Proceedings of the 2008 IEEE/ASME
International Conference on Advanced Intelligent Mechatronics
July 2 - 5, 2008, Xi'an, China

Lux - An Interactive Receptionist Robot for University Open Days

Nicola Bellotto, Susanna Rowland, Huosheng Hu

***Abstract*— Interactive service robots are ready to serve us both in public and private places soon. This paper presents a multisensor framework to the simultaneous tracking and recognition of humans so that a service robot can implement reception tasks interactively. Both a laser range scanner and a camera have been deployed in the framework. By fusing range and visual data, the robot can detect and identify applicants in the university environment in order to provide dedicated services. The robustness of the robot perception is increased by the joint tracking and recognition of face, clothes and height, implemented using a bank of filters and probabilistic data association. Experiments in the university open days demonstrate the effectiveness of the proposed solution.**

I. INTRODUCTION

As the advancement of artificial intelligence and intelligent robots, a growing number of mobile service robots has been deployed in human environments for entertainment or other applications, including the robot for elderly and patient care [1], tour-guide for visitors [2], security tasks [3] and so on. However, in many cases, these service robots are unable to recognize and remember the persons being served, and often got confused if it is expected to serve multiple people. Therefore. the identity of the people in the robot environment becomes very important for a service robot to implement user-oriented tasks and other advanced behaviors.

In general, a successful service robot should be able to track and recognize moving people in order to provide service, and also know their specific needs. Several biometric features can be observed, using different sensors, to recognize people. Vision-based solutions are obviously the most common, and include human identification from face, iris, ear shape, gait or clothes. Other sensors are also used for the recognition of fingerprints, voices, odors, etc. [4]. Since cameras are installed on almost every modern robot, their people recognition is mainly vision-based.

The techniques adopted in most of the recognition systems is usually vision-based, and consists in selecting a video frame where the subject to recognize satisfies some criteria, like pose, size or number of visible features, using then some standard comparison versus a database of known people. The solution implemented in this paper, instead, makes use of the tracking estimation to integrate spatial and time information during the recognition process. The task is performed with a bank of Bayesian estimators that fuses laser and visual data to estimate, simultaneously, position and identity of the people in the environment. Eventually, this information is used by our service robot, Lux, to introduce the facilities and courses to the applicants during the university open days.

Nicola Bellotto, Susanna Rowland and Huosheng Hu are with Department of Computing and Electronic Systems, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom nbello@essex.ac.uk, slmrow@essex.ac.uk, hhu@essex.ac.uk

Fig. 1. Lux, the interactive service robot.

The remainder of the paper is organized as follows. Section II introduces the experimental platform used in this research, including the interaction algorithm and the multisensor solution to human detection and recognition. Section III summarizes the problem of joint target tracking and identification, illustrating then the implementation of the bank of filters and the procedure for data association adopted. The performance of the system in real situations are illustrated by some experiments in Section IV. The paper finally concludes with a summary of the progresses achieved in Section V, and with suggestions for future research.

II. EXPERIMENTAL PLATFORM

The aim of this research is to create a suitable human-robot interaction system for use at university open days to provide potential students with required information in a pleasant and friendly way. The system will response to user input from a GUI as well as some basic spoken commands. The information provided will be both texture and spoken formats based. It should respond to the user movement, turning to face the user thus making the system appear more lifelike.

A. System Configuration

Fig. 1 shows Lux, our interactive service robot, which is based on a Scitos G5 mobile platform provided with SICK laser, color camera, touch screen and iCat on the top. Both touch screen and iCat animations are used for interaction, providing visual information and generating speech

978-1-4244-2495-5/08/$25.00 © 2008 IEEE.

