IEEE TRANSACTIONS ON

# INTELLIGENT TRANSPORTATION SYSTEMS

A PUBLICATION OF THE IEEE INTELLIGENT TRANSPORTATION SYSTEMS SOCIETY

JULY 2016 VOLUME 17 NUMBER 7 ITISFG (ISSN 1524-9050)



*(Contents Continued on Back Cover)*

IEEE

*(Contents Continued from Front Cover)*

SHORT PAPERS



GUEST EDITORIAL



SPECIAL ISSUE PAPERS

# On-Road Vehicle Detection and Tracking Using MMW Radar and Monovision Fusion

Xiao Wang, Linhai Xu, Hongbin Sun, *Member, IEEE*, Jingmin Xin, *Senior Member, IEEE*, and Nanning Zheng, *Fellow, IEEE*

***Abstract*—With the potential to increase road safety and provide economic benefits, intelligent vehicles have elicited a significant amount of interest from both academics and industry. A robust and reliable vehicle detection and tracking system is one of the key modules for intelligent vehicles to perceive the surrounding environment. The millimeter-wave radar and the monocular camera are two vehicular sensors commonly used for vehicle detection and tracking. Despite their advantages, the drawbacks of these two sensors make them insufficient when used separately. Thus, the fusion of these two sensors is considered as an efficient way to address the challenge. This paper presents a collaborative fusion approach to achieve the optimal balance between vehicle detection accuracy and computational efficiency. The proposed vehicle detection and tracking design is extensively evaluated with a real-world data set collected by the developed intelligent vehicle. Experimental results show that the proposed system can detect on-road vehicles with 92.36% detection rate and 0% false alarm rate, and it only takes ten frames (0.16 s) for the detection and tracking of each vehicle. This system is installed on Kuafu-II intelligent vehicle for the fourth and fifth autonomous vehicle competitions, which is called "Intelligent Vehicle Future Challenge" in China.**

***Index Terms*—Intelligent vehicle, vehicle detection and tracking, sensor fusion, MMW radar, monocular camera.**

## I. Introduction

INTELLIGENT vehicle technologies have significantly improved owing to the effort of both the automotive industry and many other fields in recent years [1]. The increasing popularity of intelligent vehicles and the advanced driver assistance system (ADAS) offers the potential to significantly enhance the safety and convenience of drivers. Currently, ADAS has been installed in luxury cars and even in several entry-level cars as an intelligent vehicle application. In ADAS or intelligent vehicle systems, on-road vehicle detection and tracking is the first step to understand the surrounding environment. Its robustness and reliability can have a direct impact on the stability and safety of the entire system. Thus, the design of vehicle detection and tracking system has elicited considerable attention from both the academe and industry.

The millimeter-wave (MMW) radar and monocular camera are two mainstream sensors for vehicle detection and tracking in intelligent vehicles. The MMW radar detects objects by emitting a millimeter-wave radio signal and analyzing the frequency shift in reflections. It can operate fairly consistently in different weather conditions, hence many researches use it for vehicle application and navigation [2], [3]. The monocular camera is also widely utilized because it can provide a rich data source from which additional information and context can be surmised for advanced surround perception. In addition, given that cameras are less expensive than other sensors, vision-based vehicle detection for driver assistance has received considerable attention over the last decades [4]–[7].

Despite their advantages, the drawbacks of these two sensors make them insufficient when used separately for vehicle detection and tracking. The MMW radar's measurement is limited in terms of spatial resolution and is rather noisy because of false alarm detection [8]. The monocular camera is sensitive to light and weather conditions and requires sophisticated computational intensity to guarantee high accuracy; hence, these mono-vision methods [5], [9]–[11] either can not meet the real-time requirement or have poor detection rate. Given these limitations, the combination of these two sensors [12]–[15] is considered as an efficient means to significantly increase detection accuracy while reducing detection noise. However, the majority of previously proposed fusion approaches [12], [15] fuse detection results from these two sensors in the end stage. Although the detection results are improved, the computational intensity resulting from vision computing remains high. Hence, several studies proposed the use of machine learning methods for vision-based vehicle detection to improve detection accuracy with reduced computational intensity. Liu *et al.* [16] applied a SVM-based classifier to detect shadow segments below on-road vehicles. Chavez-Garcia *et al.* [17] and Vu *et al.* [18] used HOG features and boosting-based classifiers to detect vehicles, where helps from laser sensors are also needed. However, the resulting accuracy of machine learning methods is largely dependent on the training dataset of the experimental environment; thus, poor results may be obtained in an unfamiliar environment. To address this issue, Alessandretti *et al.* [19] and Kadow *et al.* [20] proposed the fusion of these two sensors in the early stage; the MMW radar performs detection in a region of interest (ROI) on the captured image and searches for vehicle features within the ROI. This type of fusion method can significantly reduce the computational intensity of vision computing and hence is suitable for practical use in intelligent vehicles. Nevertheless, the detection accuracy of this simple early-stage fusion is unsatisfactory, especially in scenarios that require high reliability.

Manuscript received March 15, 2015; revised July 22, 2015, December 13, 2015, and February 13, 2016; accepted February 18, 2016. Date of publication April 29, 2016; date of current version June 24, 2016. The authors gratefully acknowledge the financial support from National Natural Science Foundation of China (61231018). The Associate Editor for this paper was L. Li.

The authors are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: hsun@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2533542


Fig. 1. The hardware platform of our "Kuafu-II" intelligent vehicle.

This study is also interested in fusing the sensors of MMW radar and monocular camera for on-road vehicle detection and tracking. Compared with the previous fusion-based vehicle designs, the proposed fusion method employs a more robust and efficient vision-based vehicle detection in the ROI provided by MMW radar, to improve the detection accuracy and reduce the false alarm rate from MMW radar. Moreover, the trajectories for both monocular camera and MMW radar are generated and compared in the end stage to further improve the detection accuracy and eliminate the false alarm that is very difficult to distinguish by simply using vision-based methods. Therefore, the proposed vehicle detection and tracking method achieves a much better trade-off among detection rate, false alarm rate and computational efficiency. To evaluate the proposed vehicle detection and tracking system, real-world experimental data are collected with Kuafu-II intelligent vehicle platform. The experimental results demonstrate that the proposed system can detect on-road vehicles with a 92.36% detection rate of and 0% false alarm rate. Each potential vehicle detection and tracking only takes 0.16 s at the frame rate of 60 Hz.

