Journal of Southeast University (English Edition) Vol. 27, No. 2, pp. 169 – 173 June 2011 ISSN 1003—7985

Design of a road vehicle detection system based on monocular vision

Wang Hai Zhang Weigong Cai Yingfeng

(School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: In order to decrease vehicle crashes, a new rear view vehicle detection system based on monocular vision is designed. First, a small and flexible hardware platform based on a DM642 digtal signal processor (DSP) micro-controller is built. Then, a two-step vehicle detection algorithm is proposed. In the first step, a fast vehicle edge and symmetry fusion algorithm is used and a low threshold is set so that all the possible vehicles have a nearly 100% detection rate (TP) and the non-vehicles have a high false detection rate (FP), i. e., all the possible vehicles can be obtained. In the second step, a classifier using a probabilistic neural network (PNN) which is based on multiple scales and an orientation Gabor feature is trained to classify the possible vehicles and eliminate the false detected vehicles from the candidate vehicles generated in the first step. Experimental results demonstrate that the proposed system maintains a high detection rate and a low false detection rate under different road, weather and lighting conditions.

Key words: vehicle detection; monocular vision; edge and symmetry fusion; Gabor feature; PNN network

doi: 10. 3969/j. issn. 1003 – 7985. 2011. 02. 011

Every minute, on average, at least one person dies in a vehicle crash. Auto accidents also injure at least ten million people each year, and two or three millions of them seriously. With the aim of reducing injury and accident severity, precrash sensing is becoming an area of active research among automotive manufacturers, suppliers and universities. Vehicle accident statistics disclose that the main threats to a driver are from other vehicles, especially the vehicle in front. Consequently, an on-board automotive driver assistance system aiming to alert a driver about driving environments, and possible collisions with other vehicles has attracted much attention. To achieve this, robust and reliable vehicle detection is the first and most important step. Among the existing detection ways[1-5], vision is the most effective and popular way compared to others such as using laser and sonar. Because vision can obtain much more road environment information including vehicles, road lanes and pedestrians.

Till now, much research related to vision-based rear view vehicle detection has been done[6-11]. However, most of these works are still based on theoretical and algorithmic levels. In our work, a new rear view vehicle detection system based on monocular vision is achieved. A DM642 DSP is used as the micro-controller due to the fact that it is small in structure but maintains a high calculation ability for digital images. Then a two step vehicle detection algorithm is proposed. In the first step, a fast edge and symmetry fusion based method is used to obtain all the possible vehicles with nearly a 100% detection rate and a high false detection rate. In the second step, a Gabor feature probabilistic neural network (PNN)-based classifier is built to classify the possible vehicles from the candidate vehicles generated in the first step.

1 System Description

The general structure of the system is shown in Fig. 1. The system uses TMS320DM642 as the video signal processor which focuses on digital video and audio applications. The DM642 processor maintains a high calculation ability, whose peak calculation speed can reach 4 800 MIPS. Such high calculation speed guarantees that the processor satisfies real-time video digital signal processing. The CCD camera is installed near the shield window and at the middle of driver bridge (see Fig. 2). The video signal cannot be directly processed by the DM642, so a SAA7115 chip is used as the video decoder which supports many forms of video signals like PAL, NTSC and SECAM. It can automatically translate the input video signals and send them to the DM642 to execute vehicle detection algorithms. Similarly, the SAA7105 chip is used as an encoder to encode the processed signals and output them to the LCD screen. The DSP board is viewed in Fig. 3.

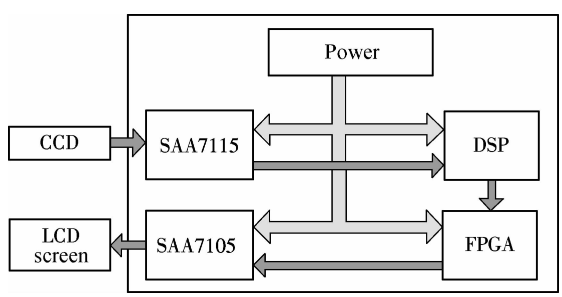

Fig. 1 General structure of system

Fig. 2 Camera installation

Received 2011-01-10.

