Xun Liu, Yin Zhang, San-yuan Zhang, Ying Wang, Zhong-yan Liang, Xiu-zi Ye, 2015.Detection of engineering vehicles in high-resolution monitoring images. *Frontiers of Information Technology & Electronic Engineering,* **16**(5):346-357. [doi:10.1631/FITEE.1500026]



# Detection of engineering vehicles in high-resolution monitoring images

**Key words:** Object detection, Histogram of oriented gradient (HOG), Dense scale-invariant feature transform (dense SIFT), Saliency, Part models, Engineering vehicles

Contact: Xun Liu
E-mail: star.liuxun@gmail.com
ORCID: http://orcid.org/0000-0002-3045-2943

# Introduction

- This paper presents a novel formulation for detecting objects with articulated rigid bodies from high resolution monitoring images, particularly engineering vehicles (Fig. 1).
- First, we detect object patches from monitoring images using a coarse detection process. It works fast with a low false negative rate and a high false positive rate.
- Second, we apply a refinement classification to determine the patches that actually contain objects.

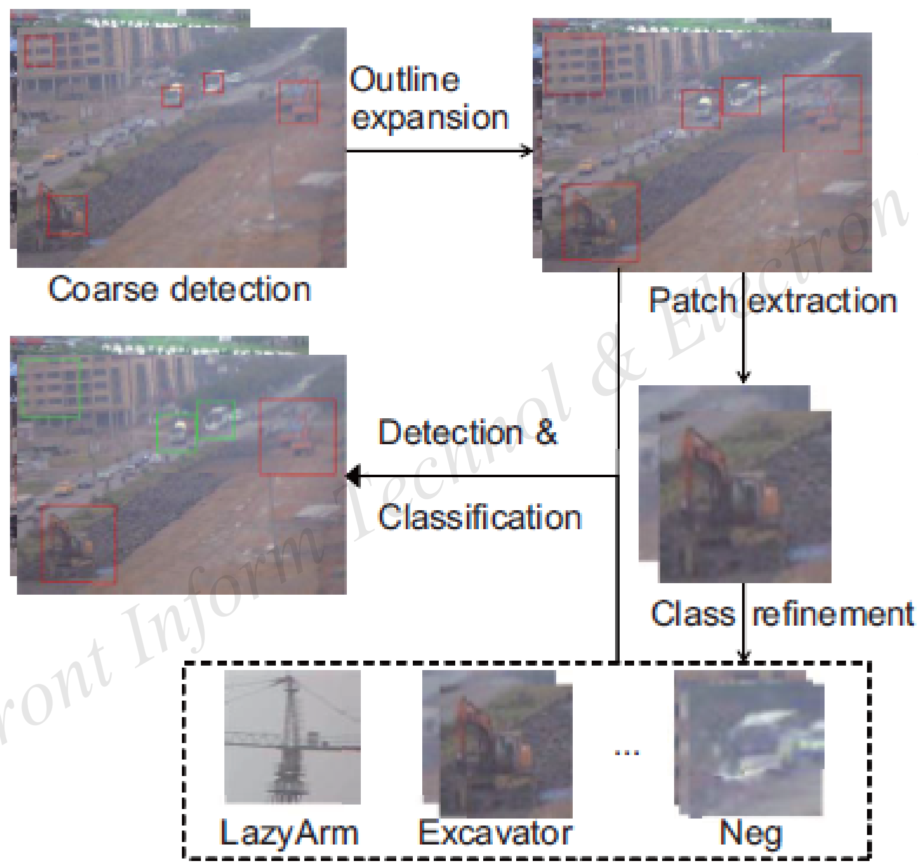


**Fig. 1 An overview of the proposed method**

# Coarse detection process

- Proposed the color frequency features, and added them to the classical HOG descriptors (Fig. 3).
- Used a linear support vector machine (linear SVM) to rapidly detect many image patches that may contain object parts.

