Human Detection Using Multimodal and Multidimensional Features

Human Detection using Multimodal and Multidimensional Features

Luciano Spinello and Roland Siegwart

ASL - Swiss Federal Institute of Technology Zurich, Switzerland
email: {luciano.spinello, roland.siegwart}@mavt.ethz.ch

***Abstract*— This paper presents a novel human detection method based on a Bayesian fusion approach using laser range data and camera images. Laser range data analysis groups data points with a novel graph cutting method. Therefore, it computes a belief to each cluster based on the evaluation of multidimensional features that describe geometrical properties. A person detection algorithm based on dense overlapping grid of Histograms of Oriented Gradients (HOG) is processed on the image area determined by each laser cluster. The selection of HOG features and laser features is obtained through a learning process based on a cascade of linear Support Vector Machines (SVM). A technique to obtain conditional probabilities from a cascade of SVMs is here proposed in order to combine the two information together. The resulting human detection consists in a rich information that takes into account the distance of the cluster and the confidence level of both detection methods. We demonstrate the performance of this work on real-world data and different environments.**

I. Introduction

According to National Highway Traffic Safety Administration (NHTSA) report [1] there were 4784 pedestrian fatalities in United States during the year 2006, which accounted for 11.6% of the total 42642 traffic related fatalities. In countries of Asia and Europe, the percentage of pedestrian deaths is even higher. Intelligent vehicle systems should have the capability to reduce pedestrian injuries. Human detection is the next logical step after the development of a successful navigation and obstacle avoidance algorithm in urban environment. However humans have been proved to be a difficult object to detect because of the wide variability in the appearance due to clothing, illumination and view point variant shape characteristics. To be supportive to a navigation module we want to detect pedestrians and localize them in 3D at any point in time and as fast as possible. Since we cannot control the vehicle path, nor the environment it passes through, the detector needs to be robust to a large range of lightning variations, noise and partial occlusion. Sensor characteristics reveal that each sensor can only perceive certain characteristics of the environment, therefore a single sensor is not sufficient enough to comprehensively represent the driving environment. A multisensor approach has the potential to yield a higher level of reliability and security. In this paper we present a system which addresses human detection using a laser rangefinder-camera Bayesian sensor fusion approach.

Laser range data contain little information about people, specially because it typically consists of two-dimensional range information. However, range measurements that correspond to humans have certain geometrical properties such as size, convexity etc. This paper extends the key idea of Arras [2] to determine a set of scalar features that quantify pedestrian properties using an AdaBoost learning approach, expanding the feature set and considering multiple dimensional features learned through a boosted cascade of Support Vector Machines (SVM). Neither the selection of features nor their threshold are determined by manual design or hand tuning but they are statistically learned.

Dalal & Triggs [3] and then Zhu & Avidan [4] presented a image based human detection algorithm with excellent detection results and very good performance in terms of execution speed. This method is based on the classification of Histogram of Oriented Gradients computed over blocks of different sizes and scales in the detection window. A classification method based on a rejector-based SVM cascade is proposed to discriminate the presence of a human in the detection window.

In this paper we compute classified geometrical features on range data and HOG features in image data to give a probability estimation of human detection. Namely, laser data analysis (*structure information*) groups the data points and computes a belief to each cluster based on geometrical properties. An image detection (*appearance information*) algorithm based on HOG [3] is then processed on the image area determined by 3D planes, defined by laser clusters with a given height projected on the current image frame. Both range data features and HOG features are classified according to an Adaboost cascade based on linear C-SVMs. In this work it is proposed a strategy to obtain a probability to formulate a Bayesian fusion model.

The novelty of this work consists in:

1) combining together, with a Bayesian sensor fusion model, two of most recent and reliable human detection methods using two different sensors: camera and laser.
2) increasing the feature set of the previous work of Arras [2] for range based human detection, considering 2D features and n-dimensional features (shape factors).
3) the implementation of a novel fast graph-cut based segmentation for range data
4) the development of the method on a moving vehicle considering real-world experiments.

