Visual tracking of a moving person for a home robot

Kai-Tai Song and Chen-Chu Chlen

Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 2005 219: 259
DOI: 10.1243/095965105X9597

Visual tracking of a moving person for a home robot

Kai-Tai Song* and **Chen-Chu Chien**
Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, Republic of China

The manuscript was received on 22 January 2004 and was accepted after revision for publication on 25 February 2005.

DOI: 10.1243/095965105X9597

Abstract: This paper presents a visual tracking system for a home robot to pursue a person. The system works by detecting a human face and tracking a person via controlling a two-degree-of-freedom robot head and the robot body. An image processing system has been developed to extract facial features using a complementary metal-oxide semiconductor (CMOS) web camera. An algorithm is proposed to recognize a human face by using skin colour and elliptical edge information of a human face. A digital signal processing (DSP)-based motor control card is designed and implemented for robot motion control. The visual tracking control system has been integrated on a self-constructed prototype home robot. Experimental results show that the robot tracks a person in real-time.

Keywords: tracking control, human–robot interaction, intelligent robots, image processing

1 INTRODUCTION

In recent years, many researchers have been interested in developing home robots for providing various services and entertainment functions in a home setting. These new types of robot might become popular home appliances in the near future. The main difference between home robots and conventional household appliances is in the way that robots interact with family members. Speech communication and vision recognition are important factors for intelligent human–robot interaction. Charge coupled device (CCD) cameras have been the most utilized sensing devices for autonomous robotic systems. A robotic vision system interacts with a person by detecting, recognizing, and tracking a person [**1–3**]. Visual tracking of a home robot gives people a feeling of awareness and thus an important technology for human–robot interaction. The personal service robot Flo [**1**] is equipped with two colour cameras actuated by two independent servo motors. Flo tracks a person by controlling the head and the cameras. It also recognizes a human face. The mobile robot Yamabico is equipped with a Sony EVI D30 CCD camera [**2**], which is used for human detection and localization. The home robot R100 developed by NEC Corp. is equipped with two colour CCD cameras for detecting, tracking, and recognizing humans. R100 can recognize about ten persons and records observed objects. The robot Colin uses stereo vision for detecting, tracking, and pursuing a human [**3**].

Several approaches to robot visual tracking have been reported. In references [**4**] to [**6**], robotic visual tracking systems were designed to detect a human face using skin colour. In reference [**7**], shape information and facial features (eyes, mouth, nose) are utilized. Using one cue for detecting a human face is normally not reliable enough. Several systems use a combination of multiple cues to find out human faces from the background [**8–10**]. For controlling the robot head to track interested objects, a feedback control system needs to be constructed. Approaches employing fuzzy controllers [**3**, **10**, **11**], as well as conventional proportional-integral-derivative (PID) controllers have been reported. However, a real-time, embedded visual tracking control system for a home robot needs to be investigated. Moreover, to realize a reliable visual tracking system to follow a person, a study on the selection of features to search for a human head deserves special attention.

In this paper, a control scheme is proposed for real-time robot tracking of a human face. A calibration for an image-based pan–tilt robotic head is presented. This paper also presents an efficient visual tracking system for a home robot based on skin

* *Corresponding author: Department of Electrical and Control Engineering, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 300, Taiwan, Republic of China. email: ktsong@cc.nctu.edu.tw*

colour and head contour. The experimental results on real-time tracking of a moving person will be shown. The robot head detects and tracks a human face, while the robot body pursues the person, keeping a predefined distance to the person and directing the robot to the person.

The rest of this paper is organized as follows: in section 2, the system architecture of the robotic face detection and tracking system is presented. Section 3 describes the proposed algorithms for human head detection and localization. The tracking controller design and realization are presented in section 4. Section 5 presents experimental results of the developed system. Section 6 summarizes the contributions of this work.

2 SYSTEM ARCHITECTURE

The developed visual tracking system is implemented on the self-made prototype home robot H2. The home robot H2 is equipped with an on-board industrial PC (IPC), which is connected to the internet via a wireless local area network (LAN) card. On top of the robot body, a two-degree-of-freedom robotic head was installed with a complementary metal-oxide semiconductor (CMOS) camera on top of it. The camera is connected to the IPC via a universal serial bus (USB) port. For building environmental map and collision avoidance, H2 is equipped with 12 ultrasonic sensors. Two stand-alone motor control modules are provided for servo control of robotic head motors and body wheel motors. The motor control module consists of a self-designed motor control card and a servo driver. The motor control card is based on a TI TMS320F240 digital signal processing (DSP) chip, which handles two d.c. servo motors. The DSP chip communicates with the on-board IPC via a serial port to receive motion commands. Figure 1 shows a picture of the home robot hardware prototype.

The software structure of the home robot is illustrated in Fig. 2. The programming language used is C++ under a Windows XP environment. In this structure, the visual tracking system is integrated with the robotic head control system and body navigation system. The complete system consists of five modules. Module 1 is the image preprocessing module. For skin colour segmentation, hue and saturation information of the HSV (hue, saturation, and brightness value) colour space is used. This module transforms a red–green–blue (RGB) colour image into a grey-level image and an HSV colour space image separately. Edge detection and skin colour segmentation are performed using the result