Biographies: Wang Hai (1983—), male, graduate; Zhang Weigong (corresponding author), male, doctor, professor, zhangwg@seu. edu. cn.

Foundation items: The National Key Technology R& D Program of China during the 11th Five-Year Plan Period (2009BAG13A04), Jiangsu Transportation Science Research Program (No. 08X09), Program of Suzhou Science and Technology (No. SG201076).

Citation: Wang Hai, Zhang Weigong, Cai Yingfeng. Design of a road vehicle detection system based on monocular vision[J]. Journal of Southeast University (English Edition), 2011, 27(2): 169 – 173. [doi: 10. 3969/j. issn. 1003 – 7985. 2011. 02. 011]

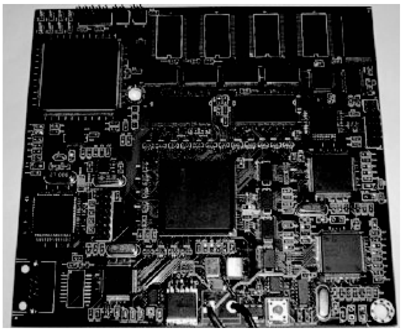

Fig. 3 DSP board

2 Vehicle Detection Algorithm

There are two kinds of vehicle detection methods. One uses prior knowledge of vehicles like horizontal/vertical edges, symmetry, color, shadow and texture. The other is the training-based method which learns the characteristics of the vehicle class from a set of training images. The training-based method captures the variabilities of vehicle appearance. First, each training image is represented by a set of local or global features. Then, the decision boundary between the vehicle and the non-vehicle classes is learned by a trained classifier. Generally, the prior knowledge-based method has a fast detection speed but a low detection rate. Contrarily, the training-based method has a low detection speed but a high detection rate. In the proposed system, a two-step detection algorithm is proposed, which combines both the prior knowledge-based and training-based methods. In the first step, an edge and symmetry fusion algorithm is used to obtain all the possible vehicles with a nearly 100% detection rate and a high false detection rate. In the second step, a classifier generated by the Gabor-feature-based PNN is trained to recognize the vehicles from the candidate vehicles generated in the first step.

2.1 Prior knowledge based candidate vehicle detection

2.1.1 Vertical edge extraction

Edge detection is used to locate the zone that changes rapidly in the image, which means that the first-order derivative of brightness is greater than a threshold. Because the vertical edge of a vehicle maintains the strongest symmetry feature, the proposed method just uses vertical edge information. The Sobel operator vertical mask is used to calculate the vertical gradient of the image.

$$G(x,y)=\left|f(x+1,y-1)+2f(x+1,y)+f(x+1,y+1)-f(x-1,y-1)-2f(x-1,y)-f(x-1,y+1)\right| \tag{1}$$

where $f(x, y)$ is the gray value of the pixel point (x, y) and $G(x, y)$ is the approximate gradient of this pixel point in a vertical direction. Then we use OTSU adaptive threshold segmentation theory to adaptively obtain a threshold so that pixels can be separated well with strong edge and weak edge information. Figs. 4 (a) and (b) show the original image and the binary edge image, respectively. The vertical gradients of all the pixel points in a vertical direction are summed.

$$B(i)=\sum_{j=1}^{H-1}G(i,j) \tag{2}$$

where H is the height the of detection area and B is the sum of the vertical gradient of this area.

(a) (b)

Fig. 4 Original and binary image. (a) Original; (b) Binary

2.1.2 Symmetry analysis

The gradient projection function $B(i)$ in Eq. (2) can be considered as a one-dimensional function $g(x)$. Assume that the coordinate of its symmetry axis is x_s and w is the width of the symmetry zone.

