2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

# A Lazy Decision Approach Based on Ternary Thresholding for Robust Target Object Detection*

Jae-Yeong Lee[1], Wonpil Yu[1], Jungwon Hwang[2], and ChangHwan Kim[2]

***Abstract*— One of the main problems of binary classification of overlapping distributions is that there always exist misclassification errors with any value of threshold. In this paper, we propose a novel lazy decision approach for robust object detection and tracking, where decision on an uncertain observation whose evaluation lies between low and high thresholds is postponed until a clear evidence appears. As a practical application of the proposed approach, we present a sensor fusion pedestrian detection system for safe navigation of UGVs in driving environment. We combine a laser-based detection of target candidates and vision-based evaluation within the proposed lazy decision framework. Experimental results on real test data demonstrate effectiveness of the proposed approach, showing significant improvement of precision-recall performance.**

## I. INTRODUCTION

The problem of target object detection in an image can be viewed as a problem of binary classification, where each image region is classified as target object or not. An ideal case is one that feature distributions of target object and background have clear separation and thus can be classified completely by a single threshold. In most real cases, however, problems are more complex and distributions are often largely overlapped, giving inevitable misclassification errors.

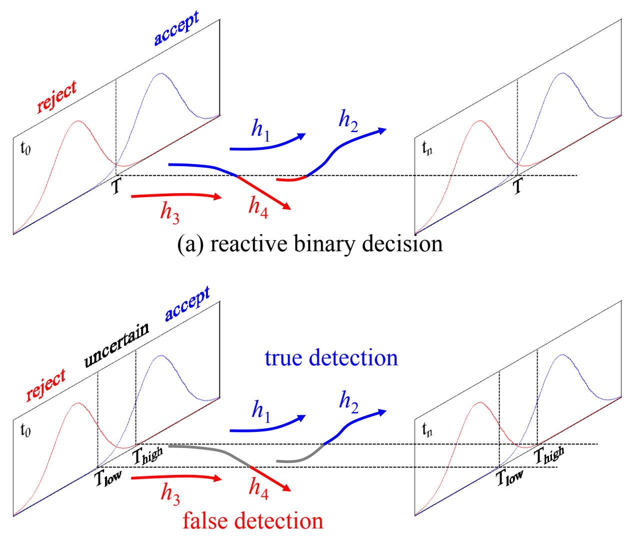
Conventional approaches have mainly focused on finding discriminant features [1], [2] or optimal threshold [3], [4] to minimize misclassification errors. However, it generally is hard to find discriminant features and often impossible. The other approach of optimal thresholding also has a limitation in that there always exists a misclassification error with any value of threshold for overlapping distributions.

In this paper, we propose a novel lazy decision approach for object detection and tracking, where decision on an uncertain observation whose evaluation lies between low and high thresholds is postponed until a clear evidence appears. In the proposed approach, we do not use typical binary classification by a single threshold. Instead, we use so called ternary thresholding where detections larger than a high threshold are classified as true positives and detections below than a low threshold as false positives. Detections between the low and high threshold are classified as unknown and decision on them is postponed for later evaluation.

*This work was supported by the R&D program of the Korea Ministry of Trade, Industry and Energy (MOTIE) and the Korea Evaluation Institute of Industrial Technology (KEIT). (The Development of Low-cost Autonomous Navigation Systems for a Robot Vehicle in Urban Environment, 10035354)

[1]J. Lee and W. Yu are with Intelligent Robot Research Department, Electronics and Telecommunications Research Institute, Daejeon, South Korea 305-700. jylee at etri.re.kr

[2]J. Hwang and C. Kim are with Center for Bionics, Korea Institute of Science and Technology(KIST), Seoul, South Korea.



(a) reactive binary decision

(b) proposed lazy classification

Fig. 1. (a) Convention classification framework (reactive binary decision). (b) Proposed lazy decision framework based on ternary thresholding.