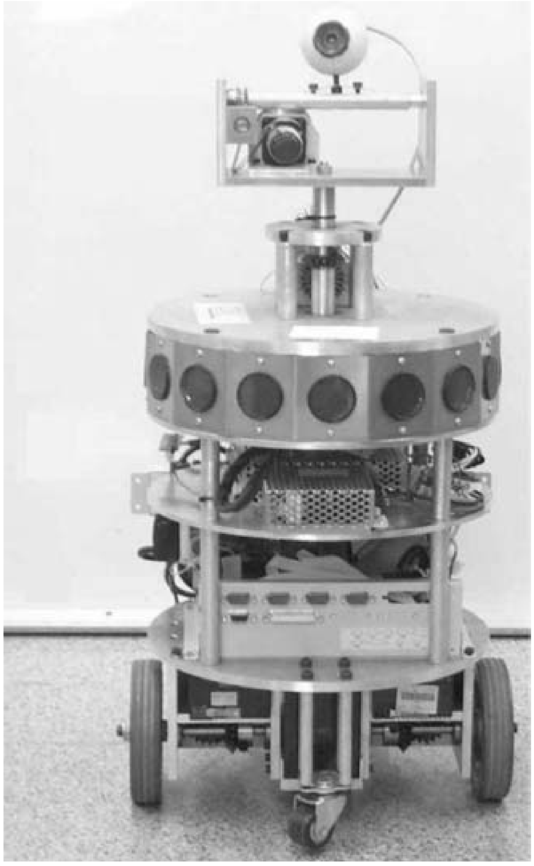

Fig. 1 From view of home robot H2

of image preprocessing. Module 2 performs the facial feature searching and localization. This module is responsible for searching the ellipse contour and skin colour region. It determines the maximal likelihood from the candidates. The result of this module is used to calculate the desired angular position of the pan and tilt motors of the robotic head. Module 3 is the robotic head motion controller. The controller ensures that the pan and tilt motors move to the positions calculated in the previous sample instant. Module 4 is the robot pose estimation and human face measurement module. In this part, the size of the head is extracted and the robot pose is estimated. Module 5 is the robot navigation controller. From the information of the detected head size and robot pose, the controller decides the action of the robot body. When an image is input into the system, module 1 works to obtain the edge and skin colour image. Then the interested human head is searched and localized. If no human head is found in the image, the system then restarts in order to acquire new image input. Otherwise the system calculates the pan angle θ_{pan} and tilt angle θ_{tilt} in order to control the robot head. At the same time, module 4 estimates robot pose and measures the size of the human head via the result

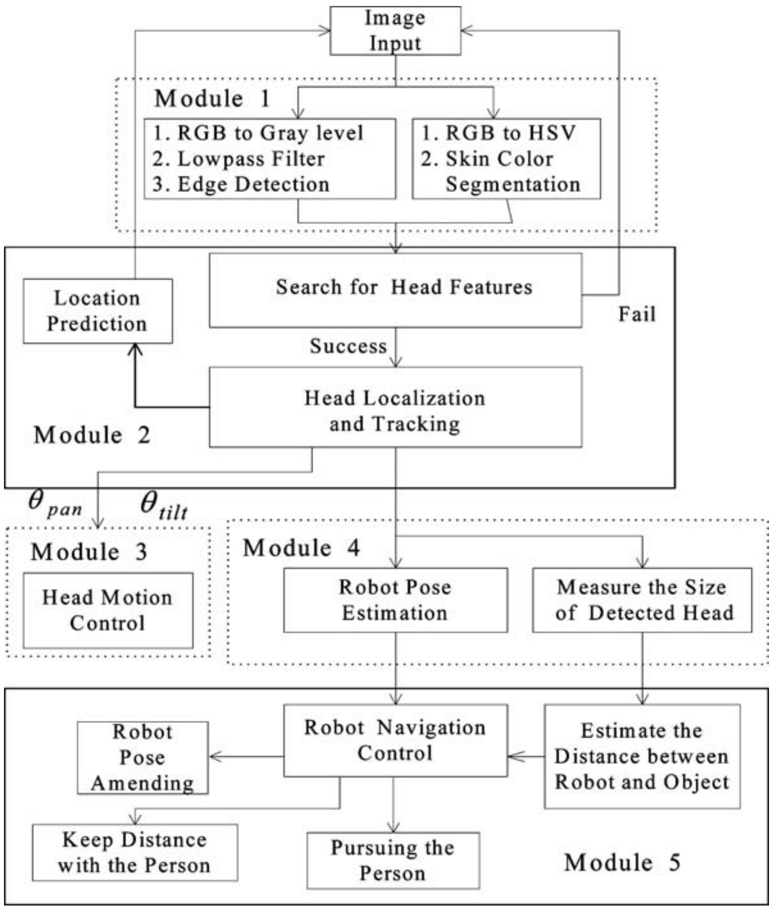

Fig. 2 System architecture of the visual tracking system

of module 3. Finally, module 5 obtains the result of module 4 to control the robot body. Thus the system simutaneously controls the robot head and robot body for visual tracking.

3 HEAD DETECTION AND LOCALIZATION

In the visual tracking system, the human head shape and skin colour information are utilized to search for and localize the human head. An elliptical shape is adopted to model the human head contour [**8**]. Using the ellipse model facilitates fast and effective human head searching. A bound box is employed to detect skin colour inside the head region. For searching and localizing the human head, the system extracts edge and skin colour information of the image. To obtain the edge information, colour space transformation is used to transform RGB colour space images to grey-level ones. A low-pass filter is then applied to reduce noise [**12**]. Finally, edge information is obtained using the Sobel algorithm. To obtain the skin colour information, the RGB colour space is also transformed to the HSV colour space. The system segments the skin colour pixels by using thresholds of hue and saturation. The human head is found and localized subsequently.

3.1 Head contour detection

An elliptic shape was adopted to model the human head contour [**8**, **9**]. The ratio of the long and short axes of the ellipse is selected to be 1.2. At each sample instant, at position (x, y), the average gradient magnitude is calculated around the perimeter of the ellipse. If the average value is higher than a threshold, the ellipse is then a candidate. The short axis of the ellipse is set to a value between 30 and 50 pixels for the system to capture a human face between 0.5 and 1.5 m away from the robot. The gradient magnitude around the perimeter of an ellipse $\phi_g(s)$ is calculated as

$$\phi_g(s) = \frac{1}{N_\sigma} \sum_{i=1}^{N_\sigma} |g_s(i)| \tag{1}$$

where $g_s(i)$ is the binarized intensity gradient value at perimeter pixel i of the ellipse represented by s. N_σ is the number of pixels on the perimeter of the ellipse and $\phi_g(s)$ is the average gradient magnitude around the perimeter of the ellipse. The value of $\phi_g(s)$ is between 0 and 1.

