

# Who am I talking with? A Face Memory for Social Robots

Marc Hanheide, Sebastian Wrede, Christian Lang, and Gerhard Sagerer

Faculty of Technology, Applied Computer Science, Bielefeld University, 33594 Bielefeld, Germany

{mhanheid, swrede, clang, sagerer}@techfak.uni-bielefeld.de

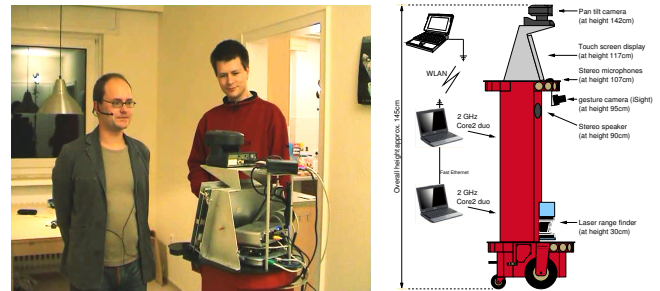
**Abstract**—In order to provide personalized services and to develop human-like interaction capabilities robots need to recognize their human partner. Face recognition has been studied in the past decade exhaustively in the context of security systems and with significant progress on huge datasets. However, these capabilities are not in focus when it comes to social interaction situations. Humans are able to remember people seen for a short moment in time and apply this knowledge directly in their engagement in conversation. In order to equip a robot with capabilities to recall human interlocutors and to provide user-aware services, we adopt human-human interaction schemes to propose a face memory on the basis of active appearance models integrated with the active memory architecture. This paper presents the concept of the interactive face memory, the applied recognition algorithms, and their embedding into the robot's system architecture. Performance measures are discussed for general face databases as well as scenario-specific datasets.

## I. INTRODUCTION

As the identity of an interaction partner is one of the most important contextual factors for any conversation, humans have developed very advanced and specialized capabilities to acquire models of human faces. Consequently, robots that interact with different human users can draw huge benefit from discerning the respective communication partner. A robot that knows ones specific preferences, e.g., with regard to drinks or entertainment, can provide more personalized services and can interact in a more efficient and natural means. Hence, social robots need to recognize their interaction partners, especially in scenarios they have to cope with a number of different ones as illustrated in Fig. 1(a).

Face recognition is nowadays mostly studied as a task to retrieve records from huge face databases. This discipline has been boosted in the last decade due to the arising need in security technology. In this application field, the face models are usually obtained from well-illuminated and carefully taken face pictures.

However, in social interaction the situation is a different one. A facial model needs be acquired during the initial contact already and no large and well captured set of facial images will be available. On the other hand the number of people to be discerned is usually much smaller compared to security databases. Face recognition for social robotics needs to be integrated with the interactive capabilities of a robot like its dialog subsystem in order to provide valuable assets for improved human-robot-interaction. So, identifying the interlocutor is not only a pattern recognition challenge but a systemic human-robot interaction task. Different sources of information need to be exploited to obtain the required information. The identity of a person might be ascertained



(a) Two persons interacting with BIRON

(b) Hardware

Fig. 1. The home robot.

by means of face classification but as well by asking the human for her name during the discourse. And both these use cases affect and control each other mutually.

In this paper, we present our approach of tightly coupling interaction, model acquisition and pattern recognition in a system architecture on the basis of the active memory concept [21]. The presented system of a face memory is part of a robot system architecture for an indoor robot companion sketched in Fig. 1(b). The development of the robot is framed by the so-called home-tour scenario which is driven by the vision of future household robots being introduced for the first time use after purchase. A robot needs to get to know its new working environment which cannot be pre-programmed, but needs to be explored together with naïve users. Hence, human-robot interaction about the spatial and functional environment is in the focus of research in such home-tour scenarios. Natural interaction Capabilities of the home-tour robot comprise understanding of spoken utterances, co-verbal deictic reference, verbal output, referential feedback, as well as person attention.