The advantages of using camera and laser sensor fusion principally are:

- direct, precise and instantaneous distance measurement of the detection (due to laser).
- sensors complementary characteristics: a moving person is a complex deformable object and for certain poses it can be described with a high confidence by one sensor or the other due to sensor peculiar characteristics.

II. Previous Work

Several approaches exist in the literature in to detect a person in range data including analysis of local minima [5], geometric rules [6], or maximum-likelihood estimation [7].

Most similar to our work is the approach of [2] which clusters the laser data and learns an AdaBoost classifier from a set of monodimensional geometrical features extracted from the clusters.

In the area of image-based people detection, there mainly exist two kinds of approaches (see [8] for a survey). One uses the analysis of a *detection window* or *templates* [9], [10], the other performs a *parts-based* detection [11]

Combined camera and laser rangefinder pedestrian detection methods use hard constrained approaches or hand tuned thresholding. Cui [12] uses multiple laser scanners at foot height and a monocular camera to obtain people tracking by extracting feet and step candidates. Zivkovic [13] proposes a probabilistic part based approach using camera and laser, principally relying on detection methods suitable for indoor environments.

III. Overview of the method

This sections gives an overview of our pedestrian detection method. The work described in this paper is divided in two phases: training and detection. In the training phase a supervised learning technique, based on cascades of n-dimensional Support Vector Machines, is developed in order to discriminate features that are characteristic of pedestrians in laser rangefinder data and camera images. Then two cascades of SVMs are built: one to classify laser data and the other to classify image data.

In the detection phase, laser data (*structure information*) is clustered according to a novel segmentation method here proposed (Sec IV). A prior probability is assigned to each cluster; each cluster is then projected as 3D plane in the image and evaluated using the trained laser data classifier to obtain a belief. Parallelly, the Histogram of Oriented Gradients (HOG) detector is run on the image-projected laser clusters in order to obtain an image based human detection probability (Sec V). The following boosted support vector machine cascade is explained in Sec VI. A Bayesian sensor fusion approach addresses the problem of fusing the information between the laser detection and image detection conditional probability (Sec VII). Experimental results are shown in Sec VIII.

IV. Structure Information from Laser Data Analysis

We assume that the robot is equipped with a laser range sensor that provides 2D scan points $(\mathbf{x}_1, ..., \mathbf{x}_N)$ in the laser plane. We detect a person in a range scan by first clustering the data and then applying a boosted classifier on the clusters, which we describe as follows.

A. Clustering

Jump distance clustering is a widely used method for 2D laser range data in mobile robotics (see [14] for an overview). It is fast and simple to implement: if the Euclidean distance between two adjacent data points exceeds a given threshold, a new cluster is generated. Although this approach performs well in indoor scenarios, it gives poor results for outdoor data, because the environment is geometrically more complex and bigger distances, reflections and direct sunlight effects usually occur. This often leads to over-segmented data with many small clusters. To address this problem, we use a simple and effective technique that extends the classic jump distance method. It consists in the following steps:

1) Perform jump distance clustering with threshold ϑ. Each cluster $\mathcal{S}_i$ is defined by its left border $\mathbf{x}_i^l$, its central point $\mathbf{x}_i^c$, and its right border $\mathbf{x}_i^r$:
$$\mathcal{S}_i = \left\{\mathbf{x}_i^l, \mathbf{x}_i^c, \mathbf{x}_i^r\right\} \tag{1}$$
2) Compute a Delaunay triangulation on the cluster centers $\mathbf{x}_i^c$.
3) Annotate each edge $\mathbf{e}_{ij} := (\mathbf{x}_i^c, \mathbf{x}_j^c)$ of the Delaunay graph with the Euclidean distance between $\mathcal{S}_i$ and $\mathcal{S}_j$.
4) Remove edges with a distance greater than ϑ and merge each remaining connected component of the graph into a new cluster.

Note that the same threshold ϑ is used twice: first to define the minimum jump distance between the end points of adjacent clusters and then to define the Euclidean distance between clusters. Experimental results showed that this reduces the cluster quantity of $25\% - 60\%$, significantly reducing overclustering. The additional computational cost due to the Delaunay triangulation and distance computation is lower compared to a full 2D agglomerative clustering approach.

