Chapter 6
Human Interaction Techniques for Indoor Person Following

6.1 Introduction

After the vehicle can follow a person by the proposed methods of human detection and following, 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 6.2. 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 this method in Section 6.3. A flowchart of these methods is shown in Figure 6.1 and the detailed process of human interaction is described in the following as an algorithm.

Algorithm 6.1 *Process of human interaction*.

Input: Current image I_c.

Output: The result of a response to the person's interaction.

Steps:

Step 1. Judge whether the person turns to face a side or waves his/her hand in I_c.

Step 2. If the person turns to face a side, then compute the facing direction of the person; else, go to Step 4.

Step 3. Make an audio introduction to what the person is interested in.

Step 4. If the person waves his/her hand, then move the vehicle closer to the person. After that, back to the position where the vehicle stood at the last cycle.

Step 5. End the process of human interaction.

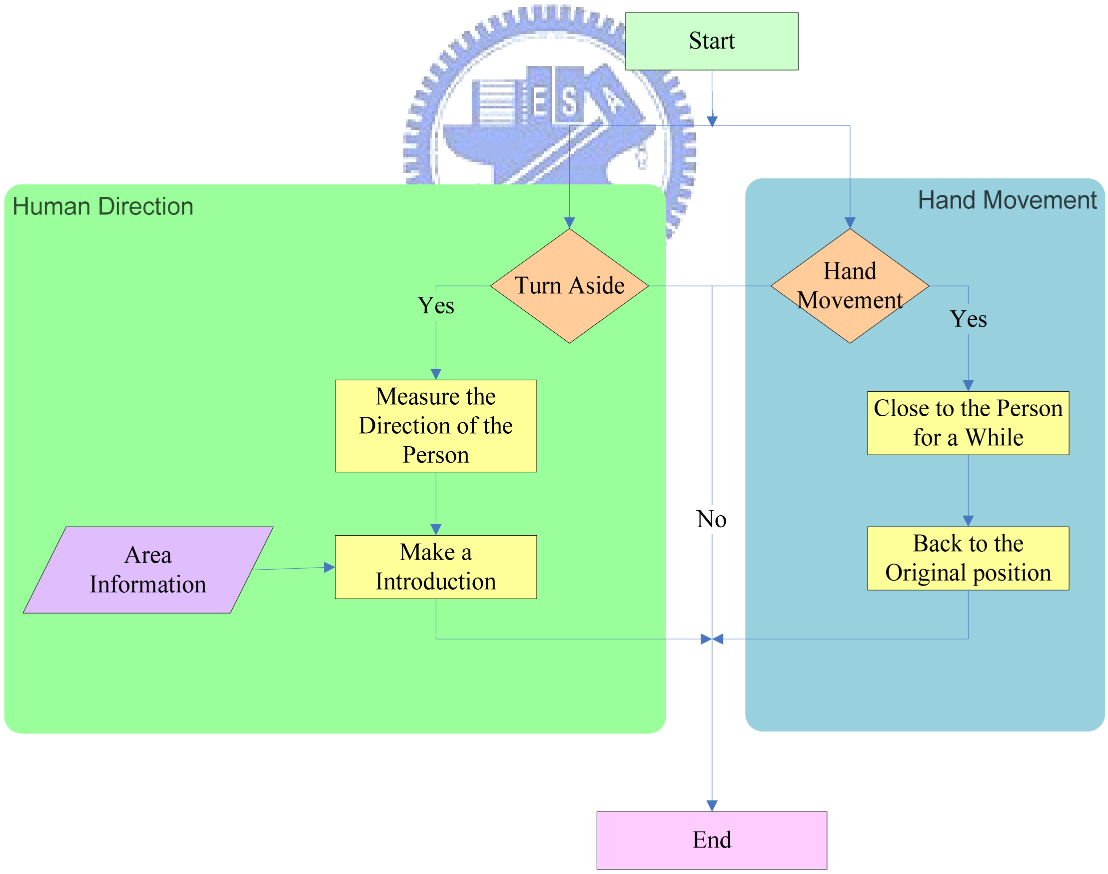


Figure 6.1 A flowchart of human interaction.

