Proceedings of 2013 IEEE
International Conference on Mechatronics and Automation
August 4 - 7, Takamatsu, Japan

Human Detection and Following by a Mobile Robot using 3D Features

Badar Ali, Khawaja Fahad Iqbal, Yasar Ayaz and Naveed Muhammad

***Abstract*—Human-robot interaction is one of the most basic requirements for service robots. In order to provide the desired service, these robots are required to detect and track human beings in the environment. This paper presents a novel approach for classifying a target person in a crowded environment. The system used the approaches for human detection and following by implementing the multi-sensor data fusion technique using stereo camera and laser range finder (LRF). Our system tracks human being by gathering features of human upper body and face in 3D space from stereo camera and uses laser rangefinder to get legs data. Using these data our system classifies the target person from other human beings in the environment. We used Haar cascade classifiers for the detection of upper body and face, and used stereo camera for getting dimensions in 3D space. The approach for gathering legs data is based on the recognition of legs pattern extracted from laser scan. Tracking of target person is done using Cam Shift theorem. Using all these techniques we present a novel approach for target person classification and tracking. Our approach is feasible for mobile robots with an identical device arrangement.**

***Index Terms*—Service robots, mobile robots, body detection, face detection, leg detection, Haar cascade classifiers.**

I. Introduction

THE detection and tracking of human being is a problem of very high importance in Robotics. The applications include surveillance, intelligent transportation, human robot interaction and many more. A great deal of work has been done in this area and various methods have been applied to reach solution. Generally laser range finder and cameras have been the most common devices used for human detection. Some of the earlier methods used only laser data for human detection but it only provides 2D information. Modern approaches use both laser and camera fusion techniques. Songmin Jia et al. [1] present a human detection method based on extraction of human model from disparity image by stereo camera. Further improvement in the robustness and real-time performance of their algorithm has been presented in [2]. Mehrez Kristou et al. [3] Present an approach that uses omnidirectional camera and laser range finder to detect human being. Their approach uses human's cloth pattern to identify the target person and the process consists of two stages: registration stage and identification and localization stage. They further improved their approach by merging the registration stage in human tracking phase [4]. Nicola Bellotto and Huosheng Hu used laser range finder to detect human legs and PTZ camera to find faces [5]. They used Unscented Kalman filter approach to perform real-time tracking. Orasa Patsadu et al. [6] present another approach for recognition of human gesture using Kinect camera.

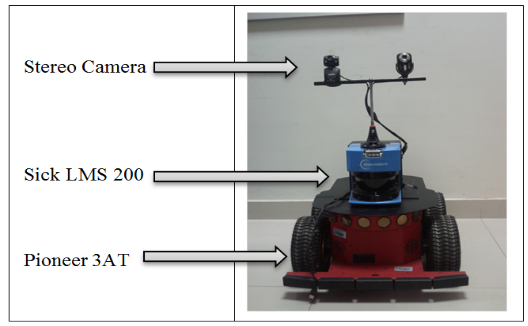


Fig. 1 System Components

In this paper, we present an approach for human detection using multi sensor data fusion technique. Our system use Stereo Camera to get the 3D information of human being and Laser Range Finder to get human legs data. The information from stereo camera includes the calculation of width and height of upper body and face. LRF will give the width of legs in plane. So the system uses this information to classify the target person from other people in the environment. Novelty in our approach is to get more features of target person in term of his body, face and legs data so that the classification between target person and other humans become easy.

This paper is organized as follows. Section II introduces the overall components of our system. Section III describes briefly the related data. Section IV explains human detection and tracking approach in detail. Section V presents the implementation and analyzes the results. Section VI gives a discussion to robustness and accuracy of the system. Finally, conclusions and future work are illustrated in Section VII.

B. Ali, K. F. Iqbal, Y. Ayaz and N. Muhammad are with the Robotics and Intelligent Systems Engineering (RISE) Laboratory, Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), Main Campus, Sector H-12, Islamabad, 44000, Pakistan.
E-mails:
B. Ali: badarali@live.com
K. F. Iqbal: fahadiqbal@smme.nust.edu.pk
Y. Ayaz: yasar@smme.nust.edu.pk
N. Muhammad: naveed.muhammad@smme.nust.edu.pk

II. SYSTEM COMPONENTS

Our system, as shown in fig.1 consists of a mobile robot Pioneer 3AT, Sick LMS 200 laser rangefinder and stereo camera. In support to this hardware, we have majorly used two libraries: Player and Open CV. We have further used MATLAB to calibrate the stereo system [23].

