A Healthcare Mobile Robot with Natural Human-robot Interaction

J. Liu, J. Correa, S. McKeague, E. Johns, C. Wong, A. Vicente, G.-Z. Yang

The Hamlyn Centre, Imperial College London

[j.liu, jc310, sjm05, ej09, charence, apv110, gzy]@imperial.ac.uk

INTRODUCTION

The ageing population presents a significant challenge to the future of healthcare. By 2050, one-in-four people living in the UK will be over the age of 65 [1]. One potential solution to this challenge is the development of assistive mobile robots. Although the very concept of mobile assistive robots have been explored for the last decade [2], most systems are still restricted to laboratory use. Two of the biggest challenges of the current systems are the lack of (i) a natural means of Human-Robot Interaction (HRI) and (ii) an adaptive autonomous scheme for the robot to deal with dynamic environments. In this paper, a mobile robot that provides natural human-robot interaction and an adaptive learning ability is presented.

Five key technical issues are explored in the system: (i) autonomous navigation with human-like behaviour, (ii) gesture recognition within crowded environments, (iii) distant speech recognition with room reverberation and background noise, (iv) adaptive visual scene recognition and (v) the integration of Body Sensor Networks (BSN) with mobile robots. Figure 1 shows a concept of the system when used in a hospital environment, demonstrating the range of different modes of interaction between human and the robot.

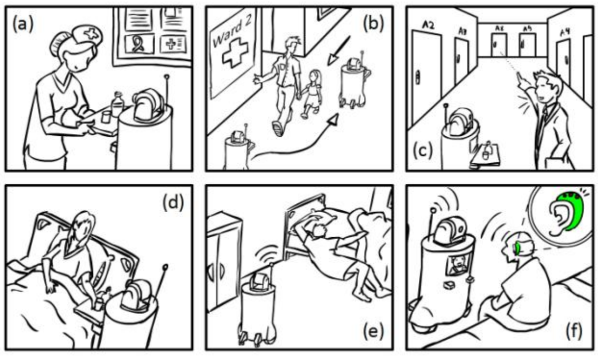

Figure 1: Concept of a healthcare robot in a hospital environment. The robot delivers drugs (a). En route, it avoids people with minimal disturbance (b). The robot understands natural gesture and speech commands (c) and plans the route to the destination (d). Emergency situations, such as falls detected through pervasive sensors, can alter a robot's priorities (e) and the robot can act as a remote presence between doctors and patients (f).

SYSTEM ARCHITECTURE

Figure 2 shows a schematic diagram of the structure of the whole system. It is hierarchically organised between low-level sensor inputs and modular functions. In this paper, the main emphasis is on HRI and adaptive algorithms for dynamic environments.

Robot Navigation using Human-like Behaviour

For assistive robots, failing to consider the dynamics of an environment or the implicit social rules while navigating can yield sub-optimal solutions. In the worst case scenario, the robot may freeze as no valid solution can be found. In practice, the robot is not only expected to move in the environment, but also to follow some basic social rules. Not getting to close to people or disturb a group of people talking or moving together is undesirable. To take into account the movement of people and social rules, the Dynamic Window Approach (DWA) cost function was extended to include a dynamic and a social component. The dynamic component takes care of avoiding moving people in the environment, while the social component takes care of doing it in a socially acceptable way. Simulations were carried out to test the performance of the proposed social navigation approach. These simulations show the robot is able to complete a path in a dynamic environment faster, while keeping a greater distance to people. To compare our approach, the classic DWA with no social or dynamic considerations was used.

Reverberation-robust Speech Recognition

Although artificial speech recognition has been well explored for human-computer interaction for many years, it still has many limitations when applied to mobile robots in realistic environments with unavoidable artefacts due to reverberation. Existing models trained with anechoic speech signals can easily deteriorate when people talk to the robot located a few meters away. The speech recognition needs to be robust to these changes as the robot roams around in different rooms. Inspired by the finding on precedence effects of humans, an attenuated statistical room impulse response (IR) model is proposed that can match the real room IR according to the reverberation time T_{60}, i.e. the time it takes for the energy of the sound to reduce by 60 dB after a sound source is switched off. The proposed speech recognition model is trained with simulated room IRs with different T_{60}s and tested on real room IRs. See system details and results in [4].

Gesture Recognition in Cluttered Background

Gesture recognition provides a natural means for non-technical users to command and control mobile robots. This is particularly important in noisy, dynamic environments. For example, a doctor may want, for example, to point towards the room where he wants a robot to travel to in order to clarify their verbal command.