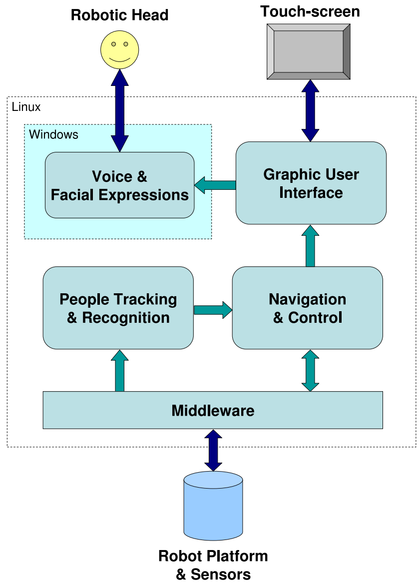


Fig. 2. System configuration for the robot.

and facial expressions. The on-board PC is a Core Duo 1.66GHz running Linux OS, on which most of the software modules are running, as illustrated in Fig. 2. These are the tracking and navigation modules, written in C++, plus the GUI module, implemented in Java. The Player[1] middleware provides the interface to communicate with the robot and its devices, including laser and camera. The iCat instead is controlled by scripts running in Windows, so the latter is installed on a VMWare[2] virtual machine for Linux. All the modules are independent processes that communicate through TCP connections.

B. Interaction Algorithm

The following is a basic run through of the system and how it responds to various forms of input from the user, as well as how the system proceeds while no input is received both before and after interaction has been started. For a diagram of the algorithm please see Fig. 3. All activities within the dashed line occur only once user interaction has commenced. If the robot locates a potential user, a string is sent via TCP to the interaction system, and if it is not currently being interacted with, a greeting message is output through the iCat. This message is personalized if the user has been identified.

The Countdown Timer is started once interaction has begun, and runs separate from the main system; the timer is set to loop for approximately two minutes. When the timer reaches zero, it is assumed that interaction has ceased, and the interface is reset to the Start screen shown in Fig. 4 . Each time a button on the user interface is pressed by the user, the display is updated and, if required, a command is sent via a TCP connection to the animation module controlling the iCat.

[1]http://playerstage.sourceforge.net
[2]http://www.vmware.com

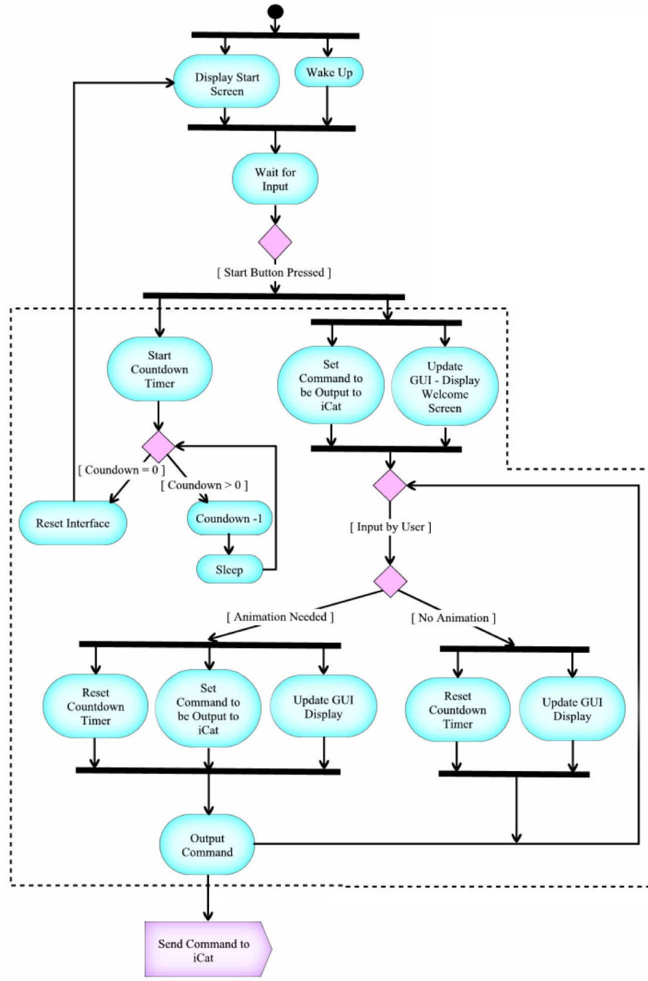


Fig. 3. System activity diagram.

