Car Detection Based on Multi-Cues Integration

Zhenfeng Zhu[1], Hanqing Lu[1], James Hu[2], Keiichi Uchimura[2]
[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China
[2]Computer Science Dept., Kumamoto University, Japan
{zfzhu, luhq}@nlpr.ia.ac.cn ; {hu, uchimura}@cs.kumamoto-u.ac.jp

Abstract

In this paper we present a novel fast multi-cues based car detection technique in still outdoor images. On the bottom level, two novel area templates based on edge cue and interest points cue are first designed, which can rapidly reject most of the non-car sub-windows at the cost of missing few of the car sub-windows. On the top level, both global structure cue and local texture cue are considered. To character the global structure property the odd Gabor moments are introduced and trained by SVMs. The multi channels even Gabor based local texture property extracted from corner area is modeled as a Gaussian distribution. The final experiment results show that the integration of global structure property and local texture property is more powerful in discrimination between car and non-car objects and a high detection accurate 93% is obtained.

1. Introduction

Object detection in two dimension images has been a deeply investigated field in vision community. Performance of object detection depends on the properties of extracted features and the classifier used. Here we focus our attention on car detection in still outdoor images.

A statistical method [1] for vehicle detection based on local features was investigated, in which the EM algorithm is adopted to learn the parameters of the probability distribution of the constellation of local features. In [2], a different sparse parts-based car detection system is proposed, in which the SNoW architecture is learned based on the parts automatically extracted from some Forstner interest points and the relations between parts. To the extension of this work, A.Garg et al. [3] proposed a method by fusing global (ICA-based) and local (parts-based) information, in which high object detection accurate was obtained. But both them pay no more attention to the computing expenses.

In [4] Schneiderman et al. successively proposed a wavelet coefficient histogram based statistic method for multi views face and car detection. But high computing expenses are its main problem. Papageorgiou et al. [5] proposed a general object detection scheme using Harr wavelet and *SVM*s and applied it to face, pedestrian and car detection. Viola et al. [6] proposed a more popularly referred method for face detection, in which the adaboost classifier was trained on a series of Harr wavelet coefficients and real time is achieved by using integral image and cascade technique.

Betke et al. [7] used motion and edge information to hypothesize the vehicle locations and template matching for detection. Sun et al. [8] provided an on-road vehicle (rear views) detection using Gabor filter and support vector machines. But it relies on some other vehicle location method and only can be taken as a verification step.

In this paper we present a novel fast multi cues based car detection system in still outdoor images. Both global structure cue and local texture cue are introduced in our work. To character the global structure property we propose an odd Gabor moments obtained by the Angular Radial Transform (***ART***) on odd Gabor responses. In addition, the multi channels even Gabor based local texture is modeled as a gaussian distribution. To rapidly reject large parts of background with true car object parts kept, two novel low level cues based templates, Edge Area Template (***EAT***) and Corner Area Template (***CAT***), are designed. The experiment results show that the proposed scheme can obtain a high detection accurate and low computational costs compared to [2] [3].

2. Edge area and corner area templates

In most cases of object detection, there exist a huge of non-objects that can be rapidly eliminated only based on some low level features without need to introduce higher-level mechanism. In the first phrase of our object detection, both edge area template and corner area template are constructed to perform such task.

2.1 Edge Area and corner area templates

The original idea of constructing area templates came from object matching based on hausdorff distance [9]. To extract the edge and interest point maps canny edge detector and Harris corner detector [10] are used. To consider the position deviation, the corner map was dilated with 5×5 structure and the edge map was dilated with 3×3 structure. Thus we define the edge area template (***EAT***) T_e and corner area template (***CAT***) T_c as

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

$$T_c(x,y)=\begin{cases}1 & \frac{1}{N}\sum_{n=1}^{N}C_n^d(x,y)\geq 0.25\\ 0 & \frac{1}{N}\sum_{n=1}^{N}C_n^d(x,y)<0.25\end{cases},\ T_e(x,y)=\begin{cases}1 & \frac{1}{N}\sum_{n=1}^{N}E_n^d(x,y)\geq 0.7\\ 0 & \frac{1}{N}\sum_{n=1}^{N}E_n^d(x,y)<0.7\end{cases} \quad (1)$$

where N is the number of positive examples, $C_n^d(x,y)$ denotes the dilated corner map, and $E_n^d(x,y)$ the dilated edge map. Fig.1 shows the constructed corner area template and edge area template. To consider the influence from background, the corner map is filtered by an ellipse mask.