In the literature, the notion of delayed decision making or lazy evaluation has mainly been used for denoting a coarse-to-fine strategy [5], [6] or decision making based on temporal context [7], [8]. In coarse-to-fine approach [5], [6], decision is made progressively by hierarchical analysis of the problem or by cascade classification, where the main purpose is for computational efficiency. In the other approach [7], [8], which is more close to our approach, the decision is delayed until final stage of the detection process or multiple observations available. The behind idea is that decisions that are difficult based on single observation can be made much easier using data from multiple observations.

Our lazy decision approach is different from [7], [8] in that the decision is delayed only when it is currently ambiguous. That is the decision can be made instantly even at the beginning of the detection process if the observation or evidence is sufficiently clear. With this scheme we thus are able to minimize the time delay caused from lazy decision. When being applied for object detection in videos, the proposed lazy decision approach has an effect of reducing false alarms while preserving high detection rate.

As a practical application of the proposed approach, we present a sensor fusion pedestrian detection system for safe navigation of UGVs in driving environment. The developed pedestrian detection system firstly detects target candidates

978-1-4799-3684-7/14/$31.00 ©2014 IEEE

from laser range data and then applies a vision-based evaluation for identifying true pedestrians within the proposed lazy decision framework.

The paper is organized as follows. In Sect. II we describe our lazy decision approach with the developed pedestrian detection system. In Sect. III experimental results on real video sequence is given with discussions. Finally, we conclude the paper in Sect. IV.

## II. PEDESTRIAN DETECTION SYSTEM

Many recent works for pedestrian detection is vision-based [2], [9], [10], [24]. However, vision-based approaches have limited performance in terms of robustness and processing time and hard to obtain relevant accuracy concerning depth. Another popular approach is to detect pedestrians based on a radar or laser sensor [11], [12], [13]. However, they also have a drawback in that it is hard to discriminate true pedestrians from other vertical objects like trees and lamp posts especially in driving environment. Compared with these approaches, sensor fusion approach [14], [15], [16], [17] that combines 2D laser sensor and camera is very promising in that it is able to utilize both accuracy and robustness of laser sensor and plentiful information of camera image.

Our pedestrian detection system belongs to a sensor fusion approach of laser and camera and originally developed as a moving obstacle detector for an unmanned ground vehicle for safe navigation. The developed pedestrian detection system utilizes a pair of front camera and single channel laser scanner mounted on a vehicle.

### A. System Overview

Figure 2 shows overall architecture of the developed pedestrian detection system. Range data processing module detects and tracks candidate objects from successive range data. Detected candidates from the range data processing module are then mapped onto the camera image, defining a region of interest (ROI) for vision-based pedestrian detection. In the image processing module, pedestrians are detected from the ROIs by applying a HOG (Histogram of Oriented Gradient) [2] detector. Detection results from the image processing module (at every frame) are matched with the maintained object hypotheses based on the tracking information provided from the range data processing module. The matched detections then are used for updating the matched hypotheses. For each of unmatched detections, a new hypothesis is created and maintained. Finally, the updated object hypotheses are evaluated for determining true positives within the proposed lazy decision framework.

One characteristic of our sensor fusion architecture is that tracking information from range data processing module is utilized not only to find candidates but also to enhance detection rate of the system. Conventional sensor fusion approaches of laser and camera [14], [15], [16], [17] use clustered detection results of range data mainly for defining ROIs for further vision-based detection. In our approach, we utilize the tracking information of the detected clusters of range data for compensating the possible failure of vision-based detection as well as for defining ROIs, which is described in more detail in Subsect. II-D.

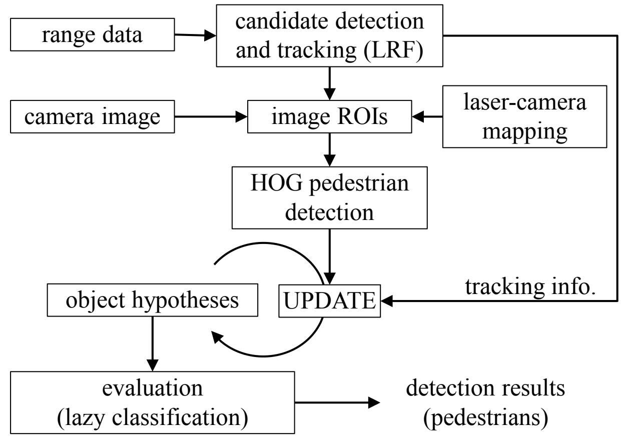



Fig. 2. The sensor fusion architecture of the developed pedestrian detection system.

