

A STUDY ON AUTONOMOUS VEHICLE GUIDANCE FOR PERSON FOLLOWING BY 2D HUMAN IMAGE ANALYSIS AND 3D COMPUTER VISION TECHNIQUES*

Kuan-Ting Chen (陳冠婷)¹ and Wen-Hsiang Tsai (蔡文祥)^{1, 2}

¹ Institute of Multimedia Engineering, College of Computer Science
National Chiao Tung University, Hsinchu, Taiwan 300

² Department of Computer Science and Information Engineering
Asia University, Taiwan, Taiwan 413

E-mail: whtsai@cis.nctu.edu.tw

ABSTRACT*

An intelligent vision-based vehicle system for person following is proposed. To follow a target person who turns fast at a corner in a narrow path, a technique of using crossroad points and the image information taken by the robot arm to search the disappearing person is proposed. In addition, techniques for intelligent interaction with humans to provide more intelligent services are also proposed, which can be employed to detect the facing direction and the hand movement of a person. Good experimental results show the flexibility and feasibility of the proposed methods for the application of indoor person following.

1. INTRODUCTION

Recently, vision-based autonomous vehicles and mobile robots have been used in more and more human environments to play various roles. For playing the role of a guide by following a person, which we call a *person-following guide* in the sequel, if we use stationary cameras, we have to install many of them and there might be still corners where the cameras cannot cover. But if we use an autonomous vehicle equipped with a video camera, the problem can be solved because the vehicle can follow people to go everywhere like a guide. When a vehicle follows a person, some unexpected situations might still be encountered, such as the case that the person makes a fast turn so that the vehicle cannot see the person or the case that the person walks into a narrow path so that the vehicle is apt to strike the wall and difficult to keep following the person.

For the application of person guidance, the vehicle needs the additional functions of making some introductions to the environment or surrounding objects, or interacting with the person to give some responses to

the person's actions without the use of voices and touches.

To achieve the mission of person following in indoor environments, many methods have been proposed to build a skin color model for human face detection. Wang and Tsai [1] proposed a method which uses an elliptic skin model to detect human faces by color and shape features in images. Ku and Tsai [2] proposed a method that uses sequential pattern recognition to decide the location of the person related to the vehicle and detects the rectangular shape attached on the back of the person to achieve smooth person following. In applications of person following, Kwolek [3] proposed a method that determines the position of a mobile robot by laser readings. Hirai and Miroguchi [4] studied person following to provide people with services by the use of a human collaborative robot which tracks the back and shoulder of a person. Kulyukin, Gharpure, and Nicholson [5] proposed an autonomous vehicle for the blind in grocery stores and conducts navigation by using lasers.

The goal of this study is to design an intelligent system for person following by the use of a vehicle equipped with a video camera. It is desired to devise the system to be capable of human following, as well as interaction with humans. By these functions, more applications of the vehicle can be created, like being used as an autonomous shopping car to help people carrying heavy goods or as a guide in a museum, an exhibition, a tour, and so on.

2. SYSTEM CONFIGURATION

In our system, we use the Pioneer 3, a vehicle made by ActiveMedia Robotics Technologies Inc., as a test bed. The vehicle is equipped with a robotic arm which can reach up to 50 cm. The tip of the arm is used to hold a digital web camera, AXIS210. The camera is IP-based and has a build-in web server of which we can adjust some camera parameters in the graphical user interface of the website, such as image resolution, image format, and so on.

* This work was supported by the Ministry of Economic Affairs under Project No. MOEA 94-EC-17-A-02-S1-032 in Technology Development Program for Academia.



Figure 1. The vehicle Pioneer3 used in this study.

3. HUMAN FOLLOWING TECHNIQUES BY INDOOR AUTONOMOUS VEHICLE

When the vehicle follows a person, if the followed person turns fast at a corner, the system will lose the track of the person. We propose a method to solve this problem in this study. In addition, the vehicle is enabled to adjust its orientation for monitoring the surrounding when it follows the person. Therefore, if the person walks in a narrow path, the vehicle will not hit the wall. In Section 3.1, we will describe the details of the above-mentioned techniques.