(a) Corner Area Template (b) Edge Area Template

Figure 1 The constructed Corner Area Template (***CAT***) and Edge Area Template (***EAT***)

2.2 Discrimination based on edge area and corner area templates

Given example $\boldsymbol{I}(x,y)$ to be discriminated, let n_1^c denotes the number of corners contained in corner area, n_2^c the number of whole corners, n_1^e the number of edge pixels contained in edge area, and n_2^e the number of edge pixels contained in non-edge area. Four qualifications are defined as:

$$\boldsymbol{Q}_1^c(I)=\begin{cases}1 & n_1^c>5\\ 0 & else\end{cases},\ \boldsymbol{Q}_2^c(I)=\begin{cases}1 & \frac{n_1^c}{n_2^c}<0.3\\ 0 & else\end{cases}$$

$$\boldsymbol{Q}_1^e(I)=\begin{cases}1 & 100<n_1^e<170\\ 0 & else\end{cases},\boldsymbol{Q}_2^e(I)=\begin{cases}1 & 1<\frac{n_2^e}{n_1^e}<2\\ 0 & else\end{cases}$$

Thus the final qualification $\boldsymbol{Q}$ of $\boldsymbol{I}(x,y)$ can be obtained

$$\boldsymbol{Q}(I)=\boldsymbol{Q}_1^c(I)\cap\boldsymbol{Q}_1^c(I)\cap\boldsymbol{Q}_1^e(I)\cap\boldsymbol{Q}_2^e(I) \quad (2)$$

When $\boldsymbol{Q}(I)$ is true it will be passed to the next phase, otherwise it will be rejected.

3. Classification based on odd Gabor moments by using SVMs

How to represent an object is a key step in object detection task. There exist many different ways to represent an object. To obtain an effective representation in our case, the structure property of an object is extracted by using odd Gabor filter moments.

Gabor filters, which have been shown to fit well the receptive fields of the majority of simple cell in the primary visual cortex [11], are modulation products of Gaussian and complex sinusoidal signals. A 2D Gabor filter oriented at angle θ is given by [12]:

$$\boldsymbol{G}_j(x,y)=k_j^2\exp(-\frac{k_j^2x^2}{2\sigma^2})\exp(i\vec{k}_j\overline{x}) \quad (3)$$

$$\vec{k}_j=\begin{pmatrix}k_{jx}\\ k_{jy}\end{pmatrix}=\begin{pmatrix}k_v\cos(\theta_u)\\ k_v\sin(\theta_u)\end{pmatrix},\ k_v=2^{-\frac{v+2}{2}}\pi$$

where σ is the standard deviation of the circle gaussian along x and y, and $\vec{k}_j$ denotes the spatial frequency. The even real part and odd imaginary part of 2D Gabor filter is shown in Fig.2.

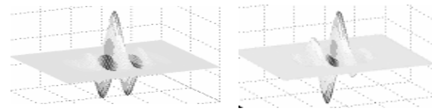

Figure 2 (a) Even real part and (b) Odd imaginary part of 2D Gabor filter

For a given input image $I(x,y)$, the Gabor filter response is denoted by

$$\boldsymbol{R}_j(x,y)=\boldsymbol{G}_j(x,y)*\boldsymbol{I}(x,y) \quad (4)$$

where * denotes two-dimensional convolution operator. Here we denote R^e as even Gabor response, which will be adopted to construct probable model in next section, and R^o as odd Gabor response. Thus the final energy map for odd Gabor response, which can effectively reflect the structure property [13], is defined as

$$\boldsymbol{E}(x,y)=\left\|\boldsymbol{R}_j^o(x,y)\right\|_2 \quad (5)$$

where $\|\cdot\|_2$ denotes Euclidean norm. Note that in odd Gabor case only two orientations and one scale are adopted, i.e. $\theta_u=\frac{u\pi}{2}, j=u+v, u=0,1, v=0$.

