2014 International Conference on Robotics and
Emerging Allied Technologies in Engineering (iCREATE)
Islamabad, Pakistan, April 22-24, 2014

Long-Term Estimation of Human Spatial Interactions Through Multiple Laser Ranging Sensors

Amir Salarpour, Hassan Khotanlou
Department of Computer Engineering
Bu-Ali Sina University
Hamedan, Iran
{a.salarpour, khotanlou}@basu.ac.ir

Nikolaos Mavridis
Interactive Robots and Media Lab
NCSR Demokritos
Athens, Greece
nmav@alum.mit.edu

***Abstract*—For robots to be able to naturally co-habitate human spaces, as well as to interact with single humans or groups of humans, they should be able to navigate in ways that are human-friendly, and appropriate to human spatial interaction social norms, such as keeping personal spaces. For that purpose, we are developing a special theory which extends path planning, which we call social path plan, which allows humans or groups as obstacles or goals. In order to provide tuning for our simulation results, we are acquiring a natural human interaction dataset, through measurements from multiple laser ranging sensors positioned at a cross-roads indoor space. We thus describe our system consisting of spatial and temporal alignment algorithms for multiple laser sensors, as well as foreground detection, sensor data fusion, segmentation, tracking, two-legged position and pose estimation, and event detection. The method presented can be easily extended to larger spaces and applied for many other application domains beyond our main goal of learning optimal spatial interaction behaviours for human-robot interaction.**

Keywords— Human-Robot Interaction, Spatial, Laser Range Finder, People Detection, People Tracking

I. INTRODUCTION

People tracking has been widely used in various applications, such as surveillance, activity recognition, building security and traffic flow analysis. There is a lot of developed tracking system for people tracking based on video streams [1], [2]. These systems face difficulties when dealing with many people in a relatively large area, due to the frequent occlusion among people and the difficulty in integrating video streams from multiple cameras. The laser range finder has received increasing attention for tracking problems in recent years. A laser range finder is a device which uses a laser beam to determine the distance to an object. The most common form of laser rangefinder operates on the time of flight principle by sending a laser pulse in a narrow beam towards the object and measuring the time taken by the pulse to be reflected off the target and returned to the sender.

Because of good precision and fast sensing ability, laser range finder has become one of the most popular equipment in the robotics community. Knowledge about the presence, position, and motion state of people enables robots to better understand and anticipate human intentions and actions. Apart from human-robot interaction and cooperation scenarios, applications of laser-based people tracking include also surveillance, crowd control, or pedestrian detection for intelligent cars.

Additionally, tracking persons with laser range finders does not violate personal rights and puts privacy law advocates at ease to a certain degree. Using only one static LRF limits the area for tracking to its maximum detection range. Also occlusions may disturb the tracking. Both problems can be avoided by adding more laser range finders to the scene, however, then special algorithms for spatio and temporal alignment, as well as data fusion, should be included in the processing pipeline of the people tracking system.

Several range scan registration techniques based on the Iterative Closest Point (ICP) have been introduced and successfully adopted in mobile robot localization [3], [4]. The ICP algorithm uses an iterative process in the following steps: first, a set of points in one or both range images are selected; then, correspondences between two range scans are established; finally, an error metric is defined and minimized to compute the rigid-body transformation. This “select-match-minimize” procedure is repeated until two range scans are converged. Since true correspondences are generally unknown for range scan registration, the ICP algorithm utilizes the closest points in Cartesian coordinate frame as an approximation of true correspondences.

Towards enabling human-friendly and socially appropriate human-robot spatial interaction, we have thus introduced the “social path planning” problem formulation in [5], enumerating six specific subproblems, which include humans or groups of humans either as obstacles or as goals of the social path planner. We have furthermore provided an initial solution of the problem using the fast-marching squares algorithm [6]. However, in order to be able to tune the parameters of our solution so that they are closer to actual human behaviors respecting behavioral norms such as personal spaces, we need datasets of appropriate format including longer-term recording of human spatial interactions.