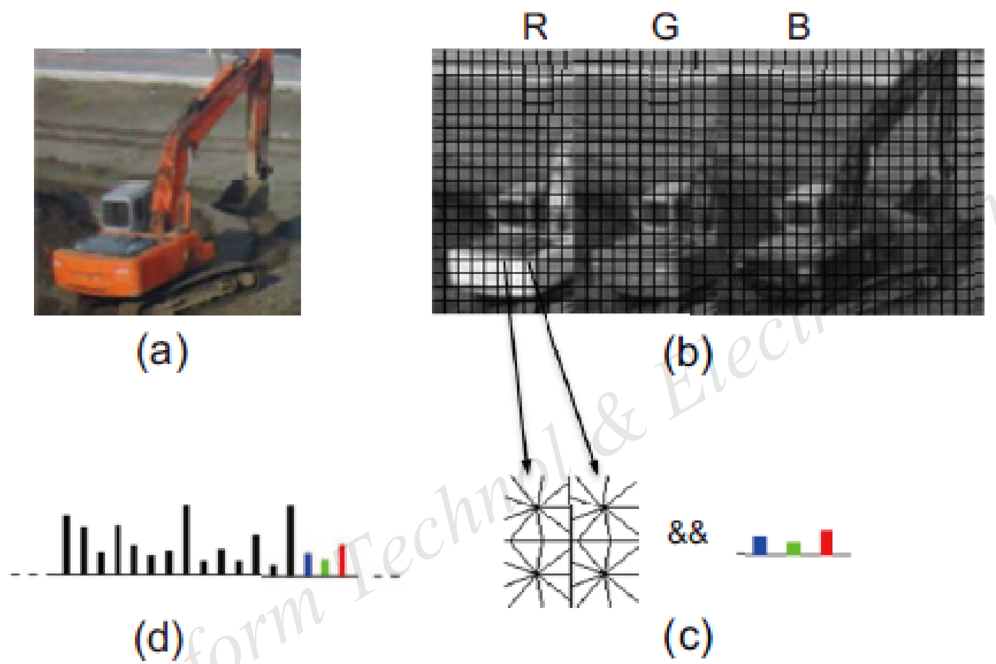


Fig. 3 HOG descriptor with color frequencies: (a) an image patch; (b) the RGB color spaces of (a); (c) the HOG features and color frequencies of (b); (d) the HOG descriptor added with color frequencies. References to color refer to the online version of this figure

# Refinement classification process

- Increased the size of the image patches so that they include the complete object using models of the object parts.

- Improved a saliency detection algorithm and used it to improve the performance of the dense SIFT descriptor (Fig. 8).

- Applied HIK to optimize the feature vector.