## II. A FACE MEMORY FOR SOCIAL ROBOTS

When we consider the ability to learn the identity of interlocutors as a core cognitive function for a social robot we have to ask the question, how the mutual introduction and the recalling of faces is embedded into the general interaction scheme. Different sources of information are available, as the identity of a person might be the result of the current conversation (“What is your name?”) or obtained from analyzing the person’s appearance. Which knowledge source to consider is depending on the current content of the face memory and the conversational state. The memory here serves as the central knowledge source. It allows the system to determine whether a person is known or not. In case the person is already known and the robot is certain

about her identity it can just activate its knowledge about this person, while otherwise it has different options. First, it can take initiative and ask the new person for her name. Alternatively, it may continue conversation with an implicit but not yet named user model and wait for the name of the person to be mentioned sometime. Currently, the robot asks for the human’s name whenever a yet unknown persons is engaged in conversation, resulting in a certain curiosity.

However, it should be noted that the robot does not require to know every person in its vicinity. Besides recognizing and memorizing people’s faces the system also comprises a person tracking and attention functionality called person anchoring [7] to establish and maintain hypotheses of surrounding persons. Storing face views of these persons of interest alongside in a memory allows to immediately compute a new face representation on the basis of the last seen face patches. This is again in accordance to human behavior, as we do not start looking at someone’s face after hearing the respective name. Instead, we already memorize their appearance when initiating the conversation.

In the COGNIRON [4] project, we target at developing robot companions that exhibit social interaction capabilities. The robot is enabled to focus on its current interaction partner and to detect other persons of interest. Its person attention mechanism allows it to align its pan-tilt camera accordingly, and in consequence zoom and focus the face of the respective interaction partner.

Face recognition for conversational robots has already some tradition in human-robot interaction. Wong et. al. [20] already presented a robot system that is able to recognize faces more than ten years ago. However, their approach is limited to six pre-trained faces and sensitive to lighting conditions. The humanoid robot ROBITA has also been equipped with a person attention system comprising face recognition [13] to discern different users. A special challenge for face recognition arises from the fact, that faces are usually not well-aligned in the visual field of the robot. Therefore, research focuses on methods that allow face recognition also from different viewpoints [10]. In the remainder of this paper, our proposal for the recognition of faces from a robot’s perspective will be outlined, followed by the explanation of the system architecture facilitating the outlined interaction scheme and recognition results.

### III. FACE RECOGNITION

For a face memory to be useful, a process that identifies a person’s face on the basis of acquired face models is required. Face recognition in general is considered a major topic in computer vision for quite a while. For the classification of faces and facial expressions a comprehensive survey [23] lists different algorithms such as principal component analysis, independent component analysis (ICA), hidden markov models, and neural networks. While all the different approaches have different pros and cons, robotic applications in real world settings demand for specific requirements. First, faces have to be robustly detected also in cluttered background to be applicable in home environments. Second,

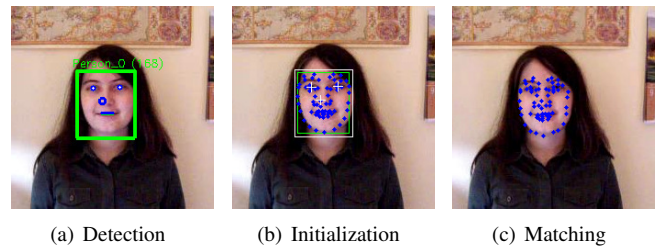


Fig. 2. Face Recognition with AAM from a robot’s first person perspective.

light changes must not impede the performance of the face recognition. And finally the number of training samples should be rather small to account for the interactive, online acquisition of the underlying models. In order to account for these challenges a three stage procedure comprising face detection, feature extraction using active appearance models (AAM), and a successive support vector machines (SVM) classification step is applied consecutively.