For that purpose, we are presenting in this paper a system utilizing multiple laser ranging sensors, which was installed in an indoors highly frequented cross-roads space. Our system includes stages for spatial and temporal alignment, as well as foreground detection, sensor data fusion, segmentation, tracking, two-legged position and pose estimation, and event detection. The methods presented can be easily extended to larger spaces and applied for many other application domains beyond our main goal of learning optimal spatial interaction behaviors for human-robot interaction.

We will thus proceed in this paper by providing background on relevant existing research, including the basics of our six sub-problem decomposition of the social path planning problem. Then, in section III, we will describe our data set capture conditions and specifications. In the next section (IV), we will present spatial and temporal alignment algorithms and results when applied to our dataset. Then, in section V, detection and tracking will be covered, all the way to position and pose derivation from feet tracking, and event derivation, followed by a forward-looking concluding section.

II. BACKGROUND

For multi LRF matching, at least four categories of methods that could potentially be used. First, known approaches in mobile robotics such as [7], [8]. Unfortunately these methods cannot be used without robots, since these methods use odometry information to know the relative movement. Second, methods utilizing moving object trajectories over time, towards matching. These methods extract for example a walking people trajectory and use point registration approaches as a matching tool to know the constraints between LRFs [9], [10]. However, the main hypothesis in these methods is the LRF's observations are temporal aligned, which is rarely the case in real-world systems. Third, one could utilize static objects in scan data to match the LRFs. Static objects that are observable from LRFs could be used as a matching reference. Two famous methods for static objects matching are ICP [3], [4] and polar scan matching (PSM) [11]. Also, computer vision point registration methods could be used [12]. Occlusion and symmetries in environment could severely deteriorate the performance of such kind of methods, although an adequate pre-existing background model might be used as a solution to such problems.

Most of human detection studies emphasize motion tracking. In [13] kalman filtering is used to track the movement of multiple fixed LRFs. In [14], [15], [16] a walking model is proposed to extract the leg position and track the legs using an extended Kalman filter (EKF). In [17], [18] multiple hypothesis tracking (MHT) is applied to the problem of human leg tracking. Another scheme based on target tracking is the sample-based joint probabilistic data association filter (SJPDAF) in [19]. As we also need to track the people who are standing and but not moving, foreground detection in our work is applied based on difference from the background model, and not just the movement of objects.

In our previous work which introduced the definition and subproblems of "social path planning" [5], a novel classification and a mathematical formalization of all the different cases was proposed, and decomposed to 6 different types of subproblems:

1) Single human, individual:

a) Robot to point. Regular path planning with special consideration of humans as obstacles.

b) Full interaction: 1) approach human, 2) interact, keep interaction, 3) disengage.

c) Follow human.

2) Group of humans:

a) Robot to point. Regular path planning considering groups of humans as special obstacles.

b) Observe group, ask for permission to enter.

c) Full interaction: 1) enter the group, 2) interact, keep interaction, 3) disengage.

III. DATA SET DESCRIPTION

Our dataset is composed of two sets of data which were initially acquired from two Laser Range Finders (LRF). The data were acquired in the first level hall of IIT building in the NCSR Demokritos research institute on 30th July 2013 between 2 pm and 5 pm, in an indoor location with an area of about 6m×19m, with 4 hallways for access to other parts of the building, a staircase, 2 room entrances/exits, and 2 vending machines that are shown in Fig. 1. Also two different laser ranging finders (SICK and Hokuyo) were used to cover the hall, the blue dots and red dots showing the range that acquired by SICK laser and Hokuyo laser, respectively.

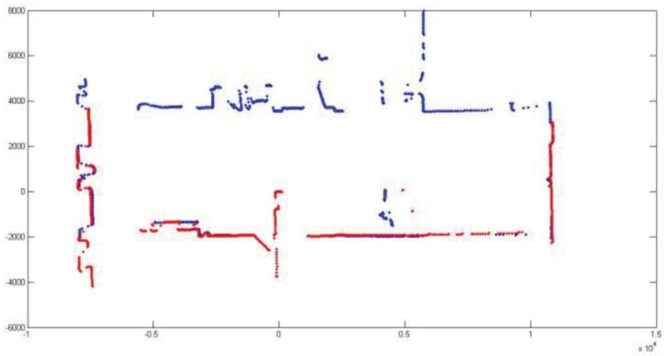


Fig. 1. the hall in which the data was acquired with detail.

