Multimodal Approach to Human-Face Detection and Tracking

Prahlad Vadakkepat, *Senior Member, IEEE*, Peter Lim,
Liyanage C. De Silva, Liu Jing, and Li Li Ling

***Abstract*—The constructive need for robots to coexist with humans requires human–machine interaction. It is a challenge to operate these robots in such dynamic environments, which requires continuous decision-making and environment-attribute update in real-time. An autonomous robot guide is well suitable in places such as museums, libraries, schools, hospital, etc. This paper addresses a scenario where a robot tracks and follows a human. A neural network is utilized to learn the skin and nonskin colors. The skin-color probability map is utilized for skin classification and morphology-based preprocessing. Heuristic rule is used for face-ratio analysis and Bayesian cost analysis for label classification. A face-detection module, based on a 2-D color model in the YC_rC_b and YUV color space, is selected over the traditional skin-color model in a 3-D color space. A modified Continuously Adaptive Mean Shift tracking mechanism in a 1-D Hue, Saturation, and Value color space is developed and implemented onto the mobile robot. In addition to the visual cues, the tracking process considers 16 sonar scan and tactile sensor readings from the robot to generate a robust measure of the person's distance from the robot. The robot thus decides an appropriate action, namely, to follow the human subject and perform obstacle avoidance. The proposed approach is orientation invariant under varying lighting conditions and invariant to natural transformations such as translation, rotation, and scaling. Such a multimodal solution is effective for face detection and tracking.**

***Index Terms*—Continuously adaptive mean shift (CAMSHIFT) tracking mechanism, face tracking, facial skin-color model, multimodal approach.**

I. Introduction

MOBILE ROBOTS are increasingly being integrated into our daily lives. In order to provide the basic navigational ability to the robot and to study the coarse structure of the environment, visual, sonar, ultrasonic, infrared, and other range sensors are required. The robots have to acquire information about the environment through various sensors. Nevertheless, the dynamic nature of the environment and the need to interact with the users have set requirements that are more challenging in robot perception. In this paper, the research focus is put on the interaction capability of the mobile robot, particularly in detecting, tracking, and following human subjects.

There are several face-detection techniques which are broadly classified as motion-based, neural-network-based, model-based, and skin-color-segmentation-based techniques. In the motion-based face-detection methods, motion is used to characterize the targets in relation to a static environment. The current captured image is compared to a reference image to segment out the moving targets [1], [2]. However, the methods assume that the targets are nonstationary in a video scene and could not deal with the stationary-object detection. Neural-network-based detection [3]–[6] requires learning from both positive (target) and negative (nontarget) examples. The training of neural networks for face detection is mostly done offline and research is ongoing to improve the real-time performance of systems employing neural networks in face detection. Model-based approaches include the use of models, with a number of parameters describing the shape or motion of the face or a combination of both to aid in estimation [7], [8].

Target-tracking techniques include Kalman filter [9]–[11] and fuzzy control [12], [13]. Kalman filter assumes a dynamic model of the target and that the noise affecting the system is zero mean. However, when the target moves randomly, it is difficult to obtain the dynamic model of the target. The fuzzy-logic approach requires a fuzzy-logic knowledge base to control the specified system. Informally, a knowledge base is a set of representations of facts about the environment. However, the fuzzy-logic approaches are based on the known environment and not robust to variations in environment. Lately, research has adopted the mean-shift algorithm [14]–[16], which is used for the analysis of probability distributions, into the field of real-time visual tracking. Yang and Waibel [17] use a model to estimate image motion and to predict the search window; integrated with a camera model to predict and compensate for camera motion.

In order to achieve near real-time performance and simplicity in design and implementation, the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm [18], [19] is adopted in this paper. The CAMSHIFT algorithm is computationally efficient and can deal with the challenges of irregular object motion due to variations in perspective and image noise, such as facial countenance and other skin-colored objects in the scene. CAMSHIFT is designed as a computer interface for controlling commercial computer games. The motion of the moving camera is not considered in CAMSHIFT.

The multimodal [20]–[22] implementation can extend beyond the use of a mouse, keyboard, and camera to include

Manuscript received September 19, 2006; revised June 7, 2007.

P. Vadakkepat, P. Lim, and L. Jing are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576 (e-mail: elepv@nus.edu.sg; elelj@nus.edu.sg).

L. C. De Silva is with the Institute of Information Sciences and Technology, Massey University, Palmerston North, 11-222, New Zealand (e-mail: L.desilva@massey.ac.nz).

L. L. Ling is with the TECH Semiconductor Singapore Pte. Ltd., Singapore 738799.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2007.903993

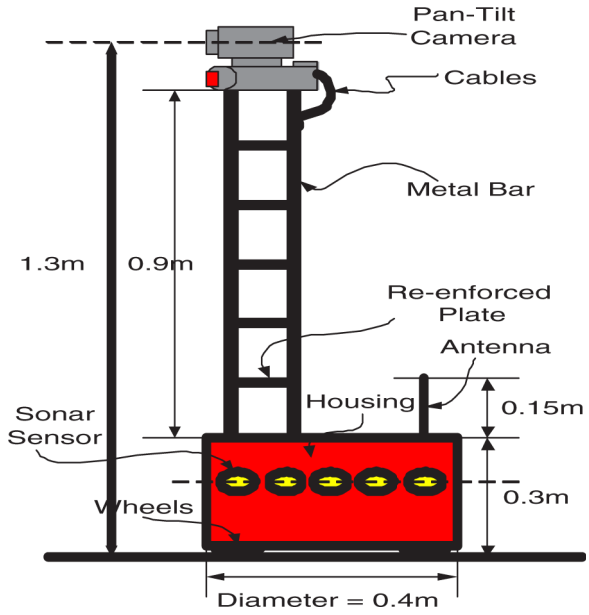


Fig. 1. Magellan Pro mobile robot equipped with 16 ultrasonic and 16 tactile sensors with 360° coverage.

other sensory modalities such as tactile and auditory. In this paper, tracking and robot obstacle avoidance are done with the information from the sonar sensors, tactile sensors, and the conventional visual cues.