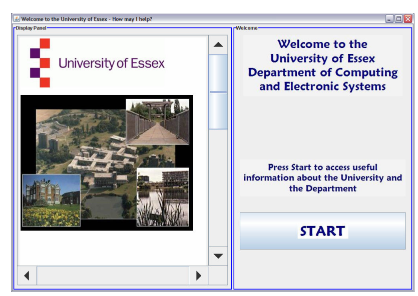


Fig. 4. Start screen of the GUI.

C. Multisensor Detection

Lux can observe people using two different sensors, which are a laser range finder, mounted a few decimeters from the floor, and a colour camera, installed on the top of the robot. These devices are used to detect and recognize humans in the surroundings as explained next. Thanks to an efficient algorithm for human legs detection [5], people in front of the robot can be located in the range $\pm 90°$. Also, the system is able to recognize different legs postures, namely legs-apart, forward-straddle and legs-together (or single-leg), simplifying the discrimination of false positive.

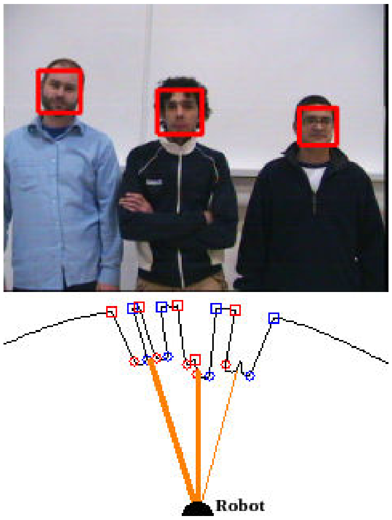


Fig. 5. Legs and face detection using laser and camera.

The camera on the top, approximately 1.70m high, can detect faces thanks to a popular algorithm for object recognition [6]. Using a simple pin-hole camera model, information like direction and size of the detected faces can be derived and later used for people tracking. The illustration in Fig. 5 shows an example of example of human detection with different subjects in different poses.

D. Human Recognition

Histogram comparison is a common technique used for the identification of people wearing different clothes. From the current camera frame, the region containing all or part of the human body must be selected. There are of course several difficulties, for example regarding varying light conditions. Another challenging problem is the correct selection of the region of interest, in particular when both robot and people are moving. If this region is not accurate, the histogram considered might be completely wrong, with a consequent recognition failure. To reduce the probability of errors, in a previous work we implemented a selection procedure that takes into account the uncertainty of the human position estimate, extracting then the best matching histogram and the direction where this occur [5].

Once detected, faces are recognized using PCA and standard MahCosine distance [7], [8]. Before the actual recognition, a pre-processing of the face image is performed. This consists in locating the eyes and scaling the face to a fixed size (i.e. 24×24 pixels in our case), applying then an elliptical mask and equalizing the image using a CLAHE procedure [9]. The height of the subject being tracked (more precisely, the height of his/her face center) is also determined combining range and elevation measurements, provided respectively by laser and camera. Human height, indeed, is also an important biometric information that, in conjunction with face and clothes identification, can be used to improve the recognition performance.

III. JOINT PEOPLE TRACKING AND RECOGNITION

At time step k, the tracking system provides the estimate $\mathbf{x}_k = [x, y, z, \phi, v]_k^T$, where (x, y) is the 2D position, z is the face's height, ϕ the orientation and v the velocity of the human target. Such quantities can be estimated using a classic Bayesian filter, like UKF [10], that integrates the anonymous observations provided by the legs and the face detection. However, since we want to know also the identity of the subject being tracked, the joint state $\mathbf{x}_k^i = \{\mathbf{x}_k, c_i\}$ must be considered, where c_i is a time-invariant feature that depends on the identity of the person. For the purpose, a solution based on a bank of filters (BoF) can be implemented.

