Learning-based Vehicle Detection Using Up-scaling Schemes and Predictive Frame Pipeline Structures

Yi-Min Tsai, Keng-Yen Huang, Chih-Chung Tsai, and Liang-Gee Chen
DSP/IC Design Lab, Graduate Institute of Electronics Engineering,
National Taiwan University, Taipei, Taiwan
{*ymtsai, kyhuang, cctsai, lgchen*}*@video.ee.ntu.edu.tw*

Abstract

This paper aims at detecting preceding vehicles in a variety of distance. A sub-region up-scaling scheme significantly raises far distance detection capability. Three frame pipeline structures involving object predictors are explored to further enhance accuracy and efficiency. It claims a 140-meter detecting distance along proposed methodology. 97.1% detection rate with 4.2% false alarm rate is achieved. At last, the benchmark of several learning-based vehicle detection approaches is provided.

1. Introduction

Nowadays, electronic vehicular technologies have become dominant roles in vehicle industry. On account of the maturity of vision sensors, the vision-based Advanced Driver Assistance System (ADAS) becomes an emerging application for guiding drivers. Collision Warning Systems (CWS) attempt to prevent vehicles from crashes. All these applications require robust detection and recognition of on-road vehicles. Therefore, this research field has drawn intensive attention recently.

2. Prior Arts

Sun et al. made an overview for vision-based on-road vehicle detection [8]. Knowledge-based methods utilize appearance cues including edge [1], corner [2], and symmetry [1, 2] for detection. However, their performance may decisively rely on complexity of content. Motion-based methods use motion vectors such as optical flow [10] to locate objects with large displacement but such methods suffer from correspondence problems.

In recent years, machine-learning makes progress in object detection and recognition [3–7, 9]. Features like Haar-like [6, 9], histogram of oriented gradient (HOG) [4, 6], and Gabor [3, 7] are adopted in object classification frameworks. It is believed that learning-based methods have decent performance for vehicle analysis.

However, most researches are primarily focused on near (0∼30 meters) or medium (30∼60 meters) distance detection while few investigate far distance (over 60 meters) part. In addition, few literatures well organize spatial and temporal information into one object detection flow.

In this paper, a sub-region up-scaling detection scheme is proposed to improve far vehicle detection accuracy. The relationship between object size and detecting distance is revealed. Then we introduce predictive frame pipeline structures for collaboration of object predictors and sub-region up-scaling detection routine.

This paper is organized as following. We briefly introduce an Adaboost-based approach in Sec. 3. The proposed method is presented in Sec. 4. In Sec. 5, experimental results are discussed. Sec. 6 summarizes our exploitation and contributions.

3. Adaboost-based Approach: Overview

Viola and Jones [9] proposed an object recognition method using Adaboost with Haar-like features and combine it with a sub-window scaling detection scheme.

Detection: To detect both small-sized and large-sized objects, sub-windows progressively scan entire image and are up-scaled for the next scan phase (Fig. 1).

Recognition: To classify these scaled sub-windows, a stage cascade composed of trained weak classifiers is constructed. Non-object sub-windows can be rapidly rejected with early stage classifiers. More complex stage classifiers are performed subsequently on object-like sub-windows only if they passed through previous stage classifiers.

Grouping: Eventually, each detected object contains multiple overlapped neighboring sub-windows, which indicates multiple hits. Neighboring sub-windows are grouped into a single window representing the detected object. The more neighboring sub-windows are combined, the higher confidence the object has.

Limitation: However, as for distance issues, this approach may fail to detect far vehicles because of higher grouping thresholds, especially as sub-window scaling factor is close to one. On the other hand, lower thresholds tend to generate much false detection.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.759

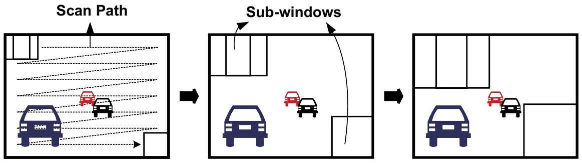


Figure 1. Scaled sub-window detection routine for VJ's method.