To design a robust system, the tracking algorithm is not activated until a face is detected in the video scene. The conventional approach is to gather images of faces and segregate the skin and nonskin areas. A skin-color model is incorporated in the detection module for face detection. Neural networks are used to build the skin-color models in the YC_rC_b and the YUV color spaces.

This work discusses the challenges and approaches with respect to human-face detection and tracking. Two important issues, "what to detect and how to track," are explored. Section II addresses the issue of "what to detect" and presents the skin-color model in the YC_rC_b color space and the YUV color space for human-face characterization. The color models can detect both stationary and moving human faces with different skin colors under varying lighting conditions. Section II also discusses the development of the color-tracking algorithm to address the issue of "how to track." The detection module and the visual-tracking module are integrated into a mobile robot, which has a Sony EVI-D30 pan–tilt camera and 16 sonar and tactile sensors. Section III presents the performance of the proposed system in the UV and YC_rC_b color spaces.

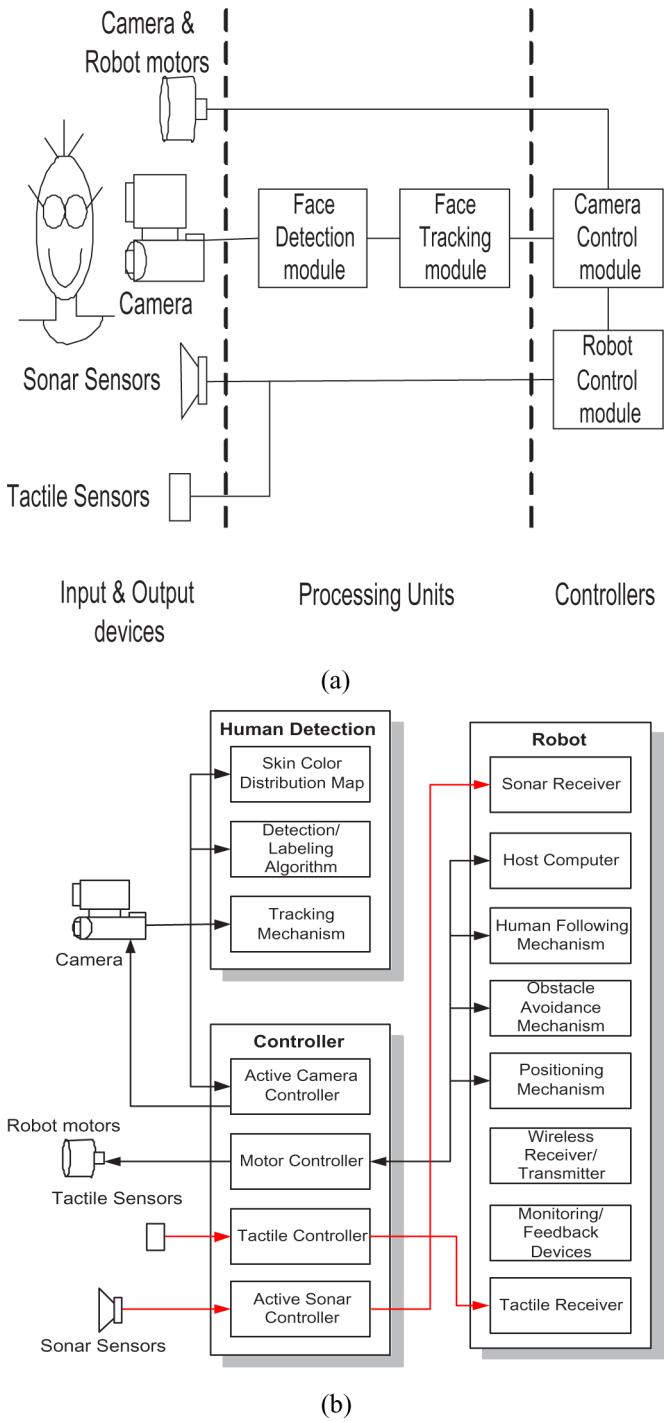


Fig. 2. System overview. (a) Robot-camera control system. (b) System block diagram of the various modules.

II. System Overview

The proposed system is implemented on a Magellan Pro mobile robot (Fig. 1). The robot is equipped with 16 ultrasonic and 16 tactile sensors with 360° coverage. It has an onboard Pentium II computer with Red Hat 6.2 Linux Operating System and operates on a 24-V rechargeable-battery supply. A wireless Ethernet port is used to control the robot from a remote computer terminal, as required. A Sony EVI-D30 pan–tilt camera is installed onto the robot for visual sensing and is controlled via the serial RS232 communication port.

The robot control system is designed to concurrently control both the camera and the robot while tracking a human subject. The human tracking and obstacle avoidance are done with the information drawn from the image captured by the camera, the sonar-scan data, tactile-sensor data, and the camera readings. Fig. 2(a) and (b) shows the block diagrams of the overall system and the associated modules.

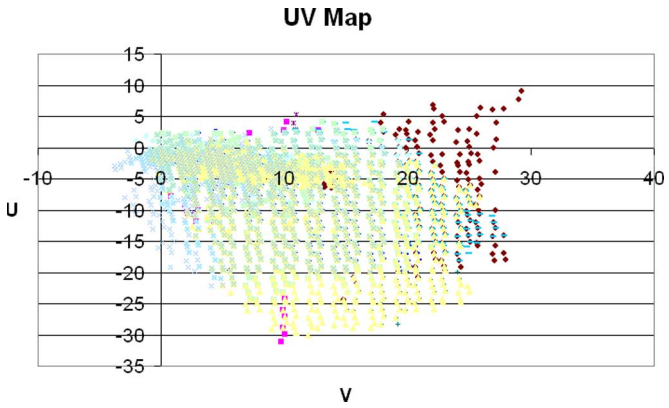


Fig. 3. UV 2-D color map.

The UV color space is used for face detection. The Hue, Saturation, and Value (HSV) color space is utilized in the tracking module. A modified CAMSHIFT algorithm controls the camera motion.