III. RELATED WORK

In past, a lot of different methods have been used for human detection. In [1], Hu moment based method is used to detect human being. In [3] Panoramic vision based approach have been used. Another method of people identification based on color histogram comparison is used in [5].

In our work, we have used object based detection by using Haar Cascade Classifier. This approach has been presented in [11-12] and has been used for the detection of Human Face and Eyes in [14].

Haar Cascade classifier is composed of several simple classifiers which are applied one after the other to a region in an image. These basic classifiers may use one of the four boosting techniques. These boosting techniques are Discrete Adaboost [19], Real Adaboost [20], Gentle Adaboost [21], and Logitboost [22]. In order to make use, these classifiers are first trained with several hundred examples of the object and then the result can be applied to any region of interest to detect that particular object. More details about these classifiers have been explained in [12].

Papers [7-10] describe approaches for detecting legs. In [9] circle detection algorithm has been presented which further has been used to detect legs after adding distance constraint to it. This approach requires both legs to be exact circle with inscribed angle between 90^0 and 135^0. Hence legs detection using this approach is difficult to implement. The one which we have used is based on looking for one of the three common legs patterns in a laser scan. Laser scan is initially stored in a vector of distance and angle, and edges are extracted from these scan. Then sequence of edges is searched in the vector of edges. All the three legs pattern have separate sequence of edges and the total length of the object regarded to as leg also lie in particular range.

In our system, we have made a stereo camera by using two USB cameras. In order to use them as stereo camera, we were required to perform stereo calibration. Hence we have used MATLAB Camera Calibration toolbox [23] to perform stereo calibration by taking 30 samples of Checkerboard images by each camera. At first two individual calibration files are generated by separately calibrating the two cameras using the standard procedure described in [24]. Then stereo calibration toolbox is used to perform stereo calibration [25]. This toolbox will give the intrinsic parameters of each camera and extrinsic parameter of one camera with respect to other.

More information on Camera Calibration Toolbox can be found at [23].

Triangulation is a technique used for finding the 3D location of a point in space based on its position in left and right stereo camera. Various triangulation methods have been presented in [15]. Commonly used methods are Linear Triangulation, Iterative Linear Methods and Mid-point Method. In our system, we have used Linear Triangulation because of its simplicity.

Bradski [16] proposed a method to track objects using mean shift. This method has been implemented in Open CV. At first, initial histogram is calculated which identify the region being tracked. Then this initial histogram is applied to a new frame of input video using back-projection to get a grayscale image containing the size of that color histogram in the new frame for the corresponding pixels. Then this back-projection is searched to get the region with highest intensity that associates the region where tracked object most probably lies.

This method has been used by Atsushi Yamashita et al. in [17] for face tracking.

Cam-Shift (Continuously Adaptive Mean-SHIFT) method which we have used, finds the object center using simple mean-shift. It then adjusts the window size and finds optimal rotation. The cam-shift function in Open CV [18] returns the rotated rectangle structure that includes the object position, size, and orientation.

IV. HUMAN DETECTION AND TRACKING

This part describes human detection and tracking in detail. At first overview of the method is presented. Then Legs and human detection are focused and how they are used to classify the target person. Finally the target person tracking is explained.

A. *Overview of Method*

Following flowchart provides the overview of method;