B. Features description

We define a feature as a function $f_j : S_i \rightarrow \Re^n$ that takes a cluster S as an argument and returns a n-dimensional value. Multidimensional features and shape descriptors that describe geometrical and statistical properties of the cluster are here considered:

Number of points	Standard deviation
Mean average deviation	Width
Height	Linearity
Circularity	Radius
Boundary length	Mean curvature
Mean angular difference	Kurtosis
PCA based shape factors	N-binned histogram
PHG	Boundary regularity

The distance between clusters is not considered as a feature because we aim to achieve a detection just based on the characteristics of the cluster. 2D features are created adding

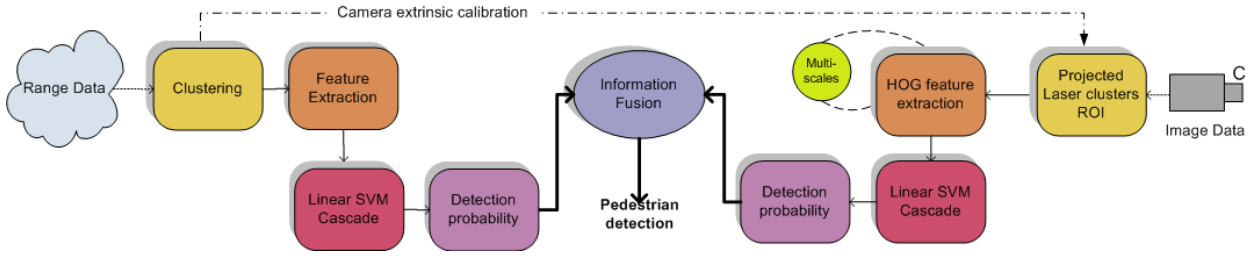

Fig. 1. An overview of the proposed pedestrian detection method.

cluster-observer distance to 1D features as another dimension in order to learn how the feature value changes with respect to the distance. The feature set is composed in total by 50 features.

V. Appearance information: image data analysis

The image based human detection is based on a method that relies on the classification Histogram of Oriented Gradients (HOG) features in a detection window (Dalal [3]). This section gives an overview of the feature extraction method. Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. The image window is divided into *cells*. For each *cell* a local histogram of gradient directions over the pixels of the cell is accumulated. For better invariance to illumination, it is also useful to contrast-normalize the local responses before using them. This can be done by grouping the *cells* in *blocks* and normalizing the included cells histograms. The Dalal & Triggs algorithm uses an overlapping fixed scale block tessellation with the use of fairly small block size (16 pixels) which might miss the "big picture" of the entire detection window. To overcome this limitation and to accelerate the detection process, we implemented the method proposed by Zhu & Avidan [15]. Multiple scales blocks, in different locations and aspect ratios, are added to enlarge the feature set and capture more information in the detection window. The ratio between block width and block height can be any of the following ratios $(1:1)$, $(1:2)$ and $(2:1)$ and we consider all blocks whose size ranges from 12×12 to 64×128 using an increasing step of $\{4, 6, 8\}$. Each histogram is a vector of $36D$ that is the concatenation of 9 orientation bins in the included 2×2 cells. 5245 HOG features are present in each window.

To support a fast evaluation of specific blocks we use the integral image representation to efficiently compute HOG of each block. Porikli [9] suggested the *Integral Histogram* to efficiently compute histograms over arbitrary rectangular image regions. Inspired by the work of Porikli [16] and Viola [17], it is possible to quickly compute an HOG feature. We discretize each pixels orientation into 9 histogram bins and store an integral image for each bin of the HOG (resulting in 9 images in our case) and use them to compute efficiently the HOG for any rectangular image region. A L_1 (Manhattan distance) histogram normalization is executed in each HOG feature block.