Let $u=x-x_s$, then $g(u)$ is a function whose coordinate origin is x_s. For any function $f(u)$, it can be expressed as the sum of an odd function and an even function. Then we can define the odd function and even function components of function $g(u)$ and $g(x_s+u)$ as

$$\left.\begin{aligned} O(u,x_s)&=\frac{g(x_s+u)-g(x_s-u)}{2}\\ E(u,x_s)&=\frac{g(x_s+u)+g(x_s-u)}{2}\\ &-w/2\leqslant u\leqslant w/2\end{aligned}\right\} \tag{3}$$

Then the even function is normalized as

$$E_n(u,x_s)=E(u,x_s)-\frac{1}{w}\int_{-w/2}^{+w/2}E(u,x_s)\,\mathrm{d}u \tag{4}$$

So the mean value of the even function is 0.

The symmetry value can be evaluated by

$$S(x_s)=\frac{\int_{-w/2}^{+w/2}E(u,x_s)^2\mathrm{d}u-\int_{-w/2}^{+w/2}O(u,x_s)^2\mathrm{d}u}{\int_{-w/2}^{+w/2}E(u,x_s)^2\mathrm{d}u+\int_{-w/2}^{+w/2}O(u,x_s)^2\mathrm{d}u} \tag{5}$$

The value of $S(x_s)$ is between -1 and 1. When $S=1$, it means 100% symmetrical. On the contrary, $S=-1$ means no asymmetric.

On the left and the right sides of the vehicle, the vertical edge reaches its maximum value. Based on this, a big area where a vehicle may exist can be generated. In this big searching area, we begin to search the symmetry axis by changing the value of x_s. When $S(x_s)>0.3$, a vehicle symmetry axis is considered to be found. Then a small area with a high vehicle existing probability is verified.

2.2 Training based vehicle detection

2.2.1 Design of Gabor filter

Motivated by biological findings on the similarity of the two-dimensional (2-D) Gabor filters and receptive fields of neurons in the visual cortex, there has been increasing interest in deploying Gabor filters in various computer vision applications. One of the most important properties is that they have optimal joint localization in both spatial and frequency domains. The general functionality of the 2-D Gabor filter family can be represented as a Gaussian function modulated by a complex sinusoidal signal. Specifically, a 2-D Gabor filter $G(x, y)$ can be formulated as

$$G(x, y) = \frac{1}{2\pi\gamma\sigma^2}\exp\left[-\frac{(\bar{x}/r)^2 + (\bar{y})^2}{2\sigma^2}\right]\exp(2\pi j f \bar{x}) \quad (6)$$

where

$$\bar{x} = x\cos(\theta) + y\sin(\theta),\ \bar{y} = -x\sin(\theta) + y\cos(\theta) \quad (7)$$

The Gabor filter has a good ability to express the direction feature of an image which is demonstrated by

$$\theta_k = \frac{k\pi}{n} \qquad k = 0, 1, \dots, n-1 \quad (8)$$

where n is the number of all the directions, and θ_k is the k-th direction.

In vehicle imaging, there are many vertical and horizontal edges like doors, windows and bumpers. So, $\theta = 0$ and $\theta = 2\pi$ are chosen to extract Gabor features of vertical and horizontal edges. Through observation, it is found that many vehicles like sedans also contain some 45° edges, then $\theta = \pi/4$ and $\theta = 3\pi/4$ are selected. Besides, the setting of the direction parameter θ should fully and evenly cover the whole image. Based on the above analysis, the number n is equal to 8. f is the space frequency of the cosine function in Eq.(6) which determines the frequency selection feature of the Gabor filter. f is related to frequency bandwidth B and deviation σ of the Gaussian function. It is proved that when $B = 1$, the Gabor filter is the closest to the human vision receiving system, so we set $B = 1$. There is a relationship among σ, f and B: $\sigma F = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^B+1}{2^B-1}$, so $\sigma f = 0.56$, and f decreases progressively with a scale factor $1/\sqrt{2}$. The function of Gabor filter groups is

$$f_k = \sqrt{2}^{-k} f_{\max} \qquad k = 0, 1, \dots, m-1 \quad (9)$$

where m is the number of frequencies; f_k is the k-th frequency, $f_{\max} = 0.25$. Here, we choose $m = 5$. Since $\sigma f = 0.56$, when f is verified, σ is obtained.