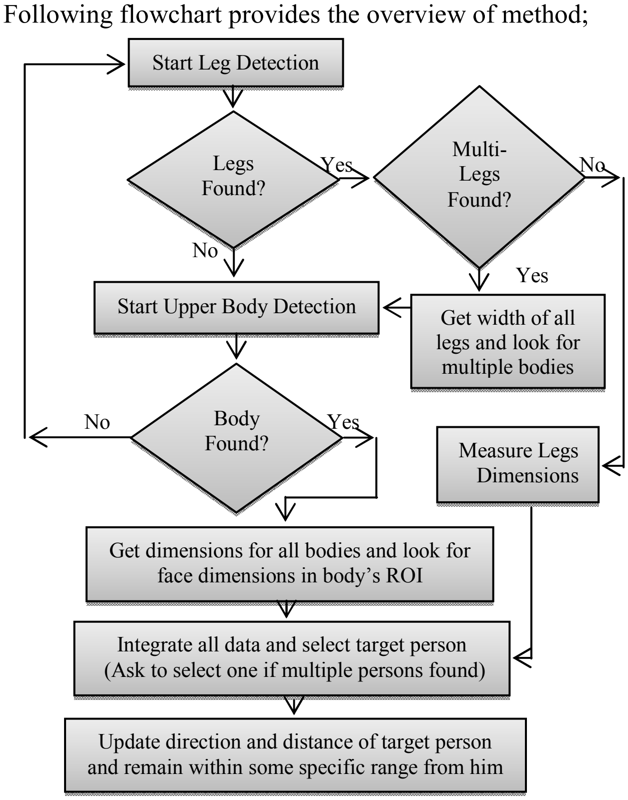


Fig. 2 Overview of Method

In our system, we have implemented our approach by using Haar Cascade Classifier [11] [12] for Upper body and face detection [14] and Legs pattern based Legs detection [10]. Our system uses these methods to implement an application of human detection and following. Our system initially looks for the legs in the environment. In case of multiple bodies, the system will store the width and position of all the legs found. If no legs are found, the system still looks for bodies. When bodies are detected, then it calculates the body width and height based on stereo vision. Then the system looks for face inside the body's region and if found, it will calculate the face's width and height. When multiple bodies are found, it will calculate the features for all bodies and then will ask to choose the target person. Hence it will record all the features for target person and discard the features of other human beings. Then the system will calculate the position of target person by getting the polar coordinates of distance 'r' and angle 'θ' from Robot. Then it will start following the human being by maintaining a distance '$r \pm r_{th}$' and angle '$\theta \pm \theta_{th}$' between Robot and Target Person.

Next parts of paper will describe the individual methods in details.

B. Laser based Leg Detection

In [7-10] different approaches for laser based leg detection has been described. In [9], algorithm for circle detection has been presented which is further used to detect legs by adding distance constraint in circle detection algorithm. Papers [7] and [10] identify legs by looking for typical legs pattern in laser scan.

The approach which we have used extracts the necessary legs features from a single laser scan by identifying the typical legs patterns. These patterns are shown in Fig. 3. Typically there can be three patterns: two legs apart (LA), forward straddle (FS) and two legs together (SL). LA is usually when the person is standing in front of LRF, FS is usually when the person is walking and SL is the one which covers all other patterns like person standing sideward or person standing with closed legs.

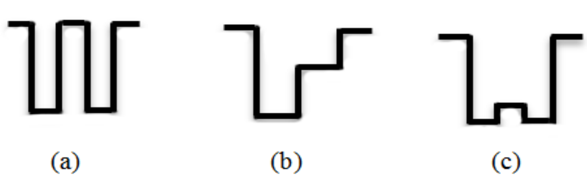


Fig. 3: Typical Legs patterns (a) Legs apart, (b) Forward Straddle, (c) Single Leg

Let the angular step size of laser scan is constant, the laser data is stored in two vectors **R** and **θ**.

Where $\mathbf{R} = \{r_1, r_2 \dots r_i \dots, r_n\}$ is the range at each sample, $\boldsymbol{\theta} = \{\theta_1, \theta_2 \dots \theta_i \dots, \theta_n\}$ is the angle of each sample and n is the total no of samples.

So from the above data, we can identify the sequence of edges. The point is detected as an edge if Euclidean distance between r_i and r_{i+1} is greater than some threshold. Furthermore there can be two types of edges: Left Edge and Right Edge. When $r_i > r_{i+1}$, then it is a Left Edge $\mathbf{L_i}$ and if $r_i < r_{i+1}$, then it is a Right Edge $\mathbf{R_i}$. We store the edges data in a vector **E**, where $\mathbf{E} = \{e_1, e_2 \dots e_i \dots, e_m\}$ where e_i is the i^{th} edge and m is the total no of edges found in a scan. e_i can either be **L** or **R** depend upon the left or right edge.

In order to extract one of the three legs pattern, we look for the ordered **L/R** sequence of consecutive edges. The above mentioned three patterns will have the following sequence of edges;

1. LA: {L, R, L, R}
2. FS: {L, L, R} or {L, R, R}
3. SL: {L, R}

Hence we extract these sequences of edges from **E** and if any of the sequences is within 25cm to 55cm, then it is corresponded as a leg. Every edge is removed from **E** when it is detected as a part of any of above sequence. LA patterns are most reliable ones and hence are searched first. Whereas SL patterns are easiest to get false and thus are searched at the end.