6.2 Human Facing Direction Detection

For detection 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 to the right or left, the shape of the clothes extracted in the image becomes thinner. The detail of finding the shape of the clothes and the aspect ratio of the person's body will be described in Section 6.2.1. Then 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 will describe this method in Section 6.2.2. In Section 6.2.3, we will show some experiment results of detecting the person's facing direction. As an application of using the facing direction information and the position of the person, the vehicle can know what the person is interested in, as mentioned previously. Consequently, when the vehicle follows a person, say, in a museum, the vehicle can make an audio introduction to an exhibition which the person is interested in, like a tour guide. We will describe this application in Section 6.2.4. The detailed process of human facing direction detection is described in the following as an algorithm. An illustration of human facing direction detection is shown in Figure 6.2.

Algorithm 6.2 *Process of human facing direction detection*.

Input: Current image I_c.

Output: The result of a response to the person's interaction.

Steps:

Step 1. Measure the aspect ratio of the person body in I_c.

Step 2. Judge whether the person turns to the right or left. If it is, go to Step 3; else, finish the detection.

Step 3. Find the center of the person's face.

Step 4. Measure the distribution of the hair and skin in the face.

Step 5. Apply human facing direction detection.

Step 6. Record the information of the person's facing direction for the application of being as a tour guide.

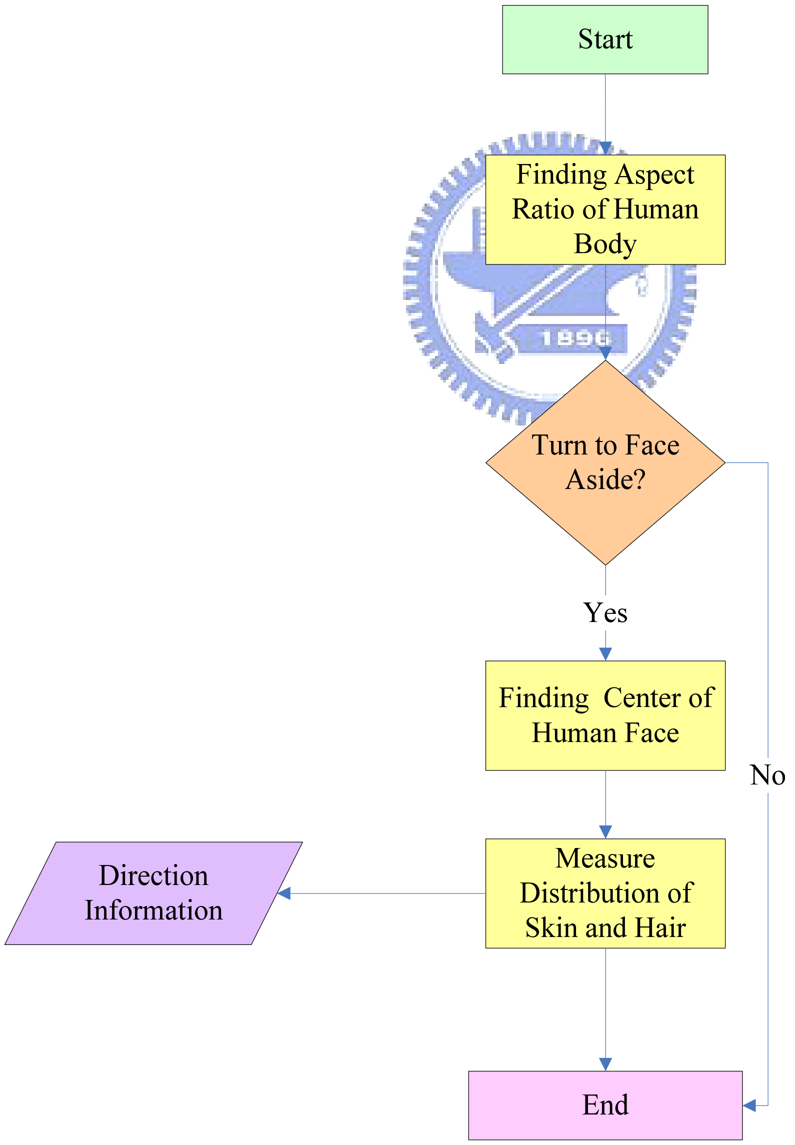


Figure 6.2 An illustration of detection of a person's facing direction.