3.1. Fast Human Turning at A Corner in A Narrow Path

When the person turns fast at a corner in a narrow path, it is difficult for the vehicle to keep following the person. We propose here a method to solve this problem. The detail is described in the following.

3.1.1. Adaptation to fast turning by use of crossroad points

In a new environment, the vehicle is instructed to learn the *crossroad points* of the environment first. So if the person disappears for a while, the vehicle can use these points to search the person. To choose the correct spots as these crossroad points, the system always has to record the position of the person in the real world. Then once the person disappears, the system can find the crossroad point nearest to the last position of the person. The detail of the proposed method is described in the following as an algorithm.

Algorithm 1: Adaptation to fast turning.

Input: The current image I_c , a reference image I_r , the disappearance time T of the person, a crossroad point set $P_{cross} = \{P_{c1}, P_{c2}, \dots, P_{cn}\}$, the distance between the vehicle and the followed person D_{last} , the position (W_{vx}, W_{vy}) of the vehicle and the position (W_{px}, W_{py}) of the person in the real world.

Output: Command the vehicle to go forward to the correct crossroad point $(Cross_x, Cross_y)$.

Steps:

- Step 1. Detect the person in the reference image I_r .
- Step 2. Measure the distance between the vehicle and the person D_{last} .
- Step 3. Measure the position of the person in the real world by Eqs. (3.1) and (3.2) as follows:

$$W_{px} = W_{vx} + D_{last} \cdot \sin\theta; \quad (3.1)$$

$$W_{py} = W_{vy} + D_{last} \cdot \cos\theta. \quad (3.2)$$

- Step 4. Detect the person in the current image I_c . If the person is detected, set the I_c as I_r . and go to Step 1; else, compute the disappearance time T of the person. If T is larger than 3 seconds, go to Step 5; else repeat Step 4.
- Step 5. Let (P_x, P_y) denote the values of $P_{cross\ ij}$. Find out the point, $(Cross_x, Cross_y)$, from crossroad point set P_{cross} , which is nearest to the point (W_{px}, W_{py}) according to the distance measure Eq.(3.3) below:

$$D = \sqrt{(W_{px} - P_x)^2 + (W_{py} - P_y)^2}. \quad (3.3)$$

- Step 6. Command the vehicle go to the point, $(Cross_x, Cross_y)$.

3.1.2. Recording direction of fast human disappearance by a vehicle arm

If the person walks too fast, the vehicle cannot keep up with the person because the turning speed of the vehicle has a limit. Therefore, we use the robot arm which holds the camera to turn first and record the turning angle of the arm. By this method, when the person walks in a narrow path, the vehicle can be more sensitive to the changes of the person's position and adjust the orientation of the vehicle faster to prevent the vehicle from hitting the wall. The detail of the proposed method is described in the following as an algorithm.

Algorithm 2: Recording the turning angle of the robot arm.

Input: The center of the image $C_{image}(i_{cimage}, j_{cimage})$, the center of the clothes $C_a(i_{ca}, j_{ca})$ and a threshold T_3 .

Output: The turning angle θ of the arm.

Steps:

- Step 1. Detect the center of the clothes, $C_a(i_{ca}, j_{ca})$, by Algorithm 5 described later.
- Step 2. Compare the image center $C_{image}(i_{cimage}, j_{cimage})$ with the clothes center $C_a(i_{ca}, j_{ca})$ by the following inequalities:

$$|i_{cimage} - i_{ca}| < T_3; \quad (3.4)$$

$$|j_{cimage} - j_{ca}| < T_3. \quad (3.5)$$

where T_3 is a pre-selected threshold value. If Inequalities (3.4) and (3.5) are satisfied, we don't have to adjust the orientation of the camera because it means that the person does not move too much; else, go to Step 3

- Step 3. Compute the longitude angle L_θ of $C_a(i_{ca}, j_{ca})$, adjust the arm orientation for the value of L_θ , and set the desired value θ as the value L_θ .

Because the system records the turning angle of the arm all the time, when the person disappears, the

vehicle can turn to the correct direction to search the person. When the person disappears, we use the turning angle of the arm to find out the side on which the disappearing person is. The detail of the proposed method is described as Algorithm 3 below.