3.2 Skin colour segmentation

Skin colour is a useful feature and one of the most important characteristics of a human face. It is often used to locate and track a human face. However, skin colour is influenced by many factors. One is that different people have different colour appearances. Previous studies show that such a difference can be reduced using normalized RGB or HSV colour space calculations. Another factor is lighting variations. The influence of lighting can also be reduced by using colour space approaches. In this design, a bound box is set up to detect skin colour inside the head region. The height and width of the bound box are set to be double those of the long and short axes of the ellipse. The skin colour distribution factor ϕ_{skin} is obtained using the expression

$$\phi_{\text{skin}} = \frac{1}{N_s}\left[\sum_{i=x-E_x}^{x+E_x}\sum_{j=y-E_y}^{y+E_y} I_{\text{skin}}(i,j)\right] \tag{2}$$

where ϕ_{skin} is the skin colour distribution factor, N_s denotes the number of pixels in the test region represented by $s=(x, y)$, E_x and E_y are, respectively, the long and short axes of the ellipse, and ϕ_{skin} is a value between 0 and 1. If ϕ_{skin} is equal to 0, there is no skin pixel in the bound box. If ϕ_{skin} is equal to 1, all the pixels in the bound box are skin colour pixels. I_{skin} is the segmented skin colour image and is a binary image. In the I_{skin} image, if a pixel in (x, y) is a skin colour pixel, then $I_{\text{skin}}(x, y)$ is equal to 1; otherwise $I_{\text{skin}}(x, y)$ is equal to 0.

At each sample instant, searching for a human head is first performed in the image frame. Then the average intensity gradient along the head boundary ϕ_g and the skin colour distribution of the interior of the head region ϕ_{skin} are calculated. The human head position in the image is the maximum of the sum of ϕ_g and ϕ_{skin}. This procedure is given by

$$s^* = \max_{s_i \in S}\left[\phi_g(s_i) + \phi_{\text{skin}}(s_i)\right] \tag{3}$$

where s^* denotes the obtained head location, S_i denotes the candidate head location, and $\boldsymbol{S}$ is the search space in the image.

The rotation angles of pan and tilt motors are determined by the difference between the estimated head position and the centre position of the image

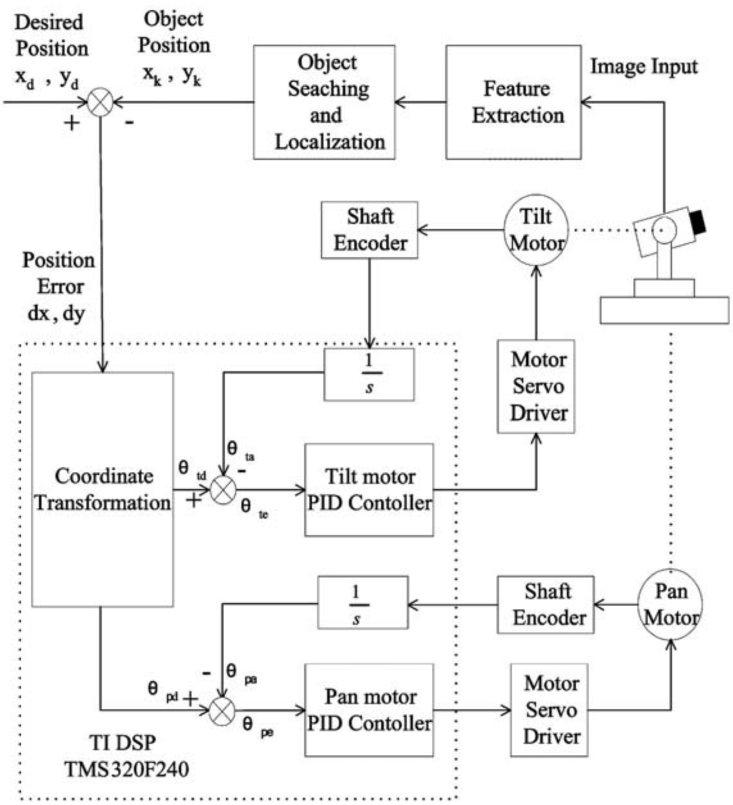

Fig. 3 System architecture of the robotic head controller

frame

$$\theta_{\mathrm{pan}} = \frac{x_{\mathrm{k}} - x_{\mathrm{c}}}{\psi_{\mathrm{pan}}}, \qquad \theta_{\mathrm{tilt}} = \frac{y_{\mathrm{k}} - y_{\mathrm{c}}}{\psi_{\mathrm{tilt}}} \tag{4}$$

where $(x_{\mathrm{c}}, y_{\mathrm{c}})$ is the centre point of the image frame, which is (160, 120), and $(x_{\mathrm{k}}, y_{\mathrm{k}})$ is the head position in the image plane. In this design, ψ_{pan} and ψ_{tilt} are, respectively, the scaling factors of pan and tilt rotations. These two scaling factors are determined using a calibration procedure, which is presented in section 4.1. The angles θ_{pan} and θ_{tilt} are the inputs of the pan and tilt motors of the robot head control system, which rotates the pan and tilt motors to the desired angle via, respectively, the pan and tilt PID controllers. The controllers thus ensure that the camera observes the direction of the tracked human head.

4 VISUAL TRACKING CONTROLLER

The purpose of the visual tracking controller is to ensure that the camera tracks the intended human face by controlling the pan and tilt motors of the robotic head. An image-based control scheme is proposed in this paper to perform the image tracking control. This design uses image features directly. It does not need to compute the relative positions and orientations of the target in Cartesian coordinates. The block diagram of the control system is illustrated in Fig. 3. The system consists of two parts: an image feedback outer loop and servo inner loops.

