

Fusion of Double Layered Multiple Laser Range Finders for People Detection from a Mobile Robot

Alexander Carballo, Akihisa Ohya and Shin'ichi Yuta

Abstract—This work proposes a new method people detection and position estimation from a mobile robot by fusion of multiple Laser Range Finders arranged in two layers. Sensors facing opposite directions in a single row (layer) are fused to produce 360° scan data of robot's surroundings, then data from every layer is further fused to create a 3D model of people and from there their position. The main problem of our research is an autonomous mobile robot acting as member of a people group moving in public areas, simple and accurate people detection and tracking is an important requirement. We present experimental results of fusion steps and people detection in an indoor environment.

I. INTRODUCTION

Companion robots are becoming more part of daily life and are designed to directly interact with people, from a pet robot to aid the development process in children as to provide company to lonely elder people, to complex humanoid robots programmed with verbal interaction and providing services like guiding, entertainment and company. One necessary subsystem for such robots is detection, recognition and tracking of people as well as obstacles in the environment.

Laser Range Finders (LRF), besides being used for obstacle detection are also an important part of tracking systems, with important advantages over other sensing devices, like high accuracy, wide view angles, high scanning rates, robustness to changes in environment, usage simplicity and relatively less computing power to process data. However, 2D LRF data is not enough to solve important problems like occlusion in scan data, a set of multiple laser finders in different locations inside an area (for example [1], [2]) may reduce occlusion problems and effectively track multiple people, however detection is limited to the selected area and not suitable for tracking from a mobile robot.

Current approaches based on LRFs ([3], [4], [5], [6]) place the sensors in the same height (single row), some 20cm to 50cm from the ground, to detect and track people's legs. However, detection of legs in cluttered environments is difficult especially if people are standing still. In Fod *et al* [1] a row of several LRFs on different positions in a room were used for tracking moving objects, blobs (segments) are extracted and future positions is estimated according to a motion model. Xavier *et al* [4] focused on people detection using a fast method for line/arc detection but from a fixed position. Zhao *et al* [2] proposed a walking model to improve position prediction by including information about

leg position, velocity and state. The later model was then used by Lee in [5] but this time from a mobile robot. Montemerlo *et al* [7] also uses LRF from a mobile robot for people tracking and simultaneously robot localization by using conditional particle filters. Also in [3] a mobile platform is used for a monitoring system based on a cart with two LRFs on a single row, to monitor people motions; their system also helps covering blind spots by moving the cart. Finally, Arras *et al* [6] considers the problem of how to set the necessary threshold values and which features to use to successfully detect people from a mobile robot.

We propose in this work a new method for multiple people detection and position estimation by fusion of several LRF sensors, as other approaches, but installed in a layered (multirow) arrangement. The idea is to obtain simultaneously different-but-complementary features to better detect people even in cluttered environments. This approach allows robust people detection even in presence of occlusion and is simple to implement using a mobile robot by placing sensor layers in the robot body at different heights from the ground depending on the features to detect.

Fig. 1 represents our layered approach, every layer has two sensors facing opposite directions for 360° scanning (Fig. 1(a)), and two layers are used to extract features from upper and lower parts of a person's body (Fig. 1(b)).

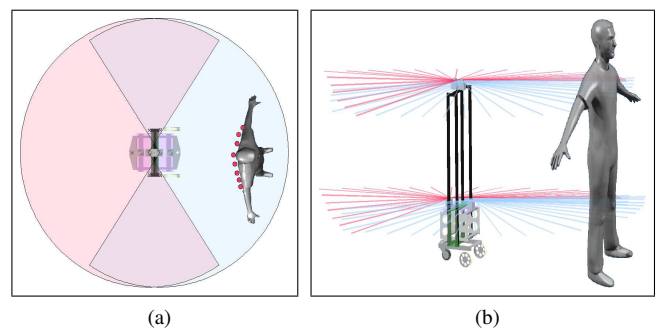


Fig. 1: Scanning from a double layered approach: (a) opposite facing sensors (top view) and (b) two layers of sensors (lateral view).