The spatial functioning scope of the Gabor filter is an ellipse, and the ellipticity is expressed by parameters. Its value range is $(0, 1)$. When $\gamma = 1$, the functioning scope is a circle. In most cases, the shape of a vehicle in the image is close to a rectangle and the length to width ratio is around 0.6. Besides, the receiving field of the Gabor filter is close to that of human beings as the best receiving field of the human is 0.3 to 0.7. Considering these two factors together, we choose $\gamma = 0.55$. Till now, a 5×8 filter group is established (see Fig. 5).

The training image samples are all resized to 24×24. The dimensions of the feature vector of each image is $24 \times 24 \times 40 = 23\,040$, which is too large for training. So we sample it to 2 560 as the feature vector of each image.

2.2.2 Design of PNN network

The PNN network developed by Donald Specht is a model-based competitive learning method with a concept of "winner takes all attitudes" and the core concept is based on multivariate probability. The PNN provides a general solution to pattern classification problems by an approach developed in statistics, called Bayesian classifiers.

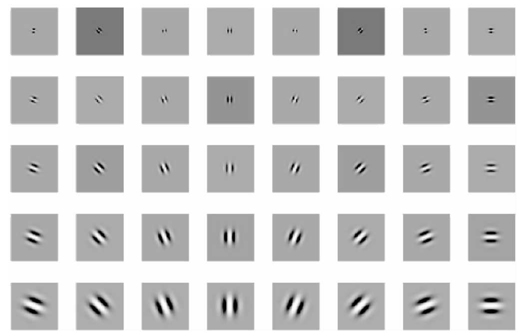

Fig. 5 Gabor features

Former research has demonstrated many good features of the PNN: 1) It is easy to train compared with other neural networks like RBFNN and has a high convergence rate; 2) It can achieve any nonlinear transform and the decision surface is close to that of the Bayesian optimal rule; 3) It has very strong fault-tolerance capability; 4) The transfer functions in the pattern layer can choose many kernel functions and the classification results are not sensitive to it.

As shown in Fig. 6, in the radial basis layer, vector $\boldsymbol{b}$ is combined with $\|\boldsymbol{W} \cdot \boldsymbol{p}\|$. The result is denoted as $\boldsymbol{n} = \|\boldsymbol{W} \cdot \boldsymbol{p}\| \times \boldsymbol{b}$. The transfer function in the PNN is built with a distance criterion with respect to a center. In this paper, we define it as

$$\text{radbas}(\boldsymbol{n}) = \exp(-\boldsymbol{n}^2) \quad (10)$$

Each element of $\boldsymbol{n}$ is substituted into Eq. (10) and produces corresponding element of $\boldsymbol{a}$, the output vector of the radial basis layer. We can represent the i-th element of $\boldsymbol{n}$ as

$$a_i = \text{radbas}(\|\boldsymbol{W}_i \cdot \boldsymbol{p}\| \times \boldsymbol{b}_i) \quad (11)$$

where $\boldsymbol{W}_i$ is the i-th row of $\boldsymbol{W}$ and $\boldsymbol{b}_i$ is the i-th element of bias vector $\boldsymbol{b}$. There is no bias in the competitive layer. In this layer, vector $\boldsymbol{a}$ is first multiplied by layer weight matrix $\boldsymbol{M}$, producing an output vector $\boldsymbol{d}$. The largest element of $\boldsymbol{d}$ is d_j, and the competitive function C set a_j of $\boldsymbol{a}$ as 1 while other elements of $\boldsymbol{a}$ as -1. $\boldsymbol{M}$ is set to $K \times Q$ matrix. If the i-th sample in the training belongs to class j, then the value of the j-th row and the i-th column of $\boldsymbol{M}$ is set as 1.