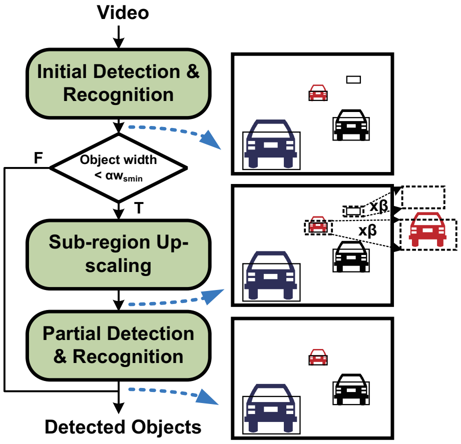

Video
Initial Detection & Recognition
Object width < αw$_{smin}$
F
T
Sub-region Up-scaling
Partial Detection & Recognition
Detected Objects
xβ
xβ

Figure 2. Proposed sub-region up-scaling (SRUS) flow.

4. Proposed Methods

4.1. Sub-region up-scaling

Adaboost with Haar-like features is adopted for recognizing rear part of vehicles. We derived a sub-region up-scaling (SRUS) approach based on Sec. 3. Fig. 3 describes that object width in pixel w_o is about inversely proportional to distance D between host vehicles and preceding vehicles. The relationship among detecting sub-window size, object size, and distance is given by,

$$\begin{aligned} w_o \times D &= K \times W \\ &= (w_{s_{avg}} + w_{\Delta}) \times D \\ &\cong \frac{\sum_{i=0}^{G-1} w_{s_i}}{G} \times D \\ &= \frac{\sum_{i=0}^{G-1} (w_{s_{min}} \cdot SF^{p_i})}{G} \times D \end{aligned} \tag{1}$$

where K is a constant ranging from 2.0 to 2.5 depending on pixel aspect ratio and sensor parameters. W is image width in pixel. The minimum sub-window width is $w_{s_{min}}$. $w_{s_{avg}}$ is the window width after grouping G neighboring sub-windows with width w_s. w_{Δ} is the width error that is close to zero as scaling factor SF approaches 1 and step index p approaches a large number. Hence, distance D can be approximately estimated as $K \cdot W / w_{s_{avg}}$. If p equals 0, Eq. (1) results in a theoretical maximum detecting distance, $K \cdot W / w_{s_{min}}$.

The strategy (Fig. 2) follows above observations. Firstly, after initial detection and recognition, two grouping thresholds for object confidence c are set according to scaled sub-window width. The lower one is adopted for accepting sub-windows with width smaller than $\alpha \cdot w_{s_{min}}$, which leads to increase of detection as well as false alarms at the distance longer than $K \cdot W / \alpha \cdot w_{s_{min}}$.

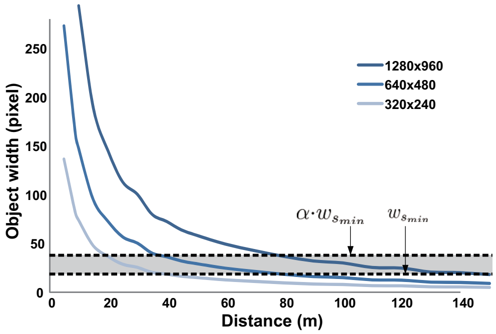


Figure 3. Relationship between distance and object width in pixel.

Secondly, regions of the detected windows (after grouping and width less than $\alpha \cdot w_{s_{min}}$) are up-scaled with a factor β. Detection and recognition routines are re-applied to these up-scaled regions with the previous high threshold. For each region, this verification procedure returns either original or no object as result and update the re-detected object's confidence.

Therefore, with a specific resolution, the SRUS scheme attempts to improve detection accuracy at the distance longer than $K \cdot W / \alpha \cdot w_{s_{min}}$ by adjusting the α value.