Our method involves two fusion steps: fusion of sensors in a single layer and then fusion of layers. In the first step, sensors facing opposite directions in the same layer are fused to produce a 360° representation of robot's surroundings. There is overlapping of scan data from both sensors (darker areas in fig. 1a) so this fusion step must deal with data duplication. Then, in the multiple layer fusion step, raw data from every

layer is processed to extract features corresponding to people, then a 3D representation (a people model) is obtained and from it allowing people detection and person position and direction estimation.

The main idea of our approach is depicted in Fig. 2, after fusion of sensors in every layer, geometrical features are extracted (Fig. 2(a)): in the upper one large elliptical shapes (chest areas) are seen while in the lower layer circular shapes (legs) are also visible. Then fusion of the extracted features from both layers define a 3D volume (Fig. 2(b)) enclosing every person detected.

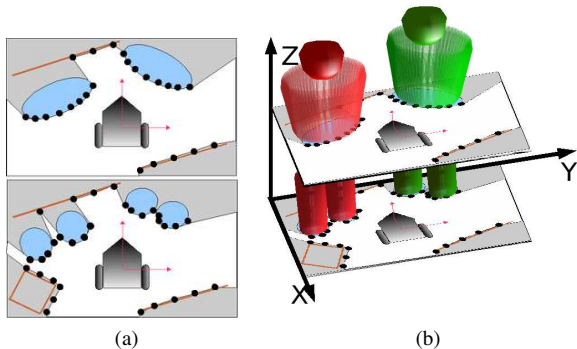


Fig. 2: (a) Projection of two 2D planes for a 3D representation, (b) a volumetric representation of people.

A simple yet logical assumption is that an elliptical shape corresponding to a chest is always associated to one or two circular shapes corresponding to legs, and that the large elliptical shape (chest) is always *over* the set of small circles (legs). Fusion of extracted features allows creating a 3D volume (fig. 2b) and from it the estimated person position is computed.

Our main research goal aims to develop a companion robot with the objective to study the relationship of an autonomous mobile robot and a group of multiple people in a complex environment like public areas, where the robot is to move and behave as another member of the group while achieving navigation with obstacle avoidance.

The robot used for our research is depicted in Fig. 3. It is based on *Yamabico* robotic platform [8]. Two layers of LRF sensors are used, the lower layer is about 40cm from the ground while the upper layer is about 120cm. Every layer consists of 2 LRF sensors, one facing forwards and another facing backwards for a 360° coverage (figures 1 and 3). The sensors used in our system are the *URG-04LX* range scanners ([9] provides a good description of the sensor's capabilities).

The rest of the paper is organized as follows. In section II we describe the step of fusion of sensors in a single layer into a 360° representation. Section III presents our approach for fusion of multiple sensor layers, including feature extraction, people detection and position estimation. Section IV presents experimental results for the different fusion steps and for people detection. Finally, conclusions and future work are left for section V.



Fig. 3: Our robot system for multiple people detection and tracking, four *URG-04LX* are used.

II. FUSION OF LRF SENSORS IN SINGLE LAYER

The robot scans its 360° surroundings from both sensor layers having two sensors in each layer, one facing forward and one facing backwards. A total of 4 LRFs are used, as presented in Fig. 3, raw scan data from every sensor is read, timestamped and integrated with odometry information in pairs. According to the top view representation in Fig. 1a, real scan data obtained from both sensors in one layer is presented in Fig. 4 (data from the upper layer).

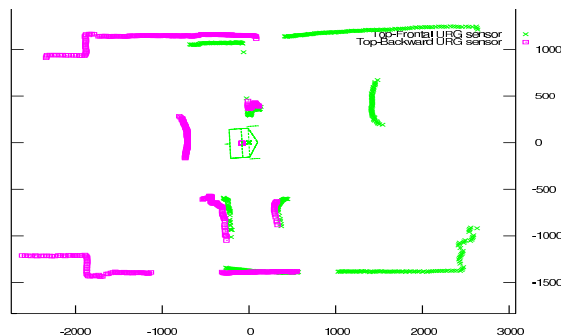


Fig. 4: Raw scan data from the upper layer

A previous step is to ensure sensor-to-sensor alignment which is to properly fix the pose of the two sensors so as the resulting combined scan data appears to be obtained from a single sensor. In other words scan data from areas where the sensors overlap should match and ideally be indistinguishable. A top view of real scan data is presented in Fig. 4, this overlap problem at left and right sides of the robot is also present.

