Fast, Reliable, Adaptive, Bimodal People Tracking for Indoor Environments

M. Scheutz*, J. McRaven**, Gy. Cserey**
* Department of Computer Science and Engineering
** Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA

***Abstract*— We present a real-time system for a mobile robot that can reliably detect and track people in uncontrolled indoor environments. The system uses a combination of leg detection based on distance information from a laser range sensor and visual face detection based on an analogical algorithm implemented on specialized hardware (the CNN universal machine). Results from tests in a variety of environments with different lighting conditions, a different number of appearing and disappearing people, and different obstacles are reported to demonstrate that the system can find and subsequently track several, possibly people simultaneously in indoor environments. Applications of the system include in particular service robots for social events.**

I. Introduction

The detection and reliable tracking of people in real time is a difficult problem in dynamic indoor environments (such as used for receptions, exhibitions, etc.), where lighting conditions are not controlled, multiple people are move at different speeds in different directions, and temporary occlusions can occur at any time. The problem typically becomes intractable on mobile robots with limited computational resources, where standard algorithms are not applicable because of their high computational demands. Yet, for many service tasks such as those performed by a waiter offering appetizers to guests at a reception or a tour guide keeping track of a group of visitors, being able to track people for a certain time is essential.

Recently, several systems have been proposed for mobile robots to detect and subsequently track people based on a combination of distance information obtained from laser range data and visual information obtained from a camera [24], [9], [12]. While these systems are capable of detecting and tracking individual people to varying degrees, none of them is capable of tracking multiple people at the same time. And even in the case of single person tracking it seems that these systems will fail to reliably track a person moving at a normal speed given that their overall update rates are very low. Moreover, given that the employed face detection algorithms are often dependent on skin color detection alone (as opposed to the detection of other facial features), they are prone to exhibit high classification errors (i.e., *false positives*) in environments that contain skin-like colors. Finally, the proposed systems typically cannot track people while the robot is moving, which effectively excludes the applicability of such systems in highly dynamic environments (e.g., where the robot has to move frequently in order to avoid obstacles or to follow people around).

In this paper, we present a novel bimodal, combined digital-analog approach that is intended to overcome these difficulties. Two subsystems use visual information from a simple web camera for face detection and distance information from a 2D laser range finder for leg detection, respectively. Both systems can track people independently, but use information from the other system to constrain their set of possible candidates for tracking and to overcome temporary occlusions of faces or legs dues to obstacles in the environment or movements of the robot. For the visual subsystem a parallel processor for fast vision processing operations on 64 x 64 pixel images, the *cellular neural network* (CNN) universal machine [1], [2], [3], [4], was integrated. This processor runs a sensory pre-processing analogical algorithm, which in combination with standard digital algorithms allows for fast, reliable detection of faces [23].

The paper is organized as follows: after a quick review of four recent approaches for person detection and tracking, we present an overview of our proposed system. We describe both the face-detection and leg-detection subsystems in more detail, and present results from various evaluation experiments on a mobile robot, which demonstrate different capacities of the system, from fast detection of individuals entering a room, to tracking of multiple individuals with partial occlusions of faces and legs. The subsequent discussion briefly addresses how the proposed system could be integrated into a larger robotic architectures (especially for service robots).

II. Background on People Tracking in Indoor Environments using Visual and Distance Information

Much work on detecting and tracking people with mobile robots has focused on visual methods (e.g., [10], [19], [18], [13]). However, there are also recent approaches that make use of laser range finders to detect and track people (e.g., [17], [14], [15], [16]). Most recently, a few approaches have attempted to combine visual and range data information from laser and sonar sensors to obtain better detection and tracking (e.g., [24], [9], [12], [20]).

The basic idea common to these "bimodal" systems it to use distance information to find the legs or the body of a human person and then subsequently use this information to confine the visual search for faces or human bodies. The individual systems differ in how they make use of distance information and how distance information and visual information are integrated.

The “robot photographer” in [24], for example, uses laser range data to verify that skin-colored pixels, which have been isolated by a color-blob detection algorithm as possibly belonging to faces, do indeed belong to a face. By computing the height of people and the size of their faces based on the distance information, the algorithm excludes pixels that do not correspond to objects within given distance and size parameters, while objects that satisfy the parameters are interpreted as faces.

