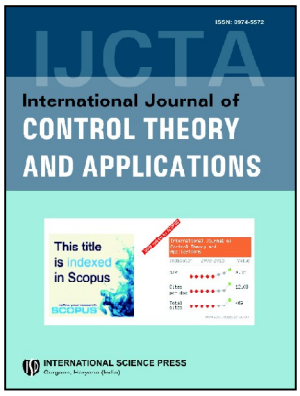

International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 9 • Number 49 • 2016

Pedestrian Detection for Surveillance Network- Computer Vision Task

Sharma S., Sameera Shaik and Gudapati Ramyasri

Department of Electronics & Department of Computer science, Vignan's University, Vadlamudi, Guntur, India
E-mail: Sharma_ece@vignanuniversity.org

Abstract: The paper presents a real time pedestrian detection with improvement in speed and detection quality. By efficient handling of geometric information provided by cameras in transfer of different scales and computation from test to training time, detection time is improved. It is a non-trivial task when extraction of information is done at variable speeds more than 100 Hz. A new method is proposed to estimate the ground-obstacles boundary (along with the distance), fastest stixel based people detection without depth map computation is proposed. Using the machine that could handle the processing speeds of the techniques developed it is possible to process 130 fps from rectified input to detections output, while in case of monocular images system could provide detections with high quality at 50 fps for debugging process. This analytical approach provides a means to achieve the best detection performance on different challenging datasets like INRIA, Caltech-USA datasets and also for the real time captured videos, reducing the average miss rate over HOG + SVM by more than 30%.

Keywords: stixel, pedestrian detection, fps, classifiers, speed.

1. INTRODUCTION

Fast detection of objects is one of the most important application will result in the increase of quality and high speed that enables inclusion into larger systems extensive subsequent processing (initialization for segmentation or tracking).It also focuses to minimize the false positive and false negative detections to improve secure applications.

The proposed method has two new algorithmic speed-ups, one approach is better scale handling in monocular images and the other is based on exploiting the geometric information of frames in high quality detections. The main aim is to detect the people at 100 fps (frames per second) [1] without any loss in the quality [10] of results. Using direct processing methods for the estimation of the objects above the ground called as "Stixel World" estimation [2, 3] could help in detection of pedestrians in terms of milliseconds.

2. RELATED WORK

Numerous studies have proposed literal solutions to people counter using Computer Vision. The primary factor is to consider the position and orientation of cameras used for detection as many problems emerge depending on

the way they are captured. In order to provide solution different stages of algorithms are implemented to improve the quality of detections.

(A) Better Features and Classifier

To quickly compute the rectangular averages [7] using integral images has been introduced by Viola and Jones. The proposed method has the computational features that capture the best input image information with least cost. It is also known that exploiting depth and motion [9, 11] cues improves the quality of detection [5, 6] but with reduced speed compared to the actual one. Linear classifiers such as Adaboost and Random/Hough forests are generally preferred as the non-linear classifiers suffer from the low speed.

(B) Cascades

The average computation time is reduced by the reduction of number of false positives in the initial stages of processing. To achieve this efficient classifier is used that could split the speed classifiers into sequence of simple classifiers.

3. PROPOSED ALGORITHM

To provide a pedestrian detection with high quality at nearly 100 fps a detector that is almost 20 times faster and reduced false positives on the INRIA dataset is obtained at high speed. This work is based on the following core concepts.

(A) Ground Plane Estimation

Here the ground plane is estimated at q,-disparity method but the collection of actual evidence is from the matching different disparities of left and right image rows. This evidence is collected from each row finds the ground plane using the robust line fitting. The ground plane is represented as f ground: $V \rightarrow D$ that maps the image rows q, to specific disparity d = f ground (q,). Computation of one-out-of-N rows is done to avoid collecting the information from every row below the horizon which improves the speed. Even when handling large images this speeding up does not degrade the quality.

(B) Stixel Distance Estimation

In general the stixel distance estimation uses dynamic programming problem to collect the evidence column-wise in the image to estimate the distance using in μ-disparity domain.

This approach in general suffers from different problems:

(I) Ignores Horizontal Gradient

When combining the left and right images the information used is aligned to the vertical gradient for different disparities since the horizontal gradient do not have information for stereo matching. The q, - disparity cost matrix [2] contains half of the information present in the image.

(II) Computes more than Required Information

Due to quantization effect, unwanted quantizations and also more number of computations are carried out than required. This causes redundant computations that reach the resolution of sub-pixel close to the horizon. Hence unwanted computations are avoided in object detection to estimate the position of the object.