VI. Boosted Support Vector Machine Cascade

Boosting is a general method for creating an accurate strong classifier by combining a set of weak classifiers. The requirement to each weak classifier is that its accuracy is better than a random guessing. The adaboost algorithm introduced by Freund and Schapire [18] was extended by Viola [17] who introduced the attentional cascade. This method radically reduces the computation time: the key insight is that smaller, and therefore more efficient, boosted classifiers can be constructed, which reject many of the negative samples while detecting almost all positive instances. The overall form of the detection process is that of a degenerate decision tree (or cascade). A positive result from the first classifier triggers the evaluation of a second classifier which is adjusted to yield very high detection rates and so on. A negative outcome at any point leads to the immediate rejection. It is important to note that stages in the cascade are constructed by training classifiers and adjusting the threshold to minimize false positives.

Because most of laser features and HOG features lie in a $\Re^1 - \Re^{36}$ space, we implemented as weak classifier the separating hyperplane computed using a C-SVM with linear kernel [19].

The training set is constituted by a set of n-samples of m-kind of features f_i. A sampling method is used in the case of classification of HOG features to reduce the training time: 5% of the features is randomly sampled and evaluated each round. All the features of the set are considered in the laser feature classification case.

A. Parameters selection

SVM parameter selection is an important issue to consider. In C-SVM only parameter C has to be chosen:

$$\min_{w,b,\xi} = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

subject to:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \tag{2}$$

$$\xi_i \geq 0, i = 1, \ldots, l. \tag{3}$$

The parameter C is chosen using the value of C that minimizes the cost function in the v-fold cross validation:

$$\Gamma = a_1 \varepsilon_p + a_2 \varepsilon_n \tag{4}$$

where ε_p and ε_n in equation 4 express respectively the false positive rate and the false negative rate. This cost function differs from the classic approach to maximize an accuracy based cost function. Due the structure of the boosted cascade it is important to have many *hard examples* in order to have sufficient negative samples each time a stage is completed and a new one is trained. If we run the parameter selection on accuracy, the resulting SVM will be naturally biased towards a high occurrence of false negatives.

B. Cascade Probability Estimation

We denote the detection of a person using a binary random variable π that is true whenever a person is detected. Each of the L cascaded SVM-classifiers h_i yields either 1 or 0 for a given input feature vector $\mathbf{f}$. The overall detection probability can then be formulated as

$$p(\pi \mid \mathbf{f}) = \sum_{i=1}^{L} w_i h_i(\mathbf{f}) \tag{5}$$

In the learning phase, the weights w_i and the hyperplanes are computed for each SVM classifier h_i. The probability is evaluated when all the stages of the cascade are successfully passed.

VII. Information fusion

This section explains the method used to estimate a pedestrian detection using the structure and appearance information.

Camera is intrinsically and then extrinsically calibrated with respect to laser rangefinder using the method described by Pless [20].

Each segmented cluster in laser data is projected into image frame as a 3D plane: the enlarged cluster width ($1.5m$) is used to define the width of a plane surface with constant height of $3m$, the projection of the extremal points of this surface create a region of interest in the image frame, and namely define the prior of a pedestrian at that image location:

$$p(\pi) \approx p(\pi|r) \tag{6}$$

Equation 6 modulates the uncertainty relating it to the distance r from the observer. This is a reasonable assumption because the quantity of information decreases with the distance: a far away pedestrian is described by few points in the range data and few pixel in the image data.

Each cluster S_i is evaluated using the classifier trained on laser data that describes $p(\pi|\theta_l)$, that is the probability of human detection given laser data analysis.

The image data classifier is run on the regions of interest defined by each laser cluster. The HOG image detector detects classified features at multiple scales and locations inside that image space. This significantly reduces the image detection process due to a reduced research space. The HOG classifier describes the probability of pedestrian detection given image data analysis: $p(\pi|\theta_c)$

Structure and appearance information are considered cues of same importance, thus the same confidence level of detection should be given in the fusion information process. The information fusion is addressed using a Bayesian modeling approach.

