# Towards Socially Assistive Robots for the Elderly

Chieh-Chih Wang, Shao-Wen Yang, Ko-Chih Wang, Yu-Chun Wu, Jiun-Fu Chen and Chung-Che Yu

***Abstract*— The NTU-PAL1 and NTU-PAL2 robots for the elderly are introduced in this paper. The emphasis of NTU-PAL1 is offering increased mobility for elderly people who are less-mobile. NTU-PAL2, an emotional expressive robot, is used to stimulate interaction, to supply relaxed companionship and to fulfill a sense of affection for senior citizens staying alone. The core software modules in terms of dynamic scene understanding and human understanding are summarized to demonstrate our progress towards socially assistive robotics for the elderly.**

## I. Introduction

Nowadays in Taiwan and other countries around the world, elderly care is an imperative issue since our society appears to be aging rapidly. Statistics show that citizens over 65 years old are accounting for 10.4 percent of the total population. It is estimated that the number will grow to 20 percent by 2025. Resulted from the aging trend is a dropping aged dependency ratio - the ratio of the number of people in the workforce to the number of people aged 65 and over. The averaged number of active workers supporting one elderly person is currently around seven and is going to drop to 3.2 by 2026. This demographic shift brings us to face the problem that a growing number of elderly people demand urgent attention but insufficient human resource is available. Aiming to relief this burden and improve the living quality of senior citizens, we have been undertaking the development of socially assistive robots for elderly care. It is our hope to promote elderly independence and provide company in their daily life with the help of robotics technology.

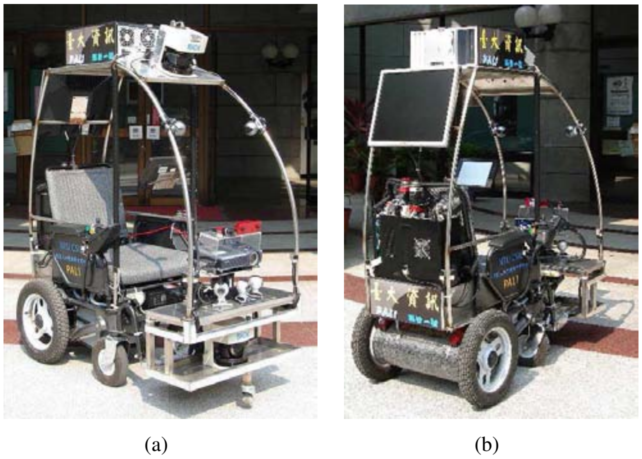


Fig. 1. The NTU-PAL1 robot.

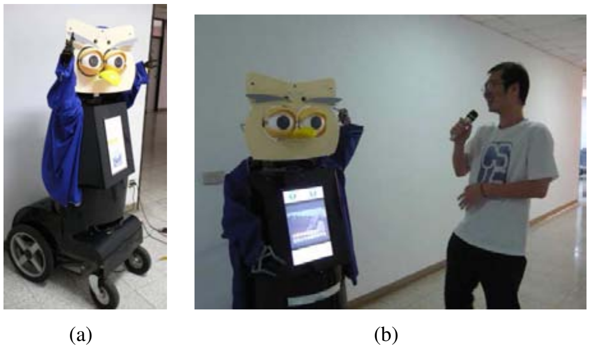


Fig. 2. The NTU-PAL2 robot.

For achieving the goals of socially assistive robotics, we have built two robotic systems with advanced perception and action capabilities. The first robot for the purpose of elderly care is the NTU-PAL1 robot, which laid its emphasis on offering increased mobility for elderly people who are less-mobile. Figure 1 shows the intelligent wheelchair robot. Heterogeneous sensors such as cameras, laser scanners and sonar equipped on this robot are integrated, and the corresponding algorithms are implemented to give assistance to the elderly in advanced perception. A camera-projector system was developed for human robot interaction as well as perception. In addition, the electric wheelchair-based platform makes it feasible for our NTU-PAL1 robot to move in both indoor and outdoor environments in which the elderly could obtain greater action capabilities. By enhancing their mobility, the elder users can achieve higher independence and will be exposed to more opportunities to interact with and connect to their social surroundings.

