Developing visual competencies for socially assistive robots: the HOBBIT approach

K. Papoutsakis, P. Padeleris, A. Ntelidakis, S. Stefanou,
X. Zabulis, D. Kosmopoulos, A.A. Argyros
Institute of Computer Science (ICS)
Foundation for Research and Technology – Hellas (FORTH)
{papoutsa, padeler, ntelidak, stevest, zabulis, dkosmo, argyros}@ics.forth.gr

ABSTRACT
In this paper, we present our approach towards developing visual competencies for socially assistive robots within the framework of the HOBBIT project. We show how we integrated several vision modules using a layered architectural scheme. Our goal is to endow the mobile robot with visual perception capabilities so that it can interact with the users. We present the key modules of independent motion detection, object detection, body localization, person tracking, head pose estimation and action recognition and we explain how they serve the goal of natural integration of robots in social environments.

Categories and Subject Descriptors
I.2.10 **[Vision and Scene Understanding]** *3D/stereo scene analysis, Shape*

I.2.9 **[Robotics]** *Commercial robots and applications*

General Terms
Algorithms, Measurement, Performance, Design

Keywords
Tracking, action recognition, object detection, head pose estimation

1. INTRODUCTION

In this paper we present the design concept and the methods developed to realize the first robotic prototype (PT1) of the HOBBIT project[1]. HOBBIT aims to develop a robotic platform, which will observe visually the users in indoor environments and interpret their actions, so that assistive services can be provided. The development involves a rich set of robot functionalities for human detection, localization, 3D human tracking and action recognition that capacitates the interpretation of important aspects of the user's presence, behavior and intentions. The ultimate goal is to develop effective mechanisms for visual perception and interaction with the users. This interaction is intended to contribute towards the involvement of a *mutual care relation* and bonding between the user and the robot, which constitutes a fundamental concept in the HOBBIT project.
In the next section we briefly present related efforts concerning projects funded by the EU as well as by the NSF. In section 3 we give an overview of the proposed architecture that enables the robot to perceive humans. Section 4 describes the key components that were developed so far, i.e., independent motion detection, object detection, body localization, tracking, head pose estimation and human action recognition. Section 5 gives an overview of our future plans and section 6 concludes the paper.

[1] http://www.hobbit-project.eu

Figure 1: The first robotic prototype (PT1) of the HOBBIT project.

2. RELATED PROJECTS

The development of mobile robots that will be able to provide assistive services has been a goal pursued by several funded projects. Some of the most representative ones are described in the following.

The **KSERA (Knowledgeable SErvice Robots for Aging)** project [6] aims to develop a socially assistive robot to help elderly people. The Nao robot [7] was utilized to enable human-robot interaction, employing various computer vision methods regarding face tracking, motion tracking and head pose estimation. Face tracking enables Nao to detect the location of the user's face and direct its head towards the user, while motion tracking enables it to direct its gaze away from a person to other objects or to a random direction for a certain amount of time during communication.

The **DOMEO (Domestic robot for elderly assistance)** project [8] focused on the development of an open robotic platform for the integration and adaptation of personalized homecare services,

as well as on the cognitive and physical assistance in an AAL-enabled home environment. A novel human detection system has been proposed in the scope of this project, utilizing a laser based leg detector, a body detector and an upper-body detector, both based on vision. Using a grid based approach and Gaussian Mixture Models (GMMs), their output probabilities are fused to provide efficient human detection. Upon detection, the robot may engage itself in a dialogue with a potential speaker exploiting rich vision-based extracted information regarding face detection and tracking, as well as mouth and lips detection and tracking.

The **COGNIRON (Cognitive Robot Companion)** project [9], [10] focused on the development of a robot whose ultimate task was to serve humans as a companion in their daily life. It aims to adapt robot behavior in changing situations and for various tasks. One of the project's prominent objectives regards the detection and understanding of human activities. Detection and understanding of human activity explores modeling, observation and semantic interpretation of human activities in the vicinity of and in interaction with the robot companion. The focus was set on non-verbal characteristics such as position, movement and pose. For humans in the far field, a part-based people detection algorithm has been developed based on omnidirectional camera images to track their location and motion as a whole. A mid-range skeleton based 3D human motion tracking approach has also been developed based on a geometric body model. It relies on depth images acquired using a time-of-flight camera and laser scanner data. Thus, comprehensive perception of humans can be achieved in terms of articulated 3D body tracking and pose estimation.

