2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

Tracking of Facial Features to Support Human-Robot Interaction

Maria Pateraki, Haris Baltzakis, Polychronis Kondaxakis, Panos Trahanias
Institute of Computer Science,
Foundation for Research and Technology - Hellas,
Heraklion, Crete, Greece
{pateraki,xmpalt,konda,trahania}@ics.forth.gr

Abstract— **In this paper we present a novel methodology for detection and tracking of facial features like eyes, nose and mouth in image sequences. The proposed methodology is intended to support natural interaction with autonomously navigating robots that guide visitors in museums and exhibition centers and, more specifically, to provide input for the analysis of facial expressions that humans utilize while engaged in various conversational states. For face and facial feature region detection and tracking, we propose a methodology that combines appearance-based and feature-based methods for recognition and tracking, respectively. For the stage of face tracking the introduced method is based on Least Squares Matching (LSM), a matching technique able to model effectively radiometric and geometric differences between image patches in different images. Thus, compared with previous research, the LSM approach can overcome the problems of variable scene illumination and head in-plane rotation. Another significant characteristic of the proposed approach is that tracking is performed on the image plane only wherever laser range information suggests so. The increased computational efficiency meets the real time demands of human-robot interaction applications and hence facilitates the development of relevant systems.**

I. INTRODUCTION

A key enabling technology for next-generation robots for the service, domestic and entertainment market is Human-Robot-Interaction. A socially interactive robot, i.e a robot that collaborates with humans on a daily basis (be this in care applications, in a professional or private context) requires interactive skills that go beyond keyboards, button clicks or metallic voices. For this class of robots, human-like interactivity is a fundamental part of their functionality. Some of the greatest challenges towards this goal are related to how robots perceive the world. As pointed out in [1], in order to interact meaningfully with humans, a socially interactive robot must be able to perceive, analyze and interpret the state of the surrounding environment and/or humans in a way similar to the way humans do. In other words, it must be able to sense and interpret the same phenomena that humans observe.

Unlike humans that mostly depend on their eyes, most current robots, in addition to vision sensors, also utilize range sensors like sonars, infrared sensors and laser range scanners. Approaches based on range sensors are very popular for tasks like autonomous navigation [2], [3], [4], [5] and 2D people tracking [6]. The main advantage of such sensors over vision ones is that they are capable of providing accurate range measurements of the environment in large angular fields and at very fast rates. On the other hand, for tasks like gesture recognition and face detection, i.e tasks that require richer information (e.g intensity, color) or information beyond the 2D scanning plane of a typical range sensor setup, vision is the only alternative [7].

In this paper we present a novel methodology for detection and tracking of facial features like eyes, nose and mouth in image sequences. The proposed approach is intended to support natural interaction with autonomously navigating robots that guide visitors in museums and exhibition centers and, more specifically, to provide input for the analysis of facial expressions that humans utilize while engaged in various conversational states. The operational requirements of such an application challenge existing approaches in that the visual perception system should operate efficiently under unconstrained conditions regarding occlusions, variable illumination, moving cameras, and varying background. The proposed approach combines and extends multiple state-of-the art techniques to solve a number of related subproblems like (a) detection and tracking of people in both the ground plane and the image plane, (b) detection and tracking of human faces on the image plane and, (c) tracking of specific facial features like eyes, nose and mouth on the image plane.

People tracking, given the constrains of the application at hand, is a very challenging task by itself. This is because the applied method must be computationally efficient, in order to perform in almost real-time, and robust in the presence of occlusions, variable illumination, moving cameras and varying background. A thorough survey on vision-based approaches to *people tracking*, can be found in [7]. The referenced methods rely to a great extent on visual detection of head or face and tend to be time-consuming and less robust in uncontrolled lighting conditions. Laser-based detection and tracking can provide a more reliable automatic detection of humans in dynamic scenes, using one [8], [9], [10] or multiple registered laser scanners [11]. However, the lack of color information, causes difficulties in laser- based methods, e.g. to maintain tracked trajectories of different objects when occlusions occur. Therefore, the combination of distance and angle information, obtained from a laser scanner with visual information, obtained from a camera, could support vision-

This work has been partially supported by the EU Information Society Technologies research project INDIGO (FP6-045388). and the Greek national GSRT project XENIOS.

based methods for faster and more reliable human tracking. In the field of robotics, *"hybrid"* methods combining laser and camera data have appeared recently, and in [12] representative methods, e.g. [13], [14], are discussed.