The rest of this paper is organized as follows. Section II briefly describes the system platform and sensor calibration. Section III presents the proposed vehicle detection and tracking algorithm. The performance and efficiency of the proposed method are extensively evaluated in Section IV. Section V presents the conclusion.

## II. System Platform and Sensor Calibration

### A. System Platform

The intelligent vehicle platform for the presented techniques and experiments in this study is based on a 1.6 L Tiggo SUV manufactured by Chery Automobile Co. and called "Kuafu-II." Kuafu-II was developed in 2008 and participated in the 4th and 5th "Intelligent Vehicle Future Challenge (IVFC)," a unmanned vehicles competition organized by Natural Science Foundation of China (NSFC), since 2009. In the Kuafu-II intelligent vehicle, an MMW radar and a monocular camera are installed. The MMW radar is a Delphi ESR bi-mode radar mounted on the front bumper, and the camera is a PointGray FMUV-03MTC color camera mounted behind the front windshield (as shown in Fig. 1). The detailed specifications of these two sensors are presented in Table I. Both the MMW radar and monocular camera are connected to a notebook computer with an Intel i7 3.0 GHz CPU and 8 GB DRAM. The computer is placed in the trunk as the computation center. The MMW radar detects the vehicles in front of the vehicle and transmits the relative speed and position of objects to the computer through a CAN bus. The camera captures the road scene at the frame rate of 60 Hz and transmits the uncompressed video sequences to the computer through a USB cable. The notebook computer is in charge of sensing data processing, vehicle detection and tracking.

TABLE I
Specifications of MMW Radar and Monocular Camera

| MMW radar | | Camera | |
|---|---|---|---|
| Product model | Delphi ESR | Product model | PGR FMUV |
| Detection speed | 60Hz | Frame fate | 60Hz |
| Max objects | 64 | Resolution | 640×480 |
| Max distance | 175m | Focus length | 6mm |
| Detection range | 90°/20° | Horizontal FOV | 53.1° |
| Max speed | ±81.92m/s | Vertical FOV | 40.0° |

Fig. 2. Illustration of the spatial calibration process. The poles are regularly arranged and detected by the MMW radar and the monocular camera.

### B. Spatio-Temporal Calibration

The relationship between the detection results of MMW radar and the corresponding image regions can be described by the prospective transformation. MMW radar scans on a horizontal plane (radar plane) and produces the radial distance $r$, the angle $\theta$ and the relative radial velocity $v$ of each target in front of the equipped vehicle. The detected result can also be described as $(x, y)$, where $x = r\cos(\theta)$ and $y = r\sin(\theta)$. For the monocular camera, the detected position of the target can be regarded as $(u, v)$ on the image plane. A calibration method is proposed to estimate the transformation between radar plane and image plane through equation (1). The transformation matrix is illustrated in equation (2).

$$\begin{cases} u = \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + a_9} \\ v = \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + a_9} \end{cases} \tag{1}$$

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix}. \tag{2}$$

The procedure of the proposed calibration method is briefly discussed below. As shown in Fig. 2, three tall and thin poles are

regularly arranged and detected by the MMW radar. The image frame is simultaneously captured by the camera. Assuming that the reflector's reflection points on the detected pole are at the same height as the mounted position of MMW radar, locating the object detected by MMW radar on the image plane becomes convenient, and the pixel of the reflection point can be selected. With deferent spatial arrangements of poles, nine datasets of $(x, y)$ and $(u, v)$ are obtained for equation (3) to estimate the projective transformation matrix.

$$\begin{bmatrix} x-u*x & 0 \\ y & 0 \\ 1 & 0 \\ 0 & x-v*x \\ 0 & y \\ 0 & 1 \\ -u*x & -v*x \\ -u*y & -v*y \\ -u & -v \end{bmatrix}^T \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \end{bmatrix} = [0 \quad 0]^T. \qquad (3)$$

The frame rate of camera is 60 fps, which is the same as the detection speed of MMW radar. Temporal calibration was performed by externally triggering camera to allow for synchronous detection with MMW radar.

## III. Proposed Vehicle Detection and Tracking

### A. Overall Architecture for Radar and Vision Fusion

Human vision system presents an interesting and efficient obstacle detection and tracking system. Two types of cells in the retina, i.e., rods and cones, are involved in obstacle detection and tracking. The rods are highly sensitive to moving objects and can quickly identify the position of detected objects. The cones are highly sensitive to features, including color and texture, and is thus suitable for obstacle detection and tracking. These two types of cells interact with each other in the back-end process to further improve the accuracy of detection and tracking. Owing to the "collaboration" between the rods and cones, human vision system can efficiently achieve high-speed and accurate vehicle detection. MMW radar and camera in intelligent vehicles act similarly to the rods and cones in human vision system. This similarity motivates us to design the proposed radar and vision fusion system for on-road vehicle detection and tracking.

The overall architecture of the proposed detection and tracking approach is illustrated in Fig. 3. Similar to the rods in human vision system, MMW radar detects on-road vehicles and transmits the location and size of the region of interest (ROI) to the image sequences captured by the monocular camera. Then, the vision processing module generates a square boundary in the image frame according to the transmitted information on ROI and employs the active contour method to detect vehicles within the square boundary. If the active contour method fails, it is a false alarm of MMW radar and vision processing module should eliminate this detection; otherwise, the detected contour is utilized by the following vehicle tracking and trajectory generation in the video sequences. The radar processing module also generates a trajectory using the information from MMW radar. These two trajectories, generated from vision and radar, were then compared and verified to confirm whether the detection and tracking are valid. The interaction between vision and radar should be on the same reference plane. This condition can be achieved by employing projective transformation, the parameters of which are generated during sensor calibration.