The model of the first LRF was a SICK LMS101 with 0.25° angular resolution and 25 Hz frequency. For 270 degree range there are 1080 readings in millimeters. We used the "SICK engineering tool" that is given by the laser manufacturer. The capture data format is:

Timestamp (millisecond accuracy) [semicolon] (column 2 to 22 header) [semicolon] 1081 readings with semicolon between them (millimeter accuracy) [newline character]

The output was in CSV format. We imported the file in Matlab, removed columns 2 to 22 and saved for further use. The sample outputs of SICK laser are shown in Fig. 2, in Cartesian and Polar coordinates.

The second LRF was a Hokuyo UTM-30lx with 0.25° angular resolution and 10 Hz frequency. For 180 degree range there are 720 readings in meters. The capture data format is:

Time (the world "time") [space] Timestamp (millisecond accuracy) [space] 721 readings with spaces between them (meter accuracy) [newline character]

The file was converted to the desired format using Matlab, i.e. useless data were removed and distances converted to mm.

Fig. 3 shows the sample output of the Hokuyo laser in Cartesian and Polar coordinates.

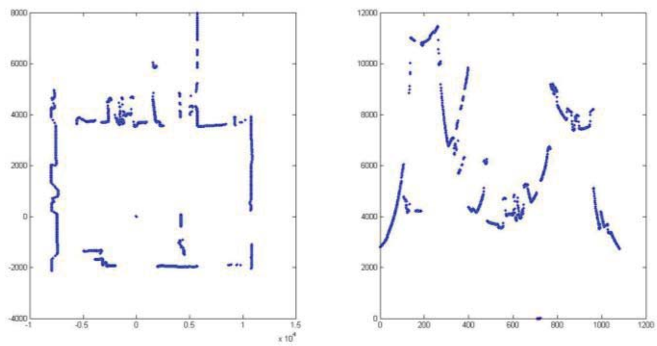


Fig. 2. The sample output of SICK laser. a) Cartesian coordinate. b) Polar coordinate.

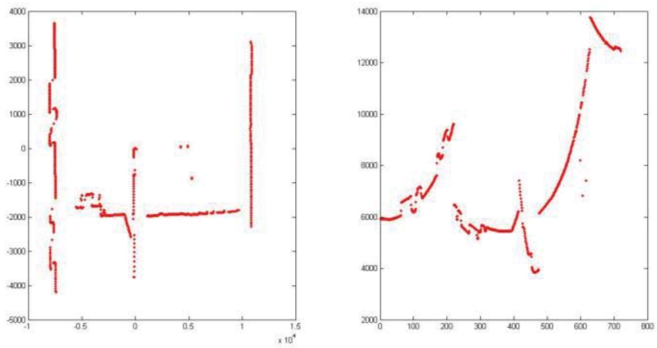


Fig. 3. The sample output of Hokuyo laser. a) Cartesian coordinates. b) Polar coordinates.

Our data set after preprocessing and removing useless information consisted of 4 files: SICK1 and SICK2 were the first and the second part of the SICK laser output, respectively, and the Hokuyo data were in HOK1 and HOK2.

The characteristics of these files collected in Table I.

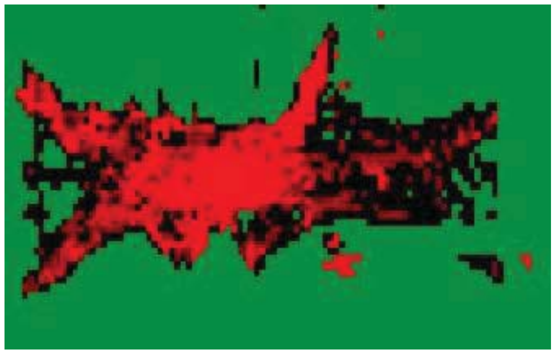
TABLE I. THE CHARACTERISTICS OF FOUR PREPROCESSED FILES.