Fig.4 shows briefly the flowchart of leg detection method.

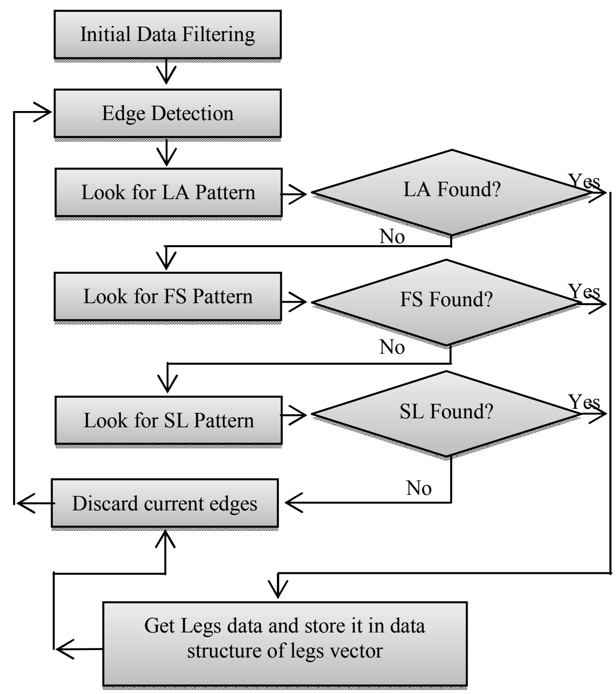


Fig. 4 Leg Detection Algorithm

We stored the legs data in a vector of data structure **Legs** where the leg pattern's widths, leg type, start and end position (with respect to robot's origin) of all the legs found are stored.

C. Vision based Human Detection

This portion consists of feature detection based on Haar Cascade Classifier and Triangulation based on stereo vision.

I. Haar Feature-based Cascade Classifier

The feature based object detection was initially proposed by Paul Viola in [11] and improved by Rainer Lienhart in [12]. Papers [13] and [14] also presented Face and Eyes detection approach.

According to this method, a classifier (cascade of boosted classifiers using haar-like features) is trained with various samples of positive and negative examples. After training, the classifier can be applied to a region of an input image. To search for the particular object in an image, the classifier window is moved across the whole image to check every location. The classifier then gives the output relative to whether it has detected object or not.

Open CV provides different classes for feature based object detection. In our approach we have used trained sets of cascade classifiers (.xml files) to detect upper body and face over a stream of stereo camera. The approaches used in these classifiers are presented in [14]. When the objects are detected, resulting data gives the starting pixel point of the object's region and width and height of the detected region. These data has been stored in a vector of rectangles in the control program.

Flow chart in Fig. 5 briefly sums up the human detection procedure.