To solve these problems the parameterization has to be changed from μ-disparity domain [2, 12] to the μ-q, domain.

To exploit the horizontal gradient information particular boundary is to be weighted by the gradient magnitude at some pixels so that all the pixels lie below the horizon. The alternate approach is to split the image in vertical row bands and in each band the pixel with maximal horizontal gradient is to be selected that increases the chances of finding the objects accurately.

The row band 1 is termed as *bi* . The row that is selected inside band *bi* and stixel *q j* is given as *v*(*qj* ,*bi*) and in the ground plane model

$$d(q_j , b_i) \; ? \; f_{ground}\,(v(q_j, b_i)) \tag{1}$$

Object of interest have an image width greater than one column computing evidence for each column is highly redundant. Each stixel *q j* is located at column *u*(*qj*) ? *j*.*stixel*_ *width* . By selecting the different stixel widths and row band heights the amount of data extracted from an image can be controlled.

1. Dynamic Programming Formulation

The goal is to find the optimal row band for each stixel

$$b_s^*(q) = \arg\min_{b(q)} \sum_q c_s(q, b(q)) + \sum_{q_a, q_b} s_s(v(q_a, b(q_a)), v(q_b, b(q_b))) \tag{2}$$

Where q_a , q_b are neighbours (|a-b| = 1), C_s is the data term and S_s is the smoothness term.

This problem has an efficient solution using Dynamic Programming [2, 13]. For each stixel column q and row band b, the evidence supporting the presence of stixel in the left image by computing the cost *cs* (*q*, *b*) which is "stixel cost" is calculated. The lower the cost is more chance that the stixel is present.

$$c_s\,(q, b) \; ? \; c_0\,(u(q), d(q,b)) \; ? \; c_g\,(u(q), d(q,b)) \tag{3}$$

Where $c_0\,(u, d)$ is the "Object Cost", the cost of vertical object present and $c_g\,(u, d)$ is the "Ground Cost", the cost of the supporting ground present [2].

The ground plane estimate is used to wrap the right image such as cg can be computed between left images and wrapped right image for efficient implementation. This wrapping can be done even for Non-linear ground plane models.

The smoothness term ss enforce to respect the left-right occlusion rstrictionss and promote ground object boundaries with a few amounts of jumps.

$$s_s(v_a, v_b) = \begin{cases} \infty & if\ d(v_a) < d(v_b) - 1 \\ c_o(u_a, d(v_a)) & if\ d(v_a) \approx d(v_b) - 1 \\ -w.c_o\,(u_a, d(v_a)) & if\ q_a = q_b \\ 0 & if\ d(v_a) > d(v_b) - 1 \end{cases} \tag{4}$$

(A) Object Detection using Stixels

The previous frame stixels are used as the guidelines for the current frame detections. The proposed detector uses "Very Fast Detector" [1]. The available information about the centre position of the stixel and its expected scale is used by detector to search few pixels up and down in the column. Since the algorithm evaluates only useful rows*columns instead of evaluating every row and column the search space and time is reduced. The depth information method [2] called as "Stixel World" is used to accelerate the object detection.

1. Object Detection without image resizing

The main idea is to move the resizing of the image from test time to train time using the insight of the FPDW detector [4] and reverse it. To adjust the given classifier to classify correctly the feature response approximation has to be computed at different scales. The strong classifier is to be built from a set of decision trees, each containing the set of three stump classifiers. Each stump classifier is defined by a channel index, a rectangle over such a channel and decision threshold*?* . To rescale a stump the channel index needs to be kept constant, scale the rectangle by relative scale factor s and update the threshold as

$$\tau' = \tau . r(s) \tag{5}$$

A canonical classifier is converted into different scales. To train the baseline detector, train $N / K(\sim 5)$ classifiers, one for each octave. By using the approximation to transform the N / K classifiers into N classifiers the integrated channel features on the output image and the response for each scale using N classifiers can be computed.

2. Algorithmic Speed-up

By skiping the estimation of features multiple times it is interesting to make the detector compute scaling features for first half of the time and the remaining half to evaluate the classifier responses computing features only once that provides the 1.9 times more speed than the algorithm.

Compared to FPDW assuming canonical scales 0.5,1,2,4 and avoiding the image resizing and then using Very fast Classifier [1] (2.68 Hz) instead of FPDW [4] (1.55 Hz) may increase the speed of 1.57 times the actual speed.