File name	Time (millisecond)	Number of rows	Average time differences of rows
SICK1	7200063	171920	41.8803
SICK2	3600063	86498	41.6202
HOK1	6231290	50968	122.2589
HOK2	4049774	33107	122.3238

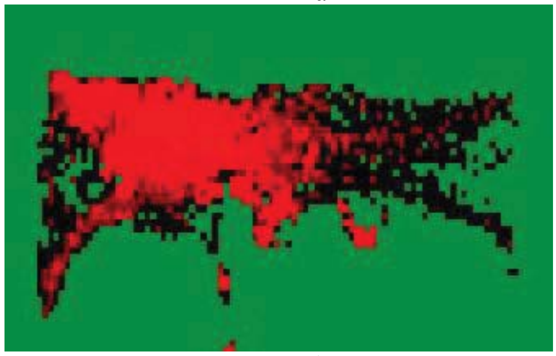
In Fig. 4 the spatial occupancy heat maps of two lasers are shown, indicating how frequently a position was occupied.

IV. SPATIOTEMPORAL ALIGNMENT

One the most important preprocessing phases is spatiotemporal alignment, which is divided to two subsections: spatial aligning and temporal aligning.



a



b

Fig. 4. Heat map of two lasers. a- SICK laser. b- Hokuyo laser.

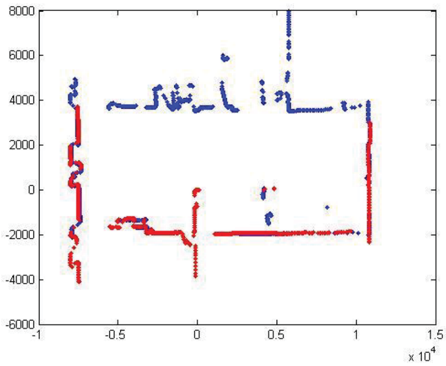
A. Spatial Alignment

At first we need to model the background of two lasers. Since laser range measurements give us the distance to an object, we can use simple rules to classify background. We assume that the farthest known stationary object is part of the background. And then we update the mean and variance of background model with (1) and (2):

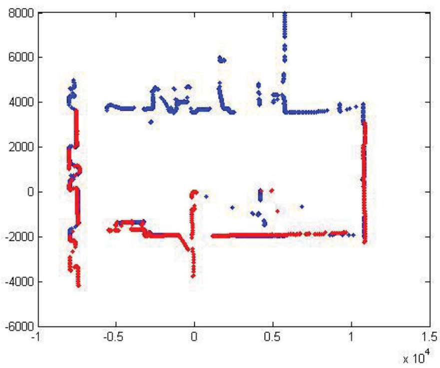
$$Mean_{n+1} = Mean_n + \frac{x_{n+1} - Mean_n}{n+1} \tag{1}$$

$$Var_{n+1} = \frac{n-2}{n-1} \times Var_n + \frac{(x_{n+1} - Mean_n)^2}{n} \tag{2}$$

After extracting the background model for the two LRF, we can use ICP (Iterative Closest Point) to find the rotation and translation which are needed in order to perform the optimal spatial match between the two laser range images. In our case due to the structure of the environment and fixed LRF we just need a single rotation and translation vector. ICP is an algorithm employed to minimize the difference between two clouds of points. The algorithm is conceptually simple and is commonly used in real-time. It iteratively revises the transformation needed to minimize the distance between the points of two raw scans. ICP has a rejection parameter to ignore the specified percent of worst match points. We get utilized a simple genetic algorithm (GA) to tune the rejection parameter of ICP. Fig. 5 shows the result of ICP-GA algorithm compare with ground truth. In our work we don't use background updating due to the possibility that a human stays in one place for a long time: if we had used background updating, he would be considered as a part of background.



a



b

Fig. 5. a- The matching result by using ICP-GA method b- Ground truth

B. *Temporal aligning*

For temporal alignment, we first convert our data to Cartesian coordinates. Then we try out multiple time delay choices and use a correlation metric to find the best correspondence. We start with 60 seconds and then 10 seconds to 1 millisecond, in a multi-resolution coarse to fine method. In highest level we compare two input stream in 60 second parts and find the best match then we continue it in 1 second and millisecond. We have found that the optimal temporal alignment for our case was at 25364 milliseconds.