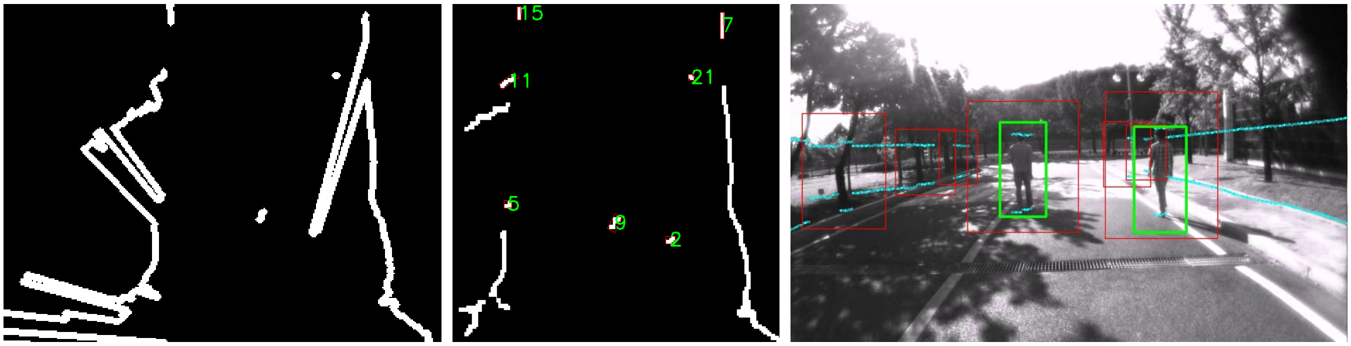
### B. Laser-based Detection and Tracking

The range data processing module detects and tracks candidate objects which are expected to be a pedestrian by analyzing range scan data. A raw laser scan data is first projected on a grid map [18], forming an occupancy map. Next, candidate objects for pedestrian are detected from the segmented clusters of occupied cells by considering the size and shape condition. The candidate objects detected are then tracked by using a particle filter [19], [20]. Particles are predicted based on a constant velocity motion model and their likelihood are evaluated by Mahalanobis distance between particle and sensor data. We adopt a Monte Carlo sampling to reduce computational load and use a recursive Bayesian filter to propagate probabilities.

### C. Vision-based Detection

Image ROIs are obtained by mapping detected clusters of range data into the camera image. We adopt HOG (Histogram of Oriented Gradient) features [2] and apply a sliding windows method for each of image ROIs to detect pedestrians. Since range data gives accurate depth information of the detected clusters, we do not need to build image pyramid and simple sliding window detection on one scale image is sufficient. It is because we can determine a possible scale of pedestrian in the image from the corresponding depth information. Consequently the search space for vision-based detection is greatly reduced in both image searching area and the scale, speeding up the search greatly and also reducing possible false alarms.

Let $M(\cdot)$ be a mapping function from metric coordinate to image coordinate and $M_h(\cdot)$ be a mapping function that computes an estimated pixel height of pedestrian in the image. For a given metric coordinate $p = (x, y)$ of a candidate object in vehicle coordinate system, $M(p)$ and $M_h(p)$ gives corresponding image pixel coordinate of the center of bottom




(a) (b) (c)

Fig. 3. An example of sensor fusion pedestrian detection. (a) Input range data. (b) Detection and tracking results of candidate objects in range data processing module. (c) Projection of raw range data into the image (cyan), extracted ROIs (red box), and the detection results (green box) of vision-based processing.

position of the object on the ground and estimated pixel height of the object, respectively. The mapping functions are obtained from a process of sensor calibration of the camera and laser. The image ROI is then defined as

$$[x'-h/2-\delta_x \leq x \leq x'+h/2+\delta_x, y'-h-\delta_y \leq y \leq y'+\delta_y], \quad (1)$$

where $M(p) = (x', y')$ and $M_h(p) = h$ and $\delta_x$ and $\delta_y$ are predefined constants.