Algorithm 3: Turn to the direction of the disappearing person.

Input: Current image I_c , the direction of the vehicle θ_v , the turning angle of the arm θ , and the turning direction *Direction*.

Output: Command the vehicle to turn to the correct direction *Direction* with the angle θ_y for searching the disappearing person.

Steps:

Step 1. Detect the person in the current image I_c . If the person is considered to have disappeared, go to Step 2; else, go to Step 1.

Step 2. Decide the turning direction *Direction* in the following way;

$$\text{if } \theta > 0, \text{ then set } \textit{Direction} = \text{'right.'} \quad (3.6)$$

$$\text{else, if } \theta < 0, \text{ then set } \textit{Direction} = \text{'left.'} \quad (3.7)$$

Step 3. Compute θ_v' as the turning angle of the vehicle for the following cases.

Case 1: if the direction of the vehicle is between -45° and 45° , then set $\theta_v' = 0^\circ - \theta_v$.

Case 2: if the direction of the vehicle is between 45° and 135° , then set $\theta_v' = 90^\circ - \theta_v$.

Case 3: if the direction of the vehicle is between -45° and -135° , then set $\theta_v' = -90^\circ - \theta_v$.

Case 4: if the direction of the vehicle is between 135° and 180° , then set $\theta_v' = 180^\circ - \theta_v$.

Case 5: if the direction of the vehicle is between -180° and -135° , then set $\theta_v' = -180^\circ - \theta_v$.

Step 4. If *Direction* is 'right,' then set $\theta_y = 90^\circ + \theta_v'$; else, set $\theta_y = 90^\circ - \theta_v'$.

Step 5. Command the vehicle to turn to the correct direction *Direction* with the angle θ_y and search the followed person.

4. HUMAN INTERACTION TECHNIQUES FOR INDOOR PERSON FOLLOWING

After the vehicle can follow a person, we want to let the vehicle be of more widespread uses. If a person can communicate with the vehicle, then the vehicle can provide the person with more intelligent services. It is desired to enable the vehicle to figure out the person's thoughts by vision-based analysis of the person's actions. For this goal, we deal with two kinds of human actions in this study. The first is the facing of a person, i.e., looking at the left or the right. By detecting the facing direction of a person, we can know the place where a person faces to and what the person is

interested in. The details will be described in Section 4.1. The second kind is the movement of a person's hand, i.e., hand waving. When the person waves his/her hands, we want to let the vehicle know that the person is calling it, and move closer to the person to do some services. We will introduce a method to achieve this goal in Section 4.2.

4.1. Human Facing Direction Detection

To detect of the person's facing direction, we find that the ratio of the width and height of the person's clothes changes when the person turns to the right or left. When the person turns, the shape of the clothes extracted in the image becomes thinner. We recognize the person's turn by utilizing this property. Next, we use the distribution of the person's hair and skin to judge whether the person turns to the right or left. When the person turns to the right, the hair is at the left side of the face and the skin is at the right side, and vice versa. For this purpose, we find the center of the face and use the colors of the pixels in a horizontal line which passes this point to compute the color distribution of the hair and skin in the person's face.

We apply the proposed human facing direction detection technique to the application of tour guiding. We define the area information in the environment first. For example, if the person is in a museum, then we define a specific region in front of each art in advance. When the vehicle plays a role as a tour guide, it judges whether the person continues to stand at an identical position. If the person does, then we will measure the facing direction and the position of the person. Finally, the vehicle will introduce the information of the object which the person is interested in, as shown in Figure 2. The detail of finding the facing direction of the person is stated in the following as an algorithm.

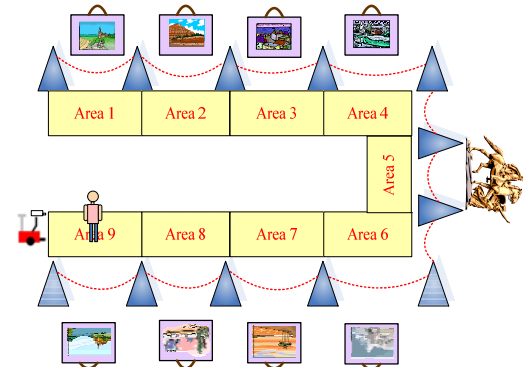


Figure 2. An illustration of the vehicle being used as a tour guide.