As we can see that the dimension of $E(x,y)$ keeps the same as the $m\times n$ input image $I(x,y)$ after the above Gabor transform. To obtain a low dimension representation, a further Angular Radial Transform (ATR)[14] is adopted and we named F_{lp}, the Angular Radial Transform on the above Energy map $\boldsymbol{E}(x,y)$, as odd Gabor moment. We have

$$F_{lp}=\sum_{x=-n/2}^{x=n/2}\sum_{y=-m/2}^{m/2}\boldsymbol{V}_{lp}(\rho,\theta)\boldsymbol{E}(x+n/2,y+m/2) \quad (6)$$

$$\boldsymbol{V}_{lp}(\rho,\theta)=\frac{1}{2\pi}e^{jp\theta}R_l(\rho),\ R_l(\rho)=\begin{cases}1 & l=0\\ 2\cos(l\pi\rho) & l\neq 0\end{cases}$$

where $\rho=\sqrt{x^2+y^2}$ and $\theta=\arctan(y/x)$

Note that only real part of $\boldsymbol{V}_{lp}$ are adopted in our case and $l=(0\sim 2), p=(0\sim 5)$. Thus we obtain an odd Gabor moment vector $\boldsymbol{F}=\{F_{lp}\}$

As an eminent binary classifier, the nature of SVM is to find a discriminate hyper-plane that optimally separates two classes of objects by using structural risk minimization. The final discriminate hyper-plane can be defined as:

$$\boldsymbol{S}(x)=\theta(\sum_{i=1}^{n}y_i\alpha_ik(x,x_i)+b),\ \theta(z)=\begin{cases}+1 & if\quad z>\tau_1\\ -1 & if\quad z<\tau_1\end{cases} \quad (7)$$

where $k(x,x_i)$ is a kennel function, $x_i\in R^D, y_i\in\{-1,+1\}$

is the class label of the example. Any example x_i corresponding to nonzero α_i is a support vector. Thus we can give the label for $\boldsymbol{F}$ by the sign of $\boldsymbol{S}(F)$.

4. Modeling multi channels Gabor responses in corner area

The odd Gabor moments introduced in last section effectively reflect the structure property, but to have an enhanced object representation some other properties, e.g. local texture property, should be integrated. Since Gabor filter shows a good performance in charactering texture property, the multi channels even Gabor filters with Dc-free were adopted in our case.

Let R_j^e be a jet of multi channels even Gabor responses with $\theta_u = \frac{u\pi}{2} + \frac{\pi}{4}, j = u + 2v, u = 0,1, v = 0,1,2$. Here we only consider the local texture extracted from the corner area, then a local texture vector $\boldsymbol{f} = \{f_j\}$ can be obtained with $f_j = R_j^e \bullet T_c$, where $\bullet$ denotes dot product. For simplicity, here we assume that $\{f_j\}$ has a gaussian distribution with mean μ_f and covariance Σ and we have

$$\boldsymbol{P}(f) = (2\pi)^{-\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} \exp[-\frac{(f-\mu_f)\Sigma^{-1}(f-\mu_f)^T}{2}] \quad (8)$$

Then a threshold τ_2 can be set on $\boldsymbol{P}(f)$ for discrimination. The multi Gaussian distribution has also been performed but gained no improvement for our final experimental results.

5. Experimental results

The proposed scheme is evaluated on the car image database downloaded from [15]. An additional 500 negative examples collected from website are added to the database. Thus the whole training database contains 550 positive and 1000 negative images, and 170 test images containing 200 cars in all.

Note that all images in the database are natural. They are taken from different sources and include occlusion and cluttered backgrounds. In addition, the size of training images is clipped to 30×90 manually from 40×100 to eliminate the influence from background and all positive examples are adjusted to the same direction. Fig.3 gives some positive and negative examples.