#### A. Detection

Detecting faces is a required as a first step to choose the image patches for successive analysis. Most recent detection methods have been influenced by the work of Viola and Jones [17] and improved ever since. The method applied in our system also belongs to this family of approaches and combines detection and tracking not only of the full facial region but also the separate feature such as mouth, nose and eyes [2]. The detection rate of this approach is good enough to allow its application in cluttered environments such as apartments and hence accounts for the first requirement outlined above. The result of this detection step is displayed in Fig. 2(a).

#### B. Feature Extraction

As the variance of captured images of human faces is immense and to a large extend bound to unwanted factors such as illumination and position of as well the robot as the human, the image patches itself are not well suited as raw features for the classification. Instead, a dedicated feature extraction step is needed and several methods have already been developed in the past years. Wide spread are on the one hand direct statistical analysis such as ICA and generative models for faces. A comparison of ICA and AAM — as one of the most prominent generative approaches — reports a slightly better performance in face recognition for the first [19]. However, the differences are minor and ICA demands for precisely aligned images and is considered to be less robust in real world settings. Consequently, we decided for active appearance models which have been originally proposed by Cootes [5]. These models are used to capture the relevant variances of the individual’s faces and to compute suitable features by a generative pattern matching technique. AAMs are statistical models of shape and gray-level appearance that are trained from annotated example images with certain similarities to morphable models [18]. Applying an “interpretation by synthesis” scheme, appearance models become *active* by means of a matching procedure that fits a

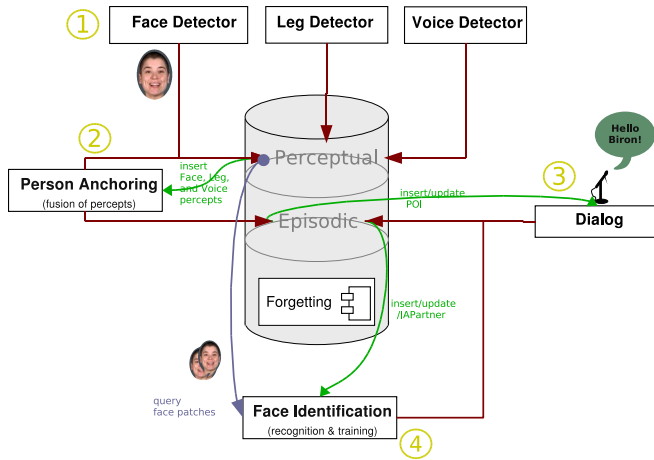


Fig. 3. System architecture for an interactive face memory.

rendered face to a given image and computes a corresponding parameter vector iteratively.

In our system we apply 2D AAM models for frontal faces following the assumption that we are interested in communication situation where the human to be recognized is facing the robot. Using the detection results to initialize the AAM for a given face image as shown in Fig. 2(b) the iterative optimization is carried out until it converges. The optimal match (Fig. 2(c)) provides the resulting parameter vector.

### C. Classification

The classification is carried out as a linear three stage process. First, the reconstruction error of the AAM fitting algorithm is used to reject badly aligned and matched face images from recognition based on a threshold. Second, a nearest-neighbor classifier on the basis of a mahalanobis distance with a rejection threshold is applied for the binary decision whether a face is known or not. If the person is considered to be known it is classified using an one-vs-rest SVM classifier with linear kernels. “It” for person? By means of this three stage approach, the system is certain that it either does know the person correctly, or that it cannot classify this face patch appropriately. It shall be noted, that a high rejection rate is not necessarily a drawback for online system because new patches arrive at frame rate, allowing dropped frames to be compensated quickly.

## IV. SYSTEM DESIGN

The extension of our interactive robot BIRON by a face memory towards a robot capable of improved social interaction quickly turns out to be an architectural challenge as face recognition on its own does not pay off in a smarter robot. As already argued in the introduction a system perspective is needed, focusing at the interplay of perceptual and deliberative processes.