The system in [12] obtains laser range data first to build a histogram of the background in order to distinguish moving from non-moving objects and subsequently to determine where people (i.e., their legs) are. Then the position information is used to perform face detection on the subimage using a neural network-based face detection algorithm [21].

Visual and laser range information is used to independently obtain information about faces and legs in [9]. These individual “percepts” are subsequently combined or “anchored”: when legs or a face or both are detected in a particular position, the system counts this as evidence that a person has been detected.

Finally, sonar and visual information from a 360 degree camera are fused in [20] to determine the closest person using a variant of the condensation algorithm in [22] for vision-based tracking. Samples of the environment that are only supported by one modality (i.e., either sonar or vision) are eventually discarded.

While all four approaches have their individual strengths, they also have their individual shortcomings. The skin detection algorithm in [24], for example, needs to be re-trained in each new environment. The system in [12] will fail to detect people if the laser scan fails to detect them, and, moreover, only attempt to detect people when the robot is stationary. Finally, the systems in [9] and [20] will have difficulties detecting and tracking people in environments with moving people, where low frame rates are insufficient to keep track of environmental changes (especially the system in [20] given its computationally expensive face and “head-shoulder” detection methods–a detailed assessment of the system’s potential was unfortunately not possible due to the lack of information about the frame rate or any experiments that were performed to evaluate the system).

Independent of these individual challenges, all systems share three important limitations: (1) they can either not deal at all or not deal well with a moving robot, (2) they cannot deal at all with temporary occlusions (e.g., with the absence of one of the two modalities), and most importantly, (3) they can only track one person at a time (e.g., in the presence of multiple people, some chose to track the closest person, e.g., [12] and [20]).

In the following, we will present our approach to people tracking that was designed to overcome all three limitations.

III. An Architecture for Adaptive Bimodal People Tracking

The basic idea underlying the design of the proposed solution, which also distinguishes it from the above described systems, is that the face detection and leg detection systems operate asynchronously and are to some extent capable of tracking faces and legs independently. As will be seen, this leads to superior performance compared to previous solutions, as movements of the robot do not have any significant impact on the performance of each subsystem. Specifically, the robot might be moving due to a violation of the “safety zone” defined around it (e.g., because a person stepped up too close) and still be able to track legs and faces, because both the leg tracking system and the face tracking subsystem will automatically and independently adjust for any robotic movements. We start with an overview of the architecture, which is depicted in Figure 1.

A. Overview of the Architecture

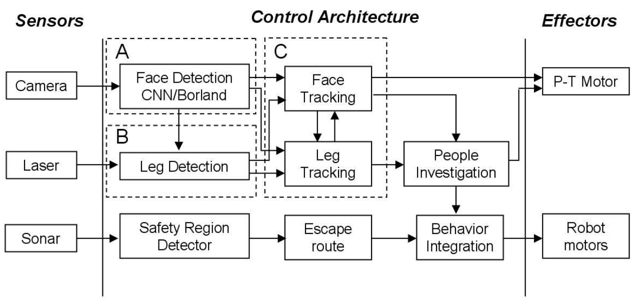

Fig. 1. The basic architecture for people tracking. The dashed areas indicate separate, concurrently operating subsystems for face detection (A), leg detection (B), and face and leg tracking, respectively (C).

Three different sensors are employed in current implementation: a web camera, a 2D laser range sensor, and three sets of 8 sonar sensors each. Sensory data streams are processed separately in parallel (i.e., a visual 2D image stream from the camera, a 1D distance data stream from the laser, and three sets of 1D distance streams each from the sonar sensors). For each stream, there is a perceptual processing module that extracts relevant information from the stream: a face detection module (A), a leg detection module (B), and the “safety region detector”, which detects objects that enter a region of 30 cm around the robot at two different heights. Perceptual information from both modules (A) and (B) is then fed into independent face and leg tracking modules (C), which in turn can constrain each other’s operation. The functional details of these modules will be described below.

The face tracking module directly controls the motors of the pan-tilt unit (on which the web camera is mounted), while the leg tracking module can generate “people investigation” behavior. This behavior is triggered if a candidate region for legs of people is too far away to allow for visual verification about whether the region also has an associated face. If a set of possible legs is to be further investigated, the robot's movement towards the potential person is combined with the ongoing escape route computation that is based on sonar information about objects entering the robot's safety zone. If there is a conflict between the robot's safety and investigation of new potential people (e.g., because they would make the robot move in opposite directions), the safety routes always take precedence in the behavior integration module.