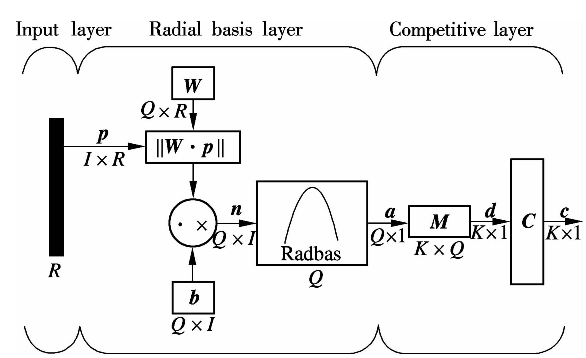

Fig. 6 PNN structure

The training set contains 1 000 vehicle and 5 000 non-vehicle sub images which are extracted manually from pictures taken in different places under different lighting conditions. Each sub image in the training set and the test set is scaled to 24×24 and preprocessed to account for different lighting conditions. And also 500 test images are prepared. The training accuracy is 98.05% and the test accuracy is 96.2%, which is better than those using the Haar wavelet feature and Adaboost[12] whose accuracy is 92% and the Gabor feature and SVM whose accuracy is 94.81%[13].

3 Experimental Results

The experiments are conducted on many freeways. In the first step, the candidate cars are selected with symmetry constraints (see Fig. 7). In this step, a low symmetry threshold is set to make sure that all the vehicles can be detected but with a high false detection rate. In the next step, all the candidate cars are classified by the trained Gabor based PNN network. In this step, most of the false detected cars in the first step are rejected with real vehicles being recognized (see Fig. 8). In these experiments, we can see that the proposed system can detect different vehicles (sedan, bus and truck) not only under good lighting conditions but also under bad lighting conditions (see Fig. 7(b) and Fig. 8 (b)). The average calculation time of each image is around 80 ms, which can satisfy real time application (see Tab. 1).

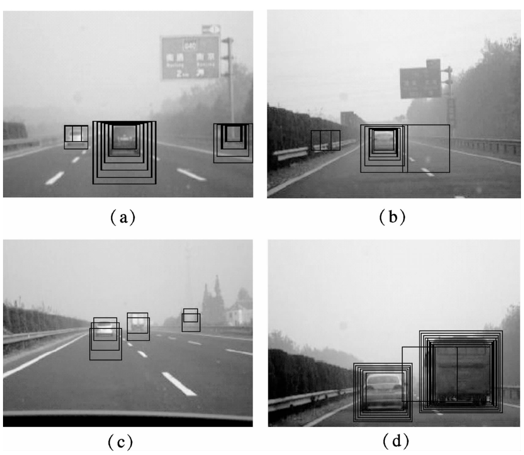

Fig. 7 Candidate vehicles in continuous video. (a) Frame 245; (b) Frame 384; (c) Frame 659; (d) Frame 836

Fig. 8 Detected vehicles in continuous video. (a) Frame 245; (b) Frame 384; (c) Frame 659; (d) Frame 836

Tab. 1 Experimental results

Weather condition	Total vehicles	Error rate/%	Average time cost/ms
Good	173	2.89	73.82
Bad	145	4.14	83.56

4 Conclusion

In this paper, a new vehicle detection system based on monocular vision is achieved. A small and flexible hardware platform based on the DM642 DSP micro-controller is built. A two-step vehicle detection algorithm is used in this system. In the first step, a vertical edge and symmetry fusion algorithm is used to obtain all the possible vehicles with a nearly 100% detection rate and a high false detection rate. In the second step, a Gabor feature based PNN is built to classify the possible vehicles from the candidate vehicles generated in the first step. The experimental results demonstrate that the system can work well under different road, weather and lighting conditions and can meet real-time requirements of application.

