# HOG Feature Human Detection System

Matt Davis
Dept. of Electrical and Microelectronic Engineering
Rochester Institute of Technology
Rochester, NY 14623, USA
msd8744@rit.edu

Ferat Sahin
Dept. of Electrical and Microelectronic Engineering
Rochester Institute of Technology
Rochester, NY 14623, USA
feseee@rit.edu

**Human detection systems are becoming more important as more automatic and robotic systems are being used in the world. RGB, depth, and thermal images can be used together to produce a better detection system that works in situations where one of the sensors might not produce valuable data. HOG features can provide valuable information for detecting humans in an image and that data can be used to train individual classifiers to detect humans in a scene. The combination of sensor modalities in conjunction with individual classifiers can create a human detection system that can detect partially obscured humans. The multi-layer classifier that was created provided a high level of accuracy when tested against untrained data. The multi-layer classifier performed better than eleven of the twelve individual classifiers, but did not overcome the SVM thermal HOG classifier. The multi-layer classifier had a much tighter standard deviation and fell within the band of the SVM thermal classifier.**

## I. Introduction

Robots are following a path similar to computers where they are increasingly becoming part of daily life. Just as computers arose in research labs and spread to every home and industry, robots seem to be following a similar trajectory. Robots are able to complete tasks that humans might not want to or be able to complete. Today, robots are used to manage warehouses full of goods, clean houses, pilot autonomous vehicles, and construct objects. In the future, robots will be used even more to help humans with their daily lives. Robots can incorporate machine learning in order to improve on tasks set for them or to better predict actions to take based on the presented choices.

Robots can also be useful for rescue operations. A robot can enter into dangerous areas that would be hazardous for a human to go. Once there, autonomous robots can search the area for points of interest, survivors, or general reconnaissance. The ability for a robotic system to detect humans in a video or image is extremely important for robots interacting with humans and especially important for rescue operations. A system needs to be able to accurately determine if a human is in a frame even if the body is partially obscured or oriented in an unusual position. A rescue robot that has the ability to find a human whilst only seeing a portion of a body not obscured by debris will greatly increase the speed of rescue operations and reduce danger to first responders.

The chances of survival for endangered individuals greatly increases when they can be located quickly, removed from the dangerous situation, and provided medical assistance. An efficient human detection system will allow first responders to find endangered victims that might otherwise be overlooked in the early stages of a rescue operation.

Autonomous rescue operations can be improved with the combination of RGB-D, and thermal imaging camera (TIC) images. Given the RGB, depth, and thermal image data, a partially obscured human detection system can be created by using feature descriptors such as Histograms of Oriented Gradients (HOG) and Histograms of Oriented Depths (HOD) along with various supervised machine learning classification algorithms. A human detection system used in conjunction with Robot Operating System (ROS) will allow a robot to identify humans within view and respond to their presence.

## II. Related Work

### A. Human Detection

Autonomous human detection is difficult because of many different variables that need to be accounted for [1]. The orientation of a person, the type of image data used as an input, the amount and type of light that is captured can all vary drastically leading to a very complex problem. Another major factor in the difficulty of detecting humans is if the human is partially obscured by another object [2], [3]. Dalal and Triggs [4] focused on detecting the entire body while others such as Xia et al. [1] focused on detecting a portion of the body. The addition of obstructed portions of a human creates an even more complicated problem because of the non-uniform nature of incomplete humans. This represents another variable in addition to changing light conditions, variation in the way people look, distance, and point of view. Partially obscured humans are particularly difficult to detect because color based segmentation doesn't work [5].

### B. RGB, Depth, and Thermal Images

A typical RGB-D camera consists of an infrared (IR) projector to emit an IR dot pattern, an IR camera to record the reflections of the projected IR, and a color camera [6]. The resolution of the IR data is greatly affected by the distance from the IR camera as well as the surface's ability to absorb IR [7]. The IR projector and IR camera produce a depth map showing the distance from each reflected IR dot to the camera [6]. Thermal images show the temperature distribution of the environment with each pixel corresponding to a certain temperature value [8]. The RGB-D camera presents a cheap

978-1-5090-1897-0/16/$31.00 ©2016 IEEE

solution for acquiring range data as well as RGB images. Until thermal cameras become more prevalent, laser sensors to track distances are the second most used sensor for human detection after cameras. A combination of RGB-D and a thermal sensor showcases the fusion of two easily obtainable sensors with a slightly rarer sensor.