### A. Information-Driven Integration with Active Memories

In order to further enhance our robot companion with additional features, we incrementally adapted the previous integration and control architecture presented at ICRA05 [8] towards an extended use of the active memory concept [22]

allowing for *information-driven integration*. The basic idea of this approach is to track the interaction context in terms of perceived episodes, events, and scenes and coordinate modules within a system architecture by an event-driven integration scheme through a hierarchy of active memories. Information-driven integration and the underlying software framework has been particularly suitable for the development of systems aiming at advanced human-machine-interaction [16] and cognitive robotics [14].

Within our realization of this concept, almost all information flowing through the distributed system is mediated through instances of so called active memory servers. These infrastructure components provide the basis for coordination and shared data management within architectural layers. Coordination is facilitated by an event-driven software architecture co-located with the database backends of the memory server. Following an event-condition-action model, conditions can be specified by using XPath as subscription language for subscribing to different types of XML data as indicated in Fig. 3. A condition is usually supplemented by a memory action, e.g., insert or update. Whenever one of these conditions matches, the active memory server triggers the registered action of a remote client delivering an asynchronous event. Without going into the details of the active memory concept [9], [21], a memory process to be mentioned is the “forgetting” as an inherent part of the memory. It is applied to clear the memory instances from outdated or unreliable information, such as recorded face patches no longer needed.

### B. System Architecture

For the realization of the face memory we enhanced the previously existing system architecture by the face recognition component introduced in Sect. III and a person anchoring component, which is inspired by a former approach [7] that yields good results for the application in typical home environments. Fig. 3 shows the part of BIRON’s system architecture that is responsible for the realization of the face memory functionality. Utilizing the information-driven integration approach no changes have been necessary to reuse the already existing dialog subsystem. The coordination between the different processes is solely event-driven as explained in the previous section. While the perceptual memory is optimized for short-term memorization of hypotheses generated from the stream of low-level sensor data, the episodic memory stores and processes higher-level symbolic information which is valid for longer periods of time. In the following, we will shortly describe the functionality of the different integrated processes as enumerated in Fig. 3.

1) *Perceptual Processing*: Within our system architecture a number of different components perform bottom-up processing of incoming sensor data. They implement most of BIRON’s perceptual capabilities. For the face memory functionality we focus on three of these processes. First, a voice detection component analyzes the cross-power spectrum phase to estimate the relative locations of multiple speakers. As soon as a speaker is detected, this and the spatial origin of

the corresponding audio signal is submitted to the perceptual memory. The second process scans the vicinity of the robot for pairs of legs by analyzing the data available from the attached laser scanner in order to generate hypotheses about possible human interaction partners standing or moving in front of the system. The face detection component as the third process in this layer has already been introduced in the course of Sect. III. For each video frame the detected facial regions are extracted and inserted in the perceptual memory of our robot together with a referring face hypothesis for subsequent processing.

2) *Person Anchoring*: A key component facilitating a face memory for an interactive robot is a multi-modal person anchoring process. Our realization is inspired by the approach introduced by Coradeschi & Safiotti [6]. Anchoring in general is a process that links perceptual information about real world entities, e.g., faces, to symbols that reliably represent the found entity over a certain period of time. Anchoring processes in active memory architectures are employed to populate the episodic layer with information generated from the data available in the perceptual layer. Within BIRON's face memory, information generated by the individual modalities in the perceptual layer (face, legs and speakers) is anchored separately in the person anchoring component itself and is afterwards assigned to a person-of-interest (POI) hypothesis. A new POI will be created iff one of the input percepts does not match any of the existing modality anchors. Existing POI hypotheses are maintained as long as at least one of its three modality anchors can be tracked continuously. New POI hypotheses and their updates are submitted as episodic information to the corresponding active memory instance. Additionally, the person anchoring component frequently updates the face information in the perceptual memory with an ID originating from the corresponding POI for subsequent use through the face recognition component.