A. Sensor Fusion

We require scan data from the complete surroundings of the robot thus we need to combine two opposite facing sensors' readings into a 360°. Ideally if the sensors we perfectly aligned data in the overlapping areas will be indistinguishable, however the different poses of the sensors

provide different viewpoints with different problems, while a sensor's beam may suffer from reflection problems in one point the other sensor's beam (if both hitting the same point) may not. Our approach to fuse not-perfectly aligned scan data from two sensors in the same layer is described in next paragraphs.

A sensor s provides range scan points p_i^s from a set \mathcal{P}^s consisting of range and direction $[r_i^s \ \theta_i^s]$ for every beam $i, i \in 1..N$ in a local coordinate system \mathcal{L} where direction is in the range $\{-120^\circ..120^\circ\}$. Data from sensors facing opposite directions is transformed into a global system \mathcal{G} such that $[r_i^s \ \theta_i^s]$ is converted into $[R_k \ \phi_k]$ for each sensor.

Figure 5 represents this idea, the pair of sensors 1 and 2 facing opposite directions and separated in the vertical axis over a distance $d_{1 \rightarrow 2}$ are combined by transforming their scan data into \mathcal{G} .

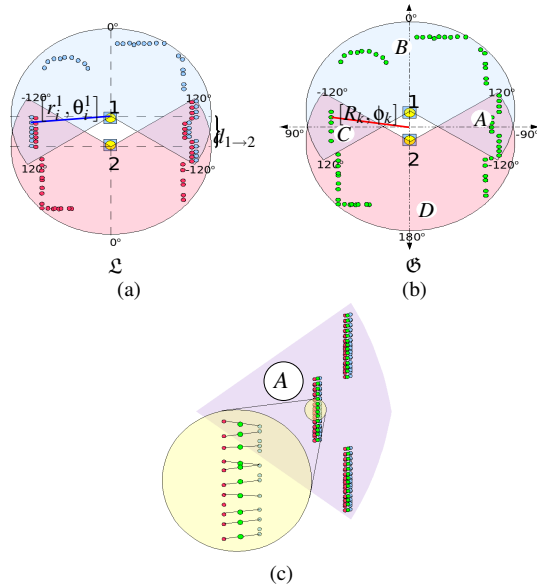


Fig. 5: Fusing scan data of two sensors in the same layer, (a) sensors in local coordinate system \mathcal{L} , (b) converted to global system \mathcal{G} and fused, (c) averaging of matching points in overlapping areas A and C.

However sensors in this arrangement share scan points in *overlapping* areas (dark areas in Fig. 5(b) labeled as A and C, with points in range from -120° in one sensor to 120° in the opposite) thus a problem of duplication of data exists. *Non-overlapping* areas (B and D, right after the end of one overlap to 0° to the start of the next overlap, in both sensors) correspond to sensor's independent observations but those in the overlapping areas include both independent and shared observations, for the difference in pose of the sensors allows different points of view of the same object.

To cope with this duplication of data in overlapping areas, a method was implemented to find for every scan data of one sensor the closest on the other sensor and obtain the average of their positions in \mathcal{G} . Data in non-overlapping areas was left unchanged in \mathcal{G} .

After joining every pair of sensors, scan data addition-

ally transformed to include robot odometry information $[x \ y \ \phi]^T$. The 2D representation of sensors in both layers (chest and legs) can be considered as two different XY planes in a 3D representation (fig. 2), thus the Z component for every plane is the actual sensor height in the robot body (sensors are not tilted).

III. FUSION OF DOUBLE LAYERED LRF SENSORS

Sensors in the same layer are facing opposite directions, individual scan data are combined into a 360° representation in the previous section. The next step is fusion of both sensor layers, here data will be divided into clusters with a segmentation function and then clusters will be classified according to their geometrical properties. Finally only those segments that match people features will be selected and joined into a 3D model from where people position is obtained.

A. Segmentation

Data clustering can be considered as the problem of breakpoint detection and finding breaking points in scan data can be considered as the problem of finding a threshold function \mathcal{T} to measure separation of adjacent points. Every pair of neighboring points p_j and p_k are separated by an angle α which is proportional to the sensor's angular resolution (true for points of two adjacent scan steps) and by a distance $\mathcal{D}(p_j, p_k)$. Points are circularly ordered according to the scanning step of the sensor.

