The Study on Vehicle Detection Based on DPM in Traffic Scenes

Chun Pan, Mingxia Sun and Zhiguo Yan

Abstract After the HoG feature was proposed, a lot of detectors were developed based on the feature. But HoG feature has its defects, as high dimensional data leading to inefficiency, complex scenes leading to poor performances and so on. In this article, we proposed a vehicle detector based on DPM (Deformable Part Model). This detector uses a deformable part model to classify the front and the rear of the vehicles.

Keywords Vehicle detection · DPM · Traffic scenes

1 Introduction

Object detection is one of the most popular researching fields in computer vision, such as vehicle detection, pedestrian detection and so on. Normally, the common detecting solutions are using HoG, Sift or Haar to extract features and using SVM or Adaboost as classifiers. In this article, we propose a solution by using DPM to detect Vehicles. In consideration of the variety of appearance of the vehicles are affected by many factors, as the changes of illuminations or angle of view. The traditional detecting algorithms are hard to overcome the rigid deformations. DMP uses mixture of multiscale deformable part models to describe an object detection system which represents highly variable objects [1], which has better robustness against deformation.

DPM (Deformable Part Model) as one of the most successful detection algorithms, was proposed by Pedro Felzenswalb in 2008 and was awarded the PASCAL VOC "Lifetime Achievement" Prize in 2010. Due to Felzenswalb's paper, the resulting system is both efficient and accurate, achieving state-of-the-art results on PASCAL VOC benchmarks and the INRIA Person dataset in 2007 [2]. The strong low-level features of DPM are based on the HoG (Histograms of

C. Pan · M. Sun (✉) · Z. Yan
The Third Research Institute of the Ministry of Public Security, Shanghai, China
e-mail: mingxiasun@163.com

N.Y. Yen and J.C. Hung (eds.), *Frontier Computing*, Lecture Notes in Electrical Engineering 422, DOI 10.1007/978-981-10-3187-8_3

Oriented Gradients) features. So the DPM can be considered as an upgrade of HoG in some ways [3].

As showed in Fig. 1, the upgraded HoG feature in DPM kept the "Cell" concept in HoG feature, but altered the normalization progress. The result shared similarity with the result of HoG feature as the upgraded HoG feature normalized the region which consisted of the target cell and the four surrounding cells. In order to reduce the feature dimension, P. Felzenswalb used PCA (Principal Component Analysis) [4] to analyze the unsigned gradients. As illustrated in Fig. 1, there are 31 dimensional features.

In P. Felzenswalb's work [5], he showed pedestrian detection model as following.

In Fig. 2, figure (A) shows the pedestrian, figure (B) is a root filter, figure (C) shows several part models with high resolution, figure (D) shows the spatial relationships of the part filters.

DPM uses a root filter and several part filters and the corresponding deformable model, the construction of the whole model is based on the pictorial structures. Normally, the part models use higher resolution than the root filter, about two times. Figure 2B, C illustrate the visual structure of the root and part models, it shows the weighted sum of SVM coefficients which oriented in gradient direction, and the brightness is proportional to the value. In order to reduce the complexity of the