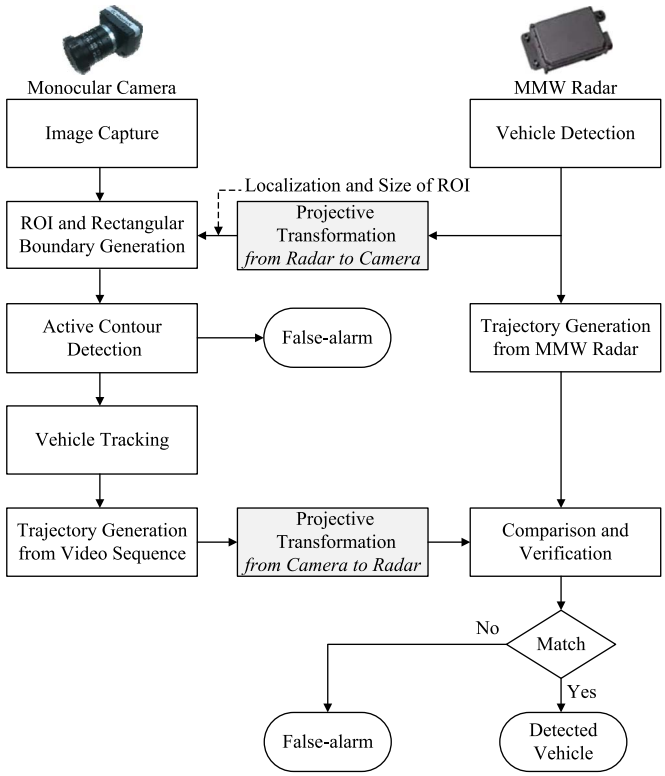



Fig. 3. Overall architecture of the proposed vehicle detection and tracking system.

Compared with conventional vision-radar fusion approaches, the proposed approach has the following advantages.

1) The computational intensity of vision processing can be significantly reduced by using the ROI provided by MMW radar.
2) The active contour detection method employed in vision processing can effectively eliminate the false alarm incurred by MMW radar.
3) Comparison and verification of trajectories generated by vision and radar can improve the overall accuracy of vehicle detection and tracking.

The overall architecture can achieve an optimal balance among accuracy of detection and tracking, real-time performance and computational efficiency.

### B. Vehicle Detection

Because of the accurate ROI provided by MMW radar, the sensitivity of vision processing to complex environment is largely compensated. Therefore, we can use the efficient feature detection method to detect vehicle within ROI. The proposed vehicle detection method consists of two modules, namely, rectangular boundary generation and active contour detection.

The rectangular boundary generation marks a rectangular of the vehicle according to the ROI, and the active contour technique detects a vehicle within the marked rectangular. The ROI is a square with side length 3 m. Its mapping in image sequences also depends on the distance of the vehicle. The symmetry features of ROI provide a remarkable clue for the existence of a vehicle. Owing to the visual appearance or aesthetics of automobiles, the image of the rear of the vehicle shows a visual appearance of symmetry about a vertical axis, which is usually similar to the vertical axis of ROI. Finding such symmetry axis to localize the vehicle is the first step after rectangular boundary generation. Conventional gray-level symmetry detection is based on the differences among gray-level intensities of pixels, which are time consuming and sensitive to noises in the outside environment. Broggi *et al.* [21] proposed a symmetry detection algorithm based on edge symmetry to reduce the computation cost and increase the detection accuracy. However, this method cannot efficiently discriminate the edge symmetries of the vehicle from that of other unrelated local details. In this study, we developed a symmetry detection algorithm using the statistics of symmetrical edge point pairs and a properly sized bounding box to generally represent the symmetrical relationship among pixels.

The proposed symmetry detection algorithm is described below. First, the Sobel operator is utilized to calculate the edge image. Second, for each line of the edge image, the numbers of symmetrical and non-symmetrical point pairs $\mathrm{Pos}(x)$ and $\mathrm{Neg}(x)$ with respect to $x$ point are calculated as follows:

$$\begin{cases} x = \frac{(x_i + x_j)}{2} \\ \mathrm{Pos}(x) = \mathrm{Pos}(x) + 1(x_i, x_j \in O) \\ \mathrm{Neg}(x) = \mathrm{Neg}(x) + 1(x_i \notin O || x_j \notin O) \end{cases} \tag{4}$$

where $O$ is the edge point space, $i$ and $j$ represent different columns of points, and $x_i$ and $x_j$ are point pairs selected from the same row in the covered bounding box whose height and width equal to half of the height and width of ROI, respectively. Third, $\mathrm{Pos}(x)$ and $\mathrm{Neg}(x)$ calculated from equation (4) are utilized to calculate the intensity of symmetry as follows:

$$Sr(x) = \frac{2\mathrm{Pos}(x)}{S(x)} \left(1 - \frac{2\mathrm{Neg}(x)}{W(x)}\right) \tag{5}$$

where $S(x)$ is the total number of symmetrical point pairs and $W(x)$ is the total number of point pairs. These two variables are theoretically two times larger than the maximum value of $\mathrm{Pos}(x)$ and $\mathrm{Neg}(x)$ in the same bounding box. Global symmetrical intensity is calculated with respect to the entire ROI using the following equation.

$$Sg(c) = \mathrm{Pos}(c) - \mathrm{Neg}(c) \tag{6}$$

where $c$ is the position of each vertical line in ROI. Finally, $Sr(x, w)$ in the same column and the corresponding $Sg(c)$ with $c$ equals to $x$ are weighted and summarized, the column with the peak value is the detected position of the vertical symmetry axis in ROI. This procedure can locate the position of the vehicle horizontal center with high accuracy under various outside environment.

Based on the assumption that the distribution of color in the rear image of the detected vehicle is symmetrical and simple, a histogram algorithm was utilized to precisely locate the left and right boundaries of the vehicle with the help of the detected center position. The histogram algorithm consists of three steps. Let $h$ and $w$ represent the height and width of ROI respectively.

- In step 1, the 32-bins 3D color histogram arrays $h_s$ in the local window $(h \times w)$ are counted as the histogram template of the detected vehicle.
- In step 2, the 32-bins 3D color histograms of left part $\mathrm{hist}_l(x)$ and right part $\mathrm{hist}_r(x)$ in local windows (from $h \times w/8$ to $h \times w/2$) are counted. The similarities between histogram template $\mathrm{hist}_s$ and candidate histograms $\mathrm{hist}_l(x)$ and $\mathrm{hist}_r(x)$ are computed with the following equation:

$$S(x) = \frac{\mathrm{hist}_*(x) \cdot \mathrm{hist}_s}{\|\mathrm{hist}_*(x)\| \, \|\mathrm{hist}_s\|}. \tag{7}$$

- In step 3, we calculate the leftmost position $x_l$ and rightmost position $x_r$ of $S(x)$ with regard to a threshold $\theta$. In order to get a proper value for $\theta$, a course-to-fine strategy is applied. We firstly divided the [0 1] interval into ten periods. After voting by similarity values, the period with highest vote is treated as the threshold region. Finally, $\theta$ equals to the weighted mean of this threshold region. Let $s$ is the center position of the detected vehicle. The detected left and right vertical boundaries of the vehicle can be described as $s - x_l$ and $s + x_r$ on the image plane.