B. The Adaptive Face Detection Module

The face detection algorithm was developed for indoor settings, where faces have to be detected against backgrounds with varying color and uncontrolled lighting conditions (e.g., flickering lights). In such an environment, the variety of background objects makes it often difficult to distinguish objects based on contours, hence color information is used for an initial estimation of whether a face could be located in an image.

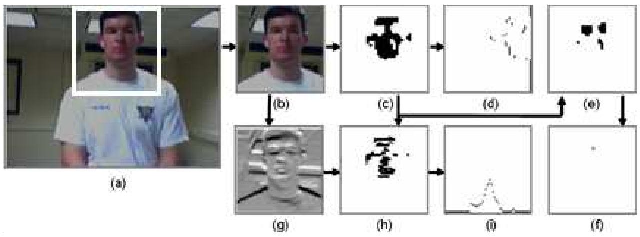

Fig. 2. A diagram of the Analogical CNN Algorithm.

First, a 64 x 64 subimage of the larger 160 x 120 image is selected (step (a) in Figure 2). This region is dependent on whether a face was detected previously in the region or whether the face detection module is in search mode for a face (in which overlapping regions are considered consecutively). This subimage is then pre-processed with the CNN processor to detect faces. Specifically, in a first step the influence of brightness is factored out by taking the difference between the red and green color channels. A threshold function with a parameter range between 0-255 (where 128 is an equal amount of red and green, and 255 is all red and no green) is applied (a typical range for facial color was determined to be 180-240). To make the system more robust against sudden changes of lighting, an adaptation mechanism was integrated that can quickly adjust the lower boundary of the difference range (steps (b) and (c) in Figure 2).

To ensure that a given color patch is a face, *eye detection* is performed, first based on the color information by picking out eyes as gaps in the color patch. To make sure the gaps are inside the face, shadow operators [8] are used to perform shadow operations independently up, and down, then left and right for the respective eyes. A logical AND operation is performed on all shadowed images and the original color data is subtracted from the result to find the gaps inside the face. The left and right shadow operators are then independently combined via an AND operation with the result, from which a centroid for the two eyes is obtained (steps (e) and (f) in Figure 2).

A second method of eye detection is also employed, which uses a horizontal projection of the color information to determine the distribution of the color data. The horizontal projection is analyzed to find the minima and check that the results correspond to the eyes found in the earlier method. This horizontal histogram is also used to find the height of the face for later use (step (d) in Figure 2).

While information about the possible location of eyes in a color patch can significantly improve the detection of faces, background objects with colors similar to facial colors can still interfere enough to reduce the performance of face detection based on eye information. To reduce this interference, edge detection of the area, already determined to be of the correct color, was used to find more complicated structures. The peak in a vertical histogram of the edges is determined to be the center of the face, in a method similar to that used by [7]. The vertical histogram is also used to find the width of the face. A ratio of the height, found earlier, to the width of the face is then used as a final decision making factor (steps (g), (h) and (i) in Figure 2) to determine the *instantaneous probability* that a face was detected, which is computed for every frame (for details of the whole process, see [23]).

If no face was found, but there is skin-like color in a region, the color threshold values are adjust and the process is repeated. This mechanisms helps to eliminate some of the variations due to different lighting conditions.

C. The Leg Detection Module

The leg detection system is based on 180 distance readings at about 25 cm above ground (one per degree) obtained from the laser range finder for the area in front of the robot. Because there are many objects in indoor environments that might qualify as human legs from this simplistic perspective (e.g., legs of chairs or tables, protruding door door frames), objects that are too close to the walls are discarded (this is determined by subtracting subsequent sonar readings of the environment) and only objects that are at some minimum distance away from walls are considered.

Humans can be in two poses from the robot's perspective: in the first both legs are distinguishable, because there is an empty space between them, in the second they appear as one "column". The leg detection system searches for legs of the appropriate with and a possible gap between legs scaled by the distance. If two legs with a gap of appropriate size are found, they are marked as detected. If only one leg is found, there are three cases: it could be two legs close together without a gap (if it is twice the width of a human leg at that distance–this typically happens at greater distances where the 1-degree resolution is insufficient to pick up the gap between the legs), or one leg could occlude the other (e.g., the person is standing sideways), or it is not a human leg after all.