The **CompanionAble** project [13] aimed to provide the synergy of robotics and ambient intelligence technologies and their semantic integration to provide for a care-giver's assistive environment supporting the cognitive stimulation and therapy management of the care-recipient. A set of key services of the envisioned robotic platform involved day time management, cognitive stimulation / therapy management, detection of critical situations, video-conferencing and situation awareness. Vision and laser sensors are considered and fused to perform vision-based human body observation and pose analysis, fall detection and activity recognition of care-recipient at any time. Moreover, long-term behavior pattern analysis is supported.

The **SRS (Multi-Role Shadow Robotic System for Independent Living)** project [11] was based on the Care-O-Bot robotic platform [12]. Human motion tracking and analysis was a vital part of the perception mechanism of the system enabling the robot to recognize and trace user in the environment. The SRS robot is able to observe the location, pose and actions of a human in order to deduct information about his movements, gestures and intentions. A vision-based mechanism, called Human Presence Sensor Unit, is considered involving a camera with multidirectional view, human motion algorithms based on acquired 3D information and basic gesture recognition algorithms.

Care-O-Bot [13] is a long term project, developed by Fraunhofer IPA and is already in the third generation of the platform. It regards a mobile robot assistant able to assist human in their daily life. Providing open-source interfaces and a rich set of visual sensors, including stereo cameras and a time-of-flight sensor it is utilized by many other research projects as a base to develop advanced technologies and application regarding social assistive robots. **Accompany** [17] is such a project, where an elaborate

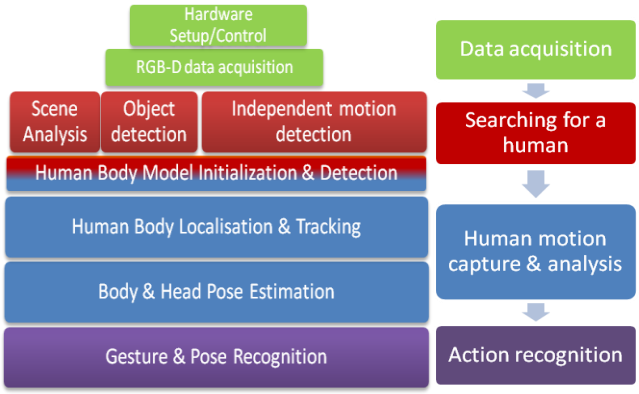


Figure 2: Overview of the system architecture for human observation realized for the HOBBIT (better viewed in color). The green modules correspond to data acquisition, the red ones to searching for humans the blue to human motion capture and analysis and the purple to action recognition.

algorithm for localization of humans using ambient cameras and robot-mounted Laser Range Finders has been developed.

Significant research efforts have also been conducted worldwide on socially assistive robotics by many research laboratories based on several projects and robotic platforms. The Healthcare robotics laboratory [18] at Georgia Institute of Technology has developed efficient methods for human motion analysis and object detection applied to a variety of home assistive robotic platforms.

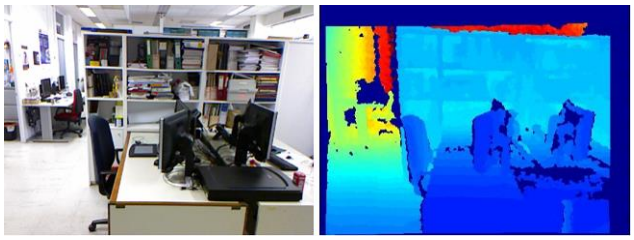
The **HERB** robot [19] is developed in CMU as an autonomous mobile manipulator that performs useful manipulation tasks in the home. A rich source of robotic platforms and state-of-the-art methodologies related to assistive social robots are provided by USC Interaction Lab [20] and ASORO lab at A*Star [21].

The HOBBIT project aims to develop a socially assistive robot for elderly based on a rich set of efficient visual perception capabilities. These will rely, among others, on methods for 3D image/scene analysis, 3D localization of humans and modeling, 3D tracking of human hands and body, posture/gesture and activities recognition, face recognition, 3D head pose and gaze estimation based on RGB-D data and object detection.