The shadows under the vehicle are remarkable features that always exists under various light conditions. Hence, the localization of the bottom boundary of the vehicle in ROI was performed by detecting the position of the shadow under it. The appearance of a shadow is always a dark area in the image. Thus, the upper bound of this area can be efficiently located by calculating the first-order and second-order derivative of the gray-level intensity. We let $I$ represent the Sobel edge image and $(u, v)$ represent the 2D coordinates on the image plane. The calculation of the horizontal gradient $\mathrm{Grad}(x)$ of the gray-level intensity is from the bottom to the middle of ROI.

$$\mathrm{Grad}(x) = \sum_{u=1}^{W} \|I(u, x+1) - I(u, x)\|. \tag{8}$$

Then, for each row of the image, if $\mathrm{Grad}(x)$ is larger than $\sigma_1$ and the $\|\mathrm{Grad}(x) - \mathrm{Grad}(x-1)\|$ is larger than $\sigma_2$, current line $x$ is regarded as the candidate position for the bottom boundary. The calculation is continued until the next candidate position $x'$ is detected. If $x - x' < \sigma_3$, then $x'$ is discarded and the second step is repeated. Otherwise, $x$ is outputted as the detected position of the bottom boundary in ROI.

After these procedures, the rectangular boundary is generated by the left and right boundaries $s - x_l$ and $s + x_r$, the detected bottom boundary, and the top boundary of ROI. Inside this rectangular boundary, the active contour algorithm [22] is employed to extract the accurate contour of the rear image of the detected vehicle. We resized the rectangular boundary to

less than $50 \times 50$ pixels to reduce the computational intensity. Although this algorithm provides the contour of the vehicle for each radar detection, MMW radar may perform false detection and no vehicle exists in ROI. To address this, we further confirm the detection by analyzing the relative distance size and the rectangular boundary on the MMW radar plane and eliminate the detection with extremely large or small size.

### C. Vehicle Tracking

The proposed vehicle tracking design is a typical generative tracking method [23], which usually includes the following five components, i.e., target region, appearance representation, motion and position representation, method and model updating. In the proposed vehicle tracking method, the target region is described as a bounding box. The locality sensitive histogram approach and the uniform motion model are used for appearance and motion representation respectively. Histogram matching is chosen as the method, and the model updating is done by selecting the best matching position of the candidate histogram.

It should be noted that, the conventional image histogram does not have the spatial information of pixels and thereby is sensitive to noise. The locality sensitive histogram is more appropriate according to the evaluation in [24]. Hence, the locality sensitive histogram is employed for the appearance representation and can be described as follows.

$$\text{Weight}(u,v) = 1 - \frac{(u-c_u)(v-c_v)}{(u_{\max}-u_{\min})(v_{\max}-v_{\min})} \quad (9)$$

where $(u, v)$ is the 2D coordinates of the pixel, $(c_u, c_v)$ is the center position of the local window, and $(u_{\max}, v_{\max})$ and $(u_{\min}, v_{\min})$ are the bottom right and top left corners of the local window. In the meanwhile, the predicted position of the tracked vehicle is calculated on the MMW radar plane for a better accuracy, thus the projective transformation is deployed before and after the prediction.

The proposed vehicle tracking algorithm is described below in detail. We let $(u_1, v_1)$ and $(u_2, v_2)$ represent the mass center of the extracted rear image in the first and second image frames after the vehicle was detected. $R_1$ and $R_2$ represent the external rectangle of the considered rear image. First, $(u_1, v_1)$ and $(u_2, v_2)$ on the image plane are transformed to $(x_1, y_1)$ and $(x_2, y_2)$ on the MMW radar plane through projective transformation. Second, the uniform motion model is utilized to predict the position $(x_{3r}, y_{3r})$ in the next image frame, where two pre-calculated positions $(x_1, y_1)$ and $(x_2, y_2)$ are required.

$$\begin{bmatrix} x_{3r} \\ y_{3r} \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}. \quad (10)$$

After the prediction, the estimated position $(x_{3r}, y_{3r})$ on the MMW radar plane is transformed to $(u_{3r}, v_{3r})$ on the image plane through projective transformation. In the meanwhile, the rectangle $R_{3r}$ in the next frame with the same size as $R_2$ around the candidate position $(u_{3r}, v_{3r})$ is selected as the candidate area of the external rectangle. Next, the locality sensitive histogram in $R_2$ which is regarded as the histogram template of the detected vehicle, is calculated. The comparison between the histogram template and local histogram is operated repeatedly in each local window which has the same size as $R_{3r}$ and contains the candidate position $(u_{3r}, v_{3r})$ in the third frame. The similarity is calculated as equation (7). The matching window has three different scales, i.e., $0.9\times$, $1.0\times$, and $1.1\times$ with regard to the original window size, to obtain a more precise external rectangle $R_3$. Finally, the mass center of $R_3$ is selected as the predicted position $(u_3, v_3)$ of the detected vehicle in the third frame of the image sequences. This detection result is regarded as an input, and the estimating and matching processes are repeated for the next few frames to obtain consecutive detection results.

The trajectory of the detected vehicle on the image plane is constructed using the consecutive detection results above. This trajectory is further transformed to the MMW radar plane through projective transformation for the following comparison and verification work.

### D. Trajectory Comparison and Verification

Although the active contour method can eliminate the majority of false alarm detections from MMW radar, a few invalid detections can still exist. These can be further filtered by using trajectory comparison and verification. To achieve zero false alarm detection, the trajectories generated by the vision and radar processing modules are further compared and verified to confirm whether the detection and tracking are valid. The data association between radar and vision sensors is achieved by calculating the trajectory error which is defined as follows, where $F$ represents the total amount of frames during vehicle tracking, $n$ represents the index of frame number, $V(n)$ represents the trajectory generated from video sequences, $R(n)$ represents the trajectory generated from MMW radar and $E$ represents the trajectory error. $E$ is calculated by the Euclidean distance between two position trajectories and two speed trajectories estimated by adaptive Kalman filter [25].

$$E = \sqrt{\frac{1}{F}\left[\sum_{n=1}^{F}(V(n)-R(n))^2 + \sum_{n=1}^{F}\left(\frac{dV(n)}{dt} - \frac{dR(n)}{dt}\right)^2\right]}. \quad (11)$$

To evaluate the trajectory error $E$, the distribution of the detection error of each sensor is assumed to be Gaussian distribution. According to the data sheet of MMW radar, its detection standard deviation is 1.3 m. Meanwhile, the detection standard deviation of the proposed vision processing module is 2.36 m according to the statistics from our experimental dataset. The probability distribution of the trajectory error is shown in Fig. 4.