Our second robot, NTU-PAL2, is an emotional expressive robot as shown in Figure 2. This robot is designed to have an expressive face and other actuators such as arms. NTU-PAL2 can provide functionalities that assist daily life of the elderly such as schedule reminder, memory training and entertainment. Intended to serve as a companion for the elderly that needs to perform a lot of high-level interactions with users every day, this robot is given an animal-like appearance and the ability to present emotions to make interacting with the robot more natural and interesting. The human-like but simplified behaviors of the robot stimulate interaction, supply relaxed companionship and fulfill a sense of affection for senior citizens staying alone.

C.-C. Wang is with the Department of Computer Science and Information Engineering, and the Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan bobwang@ntu.edu.tw

S.-W. Yang, Y.-C. Wu and J.-F. Chen are with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan {any,yuchun,jeffchen}@robotics.csie.ntu.edu.tw

K.-C. Wang and C.-C. Yu are with the Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan {casey,fish60}@robotics.csie.ntu.edu.tw

With the use of these two robots, we have been developing the core perception and action capabilities such as motion control in dynamic environments, dynamic scene understanding and human understanding which are essential to socially assistive robotics. Regarding motion control in dynamic environments, imitation learning is applied to increase the flexibility of the existing obstacle avoidance algorithm such as the nearness diagram navigation algorithm as well as to reduce the burden of parameter tuning. In [1], a mapping between the environment information and the control comments from humans is learned and is later used to control the robot. It is our hope that the elderly may feel more comfortable with the proposed approach compared the classical control approaches based on predefined cost functions.

In this paper, the core perception software modules are summarized to demonstrate our progresses towards socially assistive robotics. Regarding dynamic scene understanding, our ladar-based solutions to the mapping, localization and tracking problems are described in Section III. In addition, our probabilistic structure from sound (PSfS) and probabilistic sound source localization (PSSL) algorithms using microphones are addressed. With regard to human understanding, 3D face alignment in 2D images and hand posture recognition using cameras are described in Section IV. In addition to these perception software modules, the camera-projector system and the expressive robot face and arms are described firstly in Section II.

While the wheelchair robot increases physical mobilities of the users and thus enlarges the world they can touch, the emotional expressive robot offers both useful assistance in life and certain level of mental care. We have applied robotics technology to construct systems which cope with issues of elderly care we are confronted with in our aging society. Although a robot can never replace a human being, the existence of these robots still ease human resource shortage and contribute to better life quality when designed and applied with care.

## II. Systems for Human Robot Interaction

Instead of describing the whole hardware systems of NTU-PAL1 and NTU-PAL2, only systems for human-robot interaction are addressed. In this section, the camera-projector system on NTU-PAL1 and the expressive robot face and arms on NTU-PAL2 are described.

### A. The Camera-Projector System

The camera-projector systems have been shown to be effective for human computer/robot interaction as illustrated in Figure 3. However, an ideal display surface for a projector is often unavailable for robotics applications. With the geometry and color of the scene, projected images or videos can be adjusted accordingly. It has been shown in the computer vision literature that the camera-projector systems can be a low-cost 3-dimensional range sensor in which the geometry and color of the scene can be obtained via triangulation. The environment-adaptive display can be accomplished in which it is feasible to project images on curved and non-planar surfaces using the camera-projector systems. Figure 4 shows a result using the one-shot scanning approach with our system [2]. An active stereo system was also implemented in which a digital light processing (DLP) projector and a stereo camera are used. Figure 5 shows the result of our active stereo system in which the featureless regions are well modeled.

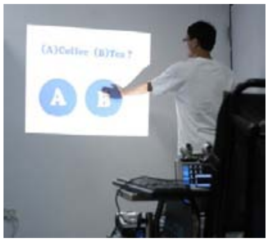


Fig. 3. The camera-projector system for human robot interaction.

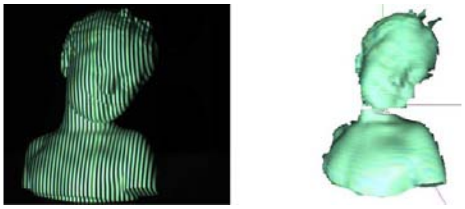
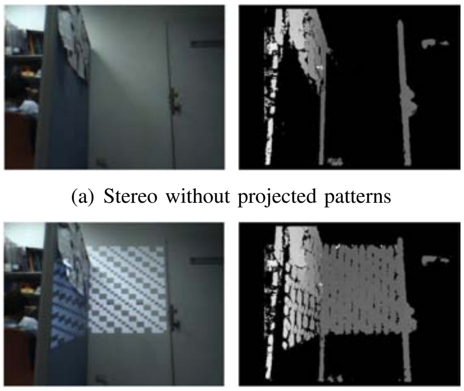


Fig. 4. The one-shot scanning approaches.