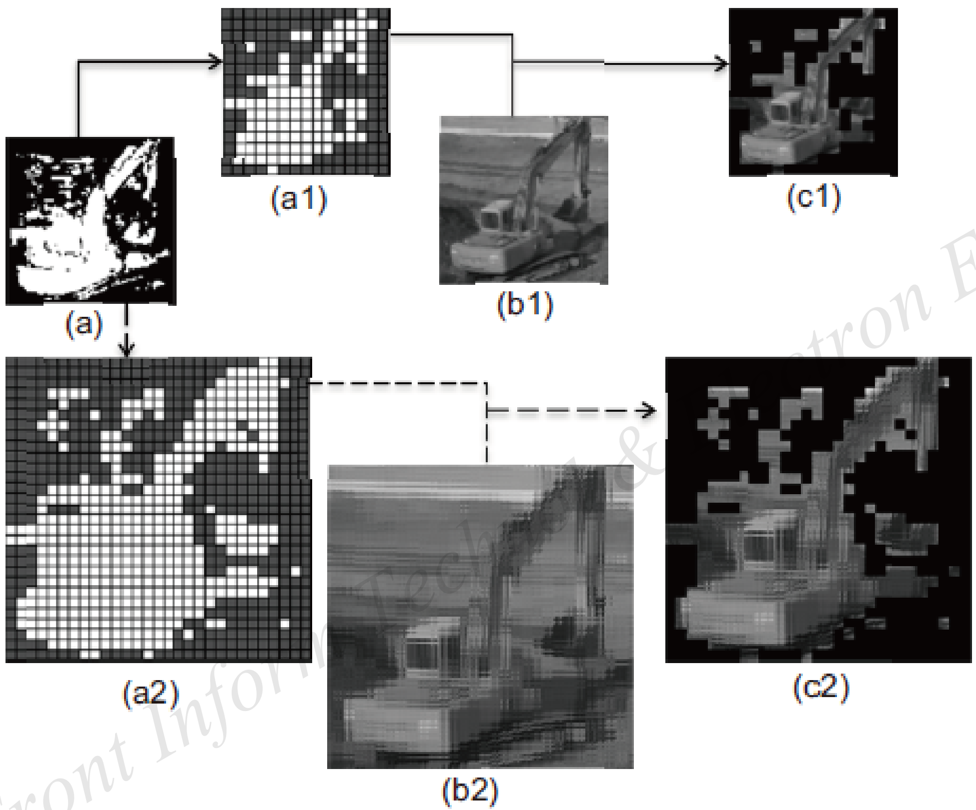




**Fig. 8** Process of using the saliency mask. There are two strategies: (a) is the binary saliency map, (a1) and (a2) are path maps of the binary saliency map, (b1) and (b2) are patch maps of the gray image, and (c1) and (c2) are saliency patch images

# Experimental results

**Table 1 Accuracy of our classification in the second phase (refinement classification phase) for the engineering vehicle dataset (%)**

| | CC | Crane | ET | LazyArm | Neg |
|---|---|---|---|---|---|
| CC | 87.1 | 0.0 | 10.3 | 2.6 | 0.0 |
| Crane | 14.2 | 82.6 | 3.3 | 0.0 | 0.0 |
| ET | 4.9 | 4.3 | 88.1 | 0.0 | 2.7 |
| LazyArm | 0.0 | 0.0 | 1.7 | 97.0 | 1.3 |
| Neg | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 |

CC: cement car; ET: excavator

# Experimental results (Con'd)

**Table 2 Comparison of the proposed method with existing methods for the engineering vehicle dataset (%)**

| Category | Ours 0.96 rate | Ours 0.92 rate | HOG | DPM 0.96 rate |
|---|---|---|---|---|
| Cement car | 83.7 | 80.2 | 75.7 | 77.5 |
| Crane | 79.5 | 76.2 | 72.3 | 75.9 |
| Excavator | 84.6 | 81.1 | 76.2 | 81.8 |
| LazyArm | 93.1 | 89.2 | 84.6 | 92.7 |

'Ours 0.96 rate' means that the detection rate in the first phase is set to 0.96; 'Ours 0.92 rate' sets the detection rate to 0.92; 'HOG' applies the classical gray-HOG descriptor in the first phase and the classical SIFT descriptor only with the pyramid HIK in the second phase; 'DPM 0.96 rate' uses the DPM method (Felzenszwalb *et al.*, 2010b) with a detection rate of 0.96