6.2.1 Finding aspect ratios of human bodies

In the human detection process which is mentioned is Chapter 4, if the person is detected, then the system extracts the region of the clothes and learns the information of the clothes: the four corner points denoted as $P_{TopLeft}(i_{tl},j_{tl})$ $P_{TopRight}(i_{tr},j_{tr})$, $P_{BottomLeft}(i_{bl},j_{bl})$, and $P_{BottonRight}(i_{br},j_{br})$ and the center denoted as $C_a(i_c, j_c)$. By these four points, we can measure the width, denoted as W_{cloth}, and the length denoted as L_{cloth}, of the person's clothes. We also define a ratio, denoted as $R_{standard}$ to measure whether the width of the person's clothes is lager than the length. The detail is stated in the following as an algorithm.

Algorithm 6.3 *Judge whether or not the person turns to face aside.*

Input: The four corner points $P_{TopLeft}(i,j)$ $P_{TopRight}(i,j)$, $P_{BottomLeft}(i,j)$, and $P_{BottonRight}(i,j)$ of the clothes region and a ratio $R_{standard}$.

Output: The result T_{aside} of judging whether or not the person turns to face aside.

Steps:

Step 1. Measure the width W_{cloth} and the length L_{cloth} of the person's clothes below:

$$W_{cloth} = P_{TopRight}(i, j) - P_{TopLeft}(i, j); \tag{6.1}$$

$$.\, L_{cloth} = P_{BottomRight}(i, j) - P_{TopRight}(i, j). \tag{6.2}$$

Step 2. Measure the aspect ratio R_{cloth} of W_{cloth} and L_{cloth} as follows:

$$R_{cloth} = \frac{L_{cloth}}{W_{cloth}}. \tag{6.3}$$

Step 3. If the value of R_{cloth} is larger than $R_{standard}$, set the value T_{aside} as "true;"

else, as "false".

6.2.2 Finding human face centers

Before finding the facing direction of the person, we have to measure the length of the person's face first. When we detect the face of the person, the system learns the length of the face, denoted as L_{FaceI}, in the image and the distance between the person and the vehicle, denoted as D_{faceO}. The length of the face can be computed in the following way.

First, we can obtain Eqs. (6.4) and (6.5) by Eq. (3.3) as follows:

$$M_{face} = \frac{L_{faceI}}{L_{faceO}} = \frac{D_{faceI}}{D_{faceO}}; \tag{6.4}$$

$$M'_{face} = \frac{L'_{faceI}}{L'_{faceO}} = \frac{D'_{faceI}}{D'_{faceO}}. \tag{6.5}$$

Because the distance between the image plane and the lens in a camera will not change, the values D_{FaceI} and D'_{faceI} are the same. And the real length of the person's face is the same in the following process, so the values L_{FaceO} and L'_{FaceO} are the same. So we can obtain Eqs. (6.6) and (6.7) below:

$$D_{faceI} = D'_{faceI}; \tag{6.6}$$

$$L_{faceO} = L'_{faceO}. \tag{6.7}$$

By Eqs. (6.6) and (6.7), we can combine Eqs. (6.4) and (6.5) to get Eq. (6.8) as follows:

$$L_{faceI} \cdot D_{faceO} = L'_{faceI} \cdot D'_{faceO}. \tag{6.8}$$

Then, we can rewrite Eq. (6.8) as Eq. (6.9) below, by which we can compute the length of the person's face in the image:

$$L'_{faceI} = \frac{L_{faceI} \cdot D_{faceO}}{D'_{faceO}}. \tag{6.9}$$

We assume the color of the hair of the person is black. And we also define a region of the values of C_b and C_r as black with a threshold T_1. We use the skin color model, denoted as *Skin*, which is mentioned in Chapter 3. To measure the distribution of the colors of the hair and the skin, we use two thresholds T_2 and T_3. In a scan line going through the center of the face, if the number of pixels with the color of the hair is smaller than T_2 and the skin color is larger than T_3, the system will decide that the person is facing to the vehicle, as shown in Figure 6.3. The detail of finding the direction of the person is stated in the following as an algorithm.