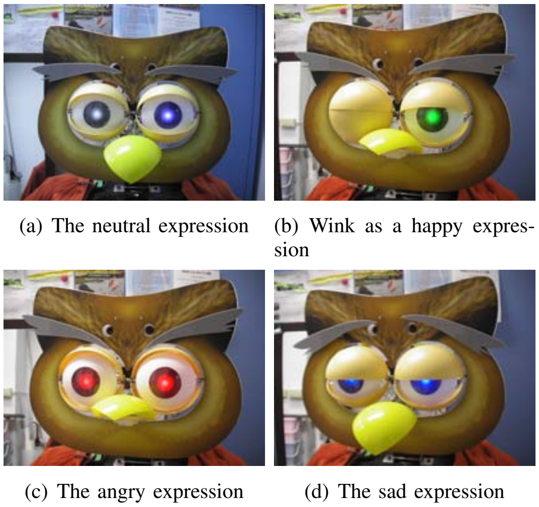
Our experiments show that the camera-projector system with the proposed algorithms can be reliable for 3D modeling, and can be a useful device for human robot interaction. However, we observed that the data association problems of the camera-projector systems in ambient lighting environments are daunting [2]. The power consumption is another critical factor to use the camera-projector systems on mobile robotic platforms. New projector systems such as small portable laser-light projectors may effectively resolve some of these issues.



(a) Stereo without projected patterns

(b) Stereo with projected patterns

Fig. 5. The active stereo approach.



(a) The neutral expression

(b) Wink as a happy expression

(c) The angry expression

(d) The sad expression

Fig. 6. Emotional expressions of the robot face.

### B. Expressive Robot Face and Arms

It is important to provide the user a simple and comfortable channel to recognize the state of the robot [3]. To this end, we designed an expressive robot face and arms on our NTU-PAL2 robot to mimic a natural way that humans use to communicate every day. As shown in Figure 6, the expressive robot face has seven servo motors mounted to control two eyebrows, four eyelids and its beak respectively. Furthermore, it has full color LED lighting inside the eyes. Subtle movements of these facial features, together with congruous colors of the lighting [4], we are able to endow the robot with the ability to present several different emotional expressions such as happy, angry and sad. These emotional expressions are useful since traditional text or voice channel to convey messages can be replaced. For example, when a user is performing an improper action which is likely to damage the robot system, an angry expression can be used to warn the user to stop the unwanted action. In addition to the robot face, a set of two arms is also available to help the robot make use of body language. Each arm has six degrees of freedom, making the robot able to imitate human gestures such as waving. Figure 7 shows a shot of our NTU-PAL2 robot showing a waving gesture.

The robot face and arms are chosen for the reason that they convey non-verbal messages that are common in daily usage and easy to understand. Moreover, this kind of representation is eye-catching and often elicits emotional responses from the user, and thus provides more enjoyable interacting experiences [5].

## III. Dynamic Scene Understanding

Establishing the spatial and temporal relationships among a robot, stationary objects and moving objects in a scene serves as a basis for scene understanding. In this section, we summarize the capabilities of NTU-PAL1 and NTU-PAL2 in terms of dynamic scene understanding. In particular, our ego-motion estimator, interacting object tracking framework and probabilistic structure from sound algorithm are described.

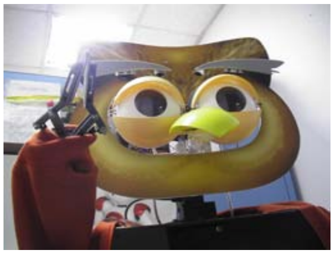


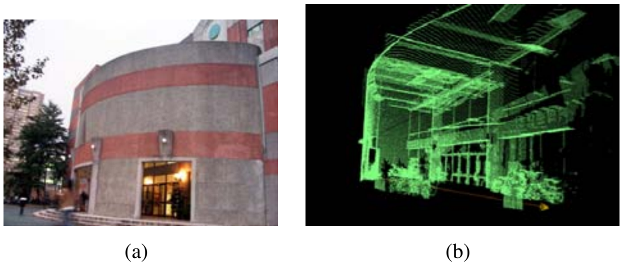
Fig. 7. The robot showing a waving gesture.