A. Joint Estimation

The recursive equations for the joint estimation of the position and the identity can be written as follows [11]:

$$\begin{aligned} p(\mathbf{x}_k^i|\mathbf{Z}_{k-1}) &= \int p(\mathbf{x}_k, \mathbf{x}_{k-1}^i|\mathbf{Z}_{k-1})\ d\mathbf{x}_{k-1} \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}^i) p(\mathbf{x}_{k-1}^i|\mathbf{Z}_{k-1}) d\mathbf{x}_{k-1} \quad (1) \\ p(\mathbf{x}_k^i|\mathbf{Z}_k) &= \frac{p(\mathbf{z}_k|\mathbf{x}_k^i, \mathbf{Z}_{k-1})\ p(\mathbf{x}_k^i|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \quad (2) \end{aligned}$$

where the denominator in (2) is just a normalization factor. The identity probability can also be updated recursively as follows:

$$\begin{aligned} p(c_i|\mathbf{Z}_k) &= \frac{p(\mathbf{z}_k|\mathbf{Z}_{k-1}, c_i)\ p(c_i|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \quad (3) \\ &= \frac{\lambda_k^i\ p(c_i|\mathbf{Z}_{k-1})}{\sum_{i=0}^{N} \lambda_k^i\ p(c_i|\mathbf{Z}_{k-1})} \quad (4) \end{aligned}$$

In case of Kalman filters, the likelihood function $\lambda_k^i = p(\mathbf{z}_k|\mathbf{Z}_{k-1}, c_i)$ corresponds to the "mode likelihood function" of the static Multiple Model estimator [12]:

$$\lambda_k^i = \mathcal{N}\left(\nu_k^i; \mathbf{0}, \mathbf{S}_k^i\right) \quad (5)$$

where ν_k^i is the innovation and $\mathbf{S}_k^i$ the relative covariance matrix. Note that usually (5) requires linear-Gaussian assumptions, but in practice the same expression is used even when these assumptions do not hold.

B. Bank of Filters

The simultaneous people tracking and recognition is performed by the system illustrated in Fig. 6. The solution is based on a BoF, where sensor data are processed by the legs and the face detector to provide positional information to all the estimators. Each one of these corresponds to a single human identity, among all the possible ones stored in the database, and receives therefore information from its particular face or clothes recognizer.

All the detectors and recognizers are described by their own observation models [5]. The MahCosine distance of the face recognition, which is simply modeled as $s_k = n_k^s - 1$ with n_k^s zero-mean Gaussian noise, is added to the observation vector of the face detection. Inside the prediction

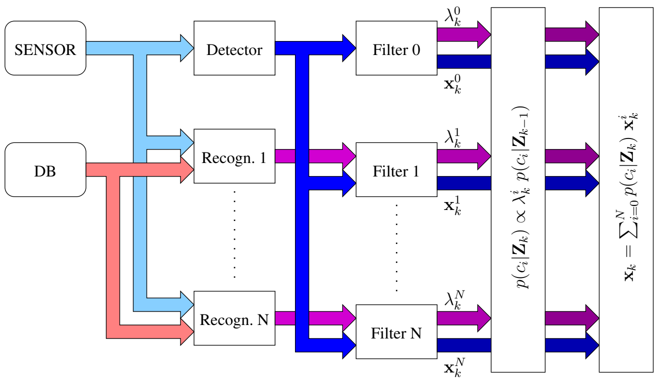


Fig. 6. BoF for joint people tracking and recognition.

model, which is an extension of the standard CV model [5], each estimator of the BoF contains also the height of the relative subject, as given by the database. Faces and histograms of the N subjects are also stored in the same database.

Note that the system illustrated in Fig. 6 includes also a special estimator, called "zero-filter". This generates the likelihood λ_k^0 for the probability of the subject to be unknown. The prediction model of the zero-filter is similar to the other estimators, except that the height component, which is obviously unknown, is modeled as a variable discrete-time Wiener process. The relative clothes observation, also undefined, is given by a "virtual" measurement formed by the predicted direction and a constant histogram distance $d^* = 2\ \sigma_d$, where σ_d is the standard deviation of the noise usually affecting this distance. This value sets a threshold on the clothes observation, assuring that only strong histogram detections influence the identity probability. A similar approach is used for the face recognition, forcing a constant MahCosine distance $s^* = 2\ \sigma_s - 1$ on the face observation vector.

At each time step k, all the filters are updated by the current observations. Also, the identity probabilities are calculated using (3), the maximum determining the current subject's identity. These probabilities are also used to calculate the following weighted estimate:

$$\mathbf{x}_k = \sum_{i=0}^{N} p(c_i|\mathbf{Z}_k)\ \mathbf{x}_k^i \tag{6}$$

which is the current position of the person being tracked. The latter is also necessary to the data association procedure for generating the expected observations, in order to handle multiple human targets.