V. DETECTION AND TRACKING

A. *Foreground detection*

The model of background that we calculated before, we are now utilizing to detect foreground. For each beam in polar coordinate if there is a significant difference between background model and current frame, we label it as foreground. After that we do some processing on the foreground result. At first we consider one or two points foreground as a noise and remove it. And the second one is if there is a foreground that is further than background, we consider as an error and remove.

B. *Merged foreground segmentation*

After that we merge the two LRF foreground points with calculated translation and rotation in a Cartesian coordinate. Mean shift clustering is then used to segment each leg of humans. The sample result of this stage are shown in Fig. 6.

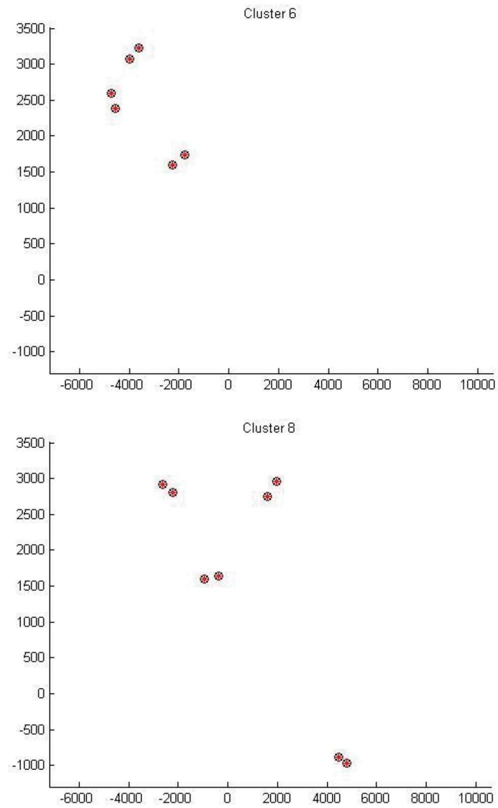


Fig. 6. The result of leg clustering

Then, at the output of the previous stage, a restricted Mean shift clustering [20] is performed: i.e. we cluster the detected legs in order to consider each person as a separate cluster (with one or two legs). The sample output is shown in Fig 7.

C. *Tracking*

The Hungarian algorithm [21] used as matching algorithm between two consecutive frame cluster centroids. The Hungarian algorithm returns the best match between two consecutive frames. A maximum distance is also considered to remove matches upper than that. Based on two level clustering the direction of each match is also calculated. Based on these outputs the trajectories extracted. Then trajectories are then processed to close potential gaps, caused by short-duration disappearance of human legs, caused by occlusions or bandwidth-parameter mismatch. Fig. 8 shows the extracted trajectories for constrained periods of time:

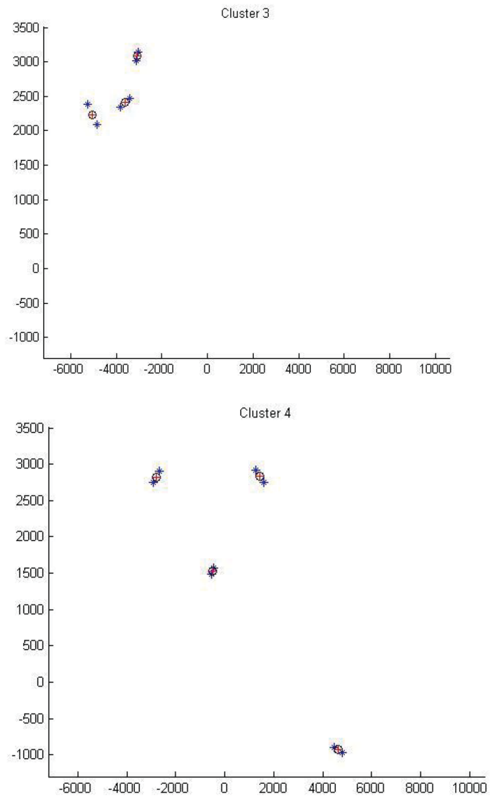


Fig. 7. The result of human clustering (second clustering)