Algorithm 6.4 *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 $P_{BottonRight}(i_{br},j_{br})$ of the clothes region, the length of the person's face L_{face} which can be obtained by Eq. (6.9), 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.

$$i_{face} = i_c; \tag{6.10}$$

$$j_{face} = j_{hair} + \frac{L_{face}}{2}. \quad (6.11)$$

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

$$NL_{Black} = NL_{Black} + 1. \quad (6.12)$$

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

$$NL_{skin} = NL_{skink} + 1. \quad (6.13)$$

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

$$NR_{Black} = NL_{Black} + 1. \quad (6.14)$$

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

$$NR_{skin} = NR_{skink} + 1. \quad (6.15)$$

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

$$NL_{black} + NR_{black} < T_2 ; \qquad (6.16)$$

$$NL_{skin} + NR_{skin} > T_3 ; \qquad (6.17)$$

$$NL_{black} > NR_{black} ; \qquad (6.18)$$

$$NL_{\mathrm{skin}} < NR_{\mathrm{Skin}} . \qquad (6.19)$$

If Inequalities (6.16) and (6.17) are satisfied, set *Direction* as "Front";
else,

if Inequalities (6.18) and (6.19) are satisfied, set *Direction* as "Right";

else, as "Left".

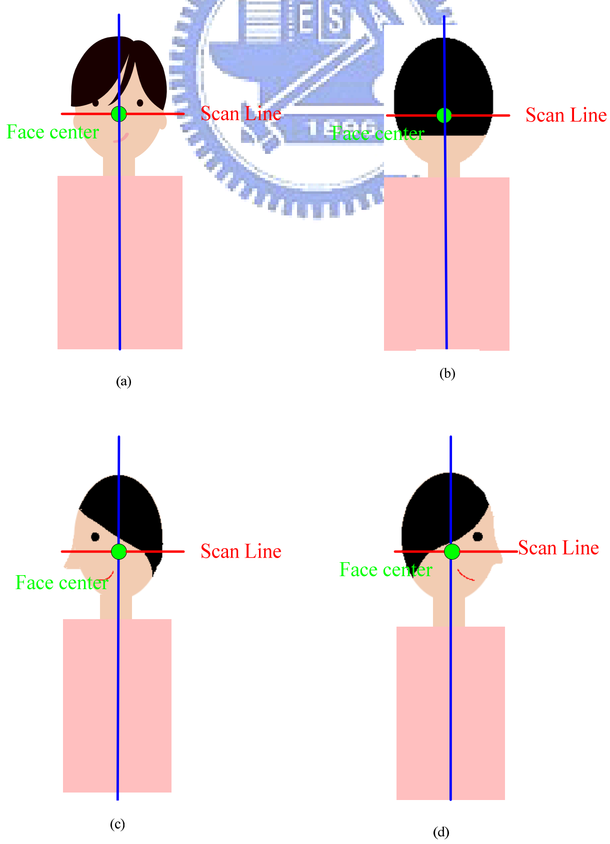
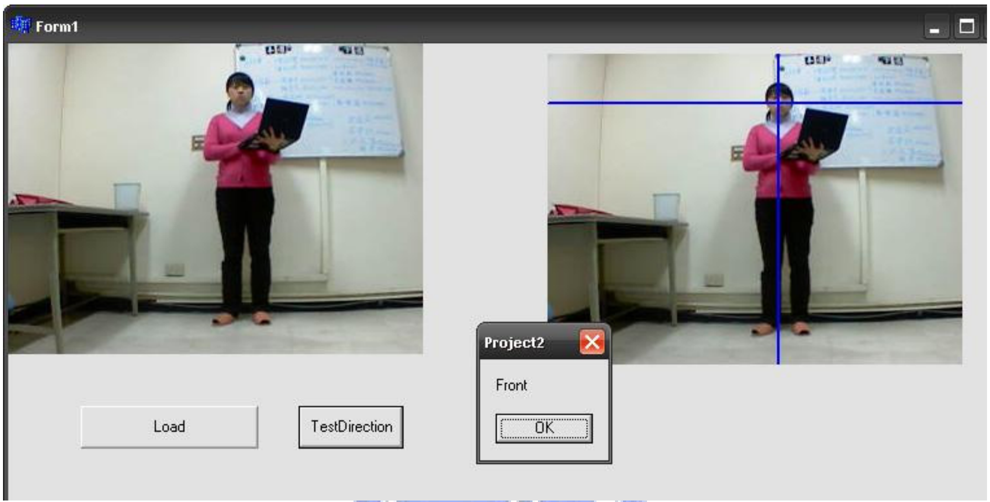