### A. Mapping and Localization

Localization is the process of establishing the spatial relationships between the robot and stationary objects and mapping is the process of establishing the spatial relationships among stationary objects, and moving object tracking is the process of establishing the spatial and temporal relationships between moving objects and the robot or between moving and stationary objects.

We previously established a mathematical framework to integrate SLAM and moving object tracking [6] in which two solutions are described: SLAM with generalized objects, and SLAM with detection and tracking of moving objects (DATMO). SLAM with generalized objects calculates a joint posterior over all generalized objects and the robot. Such an approach is similar to existing SLAM algorithms, but with additional structure to allow for motion modeling of generalized objects. Unfortunately, it is computationally demanding and generally infeasible. SLAM with DATMO decomposes the estimation problem into two separate estimators. By maintaining separate posteriors for stationary objects and moving objects, the resulting estimation problems are much lower dimensional than SLAM with generalized objects. The implementation of SLAM with DATMO was previously demonstrated using laser scanner data collected from the CMU Navlab11 vehicle at high speeds in crowded urban environments [7]. Figures 8 and 9 show 3D (2.5D) outdoor maps of our department building and the Palm Tree Boulevard in National Taiwan University using data collected from NTU-PAL1 and Figure 10 shows an indoor map of the 4th floor of our department building. Two SICK LMS 291 laser scanners mounted on NTU-PAL1 were used to collect data for building 3D maps. In a number of applications, 2D maps could be sufficient. Figure 11 shows a 2D map of the first floor of the MSRL at Taiwan ITRI.

In the SLAM with DATMO framework, a robust ego-motion estimator is essential. In [8], we proposed a random sample consensus (RANSAC) based ego-motion estimator to deal with highly dynamic environments using one planar laser scanner. Instead of directly sampling on individual measurements, the RANSAC process is performed at a higher level abstraction for systematic sampling and computational efficiency. We proposed a multiple model approach to solve the problems of ego-motion estimation and moving object detection jointly in a RANSAC paradigm.



(a)

(b)

Fig. 8. 3D Mapping: The CSIE building in NTU.

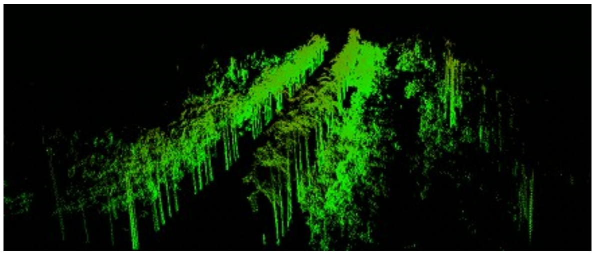


Fig. 9. 3D Mapping: The Palm Tree Boulevard in NTU

To accommodate RANSAC to multiple models - a static environment model for ego-motion estimation and a moving object model for moving object detection, a compact representation models moving object information implicitly is proposed. The experimental results show that accurate identification of static environments can help classification of moving objects, whereas discrimination of moving objects also yields better ego-motion estimation, particularly in environments containing a significant percentage of moving objects. It is feasible to build globally consistent maps of highly dynamic environments using the proposed RANSAC-based ego-motion estimator with scan matching techniques [9], [10].

For accomplishing robot localization, Monte Carlo localization [11] is implemented and run in both NTU-PAL1 and NTU-PAL2. We have successfully demonstrated Monte Carlo localization with our RANSAC-based ego-motion estimator in several demonstrations in which moving entities do not degrade the performance of localization.

### B. Interacting Object Tracking

The SLAM with DATMO framework assumes that the robot and moving objects move independently of each other to reduce the complexity of SLAMMOT enormously. This independence assumption may be unrealistic in human inhabited environments such as crowded urban areas, shopping malls and railway stations. These environments contain a large number of constraints which affect the motions of moving objects. Targets interact both with other moving objects and their surrounding environments. Interactions among moving objects and stationary objects should be of interest for higher level scene understanding.

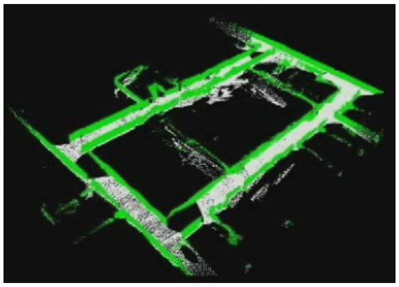


Fig. 10. 3D Mapping: The 4th floor of the CSIE building.