Algorithm 4: Finding the facing direction of the person.

Input: The current image I_c , the center $C_a(i_c, j_c)$ and the four corner points $P_{TopLeft}(i_{tl}, j_{tl})$, $P_{TopRight}(i_{tr}, j_{tr})$, $P_{BottomLeft}(i_{bl}, j_{bl})$, and $P_{BottomRight}(i_{br}, j_{br})$ of the clothes region, the length of the person's face

L_{face} , a region of the color of black, *Black*, the skin color region *Skin*, and thresholds T_2 and T_3 .

Output: The facing direction of the person, *Direction*.

Steps:

- Step 1. Scan the column of the image I_c , which contains the pixel $C_a(i_c, j_c)$ to find the first pixel $Hair(i_{hair}, j_{hair})$ with the color of the hair, *Black*.
- Step 2. Find the center of the face $C_{face}(i_{face}, j_{face})$ by the following way.

$$i_{face} = i_c; \quad (4.1)$$

$$j_{face} = j_{hair} + \frac{L_{face}}{2}. \quad (4.2)$$

- Step 3. Because the range of the width of the face will not exceed the clothes, limit the horizontal search region to be from $P_{TopLeft}(i)$ to $P_{TopRight}(i)$.
- Step 4. Measure the values of C_b and C_r of the pixels from the position i_{tl} to i_{face} . If the values falls into the region *Black*, set the number NL_{black} of black color as

$$NL_{Black} = NL_{Black} + 1. \quad (4.3)$$

If the values falls into the region *Skin*, set the number NL_{skin} of skin color as

$$NL_{skin} = NL_{skin} + 1. \quad (4.4)$$

- Step 5. Measure the color values of C_b and C_r of the pixel from the position i_{face} to i_r . If the values falls into the region *Black*, set the number NR_{black} of black color as

$$NR_{Black} = NR_{Black} + 1. \quad (4.5)$$

If the values falls into the region *Skin*, set the number NR_{skin} of skin color as

$$NR_{skin} = NR_{skin} + 1. \quad (4.6)$$

- Step 6. Check the sizes of the distributions of the colors of the skin and the hair as follows:

$$NL_{black} + NR_{black} < T_2; \quad (4.7)$$

$$NL_{skin} + NR_{skin} > T_3; \quad (4.8)$$

$$NL_{black} > NR_{black}; \quad (4.9)$$

$$NL_{skin} < NR_{skin}. \quad (4.10)$$

If Inequalities (4.7) and (4.8) are satisfied, set *Direction* as “Front”; else, if Inequalities (4.9) and (4.10) are satisfied, set *Direction* as “Right”; else, as “Left”.

4.2. Human Hand Movement Detection by Motion Analysis

For detection of the person’s hand movement, we adopt the technique of motion detection by frame

differencing. However, the vehicle might have some small quakes while the person does not move. So we have to check the range of the object’s movement. If the range is small, we consider that the movement is effected by the quakes of the camera. The detail is stated in Algorithm 5. Then we combine the information of the person’s facing direction to judge whether the person is calling the vehicle. When the person faces the vehicle, the vehicle will ‘consider’ the movement of the hand is aiming to call itself. The detail is stated in Algorithm 6.

By applying the human hand movement detection, the vehicle can play a role as a shopping cart which can carry heavy things for the person in a shopping mall, for example. When the vehicle plays this role, it judges whether the person continues to stand at an identical position and waves his/her hand. If the person does, then the system will decide the facing direction of the person. Finally by combining these two kinds of information, the vehicle can know whether the person is calling it, and if so, move close to the person to serve him/her. An illustration is shown in Figure 3.



Figure 3. An experimental result of the vehicle being a cart in the library. Note the person was calling the vehicle at $t = 2$ by waving her hand.

Algorithm 5: Motion detection by frame differencing.

Input: current image I_c , reference image I_r , a threshold t , and the size of a search window w .

Output: a difference image I_d .