C. Probabilistic Data Association

Joint Probabilistic Data Association (JPDA) [13] is known to perform well for some existing solutions of people tracking [14]. In our case, however, the standard JPDA becomes intractable in practice, as the number of possible associations grows exponentially with the size of the BoF, with further complications deriving from the multisensor implementation. We therefore combine two variants of the classic JPDA, namely the Multisensor JPDA (MS-JPDA) [15] and the Nearest-Neighbor JPDA (NN-JPDA) [13]. Therefore, at the beginning we calculate the multi-dimensional matrix of association probabilities, the elements of which are given by the product of the JPDA event probabilities for each single sensor. Then, we perform a recursive search and deletion of the assignments with maximum probabilities, similarly to the nearest-neighbor procedure. Sequences of unmatched readings are used to create new tracks, which are eventually removed when their positional uncertainty becomes too large or in case they are not updated within a certain period of time.

IV. EXPERIMENTAL RESULTS

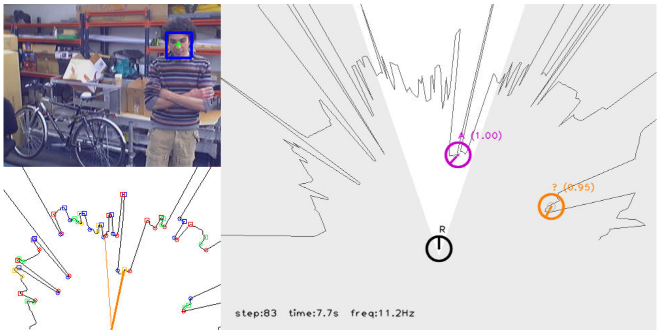
A number of experiments have been conducted with our Lux robot in real situations and also using recorded data. The information relative to 6 different persons has been inserted in the recognition database, for each one including height, torso's histogram and a set of 50 faces automatically extracted from short video files, recorded a few days earlier. The color histograms, which depend on the human clothes, have been obtained the same day, but in different rooms from those used for the experiment, with slightly different light conditions. The experiments show an analysis of the recognition robustness and accuracy, followed by a real application of Lux during the university open day, where tracking, recognition and real-time performances are essential skills.

A. Tracking and Recognition of Multiple People

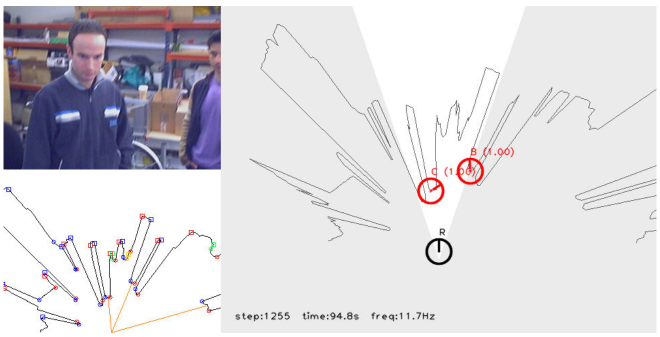
Using several minutes of recorded data, the performance of the simultaneous tracking and recognition has been tested and analyzed. During the data recording, the robot was simply wandering around the laboratory, automatically approaching every person detected along the path. Three different people were continuously moving between different locations of the environment. Lux was always able to track all of them successfully, as shown also by a few snapshots in Fig. 7. These include camera and laser inputs, together the position of the people estimated by the BoFs.

In several occasions, more than one person had to be tracked at the same time. The data association algorithm performed correctly in all these situations, even when the persons were very close to each other like in Fig. 7(b). In the graphs of Fig. 8, we report also the probability of the human recognition for each one of the subjects, A, B and C. The identification performed with the BoF is compared to that one obtained using clothes and face recognition alone.

The graphs show that the identity probability generated by the BoF is always correct and stable. Every time Lux is close enough to a person, the latter is promptly recognized and, at the same time, the probability of being unknown drops to zero. As shown in Fig. 8(a), the identity probability of subject A, using only face recognition, becomes greater than 0.5 only for a few seconds. Also, the person cannot be recognized at all using only clothes recognition. In this case, the main contribution to the successful identification of



(a) Subject A.