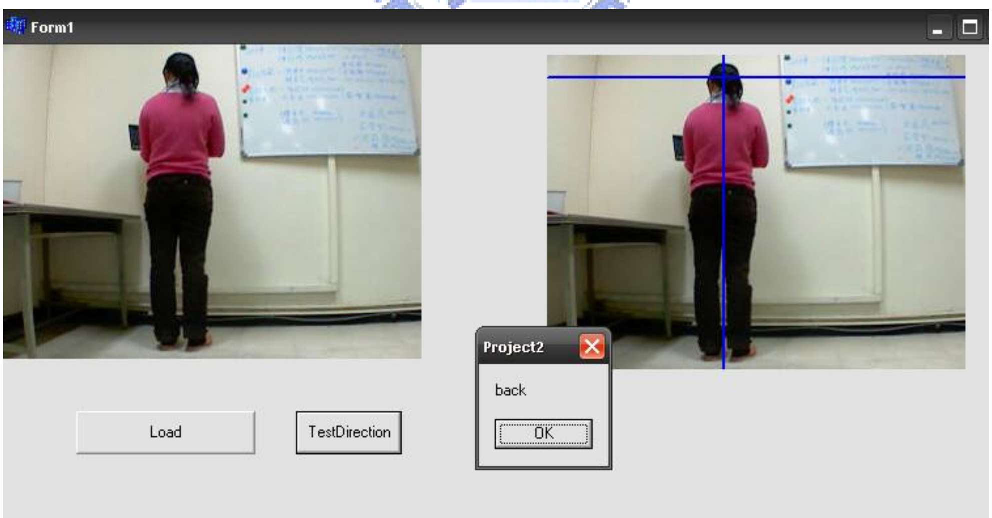
Figure 6.3 An illustration of the facing direction of the person. (a) Front. (b) Back. (c) Left. (d) Right.

6.2.3 Experimental results

In this section, we show some experimental results of facing direction detection in Figure 6.4. Figure 6.4(a) shows a case that the person is facing to the vehicle and Figure 6.4(b) is a case of backing. In Figure 6.4(c), the person is in the right direction relative to the vehicle and Figure 6.4(d) is a case of facing the left.

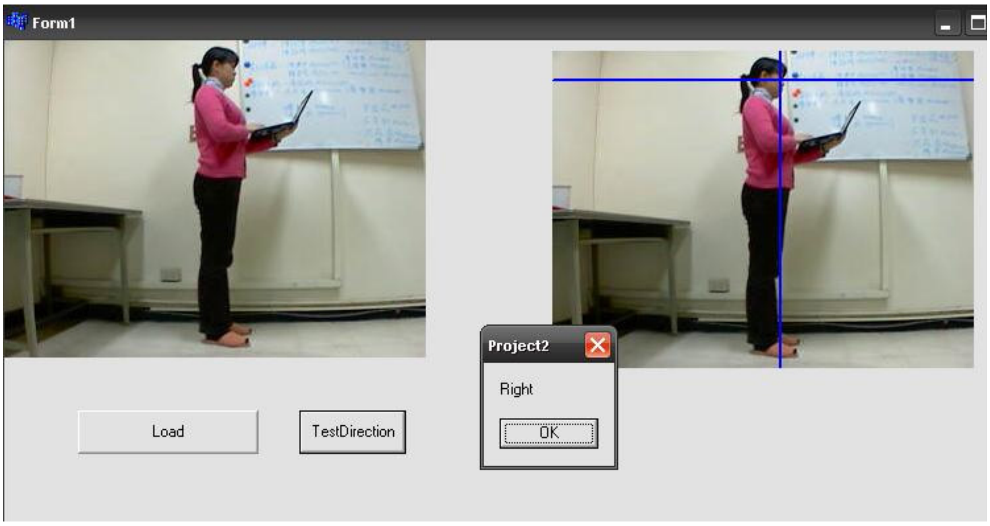


(a)