Steps:

- Step 1. Subtract the target pixel P_{ij} from the pixel at the same position in the reference image. If the difference is below the threshold t , regard the target pixel P_{ij} as *stationary*. Otherwise, go to Step 2.
- Step 2. Find the best match pixel of the target pixel P_{ij} within the search window w in the reference image by subtracting the target pixel P_{ij} from each of the pixel within the search window w .
- Step 3. If the difference between the target pixel P_{ij} and the best match pixel is below the threshold t , regard the pixel P_{ij} as *stationary*; else, *moving*.
- Step 4. Repeat Step 2 for each pixel in the image to decide the state, *stationary* or *moving*, of it.
- Step 5. Get a complete frame difference image I_d by filling the *moving* pixel with white color and the *stationary* blocks with black color.

Algorithm 6: Vehicle calling by combining the facing direction detection.

Input: A difference image I_d , the facing direction $Direction$ of the person, and the region of the clothes R_{cloth} .

Output: The result $R_{HandWave}$ of detection of hand movement and the result of calling the vehicle $R_{calling}$.

Steps:

- Step 1. Check every pixel p_i in I_d . If p_i is a ‘moving’ pixel in the difference image I_d and is inside the region R_{cloth} of the person’s clothes, regard this pixel to be in a moving point set $M = \{M_{00}, M_{01}, \dots, M_{nn}\}$; else, repeat Step 1 to check the next pixel.
- Step 2. Compute the moving region set $R_{moving} = \{R_{m,00}, R_{m,01}, \dots, R_{m,nn}\}$ by using region growing from the moving point set M with M_{00} as the start point.
- Step 3. Find the biggest region in R_{moving} and denote it as R_{hand} . If the region R_{hand} is larger than threshold T_2 , set $R_{HandWave}$ as “true”; else, as “false.”
If $R_{HandWave}$ is “true” and the facing $Direction$ is “Front”, set $R_{calling}$ as “true”; else, as “false”.

5. EXPERIMENTAL RESULTS

We show some experimental results of the proposed human following system. At first, a user controls the vehicle to learn the position of the crossroad points and the area information of the bookcases in the environment. When the vehicle

follows the person, if the person turns fast at a corner, the vehicle will go forward to a correct crossroad point and turn to the right direction to search the disappearing person, as shown in Figure 4. Besides, if the person turns to the right or left for a while, the vehicle will make an introduction about the object which the person is interested in, as shown in Figure 5. When the person waves his/her hand, the vehicle will move close to the person and wait for a while for the person to put books which the person wants to buy in the basket, as shown in Figure 6.

6. CONCLUSIONS

Several techniques and strategies have been proposed in this study. When more than one person appear in the camera view, the vehicle can follow the person by the feature of the person’s clothes. When the person makes a fast turn at a corner, the system will use the recorded information to command the vehicle to go to the correct crossroad point and turn to the right direction for searching the disappearing person in the acquired image. Therefore, if the person walks in a narrow path, the vehicle will find the person again and will not hit the wall during turning and approaching the person. In addition, human interaction techniques have also been proposed. The first technique is the change of the facing direction of a person. By detecting the facing direction of a person, we can know the place which a person faces and what the person is interested in. The second technique is the movement of a person’s hand. When the person waves his/her hand, the vehicle can know that the person is calling it and move closer to the person. The proposed strategies and methods, as mentioned previously, have been implemented on a vehicle system with a robot arm. Several suggestions and related interesting issues are worth further investigation in the future. For example, conducting human following by different features, such as texture and shape, to eliminate errors caused by the case that different people may wear clothes of the same color or that the color distribution of the clothes is similar to the surroundings. In addition, we may add a judgment of the position of a person’s body. When a person slips and falls over, the person’s body may fall down on the floor. If this situation continues for a long time, the vehicle can issue an emergence signal to call someone for help.

References

- [1] Y. T. Wang and W. H. Tsai “Indoor security patrolling with intruding person detection and following capabilities by vision-based autonomous vehicle navigation,” *Proceedings of 2006 International Computer Symposium (ICS 2006) - International Workshop on Image*

Processing, Computer Graphics, and Multimedia Technologies, Dec. 4-6, Taipei, Taiwan, Republic of China, 2006.

