Sonka M, Hlavac V, Boyle R. Image Processing, Analysis, and Machine Vision. New Delhi: Vikas Publishing House, 2002

6 Tan T N, Sullivan G D, Baker K D. Model-based localization and recognition of road vehicle. *International Journal of Computer Vision*, 1998, **27**(1): 5∼25

7 Koller D, Daniilidis K, Nagel H H. Model-based object tracking in monocular image sequences of road traffic scenes. *Internal Journal of Computer Vision*, 1993, **10**(3): 257∼281

8 Tan T N, Sullivan G D, Baker K D. Efficient image gradient-based vehicle localization. *IEEE Transactions on Image Processing*, 2000, **9**(8): 1343∼1356

9 Huttenlocher D P, Klanderman G A, Rucklidge W J. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, **15**(9): 850∼863

10 Dubuisson M P, Jain A K. Modified Hausdorff distance for object matching. In: Proceedings of 12th Internal Conference on Pattern Recognition, Piscataway: IEEE, 1994. 566∼568

11 Gao Y. Efficiently comparing face images using a modified Hausdorff distance. *IEE Proceedings of Vision, Image and Signal*, 2003, **150**(6): 346∼350

12 Kan L S, Xie Y, You S Y, Luo Z H. Non-numerical Parallel Algorithm-simulated Annealing Algorithm. Beijing: Science Press, 1994

13 He Y, Wang G H, Lu D J, Peng Y N. Multi-sensor Information Fusion and Application. Beijing: Publishing House of Electronics Industry, 2000

CHEN Ying Ph. D. candidate. Her research interests include intelligent traffic system and multi-sensor information fusion.

HAN Chong-Zhao Professor. His research interests include information fusion, remote sensing, fault detection and diagnosis, and nonlinear frequency-domain analysis.