In such circumstances, where legs cannot be identified right away, they are marked as requiring further confirmation, which subsequently will come from face detection: if a face is found in the direction of the suspected legs "above the legs", then the legs are confirmed (this confirmation process will involve the tracking modules and might cause the robot to move towards the leg area if the location of the legs is too far away from the robot's current position to be able to detect faces).

D. The Face and Leg Tracking Modules

The face tracking module attempts to track a detected face over time to gain confidence, based on the instantaneous probability, that a face is actually found. For this, it implements a simple belief revision mechanism that can deal with temporary losses of faces due to imperfections of the detection algorithm (e.g., based on lighting conditions, background objects, etc.) as well as possible occlusions. The belief revision mechanisms consists of a certainty factor C (between 0 and 1), which at any given time expresses the degree to which the system "believes" that it is tracking a face. C is initially 0 and subsequently updated according to the following differential equation: $\Delta C = G * (1 - C) * (face) - D * (1 - C) * (\neg face)$, where $face$ is a Boolean value indicating the presence of a face as determined by the face detection subsystem, and G and D are constants (determined experimentally), which influence the increase and decrease of the certainty that a face has been detected.

The leg tracking module then uses the face tracking module to verify that non-confirmed legs belong to people, and more importantly, attempts to track all confirmed pairs of legs found in the environment. Whenever a new, unconfirmed pair of legs appears (e.g., because a person enters the area observed by the laser sensor), visual confirmation is sought from the face tracking module, which will move the camera in the direction of the legs. If there is face in the area, the confidence C will eventually reach the threshold value 0.4999 used by the leg tracking module to mark the legs as confirmed.

Moving legs will be tracked by matching the closest pair of legs from the previous update cycle to the current one. Two main problems occur with movements: (1) legs are about to leave the sensory range, in which case the robot will either follow them (e.g., moving towards them) or stop tracking them (e.g., because other legs are being tracked that the)–the details of this decision making will very much depend on the particular application. The other problem (2) is that people's paths cross and sets of legs cannot be discriminated because they are occluded. In this case, the system will wait until motion occurs in the area, and then subsequently verify that still two people are present (possibly using face tracking if this cannot be achieved by leg information alone).

In general, people moving in environments will lead to temporary occlusions of legs and faces. Temporarily occluded faces typically are not a problem as long as legs can still be tracked (in that case, the person could not have disappeared). Occluded legs, on the other hand, could mean that the person left the area, and are thus marked unconfirmed.[1] Hence, the face tracking module is used to investigate the area of the last position of the legs to confirm that a face is still present. If no face can be detected, the legs are removed from the tracking list.

IV. Evaluation Experiments on a Mobile Robot

The evaluation of the proposed architecture was performed on an ActivMedia peoplebot with two PCs. The robot has an built-in PC with a 500 Mhz K2 processor and 128 MBytes of RAM running LINUX with kernel 2.6.1. The built-in PC is the main relay station via a serial port connection for the low-level embedded controller, which is in charge of the two wheel motors and also the readings from the three sets of 8 sonars each (two sets forming a sonar ring at a height of 20 cm, one set is mounted in the front of the robot at a height of 1.4 m). The PC is also connected to a SICK 2D laser range finder via another serial port and controls the pan-tilt zoom unit, on which a USB web cam is mounted, through a third. Finally, the built-in PC provides an ethernet connection through which information can be passed to the other PC, a Pentium IV 2.4 Ghz Windows XP machine, which contains a PCI card with the CNN ACE4K chip (which currently is only supported under Windows). The second PC also has a frame grabber to which the web camera is connected.

The implementation of the architecture was split over both PCs: the PC with the CNN chip implemented parts A, B, and C in Figure 1 using the CNN development environment Aladdin [8] and the Borland C++ development environment, while the built-in PC performed the rest using the JAVA-based ADE development environment [25].