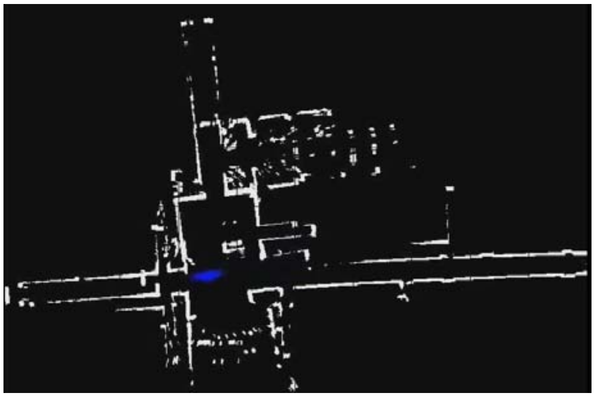


Fig. 11. 2D Mapping: MSRL, ITRI, Taiwan.

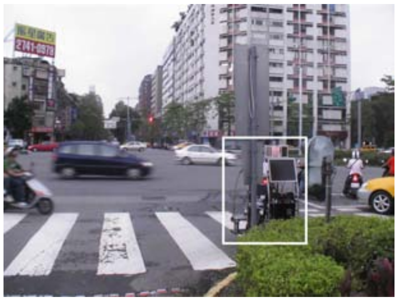


Fig. 12. Interacting Object Tracking at Crowded Traffic Intersections.

Accompanying with traditional motion modeling techniques, we introduced a scene interaction model and a neighboring object interaction model to respectively take long-term and short-term interactions between the tracked objects and its surroundings into account [12], [13]. With the use of the interaction models, anomalous activity recognition is accomplished easily. In addition, move-stop hypothesis tracking is applied to deal with move-stop-move maneuvers. All these approaches are seamlessly intergraded under the variable-structure multiple-model estimation framework [14]. The proposed approaches have been demonstrated using data from a laser scanner mounted on the NTU-PAL1 robot at a crowded intersection near the NTU campus. Interacting pedestrians, bicycles, motorcycles, cars and trucks are successfully tracked in difficult situations with occlusion. Figure 12 shows the test site and Figure 13 shows the experimental result in which a number of moving objects were successfully detected and tracked. See [12] for more information.

As indoor environments are relatively unconstrained than urban areas, interactions in indoor environments are weaker

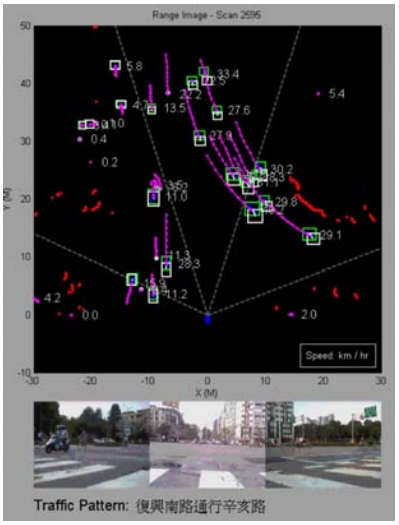


Fig. 13. An example of our interacting object tracking algorithm using one laser scanner.

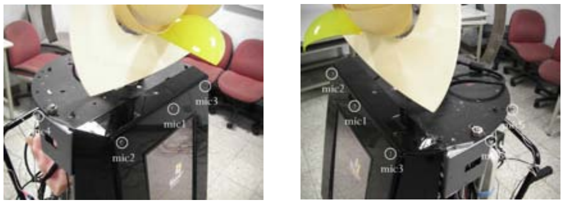
and have more variants. Weak interactions make scene interaction modeling and neighboring object interaction modeling challenging. We proposed a place-driven scene interaction model is proposed to represent long-term interactions in indoor environments. To deal with complicated short-term interactions, the neighboring object interaction model consists of three short-term interaction models, following, approaching and avoidance. The moving model, the stationary process model and these two interaction models are integrated to accomplish weakly interacting object tracking. See [13] for more information.

### C. Auditory Perception

Auditory perception is one of the most important functions for socially assistive robots. Microphone arrays are widely used for auditory perception in which the spatial structure of microphones is usually known. The structure from sound (SFS) approach addresses the problem of simultaneously localizing a set of microphones and a set of acoustic events which provides a great flexibility to calibrate different setups of microphone arrays. However, the existing method does not take measurement uncertainty into account and does not provide uncertainty estimates of the SFS results. In [15], we proposed a probabilistic structure from sound (PSFS) approach using the unscented transform. In addition, a probabilistic sound source localization (PSSL) approach using the PSFS results is provided to improve sound source localization accuracy. The ample results of simulation and experiments using low cost, off-the-shelf microphones mounted on the NTU-PAL2 robot demonstrate the feasibility and performance of the proposed PSFS and PSSL approaches. Based on this foundation, sound event recognition would be the next step to pursue.