### C. HOG and HOD

Histograms of Oriented Gradients (HOG) is one of the more popular methods used today in human detection applications [7]. A detection window slides across an image frame wherein a grid of cells is created. The gradients of the pixels in each cell are then used to create a histogram of edge orientations [2]. Local objects can be seen identified as a distribution of local intensity gradients [4]. The cells are inside of larger blocks which are used to overcome illumination variations. The final product of the HOG is a feature vector which consists of the all of the feature descriptors in the image. The larger the amount of bins, the more detail the histogram will contain. A balance must be made between the amount of desired detail and the size of the feature vectors.

Histograms of Oriented Depths (HOD) are very similar, but the depth values from the depth image are used instead of color values [7]. The idea can be extrapolated to the use of HOG on temperature data produced from thermal imagers. A HOG of temperature gradients will produce feature vectors that can be useful for human detection. With the ability to threshold a range of temperatures in the temperature data, the boundary between objects with the desired temperature and the surrounding area should be very sharp.

### D. Feature descriptors and machine learning

Hoang, Vavilin, and Jo [9] created a fast human detector using low computational cost Haar-like features. Haar-like features generally don't provide good performance with regards to human detection because of a large number of variables in images with humans. The AdaBoost algorithm was used to train a classifier by combining responses from weak classifiers. A cascade classifier is used to create a much faster system because the negative samples are discarded in the early layers. The results showed that this system is faster than HOG features and more accurate than normal Haar-like features. AdaBoost significantly decreased the computational time required. Arras et al. [10] were able to use AdaBoost to train an effective classifier using simple features of groups corresponding to people's legs using a range sensor. AdaBoost uses weak learners and linear regression to eventually converge to a strong learner.

The four basic concepts of support vector machines (SVM) involve the separating hyperplane, the maximum-margin hyperplane, the soft-margin, and the kernel function [11]. The SVM attempts to maximize the separation between two classes. A nonlinear kernel can be used to transform input features into high-dimensional feature space to separate two classes that might appear intermingled. SVM can also provide good classification results when there is noisy data.

Naïve Bayes (NB) classifiers assume that all of the features of a class are conditionally independent given the class label [12]. The assumption is usually false, but the model works well in practice even when the assumption is not valid. The training data is first used to train with the above assumption. In the case of this experiment a Gaussian distribution is used. Then the untrained test data is used to predict the probability of each sample belonging to each class based on the largest posterior probability.

The k-Nearest Neighbors algorithm (k-NN) assigns an unclassified sample to the class of its nearest neighbor that has been classified already [13]. A downside of the k-NN algorithm is if the class distribution is too skewed. If there are too many of a certain class, there is an increased probability that a sample will get classified as the most common neighbor.

### E. Sensor Fusion

Dima, Vandapel, and Herbert [14] tackled the problem of obstacle detection for autonomous off-road navigation by combining color and IR imagery with range information to create a better classifier. Five types of features were measured in each patch on the image grid. The color images were used to create two features, and the laser range data was used to create two features, and the IR data created the final feature. Three different algorithms were used to combine classifiers: committees of experts, stacked generalization, and AdaBoost. The results showed that both low-level data fusion and classifier fusion significantly improved classification performance over individual classifiers.

Kong et al. [15] sought to fuse visible and thermal IR images for a robust face recognition system. The addition of IR data was beneficial as it would not be affected by illumination changes whereas the visible images would be susceptible. The visible and IR images were fused as pairs in the discrete wavelet transform domain. The combined use of visible and IR image sensors provided an improvement over using single imaging modalities. The sensor fusion enhanced the ability to recognize a face in different lighting conditions with the addition of IR images.