Tracking of human faces and facial features on the image plane constitutes another challenging task because of face variability in location, scale, orientation (up-right, rotated), pose (frontal, profile), age and expression. Furthermore, it should be irrespective of lighting conditions and scene content. Detection can be based on different cues: skin color (color images/videos), motion (videos), face/head shape, facial appearance, or a combination of them. Comprehensive surveys on face detection and tracking are [15], [16]. Appearance- based methods avoid difficulties in modeling 3D structures of faces by considering possible face appearances under various conditions with AdaBoost learning-based algorithms, e.g. [17], [18], [19], to be the most effective so far. Color-based systems may be computationally attractive but the color constraint alone is insufficient for achieving high accuracy face detection, mainly due to large facial color variation in different lighting conditions and humans of different skin color. Other methods primarily based on color models, e.g. [20] may prove more robust in laboratory environments but in unconstrained lighting and illumination still their performance is limited and are less suitable to derive head rotations.

The most important contribution of this paper is related to the methodology used for for face and facial feature region detection and tracking which combines appearance-based and feature-based methods for recognition and tracking, respectively. For the stage of face tracking the introduced method is based on Least Squares Matching (LSM), a matching technique able to model effectively radiometric and geometric differences between image patches in different images. Compared with previous research, the LSM approach can overcome the problems of variable scene illumination and head in- and off-plane rotations.

Another significant characteristic of the proposed methodology is that visual people tracking is performed only wherever laser range information suggests so. The increased computational efficiency meets the real time demands of the specific application at hand and facilitates its application to other crucial robotics tasks as well. Moreover, since information encapsulated in visual data acts supplementary to laser range information, inherent advantages of both sensors are maintained, leading to implementations combining accuracy, efficiency and robustness at the same time.

The proposed methodology was tested extensively with a variety of real data gathered with two different robotic platforms, both in laboratory and in real museum/exhibition setups. The results obtained are very promising and demonstrate its effectiveness.

II. Methodology Overview

The basic idea in the proposed methodology is to sufficiently exploit the detection capability of laser scanners and to combine visual information (both grey-level and color images can be used) to the detection results to track people, as well as face and facial features in dynamic scenes. The temporal detection of humans relies on laser-based detection and tracking of moving objects (DATMO), i.e. of moving legs, and registration of vision-based information to localize the expected face region. After human detection, the face and facial features are detected and tracked over time. The tracking information of facial features is used for a later analysis of conversational states.

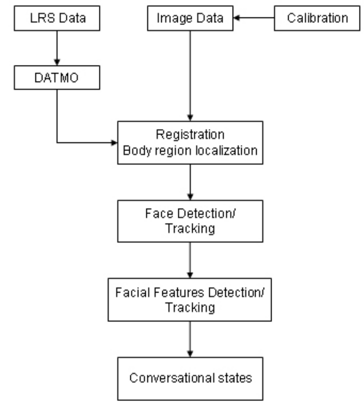


Fig. 1. Overview of laser/vision based system

The methodology is schematically depicted in Fig. 1, where it can be seen that once moving objects have been detected using the Laser Range Scanner (LRS) data, humans are tracked by integrating calibrated camera information and field of view, distance from laser scanner (baseline) and minimum and maximum man height. The face that is frontal and closest to the camera is detected and its position and rotation are tracked. Within the enclosing area of the detected face, the facial subregions, i.e. eyes, mouth and nose, are then detected by imposing anthropometric constraints. Tracking of these features is done by area-based matching approaches. In both the tracking of face and facial regions quality measures of the matching result are given and evaluated. These measures will determine if the final tracked result should be accepted or not.