A cluster C_i , where $C_i = \{p_i, p_{i+1}, p_{i+2}, \dots, p_m\}$, is defined according to a cluster membership function \mathcal{M}

$$\mathcal{M}(p_j, p_k) = (\theta_k - \theta_j) \leq \alpha \wedge \mathcal{D}(p_j, p_k) \leq \mathcal{T}(p_j, p_k)$$

such that for every pair $\langle p_j, p_k \rangle$ of adjacent points, the Euclidean distance $\mathcal{D}(p_j, p_k)$ between them is less than a given threshold function $\mathcal{T}(p_j, p_k)$ for p_j, p_k . A new point p_n is compared to the last known member p_m of a given cluster C_i as $\mathcal{M}(p_m, p_n)$.

Now, the threshold function \mathcal{T} is defined for a pair of points, as in the work of Dietmayer [11], as:

$$\mathcal{T}(p_i, p_j) = C_0 + C_1 \min(r_i, r_j)$$

with $C_1 = \sqrt{2(1 - \cos(\alpha))}$. Dietmayer's work includes the constant C_0 to adjust the function to noise and overlapping. In our case C_0 is replaced by the radius \mathcal{R} of the accuracy area for p_i as base point plus a fixed threshold value (10cm in our case). \mathcal{R} is defined according to the URG-04LX sensor specifications [9], [10] as:

$$\mathcal{R}(p_i) = \begin{cases} 10 & \text{if } 20\text{mm} \leq r_i \leq 1000\text{mm} \\ 0.01 \times r_i & \text{otherwise} \end{cases}$$

The proposed threshold function \mathcal{T} uses this accuracy information \mathcal{R} when checking for break points, if two neighboring points have a large range value, it will be most probable that they form part of the same cluster for their bigger accuracy areas.

There is also a cluster filtering step that will drop segments very small to be considered of significance.

B. Feature Extraction

The idea of *feature extraction* is to match the sensor readings with one or more geometrical models representing expected behaviour of the data. For example if a LRF sensor data scanning a wall, then the *expected behaviour* of a wall scan data is a *straight line*. Also if the same sensor is to scan a *person* then the expected behaviour is a set of points forming an *arc*. So in order to identify walls a first requirement is to correctly associate the scan data with some straight line model, for people the same: associate a set of scan points to an arc shape (a circle or an ellipse).

Before applying any fitting method, it is important to have some information about the shape of the cluster that allows selecting the method. The information about clusters is extracted as a set of indicators like number of points, standard deviation, distances from previous and to next clusters, cluster concavity/convexity, etc. ([6] presents a list of indicators useful for people detection).

One of the indicators is the cluster's *linearity*; our approach here is to classify the clusters into *long-and-thin* and those rather *short-and-thick*. The rationale behind this is that, straight line segments tend to be long and thin, round obstacles, irregular objects, etc., do not have this appearance.

Linearity is achieved by computing the covariance matrix C for the cluster C_i and then its eigenvalues λ_{max} and λ_{min} that define the scale and its eigenvectors v_1 and v_2 orientation (major and minor axes) of the dispersion of C . The ratio $\mathcal{L} = \lambda_{max}/\lambda_{min}$ defines the degree of longness/thinness of the cluster. We set threshold values for ratio \mathcal{L} and for λ_{max} .

The *ellipticality* factor is computed as the standard deviation σ of the residuals of a ellipse fitting processes using the Fitzgibbon method [12]. The distance between a cluster point and an ellipse is computed using *Ramanujan's* approximation. Only clusters with good ellipticality value are selected and segments passing the linearity criteria (that is lines) can be easily rejected since they do not belong to people.

C. People Model and Position Detection

As previously presented in Fig. 2, 3D projection of two planes of scan data from the layered sensors can be used to represent the position and direction of a person.

The set of geometrical features extracted from the former step are mostly ellipses and circles. If they belong to a person another important criteria should be met: the large elliptical segment should come from the upper layer and the small circles from the lower layer. No large ellipses are possible for a person in the leg area. The small circles can not be over the large ellipse (the person height is restricted according to the height of the upper layer).