3) *Dialog*: For social robots it is important to be capable of natural communication, e.g., by speech understanding and production. These features are realized in our robot by a dialog subsystem [12]. The basic concept followed within this component is the grounding-principle [3], which states that within a conversation the communication partners need to coordinate their mental states based on their mutual understanding. Adding up on the actual dialog functionality, a social robot must be able to distinguish between several persons in its vicinity and to identify as well as align its communication to a human interaction partner focusing its attention on the robot itself. While the former is performed in our architecture by the person anchoring module, the latter is an additional function that is provided by the dialog subsystem. As soon as the dialog is triggered by a specific initiation phrase ("*Hello Biron!*") from a person that is registered as a POI in the episodic memory, the dialog selects this one as its interaction partner (IP) and in turn submits this new information with the preserved ID from the POI hypothesis to the episodic memory. The identification of its communication partners without repetitive asking the human for his name,

significantly improves the interaction experience. This is due to the fact that the dialog manages individual user models stored in the episodic memory. In consequence, it is possible to, e.g., adapt the speech recognition component to speaker dependent profiles before a conversation starts or to optimize its interaction by not repeating instructions already known to the respective person.

4) *Face Identification*: The algorithmic details of the face recognition as the main functional enhancement to BIRON's system architecture have been described in Sect. III. It is important to note that the face identification itself demands for certain services within the architecture to achieve the envisioned embedding in the interaction scheme. Thus, the face recognition component makes use of several sources of information generated by other components, e.g., the IP hypotheses and their corresponding face patches. This information is utilized to perform the classification of the robots' communication partners. When the user has been trained previously and the classification is successful, the information available in the episodic memory is updated by the corresponding class name. Otherwise the face patches are used to train a new classifier as soon as the name of the communication partner has been acquired through the dialog subsystem. While the POI hypothesis and the name of the current interaction partner - if set by the dialog component - are retrieved through event notifications from the episodic memory, a query on the perceptual memory is performed to retrieve recent face patches that correspond to the current interaction partner.

### C. Component Interaction

Figure 4 exemplifies the dynamic interactions between the components of the system in terms of activities that are carried out in the face memory when a human user enters the robots' interaction area, looks at BIRON and finally starts the interaction by greeting it with the initiation phrase.

As soon as the human approaches the robots sensors and her legs are detected, the episodic memory notifies the person anchoring component about the new "legs" percepts. In turn, a local modality anchor is set up and a new POI hypothesis is submitted to the episodic memory. While possible in parallel, let us assume for this example that the face of the user is detected as he further approaches the robot. As a consequence of this activity, two others are triggered concurrently: the detected faces and their views are submitted to the perceptual memory and the previously established POI profile is extended by the information that the person is now "facing" the robot.

The next activity the human carries out to start a new discourse with BIRON is to address it by speech by means of an initiation phrase. This leads to an update of the corresponding POI hypothesis, which is enhanced by the information that this person is now talking. Additionally, the dialog component selects this POI as its interaction partner iff the speech understanding result provides the symbol for the initiation phrase. Dependent on this decision, the information about the selected IP is submitted to the episodic memory.