III. Laser-based Detection and Tracking of Moving Targets

To extract and track multiple moving targets (e.g. legs) from stationary background (e.g. building walls, desks, chairs, etc.) a Joint Probabilistic Data Association with Interacting Multiple Model (JPDA-IMM) algorithm is employed. Its robustness, in comparison to other techniques, is thoroughly presented in [6] and here the consecutive steps are briefly described. Initially, a single occupancy grid is created and every individual grid-cell accumulates a counter, which increases linearly when a laser measurement falls inside its occupancy area. Grid-cells with values above a dynamic threshold level are selected as stationary background. A certain background is obtained and subtracted from every laser frame leaving only the measurements that represent possible moving targets. The remaining measurements are clustered into groups and a center of gravity is assigned to each one. Finally, the JPDA-IMM initiates and allocates tracks to clusters that exceed a certain velocity level.

The technique effectively distinguishes targets which move at close proximity to each other, being also able to compensate the relative movement of the robot. The identified moving targets are treated as potential leg candidates, considering also that, apart from humans, other moving objects may appear in the scene.

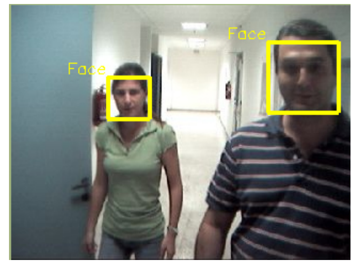


Fig. 3. Detection of faces

IV. Vision-based System

The camera is adjusted on the same robotic platform as the LRS, at an appropriate height to view humans face on, and its optical axis parallel to the LRS pointing direction. The camera is calibrated using Zhang's method [21], and calibration information is utilized at a later stage for LRS and vision data registration.

A. Human detection

Camera calibration parameters, image resolution, frame rate, known baseline between the LRS and the video camera are used to register the image with the LRS data. The laser points, indicated as leg candidates, are projected on the visual image plane and by including information on the minimum and maximum human height, e.g. 1.20 m and 1.90 m, respectively, the expected face region is localized in the image plane. The results of the method are demonstrated in Fig. 2(a), where the ellipses mark the four moving objects, corresponding to four legs, and in Fig. 2(b), where the expected face region is localized using the registered LRS and vision data and the human height constraints.

B. Face Detection and Tracking

Following body region localization the faces are detected within the given region in order to reject possible outliers arising after LRS processing and verify that there are people moving towards the robot. We assume that for the initial detection of the person interacting with the robot we have frontal views of the person. However, we still have to tackle the issues of in- and off-plane head rotations in face tracking, important in the analysis of communicative signs and variable scene illumination. With respect to illumination, our aim is to place the robot in environments with unconstrained lighting conditions.

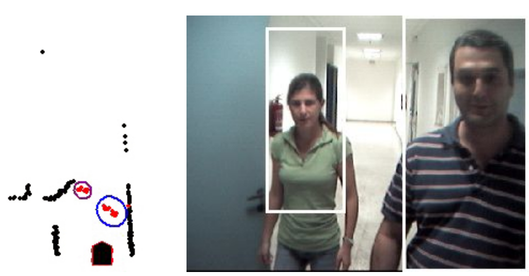
We utilize a hybrid approach by integrating an appearance-based approach for face detection and a feature-based approach for face tracking. In the introductory part the advantages of appearance-based methods have been already pointed out. The robust face detector developed by Viola and Jones [19] is employed in this work. The named detector combines four key concepts; Haar features, integral image for rapid feature detection, the Adaboost machine-learning method and a cascaded classifier to combine many features efficiently. Unfortuantelly, this approach suffers from two significant limitations: (a) inability to handle significant in-plane rotations (i.e. rotations of 30 degrees or more), and (b) increased processing time. Although some recent approaches (e.g. [18], [22]) tackle with the first limitation (inability to track in-plane rotations) using an extended set of rotated Haar-like features, the required computational power still prohibits their use to applications that involve higher framerates and/or higher image resolutions. Moreover, they aim in detecting faces in every input frame without maintaining ids of specific persons, i.e they do not perform face tracking.

Therefore, in our method the face is detected in the initial frame using the Haar method and tracked in the subsequent frames with the LSM approach, described in the next paragraph. Only if the LSM tracker fails to converge to a solution, the Haar detector will re-initialize the process. In Fig. 3 the rectangles mark the detected faces on the image by the Haar method.