For the experimental evaluation, the system was placed in a indoor environment with uncontrolled lighting conditions (flickering neon lights of different intensity). We conducted four different experiments to evaluate different aspects of the system. Experiment 1 tested simple tracking of one moving person. Experiment 2 tested detection of a moving person at a distance, which involves the "people investigation" module. Experiment 3 tested tracking of a moving person with temporary occlusions. Finally, experiment 4 tested tracking of multiple moving people. For each experiment, we show a figure containing three rows. On the first row, four snapshot images from the robot's camera depict important events from the experimental run (they are referred to by their frame numbers). The big white box drawn on the picture is the 64x64 frame that was processed by the CNN. A small white box is drawn to indicate where the system believes the face is if the instantaneous probability is high enough. In the row below the images the laser distance data from -90 to +90 degrees from the robot's heading can be seen, mapped onto a straight horizontal line (the longer the vertical line in a given horizontal position, the shorter the distance of the corresponding object detected by the laser beam from the robot). Dark parts show the objects that are considered

[1] Note that this approach to leg tracking will not work if the legs of a person are permanently occluded, e.g., because the person is wearing a long skirt.

“leg-like” (in some of the experiments the blue objects shows that the leg is confirmed by the face). The chart at the bottom of each figure shows the temporal evolution of the certainty factor of the face tracking system (as well as the instantaneous probability–individual dots–of the face detection system) for each frame during the experiment run.

A. Experiment 1: Tracking one moving person

In this baseline experiment, we tested whether the system could detect a person entering the room and subsequently track the moving person. Figure 3 depicts the results. The first image shows the person entering the sensory range of the laser, where legs are found immediately (while the camera is facing a different direction–Frame 4). Subsequently, the leg tracking module caused the camera to move to the angle at which legs were found and quickly obtained a high C value about the face detected there (Frame 68). The person was subsequently tracked by both tracking modules in different lighting conditions (in this case, the camera also followed the person since it did not have any other face to track–Frames 88 and 323).

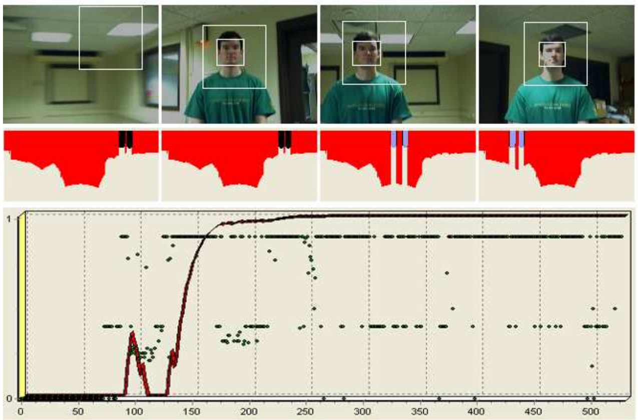

Fig. 3. Experiment 1: Tracking a moving person.

B. Experiment 2: Detecting and tracking at a distance

The second experiment tested whether the robot could find a person at a distance. Figure 4 depicts the results. Again, after the persons enters the sensory range of the laser, the face tracking module attempt to identify a face (Frame 143). Since the robot was not close enough for face tracking to obtain enough certainty, the leg tracking module caused the robot to investigate, i.e., to move towards the suspected area (Frames 446 and 586). Eventually, it came close enough for face tracking to obtain a high C value (Frame 700).

C. Experiment 3: Tracking with temporary occlusions

The third experiment tested whether the robot could track a moving person with temporary occlusions of the legs. Figure 5 shows the results. After detecting and tracking a person with both tracking modules (Frame 280), the person moved behind an obstacle (the location of which can be seen from the indentation of the distance readings in the laser data). The face tracking module immediately took over (Frame 324) and leg tracking resumed when the legs reappeared (Frame 479). As can be seen in the last picture (Frame 545), tracking continues even if the person’s legs cannot be separated (as happens when one leg temporary occlusion the other leg when a person is walking).

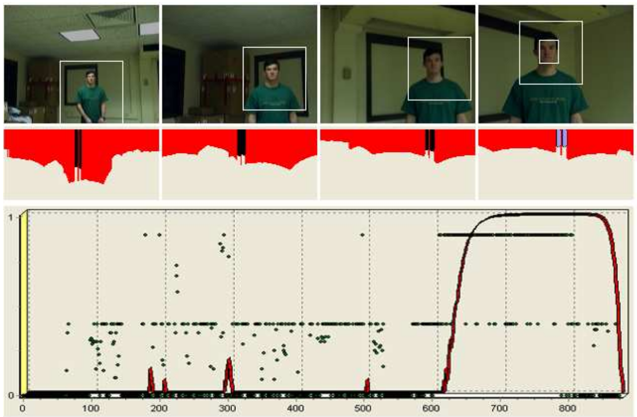

Fig. 4. Experiment 2: Detecting and tracking a person at a distance.