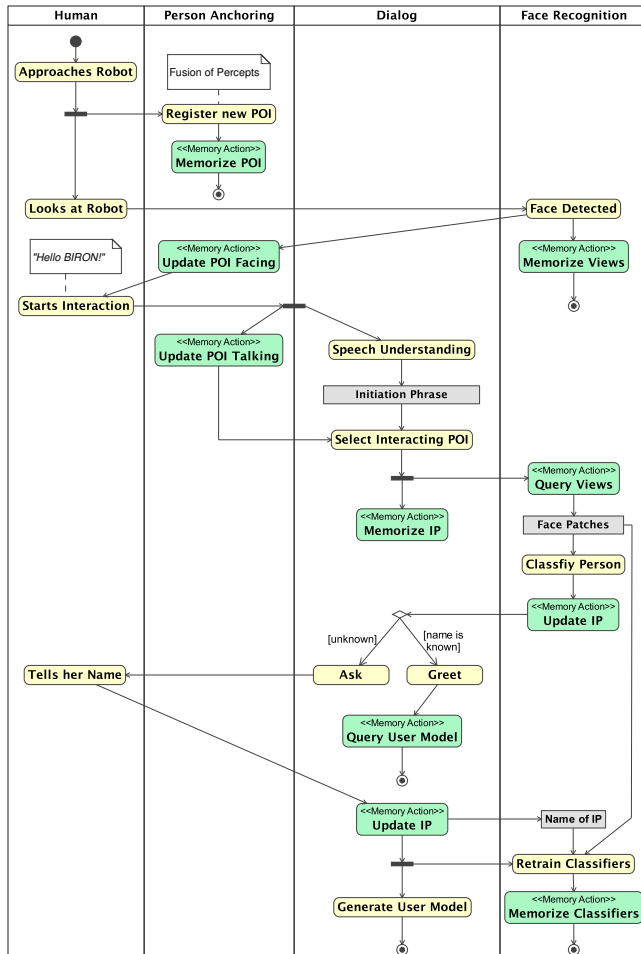


Fig. 4. Activities carried out by the human communication partner and the robot system when initiating a conversation utilizing a face memory.

Once the IP hypothesis is available, the face recognition component is activated by a corresponding event and queries the recent face patches corresponding to this interaction partner. These views are used for the following classification step that yields an update of the IP hypothesis in the episodic memory. It is either enhanced by the name of the communication partner in case of a successful classification or by a label indicating that this human is so far unknown.

The following activity, once more triggered by the update of the IP hypothesis is carried out within the dialog component in two different ways based on the name information updated previously by the face recognition available within the IP hypothesis. When the system does already know the name of its communication partner, the final activities in this example are the retrieval of the corresponding user model from the episodic memory as well as the adaptation of the dialogs' interaction strategy and the greeting of the IP using its name. In case the user could not be classified successfully from the set of recent face patches, the dialog subsystem asks the user for this information and uses the label retrieved from its speech recognition module to update the IP hypothesis with the given name. In this case the face recognition is triggered by the updated IP hypotheses and starts to train a classifier for this previously unknown person. Finally, the

(a) DaFEx		(b) Robot				
#train. samples	class %	#train. samples	rejection true pos	rejection false pos	class w/ rej	class w/o rej
10	78.6%	10	100.0%	57.0%	97.7%	86.0%
100	98.3%	100	99.5%	23.5%	97.9%	95.5%

TABLE I  
FACE RECOGNITION RESULTS.

dialog generates a new user model that is used in subsequent interactions when this communication partner is recalled by the robots' face memory system.

## V. RESULTS

While the overall interaction generally needs to be assessed in terms of user studies [11], in this paper we present a dedicated study of the face recognition approach along two different data sets. First, we analyzed the performance of the face recognition component solely on a public database. Second, the face memory implementation has been tested on images taken from the robot's perspective in a realistic environment. Our approach first requires to estimate appropriate AAM models that cover a certain variance of human faces. Therefore, we selected 333 images of 65 subjects from the Spacek database [15] as AAM training set. The resulting AAM with 41 dimensions conserved 95% of the variances in the data set. This AAM is then used with the different datasets for evaluation. No domain specific AAMs need to be retrained. The overall execution time for the classification is dominated by the nearest-neighbor classifier, whose time effort scales linear with the number of training examples and quadratic with the dimension of the feature vectors. For the tests described below, the computation including detection, feature extraction, and classification takes between 130 and 150 milliseconds on up-to-date dual core PCs.