According to Fig. 4, the range of 97% confidence interval of this error is less than 2.4 m. Under this criterion, if the distance between two generated trajectories is more than 2.4 m, this case is treated as a false alarm detection of MMW radar. Although the active contour detection can effectively reduce the false alarm from MMW radar, the false alarm rate is still unsatisfactory, as false alarm is very annoying to intelligent vehicle or ADAS. False alarm after the active contour detection

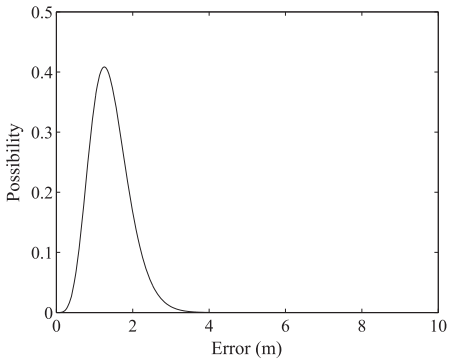



Fig. 4. Probability distribution of the trajectory error. The horizontal axis shows the calculated average Euclidean distances between trajectories according to the proposed criteria, and the vertical axis shows the corresponding probabilities.


Fig. 5. ROI estimation. Red squares indicate the calculated ROI dealing with vehicles in different ranges: (a) far range, (b) medium range, and (c) short range.

can be classified as two cases: (1) false alarms that exist during a very short period, e.g. less than two image frames; (2) false alarms that exists during a relatively long period. Both of these two cases can be addressed by using trajectory comparison and verification. In the case (1), vision processing can not generate the trajectory; In the case (2), the trajectory generated by vision processing is irregular and largely different from the trajectory from MMW radar, hence can be easily eliminated.

## IV. Experimental Results and Discussion

The proposed vehicle detection and tracking system was implemented in our intelligent vehicle Kuafu-II. Real-world radar and image sequences were captured by Kuafu-II for the experiment. Our dataset consists of 762,256 frames of the MMW radar's detection and corresponding video sequences under various traffic conditions, including many challenging scenarios such as poor illumination or low contrast. These video sequences can be considered as a general representation of the Chinese urban traffic environment. The vehicles detected by both sensors were manually labeled as ground truth to evaluate the performance of the algorithm.

### A. Vehicle Detection Results

Figs. 5–8 illustrate the procedure and results of the proposed vehicle detection, which includes ROI localization, symmetry detection, rectangular boundary generation, and active contour detection. The pre-calculated ROI is a square area whose mapping in the image depends on the distance between the targeting vehicle and the equipped vehicle. Fig. 5 depicts three calculated

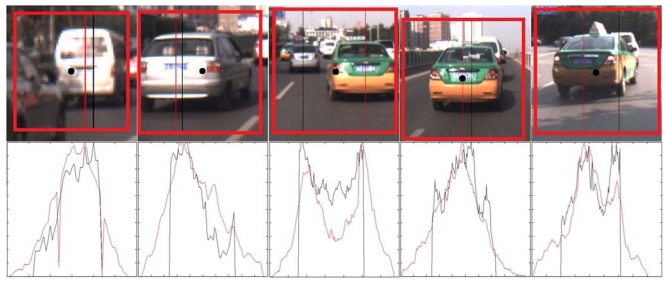


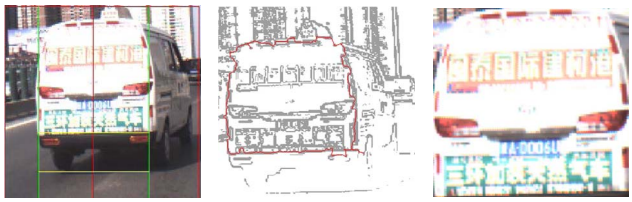
Fig. 6. Symmetry feature detection comparison between the proposed method and reference [21]. The red lines represent the detection results of the proposed method, and the black lines represent the detection results of [21].

Fig. 7. Rectangular boundary detection. Three detection results with different kinds of vehicle are illustrated. The red lines mark the central position. The left and right blue lines mark the detected horizontal boundaries. The yellow lines mark the detected vertical boundary for the bottom shadow.

Fig. 8. Active contour detection. The red curves mark the detected contour of the corresponding vehicle.

ROIs under different distance; the black dot represents the MMW radar's return point, and the red square around this point is the corresponding ROI. The size of ROI is carefully selected to fully cover the rear image of the detected vehicle. For the symmetry detection, which is the first step in rectangular boundary generation, Fig. 6 shows a comparison of the proposed symmetry detection algorithm and a reference method [21]. As the figure shows, the proposed algorithm obtains more robust and accurate results under a complicated traffic environment, which includes different types of vehicle, multiple vehicles in the same ROI and poor illumination conditions. The left and right boundary are located with regard to the detected symmetry axis. The bottom boundary is generated by the gray-level intensity information. Fig. 7 illustrates the rectangular boundary generation dealing with various types of vehicles. We choose $\sigma_1$ as 5 and $\sigma_2$ as 3 in the algorithm during our experiments. The active contour method was conducted in the generated rectangular boundary as shown in Fig. 8, where different vehicles have different shapes of contour according to the corresponding gray-level gradient information. Fig. 9 clearly describes the result in each step of the proposed vehicle detection algorithm.