IV. IMPLEMENTATION

There are two quality improvements over the baseline detector "ChnFtrs" .The release of LatSvmV4 [12] presents the significant improvements over LatSvmV2 [12]. It is comfortable to claim that ChnFtrs provides a state-of-art for single part template and is competitive with the initial versions of the part based detector. The base detector is based on the idea of "Integral Channel Detectors" that have single rectangular features that sum a filter response over a given image area. In case of pedestrian detection 6 quantized orientations, 1 gradient magnitude and 3 LUV colour channels to get a state-of-art results.

A set of two level decision trees are conducted and then linearly weighted to obtain a strong classifiers and their weights is learned via Discrete Adaboost. The strong classifiers has 2000 weak classifiers, the features are selected from a random pool of 30000 rectangles with a set of 3000 random negative samples. The classifier is trained and evaluated using the INRIA pedestrian set. By shrinking the features by a factor of 4 (after computing the feature response and before creating the integral image) the faster training and data can be evaluated. Another relevant feature of the detector is training is very fast. The training time and the memory consumption is stable [10] even the learning model has larger dimensions.

By comparison with HOG+SVM [6] this is a simple classifier that may be able to compete with sophisticated approaches such as HOG part-based models [12]. The basic difference is the use of learned features versus hand designed features. By placing the HOG cells uniformly this detector learns where to place the features to maximize the discriminating power.

(A) Stixel Estimation

The ground truth annotations of the pedestrian in the sequence Bahnhof (999 frames), the vertical distance between the bottom of the annotated bounding box and the estimated stixel bottom are computed. This may result in a cumulative absolute error in the sequence. By using a high number of row bands and vary the stixel width of 5 pixels the quality is almost unaffected, but if width increases the quality may drop.

The computing stixels in *u* ? *v* domain computes a slight quality improvement with respect to *u* - disparity domain. Using the horizontal gradient evidence directly works that improves robustness in noise in the ground plane estimate and bypasses the disparity quantization effects.

A stixel width of 1 pixel and 128 row bands is used in u-v stixel implementation we can reach 45Hz oh the high end machine. Stixel width of 3 pixels, 25 row bands and cumulating the evidence along 1 column provides the same quality as *u* -disparity stixels [2].

(B) Pedestrain Detection at 100fps

The evaluation from a baseline detector running at 0.08 Hz up to GPU detections at 135 Hz. In terms of speed the monocular images at 50Hz are more than 7 times faster than reported 6.5 Hz on the actual machine. Also the result is nearly 10 times faster than the cudaHOG results [8] reported from high end implementations and the quality is also twice as good as the cudaHOG [8]. The speed measured by the high performance machine the measured time includes all the computations, the time to download the results and non maximal suppression on the machine. The ground plane and stixels are estimated at the frame t-1 and fed to the computations at frame t resulting in high speeds when computing over the Bahnhof images (640*480 pixels) over the 55 scales, averaged over 1000 frames of the sequence.

Table 1
Relative speed-up of each aspect of the proposed detector, with respect to the baseline detector

Detector Aspect	*Relative Speed*	*Absolute Speed*
Baseline Detector	1 *	1.38 Hz
Ground Plane Detector	2 *	74.01 Hz
Stixel Detector	1.35 *	119 Hz

V. RESULTS AND DISCUSSION

(A) INRIA Dataset Results

The INRIA dataset dataset is used to train and evaluate the detector quality. the diversity of contents helps to expose the differences in performance using various methods.