Figure 3 shows an example process of the developed sensor fusion pedestrian detection system. Candidate objects detected from the range data processing module are depicted by red boxes with labels denoting their identity (tracking information) as shown in Fig. 3b. The range data processing module detects clusters of range data that meet a predefined size condition as candidate objects. The mapping of a raw range data onto the camera image for each image column is represented by a pair of points and depicted by cyan colored points in Fig. 3c. The upper cyan colored points denote the vertical upper bound of possible pixel position of pedestrians in the image (head top position) and the lower points denote the vertical lower bound of pedestrian region (foot tip position). This vertical image segments is similar with stixel world [21], [22] and bounds a possible range of image region for pedestrians. Based on this mapping relation, image ROIs for vision-based detection are computed from the positions of the laser-based detections, which are depicted by red boxes in Fig. 3c. After appropriate scaling of the image ROIs based on the depth information, we then apply HOG detector for each of scaled ROIs to find pedestrians. Note that we do not merge overlapping ROIs since each ROI defines a scale as well as position of pedestrian.

## D. Update of Object Hypotheses

We maintain single hypothesis for each pedestrian and update the hypotheses according to current processing results of range data and camera image. A hypothesis $h$ is represented by a 4-tuple

$$h =< b, p, l, f, c >, \quad (2)$$

where $b$ denotes a bounding box position in the image, $p$ denotes a metric position in world coordinate, $l$ denotes an associated label of the hypothesis, $f$ denotes a feature vector, and $c$ denotes the object class. The feature vector $f$ is a set of (image and laser) features of the object, which is used for the evaluation of the hypothesis in later lazy decision step. The $c$ denotes the current decision on the class of the hypothesis. In our application, $c$ can be one of three values of *pedestrian*, *non pedestrian*, or *candidate pedestrian*. Initially, it is set to be *candidate pedestrian* at the time of creation of the hypothesis.

**Algorithm 1** Hypothesis Update

Input

$H^{t-1} = \{h_1, h_2, ..., h_s\}$: object hypotheses

$L^t = \{\langle p_i, l_i \rangle\}$ : result of laser-based detection and tracking

$V^t = \{\langle b_j, l_j \rangle\}$: result of vision-based detection

Initialize

$H^t \leftarrow H^{t-1}$

For each $h_i$ in $H^t$,

1. If $h_i$ is successfully tracked by LRF and detected in the image, update $h_i$ with the current detection result
2. If $h_i$ is successfully tracked by LRF but not detected in the image, update $h_i$ with a predicted image position
3. If $h_i$ fails by LRF tracking, delete $h_i$ from $H^t$

For each $\langle b_j, l_j \rangle$ in $V^t$,

If $\langle b_j, l_j \rangle$ is a new detection not in $H^{t-1}$, create a new hypothesis $\langle b_j, p_j, l_j, f, c \rangle$ and add it into $H^t$

return $H^t$

The algorithm for hypothesis update is outlined in Alg 1. The laser-based detection and tracking module gives candidate objects at current frame with their relative positions and labels. If the object is a known object detected previously, the label is unchanged. However, a new label is assigned for newly detected objects at current frame. And then, the vision-

**Algorithm 2** Lazy Classification

Input
  $H^t = \{h_1, h_2, ..., h_s\}$: current object hypotheses
  $g(\cdot)$: feature evaluation function of hypothesis
  $T_{low}$, $T_{high}$ : thresholds on feature value
For each $h_i$ in $H^t$,
  1. If $g(h_i) \geq T_{high}$, update the class of $h_i$ to be *pedestrian*
  2. If $g(h_i) \leq T_{low}$, update the class of $h_i$ to be *non pedestrian*
  3. If $T_{low} < g(h_i) < T_{high}$, the class of hi is unchanged
return $H^t$

based detection module gives bounding box positions in the image if the laser-based candidates succeed to be detected by HOG detector. The hypotheses that are successfully detected by both laser and vision module are updated with the current detection results. For the hypotheses that are detected by laser module but not detected by vision module, we predict a bounding box position in the current image based on the previous state and current laser-based detection result as

$$b^t = M_h(p^t)/M_h(p^{t-1}) \left(b^{t-1} + M(p^t) - M(p^{t-1})\right). \quad (3)$$