Premebida, Ludwig, Matsuura, and Nunes [16] explored fusing LIDAR data with image data using three different fusion methods as well as a cascade classifier. The image data was composed of HOG features and covariance matrices descriptors (COV). The fusion methods used were classifiers likelihood outputs combined using the average rule, maximum rule, and a Naïve Bayes based rule. The cascade of classifiers consisted of four feature spaces, two from LIDAR data and two from visible. The fused classifiers did not produce much better results than the single classifiers with regards to the ROC curve and the cascade classifier results were unimpressive for the statistics measured.

Spinello and Arras [7] created a new way to detect humans in images with the HOD detector. The authors developed a human detector using depth data instead of color data and fused the HOG and HOD detectors together. The HOG and HOD descriptors were classified then fused using a learned SVM. The Combo-HOD detector had better accuracy than single detector schemes and was much faster in detecting humans.

Choi, Pantofaru, and Savarese [17] combine five observation models: HOG, shape from depth data, front face detection, skin color, and motion detection. The image data is gathered with an RGB-D camera and a reversible-jump Markov Chain Monte Carlo (RJ-MCMC) algorithm is used to detect people in a frame. The results show that the combination of different detection cues provides more reliable results and the advantages of different cues can overcome the disadvantages of other cues.

Susperregi et al. [18] combine a low-cost thermal sensor with an RGB-D camera to produce a better human detection system. Twenty-four different computer vision transformations are performed on each sensor dataset to produce different image descriptors. Five different supervised machine learning techniques are used to train classifiers. Finally, the best three classifiers are combined in a hierarchical classifier. The people detection accuracy was improved over single classifiers while also decreasing the FPR.

## III. Methodology

### A. Data Collection

The RGB and depth images were recorded using an Asus Xtion Pro Live camera. The camera recorded at 30 frames per second with a resolution of 640 x 480 pixels. The thermal images were recorded using an Omega OSXL-101 thermal imager. The thermal camera recorded at 1 frame per second with a resolution of 445 x 476 pixels. The two cameras were locked together to restrict movement and create a system with as little viewing angle difference as possible. This ensured that the frames would only need slight adjustment in the pre-processing phase of the experiment. The camera setup is shown in Fig 1.

Fig. 1. RGB-D camera and thermal imager

The RGB-D camera was run using the Robot Operating System (ROS) on a laptop running Ubuntu 12.04 LTS Linux. Python scripts were used to manipulate ROS and produce the image data. The raw depth images and raw RGB images were published by the ROS system and saved as individual PNG files. The saved images would be used to create the datasets for training, validation, and testing. The thermal imager was run using software supplied by Omega on a desktop running Windows 7. The thermal images were saved as individual JPG files. The saved thermal images would also be used to create datasets for training, validation, and testing.

A subject was positioned 2.5 meters from the mounted camera system. A known dataset of thermal, RGB, and depth images of people obscured by objects was not found and a custom dataset had to be generated for this experiment. In order to create a dataset of images of obscured humans a large, square piece of insulation foam was used to form a shield. Handles were added to the foam to make a less cumbersome shield. The insulation foam prevents the RGB-D camera from seeing objects behind the board as well as prevents the thermal imager from detecting temperatures from behind the shield. The subject maneuvered the shield in different positions to obscure different parts of the body during the data collection process. A database was created with images containing various parts of a human and images containing no human or a human blocked completely by the insulation board. The subject holding the shield is shown in Fig. 2.

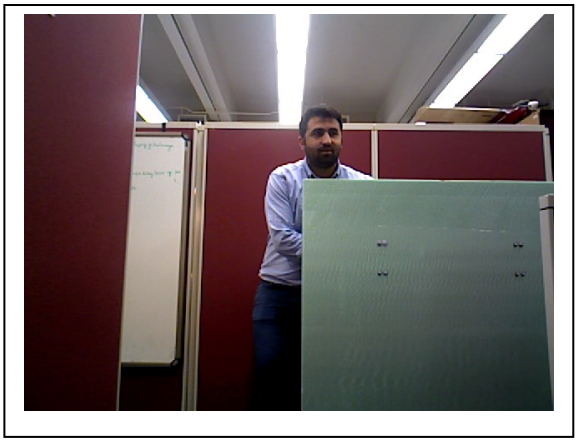

Fig. 2. RGB image of subject using shield to produce obstructed image data

A database of thousands of images was created to facilitate in the training of classifiers. The collected data was separated into separate as to whether a human was visible in the thermal, RGB, or depth image. Sets of images were chosen as positive and negative samples and labels were created for each image to use for further testing.