### A. DaFEx

We used the DaFEx database [1] because it contains faces with different facial expressions and talking persons as they are typical in conversation. The training was performed on the basis of about eight subjects with 10 and 100 different face images each. In these images, the subjects exposed a neutral facial expression and were not talking. After training the SVM 100 randomly chosen test images of the same people but engaged in conversation have been analyzed with respect to the three-stage procedure: detection, feature extraction and classification. The results reported in Table I (a) indicate a good performance in terms of correctly recognized frames and encourage the use of AAM approaches for face recognition in social interaction for small training sets even though the people were talking in these sets. This indicates good generalization capabilities of our approach at least as long as the number of training samples is not too small.

### B. Robot

As a second evaluation, the face memory is used in a realistic setting with images captured from the robot's perspective (see example image in Fig. 2). As training data either 10 or 100 images have been selected from the

beginning of a recorded video sequence. This corresponds to approximately 0.5 and 5 seconds of raw video material captured by the face detection with face patches stored in the perceptual memory, respectively. On this data set, the outlined rejection scheme is first applied to determine whether a person is known or not, and then the face is classified into the known set of persons. Therefore, a number of ten subjects was introduced to the face memory and two additional subjects have been added to the test set as unknown persons. The rejection threshold was chosen to achieve a high true positive rate, as the robot should only report reliable results about known persons. The respective true positive and false positive rejection rates are reported in table I (b) as well as the success rates of the SVM classification of non-rejected faces. The robust rejection is necessary to correctly trigger the robot's request for the human's name. Furthermore, a look at these results reveals the benefit of the rejection scheme as the classification rate is significantly higher when the previous rejection is considered (labeled "w/ rej" in Table I (b)). It shall be noted that these results are obtained on a per-frame basis. Fusing results of several frames yields even better performance, likewise does the joint classification of recorded face patches retrieved from the memory.

## VI. CONCLUSION

In this paper we presented a unique integrated face memory that links interactive introduction of interlocutors with an online learning face classification scheme. The presented results confirm the adequacy of the chosen methods especially for social robot scenarios. The presented face memory facilitates the outlined way of mutual control necessary for a socially acceptable interaction. The outlined use case also elicited the benefits of the information-driven integration approach using the active memory concept and hence underpins its suitability as a basis for building hybrid software architectures for robots with cognitive abilities.

While the classification results are already encouraging and good enough for the outlined settings our focus was currently not on optimizing this step. Literature indicates that more specialized classification approaches still promise a significant boost in recognition performance and different classification methods mentioned already in this paper are still subject to investigation in this interactive scenario. Successive re-training of face models in the memory on the basis of tracked persons is easily supported by our memory approach and will be studied in detail in the future to allow better generalization of the recognition. While currently interaction is mainly about getting to know each other, more advanced dialog strategies will allow to explicitly solve ambiguities and correct perceptual deficiency.

## ACKNOWLEDGMENTS

This work has been partially supported by the European Union within the 'Cognitive Robot Companion' (COGNIRON) project (FP6-IST-002020) [4].