Note that with this prediction step we are able to compensate for temporal failure of vision-based detection significantly. As for the hypotheses that fail to be tracked by laser processing, we simply delete the hypotheses. Finally, we create a new hypothesis as a candidate pedestrian for newly detected object (unmatched detection) and add it into the current hypotheses list.

The described sensor fusion pedestrian detection system increases accuracy and detection rate significantly with reduced runtime, compared to pure vision-based approach. The possible false alarms are further reduced by incorporating lazy decision making.

*E. Evaluation of Object Hypotheses*

The lazy classification is applied (at every frame) for evaluating current list of hypotheses of the detection system as a final step.

The algorithm for lazy classification is outlined in Alg 2. Initially, hypotheses are classified as candidate at the time of their creation as described in the previous section. The evaluation of each hypothesis then is iteratively performed based on a ternary thresholding at every frame. Hypotheses whose evaluated feature value is larger than a high threshold are classified as *pedestrian*. And, hypotheses whose evaluated feature value is lower than a low threshold are classified as *non-pedestrian*. Once a decision is made (pedestrian or non-pedestrian), it is maintained until a contradictory evidence appears. The hypotheses whose evaluated feature values lie between low and high threshold maintain their previous classification.

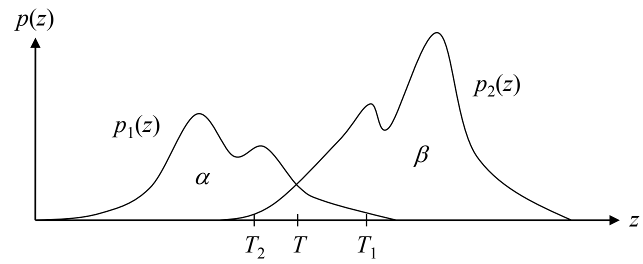



Fig. 4. An example of overlapping probability density functions.

Figure 1 illustrates the proposed lazy decision framework with a comparison with reactive binary decision. In the proposed lazy decision framework (Fig. 1b), decision on uncertain hypotheses ($h_2$ and $h_4$) is postponed until the uncertainty is resolved, avoiding possible misclassifications. However, most hypotheses usually have strong features and the decision from them is made instantly ($h_1$ and $h_3$).

This lazy decision strategy greatly reduces misclassification errors with a slight decrease of detection rate. Let $z$ be a random variable denoting feature values for discriminating two classes (background and object). And let $p_1(z)$ be a probability density function (pdf) of the background and $p_2(z)$ be the probability density function of the target object. The mixture probability density function is then given by

$$p(z) = \alpha p_1(z) + \beta p_2(z), \quad (4)$$

where $\alpha$ and $\beta$ are the prior probabilities of the two classes and $\alpha + \beta = 1$.

If we apply traditional binary thresholding with a threshold $T$, the probabilities of false positives and false negatives are given by

$$E_p(T) = \int_T^{\infty} p_1(z)dz \quad (5)$$

and

$$E_n(T) = \int_{-\infty}^{T} p_2(z)dz, \quad (6)$$

respectively. Then the overall misclassification error is given by

$$E(T) = \alpha E_p(T) + \beta E_n(T). \quad (7)$$

With single thresholding the misclassification error thus always exists for overlapping pdf's and is lower bounded by $E(T^*)$, where $T^*$ is the optimal global threshold.