### B. Pre-processing

MATLAB was used to program all of the necessary functions required for dataset control, pre-processing, feature extraction, classifier training and construction, and metrics generation. Before the features were able to be extracted, a pre-processing phase was required. Each depth image was read into MATLAB and stored as a 480x640 array with each element representing a pixel in grayscale. Each RGB image was read into MATLAB and stored as a 480x640x3 array and converted into a grayscale array. Each thermal image was read into MATLAB and stored as a 376x445 array. The thermal images had to be cropped to remove the temperature scale produced by the Omega thermal imager capture software. The thermal images also needed to be mirrored to reproduce the original

image. All three image types were resized into a uniform 240x320 array.

The thermal imaging software allowed a temperature threshold to be applied and a threshold of 80 to 100 degrees Fahrenheit was used to isolate the human subject. This threshold was chosen to focus on a temperature range produced by a living human. The RGB images were converted to grayscale to reduce the amount of features extracted and speed up extraction speed. The RGB, depth, and thermal images were resized to a standard resolution of 320x240 to prepare for feature extraction. The three pre-processed images are shown in Fig. 3.

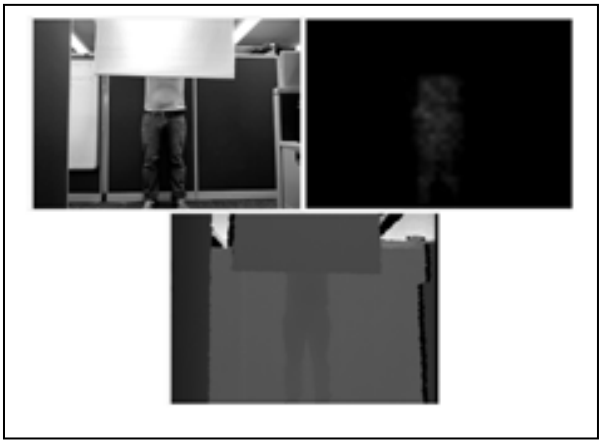

Fig. 3. Pre-processed RGB, thermal, and depth image ( CW from top left )

### C. Feature Extraction

HOG features were extracted from each image and a column of numbers was appended to the feature vectors to denote whether a human was present or not in the image. A balance of pertinent features extracted size of the feature vectors was found with a cell block size of 16 x 16 and number of bins in each cell of 7. This reduced the feature vector size greatly while losing minimal detail. The HOG features can be visualized using rose plots. Each plot shows the distribution of gradient orientations within each HOG cell. The length shows the contribution each orientation makes. The edge directions are normal to gradient directions.

After the HOG features were extracted the data was separated into three different groups to create a training set, validation set, and robust testing set. The robust set was first separated from the other two sets to provide a test set using untrained data. During each run, the remaining data was randomly shuffled and split in half to create the training set and validation set to train the classifiers with. After each shuffle, train, and validation, the robust set was used to test the classifier using untrained data.

After the HOG features were extracted from the thermal images, Principal Component Analysis (PCA) was used to reduce the feature set further. The unreduced feature vectors were first shuffled to create the training and validation sets for each run. Afterwards, PCA was performed on the training set to reduce the feature vector size and maintain 99.9% of the principal components that explain the variation of the data. The transformation matrix generated from the PCA was then used to reduce the dimensionality of the validation and robust test datasets. The reduced, extracted HOG features were then ready for model training and classification.

### D. Model Learning and Classification

Four different machine learning algorithms were used to train the individual classifiers using the feature descriptors. Each learning algorithm was used on the three different image types for a total of twelve classifiers. A binary SVM scheme was used to classify the images and determine if a human was in the image. Both linear and radial basis function (RBF) kernels were used to train the SVM classifier. A second classifier was trained using k-NN. A third classifier was trained using Naïve Bayes with a Gaussian kernel. Finally, an ensemble of classifiers was trained using AdaBoostM1 boosting algorithm and a decision tree weak learner.