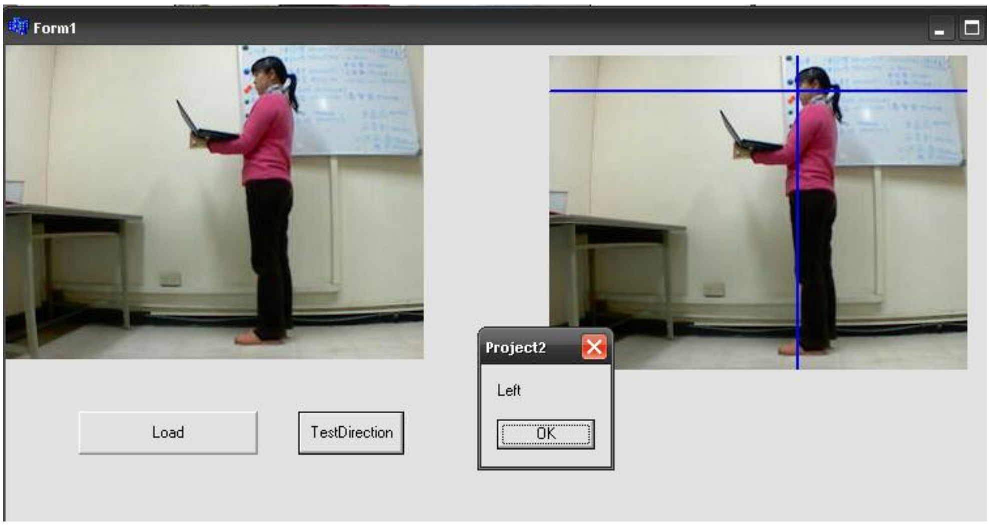


(b)

Figure 6.4 Some experimental results of facing direction detection. (a) Front. (b) Back. (c) Right. (d) Left.



(c)



(d)

Figure 6.4 Some experimental results of facing direction detection. (a) Front. (b) Back. (c) Right. (d) Left. (continued)

6.2.4 Application as a tour guide

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, like the areas in shown Figure 6.6. When the vehicle plays a role as a tour guide, it judges whether the person continues standing at an identical position. If the person does, it means the person is interested in the object in front of the person. 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. For example, in Figure 6.7, the vehicle finds that the person is in Area 9 and faces to the art for a while, then the vehicle will introduce, for example, the history of this art for the person. The process of applying human facing direction detection in this application is shown in Figure 6.5.