A. Skin-Color Model

Skin color covers only a small part of the UV plane and the influence of the Y channel is minimum [23]. Most digital images use the RGB color space; but individual R, G, and B components vary widely under changing illumination conditions. It is observed that varying skin colors lay in their color intensities rather than in the facial skin color itself [23]. The luminance part of the YUV color space is discarded as it does not provide useful information for human-face detection. Color is orientation-invariant under varying lighting conditions and under natural transformations, such as translation, rotation, and scaling. Therefore, by discarding the luminance signal Y, a more manageable UV space is obtained.

Skin-color detection has some disadvantages too. The influence of luminosity is not totally negligible. The varying skin colors are determined by the color intensities or different shades of a color [24]. This situation worsens as fluorescent lamps (indoor lighting) "flicker" resulting in non-steady luminance in the testing environment. Other objects in the environment with skin-like color also affect the detection.

The skin-color subspace covers only a small area of the UV plane. Relaxing the skin-color model leads to false detections while a rigorous model increases the number of dismissals. A tradeoff between the generic and the accurate models is required. Fig. 3 illustrates the region on the UV color map where facial skin-color pixels lie. This map is used to localize image pixels that contain facial skin color. As a result, the image pixels that contain facial skin color are segmented to form the corresponding facial skin-color regions, which are possible human faces.

It is conjectured that pixels in an image belonging to a face exhibit similar U and V values providing good coverage of all human skins [23]. One hundred face images are collected, comprised of 50 Chinese, 30 Indians, and 20 Caucasians for training and testing purposes. The face-detection neural network learning is done offline in MATLAB.

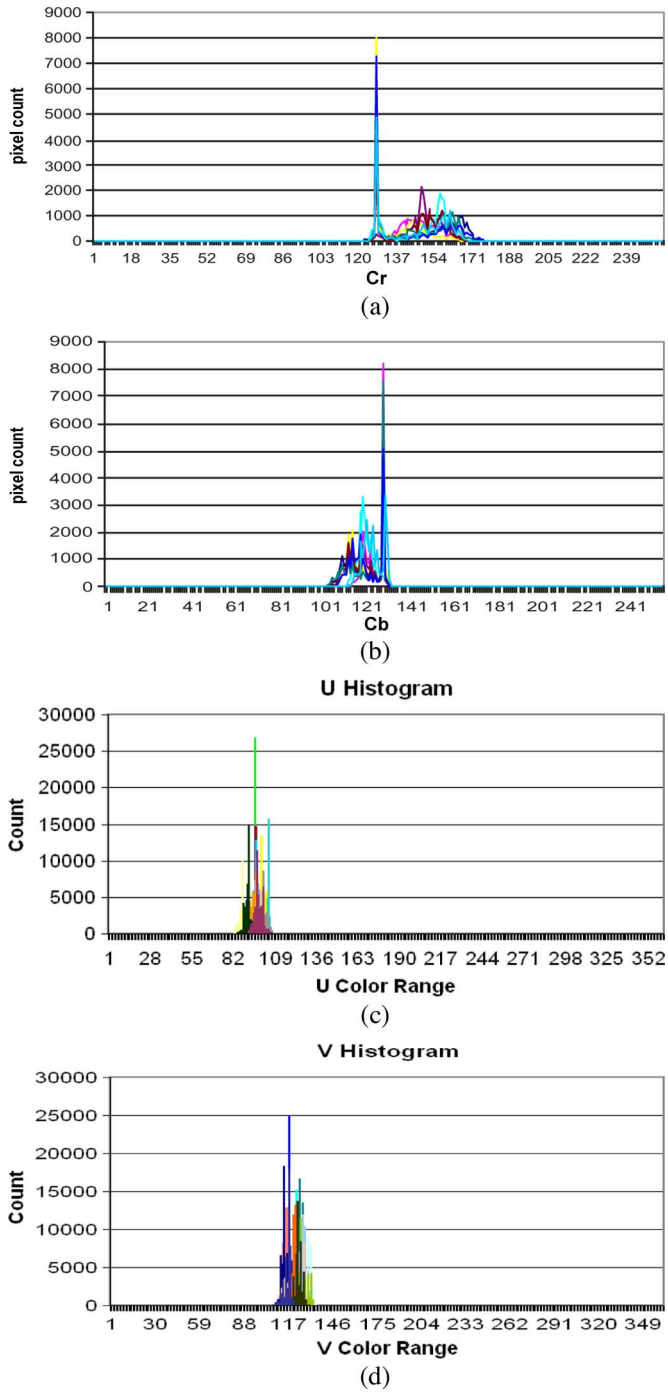


Fig. 4. Histograms of the C_rC_b values and UV values. (a) Histogram of the C_r values. (b) Histogram of the C_b values. (c) Histogram of U values under varying lighting conditions. (d) Histogram of V values under varying lighting conditions.

Histograms of the C_rC_b values [Fig. 4(a) and (b)] and the UV values [Fig. 4(c) and (d)] are built from the above 100 human subjects.

In the C_rC_b plane, from the above histograms [Fig. 4(a) and (b)], a skin-color distribution is generated [Fig. 5(a)]. By relaxing the skin-color model, as shown in Fig. 5(a), the Chrominance C_r of the skin-color-distribution map is observed to lie within a range of 138–178, resulting in (1). By using a least square estimation, two sets of linear equations are obtained,