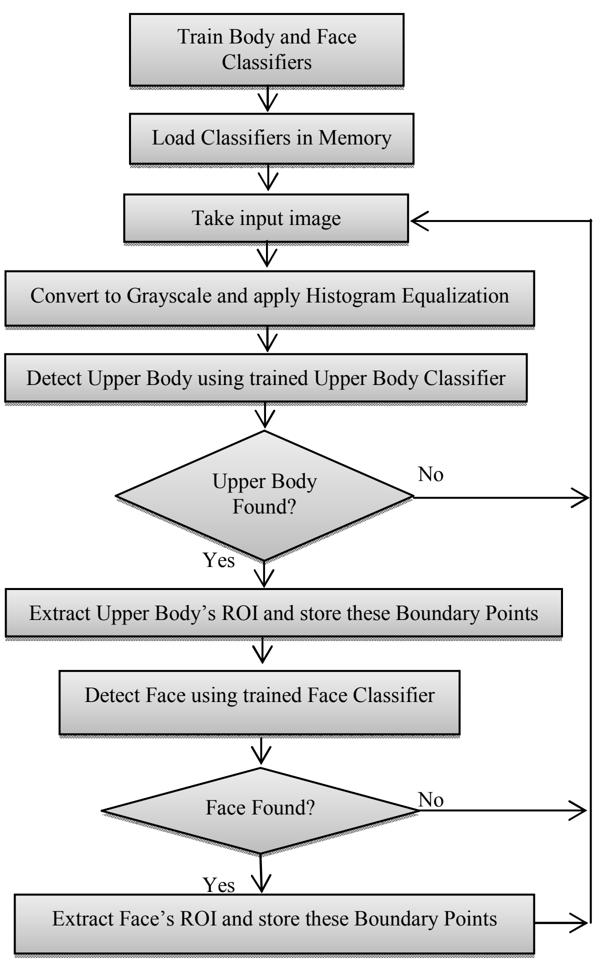


Fig. 5 Flow chart for Human Detection

II. Triangulation

This part of our system will find out the exact location of detected human body's boundaries with respect to robot's origin. Various methods of triangulation have been proposed by Richard I. Hartley and Peter Sturm in [15]. These methods include Linear Triangulation, Iterative Linear Methods and Mid-point Method.

In our system we have used Linear Triangulation to figure out the exact location of human body end points detected by left and right cameras. Fig.6 shows the schematic diagram of triangulation.

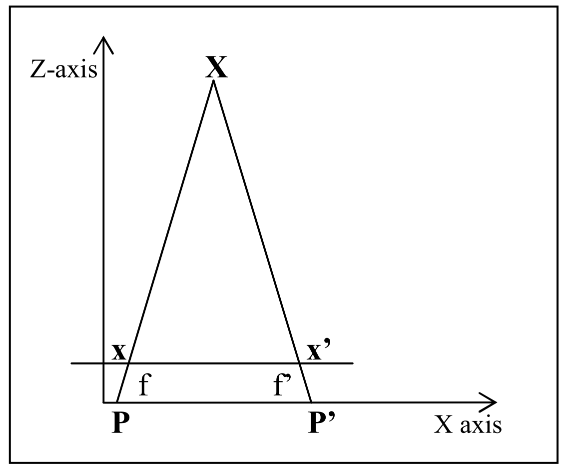


Fig. 6 Triangulation

From fig 6, **P** and **P'** are the 3x4 camera projection matrix for both cameras and **x** and **x'** are the points which we are looking for in corresponding left and right camera respectively. **X** is the point in 3D space, where **X** is of size 3x1. Hence by using Linear Triangulation technique, we can find **X** by knowing **x**, **x**' and camera projection matrices.

The triangulation results will give us the exact boundary points where the human body and face has been detected by vision system. Hence we will further use these points to extract the features which will classify the target person from other persons.

D. *Classification of Target Person*

Using the 3D points of detected human body's contour from triangulation results of stereo camera, we find the body's width and height by finding the Euclidean distance between these points. Horizontal distance will give the width of human body and vertical distance will give the height of detected upper body. Similarly width and height of face is also calculated. We also have the data of width of legs scanned by laser range finder. Hence using these body features from stereo camera and legs features detected from laser we classify the target person from other persons in the environment. The robot keeps the value of these features stored till it is following that particular target.

E. Target Person Tracking

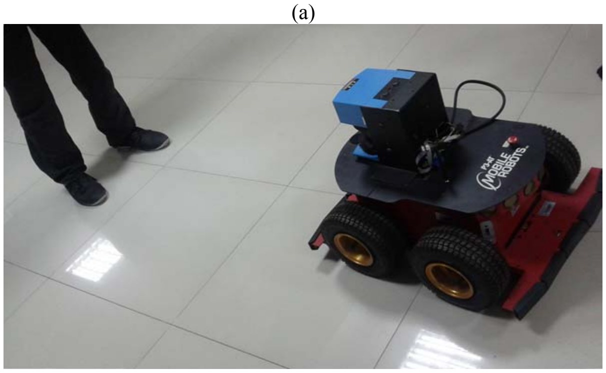
After having the above cited features, the system will start tracking the target person. This tracking is based on Cam-Shift tracker. Once cam-shift tracker is given the coordinates of human body, it will keep track of it in video stream and keep on returning the target person's position in space with respect to Robot (origin). Hence the robot is having the position of target person continuously in each frame and thus it will start following the human being.

In our system, we simply navigated the robot to follow the target person by maintaining a distance of 2.5m and an angle of $\pm 10^0$. Thus we have made an assumption that the current environment is obstacle free and hence in such environment, we only need to detect and then navigate towards target person rather than robot planning its motion. However, real systems are not obstacle free. Hence this portion of our system needs improvement in future.