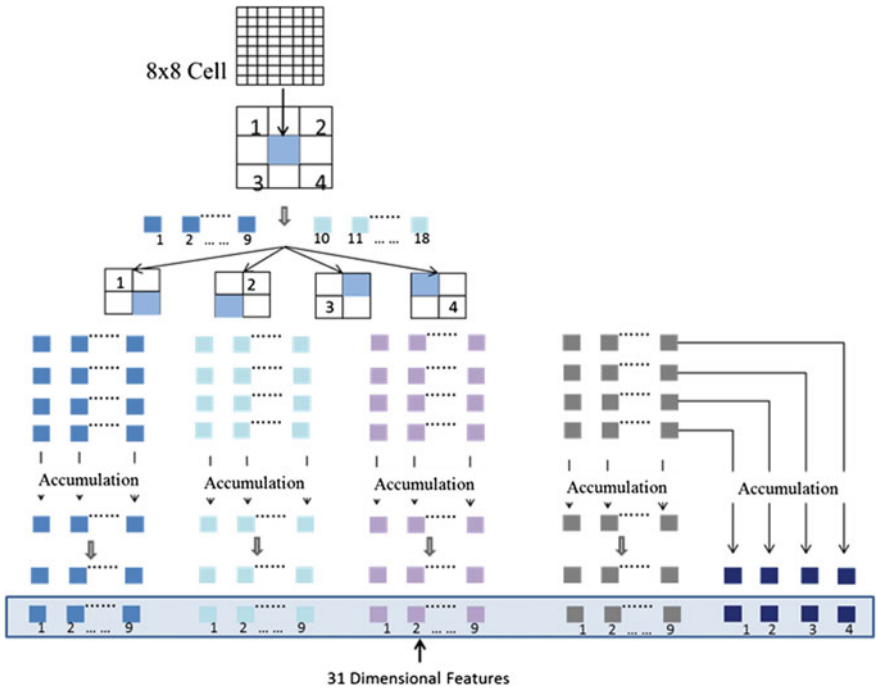

Fig. 1 The upgraded HoG feature in DPM

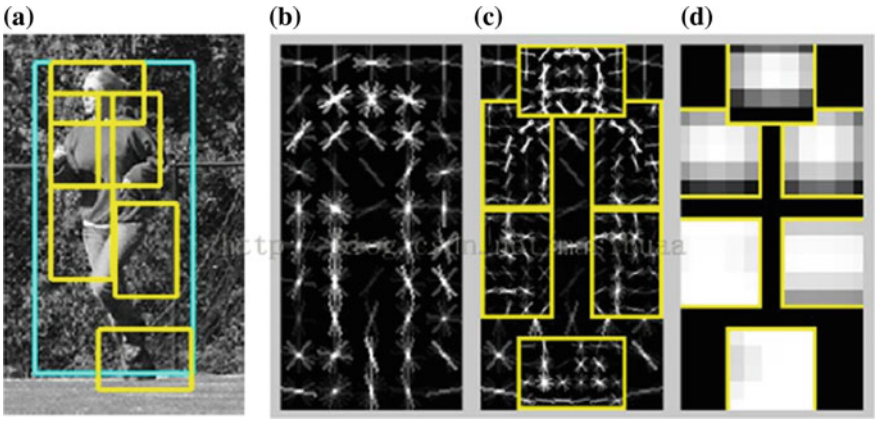

Fig. 2 DPM pedestrian detection model

whole model, the part models are symmetry. Figure 2D shows the deviation cost of the part model. The cost is zero in ideal case; the further the part model deviates, the greater the cost is. Then the target object can be represented by a collection of parts and the relative deformable position of the parts, the parts are connected by certain ways [6]. Each part describes local properties, and the spring-like connections are used to represent the relation between the deformable models [7]. As a single deformable model is not capable enough to describe an object, usually multiple deformable models are according with the request [8–15]. In this article, the variations among different vehicle types are quite significant, so the mixture of deformable models is required.

2 Methodology

In detecting progress, a scale pyramid is constructed and a scan window approach [4] is used to scan different layer of pyramid. Figure 3 shows the detecting process of DPM. In Fig. 3, the score of layer l_0 coordinate (x_0, y_0) can be calculated as follows [1]:

$$score(x_0, y_0, l_0) = R_{0,i_0}(x_0, y_0) + \sum_{i=1}^{n} D_{i,l_0-\lambda}(2(x_0, y_0) + v_i) + b.$$

$R_{0,i_0}(x_0, y_0)$ is the score of root filter, in other words, it expresses the matching degree between model and target. $\sum_{i=1}^{n} D_{i,l_0-\lambda}(2(x_0, y_0) + v_i)$ is the scores of n part filters. b is the root of set which is used to align the components. (x_0, y_0) is the