After the active contour detection, the vehicle detection rate is 95.74% and the false alarm rate is 1.18%. The proposed

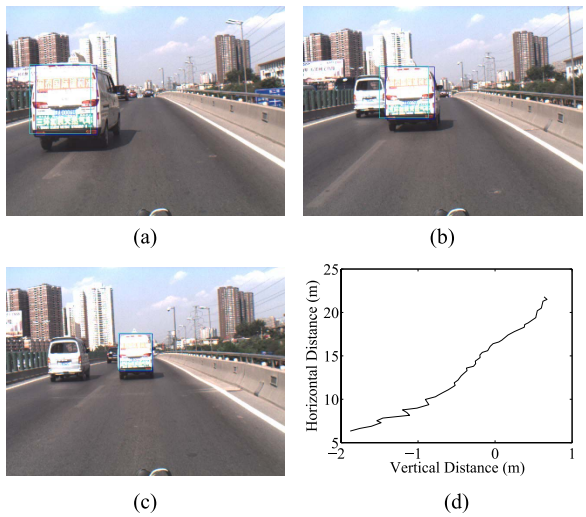


(a) (b) (c)

Fig. 9. One example of the detected results by the proposed vehicle detection algorithm. The result in each step of the algorithm is shown respectively. (a) rectangular boundary detection, (b) active contour detection, and (c) the tailored rear image.

(a) (b) (c)

Fig. 10. Challenging scenarios that the proposed approach fails.




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

Fig. 11. One 30 frames tracking example. A white minibus is tracked with three single frames of results shown as (a) the first, (b) the middle, and (c) the last frame. (d) Generated trajectory of the vehicle.

vehicle detection method fails at several challenging scenarios, as shown in Fig. 10. These scenarios generally contain the video frames with very poor illumination or contrast, hence are very challenging to computer vision algorithms. The vehicle detection rate and false alarm rate can be further optimized by the following vehicle tracking and verification stages.

### *B. Vehicle Tracking and Verification Results*

A locality sensitive histogram algorithm was operated to track the rear image of the vehicle in the image sequences. A white minibus was tracked for 30 frames as an demonstration of the tracking algorithm. The 1st, 15th, and 30th frames are shown in Fig. 11(a)–(c) respectively. The detected vehicle executed lane changing and forward-car overtaking maneuver. Fig. 11(d) shows the generated trajectory under the MMW radar's detection reference system after projective transformation, which properly reveal the maneuvers of the tracked vehicle.

Fig. 12 illustrates three consecutive frames of tracking results for four different vehicles. The rectangle in light blue represents the position of the detection in last frame, and the rectangle in dark blue represents the detection in the current frame. The selected vehicles with distinct colors and shape features (i.e., saloon, minibus, truck, and sedan) generally represent different types of on-road vehicles. The figure shows that the proposed algorithm can successfully track these vehicles under different traffic environments and light conditions. In addition, the changing of the traffic lanes in the sequential frames shows the moving of our intelligent vehicle.

The trajectories generated by the vision and radar processing modules are further compared to verify the detection. The comparison of trajectories under the same reference system of a valid vehicle detection and two false alarm detections are illustrated separately. Fig. 13(a) shows the two similar trajectories when a valid vehicle is detected. Fig. 13(b) describes the comparison of speeds calculated from these trajectories. Using the proposed criteria, the trajectory error between these two trajectories is 0.4750 m, which confirms the validation of the detected vehicle. A false alarm instance is shown in Fig. 14(a), where the black dot represents the radar return point that only exists in one frame, hence no trajectory can be established. Fig. 14(b) provides a false alarm tracking instance, where the trajectory error between MMW radar and vision processing is abnormally large. After tracking through vision processing, the trajectory error between these two trajectories is as high as 6.9725 m. Hence, this instance is treated as a false alarm.

### *C. Performance and Efficiency Analysis of the Overall System*

Three evaluation criteria, i.e., vehicle detection rate, false alarm rate, and time cost per frame, were used in our experiment with regard to the manually labeled ground truth. The experiment included 1,820 detected vehicles and 160 false alarm instances generated by the MMW radar. Table II shows the statistic results of the proposed fusion system in our experiment. The false alarm rate of MMW radar is 8.08%, which is obviously unacceptable in the practical design. On the contrary, the proposed fusion system successfully detects 1681 targeting vehicles and 160 false alarms; the vehicle detection rate is approximately 92.36%, and the false-alarm instances are all distinguished and eliminated. The reasons lead to 139 failed detection instances including challenging lighting conditions, tracking failed situations and crowded situations. All these advantages are critical to the vehicle detection module design deployed on the intelligent vehicle.

The proposed detection and tracking algorithm was then implemented with a multi-thread program framework with carefully optimized C++ codes. To reduce the time required by the vision processing module, the generated rectangular boundary was resized to no bigger than 50 $\times$ 50 for the active contour detection algorithm. The analysis of execution time of the system is shown in Table III. Fig. 15 shows the


Fig. 12. Examples of vehicle tracking results. Three consecutive frames for four different kinds of vehicles are illustrated: (a) a black saloon, (b) a white minibus, (c) a truck, and (d) a white sedan.

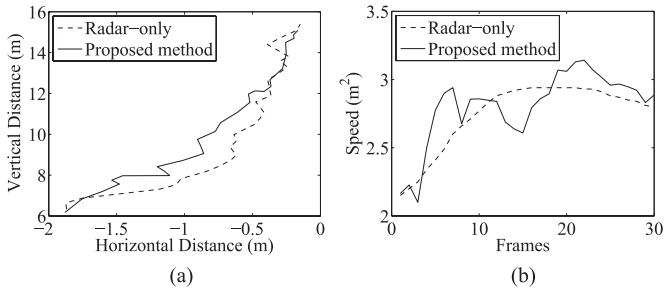



Fig. 13. An example for valid vehicle detection during trajectory comparison and verification. (a) The generated trajectories from MMW radar and vision processing, and (b) the corresponding speed trajectories.

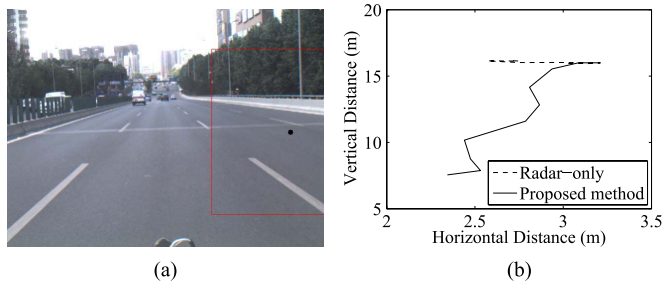



Fig. 14. Examples for false alarm detections during trajectory comparison and verification. (a) One false-alarm instance pointed by MMW radar. (b) The trajectories generated in the case of false-alarm.