V. Implementation and Results

As explained earlier, the system has been implemented on Pioneer 3AT having Sick LMS 200 and calibrated stereo camera. The detection results from legs detection by laser rangefinder are presented in fig. 7.

(a)



(b)

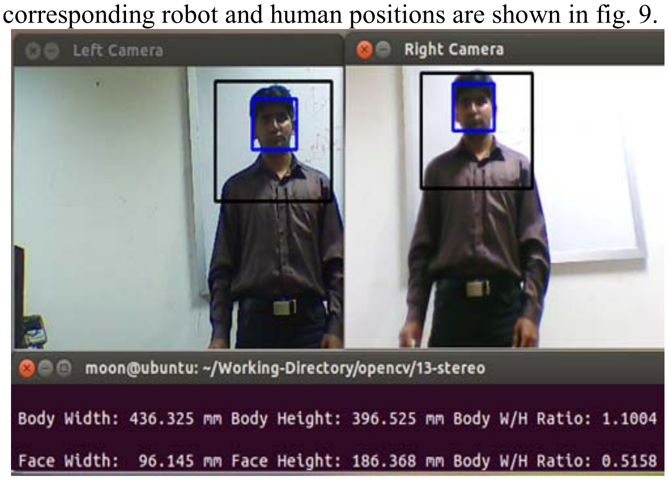
Fig. 7: Laser Leg Detection
(a) Laser Data from Player Viewer and Terminal
(b) Corresponding Robot and Human Position

The results of Body and Face Detection with corresponding human position are shown in fig. 8. Blue and Black rectangles in figure shows the boundaries of detected upper body and face.

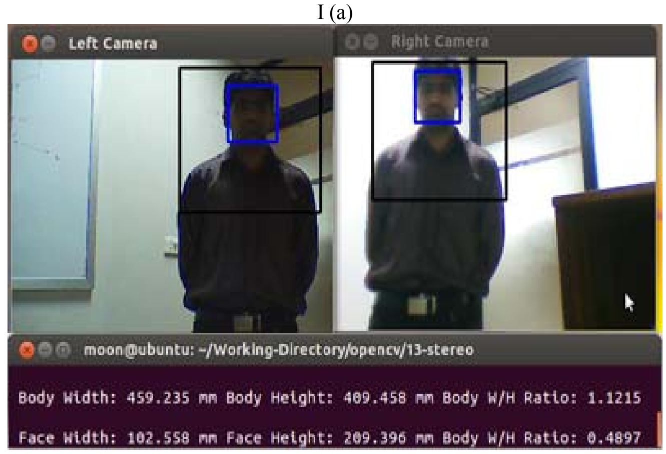
(a) (b)

Fig. 8: Body and Face Detection
(a) Output image (b) Corresponding Robot and Human Position

The results of left and right camera image with corresponding robot and human positions are shown in fig. 9.



I (a)



II (a)


I (b)


II (b)

Fig. 9: Human Detection and Following
I (a): Left & Right Image 1   I (b): Corresponding Full image for I (a)
II (a): Left & Right Image 2   II (b): Corresponding Full image for II (a)

VI. Robustness and Accuracy

The overall system is robust in a sense that it gathers many features of human being. System tracks the target person using camshaft tracker and if the target person is lost, the system has many parameters that can help in finding the target person. However, the system can only classify the human beings facing towards robot. Hence the detection process needs improvement in this portion so that the human being can be detected from sideward as well.

Moreover the accuracy of the system in the terms of camera features lies in the accuracy of calibration results. It also depends upon the quality of camera used. Hence the system will require high resolution camera with high frame rate so that the results will be more accurate.

VII. Conclusion and Future Work

In this paper we have presented an approach for human detection and tracking by extracting human's upper body, face and legs features. Vision related features are body's width and height and face's width and height and are extracted by calculation of detected body/face boundaries in 3D space w.r.t. robot. Legs features are extracted using laser range finder data. Detection results have been tested and presented.

In future we will try to test the detection results on high resolution camera having much improved frame rate (Bumble bee or stereo PTZ camera). We will also try to implement the proposed approach in an environment with more human beings. Furthermore so far we have not focused towards motion planning area. In future we can also look towards implementing an adaptive motion planning algorithm.