Several sub-problems have been explored in the overall gesture recognition system. These include the detec-

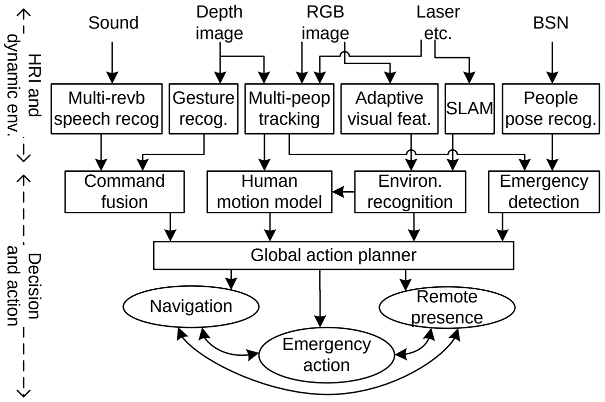

Figure 2: Schematic system structure

tion, tracking and association of hands and people. Traditional approaches to these problems have used colour cameras as an input source, which introduces problems such as illumination constraints and depth ambiguity. The proposed robot uses recently introduced depth cameras to overcome these problems and increase the discriminative capacity of the system.

Object detection in crowded environments is a challenging problem. We have introduced a hand detector specifically designed for robust detection in such situations. Furthermore, to compensate for the increased occlusions and detection ambiguities in our target environment, our hand-body association method makes use of available temporal information.

Adaptive Visual Scene Recognition

Visual scene recognition has a wide range of applications in mobile robotics including appearance-based Simultaneous Localisation and Mapping (SLAM), loop closure in metric SLAM and interaction with the environment. For ensuring natural behaviour in practical applications, it is important that the robot can update its long-term understanding of the environment's appearance as dynamic elements appear in or disappear from a scene. An approach whereby local visual features extracted from an image are tracked across adjacent nodes in a topological map and quantised to form visual words is adopted. A graphical model is then learned for the appearance of each tracked feature by incorporating the relative locations of other contextual features in an image. As dynamic changes (both short-term e.g. people and long-term e.g. furniture rearrangements) occur in the environment, the likelihood of features occurring and co-occurring can be updated to reflect the new scene appearance. Furthermore, for outdoor scenes, the appearance of each feature is learned with respect to the time of day so that illumination effects such as shadows and reflections can be incorporated.

Integration with Wearable and Ambient Sensors

Pervasive wireless sensors worn by a patient can be used to assess wellbeing. For example, to assess the likelihood of an elderly patient falling [7] and to detect falls [8]. Pervasive sensors have also been used to detect activity, monitor patient recovery and perform biomechanical analysis. Support for Body Sensor Networks (BSN) enables the robot to obtain a much more in-depth insight of a patient's health. Whilst certain aspects may be detected solely through the robot's on-board sensors, previous work demonstrated the advantages of incorporating pervasive sensors to classify abnormal gaits [9].

DISCUSSION

A mobile robot has been built to test the proposed system and each modular function. Figure 3 shows a remote interface and the robot profile. The remote user can control the robot in either autonomous or manual mode. Information about the robot, its surroundings and nearby patients can be accessed via the interface.

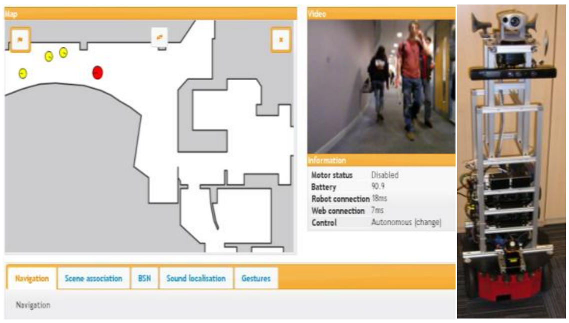

Figure 3: The remote presence interface (left) and the uncovered health care robot (right).

In future, further work will be directed to human attention detection, recognition of subtle gestures, environment dependent speech and behaviour learning, and deeper integration of BSN health monitoring with the mobile robot. More detailed results of our previous work can be found in the cited works [4][5][7][8][9].

REFERENCES

[1] M. Lali and L. Atkinson, "The Ageing Population," 2010.

[2] Y. Sakagami, R. Watanabe, et al, "The intelligent ASIMO: system overview and integration," in *IROS* 2002, pp. 2478-2483 vol.3.

[3] T. Djelani, J. Blauert, and A. Seestern, "Modelling the direction-specific build-up of the precedence effect," 2002.

[4] J. Liu, N. D. Gaubitch, and G.-Z. Yang, "Robust Speech Recognition in Reverberant Environments Using Multi-model Selection," in *IROS2012 (Submit)*

[5] E. Johns and G. Z. Yang, "Place Recognition and Online Learning in Dynamic Scenes with Spatio-Temporal Landmarks," BMVC 2011, pp. 10.1-10.12.

[6] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," 2003, pp. 1470-1477 vol. 2.

[7] R.C. King, L. Atallah, C. Wong, F. Miskelly, and G.-Z. Yang , "Elderly Risk Assessment of Falls with BSN," in *BSN, 2010*, pp. 30-35.

[8] M. Tolkiehn, L. Atallah, B. Lo, and G.-Z. Yang, "Direction Sensitive Fall Detection Using a Triaxial Accelerometer and a Barometric Pressure Sensor," 2011, pp. 369-372.

[9] C. Wong, S. McKeague, J. Correa, J. Liu, amd G.-Z. Yang, "Enhanced Classification of Abnormal Gait Using BSN and Depth," in *BSN, 2012 (to appear)*.