After the twelve classifiers were created and trained, multiple runs were completed to generate statistics and confusion matrices for each classifier. True positives (TP) were designated as people detected in a positive sample, true negatives (TN) were designated as no people detected in a negative sample. False positives (FP) were designated as people being detected in a negative sample, and false negatives (FN) were designated as people not being detected in a positive sample.

After ideal parameters were chosen for each classifier using cross-validation, ten runs were completed while shuffling the data-sets to determine the robust accuracy for each individual classifier. The results were used to determine which classifier worked best with each image type. After the best results were chosen a series of multi-layer classifiers were created using a decision tree. Twelve basic forms were chosen to test for the decision tree. Each decision tree structure uses the three image types at least once. The algorithm starts at the upper node and determines if a human is present or not in the image type. If there is a human, the algorithm moves to the right and down one level. If there is not a human, the algorithm moves to the left and down one level. An example is shown below in Fig. 4.

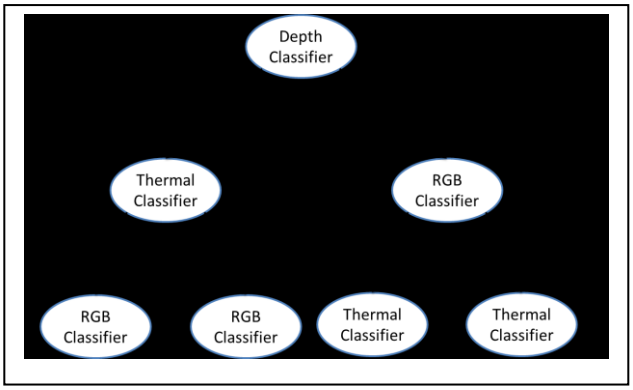


Fig. 4. Example multi-level classifier construction algorithm

Once the final multi-layer classifier was chosen, several more robust accuracy tests were performed. Metrics were then generated and compared to the different individual classifiers.

## IV. Results

### A. Individual Classifiers

The twelve individual classifiers were tested over multiple runs using extracted features from shuffled training sets. The robust accuracy results for the RGB HOG individual classifiers are shown in Table I.

TABLE I. RGB HOG individual classifier results

| | SVM | NB | k-NN | AdaBoost |
|---|---|---|---|---|
| **TP (%)** | 86.36 | 76.36 | 87.27 | **90.91** |
| **FP (%)** | 8.33 | 10.00 | 0.00 | 8.33 |
| **TN (%)** | 91.67 | 90.00 | **100.00** | 91.67 |
| **FN (%)** | 13.64 | 23.64 | 12.73 | 9.09 |
| **Robust (%)** | 88.24 | 81.18 | **91.76** | 91.18 |

The best overall accuracy for RGB images was obtained using the k-NN classifier with an accuracy of 91.76%. The k-NN classifier also had the best accuracy for images without a human present as seen by the 100% true negative rate. The ensemble classifier with AdaBoost performed almost as good with an overall accuracy of 91.18%, but with the best accuracy for detecting humans in images with humans present. Because the ensemble classifier did almost as good as k-NN for overall accuracy, but did much better for true positive rates, it was also tested in the multi-level classifier stage. The robust accuracy results for the depth HOG individual classifiers are shown in Table II.

TABLE II. Depth HOG individual classifier results

| | SVM | NB | k-NN | AdaBoost |
|---|---|---|---|---|
| **TP (%)** | **81.81** | 76.36 | **81.81** | 80.91 |
| **FP (%)** | 0.00 | 0.00 | 0.00 | 0.00 |
| **TN (%)** | **100.00** | **100.00** | **100.00** | **100.00** |
| **FN (%)** | 18.18 | 23.64 | 18.18 | 19.09 |
| **Robust (%)** | **88.24** | 84.71 | **88.24** | 87.65 |

The best overall accuracy for depth images was obtained using the SVM classifier and the k-NN classifier with an accuracy of 88.24%. Both classifiers had the same true positive rate of 81.81%. All four classifiers correctly determined that there was not a person present in the negative images 100% of the time. Both k-NN and SVM classifiers were used in the multi-level classifier stage. The robust accuracy results for the thermal HOG individual classifiers are shown in Table III.