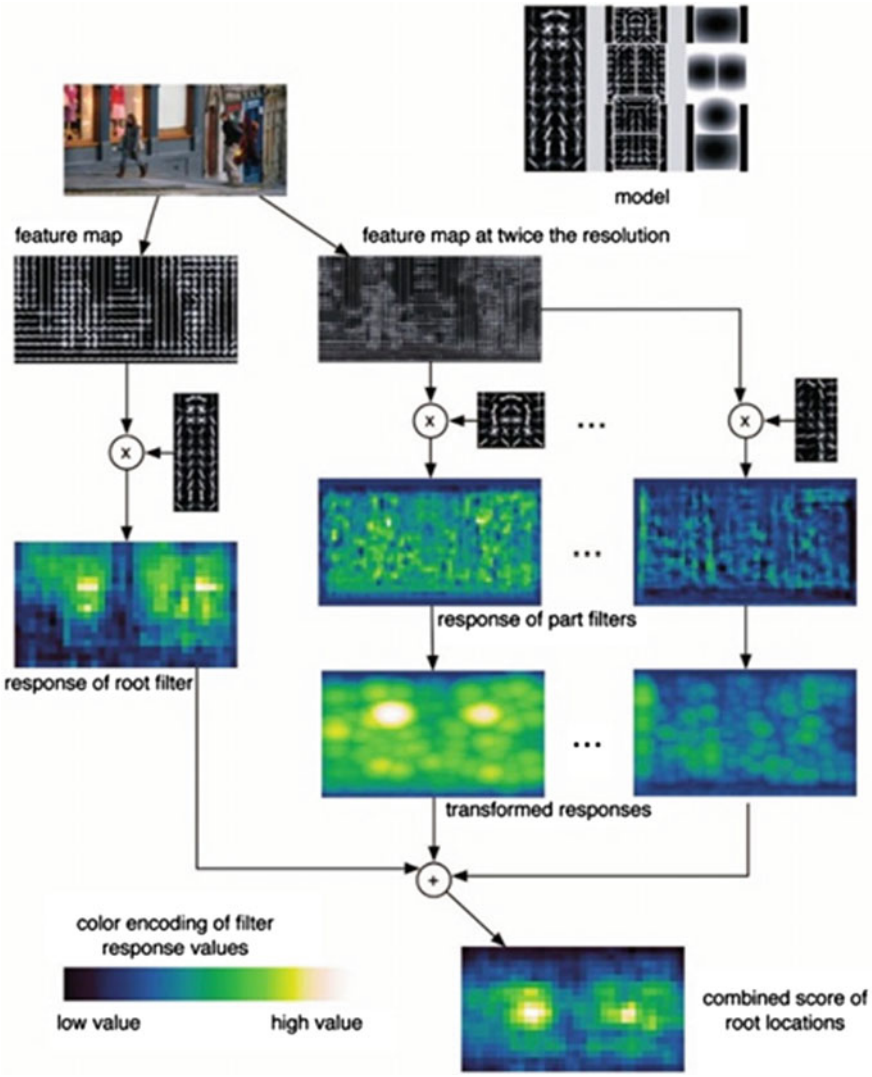

Fig. 3 DPM detection process [1]

coordinate of the root filter's left-top in the root feature map. $(2(x_0, y_0) + v_i)$ is coordinate of the i-th part filter in the root feature map.

The score of part filters can be calculated as follows [1]:

$$D_{i,l}(x, y) = \max_{dx, dy}(R_{i,l}(x + dx, y + dy) - d_i \cdot \Phi_d(dx, dy)).$$

$D_{i,l}(x, y)$ is the optimal solution of part filter, namely, it searches the anchor position and within a certain range for a proper location which has combined matching and optimal deformation. (x, y) is the ideal position of the i-th part filter in layer l. (dx, dy) illustrates the relative offset from (x, y). $R_{i,l}(x + dx, y + dy)$ is the matching score in coordinate $(x + dx, y + dy)$. $d_i \cdot \Phi_d(dx, dy)$ expresses the offset loss causing by the offset (dx, dy); $\Phi_d(dx, dy) = (dx, dy, dx^2, dy^2)$, d_i is the coefficient of offset loss, it is to be calculated in the training process. To initialize the model, $d_i = (0, 0, 1, 1)$ is the Euclidean distance between offset location and ideal location, namely the offset loss.