C. A Least-Squares Approach to Face Tracking

Cross-correlation is based on the assumption that geometric differences are modeled only by translation, and radiometric differences exist only due to brightness and contrast. Thus, its precision is limited, decreases rapidly if the geometric model is violated (rotations greater than 20° and scale differences between images greater than 30%). A generalization of cross correlation is Least Squares Matching (LSM) [23], which in its general approach can compensate geometric differences in rotation, scale and shearing.

Several approaches exist in the current literature, mainly from the photogrammetric community, using least squares for image registration, calibration, surface reconstruction etc. To the best of our knowledge, the iplementation of LSM described in this paper, is introduced for the first time face tracking in a robotic application.



(a) (b)

Fig. 2. Registered LRS and vision data. (a) The detected moving objects are marked within the ellipses, (b) the localized expected face areas

The formulation of the general estimation model is based on the assumption that there are two or more image windows (called image patches) given as discrete functions $f(x,y), g_i(x,y), \quad i = 1,...n-1$, wherein f is the template and g the search image patch in $\quad i = 1,...n-1$ search images. The problem statement is finding the corresponding part of the template image patch $f(x,y)$ in the search images $g_i(x,y), \quad i = 1,...n-1$.

$$f(x,y) - e_i(x,y) = g_i(x,y) \tag{1}$$

Equation (1) gives the least squares grey level observation equations, which relate the $f(x,y)$ and $g_i(x,y)$ image functions or image patches. The true error vector $e_i(x,y)$ is included to model errors that arise from radiometric and geometric differences in the images. For the selection of the geometrical model it is assumed that the object surface is approximated by local planar facets and an affine transformation is generally used. Radiometric corrections (e.g. equalization), for compensation of different lighting conditions are not included in the model but are applied during LSM.

In our implementation we use two images and the affine transformation is applied with respect to an initial position x_0, y_0:

$$\begin{aligned} x &= a_0 + a_1 \cdot x_0 + a_2 \cdot y_0 \\ y &= b_0 + b_1 \cdot x_0 + b_2 \cdot y_0 \end{aligned} \tag{2}$$

After linearization of the function $g(x,y)$, (1) becomes:

$$f(x,y) - e(x,y) = g(x_0,y_0) + \frac{\partial g(x_0,y_0)}{\partial x} \cdot dx + \frac{\partial g(x_0,y_0)}{\partial y} \cdot dy \tag{3}$$

With the simplified notation:

$$g_x = \frac{\partial g(x_0,y_0)}{\partial x}, g_y = \frac{\partial g(x_0,y_0)}{\partial y}$$

and by differentiating (2), then (3) results in:

$$\begin{aligned} f(x,y) - e(x,y) = g(x_0,y_0) + g_x da_0 + g_x x_0 da_1 + g_x y_0 da_2 + \\ + g_y db_0 + g_y x_0 db_1 + g_y y_0 db_2 \end{aligned} \tag{4}$$

with the parameter vector x being defined as:

$$x^T = (da_0, da_1, da_2, db_0, db_1, db_2) \tag{5}$$

The least squares solution of the system is given by (6):

$$\hat{x} = (A^T P A)^{-1} (A^T P l) \tag{6}$$

where $\hat{x}$ is the vector of unknowns, A is the design matrix of grey level observation equations, P is the weight matrix, and l is the discrepancy vector of the observations. The weights are typically diagonal with elements set to unity for all the grey level observation equations. The number of grey level observations is related to the size of the template size. If e.g. a patch size of 9x9 is selected then the number of observation equations is 81.