TABLE III. Thermal HOG individual classifier results

| | SVM | NB | k-NN | AdaBoost |
|---|---|---|---|---|
| **TP (%)** | 92.73 | **99.09** | 90.91 | **100.00** |
| **FP (%)** | 0.00 | 18.33 | 0.00 | **100.00** |
| **TN (%)** | **100.00** | 81.67 | **100.00** | 0.00 |
| **FN (%)** | 7.27 | 0.91 | 9.09 | 0.00 |
| **Robust (%)** | **95.29** | 92.94 | 94.12 | 64.71 |

The best overall accuracy for thermal images was obtained using the SVM with an accuracy of 95.29%. The SVM classifier had a true positive rate of 92.73% and a true negative rate of 100%. The ensemble classifier failed in this test and only predicted that a human was present in every image. The Naïve Bayes classifier produced a true positive rate of 99.09%, but an overall accuracy of 92.94%. The k-NN classifier also did very well with an overall accuracy of 94.12% and a true positive rate of 90.91%.

### B. Multi-layer Classifiers

The twelve multi-layer classifiers were tested over multiple runs using extracted features from shuffled training sets. The best performing individual classifiers were used for each node and tested against each other. The best multi-layer classifier algorithm is shown in Fig. 5.

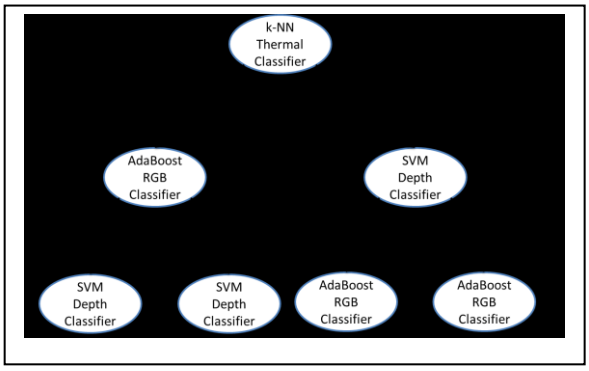


Fig. 5. Best performing multi-layer classifier algorithm

The best performing multi-layer classifier produces a robust accuracy of 92.35% with a standard deviation of 1.61%. This compares to the thermal SVM classifier at 95.29% with a standard deviation of 4.46%. Both the k-NN thermal root node and SVM thermal root node produced similar results. The robust accuracy results from the best individual classifiers for each image type and the multi-layer classifier are shown in Table IV.

TABLE IV. Multi-layer classifier results

| | SVM-T | SVM-D | k-NN-RGB | Multi-Layer |
|---|---|---|---|---|
| **Precision (%)** | 100.00 | 100.00 | 100.00 | 100.00 |
| **Recall (%)** | 92.73 | 81.82 | 87.27 | 88.18 |
| **Robust (%)** | 95.29 | 88.24 | 91.76 | 92.35 |

## V. Discussion and Future Work

This paper presented a novel method to identify partially obscured humans by combining features detected in RGB, depth, and thermal images. The HOG features extracted from the three image types provided a very useful set of data to detect obscured humans. The multi-layer classifier that was created provided a high level of accuracy when tested against untrained data. The multi-layer classifier performed better than eleven of the twelve individual classifiers, but did not overcome the SVM thermal HOG classifier. The multi-layer classifier had a much tighter standard deviation and fell within the band of the SVM thermal classifier.

The SVM thermal classifier would potentially do worse in a different experimental set up. In the data collection stage, the temperature was constant and did not fluctuate. If the temperature had been warmer and closer to the temperature of a human body, the thresholding would not have been as successful and more noise would have been introduced into the thermal images. This would have potentially decreased the robust accuracy of the SVM thermal classifier and the multi-layer classifier may have performed better than the SVM thermal classifier.

Another factor which may have reduced the accuracy of the SVM thermal classifier is if larger distances from the camera to the person were used. This would have a similar effect to increasing the ambient temperature in that the temperature would read a lower value.