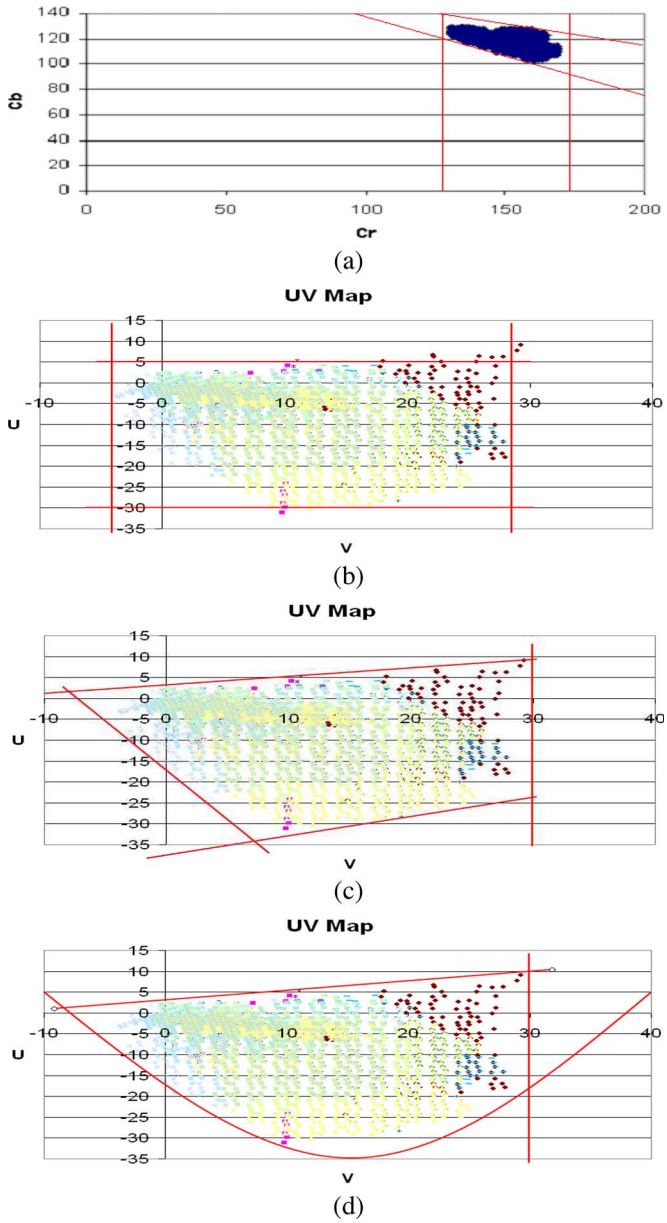


Fig. 5. Thresholding methods in the color spaces. (a) Cluster in the C_rC_b color space. (b) Simple thresholding in the UV color space. (c) Linear thresholding in the UV color space. (d) Nonlinear thresholding in the UV color space.

which are combined as in (2). The skin color in the C_rC_b color-distribution map is bounded by the planes represented by

$$138 < C_r < 178 \tag{1}$$

$$200 < C_b + 0.6C_r < 215. \tag{2}$$

In the UV plane, using the color-space thresholding approach [Fig. 5(b)], (3) and (4) are obtained. Equations (3) and (4) are simple but not effective to classify the skin and nonskin pixels. The linear color thresholding approach [(5)–(8)] is used [Fig. 5(c)] to effectively classify the skin and nonskin pixels. To improve the classification accuracy, a nonlinear thresholding approach is utilized. By relaxing the skin-color model in [Fig. 5(d)] and with the least square estimation approach, two sets of linear equations (9) and (10) and a quadratic equation

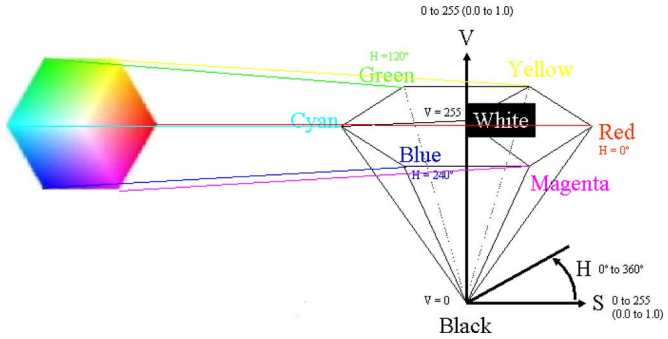


Fig. 6. HSV hexcone.

(11) are obtained. The skin-color pixels in the UV color-distribution map are bounded by these planes

$$-30 < U < 5 \tag{3}$$

$$-4.2 < V < 28.8 \tag{4}$$

$$V < 30 \tag{5}$$

$$U > 0.45\,V - 37.65 \tag{6}$$

$$U > -2.37\,V - 17.65 \tag{7}$$

$$U < 0.206\,V + 2.94 \tag{8}$$

$$V < 30 \tag{9}$$

$$U < 0.206\,V + 2.94 \tag{10}$$

$$U > 0.08\,V^2 - 2.4V - 17.2. \tag{11}$$

B. Face-Detection Operations

The quality of the source image is enhanced by removing noise via the Gaussian pyramid decomposition. All the pixels in the image are analyzed using the planes and the decision rules [Fig. 5(d)], which determine the presence or absence of a skin color. Those pixels that are classified under the skin color are set to white and the rest to black. The resultant binary image contains a few contours and spurious white pixels. To remove the noisy pixels and compact the contours, median filtering is used, followed by the morphological operations of erosion and dilation.

A face is assumed to be the dominant feature in an image. After applying the median filtering and morphological operations, the detected facial region is subjected to a simple heuristic rule. The heuristic rule is based on the geometric analysis of a human face, calculating the ratio of the major axis to the minor axis of a human face. The face ratio, for 100 training sets, is observed to lie within a range of 1.1–1.8.

The skin-color feature is fused with a geometry feature to track the human subject, which makes the tracking process more robust. The skin-color- and geometry-feature-based tracking processes are, respectively, assigned with weights of 0.7 and 0.3 in the fusion process. The fusion process puts larger weight on the color-based tracking process, which is robust to the pose of the subject and occlusion. The position of the face contour is indicated on the original source image where the face is detected.