# Experimental results (Con'd)

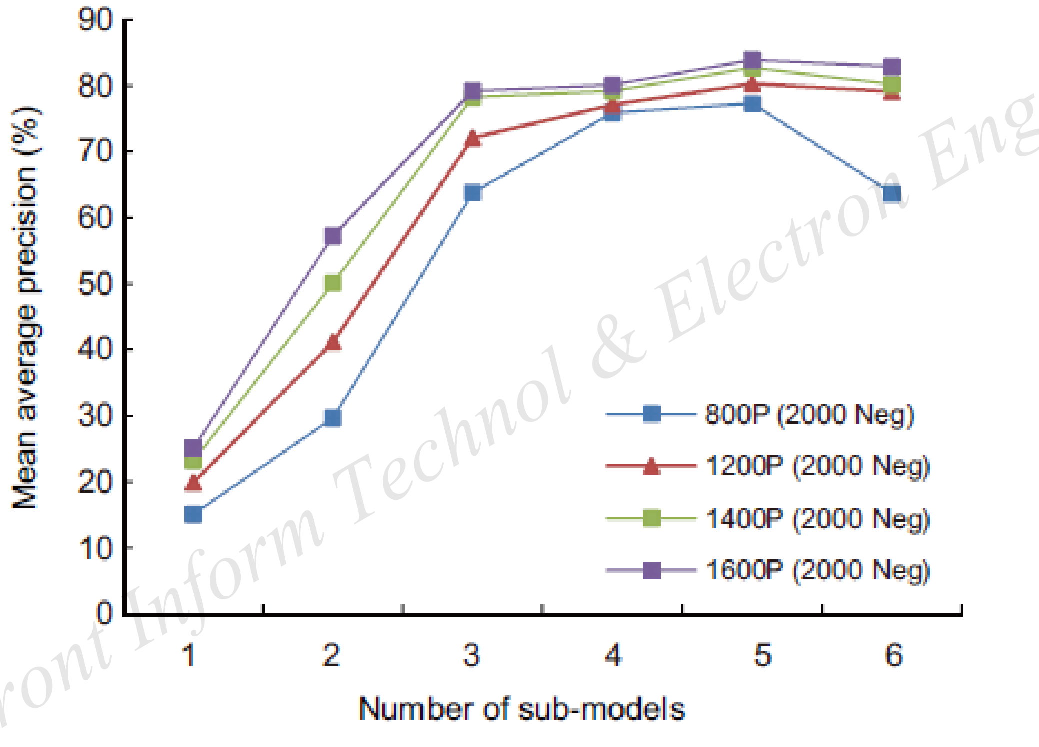


**Fig. 12** Comparison of DPM's results with training sets of different sizes, with the detection rate being set to 0.96 (Table 1). '1600P (2000 Neg)' implies that the training set has 1600 positive samples and 2000 negative samples

# Experimental results (Con'd)

Table 3 Results of our method and a comparison with the VOC 2012 winner (%)

| Category | Our method: Coarse detection phase* | Our method: Refinement classification phase | Our method: Entire detection method | VOC 2012 winner |
|---|---|---|---|---|
| Airplane | 82.3 | 89.5 | **73.7** | 65.0 |
| Bicycle | 70.1 | 84.8 | **59.4** | 54.5 |
| Bird | 33.8 | 70.5 | 23.8 | **25.1** |
| Boat | 29.3 | 72.9 | 21.3 | **24.9** |
| Bottle | 46.5 | 72.7 | **33.8** | 32.1 |
| Bus | 69.2 | 82.5 | 57.1 | **57.1** |
| Car | 68.7 | 81.7 | **56.1** | 49.3 |
| Cat | 60.2 | 83.3 | 50.1 | **53.7** |
| Chair | 28.5 | 60.3 | 17.2 | **19.5** |
| Cow | 47.8 | 78.3 | **37.4** | 35.3 |
| Dining table | 32.3 | 67.7 | 21.9 | **38.1** |
| Dog | 54.3 | 76.8 | 41.7 | **42.9** |
| Horse | 62.5 | 78.3 | 48.9 | **51.0** |
| Motorbike | 62.2 | 83.5 | 51.9 | **59.5** |
| Person | 72.2 | 84.3 | **60.9** | 46.1 |
| PP | 38.1 | 52.1 | 19.9 | **22.8** |
| Sheep | 60.6 | 71.9 | **43.6** | 40.3 |
| Sofa | 34.9 | 62.3 | 21.7 | **39.7** |
| Train | 58.3 | 83.8 | 48.8 | **51.1** |
| TM | 65.5 | 72.3 | 47.2 | **49.4** |
| MAP | 54.4 | 75.7 | 42.5 | **42.9** |

PP: potted plant; TM: TV monitor; MAP: mean average precision. A bold value means it is better. Though the method of the results of the VOC 2012 winner is not the same as the method of our results, we compared them because the detection rate of ours can be set in the coarse detection phase. * The false positive rate was set to 0.6

# Conclusions

- The method simulates human beings' behavior in looking for a target object from hundreds of different classes of objects. It is divided into two steps: coarse detection and refinement classification.
- Optimizing each phase, the method has a very good performance in the engineering vehicle dataset. Applying the method to VOC 2012 and VOC 2011 datasets indicate that the method also has a good generalization performance.