In the first part, the outer control loop was designed based on image feature feedback. An image is read into the system and the system extracts the head shape and skin colour features to search for

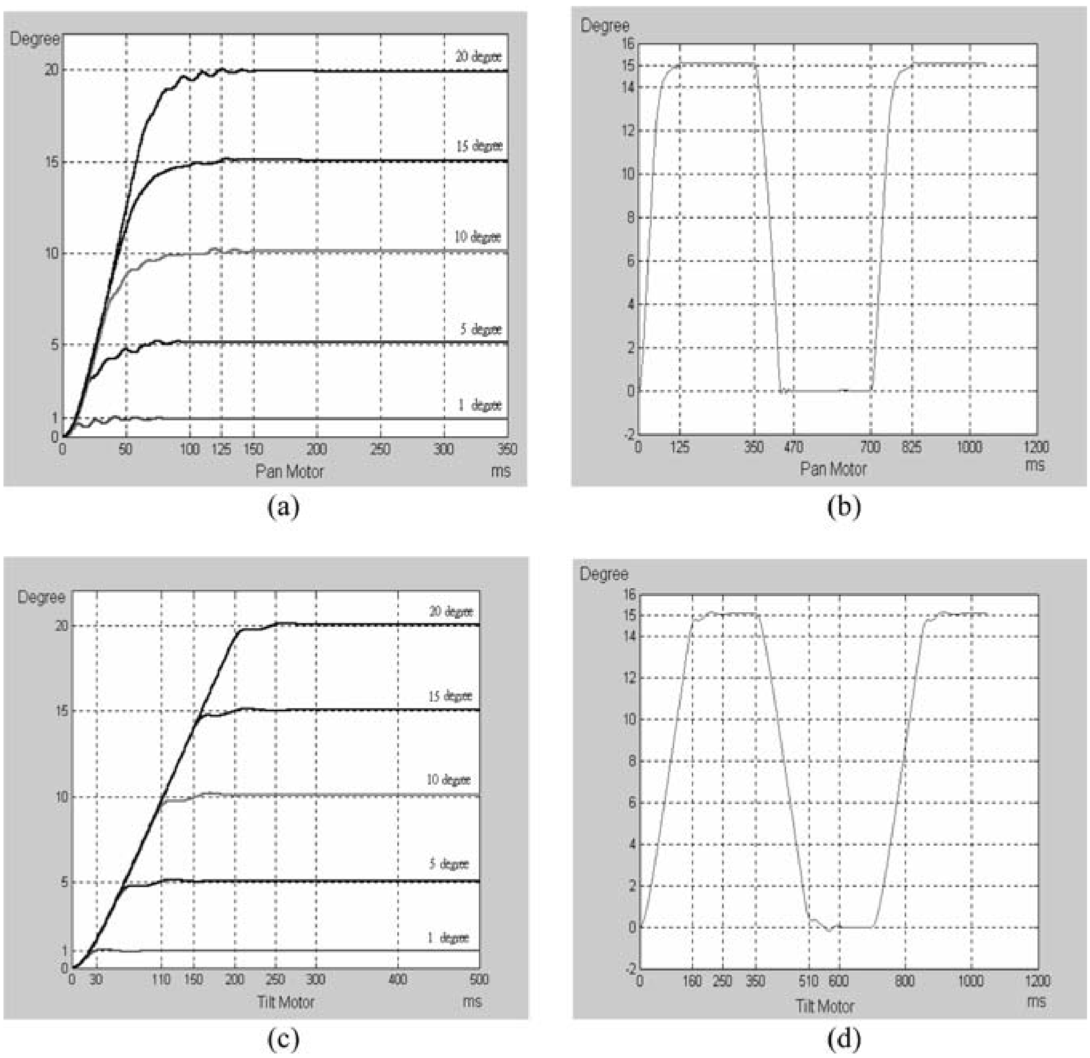

Fig. 4 Step responses of the pan and tilt motors: (a) and (b) are the step responses and the tracking response of the pan motor; (c) and (d) are the step responses and the tracking response of the tilt motor

the location of the human head. In the second part, two servo control loops are designed for the pan and tilt motors. A two-axis DSP-based motion control card is employed to complete the inner control of each axis of the robot head. The system transforms the position error in pixels to the rotation angle in degrees. The lower-level servo controller drives the pan and tilt motors to ensure that the object is in the centre of the image. Figure 4(a) shows the step response of the pan motor controller. The step inputs are 1°, 5°, 10°, 15°, and 20° respectively. From the experimental results, it can be seen that a larger input needs a longer time to reach the steady state. Figure 4(b) shows the tracking performance of the pan motor. The trajectory is a square wave of peak values 15° and -15°. The period is 700 ms. Figures 4(c) and (d) show, respectively, the step responses and the tracking performance of the tilt motor. From the experimental results, it can be seen that for a 10° angular displacement the rise-time of the tilt motor is 105 ms and that of the pan motor is 60 ms. The difference is mainly because the gear ratio of the tilt motor is 300 : 1 and that of the pan motor is 65 : 1.