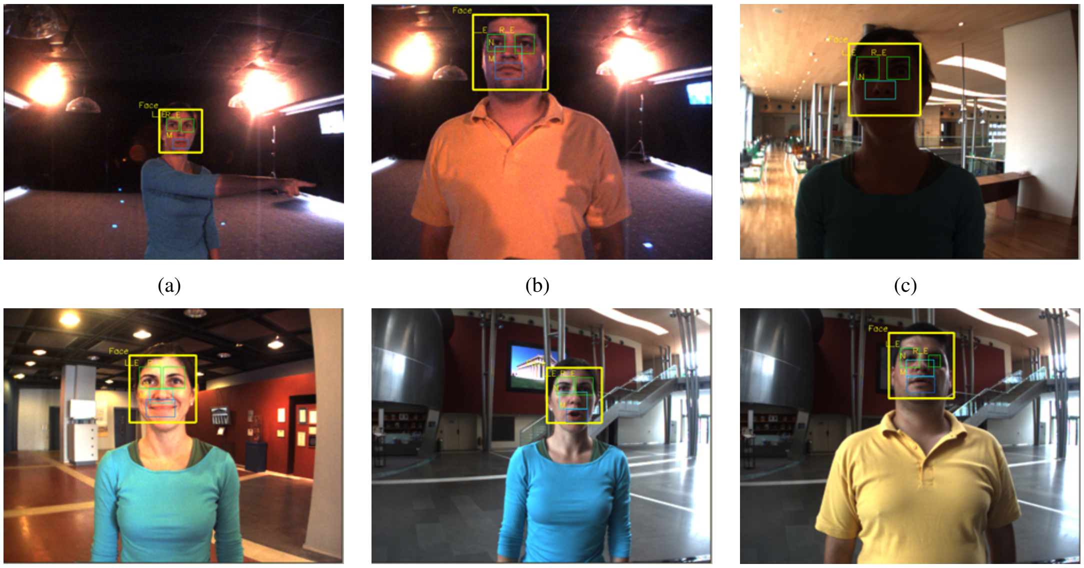
In our implementation of LSM for face tracking, the affine is constrained to conformal transformation to avoid over-parametrization of the system since only estimation of shifts, rotation and scale suffice to model the geometric differences of frontal faces in frame sequence. The geometric differences refer to: (a) face scaling, when the person moves closer or away from the robot, (b) head in-plane rotations and (c) head off-plane rotations. It is known that if there is insufficient signal content or if the initial approximations in the least squares solution are not close to the real solution, the solution will not converge. These issues can be easily handled in face tracking. Initial approximations for shifts are taken from the center of the face area detected with the Haar method that initialized the process of face localization. The template used for temporal matching, is initialized to the center of the detected face and scaled at 75% of its area. Equalization is applied both for the template and the search area during LSM to compensate for radiometric differences. The template is updated when solution converges and the process continues. Variable illumination poses less an issue since the search area expands around the initial patch at a maximum of half the size of the largest dimension of the patch. The only drawback is the size of the template when the face is very close to the camera, increasing the number of observation equations. A proposed solution is to apply LSM in images of lower resolution and transform the matching results to the original level.

As far as quality is concerned, the criteria used to evaluate matching results are the number of iterations, the alteration of the size of parameters in each iteration and the size of parameters. The number of iterations is a rather good criterion, assuming that the approximations are good. In parallel, variations in the parameter values (magnitude and sign) in each iteration have to be observed in order to evaluate the stability of the solution. The threshold for the iterations should not be set too high (maximum number of 15 iterations), considering that fast convergence should be achieved, since the initial values are close to the correct solution. The size of parameters, especially the estimated value for scales, should not exceed an upper (> 3.0) and lower value (< 0.3) and the difference to their initial values should be small. The initial values for scales are set to 1. The variation of x, y coordinates from their initial values is also checked, considering the utilized frame rate and if it is above a certain threshold the point is rejected.

D. Facial Feature Region Detection

Eyes, nose and mouth are the facial regions that are detected. However, the image resolution of the face is an important factor in facial feature region detection. When the face area is smaller than 70 x 90 pixels the facial regions become hard to detect [24]. As in the case of face detection, individual sets of Haar-like features for each region are used to detect the eyes, mouth and nose area, within the detected and tracked face region using the method from [19].

False detections may arise especially in faces of larger scale or higher resolution and anthropometric constraints are imposed for reliability of the solution. The eye, mouth and nose regions should be found in the upper half, lower half and in the central part of the face respectively, while the ratio of their respective widths to the width of the face by coarse



(a) (b) (c)

(d) (e) (f)

Fig. 4. Examples of face and facial features tracking results under variable lighting conditions and background

approximation should be close to 0.4, empirically learned from normalized measurements in a number of frontal faces. These constraints proved to enforce detection.

E. Facial Feature Region Tracking

The detected facial feature regions are tracked in the subsequent video frames by using cross-correlation of image patches. This approach is justified as there are only small deviations in the relative positions of these feature areas with respect to the position of the detected and tracked face within the image. The previously detected eye, nose and mouth regions are used as templates in the matching process. The process is computationally efficient, since Haar detection is only used for the initial detection, whereas the templates are taken from the same object, used in the tracking procedure.