TABLE II
EXPERIMENT RESULTS

| Parameter | Value |
|---|---|
| Ground truth of the number of vehicles | 1,820 |
| Detected vehicles by MMW radar | 1,980 |
| Vehicle detection rate of the proposed system | 92.36% |
| False-alarm rate the proposed system | 0% |

TABLE III
EXECUTION TIME OF EACH PROCESSING PART IN THE PROPOSED DETECTION AND TRACKING ALGORITHM

| Time cost (ms) | Max | Min | Average |
|---|---|---|---|
| Symmetrical detection | 6.324 | 0.6007 | 2.899 |
| Boundaries detection | 0.6961 | 0.0637 | 0.3402 |
| Active contour detection | 33.82 | 24.76 | 29.31 |
| One frame tracking | 9.182 | 3.457 | 7.014 |

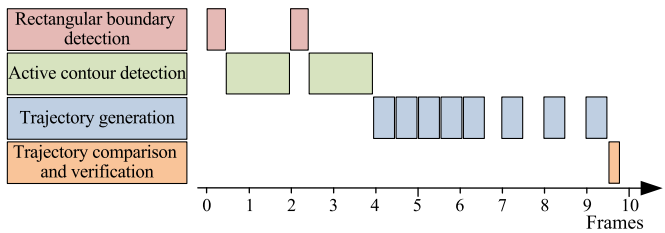



Fig. 15. The order and time consuming of each part of the proposed detection and tracking system during one detection procedure.

order and time cost of each processing part in the proposed vehicle detection and tracking algorithm during one detection procedure. In our experiment, the frame rate of camera is at 60 Hz, and each vehicle detection and tracking only takes 10 frames. Thus, the detection and tracking result can be outputted in 0.16 s.

We also evaluate the proposed vehicle detection method by comparing it with two reference methods, i.e., Alessandretti's method [19] and Kadow's method [20]. Alessandretti's method is based on symmetry features and Kadow's method is based on machine learning method. We implemented these two methods with properly optimized C++ programs. Three defined metrics, i.e., vehicle detection rate, false alarm rate, and time cost per frame, were used. Fig. 16 shows the comparison among

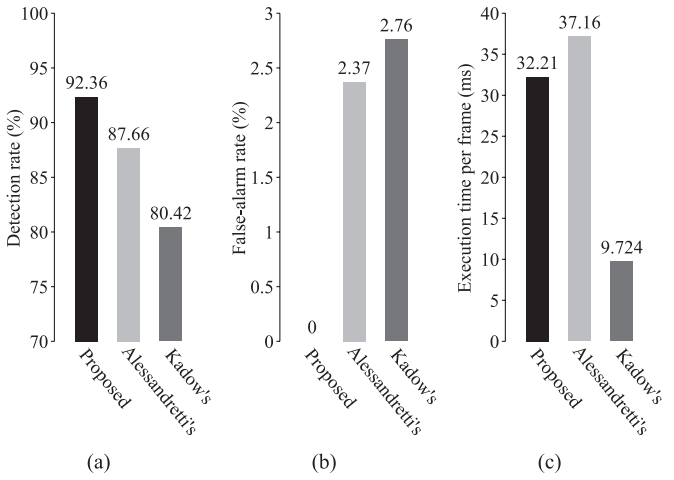



Fig. 16. Performance comparison among the proposed method, Alessandretti's method [19] and Kadow's method [20]. (a) Vehicle detection rate, (b) False-alarm rate and (c) execution time.

these methods using our real-world experimental dataset. Experimental results show that the proposed system can detect on-road vehicles with 92.36% detection rate and 0% false alarm rate. The time cost per frame of the proposed system is 32.21 ms. For Alessandretti's method, several of MMW radar's false alarm detections were verified. However, its overall performance is insufficient because of its sensitivity to unrelated details. Kadow's method provides the worst performance in our experiment, even though their time cost per frame is the shortest. Kadow's method incurs the highest false alarm rate due to the diversities and complexity in traffic scenes and traffic elements.

We note that the proposed MMW radar and mono-vision fusion system can detect and track several vehicles simultaneously by using multiple threads. Thus, each thread can detect and track one potential vehicle. In our practical hardware platform which uses 4-core, 8 thread Intel Core i7 CPU, we can achieve the detection and tracking of 3 nearest vehicles without any performance lost. As these threads are independent to each other, the detection and tracking of more vehicles can be achieved by using more powerful CPU.

## V. Conclusion

This paper aims at fusing MMW radar and monocular camera for on-road vehicle detection and tracking. The MMW radar firstly detects the potential vehicle and provides region of interest. The vision processing module employs symmetry detection and active contour detection to identify the vehicle inside the region of interest provided by MMW radar. The vehicle tracking is also employed for the two sensors. In addition, the two trajectories generated by MMW radar and monovision are further compared to verify and produce the detection and tracking result. The experimental results show that the proposed system can achieve a 92.36% detection rate and 0% false alarm rate under real-world dataset. Evaluation results clearly demonstrate that the proposed system achieves a better trade-off among detection rate, false alarm rate and real-time.

## References

[1] F. Qu, F.-Y. Wang, and L. Yang, "Intelligent transportation spaces: Vehicles, traffic, communications, and beyond," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 136–142, Nov. 2010.

[2] J. Hasch *et al.*, "Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 845–860, Mar. 2012.

[3] S. Tokoro, K. Kuroda, A. Kawakubo, K. Fujita, and H. Fujinami, "Electronically scanned millimeter-wave radar for pre-crash safety and adaptive cruise control system," in *Proc. IEEE Symp. Intell. Veh.*, 2003, pp. 304–309.

[4] A. Miranda Neto, A. Correa Victorino, I. Fantoni, and J. Ferreira, "Real-time estimation of drivable image area based on monocular vision," in *Proc. IEEE Symp. Intell. Veh.*, Jun. 2013, pp. 63–68.

[5] M. Ibarra Arenado, J. Perez Oria, C. Torre-Ferrero, and L. Renteria, "Monovision-based vehicle detection, distance and relative speed measurement in urban traffic," *IET Intell. Transp. Syst.*, vol. 8, no. 8, pp. 655–664, Dec. 2014.

[6] N. Bernini, M. Bertozzi, L. Castangia, M. Patander, and M. Sabbatelli, "Real-time obstacle detection using stereo vision for autonomous ground vehicles: A survey," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Oct. 2014, pp. 873–878.