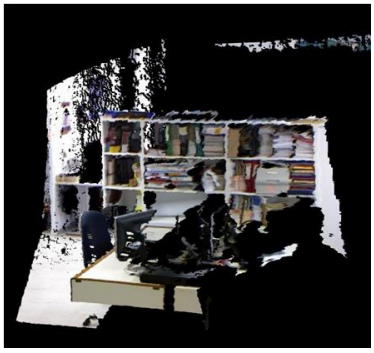
We aspire to integrate the aforementioned methodologies in a unified vision-based framework that will set the robot able to observe visually the users in indoor environments and interpret their actions. The outcome of our framework will enable HOBBIT to enhance the overall performance and the set of capabilities and applications of existing robotic platforms and the quantity and quality of assistive services that can be provided in terms of cognitive and physical assistance in an Ambient Assistive Living home environment. The architecture of our vision-based framework is described in the following section.

3. ARCHITECTURE

The organization of the basic components for human observation for the first prototype (PT1) of the HOBBIT project is illustrated in Figure 2. Several system parts realized in the current implementation of the visual human observation framework of PT1, rely on software components by the OpenNI™ API [4]. The



a b



c

Figure 3: (a-b) An RGB image and the corresponding depth image. Cold/bluish colors correspond to smaller depth values, while warm/red colors to objects further away. Unreliable depth values are set to a zero distance to the sensor and represented with pitch blue. (c) A 3D reconstruction of the scene from an arbitrary view, generated based on that RGB-D image pair. For every valid depth measurement, a 3D point is estimated and associated with the corresponding RGB value. Black areas are due to missing depth data, either due to the sensor's limited FOV or due to noise.

Human observation system of PT1 relies on an RGB-D sensor (Kinect [14] or XtionPro Live [15]). Since the cameras provide a limited field of view that is insufficient for full or upper human body observation, the sensor is attached to a tilt motorized mechanism to enable active exploration of the scene in any direction facilitating vision-based functionalities of the robot. Panning of the sensor is executed by rotating the robot on the spot. A control mechanism was developed to adjust the pose of the sensor to a certain configuration at real-time according to the needs of the system phase that is being executed.

4. COMPONENTS

4.1 Independent motion detection

We propose a method for detecting motions that are independent of the camera. The scene is captured using an RGB-D sensor. At any time, an object may move in the scene while the sensor is moving as well. The estimation of the dominant motion allows for employment of ego-motion estimation and camera stabilization, which in turn leads to the detection of motion that is independent to the sensor's motion. These techniques provide vital input and boost the robustness of algorithms applied in robotics including visual odometry, scene understanding, human detection and action recognition when the employed sensor is mounted on a

Figure 4: Correspondence establishment of 3D points between different scene views. Top-left: SIFT features are extracted on the RGB image. Top-right: The SIFT features are associated with the depth image so that a depth value for each SIFT feature is identified. Middle: SIFT features with bi-linearly interpolated depth values are indicated with a purple circle. Bottom: The establishment of correspondences between two different RGB-D pairs is performed by employing KNN similarity matching of the descriptors of SIFT features with valid 3D measurements.

moving platform. The proposed method consists of the following steps:

- Establishing correspondences of 3D points
- Dominant motion estimation
- Independent motion detection

The proposed method uses a Kinect [14] or an XtionPro Live [15] RGB-D sensor to acquire a temporal sequence of RGB and depth image pairs. The sensor is calibrated to allow the association of the RGB and the depth values. We use the sensor's intrinsic calibration parameters to estimate a 3D point in Euclidian space for every valid pixel of the depth image, and subsequently associate it with the corresponding pixel on the RGB image (Figure 3). The first allows various 3D representations of the scene within the sensor's field of view (FOV). The second allows the establishment of correspondence of 3D points among point clouds of different views of the scene, via well-established approaches applied to conventional cameras. Once correspondences between 3D points of two different views of the same scene are established, we are able to estimate the dominant 3D motion between the two views.

4.1.1 Establishing correspondences among 3D points

In order to estimate the dominant 3D motion that best describes the relation between two point clouds generated from two different RGB-D image pairs, 3D matches between these point clouds need to be established. Such correspondences are established by employing feature extraction and matching techniques on the RGB images. More specifically, SIFT features are detected and matched on each RGB image following the implementation of [1]. We then estimate the corresponding 3D points taking into account that the image coordinates of the features are in sub-pixel accuracy. Specifically, we perform bilinear interpolation using depth values from the four corresponding neighboring pixels on the Depth image grid. We use the depth interpolation result to associate it with the SIFT feature and estimate the corresponding 3D point. If there is no valid value on one of the neighboring depth pixels we eliminate the feature from the global features list. Finally, we perform nearest neighbors matching (KNN) on the remaining features and establish, indirectly, 3D correspondences between the two point clouds. The overall 3D points matching pipeline is illustrated in Figure 4.