In our ternary classification, observations below $T_2$ are classified as negative and observations above $T_1$ as positive. Decision on uncertain observations between $T_2$ and $T_1$ is postponed to next iteration of evaluation. Therefore the misclassification error for a single frame decision is given by

$$E(T_1, T_2) = \alpha E_p(T_1) + \beta E_n(T_2). \quad (8)$$

When comparing (7) and (8), we can see that the misclassification error is reduced significantly by using ternary thresholding.

Although the proposed lazy decision reduces misclassification errors significantly, it also may decrease detection rate

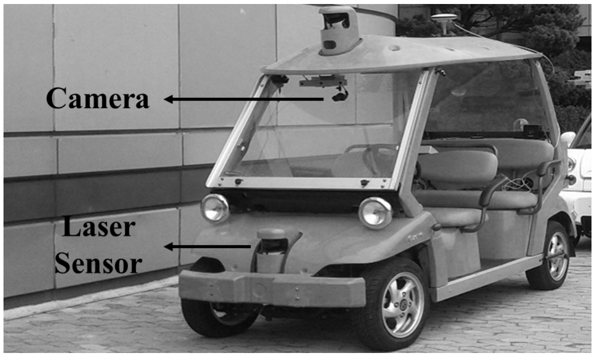



Fig. 5. A test platform and vehicle-mounted camera and laser sensor for the experiment.

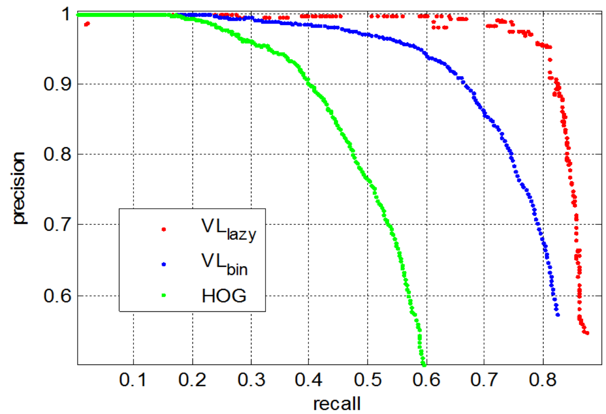



Fig. 6. Comparison of precision-recall curves of three pedestrian detection configurations.

of true target. However, the amount of decrease of detection rate of true target is not large compared to the reduction of misclassification error. Here, we give an intuitive explanation on the reason.

True targets that have strong features ($z \geq T_{high}$) are detected instantly with no delay. For the true targets that have initially weak features ($z < T_{high}$), we can imagine two cases for the reason of initial weak observations. One case is that the target has strong features but temporarily gives weak response due to the interference of background or change of view point. In this case, the target will be detected successfully within few frames. The other case is that the target has basically weak features. In that case the target gives mostly weak response but eventually it will give a response larger than $T_{high}$ sometimes. In second case the target basically is hard to be detected even with binary thresholding but in our lazy decision framework, it will eventually be detected successfully once it shows strong features.

## III. EXPERIMENT

### A. Configuration

In the experiment, we evaluate the performance of pedestrian detection system in three configurations: vision-based (HOG), simple fusion of vision and LRF with reactive binary decision (VL$_{bin}$), and the proposed sensor fusion with lazy decision (VL$_{lazy}$). In the first configuration we use a pure vision-based detector implemented based on HOG feature [2]. The pedestrian detector of the first configuration also was equally used as a sub vision module for the other two sensor fusion configurations. In the second configuration (VL$_{bin}$) the HOG detector is applied for image ROIs extracted from laser scan data and pedestrians are detected reactively at every frame whenever a feature score is larger than a predefined threshold. In the third configuration (VL$_{lazy}$) we keep track of each candidate object as hypothesis and detect pedestrians by evaluating the maintained hypotheses within the proposed lazy decision framework. In VL$_{lazy}$ we evaluate the average feature score over trajectory as well as a single feature score at each frame for the lazy classification.

The performance of each configuration of pedestrian detection system is evaluated by precision-recall. For the evaluation of precision-recall, we adopt the detection-criterion of the VOC challenge [23] given by

$$\lambda = \frac{area(b \cap gt)}{area(b \cup gt)}, \tag{9}$$

where $b$ denotes a detected bounding box and $gt$ the ground truth bounding box. A detection is considered successful (true positive) if it overlaps with a ground truth more than 50% ($i.e.$ $\lambda > 0.5$).