(b) Subject C (left) and B (right).

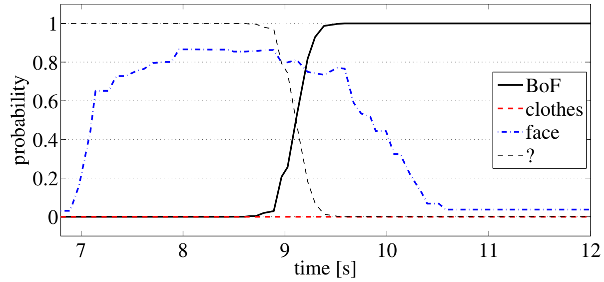
Fig. 7. Snapshots from the tracking and recognition experiment. The face and legs detections are shown on the left, while the estimated positions of the persons are illustrated on the right.

the BoF is given by the initial face recognition and by the tracking information, including person's height. The situation is different, instead, during the identification of subject B, which is illustrated by the graph in Fig. 8(b). While clothes recognition performs very well and almost like the BoF, the face recognition alone is never able to identify the person,

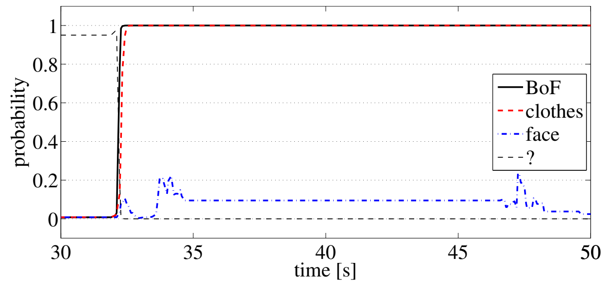
Finally, the graph in Fig. 8(c), which is relative to subject C, shows that the identity of a person can be estimated much quicker, thanks to the clothes recognition, than the case when only faces are considered. During this particular case, indeed, subject C was in the field of view of the camera just for a few seconds between time 60s and 63s, enough for the clothes and height recognition to influence the output of the BoF. The human track with his correct identity was kept even later, out of the camera view, thanks to the legs detected by the laser. Only after time 74s, the face of subject C was visible again and close enough to the robot to be processed by face recognition.

B. Human-Robot Interaction during Open Days

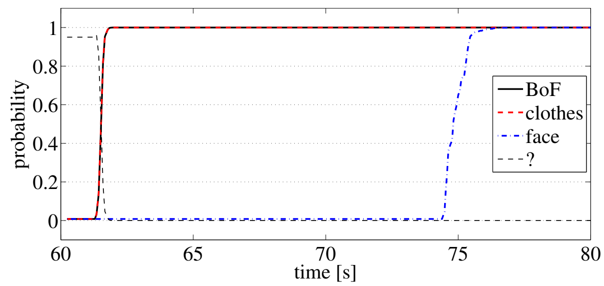
The simultaneous tracking and recognition of people becomes useful in many practical situations of human-robot interaction, in particular in all those cases where the robot's behavior and communication needs to adapt to the identity of the users. Lux is frequently used for demonstrations with young students coming to visit the university. The robot is required to approach some visitors and establish a short interaction, which includes an initial greeting, followed by an invite to use the touch-screen and get more information about the university.



(a) Identity probability of subject A.



(b) Identity probability of subject B.



(c) Identity probability of subject C.

Fig. 8. Identity probabilities. The thick black line is the probability estimated with the BoF, the red dashed line and the blue dash-pointed line refer, respectively, to the clothes and face recognition alone. The thin black dashed line is the probability of the person to be unknown.