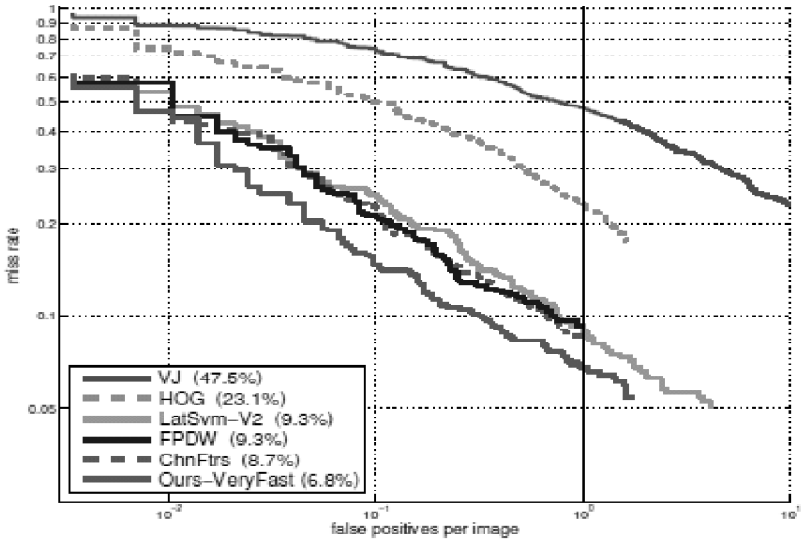

Figure 1: Quality of detector compared to other variants on INRIA people dataset

Figure 1 represents the results of N / K with other methods of evaluation. By rescaling the input images to compute the feature responses at different scales it can be referred as "Multiple Scales" Detector. The detector is competitive in terms of quality [10] with respect to ChnFtrs and provides significant improvement over the HOG + SVM algorithm. The results of INRIA datasets from ground estimation, stixel estimation and the person detection can be represented in Figure 3, where each frame could capture the detection value based on the confidence values of the algorithm.

(B) Bahnhof and other Dataset Results

The other datasets provides challenging sequences from a person who moves along the crowded side walk that allows to evaluate the stereo information [5, 14] and to detect its quality. The detector quality stays constant by using the ground estimation and stixel world detection while during pedestrian detection [6] the speed is improved in very fast detector above the HOG+SVM, but stixels from t-1 frames is used to reach the desired speed and to guide the detection to t frames which may slightly reduce the quality.

Figure 2 shows the detection of the persons from Bahnhof dataset representing the stixel estimation and pedestrian detection results with better speed and quality.

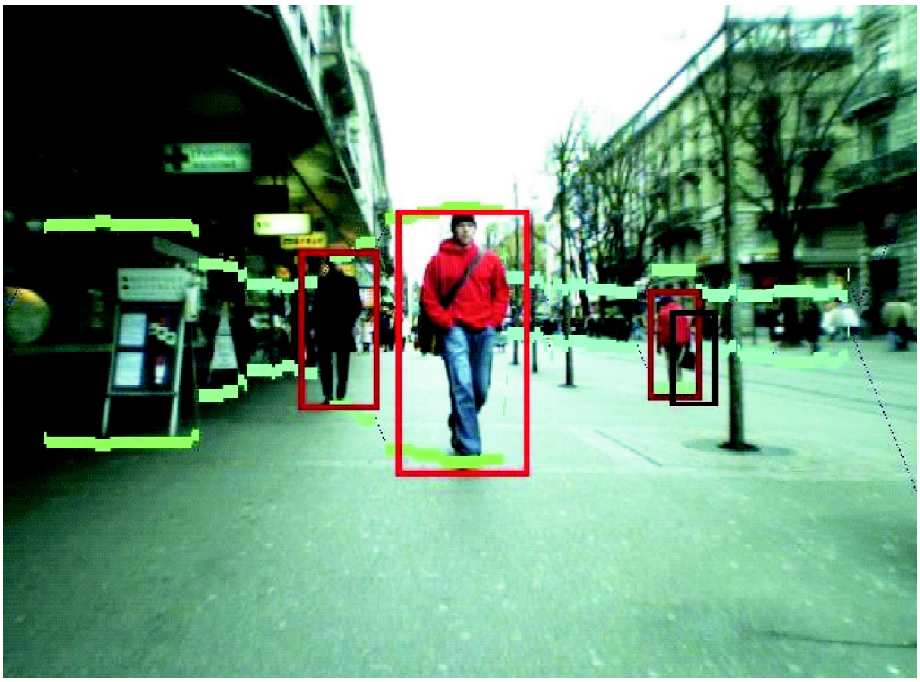

Figure 2: Stixel World and Pedestrian Detection Output for Bahnhof Dataset

The stixel representation adjusts the level of stixels based on detected height of the person or object and then it carries out the pedestrian detection at the rate of approximately 119 Hz and with the high quality of detection.

Figure 3 represents the Pedestrian detection for the INRIA database at the rate of about 105 Hz so that the each frame could capture the detection value based on the confidence values of the algorithm. Also from figure 1 it is clear that the very fast detector used has very less chances of false positives and miss rate is also least compared to all other detection approaches.