TABLE I

SELECTED PRECISION-RECALL AND RUNTIME PERFORMANCE.

| | Recall | False | Precision | Runtime |
|---|---|---|---|---|
| HOG | 39.9% | 9.9% | 90.1% | 21.3 fps |
| VL$_{bin}$ | 61.5% | 6.5% | 93.5% | 87.3 fps |
| VL$_{lazy}$ | 74.3% | 1.0% | 99.0% | 88.0 fps |

The test data consists of a video sequence and laser scan data of 2,500 frames, which are time-synchronized. The test data was recorded from a vehicle-mounted camera and laser sensor (Fig. 5). The resolution of the test video is 640 by 480 and the laser sensor is SICK LMS151 which is a single channel laser scanner for outdoor. The bounding box positions of pedestrians in the test frames were labeled manually, giving total 3,689 ground truth positions.

### B. Result and Discussion

The precision-recall graph of the three pedestrian detection configurations is presented in Fig. 6 and Table I. We are able to observe that the proposed approach (VL$_{lazy}$) obtains significantly better precision-recall graph compared to typical vision-laser sensor fusion approach (VL$_{bin}$) and vision-only approach (HOG). The proposed VL$_{lazy}$ was able to detect 74.3% of pedestrians from the test data with only 1.0% of false detections (28 false positives over 2,500 frames). The processing time was 88 fps on average.

Figure 7 shows the distribution of average HOG feature score of true positives (blue dots) and false positives (red dots) with respect to trajectory length of the hypotheses

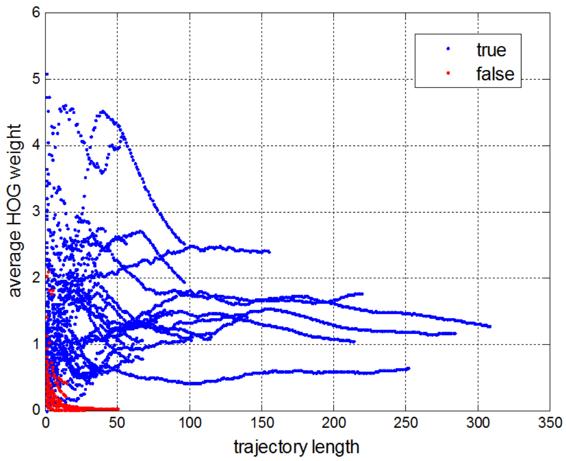



Fig. 7. Average HOG feature score of true positives (blue dots) and false positives (red dots) with respect to trajectory length.

obtained from the test data. The average scores are computed by keeping track of each candidate object (hypothesis) and averaging the feature scores over the trajectory from start to current point. In the figure, we are able to observe that the true positives and false positives are clearly discriminated as the trajectory length increases although they are largely overlapped at the beginning. This suggests that we can discriminate true positives and false positives effectively by using average feature score within the proposed lazy decision framework.

## IV. CONCLUSION

In this paper, we proposed a novel lazy decision approach based on ternary thresholding for robust target object detection. As a practical application of the proposed approach, we presented a sensor fusion pedestrian detection system based on camera image and laser range data. The experimental result confirms effectiveness of the proposed approach, showing significant improvement of precision-recall performance.

The proposed lazy decision approach also can be applied for pure vision-based problems. For example, we are able to apply the approach for detecting target object in videos by combining a detector with a visual tracker.

Future work is to analyze the effect of the proposed lazy decision on the detection performance quantitatively.

## REFERENCES


[1] R. T. Collins, Y. Liu, and M. Leordeanu, ”Online Selection of Discriminative Tracking Features,” IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 27(10):1631–1643, Oct. 2005.

[2] N. Dalal and B. Triggs, ”Histograms of oriented gradients for human detection,” in Proc. of CVPR, CA, USA, 2005.