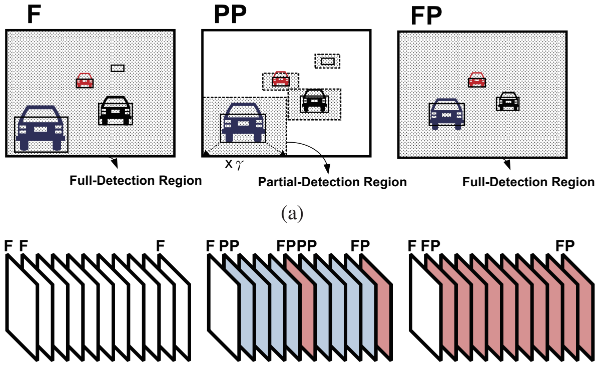
4.2. Object predictor

Kalman estimators are utilized to predict objects' characteristics in successive frames. Three attributes are measured and predicted: 2D location ($\hat{x}$, $\hat{y}$), size ($\hat{w}_{s_{avg}}$, $\hat{h}_{s_{avg}}$) and confidence $\hat{c}$. These state variables are used to describe dynamics of detected objects in 2D space. The Kalman estimator can be formulated by Eq. (2) (3),

$$\hat{\mathbf{x}}_t = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{w}_{s_{avg}} \\ \hat{h}_{s_{avg}} \\ \hat{c} \end{bmatrix}_t, \mathbf{x}_t = \begin{bmatrix} x \\ y \\ w_{s_{avg}} \\ h_{s_{avg}} \\ c \end{bmatrix}_t \tag{2}$$

$$\begin{aligned} \hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t-1} + \kappa(\mathbf{x}_t - \hat{\mathbf{x}}_{t-1}) \\ \hat{\sigma}_t^2 &= (1-\kappa)\hat{\sigma}_{t-1} \\ \kappa &= \hat{\sigma}_{t-1}^2 / (\hat{\sigma}_{t-1}^2 + \hat{\sigma}_t^2) \end{aligned} \tag{3}$$

where parameters with hat notations indicate prediction terms. σ is the estimation uncertainty and κ is the update gain. Once an object is detected, a prediction phase is immediately executed. If an object is observed and detected in frame t, its predictor had been initiated in a previous frame, such as frame $t-3$. In addition, if the detected object is not re-detected in some preceding frames, its predictor will be terminated through Kalman dynamics. Owing to object predictors, false alarms can be significantly reduced without degrading detection rate.



(a)

(b)

Figure 4. Predictive frame pipeline structure. (a)Frame types;(b)Pipeline types.

4.3. Predictive frame pipeline structure

A predictive frame pipeline structure comprises three frame types (Fig. 4(a)): full-frame-detection without prediction (F), full-frame-detection with prediction (FP) and partial-frame-detection with prediction (PP). In PP frame, partial detection regions are decided along predictors' attributes, object location and size, which are described as Eq. (4),

$$\begin{aligned} W_{DR} &= \gamma \hat{w}_{s_{avg}} \\ H_{DR} &= \gamma \hat{h}_{s_{avg}} \\ Center &= (\hat{x}, \hat{y}) \end{aligned} \tag{4}$$

where W_{DR} and H_{DR} are width and height of a partial detection region respectively. γ is a region expanding factor. Detection and recognition routines are only performed on the partial detection regions, which avoid scanning other redundant locations. Besides, the SRUS procedure is applied if $\hat{w}_{s_{avg}}$ is smaller than $\alpha{\cdot}w_{s_{min}}$. On the contrary, in F or FP frame, a full-frame detection routine is executed.

In a pipeline structure, predictors are only initiated in F or FP frame and only terminated in PP or FP frame. An F frame can be an intermediate frame or an initial frame. In Fig. 4(b), F-F-F structure indicates that detection routine is performed independently on each frame. Instead, F-NPP-FP and F-FP-FP structures both involve object prediction mechanism. N is the number of PP frames between two FP frames. A FP frame should be inserted into the two pipeline structures to ensure not to miss overtaking vehicles or reappearing vehicles from occlusion.