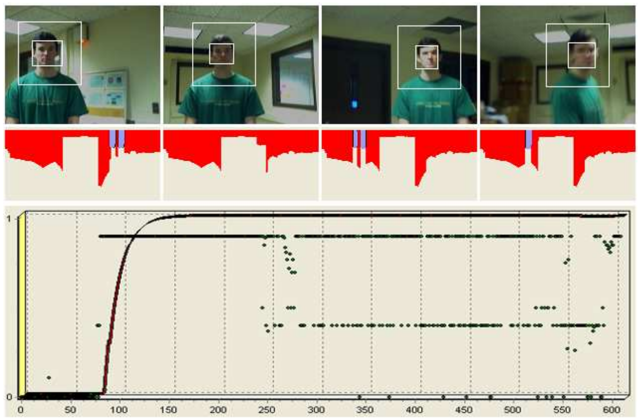

Fig. 5. Experiment 3: Tracking with temporary occlusions.

D. Experiment 4: Tracking multiple persons

The fourth experiment tested the system’s ability to track multiple people.[2] The results are shown in Figure 6. Once the first person was confirmed (Frame 178), the camera moved to the left, where the leg tracking system detected a single leg (the left black bar in the sonar range image) in order to check whether it belongs to a person. However, since the face tracking system was unable to corroborate the evidence (Frame 199), the leg was subsequently ignored. As soon as new legs were detected (Frame 282), the camera moved to check for a face and subsequently confirmed the presence of a person (Frame 627).

[2]For lack of space the two charts with the certainty factors have been omitted.

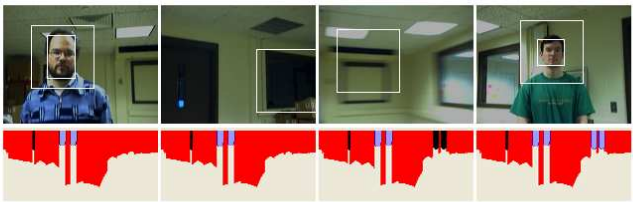

Fig. 6. Experiment 4: Tracking multiple persons.

V. Discussion

The experimental results demonstrate that the proposed system is capable of tracking multiple moving people, even when the robot is itself in motion, different from the other systems discussed in the background section. This is possible because we employ two tracking systems that work together in parallel, rather than sequentially. By using the CNN chip, we are able to achieve a very high update frequency of the visual tracking module of 30 Hz, which is the maximum frame rate of the frame grabber board. It is expected that even higher frame rates could be obtained with faster frame grabbers, while increasing the size of images in the visual stream. The leg tracking system currently operates at approximately 3 Hz, which has been found to be lowest update rate sufficient for leg tracking when people are moving at normal speeds (note that there is still room for an increased update rate for people moving quickly). We are currently evaluating the system in different indoor settings with larger numbers of concurrently present people (from 5 to 20).

The current system can easily be integrated into architectures for service robots, for example, a robotic waiter for receptions that attempts to serve appetizers and drinks to people. For such an application it is sufficient to only approximately keep track of where people are in the environment and move to all of them at regular intervals. This can be done in with the current system by using the confirmed leg information in the leg tracking module as a basis for a simple person identification mechanisms that attempts to keep track of how many times a person has been served and makes the robot move towards people who have been served least–we are currently working on an implementation of this application. Finally, we also see much room for improvement of the leg tracking modules, especially the determination of the confidence that partially or wholly occluded people are still present.

Acknowledgment

Partial support for this project was provided by the Office of Naval Research through a MURI grant.

References

[1] L. O. Chua and L. Yang, "Cellular neural networks: Theory," IEEE Trans. on Circuits and Systems, Vol.35, pp. 1273-1290, 1988.

[2] L. O. Chua & T. Roska, Cellular neural networks and visual computing, Foundations and applications, Cambridge University Press, 2002.

[3] A. Zarandy, Cs. Rekeczky, I. Szatmari, and P. Foldesy, "Aladdin Visual Computer," IEEE Journal on Circuits, Systems and Computers, Vol. 12(6), 2003.