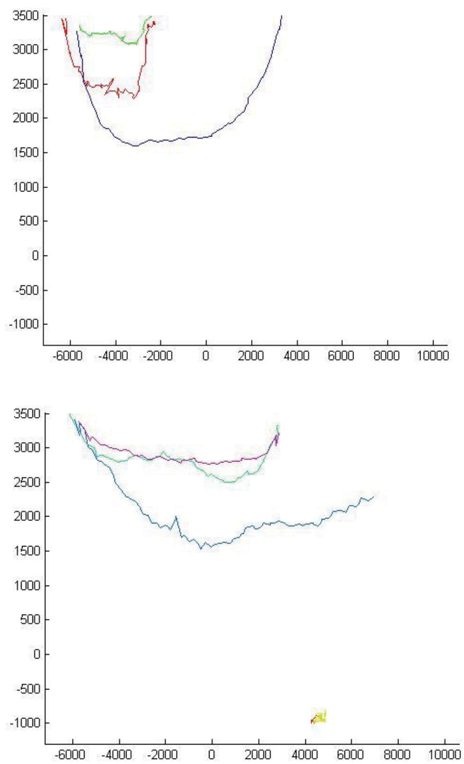


Fig. 8. Two samples of extracted trajectories

After that event analysis is performed on the extracted trajectories. Four different events were initially considered:

E1) Human Entering

E2) Human Exiting

E3) Human Walking

E4) Human Stopping

Human entering is obtained based on two conditions: near the hall entry ports and the start point of trajectory. Human exiting is based on the closeness to the hall exit ports and should be the final point of trajectory. Human walking is extracted based on the moved distance in consecutive frame and the angle of the direction. And finally, the human stopping event is extracted by measuring the distance to the previous and next human locations

In Table II the sample output of event detection is shown. The human angel in Table II is calculated based on positive direction of x-axis.

TABLE II. SAMPLE OF DETECTED EVENTS FOR FRAME 40 TO 47.

Number of Frame	Detected Human	Detected Event	Angel of Human
Frame 40	Human 1	Stop	$0.2\ \pi$
	Human 2	Walk	$-0.7\ \pi$
	Human 3	Enter	-
Frame 41	Human 1	Stop	$0.7\ \pi$
	Human 2	Walk	$-0.7\ \pi$
	Human 3	Walk	$0.1\ \pi$
Frame 42	Human 1	Walk	$0.8\ \pi$
	Human 2	Walk	$-0.8\ \pi$
	Human 3	Walk	$-0.2\ \pi$
Frame 43	Human 1	Walk	$-0.9\ \pi$
	Human 2	Walk	$-0.8\ \pi$
	Human 3	Walk	$-0.4\ \pi$
Frame 44	Human 1	Walk	$0.9\ \pi$
	Human 2	Walk	$-0.8\ \pi$
	Human 3	Stop	$-0.3\ \pi$
Frame 45	Human 1	Walk	$0.9\ \pi$
	Human 2	Stop	$0.4\ \pi$
	Human 3	Walk	$-0.2\ \pi$
Frame 46	Human 1	Walk	$-0.9\ \pi$
	Human 2	Stop	$0.3\ \pi$
	Human 3	Walk	$-0.2\ \pi$
Frame 47	Human 1	Walk	$-0.7\ \pi$
	Human 2	Stop	$0.2\ \pi$
	Human 3	Walk	$-0.2\ \pi$

VI. CONCLUSION

Robots are increasingly entering our daily lives, and cohabitating spaces with humans. However, towards safe and pleasant co-habitation, and towards interaction with single humans or groups of humans, robots should be able to navigate in ways that are human-friendly, and appropriate to human spatial interaction social norms, such as keeping personal spaces. For that purpose, we have developed a special theory which extends path planning, which we call social path planning, which allows humans or groups as special obstacles or special goals in the path planner. In order to provide tuning for our simulation results, we have acquired a natural human interaction dataset, through measurements from multiple laser ranging sensors positioned at a cross-roads indoor space.