C. Tracking Mechanism

Perceptual interfaces provide the computer the ability to sense. The perceptual user interface is able to track in real-time

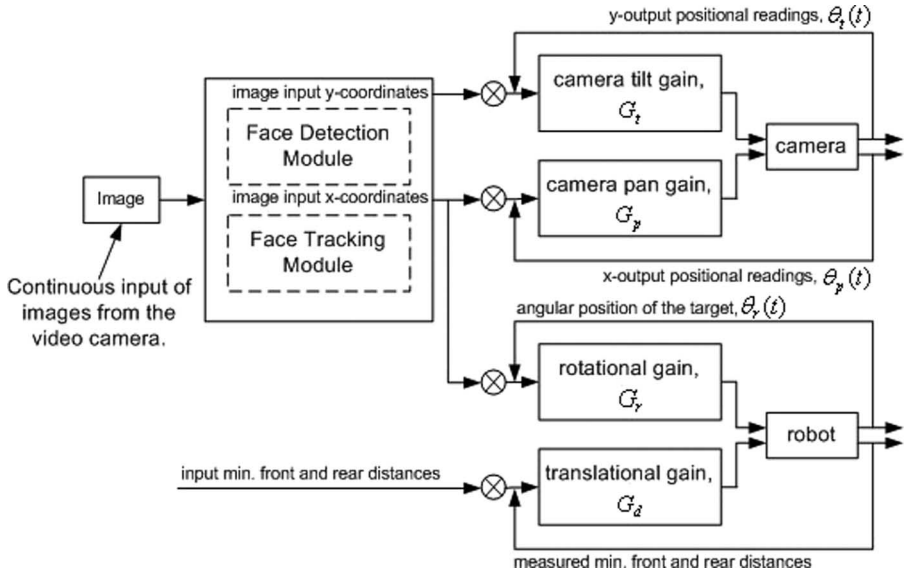


Fig. 7. Block diagram of the robot and camera control.

without taking too much computational resources. The tracking mechanism in this paper is developed based on the CAMSHIFT algorithm [18]. CAMSHIFT is based on a nonparametric technique for finding the mode of the probability distribution called the mean-shift algorithm [14], [18]. CAMSHIFT uses robust statistics, which tends to ignore outliers in a given data.

The HSV color space (Fig. 6) is used to model skin color in CAMSHIFT. This model describes the UV plane in polar coordinates using the magnitude of the saturation vector S and the rotational color angle (hue) H. The luminance V corresponds to the Y of the YUV space. It is observed that human skins have the same hue, given a 1-D color map for human skin color. A model of the desired hue is created using a color histogram. The hues derived from the skin-color pixels in the image are sampled from the H channel and channeled into a 1-D histogram. CAMSHIFT uses the HSV model to compute the probability that a pixel in the image is part of a face, and it replaces each pixel in the video frame with the probability computed.

However, Hue is ambiguously defined when the Saturation $\mathbf{S}$ or Value $\mathbf{V}$ are at their either extremes, causing tracking to be inaccurate under very bright lighting or very dim lighting conditions, as illustrated in Fig. 6. When brightness is low, $\mathbf{V}$ is near to zero, and when Saturation is low, $\mathbf{S}$ is near to zero. Hue is noisy in a small hexcone, and the small number of discrete Hue pixels are not able to adequately represent slight changes in RGB, which leads to wild swings in Hue. Integrating the tracking and the face-detection modules helps to mitigate this problem as two different color spaces are used in these two modules.

The face-detection module keeps the face in the field of view of the camera. When a face is detected and centered, the tracking algorithm is activated. If the pixel values in the region satisfy (9)–(11) and the decision rules, then the tracking continues. Otherwise, tracking stops and the face-detection module continues to search for a face.

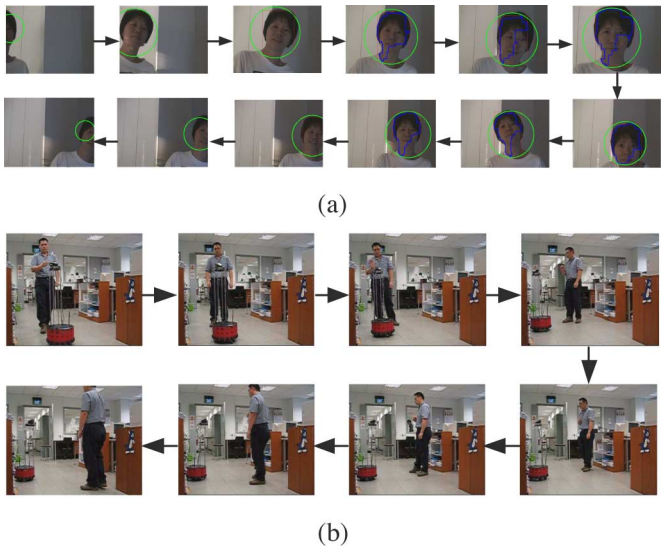


Fig. 8. Experimental results of tracking sequences. (a) Face-tracking sequence. (b) Human-following sequence.

D. Camera and Robot Control

Metal bars and supporting plates are utilized to increase the robot height from 0.3 to 1.3 m (Fig. 1) to track human subjects with a height of 1.7 m. It is assumed that the human subject's height is at most 1.7 m or at least 0.9 m. When a human subject is at a distance of 0.5 m in front of the robot, the camera tilts to keep the face of the human subject within the camera view.

A local map is used to record the sensory information provided by the 16 sonar sensors with respect to the mobile robot. This local map consists of 32 sectors, and each sector covers an angle of 11.25°. This local map provides the mobile robot with the information of the obstacle locations and distances from its current location.

The block diagram of the robot- and camera-control algorithm is shown in Fig. 7. Fig. 8(a) illustrates the face-tracking

Fig. 9. Human subjects.

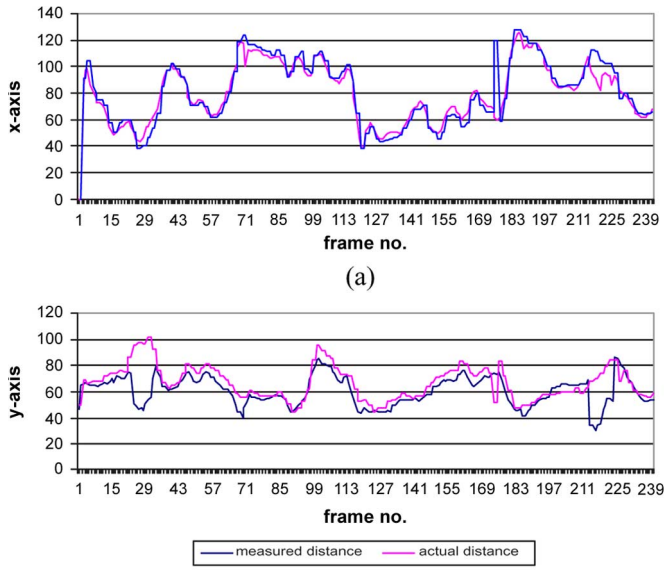


Fig. 10. Comparison of tracking accuracies (CAMSHIFT only). (a) Comparison of X tracking accuracy (CAMSHIFT only). (b) Comparison of Y tracking accuracy (CAMSHIFT only).

sequence, and Fig. 8(b) shows the robot human-following sequence.