3 Experiment

3.1 Data Preparation

The original image data are captured from traffic surveillance system somewhere in JiangSu province. The training data which are used in DPM illustrated as following:

In training process, positive samples must be labeled with bounding boxes which are illustrated in Fig. 4B. In this experiment, 1700 images of vehicle front and 1900 images of vehicle rear as positive samples which are labeled with bounding boxes and the property files are generated.

3.2 Training Procedure

The training procedure is completed by initializing the structure of a mixture model and learning parameters. The parameters are learned by training LSVM (Latent

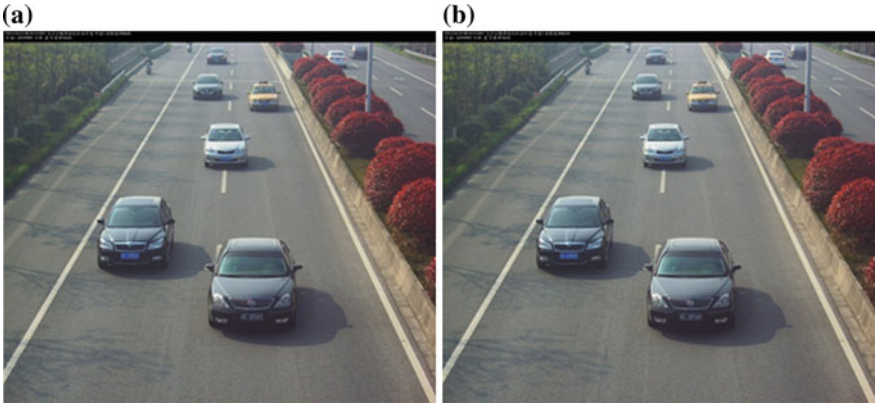

Fig. 4 DPM training data: vehicle front

Fig. 5 Training Procedure [1]

Data:
Positive examples $P = \{(I_1, B_1), \ldots, (I_n, B_n)\}$
Negative images $N = \{J_1, \ldots, J_m\}$
Initial model β

Result: New model β

$F_n := \emptyset$
for *relabel* $:= 1$ **to** *num-relabel* **do**
  $F_p := \emptyset$
  for $i := 1$ **to** n **do**
    Add `detect-best` (β, I_i, B_i) to F_p
  end
  for *datamine* $:= 1$ **to** *num-datamine* **do**
    for $j := 1$ **to** m **do**
      if $|F_n| \geq$ *memory-limit* **then** break
      Add `detect-all` $(\beta, J_j, -(1+\delta))$ to F_n
    end
    $\beta :=$`gradient-descent` $(F_p \cup F_n)$
    Remove (i, v) with $\beta \cdot v < -(1+\delta)$ from F_n
  end
end

Procedure `Train`

Support Vector Machine) [16]. The LSVM is trained by gradient descent algorithm and the data-mining approach [17, 18] with a cache of feature vectors (Fig. 5).

4 Results

In this experiment, three models are designed by DPM: vehicle front model is used to recognize the frontal side of a vehicle; vehicle rear model is used to recognize the back side of a vehicle; vehicle mixture model is used to capture either frontal or back side of a vehicle. There are two testing sets in this experiment: 100 vehicle front images and 100 vehicle rear images. The testing results of vehicle front model are illustrated as following:

DPM vehicle front model recognizing testing images of vehicle front			
Total image samples	Correctly recognized samples	Accuracy	Correctly recognized objects
100	93	93%	96

The testing results of vehicle rear model are illustrated as following:

DPM vehicle rear model recognizing testing images of vehicle front			
Total image samples	Correctly recognized samples	Accuracy	Correctly recognized objects
100	96	96%	112

In order to know which model performs better under the same conditions, we used the two models to recognize the same image objects and then outputted the results with the higher confidence degree.