4.1.2 Dominant motion estimation

Once sets of 3D matches are established, the dominant motion between the two views of a 3D scene can be estimated. Essentially, the problem that needs to be solved is the problem of robust registration of two rigid 3D point clouds. This is a well-studied problem in the computer vision literature. Our approach for solving it employs the generalized Least Squares fitting algorithm described in [2]. During the computation, we also employ RANSAC [5] to detect outliers that are due to independently moving objects. Specifically, out of the initial 3D point sets, we repeatedly draw four random pairs of corresponding points at time to estimate the rotation $\boldsymbol{R}$ and the translation $\boldsymbol{t}$ between these 3D point clouds.

4.1.3 Independent motion detection

Having established the $\boldsymbol{R}$, $\boldsymbol{t}$ parameters of the dominant motion between two different views v_1 and v_2 of a scene as well as the 3D depth values, we can generate a synthetic view v_1' of v_1 as this is observed from v_2. v_1' and v_2 should be identical, provided that there is no independent motion occurring between the two views. If an object moves independently to the sensor, then its image points will not be correctly registered. . This can be exploited to detect independent motion, because after registration, a conventional change detection algorithm can be applied to v_1' and v_2 to detect such objects . In Figure 5 (see caption for details) we demonstrate the results of such a simple change detection technique (image differencing) applied to the target and the synthetic depth image. Differences greater than a threshold are assumed to belong to independent motion.

The complete pipeline of the independent motion detection (IMD) module was tested on a PC with an Intel i7 870 processor, an NVidia GeForce GT330 graphics card and 8GB of RAM. All steps of the algorithm are currently implemented on CPU, with the exception of the KNN similarity matching which is implemented on GPU. The IMD module operates at ~1 fps. The usage of less computationally demanding ORB features allowed the IMD module to operate at ~4 fps.

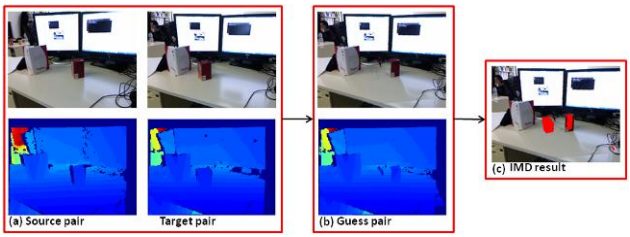


Figure 5: Independent motion detection between a source and a target RGB-D pair. (a) The RGB-D pairs. The red box in the center of the scene moves independently to the sensor. (b) The synthetic image pair generated based on the estimation of the dominant motion between the source and the target pairs. This is identical to the target pair at all points except from the ones corresponding to the independently moving object. (c) Change detection results between the target and the synthetic pairs. The detected foreground (shown in red color) corresponds to image regions with independent motion.

The proposed methodology proves sufficiently robust in estimating the dominant motion between point clouds corresponding to RGB-D image pairs. The use of other feature extraction and matching algorithms like the FAST algorithm [3] can be considered. Generation of synthetic RGB-D pairs with respect to the dominant motion is currently performed without sub-pixel accuracy and without reasoning about possible occlusions. This results in inaccuracies in the generated synthetic RGB-D pair. To improve this, the generation of the synthetic view needs to be performed with standard rendering techniques.

4.2 Object Detection

For object detection we introduced a novel method, which is described in detail in [24]. The method is scale and rotation invariant and exploits RGB information. This method provides valuable information regarding detection and localization of foreground objects and can be used for efficient detection of humans and faces, The proposed method represents an object as a Histogram of Oriented Gradients (HOG) [22]. HOGs have proven to be robust object descriptors. A variant of an existing rotation invariant HOG-like descriptor is proposed, while object detection and localization is formulated as an optimization problem that is solved using the Particle Swarm Optimization (PSO) [23]. A series of experiments demonstrates that the proposed approach results in considerable performance gains without sacrificing object detection and localization accuracy. Illustrative examples from the operation of the proposed method are provided in Figure 6.