III. Tracking Experiments

Experiments are carried out to track six human subjects (R1–R6 in Fig. 9). Each tracking run tracks one human subject and is of 40 s in duration. There are 240 sample points taken in one tracking run to check the tracking accuracy of the integrated system. The tracking runs are compared in both UV and YC_rC_b color spaces. The results are compared based on an object-distance from the origin of the image in pixels. By examining each image frame and identifying the object-center, which is the nose area, the object-distance is determined.

The comparison results are shown in Figs. 10–12. The red and blue lines denote the actual and measured distances, respectively. The x- and y-coordinates of the image center are measured in terms of pixels.

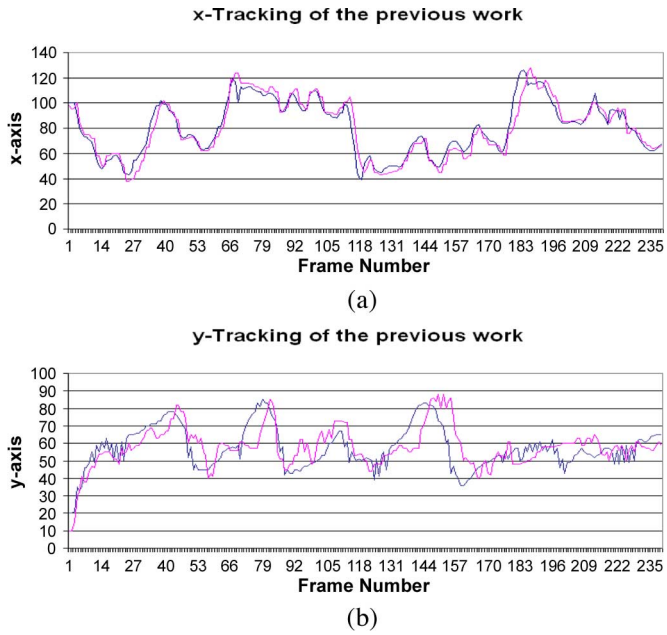


Fig. 11. Comparison of tracking accuracies (YC_rC_b). (a) Comparison of X tracking accuracy (YC_rC_b). (b) Comparison of Y tracking accuracy (YC_rC_b).

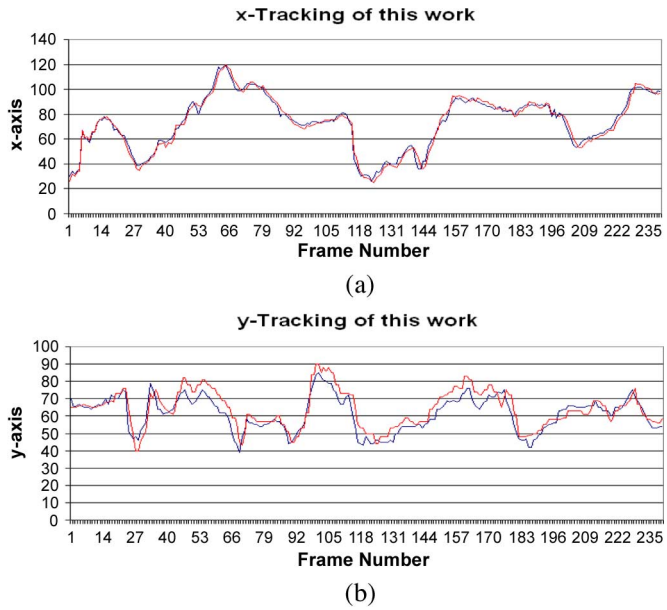


Fig. 12. Comparison of tracking accuracies (YUV). (a) Comparison of X tracking accuracy (YUV). (b) Comparison of Y tracking accuracy (YUV).

A. Performance of the CAMSHIFT Face Tracking

CAMSHIFT tracking algorithm yields reasonable tracking results, except when the camera starts to move, causing significant color shifts in subsequent images (Fig. 10). It is observed that the tracking is not quite accurate, and a significant deviation from the actual distance is observed for frames 177–181 and 217–229. The vertical tracking is generally poorer.

B. Performance of the YC_rC_b Color-Space Face Tracking

The tracking algorithm from the YC_rC_b color-space tracking yields tolerable tracking (Fig. 11). When the camera starts to

TABLE I
TRACKING EXPERIMENTS USING THE YC_rC_b COLOR SPACE

Run	Lost F	Tracked F	OFF F	Success rate (%)
R1	16	213	11	88.8
R2	113	119	8	49.6
R3	0	224	16	93.3
R4	0	234	6	97.5
R5	25	204	11	85.0
R6	11	218	11	90.8

TABLE II
TRACKING EXPERIMENTS USING THE YUV COLOR SPACE

Run	Lost F	Tracked F	OFF F	Success rate (%)
R1	6	228	6	95.0
R2	70	130	40	54.2
R3	0	235	5	97.9
R4	0	237	3	98.8
R5	3	231	6	96.2
R6	0	232	8	96.7

move, it causes significant color shifts in subsequent images and leads to poorer vertical tracking.

C. Performance of the YUV Color-Space Face Tracking

Fig. 12(a) shows that the horizontal tracking is very accurate. The vertical tracking with YUV color space is better compared to the YC_rC_b color-space tracking. The tracking with the YUV color space is significantly better for all the tracking runs.

The vertical aspect of tracking is observed to fluctuate due to the color-shifts in the image when the camera tilts upwards to "face" the laboratory-ceiling lighting. The scene captured by the camera is suddenly illuminated due to strong lighting, leading to lower tracking accuracy.