DPM vehicle front and rear models recognizing testing images of vehicle front					
DPM vehicle front model			DPM vehicle rear model		
Total image samples	Correctly recognized samples	Accuracy	Total image samples	Correctly recognized samples	Accuracy
100	39	39%	100	69	69%
DPM vehicle front and rear models recognizing testing images of vehicle rear					
DPM vehicle front model			DPM vehicle rear model		
Total image samples	Correctly recognized samples	Accuracy	Total image samples	Correctly recognized samples	Accuracy
100	47	47%	100	76	76%

As the front and rear side of the same vehicle always share certain similarity, we conjecture the probability of using one model to recognize the two sides of a vehicle. So we used DPM to design a mixture model to capture the vehicles in 2-way lanes. The testing results show as following:

DPM mixture model recognizing testing images of vehicle front			
Total image samples	Correctly recognized samples	Accuracy	Correctly recognized objects
100	70	70%	70

DPM mixture model recognizing testing images of vehicle rear			
Total image samples	Correctly recognized samples	Accuracy	Correctly recognized objects
100	90	90%	203

5 Conclusions

By comparing the results from three DPM vehicle models, the non-mixed models acquired higher accuracy. But they performed not unsatisfactory in mixture test. In order to capture the vehicles in 2-way lane, we proposed the third mixture DPM vehicle model. It is more efficient to capture vehicle vision and shows high versatility.

Acknowledgements This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014603.

References

1. P. F. Felzenswalb, R. B. Girshick, D. McAllester and D. Ramanan, Object Detection with Discriminatively Trained Part Based Models. IEEE Trans. PAMI, 32(9):1627–1645, 2010.
2. M. Everingham, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC 2007) Results." [Online]. Available: http://www.pascalnetwork.org/challenges/VOC/voc2007/.
3. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in IEEE Conference on Computer Vision and Pattern Recognition, 2005.
4. Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in IEEE Conference on Computer Vision and Pattern Recognition, 2004.
5. P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In CVPR, 2008.
6. "Pictorial structures for object recognition," International Journal of Computer Vision, vol. 61, no. 1, 2005.
7. M. Fischler and R. Elschlager, "The representation and matching of pictorial structures," IEEE Transactions on Computer, vol. 22, no. 1, 1973.
8. X. Luo, Zheng Xu, J. Yu, and X. Chen. Building Association Link Network for Semantic Link on Web Resources. IEEE transactions on automation science and engineering, 2011, 8 (3):482–494.
9. C. Hu, Zheng Xu, et al. Semantic Link Network based Model for Organizing Multimedia Big Data. IEEE Transactions on Emerging Topics in Computing, 2014, 2(3), 376–387.
10. Zheng Xu et al. Semantic based representing and organizing surveillance big data using video structural description technology. The Journal of Systems and Software, 2015,102, 217–225.
11. Zheng Xu et al. Knowle: a Semantic Link Network based System for Organizing Large Scale Online News Events. Future Generation Computer Systems, 2015, 43–44, 40–50.
12. Zheng Xu et al. Semantic Enhanced Cloud Environment for Surveillance Data Management using Video Structural Description. Computing, 98(1–2):35–54, 2016.
13. C. Hu, Zheng Xu, et al. Video Structured Description Technology for the New Generation Video Surveillance System. Frontiers of Computer Science, 2015, 9(6): 980–989.
14. Zheng Xu et al. Crowd Sensing Based Semantic Annotation of Surveillance Videos, International Journal of Distributed Sensor Networks, Volume 2015 (2015), Article ID 679314, 9 pages.
15. Zheng Xu et al. Crowdsourcing based Description of Urban Emergency Events using Social Media Big Data. IEEE Transactions on Cloud Computing, doi:10.1109/TCC.2016.2517638.
16. S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in Advances in Neural Information Processing Systems, 2003.

17. H. Rowley, S. Baluja, and T. Kanade, “Human face detection in visual scenes,” Carnegie Mellon University, Tech. Rep. CMU-CS-95-158R, 1995.
18. K. Sung and T. Poggio, “Example-based learning for view-based human face detection,” Massachussets Institute of Technology, Tech. Rep. A.I. Memo No. 1521, 1994.