The proposed method enhances object identification and manipulation. HOBBIT is planned to support learning of unknown objects and maintain a considerably large dataset of known objects which it will be able to detect, fetch and carry. Therefore, efficient object detection in domestic environments (i.e. floor, table etc.) facilitates object identification and the subsequent procedures of object grasping/manipulation.

Figure 6: Representative object detection results of the proposed approach [24]. In each row, the leftmost item indicates the query object and the subsequent items indicate the detection result for various images containing the object of interest.

Figure 7: Illustration of 3D skeleton tracking and pose estimation results for two RGB-D data frames. The human body is being tracked with respect to the 3D skeleton model and body poses are estimated based on the acquired depth data for each frame. Only the body limbs and joints are drawn in these sample colored images.

4.3 Body Pose Estimation

Body pose estimation is closely related to 3D skeleton tracking, both provided by computer vision algorithms of NITE[2] [16]. In each frame, the previous estimated body pose is being tracked to coarsely follow the new acquired depth data of an identified human body. The body estimation process performs a readjustment of the tracked 3D positions/orientations of body limbs/joints after the tracking task is accomplished, estimating the body configuration for that frame. An illustration of both tracking and pose estimation results is provided in Figure 7. Two frames are shown, where tracking of the detected human body and estimation of the body pose were performed. Thus, the trajectories

[2] NITE is an OpenNI compliant middleware component that perceives the world in 3D, based on data captured by a PrimeSense 3D sensor (Kinect and XtionPro Live). It is a freely available and proprietary (closed-source) software package that includes both computer vision algorithms that enable identifying users, tracking their movements and recognizing gestures/poses, as well as a framework API for implementing Natural Interaction UI controls that are based on user gestures.

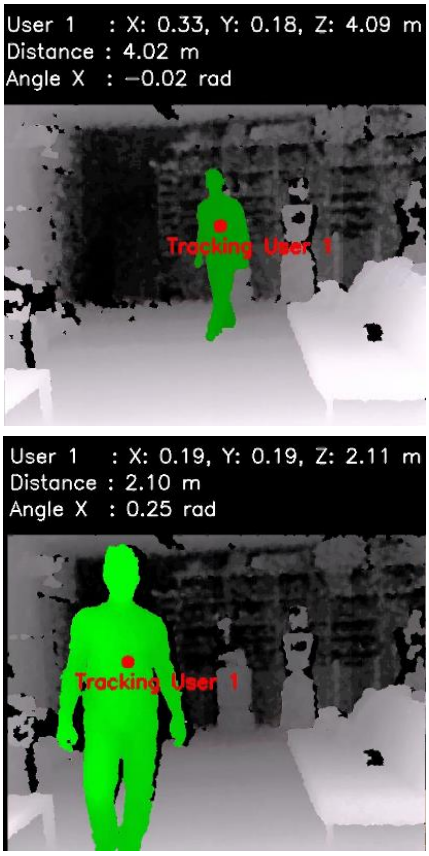


Figure 8: A detected human body is localized and tracked in the scene based on depth data acquired by the camera. Localization information is provided at the upper part of each depth frame regarding the 3D position, distance and angle of the user with respect to the camera.

of body limbs and the 3D rotations of body joints are available for each moment in time.

4.4 Body Localization & Tracking

Given the human body initialization and detection results of the previous task, 3D skeleton tracking is performed by estimating the new 3D positions and orientations of body limbs and joints in every frame based on the acquired depth data.

Body localization is performed by acquiring the 3D position of the body torso. Moreover, the distance to the sensor is calculated by projecting the 3D position of the torso joint to the extracted 3D floor plane. Based on the extracted information, a 3D bounding box can easily be calculated for each frame based on the extracted depth data of user's body. Emergency detection in case of user's fall can be performed by analyzing the length, velocity and acceleration of each dimension of the calculated 3D bounding box. Moreover, 3D body skeleton tracking provides a rich set of information that enables efficient inspection of body movements and significantly enhances action recognition. An interesting user-driven application that HOBBIT will support based on this information regards visual-assisted rehabilitation and physical therapy/exercise of elderly in their homes.