In this paper, we have described our system for deriving trajectories and events from these measurements, consisting of a novel processing chain, with spatial and temporal alignment algorithms for multiple laser sensors, as well as foreground detection, sensor data fusion, segmentation, tracking, two-legged position and pose estimation, and event detection stages. The methods presented here can be easily extended to larger spaces and applied for many other application domains beyond our main goal of learning optimal spatial interaction behaviors for human-robot interaction, towards our ultimate goal of daily collaboration and companionship with robots.

REFERENCES

[1] L. Snidaro, C. Micheloni, and C. Chiavedale, "Video security for ambient intelligence," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* vol. 35, pp. 133-144, 2005.

[2] Q. Cai and J. K. Aggarwal, "Automatic tracking of human motion in indoor scenes across multiple synchronized video streams," in *Computer Vision, 1998. Sixth International Conference on*, 1998, pp. 356-362.

[3] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Robotics-DL tentative*, 1992, pp. 586-606.

[4] C. Yang and G. Medioni, "Object modelling by registration of multiple range images," *Image and vision computing,* vol. 10, pp. 145-155, 1992.

[5] J. V. Gómez, N. Mavridis, and S. Garrido, "Social Path Planning: Generic Human-Robot Interaction Framework for Robotic Navigation Tasks."

[6] S. Garrido, L. Moreno, M. Abderrahim, and F. Martin, "Path planning for mobile robot navigation using voronoi diagram and fast marching," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 2376-2381.

[7] H.-M. Gross, A. König, and S. Mueller, "Omniview-based concurrent map building and localization using adaptive appearance maps," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, 2005, pp. 3510-3515.

[8] J. M. Porta and B. Krose, "Appearance-based concurrent map building and localization using a multi-hypotheses tracker," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, 2004, pp. 3424-3429.

[9] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Automatic position calibration and sensor displacement detection for networks of laser range finders for human tracking," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 2938-2945.

[10] K. Schenk, M. Eisenbach, A. Kolarow, and H.-M. Gross, "Comparison of laser-based person tracking at feet and upper-body height," in *KI 2011: Advances in Artificial Intelligence*, ed: Springer, 2011, pp. 277-288.

[11] A. Diosi and L. Kleeman, "Laser scan matching in polar coordinates with application to SLAM," in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, 2005, pp. 3317-3322.

[12] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends® in Computer Graphics and Vision,* vol. 2, pp. 1-104, 2006.

[13] A. Fod, A. Howard, and M. Mataric, "A laser-based people tracker," in *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, 2002, pp. 3024-3029.

[14] K. Nakamura, H. Zhao, R. Shibasaki, K. Sakamoto, T. Ohga, and N. Suzukawa, "Tracking pedestrians using multiple single-row laser range scanners and its reliability evaluation," *Systems and computers in Japan,* vol. 37, pp. 1-11, 2006.

[15] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser-based interacting people tracking using multi-level observations," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 1799-1804.

[16] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser-based detection and tracking of multiple people in crowds," *Computer Vision and Image Understanding,* vol. 106, pp. 300-312, 2007.

[17] M. Mucientes and W. Burgard, "Multiple hypothesis tracking of clusters of people," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 692-697.

[18] K. O. Arras, S. Grzonka, M. Luber, and W. Burgard, "Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 1710-1715.

[19] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People tracking with mobile robots using sample-based joint probabilistic data association filters," *The International Journal of Robotics Research,* vol. 22, pp. 99-116, 2003.

[20] Y. Cheng, "Mean shift, mode seeking, and clustering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 17, pp. 790-799, 1995.

[21] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly,* vol. 2, pp. 83-97, 1955.