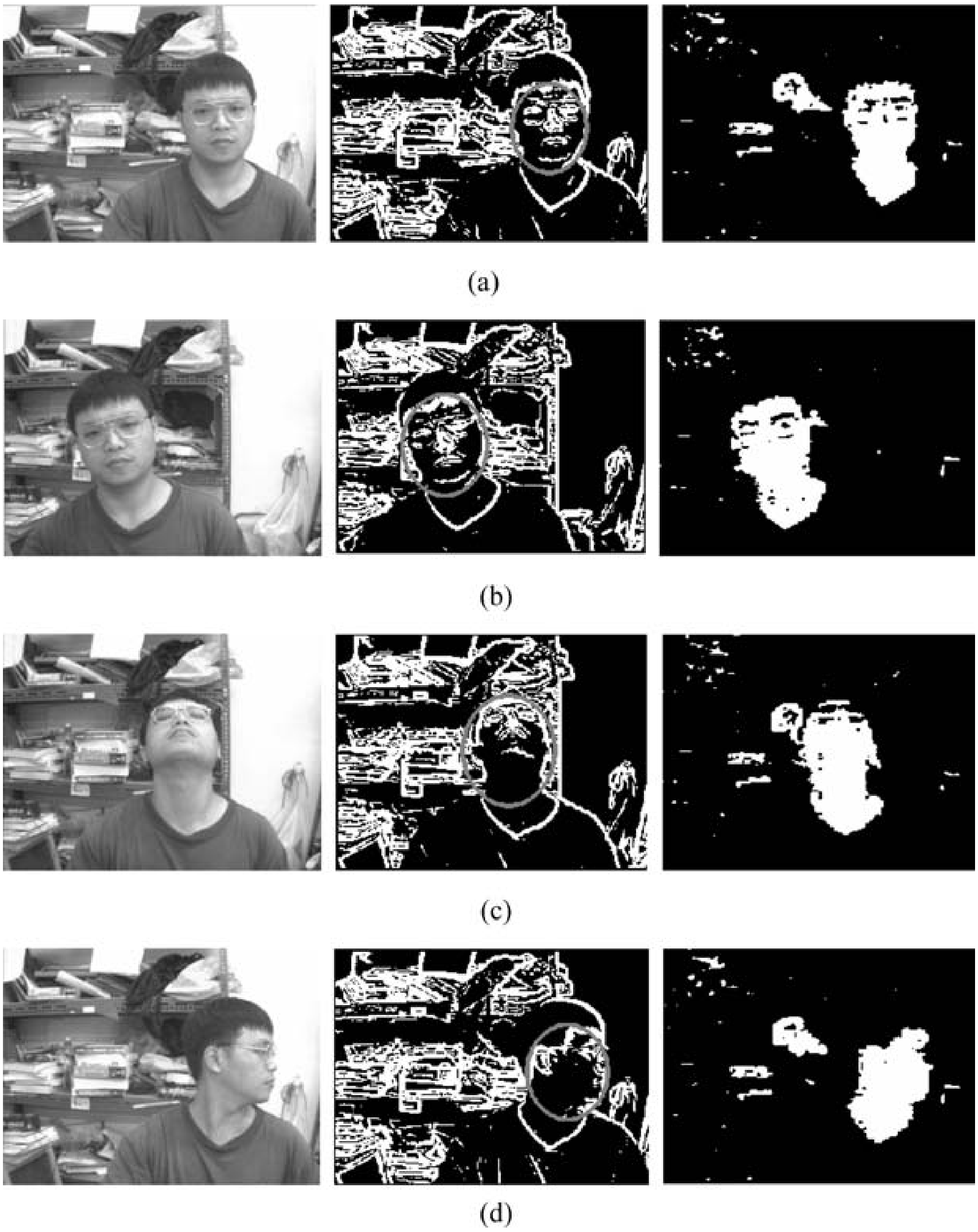

Fig. 5 The original image, edge image, and skin colour image of the static head tracking results: (a) head in front of the camera, (b) head moves from right to left, (c) head tilts up and down, (d) head turns left and right

4.1 Image calibration and coordinate transformation

The purpose of image calibration is to determine the scaling factor between the pixels in the image plane and the motor control angles for controlling the pan and tilt motors. Through a comparison with the original image and a moved image caused by rotation of the motors, the change of pixels can be obtained in the image plane. Given the rotation angular displacement, the scaling factor can be obtained from calibration measurements. The procedure for image calibration is described below.

1. Rotate the pan and tilt motors of the robot head to various angles.
2. Compare the resultant calibration pattern image with the original image.
3. Determine the pattern differences in pixels of the pan and tilt motors.
4. Calculate the scaling factor using

$$\psi_{\text{tilt}} = \frac{\sum \psi_{i-\text{tilt}}}{n}, \qquad \psi_{\text{pan}} = \frac{\sum \psi_{i-\text{pan}}}{n} \tag{5}$$

where n is the number of rotation times in the calibration test, $\psi_{i-\text{tilt}}$ and $\psi_{i-\text{pan}}$ are the scaling factors of pixels and degrees for each rotated angles of the tilt and pan motors respectively, and ψ_{tilt} and ψ_{pan} are the resultant scaling factors of the tilt and pan motors respectively. The calibrated scaling factors are obtained as follows: pan (horizontal) scaling factor $\psi_{\text{pan}} = 5.525$ pixels/degree if the horizontal angle is less than 15°; $\psi_{\text{pan}} = 6.67$ pixels/degree if the horizontal angle is greater or equal to 15°; and tilt (vertical) scaling factor $\psi_{\text{tilt}} = 4.326$ pixels/degree. Practical measurement shows that the average error in the image plane is 7.18 pixels in the horizontal direction and 7.25 pixels in the vertical direction.

4.2 Robot body motion control

For the home robot, the visual tracking system not only tracks a person by controlling the robotic head but also pursues the person by controlling the robot body. There are three navigation modes for the robot body.

1. Pursuing a person. When the robot detects a person in front of it and the distance between the person and the robot is greater than a threshold value, the robot will then move towards the person. The distance is adaptively determined according to the size of the object.