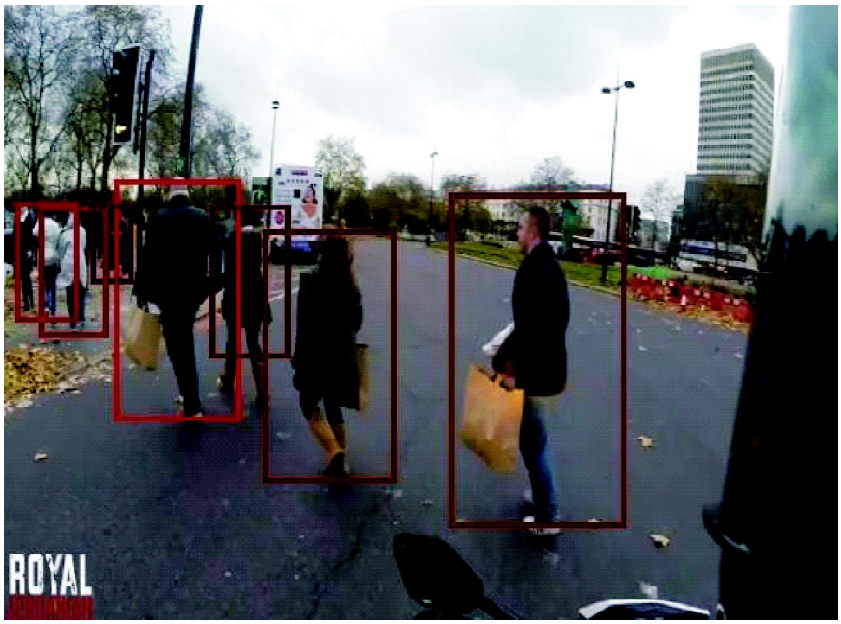

Figure 3: Pedestrian detection output for INRIA Sequence crossing the road

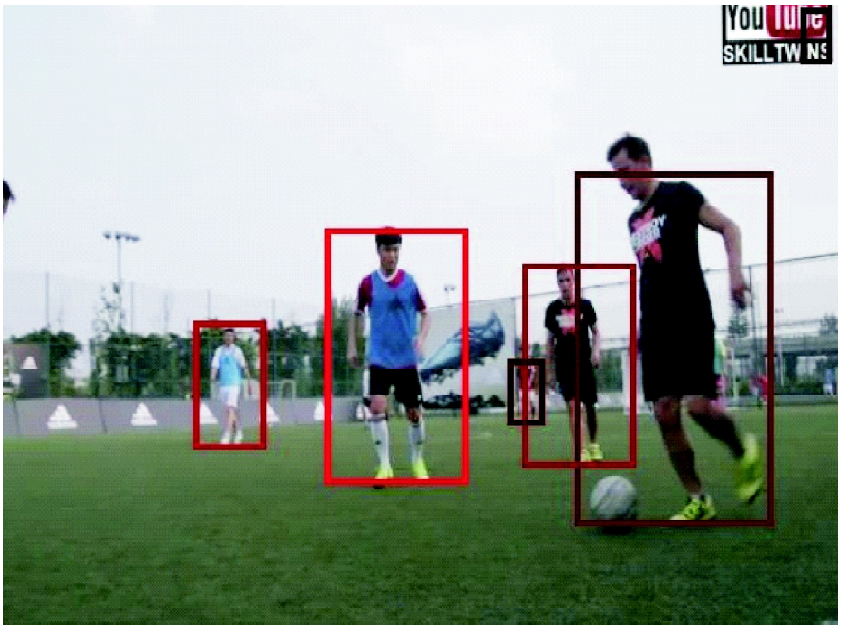

Figure 4: People detection for frame captured from the Football Match Video

Figure 4 is the dataset collected from the football match video in which the people can be easily detected completely. The detection is always based on the confidence values in which if the confidence levels are high then the person could be detected with the red box representing the person or object is detected at the point.

Figure 5 represents the input taken from the real time video taken from the webcam or it can also process the recorded video sequences which could also detect the person by high speed of operations and also without reducing the quality of the image. This real time capture can also work with more than 90% efficiency with the captured rate of approximately 97 Hz depending on the quality of video or frame being processed.

Figure 5: Frames captured from the Webcam

The detector quality stays constant while using ground plane and stixel based detection using Very fast Detector that could the high quality and high speed detections without suffering from much loss.