[4] G. Linan, S. Espejo, R. Dominguez-Castro and Rodriguez-Vazquez, "ACE4k: An analog I/O 64 x 64 Visual Microprocessor Chip With 7-bit Analog Accuracy," Intl. Journal Of Circuit Theory and Applications, Vol. 30, May-June 2002, pp. 89-116.

[5] M.-H. Yand, D.J. Kriegman and N. Ahuja, "Detecting Faces in Images: A Survey," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24. pp. 34-58, 2002.

[6] D. Balya and T. Roska, "Face and Eye Detection by CNN Algorithms," Journal of VLSI Signal Processing, Vol. 23., pp. 497-511, 1999.

[7] Z. Szlavik and T. Sziranyi, "Face Identification with CNN-UM," in Proc. ECCTD '03, Krakow 2003.

[8] Aladdin Professional, Reference Manual, Version 2.5, Budapest 2003.

[9] M. Kleinehagenbrock, S. Lang, J. Fritsch, F. L¨omker, G. A. Fink and G. Sagerer, "Person Tracking with a Mobile Robot based on Multi-Modal Anchoring", Proc. IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN), Berlin, Germany, September 2002. IEEE.

[10] D. Beymer and Konolige K. Tracking people from a mobile platform. In IJCAI-2001 Workshop on Reasoning with Uncertainty in Robotics, 2001.

[11] W. Burgard, A.B. Cremers, D. Fox, D.Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. Artificial Intelligence, 114(1-2), 1999.

[12] J. Blanco, W. Burgard, R. Sanz, J.L. Fernandez. Fast Face Detection for Mobile Robots by Integrating Laser Range Data with Vision. In Proc. of the International Conference on Advanced Robotics (ICAR), 2003.

[13] I. Horswill. Polly: A vision-based artificial agent. In Proc. of the National Conference on Artificial Intelligence (AAAI), 1993.

[14] A. Fod, A. Howard, and M. J. aMataric . Laser-based people tracking. In Proc. of the IEEE International Conference on Robotics & Automation (ICRA), 2002.

[15] B. Kluge, C. Koehler, and E. Prassler. Fast and robust tracking of multiple moving objects with a laser range finder. In Proc. of the IEEE International Conference on Robotics & Automation (ICRA), 2001.

[16] M. Montemerlo, S. Thun, and W. Whittaker. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In Proc. of the IEEE International Conference on Robotics & Automation (ICRA), 2002.

[17] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving objects with a mobile robot. In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2001.

[18] John A. Perrone, Tony Voyle, Margaret E. Jefferies "Towards a Human Tracking System for a Mobile Robot Using Neural-Based Motion Detectors", Image and Vision Computing New Zealand, 2003, 24–29

[19] Vision Based Person Tracking with a Mobile Robot C. Schlegel, J. Illmann, H. Jaberg, M. Schuster, R. Worz Proceedings of the Ninth British Machine Vision, 1998

[20] Wilhelm, T., Boehme H.-J., Gross, H.-M.: Sensor Fusion for Vision and Sonar Based People Tracking on a Mobile Service Robot. in: Proc. Int.Workshop on Dynamic Perception 2002, Bochum, 315-320, IOS Press.

[21] H.A. Rowley, S. Baluja, and T. Kanade. Neural networkbased face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1), 1998.

[22] Isard, M. and Blake, A. CONDENSATION conditional density propagation for visual tracking. International Journal on Computer Vision, 29(1):5 28, 1998. B. Jähne. Digitale Bildverarbeitung. Springer-Verlag, Berlin Heidelberg, 3. Auflage.

[23] Raven, J., Scheutz, M., Gy. Cserey, Andronache, V., and Porod, W. "Fast Detection and Tracking of Faces in Uncontrolled Environments for Autonomous Robots Using the CNN-UM". In the 8th IEEE International Biannual Workshop on Cellular Neural Networks and their Applications. (forthcoming)

[24] Zachary Byers, Michael Dixon, Kevin Goodier, Cindy M. Grimm, and William D. Smart. "An Autonomous Robot Photographer". In Proceedings of the International Conference on Robots and Systems (IROS 2003), Las Vegas, NV, October 2003.

[25] Andronache, Virgil and Scheutz, Matthias "ADE - A Tool for the Development of Distributed Architectures for Virtual and Robotic Agents". In International Journal of Artificial Intelligence Tools. (forthcoming)