References

[1] Kim S, Oh S Y, Kang J, et al. Front and rear vehicle detection and tracking in the day and night times using vision and sonar sensor fusion[C]//2005 *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Alberta, Canada, 2005: 2173 – 2178.

[2] Giancarlo A, Alberto B, Pietro C. Vehicle guard rail detection using radar and vision data fusion[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2007, **8**(1): 95 – 105.

[3] Fang J, Meng H, Zhang H, et al. A low-cost vehicle detection and classification system based on unmodulated continuous-wave radar[C]//*IEEE International Conference on Intelligent Transportation Systems Conference*. Seattle, USA, 2007: 715 – 720.

[4] Acunzo D, Zhu Y, Xie B, et al. Context-adaptive approach for vehicle detection under varying lighting conditions[C]// *IEEE International Conference on Intelligent Transportation Systems Conference*. Seattle, USA, 2007: 654 – 660.

[5] Mahlisch M, Schweiger R, Ritter W, et al. Sensor fusion using spatio-temporal aligned video and lidar for improved vehicle detection[C]//*IEEE Intelligent Vehicles Symposium*. Tokyo, Japan, 2006: 424 – 429.

[6] Sun Z, Bebis G, Miller R. Monocular precrash vehicle detection: features and classifiers[J]. *IEEE Transactions on Image Processing,* 2006, **15**(7): 2019 – 2034.

[7] Iwasaki Y, Kurogi Y. Real-time robust vehicle detection through the same algorithm both day and night[C]//*International Conference on Wavelet Analysis and Pattern Recognition*. Beijing, China, 2007: 1008 – 1014.

[8] Cheng H, Zheng N, Sun C. Boosted Gabor features Applied to vehicle detection[C]//18*th International Conference on Pattern Recognition*. Hong Kong, China, 2006: 662 – 666.

[9] Lan J, Zhang M. A new vehicle detection algorithm for real-time image processing system[C]//2010 *International Conference on Computer Application and System Modeling*. Taiyuan, China, 2010: 1 – 4.

[10] Luo W T, Jun W H, Kuo C F. Vehicle detection using normalized color and edge map[J]. *IEEE Transactions on Image Processing*, 2007, **16**(3): 850 – 864.

[11] Wen X, Zhao H, Wang N, et al. A rear-vehicle detection system for static images based on monocular vision[C]// 9*th International Conference on Control, Automation, Ro-*

botics and Vision. Singapore, 2006: 1－4.

[12] Schneiderman H, Kanade T. A statistical method for 3D object detection applied to faces and cars[C]//*IEEE International Conference on Computer Vision and Pattern Recognition*. Hilton Head Island, SC, USA, 2000: 746－751.

[13] Sun Z, Bebis G, Miller R. On-road vehicle detection using gabor filters and support vector machines[C]//14*th International Conference on Digital Signal Processing*. Reno, USA, 2002: 1019－1022.

基于单目视觉的道路车辆检测系统设计

王　海　　张为公　　蔡英凤

（东南大学仪器科学与工程学院，南京 210096）

摘要：为了减少车辆碰撞，设计了一个新的基于单目视觉的道路前方车辆检测系统. 首先，建立一个基于 DM642 数字信号处理器的小型灵活硬件平台. 然后，提出了一个 2 步车辆检测算法. 第 1 步，采用车辆边缘和对称性融合算法，设定较低阈值检测出所有可能车辆，该步骤对车辆有接近 100% 的检测率（TP）和对非车辆有较高的误检率（FP），即能获得图像中所有可能车辆；第 2 步，采用多尺度多方向 Gabor 特征对车辆样本进行表征，并用概率神经网络对大量车辆和非车辆样本进行训练，从第 1 步获得的可能车辆中识别出正确的车辆. 实验表明：所设计的系统在不同道路、天气和光照条件下都具有很高的检测率和较低的误检率.

关键词：车辆检测；单目视觉；边缘和对称性融合；Gabor 特征；PNN 神经网络

中图分类号：TP391.4