To properly establish the previous requirements, it is necessary to associate segments in the upper layer with those in the lower layer, this is to find the corresponding legs for a given chest. Even if data in both layers is aligned inside a 3D volume it is possible that legs lie inside or outside this volume according to the speed of motion and step length (length between feet when walking).

With an estimation of the maximum value for d , the separation of legs at the lower layer height, we can set a search radius of $d/2 \pm \xi$ at the center of the chest elliptical area projected into the lower layer to search for the corresponding legs for the chest. We use average walking step length from Latt *et al.* [13], at normal walking speed, to compute the value for d . This idea is represented in Fig. 6. Once associated, the position of the person will be finally set by the position of the chest ellipse center.

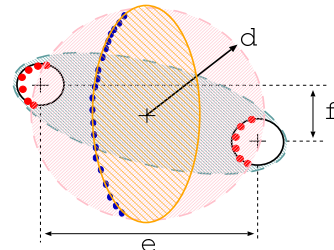


Fig. 6: Searching legs of a person in an area with radius $d/2 \pm \xi$ with center at chest ellipse in the 3D volume.

IV. EXPERIMENTAL RESULTS

The robot used for our research was presented in Fig. 3, the computer operating the robot is a Intel Pentium Core Duo based notebook running (Linux kernel 2.6.24) as operating system and robot control board is powered by a Hitachi SH-2 processor. The robot system uses 4 *URG-04LX* range scanners from *Hokuyo Automatic Co., Ltd.*[10], small size (50x50x70mm), covers distances up to 5.6m, distance resolution of 10mm and angular resolution of 0.36°, angular range of 240° operating at 10Hz. Scan data from each sensor consists of 682 points circularly ordered according to scanning step.

Data from each sensor is read every 100ms by a driver processes registered in parallel into a shared memory system (*SSM*[14]) based on IPC messaging and multiple ring-buffers with automatic timestamping, one driver process per sensor. *SSM* also allows to record raw sensor data into log files and to play it back with the same rate as the sensor (10Hz in this case).

Client processes read scan data from the ring-buffers according to sensor's pose (those in the top layer and those on the low layer), pairs of LRF sensors are processed in the fusion step, sensor layers are further fused and finally people position is computed.

The processing time for the two layers (4 sensors), from single layer fusion to people position detection, was less than 50ms, fast enough given the sensor's scanning speed.

In the following subsections we present the results of the different tasks involved in this method for people detection.

A. Fusion of LRF sensors in single layer

Results of fusing opposite facing sensors are presented in Fig. 7. The fusion method joins correctly the data from both sensors, data from areas *B* and *D* (5) is copied as it is, and data from areas *A* and *C* uses simple averaging of closest

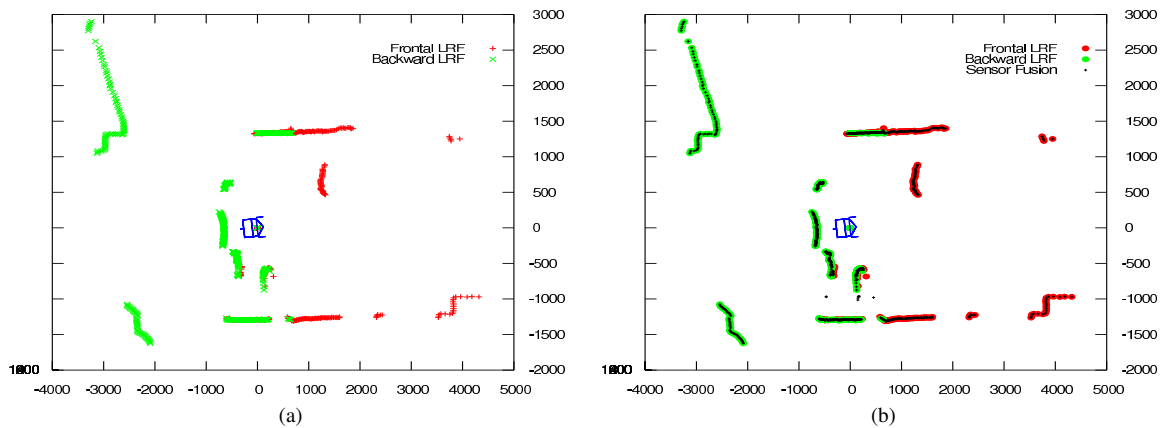


Fig. 7: Fusion of sensors in a single layer: (a) raw data (front sensor in red and backward in green points) (b) results of fusion (black points).

points from both sensors. Sensor data is joined and a 360° representation of the surrounding environment is possible.