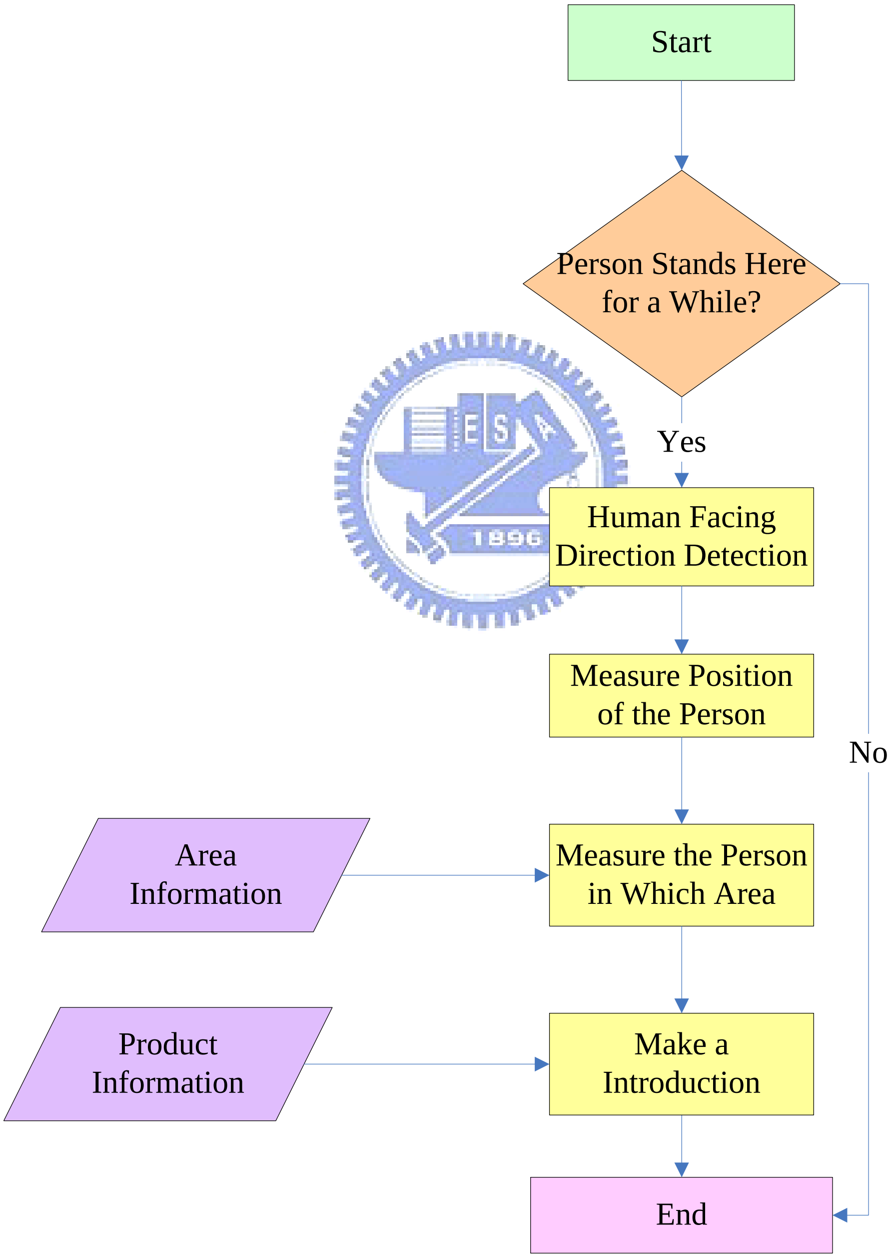


Figure 6.5 The process as a tour guide.

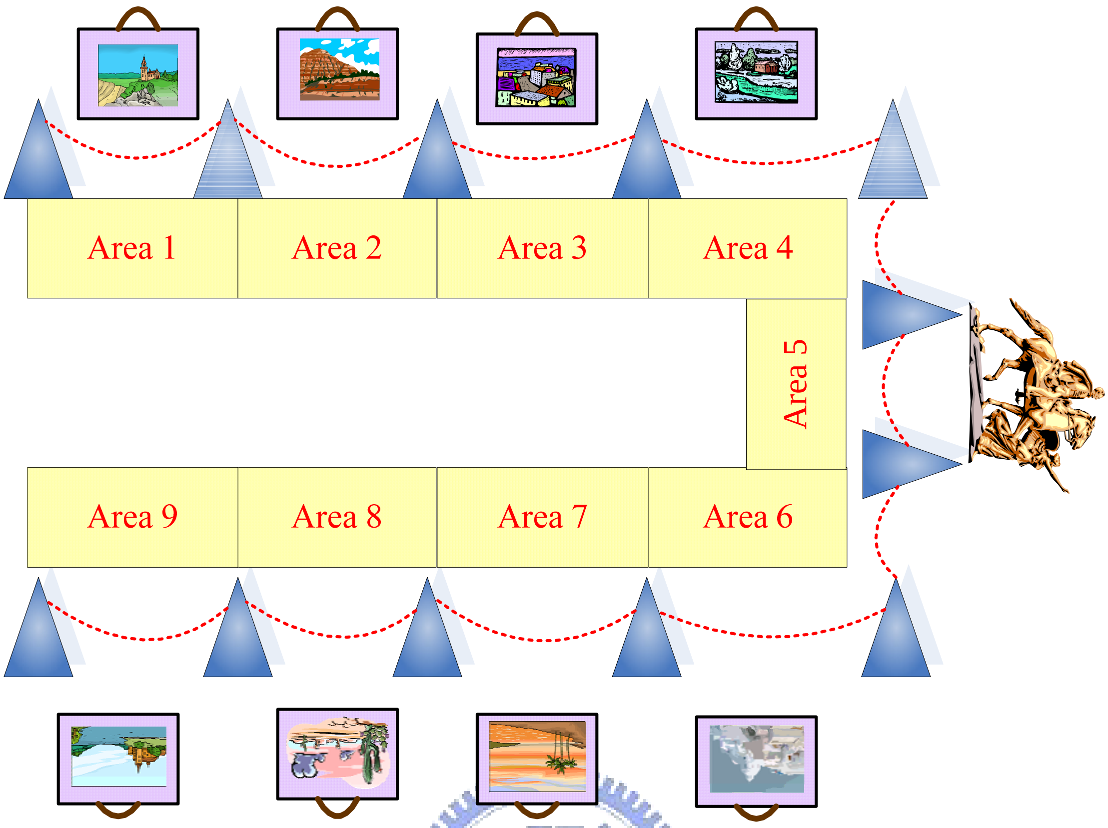


Figure 6.6 An illustration of area information.