Starting from the joint distribution and applying recursively the conjunction rule we obtain the decomposition:

$$p(\pi \wedge \theta_l \wedge \theta_c) = p(\pi)\ p(\theta_l|\pi)\ p(\theta_c|\pi) \tag{7}$$

In equation 7, the phenomenon ϕ is considered to be the main reason for the contingency of the structure and appearance information, thus knowing the cause ϕ of the readings the variables θ_l and θ_c are independent. In general, this hypothesis is not always satisfied, but it is often used in literature and it has the main advantage of considerably reducing the complexity of the computation.

The conditional probability defining the information fusion is:

$$p(\pi|\theta_l \wedge \theta_c) = \frac{p(\pi)\ p(\theta_l|\pi)\ p(\theta_c|\pi)}{\sum_\pi (p(\pi)\ p(\theta_l|\pi)\ p(\theta_c|\pi))} \tag{8}$$

VIII. Experimental Results

A. Training datasets

We trained our HOG features classifier using the well-established MIT pedestrian image database and the significantly more challenging INRIA person database [21]. To build the negative set a software has been developed to randomly crop part of images containing street and urban background from the INRIA negative dataset. The set contain in total 3123 $64x128$ positive images and 12313 $64x128$ negative images of people. The people are usually standing but appear in any orientation, against a wide variety of background including crowds. The cascade consists in 26 levels and the first three levels contains just 4 to 6 SVM classifiers each and reject circa 81% of the detection windows; this permits a fast execution time and good performances.

Laser datasets have been taken in two different outdoor scenarios: a parking lot and the university campus. The parking lot dataset consists of a staged "road like" scenario: some people pass in front of the car, a person is "shape changing" wearing a hat and eventually the car is run in an internal road with parked cars and crossing pedestrians. The university campus dataset presents a very challenging and cluttered environment with a lot of passing pedestrians with different shapes, speed and distribution in the space.

The range scan data is segmented using the algorithm explained in Sec. IV and the entire featureset is computed. Thus the clusters are manually labeled. The parking lot data set is composed by 1070 positive and 7054 negative clusters. The campus dataset is composed by by 497 positive and 3499 negative clusters

B. Experiments

The mobile platform Smartter, used to acquire the datasets, is based on a Daimler-Chrysler Smart vehicle equipped with several active and passive sensors, a camera with a wide field of view lens and a frontal laser rangefinder. An accurate camera-laser synchronization has been developed for this work. Each dataset is divided in a training set and a test set. The campus training set is composed by a random choice

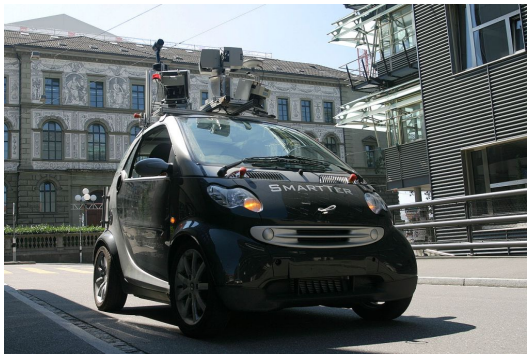

Fig. 2. Smartter platform. A camera has been placed behind the windscreen and the AlascaXT laserscanner has been mounted on the front of the car.

of 248 positive samples and 1700 negative samples from the original dataset (50%), the same method is applied to the parking lot dataset. The testing sets are composed by the remaining clusters. The range features classifier training dataset is the union of both training datasets (750 positive samples and 5250 negative samples). The resulting cascade is composed by 4 stages with a total of 8 features. The resulting selected features of the cascade that describe a human in range data are the following: *[Standard deviation, cluster distance] (2D feature)*, *[Width]*, *[Kurtosis]*, *[Radius, cluster distance] (2D feature)*, *[Histogram] (24D feature)*, *[Boundary regularity]*, *[Boundary Length, cluster distance] (2D feature)*. The resulting selected features are a balanced combination of shape description and points distribution statistics often related to the distance of the cluster from the observer. In order to study the importance of the distance between two clusters in an outdoor scenario, we added this component as another feature to the featureset. The automatic process of building the boosted cascade excluded this features due to a smaller accuracy with respect to the others. This results differ from the work of Arras [2] mainly due to a very different training set and a different segmentation method.