In the future, some changes should be made to the data collection stage. Different distances, ambient temperatures, and locations should be used for gathering data. Several different subjects should be used as well. Gathering some samples from outside would be interesting to see how the environment affects the sensors. Several different feature extraction methods beyond HOG would be a worthwhile examination.

Non-uniform blocking material such as foam debris instead of a single foam board along with different body parts buried would provide a dataset with several disconnected body parts of the same person. Multiple people in a frame while both obscured would be the next step.

Finally, optimizing and increasing the feature extraction time would be very beneficial for real-time applications. The large feature vectors took far too long to process for quick detection of endangered humans.

## ACKNOWLEDGMENT

The authors are thankful to Shitij Kumar and Ryan Bowen for their support. The authors are also thankful to Celal Savur for helping to collect the dataset.

## REFERENCES

[1] Xia, L., Chen, C. C., & Aggarwal, J. K. (2011, June). Human detection using depth information by kinect. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on (pp. 15-22). IEEE.

[2] Choi, B., Meriçli, C., Biswas, J., & Veloso, M. (2013, May). Fast human detection for indoor mobile robots using depth images. In Robotics and Automation (ICRA), 2013 IEEE International Conference on (pp. 1108-1113). IEEE.

[3] Hoang, V. D., Hernandez, D. C., & Jo, K. H. (2014). Partially Obscured Human Detection Based on Component Detectors Using Multiple Feature Descriptors. In Intelligent Computing Theory (pp. 338-344). Springer International Publishing.

[4] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.

[5] Roy, S., & Chattopadhyay, T. (2014). View-Invariant Human Detection from RGB-D Data of Kinect Using Continuous Hidden Markov Model. In Human-Computer Interaction. Advanced Interaction Modalities and Techniques (pp. 325-336). Springer International Publishing.

[6] Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. Cybernetics, IEEE Transactions on, 43(5), 1318-1334.

[7] Spinello, L., & Arras, K. O. (2011, September). People detection in rgb-d data. In Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on (pp. 3838-3843). IEEE.

[8] Susperregi, L., Martínez-Otzeta, J. M., Ansuategui, A., Ibarguren, A., & Sierra, B. (2013). RGB-D, laser and thermal sensor fusion for people following in a mobile robot. Int. J. Adv. Robot. Syst.

[9] Hoang, V. D., Vavilin, A., & Jo, K. H. (2012, October). Fast human detection based on parallelogram haar-like features. In IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society (pp. 4220-4225). IEEE.

[10] Arras, K. O., Mozos, Ó. M., & Burgard, W. (2007, April). Using boosted features for the detection of people in 2d range data. In Robotics and Automation, 2007 IEEE International Conference on (pp. 3402-3407). IEEE.

[11] Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[12] Murphy, K. P. (2006). Naive bayes classifiers. University of British Columbia.

[13] Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. Information Theory, IEEE Transactions on, 13(1), 21-27.

[14] Dima, C. S., Vandapel, N., & Hebert, M. (2004, April). Classifier fusion for outdoor obstacle detection. In Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on (Vol. 1, pp. 665-671). IEEE.

[15] Kong, S. G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B. R., Koschan, A., ... & Abidi, M. A. (2007). Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. International Journal of Computer Vision, 71(2), 215-233.

[16] Premebida, C., Ludwig, O., Matsuura, J., & Nunes, U. (2009). Exploring sensor fusion schemes for pedestrian detection in urban scenarios. Proceeding of the IEEE/RSJ Worshop on: Safe Navigation in Open and Dynamic Environments, held at the IROS2009.

[17] Choi, W., Pantofaru, C., & Savarese, S. (2011, November). Detecting and tracking people using an rgb-d camera via multiple detector fusion. In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on (pp. 1076-1083). IEEE.

[18] Susperregi, L., Sierra, B., Castrillón, M., Lorenzo, J., Martínez-Otzeta, J. M., & Lazkano, E. (2013). On the use of a low-cost thermal sensor to improve kinect people detection in a mobile robot. Sensors, 13(11), 14687-14713.