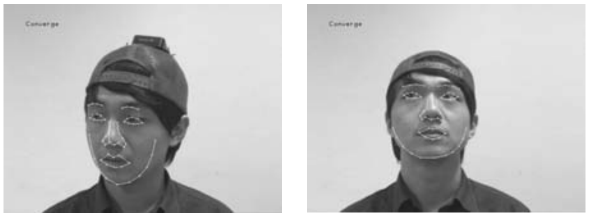
### D. Summary

Although we have demonstrated that localization, mapping, moving object tracking, and sound source localization can be reliably accomplished in both indoor and outdoor environments, the current progress of scene understanding may be insufficient for socially assistive robots. Higher level scene understanding such as activity and interaction recognition would be critical.



(a) The microphone locations. (b) The microphone locations.

Fig. 14. Microphone Array on NTU-PAL2.



(a) Yaw (b) Pitch

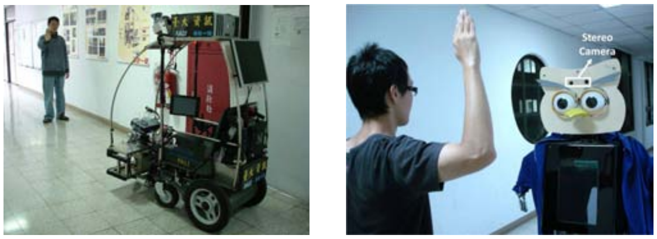
Fig. 15. 3D face alignment in 2D images.

## IV. Human Understanding

Human understanding is essential to socially assistive robotics. In this section, our algorithm to align 3D faces in 2D images and our hand posture recognition system are introduced. The results are demonstrated using 2D images from onboard cameras.

### A. 3D Face Alignment in 2D Images

Perceiving human faces is one of the most important functions for human robot interaction. The active appearance model (AAM) is a statistical approach that models the shape and texture of a target object. According to a number of the existing works, AAM has a great success in modeling human faces. Unfortunately, the traditional AAM framework could fail when the face pose changes as only 2D information is used to model a 3D object. To overcome this limitation, we proposed a 3D AAM framework in which a 3D shape model and an appearance model are used to model human faces [16]. Instead of choosing a proper weighting constant to balance the contributions from appearance similarity and the constraint on consistent 2D shape with 3D shape in the existing work, our approach directly matches 2D visual faces with the 3D shape model. No balancing weighting between 2D shape and 3D shape is needed. In addition, only frontal faces are needed for training and non-frontal faces can be aligned successfully. The experimental results with 20 subjects demonstrate the effectiveness of the proposed approach. Figure 15 shows two alignment results.



(a) A PTZ camera and a web camera are used for recognizing hand postures.

(b) A Stereo camera is used for recognizing hand postures.

Fig. 16. Hand Posture Recognition for Human Robot Interaction.

### B. Hand Posture Recognition

Hand posture understanding is essential to human robot interaction. The existing hand detection approaches using a Viola-Jones detector have two fundamental issues, the degraded performance due to background noise in training images and the in-plane rotation variant detection. In [17], we proposed a hand posture recognition system using the discrete Adaboost learning algorithm with Lowe's scale invariant feature transform (SIFT) features to tackle these issues simultaneously. In addition, we apply a sharing feature concept to increase the accuracy of multi-class hand posture recognition. The experimental results demonstrate that the proposed approach successfully recognizes three hand posture classes and can deal with the background noise issues. Our detector is in-plane rotation invariant, and achieves satisfactory multi-view hand detection. Figure 16 shows the camera systems on NTU-PAL1 and NTU-PAL2 used for hand posture recognition.

### C. Summary

The 3D AAM framework for aligning 3D face in 2D images provides a strong foundation for human expression recognition and emotion recognition. It would be critical to speed up the current 3D AAM algorithm and to increase face pose estimation accuracy in terms of large yaw and pitch motions. Regarding hand posture recognition, we have implemented hidden conditional random fields to further improve the performance [18]. While improving the performance of the camera-based systems, 3D flash ladar devices or 3D cameras are being explored to accomplish these tasks.