These results are important for the next steps since we obtained one set of points from the two sensors, from there the segmentation step in the fusion of double layers can extract clusters without having to consider the case of duplication and further merging of clusters.

B. Fusion of double layered LRF sensors

Figure 8 shows the results of an experiment for people detection and position estimation from a mobile robot. In the experiment 5 persons walked around the robot and additional person was taking the experiment video (fig. 8(a)). Log data from each sensor was recorded, people position detection tests were performed off-line by playing back this log data using our SSM system. A 3D visualization process using OpenGL was created to verify how the people detection worked; in Fig. 8(b) chest ellipses and leg circular ellipses are detected then we place a 3D wooden doll, as a representation of a person, in the estimated position the person should have. Results were verified by human operator comparing the experiment video with results.

The members have varied body sizes, from broad and tall to thin and short. Some of the members have a height a little under the average, as result their chest ellipses were not correctly detected in the people detection step. As presented in Fig. 8(b), the person to the right of the robot (represented with blue line segments) is missing although circles from legs are present.

Additional snapshots of experimental results are presented in Fig. 9, the robot is represented in all cases as blue line segments. Fig. 9(a) shows raw scan data from both layers (red for the upper layer and green for the lower one) and in Fig. 9(b) a 3D representation of the human detection and position estimation. In the cases of 3D representation, the raw scan data is plotted together with wooden dolls enclosed in the estimated people positions represented with elliptical shapes, a large one for the chest area and smaller ones for the extracted leg areas.

In Fig. 9(a) there are two rather large arc-like segments in the raw scan image and two large elliptical shapes in the 3D representation in 9(b), in both layers. That is a column inside the indoor environment where the experiment took place. The people detection method discards this elliptical object because its dimensions are larger than the expected for people, those elliptical objects are represented with red color in this figure. Also we do not expect large elliptical objects from the lower layer so discarding this column as a non human object was simple.

V. CONCLUSIONS AND FUTURE WORKS

Fusion of multiple LRF sensors arranged in a double layer structure for multiple people position detection from a mobile robot studied in this paper. The proposed approach is simple and the double layer multiple LRF approach is practical enough to be implemented on a mobile robot. Instead of fusion of different sensors with complementary capabilities, we fused the same type but at different heights (layers), giving different perspectives which also helps solving simple cases of occlusion where one sensor is occluded and the other is not.

A simple method for fusion of opposite facing sensors in the same layer was presented, from it a simple 360° representation of the surrounding environment was possible and simplified the segmentation step in the fusion of both layers.

Fusion of double layers by segmentation of fused scan data, geometrical features extraction and association in a 3D volume for every detected person allows good position estimation and a measure of the possible direction the person is facing.

The method used here easily filters out non-people segments by analyzing the key indicators such as size, compactness, ellipticity and linearity. The addition of an extra layer of LRFs to detect chest elliptical areas improve the estimation of people position as the lower part of body (the legs) move faster and wider than the chest area. The combination of both areas creates a 3D volume which helps



Fig. 8: Results of the people detection, (a) snapshot (b) 3D models in estimated positions of people in the experiment.

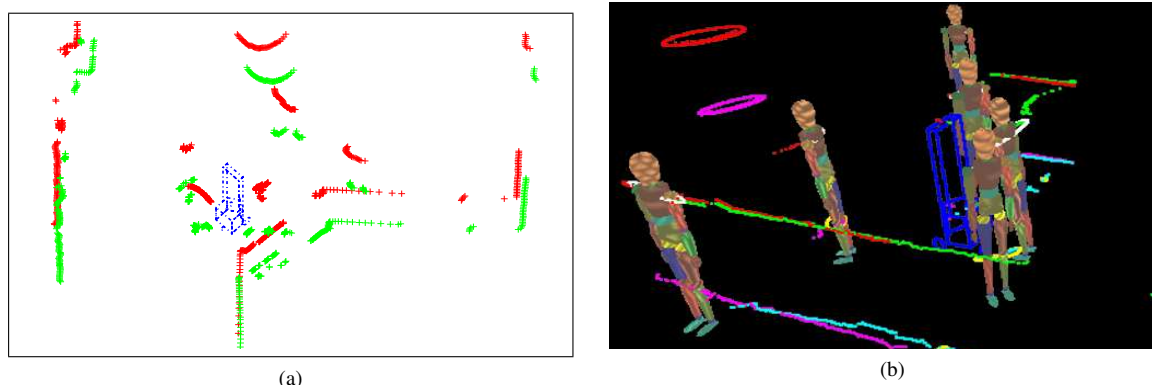


Fig. 9: Experimental results with raw scan data (a) and the corresponding people detection and position estimation (b).