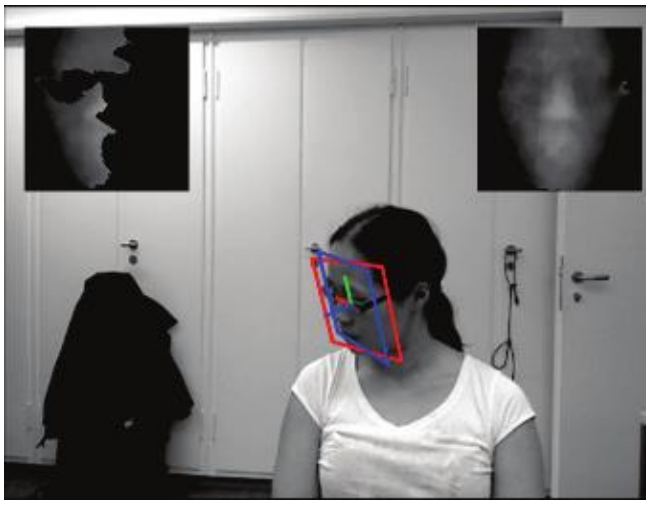


Figure 9: Head 3D pose estimation based on depth data [25]. A sample result from the experimental evaluation of the method is superimposed in the grayscale image of a human subject. The 3D head pose estimated by the proposed algorithm is shown in blue. The solution in red is the one estimated by the method described in [27].

4.5 Head Pose Estimation

Head pose estimation is an important aspect of human observation. NITE provides information regarding head pose by acquiring the position and orientation of the head joint of the employed skeletal body model. However, the efficiency of that information heavily depends on the estimation of the rest of the skeleton model and is not full, i.e., does not provide access to all the degrees of freedom of head motion. Therefore, it is inadequate to provide high accuracy and estimation of individual head movements that could indicate human's intentions.

We designed, implemented and evaluated a novel approach for 3D head pose estimation based on depth data of a detected face [25]. The method searches the 6-dimensional pose space to find a pose from which the head appears identical to a reference view acquired at initialization. This search is formulated as an optimization problem whose objective function quantifies the discrepancy of the depth measurements between the hypothesized views to the reference view. Particle swarm optimization (PSO) is utilized to search for a maximum of the objective function. The proposed method outperforms existing methods in accuracy. It is robust and tolerant to occlusions and handles head pose estimation in a wider range of head poses. Sample results of the method are shown in Figure 9.

This method will be part of the system for visual human observation enhancing functionalities provided by the robot. Combining the accurately estimated 3D head pose and the 3D direction of a pointing arm from the estimated skeleton body model both looking and pointing to a specific area in 3D, a 3D point or area in space could be determined indicating, for example, an object of interest to the user.

4.6 Action Recognition

Action recognition is a module that utilizes services provided by the previously described modules. The gesture and pose recognition components that compose this module are responsible for recognizing a number of predefined gestures or poses performed by the detected human using his full body or his body parts. Gestures refer to a series of poses performed within a time window of configurable length. This task utilizes methods for gesture and pose recognition which are available by the NITE algorithms layer, exploiting the rich set of information computed by the 3D skeleton tracking and pose estimation tasks in the previous system phase. A limited number of predefined gestures and poses are supported. These gestures and poses are utilized and mapped to specific commands and tasks to be executed by the robot. The supported gestures and poses are "Hand-Push", "Swipe up/down", "Swipe left/right", "Circle", "Waving", "Raise Hand", "Cross hands", "Hands-up", and have been determined based on extended studies of the user needs in the context of HOBBIT scenarios. Based on the recognition of these actions, an efficient vision-based user interface is feasible, mapping specific actions to robot commands, thus enabling natural interaction between the user and the robot. We refer an example of mapping between user actions and robot commands as follows: "Help-the-user" robot command is initiated after a "Cross hands" user action in front of the robot, "Stop task" robot command for a "Hand-push" user action, "Come here" robot action for a "Waving" user action, "Localize and grasp object" robot command for a "Raise Hand" and pointing with the other hand to an object of interest by the user, answering "Yes/No" dialogues with the robot using "Swipe" gestures.

Vision-based extraction of relevant information based on head pose and gaze direction estimation, object detection and scene segmentation can be integrated with information based on recognized gestures/postures to enhance the performance action/activity recognition of the user and facilitate other useful user-driven applications.

5. NEXT STEPS

Several extensions for vision-based human observation are planned within HOBBIT:

- Develop new approaches to human motion capture and gesture/posture recognition, especially for the case where a human is not fully visible by the employed camera. This will facilitate human-robot interaction, as now the distance of the robot to the camera needs to be inconveniently large for the camera to have a relatively complete view of the users' body.
- Provide a more elaborate and effective fall detection mechanism to be used for emergency detection. It is also important to be able to detect humans already lying on the floor. This will enable HOBBIT to detect emergency situations and events that did not occur in front of its camera.
- Interact with Ambient Assistive Living sensors installed in the environment (AAL-enabled home) to provide enhanced human detection, localization and tracking. This will further improve the detection rate of emergency situations.