D. Tracking Performance of the System

The tracking runs are conducted with the same environmental conditions outlined earlier and with different human subjects

$$\text{Success Rate} = \frac{\text{Total } F - (\text{Lost } F + \text{OFF } F)}{\text{Total } F} \tag{12}$$

$$\text{Average Success Rate} = \frac{\text{Success Rate}}{(\text{Total Run})}. \tag{13}$$

The equations of calculating success rate and average success rate are listed in (12) and (13), where F denotes "Frame." In the YC_rC_b color space tracking experiment, the highest success rate is 97.5% (Table I) and the average success rate is 84.2%. In the UV color space tracking experiment, the highest success rate is 98.8% (Table II) and the average success rate is 89.8%.

The system works well with different human subjects with an exception of R2 when the human subject stands up and the camera tilts upward to "face" the laboratory-ceiling lights.

E. Upper Velocity Estimation

The tracking experiments are carried out when the person moves at normal walking speed. When the subject moves faster, tracking fails as the object falls out of the camera's field of view. Although the camera can pan and tilt to center the object, due to the response-time-limit of the camera, it is required that the moving object should be present in two consecutive image frames even if the camera is stationary. The motion of the object along a direction perpendicular to the optical axis of the camera is the fastest way an object can leave out the camera's field of view. To guarantee that the object appears in two consecutive image frames, the object's position difference along the direction perpendicular to the optical axis of the camera during two consecutive frames cannot exceed the width of the camera's field of view [22].

According to the general-optics theory

$$\frac{H_i}{H_o} = \frac{D_i}{D_o} \tag{14}$$

where H_i and H_o, respectively, represent the sizes of the object in the image and the real object. D_i denotes the distance from the image plane to the rear principal plane of the lens. D_o denotes the distance from the object plane to the front principal plane of the lens. When H_i is chosen as the width of the image plane W_i, the corresponding H_o will be the width of the field of view W_{FOV}, given some specified distance D_o

$$W_{\text{FOV}} = \frac{D_o}{D_i} W_i. \tag{15}$$

The upper velocity limit V_{upper} of a moving object that can be tracked is obtained through

$$V_{\text{upper}} = \frac{W_{\text{FOV}}}{\Delta T} = \frac{D_o W_i}{D_i \Delta T} = \frac{0.5 \text{ m} \times 4.8 \text{ mm}}{3 \text{ cm} \times \frac{1}{20} \text{ s}^{-1}} = 1.6 \text{ m/s} \tag{16}$$

where $\Delta T(1/20 \text{ s})$ is the time interval between two consecutive frames after image processing. In this paper, D_o is the distance kept between the robot and the moving object and is chosen approximately as 0.5 m. Since the exact value of D_i is not known, it is approximated to 3 cm. The value of W_i is 4.8 mm. To increase the upper velocity limit of the moving object, improvements in hardware, by increasing the computing power of the onboard computer and by reducing the response time of the pan–tilt camera, can be considered.

IV. CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

This paper ventured into the possibilities of using the UV color space to model the human skin color. Experiments and design requirements, such as applications, effectiveness, and robustness, have been seriously considered in order to operate the mobile robot in a dynamic environment.

The inclusion of other more stringent criteria for face classification should be considered to improve the robustness of the face-extraction algorithm and to reduce the number of dismissals. A morphological-based processing module can be implemented to segment the eye analog [25]. The eye-analog segment is useful to aid in face detection. It is most logical to extend the system by the inclusion of the segmentation of the

eye locations, particularly with the considerable success rates reported in [25].

The facial skin-color model of the present system could be made adaptive by being able to adapt itself automatically to suit new environmental factors, such as changes in ambient light and the use of different cameras. Instead of having to derive the model manually to suit different environmental conditions, the model can have "in-built" tolerances for the variations of these parameters.

REFERENCES

[1] L. Y. Li, W. M. Huang, I. Y. H. Gu, and Q. Tian, "Foreground object detection in changing background based on color cooccurrence statistics," in *Proc. 6th IEEE Workshop Appl. Comput. Vis.*, 2002, pp. 269–274.

[2] A. Utsumi and N. Tetsutani, "Human tracking using multiple-camera-based head appearance modeling," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recog.*, May 2004, pp. 657–662.

[3] S. Rovetta and R. Zunino, "A multiprocessor-oriented visual tracking system," *IEEE Trans. Ind. Electron.*, vol. 46, no. 4, pp. 842–850, Aug. 1999.

[4] F. Yang and M. Paindavoine, "Implementation of an RBF neural network on embedded systems: Real-time face tracking and identity verification," *IEEE Trans. Neural Netw.*, vol. 14, no. 5, pp. 1162–1175, Sep. 2002.

[5] W. M. Huang, B. H. Lee, M. Rajapakse, and L. Y. Li, "Face recognition by incremental learning for robotic interaction," in *Proc. IEEE Workshop Multimedia Signal Process.*, 2002, pp. 280–283.

[6] H. Bae, S. Kim, B.-H. Wang, M. H. Lee, and F. Harashima, "Flame detection for the steam boiler using neural networks and image information in the Ulsan steam power generation plant," *IEEE Trans. Ind. Electron.*, vol. 53, no. 1, pp. 338–348, Feb. 2006.

[7] K.-H. Seo, W. Kim, C. Oh, and J.-J. Lee, "Face detection and facial feature extraction using color snake," in *Proc. IEEE Int. Symp. Ind. Electron.*, 2002, vol. 2, pp. 457–462.

[8] M. S. Alam and A. Bal, "Improved multiple target tracking via global motion compensation and optoelectronic correlation," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 522–529, Feb. 2007.

[9] S. Spors and R. Rabenstein, "A real-time face tracker for color video," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2001, vol. 3, pp. 1493–1496.

[10] B. Kwolek, "Color vision based person following with a mobile robot," in *Proc. 3rd Int. Workshop Robot Motion Control*, 2002, pp. 375–380.