locating the position of the person more closely related to the center of this 3D volume.

As future work, estimation of the person direction from motion and multiple people tracking will be considered. Also the effectiveness of our method in cluttered environments will be studied. Future steps of our research include understanding people group motion and recognition of group members.

REFERENCES

- [1] Ajo Fod, Andrew Howard and Maja J. Mataric. Laser-based people tracking. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Washington DC, U.S.A, 2002, pages 3024-3029.
- [2] Jinshi Cui, Hongbin Zha, Huijing Zhao and Ryosuke Shibasaki. Robust Tracking of Multiple People in Crowds Using Laser Range Scanners. In Proceedings of the IEEE 18th International Conference on Pattern Recognition (ICPR), Hong Kong, China, pp. 857-860, 2006.
- [3] Huijing Zhao, Yuzhong Chen, Xiaowei Shao, Kyoichiro Katabira and Ryosuke Shibasaki. Monitoring a populated environment using single-row laser range scanners from a mobile platform. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Apr. 10-14 2007, Roma, Italy, pp. 4739-4745.
- [4] 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. In Proceedings of the IEEE International Conference in Robotics and Automation (ICRA), Barcelona, Spain, April 2005, 3941-3946.
- [5] Jae Hoon Lee, Takashi Tsubouchi, Kenjiro Yamamoto and Saku Egawa. People Tracking Using a Robot in Motion with Laser Range Finder. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Beijing, China, October 2006, pp. 2936-2942.
- [6] Kai O. Arras, Óscar Martínez Mozos and Wolfram Burgard. Using Boosted Features for the Detection of People in 2D Range Data. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Apr. 10-14 2007, Roma, Italy, pp. 3402-3407.
- [7] Michael Montemerlo, Sebastian Thrun and William Whittaker. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '02), 2002.
- [8] Shin'ichi Yuta, Shooji Suzuki and Shigeki Iida. Implementation of a Small Size Experimental Self-contained Autonomous Robot - Sensors, Vehicle Control, and Description of Sensor Based Behavior. Second International Symposium on Experimental Robotics (ISER) June 25-27 1991, Toulouse, France.
- [9] Hirohiko Kawata, Satofumi Kamimura, Akihisa Ohya, Jun'ichi Iijima, Shin'ichi Yuta. Advanced Functions of the Scanning Laser Range Sensor for Environment Recognition in Mobile Robots. Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Heidelberg, Germany, pp. 414-419, 2006.
- [10] Hokuyo Automatic Co., Ltd. URL <http://www.hokuyo-aut.co.jp/>
- [11] Klaus C.J. Dietmayer, Jan Sparbert and Daniel Streller. Model Based Classification and Object Tracking in Traffic Scenes from Range-Images. Proceedings of IV IEEE Intelligent Vehicles Symposium, May 2001, Tokyo, Japan.
- [12] Maurizio Pilu, Andrew W. Fitzgibbon and Robert B. Fisher. Ellipse-specific direct least-square fitting. Proceedings of the IEEE International Conference on Image Processing 1996. v. 3, pp. 16-19, Sep. 1996.
- [13] Mark D. Latt, Hylton B. Menz, Victor S. Fung and Stephen R. Lord. Walking Speed, Cadence and Step Length are Selected to Optimize the Stability of Head and Pelvis Acceleration. *Experimental Brain Research*, 184(2), pp. 201-209, Springer Berlin, 2008.
- [14] Eijiro Takeuchi, Takashi Tsubouchi and Shin'ichi. Yuta. Integration and Synchronization of External Sensor Data for a Mobile Robot. *SICE Annual Conference in Fukui, August 4-6 2003, Fukui University, Japan*