- Gestures:
  - Hand gestures: A number of predefined or custom gestures (either single handed or bimanual) performed by a user could be recognized, providing richer interaction with the robot via a gesture recognition interface.
  - Custom body gestures/poses: A larger set of predefined gestures performed by user using his whole body or parts could be recognized.
- Support of specific applications such as:
  - Gesture control-based computer games: Users could perform body movements and gestures to play a computer game (drawing, pong, pac-man, shooting etc) shown on robot's touch screen.
  - Vision-based observation for special physical exercise programs for rehabilitation: HOBBIT may provide videos on its touch screen showing experts to perform physical exercises for rehabilitation of various health issues according to user's needs. The robot may prompt user to follow these movements providing spoken instructions. At the same time user's position, movements of the body parts will be captured, analyzed and compared to recorded "ground truth" movements/patterns to provide supervision and further spoken instructions for better performance on these exercises.

6. CONCLUSIONS

We presented our approach towards developing vision-based perceptual capabilities for socially assistive robots within the framework of the HOBBIT project. We described the key modules of independent motion detection, object detection, body localization and tracking, head pose estimation and action recognition and we explained how they serve the goal of natural integration of robots in human environments. Furthermore, we presented our plans for future work, towards more elaborate and robust visual competencies that will enable the natural, vision-based interaction of humans with socially assistive robots.

7. ACKNOWLEDGMENTS

This work is partially funded by the European Commission under contract FP7-IST-288146 HOBBIT. The head pose estimation has also received funding by the EU IST-FP7-288917 project DALi.

8. REFERENCES

[1] D.G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[2] G. Wen, Z. Wang, S. Xia, D. Zhu, "Least-squares fitting of multiple M -dimensional point sets", The Visual Computer, 22, 6 (2006), pp.387-398.

[3] E. Rosten, T. Drummond, "Machine learning for high-speed corner detection", European Conference on Computer Vision, 2006.

[4] http://www.openni.org

[5] Martin A. Fischler and Robert C. Bolles (June 1981). "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". Comm. of the ACM 24 (6): 381–395.

[6] http://ksera.ieis.tue.nl/

[7] http://www.aldebaran-robotics.com/en/

[8] http://www.aal-domeo.eu/

[9] http://www.cogniron.org

[10] http://www.cogniron.org/review2-open/files/Cogniron_D2.2005_RA2.pdf

[11] http://srs-project.eu/

[12] http://www.care-o-bot.de/

[13] http://www.companionable.net

[14] http://www.microsoft.com/en-us/kinectforwindows/

[15] http://www.asus.com/Multimedia/Motion_Sensor/Xtion_PRO_LIVE/

[16] http://www.primesense.com

[17] http://www.accompanyproject.eu/

[18] http://healthcare-robotics.com/

[19] http://personalrobotics.ri.cmu.edu/index.php

[20] http://robotics.usc.edu/interaction/

[21] http://www.asoro.a-star.edu.sg/index.html

[22] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, San Diego, USA (2005).

[23] Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: IEEE Int'l Conf.on Neural Networks. (1995) 1942-1948.

[24] S. Stefanou, A.A. Argyros, "Efficient Scale and Rotation Invariant Object Detection based on HOGs and Evolutionary Optimization Techniques", in Proceedings of the International Symposium on Visual Computing, ISVC 2012, Rethymno, Crete, Jul. 16-18, 2012.

[25] P. Padeleris, X. Zabulis and A.A. Argyros, "Head pose estimation on depth data based on Particle Swarm Optimization", in Proceedings of the Workshop on Human Activity Understanding from 3D Data (HAU3D'2012) in conjunction with CVPR 2012, Rhode Island, June 21, 2012.

[26] Noury, N.; Fleury, A.; Rumeau, P.; Bourke, A.K.; Laighin, G.O.; Rialle, V.; Lundy, J.E.; , "Fall detection - Principles and Methods," Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, vol., no., pp.1663-1666.

[27] G. Fanelli, T. Weise, J. Gall, and L. V. Gool. Real time head pose estimation from consumer depth cameras. In DAGM, 2011