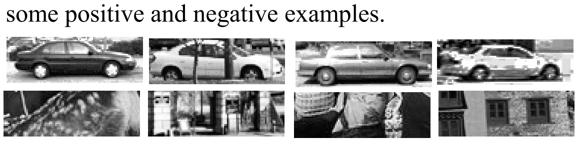

Figure 3 Several positive training examples (top) and negative ones (bottom)

The sequential multi cues based car detection architecture is shown in Fig.4.The active map is referred to the output of $S(\cdot)$ function. For space consideration, the detailed post processing (evaluation scheme on active map) procedure can be found in [2]. The performance curves of our car detection system on test set containing 200 cars are shown in Fig. 5

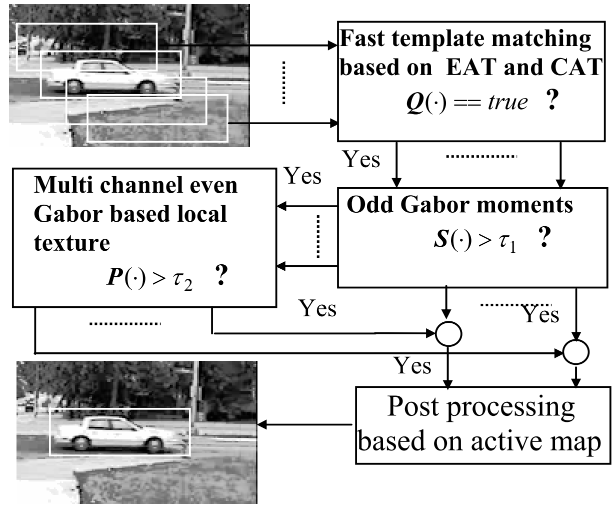

Figure 4 Sequential multi cues based car detection architecture

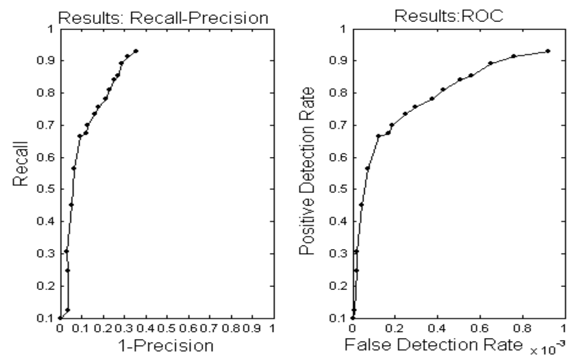

Figure 5 Performance of our car detection system on test set containing 200 cars. **Left:** Recall–precision curve showing performance of our car detection system. **Right**: ROC curve showing the same results. The definitions of different parameter and the discussions about Recall -precision and ROC can be found in [2]. The different points on curve correspond to different thresholds τ_1

The comparison of accuracy of our car detection system with S.Agarwal's [2] parts-based method is shown in Table 1, in which we can see that the performance of our detection system is some superior to S.Agarwal's. Note that to have a fair comparison, the No. of correct detection is fixed for both cases. For space consideration only some examples of correct detection are shown in Fig.6.

Note that all the test images are performed on ***Matlab* 6.1** platform with P4 1.7G Hz PC and the average computational cost is 3 seconds for 200×150 image without code optimization, whereas 8 seconds for S.Agarwal's [2] method and more for A. Garg's [3] method on ***VC++*** platform. We think our approach will be definitely faster on ***VC++*** platform.

Table 1 Comparison of accuracy of our car detection system with S.Agarwal's [2] parts based method on the test set containing 200 cars

No. of correct detection, N	Recall $\frac{N}{200}$	No. of false detections, M		Precision $\frac{N}{N+M}$		False detection rate $\frac{M}{112000}$	
181	90.5%	98	77	64.9%	70.2%	0.09%	0.06875%
178	89%	92	73	65.9%	70.9%	0.08%	0.06518%
171	85.5%	76	61	69.2%	73.7%	0.07%	0.05446%
162	81%	48	48	77.1%	77.1%	0.04%	0.04286%
154	77%	36	33	81.1%	82.4%	0.03%	0.02946%
140	70%	29	21	82.8%	87.0%	0.03%	0.01875%

*The table is adopted from [2], and the boxed parts are our results. The denominator 112000 showed in last column is the No. of negative sub-windows among 147802 sub-windows evaluated from 170 test images. The data for S.Agarwal's parts based method come from [2]