## V. Conclusion and Future Work

In this paper, we introduced the robotic systems, NTU-PAL1 and NTU-PAL2. In particular, the camera-projector system on NTU-PAL1 and the expressive face and arms on NTU-PAL2 were described. Our current progress on robot perception was summarized. Several topics and directions for future work were pointed out. Socially assistive robots for the elderly need higher level scene and human understanding. It is also critical to pursue user study in the near future to understand the needs of the elderly.

## Acknowledgment

The laboratory alumni, Tai-Liang Chen, Pei-Han Lee, Yueh-Chi Yu, De-Zheng Liu and Kao-Wei Wan, had made a great effort to construct NTU-PAL1 and NTU-PAL2. We thank for their valuable contributions. This work was partially supported by grants from Taiwan NSC (#96-2628-E-002-251-MY3, #96-2218-E-002-035, #97-2218-E-002-017, #98-2623-E-002-017-D); Taiwan CCI, MSI and Intel.

## References

[1] Y.-C. Yu, "Environment and human behavior learning for robot motion control," Master's thesis, National Taiwan University, 2008.

[2] P.-H. Lee, "3d modeling using camera-projector systems under ambient lighting condition," Master's thesis, National Taiwan University, 2007.

[3] C. Breazeal, "Toward sociable robots," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 167–175, March 2003.

[4] N. Kaya and H. H. Epps, "Relationship between color and emotion: A study of college students." *College Student Journal*, vol. 38, no. 3, pp. 396–418, September 2004.

[5] C. Breazeal, "Function meets style: insights from emotion theory applied to hri," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 34, no. 2, pp. 187–194, 2004.

[6] C.-C. Wang, C. Thorpe, S. Thrun, M. Hebert, and H. Durrant-Whyte, "Simultaneous localization, mapping and moving object tracking," *The International Journal of Robotics Research*, vol. 26, no. 9, pp. 889–916, September 2007.

[7] C.-C. Wang, D. Duggins, J. Gowdy, J. Kozar, R. MacLachlan, C. Mertz, A. Suppe, and C. Thorpe, "Navlab slammot datasets," www.csie.ntu.edu.tw/˜bobwang/datasets.html, May 2004, carnegie Mellon University.

[8] S.-W. Yang and C.-C. Wang, "Multiple-model ransac for ego-motion estimation in highly dynamic environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, May 2009.

[9] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 12, no. 2, pp. 239–256, 1992.

[10] F. Lu and E. Milios, "Robot pose estimation in unknown environments by matching 2D range scans," *Journal of Intelligent and Robotic Systems*, vol. 18, no. 3, pp. 249–275, March 1997.

[11] D. Fox, W. Burgard, F. Dellaert, and S. Thrun, "Monte carlo localization: Efficient position estimation for mobile robots," in *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI'99).*, July 1999.

[12] C.-C. Wang, T.-C. Lo, and S.-W. Yang, "Interacting object tracking in crowded urban areas," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Roma, Italy, April 2007.

[13] K.-W. Wan, C.-C. Wang, and T. T. Ton, "Weakly interacting object tracking in indoor environments," in *IEEE International Conference on Advanced Robotics and its Social Impacts (ARSO)*, Taipei, Taiwan, August 2008.

[14] X.-R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," *IEEE Transactions on Automatic Control*, vol. 41, no. 4, pp. 478–493, April 1996.

[15] C.-H. Lin and C.-C. Wang, "Probabilistic structure from sound and probabilistic sound source localization," in *IEEE International Conference on Advanced Robotics and its Social Impacts (ARSO)*, Taipei, Taiwan, August 2008.

[16] C.-W. Chen and C.-C. Wang, "3d active appearance model for aligning faces in 2d images," in *IEEE/RSJ International Conference on Robots and Systems (IROS)*, Nice, France, September 2008.

[17] C.-C. Wang and K.-C. Wang, *Recent Progress in Robotics: Viable Robotic Service to Human*. Springer Berlin / Heidelberg, 2008, vol. 370, ch. Hand Posture Recognition Using Adaboost with SIFT for Human Robot Interaction, pp. 317–329.

[18] T.-C. Liu, K.-C. Wang, A. Tsai, and C.-C. Wang, "Hand posture recognition using hidden conditional random fields," in *The IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Singapore, July 2009.