5. Experimental Results

Experimental sequences were captured from a 1080p, 24fps CMOS front-mounted camera with 640×480 and 1280×960 intermediate resolutions. For Adaboost, 2034 rear parts of vehicles and 2540 non-vehicles are trained to build a 20-stage cascade with total 764 weak classifiers. The trained window size is 20×16 ($w_{s_{min}}$=20). For comparison, we also implement VJ's method using Intel OpenCV library.

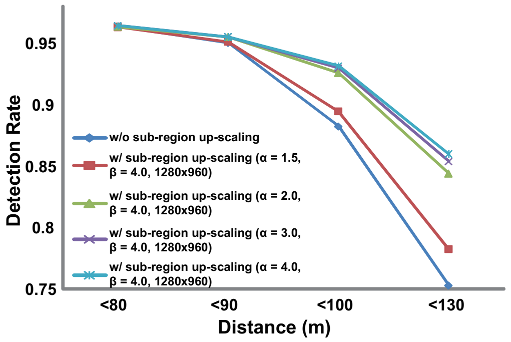


Figure 5. Degradation of detection rate with respect to detecting distance.

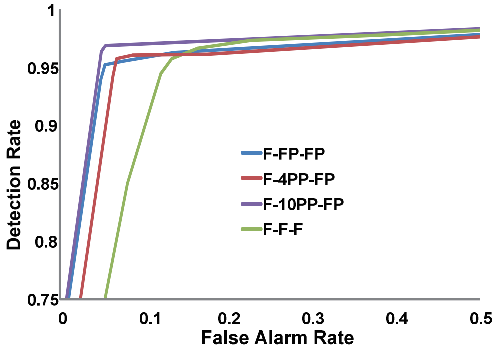


Figure 6. Accuracy evaluation. ROC curves of three frame pipeline structures.

Fig. 5 shows the detection rate (DR) descending ratio under 5% false alarm rate (FAR). The DR drastically drops as distance increases without SRUS (about 20% reduction within 30 meters). On the other hand, with SRUS, the DR drop becomes slight as α increases. It has only 11.2% and 8% reduction as α equals 2.0 and 3.0, respectively. The phenomenon saturates when α equals 4.0, which infers further raise α is less beneficial. Furthermore, β value should be greater than 3.0 to guarantee an adequate up-scaling range for SRUS.

Receiver operating characteristic (ROC) curves of three pipeline structures are constructed (Fig. 6). The F-F-F structure yields a degraded accuracy compared to that of F-NPP-FP and F-FP-FP structures. Under low FAR (less than 8%), DR of either F-FP-FP or F-NPP-FP is superior to that of F-F-F with 5 to 20% margin. The number of PPs in F-NPP-FP affects DR (Fig. 7). The results suggest a trade-off for choosing the counts of PPs, especially with much overtaking vehicle participation. For a regular scenario, there are less overtaken situations and usually their duration may not exceed 10 seconds. This explains F-10PP-FP outperforms F-FP-FP and F-4PP-FP in Fig. 6.

Methods	**Detection**	**Recognition** (feature/classifier)	**Object Predictor**	**Accuracy**(%) (DR/FAR)	**Average Processing** **Time**(s/frame) (640×480/1280×960)	**Maximum Detecting** **Distance**(m)
Fu [5]	Static ROIs	Edge/SVM	Particle filter	87.6 / N/A	N/A	<60 (*medium*)
EGFO [7]	Edge-based	Gabor/SVM	-	91.0 / 6.4	N/A	<60 (*medium*)
BGF [3]	Static ROIs	Gabor/Boosting+SVM	-	95.8 / 8.8	N/A	<60 (*medium*)
VJ [9]	Scaled sub-window	Haar-like/Adaboost	-	96.4 / 15.5	3.98 / 12.81	60∼80 (*medium*) @1280×960
Proposed	Scaled sub-window+ sub-region up-scaling	Haar-like/Adaboost	Kalman filter	97.1 / 4.2 (F-10PP-FP) 95.5 / 4.8 (F-FP-FP)	0.97 / 3.82 (F-10PP-FP) 3.47 / 11.35 (F-FP-FP)	140 (*far*) @1280×960

Table 1. Benchmark of state-of-the-arts and proposed method.