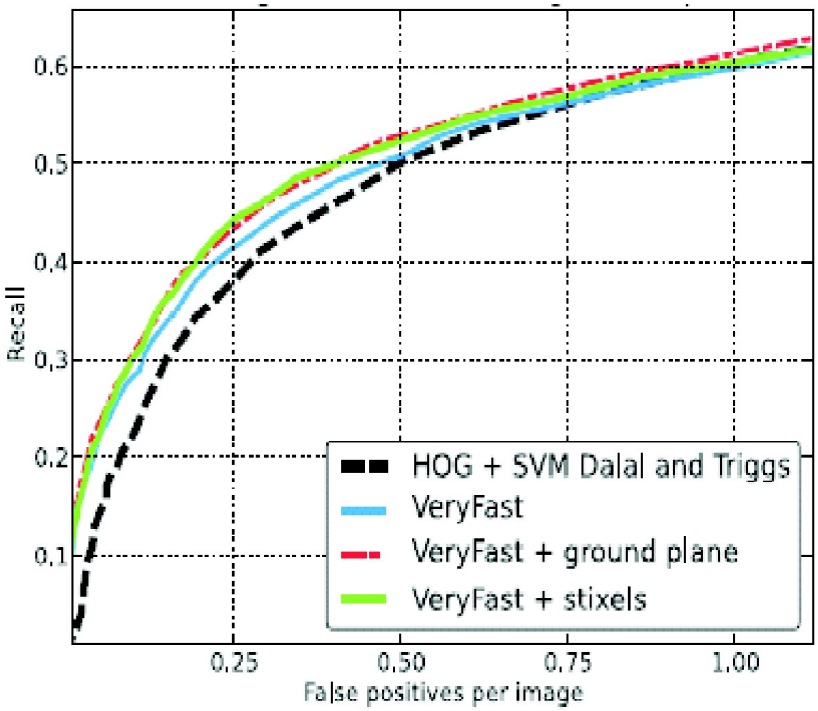

Figure 6: Results obtained on the Bahnhof Sequence

Figure 6 represents the detector performance is better than the HOG + SVM, which results in the expected quality gain.

CONCLUSION

The proposed novel based Pedestrian detector reverts the FPDW detector [4] which is running at 135 fbs in order to access the geometrical information quickly and avoid the resizing of the image from the frames captured at multiple scales by using the recent method. The approach is in coincidence with the Viola and Jones [7] idea "scale the features, not the images", applied to HOG based features.

The main advantage for the future hardware based implementations is due to high standards of parallelism in evaluation process. The current system can also extend its work for improving the quality of classifier training to multi- class/multi-view detection of vehicles and objects used for high and standard applications.

REFERRENCES

[1] Benenson, R., Mathias, M., Timofte, R., Van Gool, L.: Pedestrian detection at 100 frames per second. In: CVPR. (2012).

[2] R. Benenson, R. Timofte, and L. Van Gool. Stixels estimation without depthmap computation. In ICCV, CVVT workshop,2011.

[3] H. Badino, U. Franke, and D. Pfeiffer. The stixel world - a compact medium level representation of the 3d-world. In DAGM, 2009.

[4] P. Dollár, S. Belongie, and P. Perona. The fastest pedestrian detector in the west. In BMVC, 2010.

[5] M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. H. Matthies. A fast stereo-based system for detecting and tracking pedestrians from a moving vehicle. IJRR, 28:1466–1485, 2009.

[6] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In ECCV, 2006.

[7] P. Viola and M. Jones. Robust real-time face detection. In IJCV, 2004.

[8] P. Sudowe and B. Leibe. Efficient use of geometric constraints for sliding-window object detection in video. In ICVS, 2011.

[9] Gunyel, B., Benenson, R., Timofte, R., Van Gool, L.: Stixels motion estimation without optical _ow computation. In: ECCV. (2012).

[10] Ess, A., Leibe, B., Schindler, K., Van Gool, L.: Robust multi-person tracking from a mobile platform. PAMI (2009)

[11] Seki, A., Okutomi, M.: Robust obstacle detection in general road environment based on road extraction and pose estimation. In: IVS. (2006).

[12] P. Felzenszwalb, R. Girshick, and D. McAllester. Cascade object detection with deformable part models. In CVPR,2010.

[13] Kubota, S., Nakano, T., Okamoto, Y.: A global optimization algorithm for real-time on-board stereo obstacle detection systems. In: IVS, Turkey (June 2007)

[14] Nedevschi, S., Danescu, R., Frentiu, D., Marita, T., Oniga, F., Pocol, C., Schmidt, R., Graf, T.: High accuracy stereo vision system for far distance obstacle detection. In: IVS. (2004).