[11] T. Bucher, C. Curio, J. Edelbrunner, C. Igel, D. Kastrup, I. Leefken, G. Lorenz, A. Steinhage, and W. von Seelen, "Image processing and behavior planning for intelligent vehicles," *IEEE Trans. Ind. Electron.*, vol. 50, no. 1, pp. 62–75, Feb. 2003.

[12] H. Y. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 6, pp. 557–563, Jun. 1999.

[13] R. C. Luo and T. M. Chen, "Autonomous mobile target tracking system based on grey-fuzzy control algorithm," *IEEE Trans. Ind. Electron.*, vol. 47, no. 4, pp. 920–931, Aug. 2000.

[14] R. T. Collins, "Mean-shift blob tracking through scale space," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2003, vol. 2, pp. 234–240.

[15] C. Lerdsudwichai and M. A. Mottaleb, "Algorithm for multiple faces tracking," in *Proc. Int. Conf. Multimedia Expo.*, 2003, vol. 2, pp. 777–780.

[16] C. Shen, M. J. Brooks, and A. van den Hengel, "Fast global kernel density mode seeking: Applications to localization and tracking," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1457–1469, May 2007.

[17] J. Yang and A. Waibel, "Tracking human faces in real-time," in "Tech. Rep.," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, 1995.

[18] G. R. Bradski, "Real time face and object tracking as a component of a perceptual user interface," in *Proc. 4th IEEE Workshop Appl. Comput. Vis.*, 1998, pp. 214–219.

[19] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[20] F. Liu, X. Lin, S. Z. Li, and Y. Shi, "Multi-modal face tracking using Bayesian network," in *Proc. IEEE Int. Workshop Anal. Modeling Faces Gestures*, 2003, pp. 135–142.

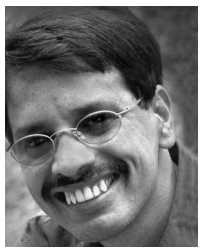
[21] B. Jensen, N. Tomatis, L. Mayor, A. Drygajlo, and R. Siegwart, "Robots meet humans—Interaction in public spaces," *IEEE Trans. Ind. Electron.*, vol. 52, no. 6, pp. 1530–1546, Dec. 2005.

[22] P. Vadakkepat and J. Liu, "Improved particle filter in sensor fusion for tracking randomly moving object," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 5, pp. 823–1832, Oct. 2006.

[23] R. C. K. Hua, L. C. D. Silva, and P. Vadakkepat, "Detection and tracking of faces in real-time environments," in *Proc. Int. Conf. Imaging Sci., Syst. Technol.*, 2002, pp. 24–27.

[24] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 696–706, May 2002.

[25] C. C. Han, H. Y. M. Liao, K. C. Yu, and L. H. Chen, "Fast face detection via morphology-based preprocessing," in *Proc. 9th Int. Conf. Image Acquisition*, 1998, pp. 469–476.



Prahlad Vadakkepat (M'00–SM'05) received the M.Tech. and Ph.D. degrees from the Indian Institute of Technology, Madras, India, in 1989 and 1996, respectively.

From 1991 to 1996, he was a Lecturer with the Regional Engineering College Calicut (currently the National Institute of Technology Calicut), Kerala, India. From 1996 to 1998, he was with Korea Advanced Institute of Science and Technology, Daejeon, Korea, as a Postdoctoral Fellow. He is currently an Assistant Professor with the National University of Singapore, Singapore. His research interests are in humanoid robotics, distributed robotic systems, evolutionary robotics, neuro-fuzzy controllers, and intelligent-control techniques.

Dr. Vadakkepat is the Founder Secretary of the Federation of International Robot-Soccer Association [www.fira.net], where he is currently the FIRA General Secretary. He was appointed as the Technical Activity Coordinator of IEEE Region 10 from 2001 to 2002. He is an Associate Editor of the International Journal of Humanoid Robotics since 2003.

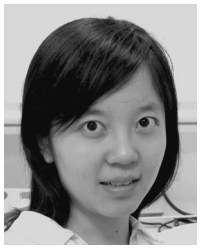
Peter Lim received the B.Tech. and M.Eng. degrees from the National University of Singapore, Singapore, in 2001 and 2005, respectively.

He was with Singapore Technologies Electronics (Training and Simulation Systems) Pte. Ltd., Singapore, as a Senior Engineer. He is currently with the Department of Electrical and Computer Engineering, National University of Singapore. His research interests are computer simulation, human performance science, and computer vision.

Liyanage C. De Silva received the B.Sc.Eng. degree (with honors) from the University of Moratuwa, Moratuwa, Sri Lanka, in 1985, the M.Phil. degree from The Open University of Sri Lanka, Nugegoda, Sri Lanka, in 1989, and the M.Eng. and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 1992 and 1995, respectively.

From 1989 to 1995, he was with the University of Tokyo. From April 1995 to March 1997, he has pursued his postdoctoral research as a Researcher with Advanced Telecommunication Research Laboratories, Kyoto, Japan. In March 1997, he was a Lecturer with the National University of Singapore, Singapore, where he was an Assistant Professor until June 2003. He is currently a Senior Lecturer with the Institute of Information Sciences and Technology, Massey University, Palmerston North, New Zealand. He has expertise in digital-image processing, speech processing, and communications theory. He has published over 100 technical papers in these areas in international conferences, journals, and Japanese national conventions. He is the holder of one Japanese national patent, which was successfully sold to Sony Corporation, Japan, for commercial utilization.

Dr. De Silva was the recipient of the Best Student Paper Award from The International Society for Optical Engineering for an outstanding paper contribution to the International Conference on Visual Communication and Image Processing in 1995.



Liu Jing received the B.Eng. and M.Eng. degrees in automatic control from the Department of Automatic Control, Northwestern Polytechnical University, Xi'an, China, in 1998 and 2000, respectively. She is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore.

Her research interests include multiple-target tracking, maneuvering-target tracking, and data fusion, with particular emphasis on the application of particle filters.

Li Li Ling received the B.Eng. degree from the National University of Singapore, Singapore, in 2003.

She is currently a Senior Systems Engineer with TECH Semiconductor Singapore Pte. Ltd., Singapore, which is a joint venture between Micron Technology, Canon, and Hewlett-Packard.