## REFERENCES

- [1] A. Battocchi, F. Pianesi, and D. Goren-Bar, "A first evaluation study of a database of kinetic facial expressions (dafex)," in *Proc. Int. Conf. on Multimodal interfaces*. ACM Press, 2005, pp. 214–221.
- [2] M. Castrillón, O. Déniz, C. Guerra, and M. Hernández, "Encara2: Real-time detection of multiple faces at different resolutions in video streams," *Journal of Visual Communication and Image Representation*, vol. 18, no. 2, pp. 130–140, 2007.
- [3] H. H. Clark, Ed., *Arenas of Language Use*. University of Chicago Press, 1992.
- [4] COGNIRON, "The cognitive robot companion," <http://www.cogniron.org>.
- [5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *PAMI*, vol. 23, no. 6, pp. 681–685, June 2001.
- [6] S. Coradeschi and A. Saffiotti, "Perceptual anchoring of symbols for action," in *Proc. Int. Conf. on AI*, 2001, pp. 407–416.
- [7] J. Fritsch, M. Kleinhagenbrock, S. Lang, T. Plötz, G. A. Fink, and G. Sagerer, "Multi-modal anchoring for human-robot-interaction," *Robotics and Autonomous Systems*, vol. 43, no. 2–3, pp. 133–147, 2003.
- [8] J. Fritsch, M. Kleinhagenbrock, A. Haasch, S. Wrede, and G. Sagerer, "A Flexible Infrastructure for the Development of a Robot Companion with Extensible HRI-Capabilities," in *Proceedings ICRA*, Barcelona, Spain, April 2005, pp. 3419–3425.
- [9] J. Fritsch and S. Wrede, *Software Engineering for Experimental Robotics*, ser. STAR. Springer, 2007, vol. 30, ch. An Integration Framework for Developing Interactive Robots, pp. 291–305.
- [10] K. Fukui and O. Yamaguchi, "Face recognition using multi-viewpoint patterns for robot vision," in *Proceedings Int. Symposium of Robotics Research*, 2003, pp. 192–201.
- [11] S. Li, M. Kleinhagenbrock, J. Fritsch, B. Wrede, and G. Sagerer, "'BIRON, let me show you something': Evaluating the interaction with a robot companion," in *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics*. IEEE, 2004, pp. 2827–2834.
- [12] S. Li and B. Wrede, "Why and how to model multi-modal interaction for a mobile robot companion," in *AAAI Technical Report SS-07-04: Interaction Challenges for Intelligent Assistants*. Stanford: AAAI Press, 2007, pp. 71 – 79.
- [13] Y. Matsusaka, T. Tojo, S. Kubota, K. Furukawa, D. Tamiya, S. Fujie, and T. Koabyashi, "Multi-person conversation via multi-modal interface: A robot who communicate with multi-user," in *Proc. Eurospeech*, 1999, pp. 1723–1726.
- [14] H. Ritter, R. Haschke, and J. J. Steil, *Perspectives of Neural-Symbolic Integration*, ser. Computational Intelligence. Springer, 2007, ch. A Dual Interaction Perspective for Robot Cognition: Grasping as a Rosetta Stone, in press.
- [15] L. Spacek, "Collection of facial images," WWW, Sept. 2007. [Online]. Available: <http://cswww.essex.ac.uk/mv/allfaces/index.html>
- [16] VAMPIRE Consortium, "VAMPIRE: Visual Active Memory Processes and Interactive Retrieval," Feb 2006, IST-2001-34401, <http://www.vampire-project.org>.
- [17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. CVPR*, vol. 1, Kauai, Hawaii, Dec. 2001, pp. 511–518.
- [18] B. Weyrauch, B. Heisele, J. Huang, and V. Blanz, "Component-based face recognition with 3d morphable models," in *Proc. CVPRW*. Washington, DC, USA: IEEE Computer Society, 2004, p. 85.
- [19] T. Wilhelm, H.-J. Böhme, and H.-M. Gross, "Classification of face images for gender, age, facial expression, and identity," in *Proc. ICANN*. Springer, 2005, pp. 569–574.
- [20] C. Wong, D. Kortenkamp, and M. Speich, "A mobile robot that recognizes people," in *Proc. Int. Conf. on Tools with Artificial Intelligence*. Washington, DC, USA: IEEE Computer Society, 1995, p. 346.
- [21] S. Wrede, M. Hanheide, C. Bauckhage, and G. Sagerer, "An Active Memory as a Model for Information Fusion," in *Proc. International Conference on Information Fusion*, no. 1, 2004, pp. 198–205.
- [22] S. Wrede, M. Hanheide, S. Wachsmuth, and G. Sagerer, "Integration and coordination in a cognitive vision system," in *Proc. of International Conference on Computer Vision Systems*. IEEE, 2006.
- [23] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surveys*, vol. 35, no. 4, pp. 399–458, 2003.