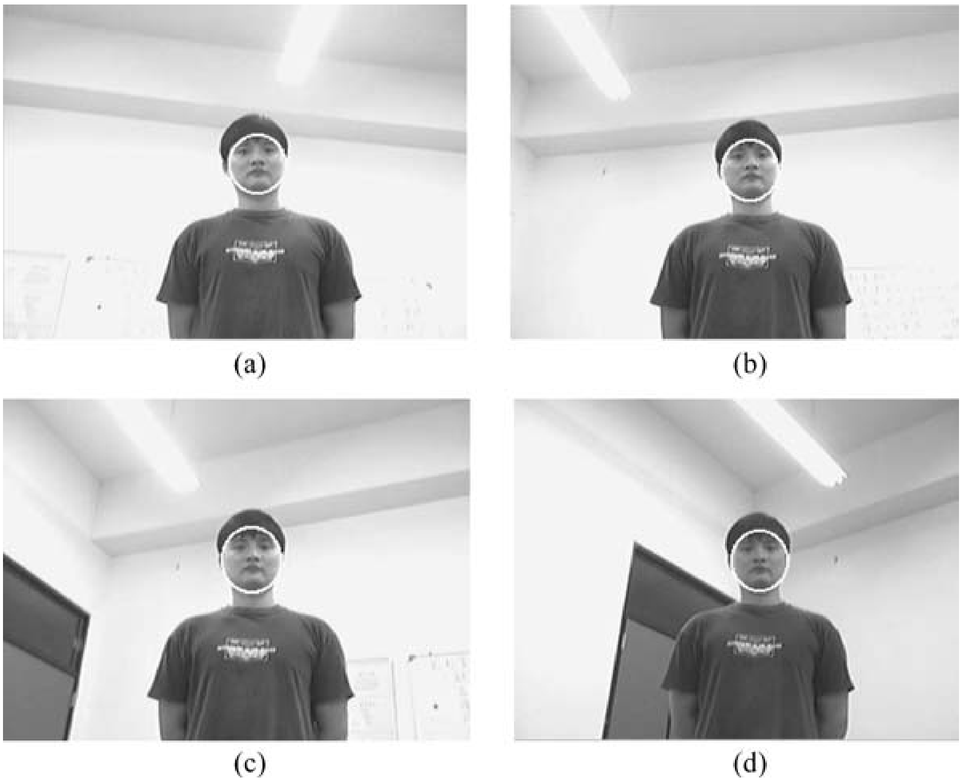

Fig. 6 The robot tracks a person walking horizontally from (a) right to (d) left in front of the robot

2. Keeping a distance with a person. When the distance between the robot and the person is smaller than a threshold value, the robot moves backwards to keep a preassigned distance with the person.
3. Body orientation correction mode. The purpose of this function is to amend the angle between the robotic head and the robot body. It works to align the robot in the direction of the person.

5 EXPERIMENTAL RESULTS

5.1 Face tracking experiments

The system was first tested for several poses without moving the robot body. Figure 5 illustrates the experimental results. In Fig. 5(a), a human face was detected on the right side of the image. The head moved from right to left, as depicted in Fig. 5(b). In the figures, ellipse contours indicate that the head was tracked. Figures 5(c) and (d) show the tracking results as the head tilted up and down, and turned left and right. It is observed that the ellipse symbol always followed the human head.

The purpose of the second experiment is to show that the two-degree-of-freedom robot head can track a person walking in front of the robot. Figure 6 shows that a person walked horizontally from the right side to the left side of the robot and the robot head moved to follow the person. In Fig. 6(a), the person stood on the right side of the robot and the robot head turned right to follow the person. In Figs 6(b) to (d) the person kept moving to the left side and the robot head followed.

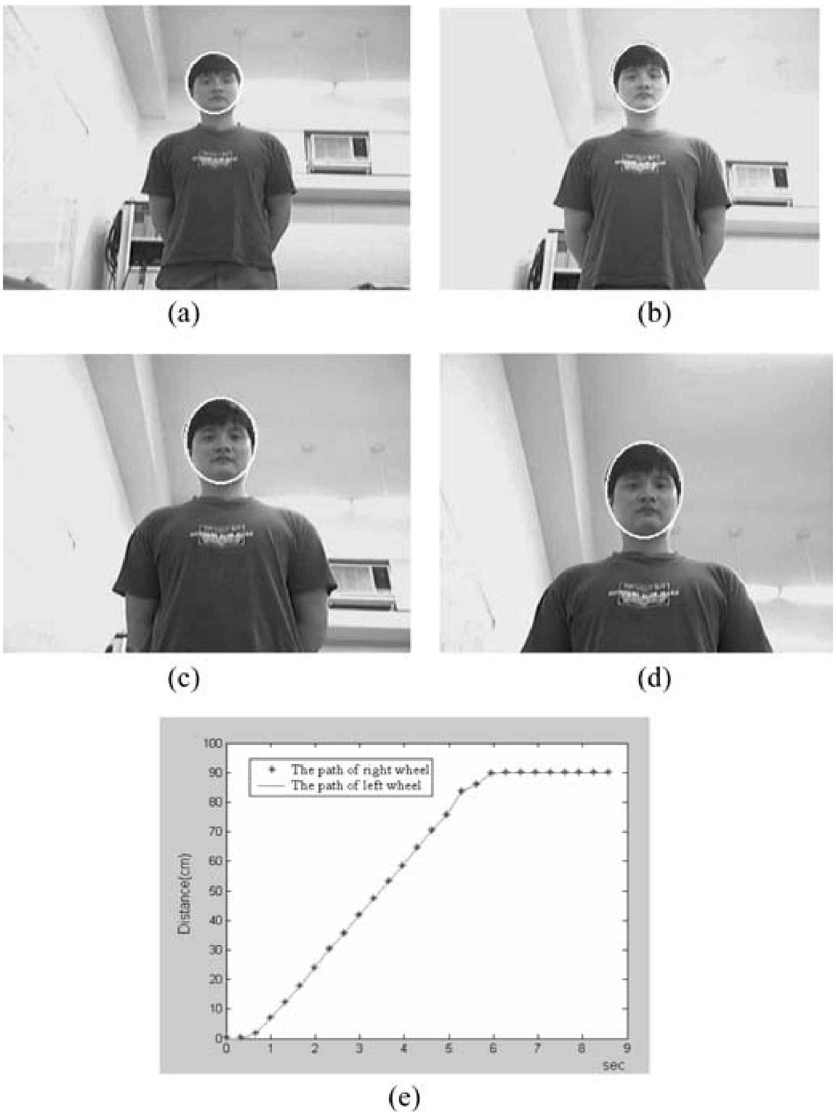

Fig. 7 (a) The robot detects a person in front of it; (b), (c), (d) the robot moves towards the person; (e) recorded trajectories of the right and left wheels