The state-of-the-arts are analyzed in six aspects (Table 1). Our method achieves a 97.1% DR and only 4.2% FAR with F-10PP-FP structures. In average, only 3.82 seconds per frame is needed for a 1280×960 resolution video. The distance bound where FAR less than 80% is defined as the maximum detecting distance. Consequently, our method provides a 140 meters detection capability. In the end, Fig. 8 illustrates far vehicle detection results.

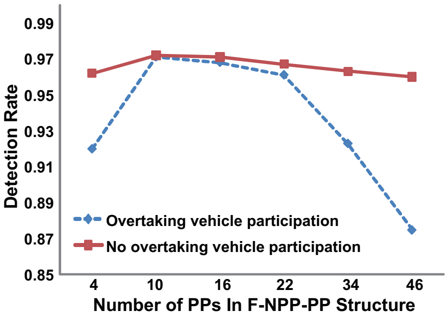


Figure 7. Effect of number of PPs in different driving situations.

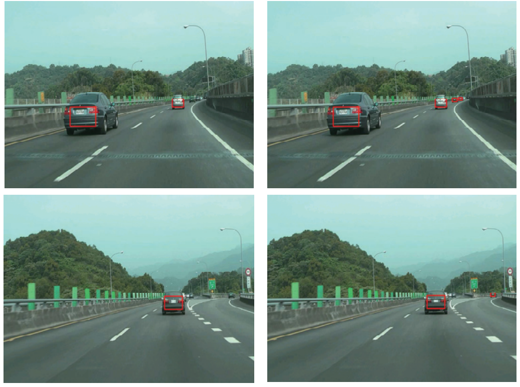


Figure 8. Detection results (red bounding boxes) without and with SRUS in F-10PP-FP structures are shown in left and right column respectively.

6. Conclusions

Our contribution is twofold. (1) We introduce sub-region up-scaling scheme for far vehicle detection based on the observation on relationship between object size and image resolution. (2) Frame pipeline structures are established for cooperation of object predictors and single-frame detection routine. The results show that the proposed method has convincing performance on vehicle detection and recognition.

References

[1] D. Alonso, L. Salgado, and M. Nieto. Robust vehicle detection through multidimensional classification for on board video based systems. In *Proc. IEEE ICIP*, volume 4, pages IV–321–IV–324, Sept. 2007.

[2] J. Arrospide, L. Salgado, M. Nieto, and F. Jaureguizar. On-board robust vehicle detection and tracking using adaptive quality evaluation. In *Proc. IEEE ICIP*, pages 2008–2011, Oct. 2008.

[3] H. Cheng, N. Zheng, and C. Sun. Boosted gabor features applied to vehicle detection. In *Proc. IEEE ICPR*, volume 1, pages 662–666, 2006.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE CVPR*, volume 1, pages 886–893, June 2005.

[5] C.-M. Fu, C.-L. Huang, and Y.-S. Chen. Vision-based preceding vehicle detection and tracking. In *Proc. IEEE ICPR*, volume 2, pages 1070–1073, 2006.

[6] P. Negri, X. Clady, S. Hanif, and L. Prevost. A cascade of boosted generative and discriminative classifiers for vehicle detection. *EURASIP Journal on Advances in Signal Processing*, 2008.

[7] Z. Sun, G. Bebis, and R. Miller. On-road vehicle detection using evolutionary gabor filter optimization. *IEEE Trans. on Intelligent Transportation Systems*, 6:125–137, June 2005.

[8] Z. Sun, G. Bebis, and R. Miller. On-road vehicle detection: A review. *IEEE Trans. on PAMI*, pages 694–711, May 2006.

[9] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE CVPR*, volume 1, pages I–511–I–518, 2001.

[10] J. Wang, G. Bebis, and R. Miller. Overtaking vehicle detection using dynamic and quasi-static background modeling. In *Proc. IEEE CVPR*, pages 64–64, June 2005.