V. Experimental Results

The method proposed in this paper has been implemented and assessed on two different robotic platforms in various laboratory and real application setups.

For laboratory testing, we utilized an iRobot-B21r platform equipped with a SICK-PLS laser range finder and a low-end camera operating at a standard resolution of 640×480 pixels. The range finder is capable of scanning 180 degrees of the environment, with an angular resolution of one measurement per degree and a range measuring accuracy of 5cm.

The platform that was utilized to collect data from the actual exhibition place, was a Neobotix NE-470 robotic platform equipped with a PointGray Bumblebee2 stereovision head, operating at the same, 640×480 resolution and a Sick S300 laser range finder. This specific range finder is capable of achieving an angular resolution of two measurements per degree and its placement in front of the robot ensured a field of view of about 210 degrees.

Prior to testing our methodology, an internal calibration procedure has been applied to both robots in order to estimate the relative positions and the intrinsic parameters of all sensors.

Examples from the human tracking results have been previously shown in Figs. 2 and 3.

The proposed methodology for face and facial feature detection and tracking was tested with the different video sequences recorded in the laboratory, as well as in various locations at the exhibition place. Fig. 4 shows results from the video sequences, recorded at various locations of the exhibition place. As it can be seen the method is able to handle severe illumination and background conditions, yet extract face and facial features reliably. Images in Fig. 4(a) and 4(b), were recorded in a room with low lighting and monitors in the background, that pose a problem for methods employing background subtraction. Even in the case of strong background lightning, as in Fig. 4(b), feature areas can be tracked. Moreover, our method is able to handle (a) dynamic backgrounds, e.g. Figs. 4(d), 4(e), 4(f), (b) scene objects being very close in color to skin, e.g. Figs. 4(b) and 4(f) (clothes in similar color to skin).

In addition the comparative advantage of the LSM method for face tracking over other commonly used methods is demonstrated. LSM is able to compensate geometric and radiometric differences between image patches. In Fig. 5, results of LSM versus pure Haar-based face tracking and the CMU tracker are shown. The indicative images are selected from a video sequence recorded in the laboratory. As can be easily observed, in this result, the Haar-based method fails to track the face when in-plane rotations occur. The CMU method also fails to provide reliable results for position as well as for rotation, whereas the LSM provides a solution, along with certain measures to evaluate the tracking result.

In all experiments conducted, including the ones presented above, the LSM tracking operated at a frame rate of 30 fps, on a Pentium Core Duo 2.8 GHz.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a novel methodology for robust detection and tracking of human faces and facial features in image sequences, intended for human-robot interaction applications.

According to the proposed methodology, the 3D locations of people, as produced by a people tracker that utilizes laser range data to track people on the ground plane, are projected on the image plane in order to identify image regions that may contain human faces. A state-of-the-art, appearance-based method is used in order to specifically detect human faces within these regions and initiate a feature based tracker that tracks these faces as well as specific facial features over time.

Experimental results have confirmed the effectiveness and the increased computational efficiency of the proposed methodology, proving that the individual advantages of all involved components are maintained, leading to implementations that combine accuracy, efficiency and robustness at the same time.

We intend to use the proposed methodology in order to support natural interaction with autonomously navigating robots that guide visitors in museums and exhibition centers. More specifically the proposed methodology will provide input for the analysis of facial expressions that human utilize while engaged in various conversational states.

Future work includes extension of the LSM temporal tracking to handle stereo vision by exploiting epipolar constraints. Moreover, the methodology presented in the paper will be employed in an integrated system for naturalistic human-robot interaction.

Fig. 5. Comparison of Haar, CMU and LSM face tracker in the presence of in-plane rotation. Results from the Haar-based detection in (a), (b), (c), the CMU tracker in (d), (e), (f) and the LSM tracker in (g), (h), (i).

REFERENCES

[1] T. Fong, I. Nourbakhsh, and K. Dautenhahn, “A survey of socially interactive robots,” *Robotics and Autonomous Systems*, vol. 42, pp. 143–166, 2003.