5.2 Robot tracking experiments

In this part, the coordination of the robotic head and robot body is demonstrated. Figure 7 shows that a person stood about 170 cm in front of the robot. In Fig. 7(a), the robot detected the person and moved towards him. In Figs 7(b) to (d), the robotic head tilted the camera to face the person as it approached him. Figure 7(e) shows the recorded trajectories of the right and left wheels. The robot moved forward 90 cm and then stopped because it was close enough to the person.

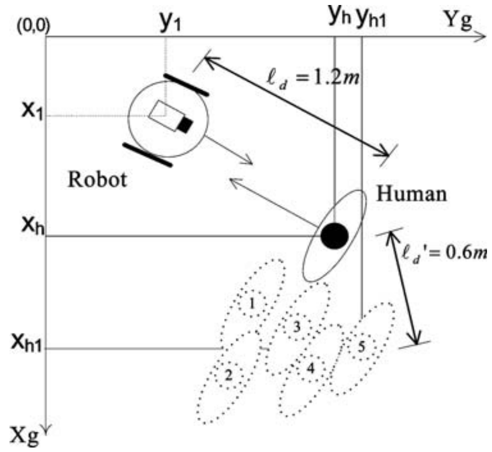

Fig. 8 The motion trajectory of the person in the robot tracking experiment

In the next experiment, a simple human–robot interaction is demonstrated. A person stood in front of the robot and moved forwards and backwards around the robot. The robotic head tracked the human head while the robot body followed the person. Figure 8 illustrates the motion trajectory of the person in the experiment. Figures 9(a) to (e) presents the tracked human image as observed by the robot in positions 1, 2, 3, 4, and 5 of Fig. 8 respectively. Figure 9(f) shows that the robot moved close to the person and tilted the robotic head upwards. Figure 10 presents the recorded trajectories of the wheels. In time intervals a, c, and d of the recorded trajectories, the robot body rotated to face the person. In time intervals b and e, the robot moved forwards to follow the person.