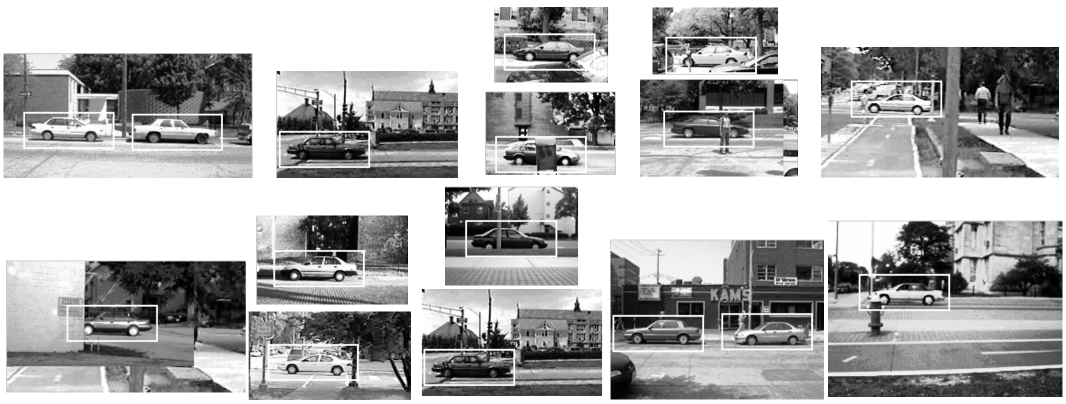

Figure 6 Examples of correct detection results corresponding to $\tau_1 = -0.25$

6. Conclusions

In this paper we present a fast multi cues based car detection system with high detection rate. The low computational cost contributes to the utilizing of the two first proposed area templates ***EAT*** and ***CAT.*** Both global structure cue based on odd Gabor moments and local texture cue based on multi channels even Gabor responses distribution are considered, which makes a final high performance car detection system. For our future work we will focus our researches on multi views and multi scales car detection by extending the approach presented in this paper.

Acknowledge

This work is funded by the NSFC under grant No. 60135020 and 60121302. We are grateful to S. Agarwal et al. for sharing their car dataset.

Reference

[1] M.Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition", In *Proc. ECCV*, 2000, pp. 18–32.

[2] S.Agarwal and D. Roth," Learning a sparse representation for object detection", In *Proc. ECCV*, 2002, pp. 113–130.

[3] A. Garg, S. Agarwal and Thomas S. Huang, " Fusion of local and global information for Object detection", *Proc. IEEE ICPR*, Quebec City, QC, Canada, August 2002.

[4] H. Schneiderman and T. Kanade, " A statistical method for 3d object detection applied to faces and cars", In *Proc. CVPR*, 2000.

[5] C. Papageorgiou and T. Poggio. A trainable system for object detection. *IJCV,* vol. 38, no. 1, pp. 15-33, 2000.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", In *Proc. CVPR*, 2001. 1988.

[7] M. Betke, E. Haritaglu and L. Davis. Multiple vehicle detection and tracking in hard real time. *IEEE Intelligent Vehicles Symposium*, 1996, pp. 351-356.

[8] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using Gabor filters and Support Vector Machines", *IEEE 14th International Conference on Digital Signal Processing*. Greece, July 2002.

[9] D.P. Huttenlocher, G.A. Klanderman, and W.J.Rucklidge, "Comparing images using the Hausdorff distance", *IEEE Trans. On PAMI,* Vol. 15(9), 1993, pp. 850-863.

[10] C. J. Harris and M. Stephens, "A combined corner and edge detector". *In Proc. of Alvey Vision Conference*, pp. 147–151.

[11] J.P. Jones and L.A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex", *Journal of Neurophysiology, 58*, 1987, pp.1233-1258.

[12] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, R. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture", *IEEE Transactions on Computers*, Vol.42(3), 1993, pp.300–311.

[13] R.Mehrotra, K.R.Namuduri, and N.Ranganathan, "odd Gabor filter-based edge detection", *Pattern Recognition, 25(12)*, 1992, pp.1479-1494.

[14] J. Fang, G. Qiu, "Face detection based on multiple regression and recognition Support Vector Machines", *BMVC*, Norwich, September 2003.

[15] http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/