[3] N. Otsu, ”A threshold selection method from gray-level histograms,” IEEE Trans. on Sys., Man., Cyber., 9 (1): 62–66, 1979.

[4] C. K. Chow and T. Kaneko, ”Automatic boundary detection of the left ventricle from cineangiograms,” Comp. and Biomed. Res., vol. 5, pp. 388–410, 1972.

[5] R. Schnabel, R. Wahl, and R. Klein, ”Efficient RANSAC for Point-Cloud Shape Detection,” Computer Graphics Forum. Vol. 26. No. 2. Blackwell Publishing Ltd, 2007.

[6] F. Fleuret and D. Geman, ”Fast face detection with precise pose estimation,” in Proc. of ICPR, vol. 1, 2002.

[7] J. Koolwaaij and L. Boves, ” Local Normalization and Delayed Decision Making in Speaker Detection and Tracking,” Digital Signal Processing, vol. 10, pp. 113–132, 2000.

[8] J. J. Leonard and R. Rikoski, ”Incorporation of delayed decision making into stochastic mapping,” Experimental Robotics VII, pp. 533–542, 2001.

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, ”Object Detection with Discriminatively Trained Part Based Models,” IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 32(9):1627–1645, 2010.

[10] R. Benenson, M. Mathias, R. Timofte, and L. V. Gool, ”Pedestrian detection at 100 frames per second,” in Proc. of CVPR 2012.

[11] K. C. Fuerstenberg, D. T. Linzmeier, and K. C. J. Dietmayer, ”Pedestrian recognition and tracking of vehicles using a vehicle based multilayer laserscanner,” in Proc. 10th World Congr. Intell. Transp. Syst., Madrid, Spain, Nov. 2003.

[12] S. Gidel, P. Checchin, C. Blanc, T. Chateau, and L. Trassoudaine, ”Pedestrian Detection and Tracking in an Urban Environment Using a Multilayer Laser Scanner,” IEEE Trans. on Intelligent Transportation Systems, 11(3):579–588, Sept. 2010.

[13] C. Mertz et al., ”Moving object detection with laser scanners,” Journal of Field Robotics, 2012.

[14] L. Spinello, R. Triebel, and R. Siegwart, ”Multimodal People Detection and Tracking in Crowded Scenes,” in Proc. 23rd AAAI Conf. on Artificial Intelligence, 2008.

[15] M. Kobilarov and G. Sukhatme, ”People tracking and following with mobile robot using an omnidirectional camera and a laser,” In Proc. IEEE Conf. on Robotics and Automation (ICRA), 2006.

[16] G. Gate, A. Breheret, and F. Nashashibi, ”Fast Pedestrian Detection in Dense Environment with a Laser Scanner and a Camera,” IEEE 69th Vehicular Technology Conference (VTC), 2009.

[17] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, ”Tracking Multiple People using Laser and Vision,” IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2005.

[18] J. Almeida, A. Almeida, and R. Araujo, Tracking multiple moving objects for mobile robotics navigation, in Proceedings of the IEEE International Conference in Emerging Technologies and Factory Automation (ETFA2005), Catania, Italy, September 2005.

[19] N. Gordon, D. Salmond, and A. Smith, Novel approach to nonlinear/non-gaussian bayesian state estimation, in IEEE Proc. F, Radar and Signal Processing, vol. 140, no.2, pp. 107–113, 1993.

[20] M. Isard and A. Blake, Contour tracking by stochastic propagation of conditional density, in Proc. 4th European Conf. on Computer Vision (ECCV), pp. 343–356, April 1996.

[21] H. Badino, U. Franke, and D. Pfeier, ”The Stixel World - A Compact Medium Level Representation of the 3D-World,” 31st DAGM Symposium on Pattern Recognition, 2009.

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

[23] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. Int. J. Comput. Vision, 88(2):303–308, 2009.

[24] H. Park, M. Park, K. Won, K. Kim, and S. Jung, In-Vehicle AR-HUD System to Provide Driving-Safety Information, ETRI journal, vol. 35, no. 6, pp. 1038-1047, 2013.