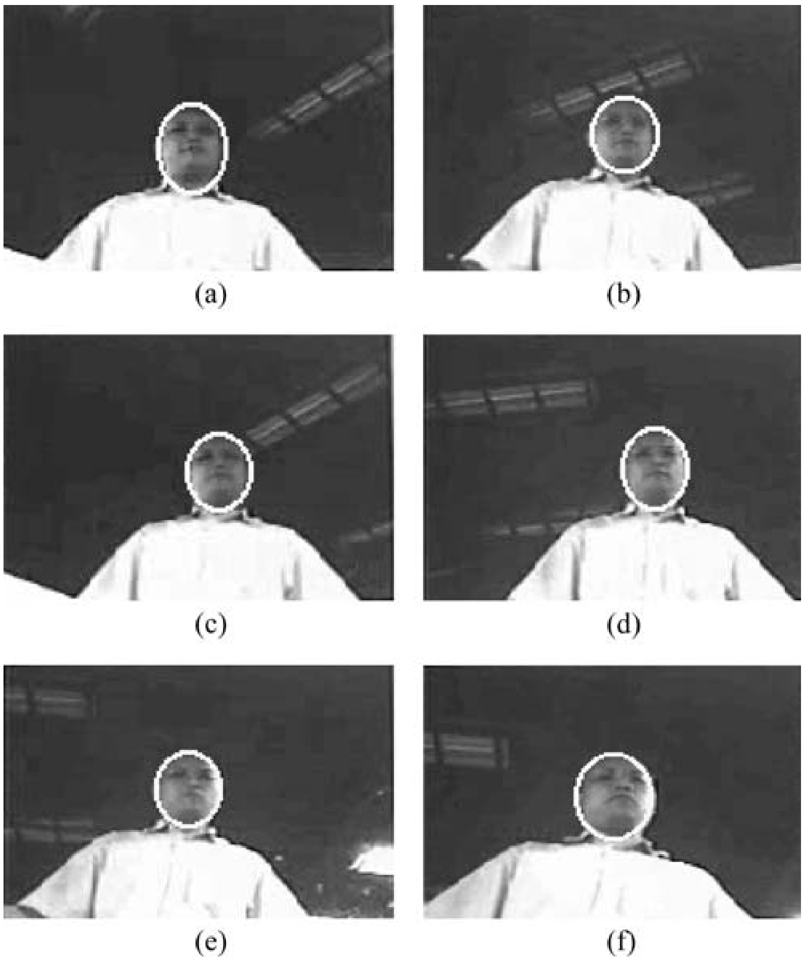

Fig. 9 The robot tracking and pursuing a person

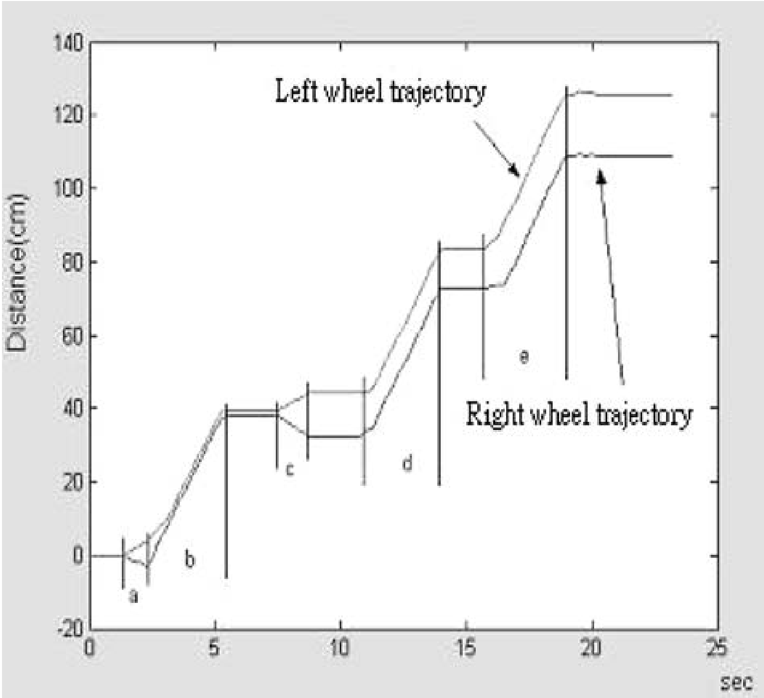

Fig. 10 The recorded trajectories of the experiment in Figs 8 and 9

6 CONCLUSION AND FUTURE WORK

A visual tracking system has been designed and implemented for a home robot to follow a person. Image processing algorithms using skin colour and head shape features have been developed for facial feature searching and localization. A tracking controller is designed using an image-based approach for following a moving person. Experimental results show that the robot tracks a person first using its head motion and then using a combined body and head motion as the person moves continuously. The real-time performance is, however, still not satisfactory. In the future, a dedicated image-processing board using a DSP chip will be added to the robot to improve the real-time tracking performance. On the other hand, a more robust image processing algorithm will be investigated to cope with the variation of lighting conditions during tracking. It will also be interesting to add a face recognition function to the home robot such that it can recognize a family member and interact with him or her accordingly.

ACKNOWLEDGEMENT

This work was supported by the National Science Council of Taiwan, under the grant NSC 91-2213-E009-027. The authors would like to thank Mr Kun-Ming Yan for his assistance in constructing the mobile robot. The suggestions and comments of referees for improving this paper are gratefully acknowledged.

REFERENCES

1 **Baltus, G., Fox, D., Gemperle, F., Goetz, J., Hirsch, T., Margaritis, D., Montemerlo, M., Pineau, J., Roy, N., Schulte, J.,** and **Thrun, S.** Towards personal service robots for the elderly. In Proceedings of the Workshop on *Interactive Robots and Entertainment* (*Wire-2000*), Pittsburgh, USA, 2000.

2 **Ghidary, S., Nakata, Y., Takamori, T.,** and **Hattori, M.** Localization and approaching to the human by mobile home robot. In Proceedings of the 9th IEEE International Workshop on *Robot and Human Interactive Communication*, Osaka, Japan, 2000, pp. 63–68.

3 **Feyrer, S.** and **Zell, A.** Tracking and pursuing persons with a mobile robot. In Proceedings of the International Workshop on *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Corfu, Greece, 1999, pp. 83–88.

4 **Schwerdt, K.** and **Crowley, J. L.** Robust face tracking using color. In Proceedings of the Fourth IEEE International Conference on *Automatic Face and Gesture Recognition*, Grenoble, France, 2000, pp. 90–95.

5 **Wang, Y.** and **Yuan, B.** A novel approach for human face detection from color images under complex background, *Pattern Recognition*, 2001, **40**(10), 1983–1992.
6 **Spors, S.** and **Rabenstein, R.** A real-time face tracker for color video. In Proceedings of the 2001 IEEE International Conference on *Acoustics, Speech, and Signal Processing*, Salt Lake City, USA, 2001, pp. 1493–1496.
7 **Lin, C. H.** and **Fan, K. C.** Triangle-based approach to the detection of human face, *Pattern Recognition*, 2001, **34**(6), 1271–1284.
8 **Birchfield, S.** An elliptical head tracker. In Proceedings of the Thirty-First Asilomar Conference on *Signals, Systems and Computers*, Pacific Grove, USA, 1997, pp. 1710–1714.
9 **Birchfield, S.** Elliptical head tracking using intensity gradients and color histograms. In Proceedings of the IEEE Computer Society Conference on *Computer Vision and Pattern Recognition*, Santa Barbara, USA, 1998, pp. 232–237.
10 **Jordao, L., Perrone, M., Costeira, J. P.,** and **Santos-Victor, J.** Active face and feature tracking. In Proceedings of the International Conference on *Image Analysis and Processing*, Venice, Italy, 1999, pp. 572–576.
11 **Chaouke, B., Tairi, H., Masmoudi, L.,** and **Radouane, L.** PI controller for visual tracking: further results. In Proceedings of OCEANS '99 MTS/IEEE on *Riding the Crest into the 21st Century*, Seattle, USA, 1999, pp. 1480–1485.
12 **Gonzalez, R. C.** and **Woods, R. E.** *Digital image processing*, Second Edition, 2002 (Prentice-Hall, Englewood Cliffs, New Jersey).

APPENDIX

Notation

E_x	long axis of an ellipse
E_y	short axis of an ellipse
g_s	binarized intensity gradient value
l_d	distance in world coordinate system
I_{skin}	binarized skin colour value after segmentation
n	number of rotations
N_s	number of pixels in the head region represented by $s(x, y)$
N_σ	number of pixels along the perimeter of an ellipse
s	perimeter of an ellipse
s^*	head location
x_c, y_c	centre point of the image frame
x_d, y_d	desired target location in the image plane
x_k, y_k	head location in the image plane
θ_{pan}	pan rotation angle
θ_{tilt}	tilt rotation angle
ϕ_g	average intensity gradient
ϕ_{skin}	skin colour distribution in an ellipse region
ψ_{pan}	scaling factor of pan rotation
ψ_{tilt}	scaling factor of tilt rotation