Range data features classifier for the parking lot dataset obtains TP:517(91.1%) FN:51(8.9%) FP:351(10.0%) TN:3153(90.0%); HOG image features cascade classifier obtains TP:197(91.53%) FN:26(8.57%) FP:137(7.62%) TN:1648(92.38%). The overall classification rate for both sensors is very high (over 90%) and false positive rate/false negative rate is low and comparable (under 10% for both). Even though the environment resembles a road scenario, from a ranged and a visual point of view, the persons silhouette and their range data remain well defined. It's noticeable that in some hard examples one sensor gives better results than the other. A pedestrian can receive a high confidence using range data features but can be rejected using the image feature classifier and viceversa. This condition occours mainly when a pedestrian is defined with few range data points or with an occluded or unconventional silhouette pose.

The confusion matrix of ranged features for the university campus dataset is composed by TP:161(64.7%) FN: 88(35.3%) FP:536(30.0%) TN:1273(70.0%); HOG cascade classifier obtains TP:119(72.6%) FN: 45(27.4%) FP:173(21.9%) TN:613(78.1%). The overall classification rate is 65% for the first and 73% for the other. This result can be explained by the complexity of the environment and, moreover, by the clutter present in this scenario: multiple pedestrians in a small space "distort" the cluster shape and point distribution; visually cluttered pedestrian contain less informative gradient information for the HOG detector. If we anyway take the maximum of the output probability of both cascade the detection rate is increased of 16%. Instead of having a "hard" method of combining the results of both

Fig. 3. Two images from parking lot dataset. Brighter rectangles indicate a high detection proability. On the left: multiple pedestrian detection is shown with different level of detection probability. On the right: a car is detected as a pedestrian with very low confidence due to a false positive in one of the detectors. Pedestrian distances are written in red.

Fig. 4. Two images from parking lot dataset. Brighter rectangles indicate a high detection proability. On the left: the far away pedestrian receive a smaller vote with respect to the foreground person. On the right: a classic pedestrian crossing in simulated. Pedestrian distances are written in red.

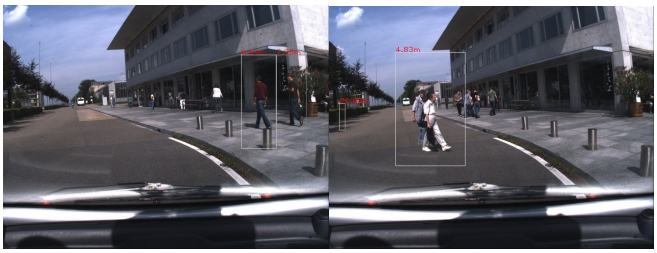

Fig. 5. Two images from university campus dataset. Brighter rectangles indicate a high detection proability. The visual and geometrical complexity of the environment decreases the overall performance of the detector. On the left: two persons are detected and the others discarded due to poor light conditions, shadows, glass reflections and generally visual and range data occlusions in the laser line of sight. On the right: foreground persons are equally well detected; the cluttered group of people on the right is ignored.

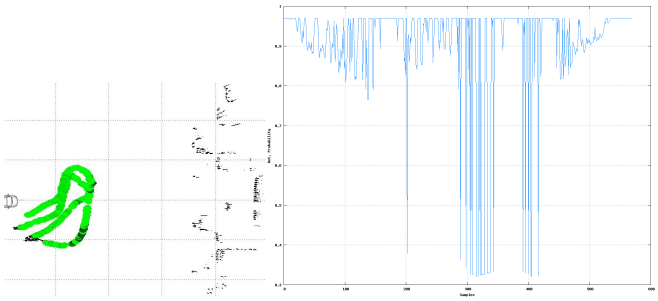

Fig. 6. The figure shows the probability progress of a pedestrian walking in front of the car. The left figure traces the path: brighter green circles depict a high probability value. On the right figure the probability value (*axis Y*) of the left figure is explicitly plotted with respect to the samples (*axis X*).

classifiers we present the method based on the Bayesian sensor fusion approach. In fig. 3, fig. 4, 5 each detection (in the range data and in the image) is labeled with its fused probability, far away clusters have a lower probability due the modulation factor of the prior present in eq. 8. This creates a reasonable dynamic traversability map for an autonomous car, taking into account the confidence of each sensor and acting consequently. If one sensor detection fails to detect a person (false negative) the result is a confidence decreasing but the affected cluster will still receive another probability measure from the other sensor. In order to show the validity of the method we depict in fig. 6 the progress of the probability estimate in time when a pedestrian is walking in front of the car. The brightness of the cluster is proportional to the detection level. It's important to notice that the detection works also when the person is not present in the image but it is estimated only using range data.