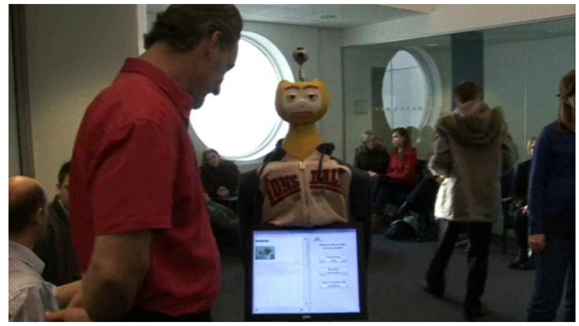
Two demonstrative cases are illustrated by the sequences in Fig. 9 and Fig. 10. The first one shows Lux approaching two different people, stopping in their proximity and starting an interaction. During the first interactions, the robot classifies the user as "unknown" (i.e. anonymous visitor) and provides the necessary information, both vocal and visual, as long as the person stays in front of it. The third snapshot in Fig. 9(b) illustrates instead Lux with a staff member. The interaction in this case is different, as the robot recognizes the subject as "known" person. Lux pronounces a greeting of acknowledgment, using the name of the staff member, and interacts differently than with visitors.

The second sequence in Fig. 10, instead, illustrates Lux that approaches a person and interact with him by inviting him to access information. During these steps, the robots first detects the subject only by legs detection, then also by face. The interaction lasts as long as the person stays in front of the touch-screen, after that Lux searches for other visitors.

(a) Interaction with a visitor.

(b) Interaction with a staff member.

Fig. 9. Snapshots from the tracking and recognition experiment during an open day event in March 2008



(a) The robot detects and unknown person.

(b) The robot approaches the subject.

(c) The person interacts with the robot.

Fig. 10. Approach and interaction during an open day in March 2008

V. CONCLUSIONS AND FUTURE WORK

In this paper, a novel solution for the integration of multisensor information is presented to perform simultaneous people tracking and recognition. It is based on a bank of Bayesian filters and probabilistic data association techniques. Real experiments with an interactive service robot were conducted to verify the feasibility and performance of the proposed system. Our future work will include a more robust algorithm to improve the face recognition accuracy and the system robustness. Also, the histogram database should be updated automatically in order to adapt to the daily changes of human clothes.

REFERENCES

[1] M. Hans, B. Graf, and R. Schraf, “Robotic home assistant Care-O-bot: Past – present – future,” in *Proc. of the 2002 IEEE Int. Workshop on Robot and Human Interactive Communication*, Berlin, Germany, 2002, pp. 380–385.

[2] P. Trahanias, W. Burgard, A. Argyros, D. Hahnel, H. Baltzakis, P. Pfaff, and C. Stachniss, “Tourbot and webfair: Web-operated mobile robots for tele-presence in populated exhibitions,” *IEEE Robotics and Automation Magazine*, vol. 12, no. 2, pp. 77–89, June 2005.

[3] A. Treptow, G. Cielniak, and T. Duckett, “Active people recognition using thermal and grey images on a mobile security robot,” in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Canada, 2005, pp. 2103–2108.

[4] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.

[5] N. Bellotto and H. Hu, “Simultaneous people tracking and recognition for interactive service robots,” in *New Research on Mobile Robots*, E. V. Gaines and L. W. Peskov, Eds. Nova Science Publishers, 2008, (in press).

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

[7] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 72–86, 1991.

[8] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper, *The CSU Face Identification Evaluation System User’s Guide: Version 5.0*, Computer Science Department, Colorado State University, May 2003.

[9] K. Zuiderveld, *Contrast limited adaptive histogram equalization*. San Diego, CA, USA: Academic Press Professional, Inc., 1994, pp. 474–485.

[10] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proc. of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[11] N. J. Gordon, S. Maskell, and T. Kirubarajan, “Efficient particle filters for joint tracking and classification,” in *Proc. of Signal and Data Processing of Small Targets (SPIE)*, FL, USA, 2002, pp. 439–449.

[12] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: particle filters for tracking applications*. Artech House, 2004.

[13] Y. Bar-Shalom and W. D. Blair, Eds., *Multitarget-multisensor tracking: applications and advances*. Artech House, 2000, vol. 3.

[14] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, “Learning motion patterns of people for compliant robot motion,” *The Int. Journal of Robotics Research*, vol. 24, no. 1, pp. 31–48, 2005.

[15] L. Y. Pao and S. D. O’Neil, “Multisensor fusion algorithms for tracking,” in *Proc. of the IEEE Conf. on Aerospace Control Systems*, 1993, pp. 618–622.