References

[1] Songmin Jia, Liang Zhao, Xiuzhi Li, Wei Cui, Jinbo Sheng, "Autonomous Robot Human Detecting and Tracking Based on Stereo Vision", *Proceedings of the 2011 IEEE International Conference on Mechatronics and Automation,* August 7 – 10, 2011.

[2] Songmin Jia, Liang Zhao, Xiuzhi Li, "Robustness Improvement of Human Detecting and Tracking for Mobile Robot", *Proceedings of 2012 IEEE International Conference on Mechatronics and Automation,* August 5 – 8, 2012.

[3] Mehrez Kristou, Akihisa Ohya and Shin'ichi Yuta, "Panoramic Vision and LRF Sensor Fusion Based Human Identification and Tracking for Autonomous Luggage Cart", *18th IEEE International Symposium on Robot and Human Interactive Communication,* Sept. 27-Oct. 2, 2009.

[4] Mehrez Kristou, Akihisa Ohya and Shin'ichi Yuta, "Target person identification and following based on omnidirectional camera and LRF data fusion", *20th IEEE International Symposium on Robot and Human Interactive Communication,* July 31 - August 3, 2011.

[5] Nicola Bellotto and Huosheng Hu, "People Tracking and Identification with a Mobile Robot", *Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation,* August 5 - 8, 2007.

[6] Orasa Patsadu, Chakarida Nukoolkit and Bunthit Watanapa, "Human Gesture Recognition Using Kinect Camera", *9th International Joint Conference on Computer Science and Software Engineering*, 2012.

[7] Nicola Bellotto and Huosheng Hu, "Vision and Laser Data Fusion for Tracking People with a Mobile Robot", *Proceedings of the 2006 IEEE International Conference on Robotics and Biomimetics,* December 17 - 20, 2006.

[8] Nicola Bellotto and Huosheng Hu, "Multisensor Integration for Human-Robot Interaction", *IEEE Journal of Intelligent Cybernetic Systems*, 2005

[9] João Xavier, Marco Pacheco, Daniel Castro, António Ruano and Urbano Nunes, "Fast Line, Arc/Circle and Leg Detection from Laser Scan Data in a Player Driver", *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2005.

[10] Nicola Bellotto and Huosheng Hu, "Multisensor-Based Human Detection and Tracking for Mobile Service Robots", *IEEE Transactions on Systems, Man and Cybernetics,* February, 2009.

[11] Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *Accepted Conference on Computer Vision and Pattern Recognition*, 2001.

[12] Rainer Lienhart, Alexander Kuranov, Vadim Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection", *MRL Technical Report*, May 2002, revised December 2002.

[13] M. Castrill´on-Santana, O. D´eniz-Su´arez, L. Ant´on-Canal´ıs and J. Lorenzo-Navarro, "FACE AND FACIAL FEATURE DETECTION EVALUATION Performance Evaluation of Public Domain Haar Detectors for Face and Facial Feature Detection",

[14] M. Castrillo´n , O. De´niz, C. Guerra and M. Herna´ndez, "ENCARA2: Real-time detection of multiple faces at different resolutions in video streams",

[15] Richard I. Hartley and Peter Sturm, "Triangulation", *6th International Conference on Computer Analysis of Images and Patterns*, September, 1995.

[16] Bradski, G.R. "Computer Vision Face Tracking for Use in a Perceptual User Interface", *Intel, 1998.*

[17] Atsushi Yamashita, Yu Ito, Toru Kaneko and Hajime Asama, "Human Tracking with Multiple Cameras Based on Face Detection and Mean Shift".

[18] "The OpenCV Reference Manual", *Release 2.4.2, July, 2012*

[19] Boris Babenko, "A Derivation of Discrete AdaBoost".

[20] Wei-Chao Lin and Michael Oakes "Real AdaBoost for Large Vocabulary Image Classification".

[21] Yang, Guo-peng, Zhengzhou "Hyperspectral Imagery Classification Based on Gentle AdaBoost and Decision Stumps", *International Conference on Information Engineering and Computer Science,* 2009.

[22] Ping Li, "Robust LogitBoost and Adaptive Base Class (ABC) LogitBoost".

[23] Matlab, Camera Calibration Toolbox. www.vision.caltech.edu/bouguetj/calib_doc

[24] Matlab, Camera Calibration Toolbox, Single Camera Calibration. http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/example.html

[25] Matlab, Camera Calibration Toolbox, Stereo Camera Calibration. http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/example5.html