The computation time required to obtain a detection depends on the number of clusters found in the laser range data scan; considering both datasets a frame rate between $3fps$ to $15fps$ is obtained. A video of the experiments is available at *http://www.asl.ethz.ch/people/sluciano/videoICRA08.mpg*

IX. Conclusions

We presented in this paper a novel human detection method that combines camera and laserscan information. A Bayesian fusion is used to fuse together two of the most recent and reliable human detection methods. One of the key points of this paper is the convergence of two different methods with different sensors to obtain a more informative detector. Even though person detection is far from being solved individually by each sensor, we have shown that the proposed sensor fusion can increase the overall detection confidence especially in hard examples. The obtained information from sensor fusion is a rich detection that takes in account the distance and the confidence level of each detector. In future work we plan to expand this work including tracking of people using dynamic models and extending the image detection part to better handle occlusions and clutters.

X. Acknowledgments

This work was conducted and funded within the EU Integrated Projects BACS - FP6-IST-027140

References

[1] "2006 traffic safety annual assessment – a preview," Traffic Safety Facts, National Center for Statistics and Analysis, July 2007, http://www-nrd.nhtsa.dot.gov/Pubs/810791.PDF.

[2] K. O. Arras, scar Martnez Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *IEEE International Conference on Robotics and Automation (ICRA07)*, 2007.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 886–893.

[4] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 1491–1498.

[5] Scheutz, Mcraven, and Cserey, "Fast, reliable, adaptive, bimodal people tracking for indoor environments," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*.

[6] J. Xavier, M. Pacheco, D. Castro, A. Ruano, and U. Nunes, "Fast line, arc/circle and leg detection from laser scan data in a player driver," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2005, pp. 3930–3935.

[7] D. Hähnel, R. Triebel, W. Burgard, and S. Thrun, "Map building with mobile robots in dynamic environments," in *"Proc. of the International Conference on Robotics and Automation(ICRA)"*, 2003.

[8] D. M. Gavrila, "The visual analysis of human movement: A survey," *Computer Vision and Image Understanding: CVIU*, vol. 73, no. 1, pp. 82–98, 1999.

[9] D. Gavrila and V. Philomin, "Real-time object detection for "smart" vehicles," in *International Conference In Computer Vision (ICCV)*, 1999, pp. 87–93.

[10] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003, p. 734.

[11] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes." Washington, DC, USA: IEEE Computer Society, 2005, pp. 878–885.

[12] J. Cui, H. Zha, H. Zhao, and Shibasaki, "Tracking multiple people using laser and vision," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2005, pp. 2116–2121.

[13] Z.Zivkovic and B. Krse, "Unifying perspectives in computational and robot vision," *D.Kragic and V.Kyrki, editors, Springer*, 2007.

[14] C. Premebida and U. Nunes, "Segmentation and geometric primitives extraction from 2d laser range data for mobile robot applications," in *Robotica 2005 - Scientific meeting of the 5th National RoboticsFestival*, April 2005.

[15] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 1491–1498.

[16] Porikli, "Integral histogram: A fast way to extract higtograms in cartesian spaces," in *CVPR '05: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2005.

[17] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, 2002.

[18] R. E. Schapire and Y. Singer, "Improved boosting using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.

[19] B. E. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Computational Learing Theory*, 1992, pp. 144–152.

[20] R. Pless and Q. Zhang, "Extrinsic calibration of a camera and laser range finder," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2004.

[21] "Inria person database," http://pascal.inrialpes.fr/data/human/.