- [2] C. H. Ku and W. H. Tsai, "Smooth Autonomous Land Vehicle Navigation in Indoor Environments by Person Following Using Sequential Pattern Recognition," *Journal of Robotic Systems*, Vol. 16, No. 5, pp.249-262, 1999.
- [3] B. Kwoltek, "Person Following and Mobile Camera Localization Using Particle Filters," *Proc. of 4th International Workshop on Robot Motion and Control*, June 17-20, Poland, pp. 265-270, 2004.
- [4] N. Hirai, and H. Mizoguchi, "Visual Tracking of Human Back and Shoulder for Person Following Robot," *Proc. of 2003 IEEE/ASME International Conf. on Advanced Intelligent Mechatronics*, July 20-24, Kobe, Japan, pp.27-532, 2003.
- [5] V. Kulyukin, C. Gharpure, and J. Nicholson, "RoboCart: Toward Robot-Assisted Navigation of Grocery Stores by Visually Impaired," *Proc. of 2005 IEEE/RSJ International Conf. on Intelligent Robots and Systems*, August 2-6, Edmonton, Alberta, Canada, pp.2845-2850, 2005.

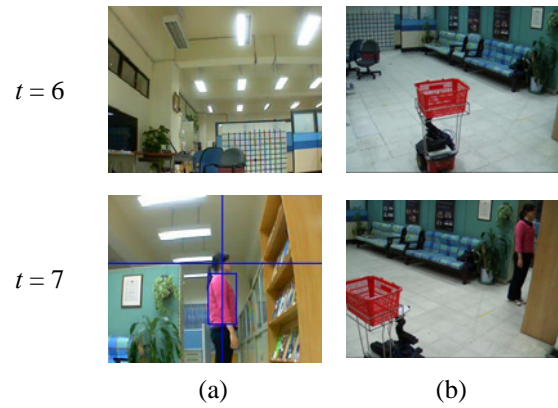


Figure 4. An experimental result of the person turning fast at a corner. (a) The input image. (b) The position of the vehicle.

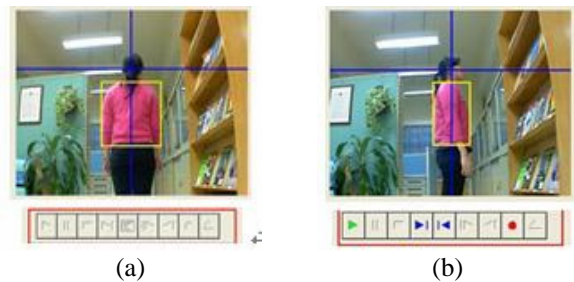


Figure 5. The vehicle makes an introduction about these books. (a) The facing direction of the person is back and the media introduction is off. (b) The person is facing some books and the media introduction is on.

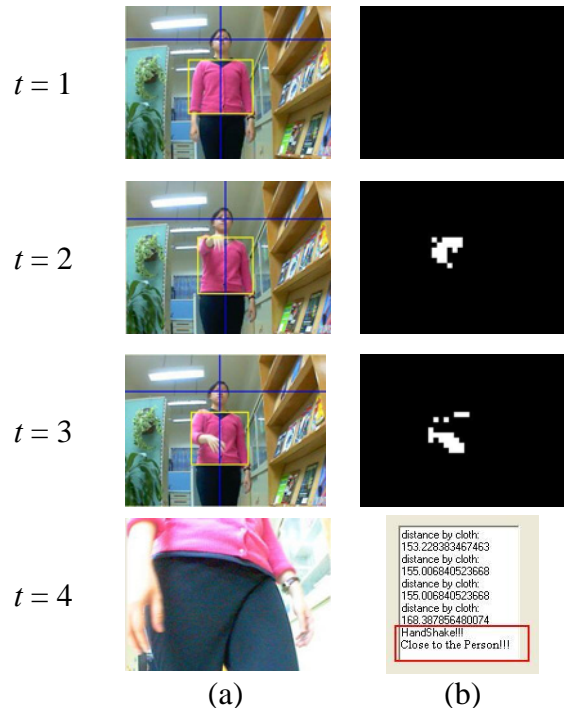


Figure 6. An experimental result of detecting the person waving her hand. (a) The input image. (b) The result of motion detection.