[2] J. Castellanos, J. Tardós, and J. Neira, “Constraint-based mobile robot localization,” in *Proc. Intl. Workshop on Advanced Robotics and Intelligent Machines*, University of Salford, Manchester, U.K., 1996.

[3] S. Thrun, A. Buecken, W. Burgard, D. Fox, T. Froehlinghaus, D. Hennig, T. Hofmann, M. Krell, and T. Schmidt, “Map learning and high-speed navigation in RHINO,” University of Bonn, Department of Computer Science, Technical Report IAI-TR-96-3, July 1996.

[4] J.-S. Gutmann and K. Konolige, “Incremental mapping of large cyclic environments,” in *Proc. IEEE Intl. Symposium on Computational Intelligence in Robotics and Automation, (CIRA)*, Monterey, CA, 1999, pp. 318–325.

[5] H. Baltzakis and P. Trahanias, “A hybrid framework for mobile robot localization. formulation using switching state-space models,” *Autonomous Robots*, vol. 15, no. 2, pp. 169–191, 2003.

[6] P. Kondaxakis, S. Kasderidis, and P. Trahanias, “A multi-target tracking technique for mobile robots using a laser range scanner,” in *Proc. IEEE Intl. Conf. on Robots and Systems (IROS’08)*, Nice, France, 2008.

[7] D. Gavrilla, “The visual analysis of human movement: A survey,” *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.

[8] A. Fod, A. Howard, and M. Mataric, “Laser-based people tracking,” in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA’02)*, Washington, DC, USA, 2002, pp. 3024–3029.

[9] D. Schulz, W. Burgard, D. Fox, and A. Cremers, “Tracking multiple moving objects with a mobile robot,” in *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR’01)*, 2001, pp. 371–377.

[10] B. Kluge, C. Koehler, and E. Prassler, “Fast and robust tracking of multiple moving objects with a laser range finder,” in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA’01)*, 2001, pp. 1683-1688.

[11] H. Zhao and R. Shibasaki, “A novel system for tracking pedestrians using multiple single-row laser range scanners,” *IEEE Trans. Syst., Man, Cybern. A*, vol. 35, no. 2, pp. 283– 291, 2004.

[12] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, “Multi-modal tracking of people using laser scanners and video camera,” *Image and Vision Computing*, vol. 26, no. 2, pp. 240–252, 2008.

[13] Z. Byers, M. Dixon, K. Goodier, C. Grimma, and W. Smart, “An autonomous robot photographer,” in *Proc. IEEE Intl. Conf. on Robots and Systems (IROS’03)*, Las Vegas, Nevada, USA, 2003, pp. 2636–2641.

[14] M. Scheutz, J. McRaven, and G. Cserey, “Fast, reliable, adaptive bimodal people tracking for indoor environments,” in *Proc. IEEE Intl. Conf. on Robots and Systems (IROS’04)*, Sendai, Japan, 2004, pp. 1347–1352.

[15] M.-H. Yang, D. J. Kriegman, and N. Ahuja, “Detecting faces in images: A survey,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 1, pp. 34–58, 2002.

[16] E. Hjelmas and B. K. Low, “Face detection: A survey,” *Computer Vision and Image Understanding*, vol. 3, no. 3.

[17] S. Li and Z. Zhang, “Floatboost learning and statistical face detection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 9.

[18] R. Lienhart, A. Kuranov, and V. Pisarevsky, “Empirical analysis of detection cascades of boosted classifiers for rapid object detection,” Intel Labs, Amherst, MA, MRL Technical Report, 2002.

[19] P. Viola and M. Jones, “Robust real-time face detection,” *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[20] F. J. Huang and T. Chen, “Tracking of multiple faces for human-computer interfaces and virtual environments,” in *IEEE Intl. Conf. on Multimedia and Expo.*, 2000.

[21] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[22] C. Messom and A. Barczak, “Fast and efficient rotated haar-like features using rotated integral images,” in *Proc. of 2006 Australian Conference on Robotics and Automation (ACRA ’06)*, 2006.

[23] Ackermann F., “High Precision Digital image Correlation,” in *Photogrammetric Week, Heft 9*. Universität Stuttgart, 1984.

[24] Y. Tian, “Evaluation of face resolution for expression analysis,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW’04)*. IEEE Computer Society, 2004, pp. 82–89.