[7] S. Sivaraman and M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[8] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle detection using millimeter-wave radar and its visualization on image sequence," in *Proc. 17th Int. Conf. Pattern Recog.*, Aug. 2004, vol. 3, pp. 342–345.

[9] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 694–711, May 2006.

[10] S. Sivaraman and M. Trivedi, "Real-time vehicle detection using parts at intersections," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Sep. 2012, pp. 1519–1524.

[11] C. Wang *et al.*, "Probabilistic inference for occluded and multiview on-road vehicle detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 215–229, Jan. 2015.

[12] R. O. Chavez-Garcia, J. Burlet, T.-D. Vu, and O. Aycard, "Frontal object perception using radar and mono-vision," in *Proc. IEEE Symp. Intell. Veh.*, Jun. 2012, pp. 159–164.

[13] T. Wang, J. Xin, and N. Zheng, "A method integrating human visual attention and consciousness of radar and vision fusion for autonomous vehicle navigation," in *Proc. IEEE Int. Conf. Space Mission Challenges Inf. Technol.*, 2011, pp. 192–197.

[14] T. Wang, N. Zheng, J. Xin, and Z. Ma, "Integrating millimeter wave radar with a monocular vision sensor for on-road obstacle detection applications," *Sensors*, vol. 11, no. 9, pp. 8992–9008, Sep. 2011.

[15] S. Wu, S. Decker, P. Chang, T. Camus, and J. Eledath, "Collision sensing by stereo vision and radar sensor fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 606–614, Dec. 2009.

[16] X. Liu, Z. Sun, and H. He, "On-road vehicle detection fusing radar and vision," in *Proc. IEEE Int. Conf. Veh. Electron. Safety*, 2011, pp. 150–154.

[17] R. O. Chavez-Garcia, T.-D. Vu, and O. Aycard, "Fusion at detection level for frontal object perception," in *Proc. IEEE Symp. Intell. Veh.*, 2014, pp. 1225–1230.

[18] T.-D. Vu, O. Aycard, and F. Tango, "Object perception for intelligent vehicle applications: A multi-sensor fusion approach," in *Proc. IEEE Symp. Intell. Veh.*, 2014, pp. 774–780.

[19] G. Alessandretti, A. Broggi, and P. Cerri, "Vehicle and guard rail detection using radar and vision data fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 95–105, Mar. 2007.

[20] U. Kadow, G. Schneider, and A. Vukotich, "Radar-vision based vehicle recognition with evolutionary optimized and boosted features," in *Proc. IEEE Symp. Intell. Veh.*, 2007, pp. 749–754.

[21] A. Broggi, P. Cerri, and P. C. Antonello, "Multi-resolution vehicle detection using artificial vision," in *Proc. IEEE Symp. Intell. Veh.*, 2004, pp. 310–314.

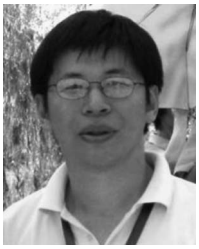
[22] C. Xu and J. L. Prince, "Gradient vector flow: A new external force for snakes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1997, pp. 66–71.

[23] A. W. Smeulders *et al.*, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

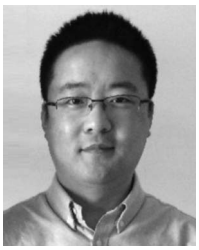
[24] S. He, Q. Yang, R. W. Lau, J. Wang, and M.-H. Yang, "Visual tracking via locality sensitive histograms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 2427–2434.

[25] S. C. Rutan, "Adaptive Kalman filtering," *Anal. Chem.*, vol. 63, no. 22, pp. 1103A–1109A, Nov. 1991.

**Xiao Wang** received the Bachelor's degree from Xi'an Jiaotong University, Xi'an, China, in 2012. He is currently working toward the Ph.D. degree in the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University. His research interests include intelligence system, pattern recognition, and vision algorithms and applications.



**Linhai Xu** received the Bachelor's degree from Shaanxi University of Technology, Xi'an, China, in 1993 and the M.S. degree from Xi'an Jiaotong University, Xi'an, in 2001. In 1993, he joined Xi'an Jiaotong University, where he is currently an Engineer in the Institute of Artificial Intelligence and Robotics. His research interests include intelligence system, robot vision, and industrial automation.



**Hongbin Sun** (M'11) received the B.S. and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2003 and 2009, respectively. From 2007 to 2008, he was a visiting Ph.D. student with the Electrical, Computer and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY, USA. From 2009 to 2011, he was a Postdoctoral Fellow with the Computer Science Department, Xi'an Jiaotong University, where he is currently an Associate Professor in the School of Electronic and Information Engineering. His research interests include memory hierarchy in computer system, VLSI architecture for video processing and computer vision, and signal processing system for new memory technology.

**Jingmin Xin** (S'92–M'96–SM'06) received the B.E. degree in information and control engineering from Xi'an Jiaotong University, Xi'an, China, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1993 and 1996, respectively.

From 1988 to 1990, he was with the Tenth Institute of Ministry of Posts and Telecommunications (MPT) of China, Xi'an. He was with the Communications Research Laboratory, MPT of Japan, as an Invited Research Fellow of the Telecommunications Advancement Organization of Japan from 1996 to 1997 and as a Postdoctoral Fellow of the Japan Science and Technology Corporation (JST) from 1997 to 1999. From 1999 to 2001, he was a Guest (Senior) Researcher with YRP Mobile Telecommunications Key Technology Research Laboratories Company, Ltd., Yokosuka, Japan. From 2002 to 2007, he was with Fujitsu Laboratories Ltd., Yokosuka. Since 2007, he has been a Professor at Xi'an Jiaotong University. His research interests are in the areas of adaptive filtering, statistical and array signal processing, system identification, and pattern recognition.

**Nanning Zheng** (SM'93–F'06) graduated from the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 1975, and received the M.S. degree in information and control engineering from Xi'an Jiaotong University in 1981 and the Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1985.

In 1975, he joined Xi'an Jiaotong University, where he is currently a Professor and the Director of the Institute of Artificial Intelligence and Robotics. His research interests include computer vision, pattern recognition and image processing, and hardware implementation of intelligent systems.

Dr. Zheng has been a member of the Chinese Academy of Engineering since 1999 and is currently the Chinese Representative on the Governing Board of the International Association for Pattern Recognition. He also serves as the President of the Chinese Association of Automation.