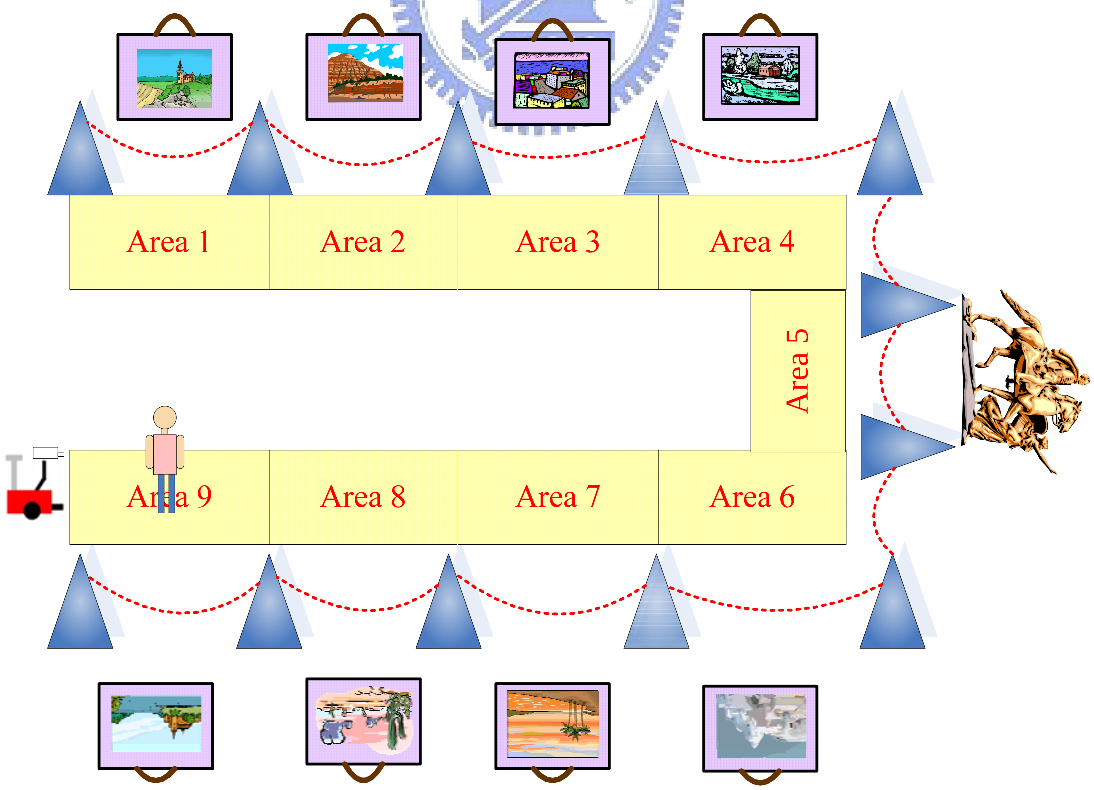


Figure 6.7 An illustration of the vehicle being used as a tour guide.

6.3 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 will be described in Section 6.3.1. As to the action of calling vehicle like a pet, the person has to face to the vehicle and wave his/her hand. Therefore, we combine the information of the person's facing direction to judge whether the person is calling the vehicle. When the person faces to the vehicle, then the vehicle will 'consider' the movement of the hand is aiming to call itself. We will introduce the proposed method to implement this 'scenario' in Section 6.3.2. In Section 6.3.3, we will show the experiment results of detecting the movement of a hand. As an application of using the human hand movement detection, the vehicle can be utilized as a shopping cart in a shopping mall. In Section 6.3.4, we will describe this application. The detailed process of human hand movement detection is described in the following as an algorithm. An illustration is shown in Figure 6.8.

Algorithm 6.5 *Human hand movement detection.*

Input: Current image I_c.

Output: The result of a response to the person's interaction.

Steps:

Step 1. Apply motion detection by frame differencing in I_c.

Step 2. Continuously record the time of detected motion, if there is any.

Step 3. If it continues for a while, regard the person to be moving his/her hand;

else, go to Step 1.

Step 4. Combine the information of the person's facing direction.

Step 5. Judge whether the person is calling the vehicle. If he/she is, end the detection of the movement of the person's hand; else, go to Step 1.

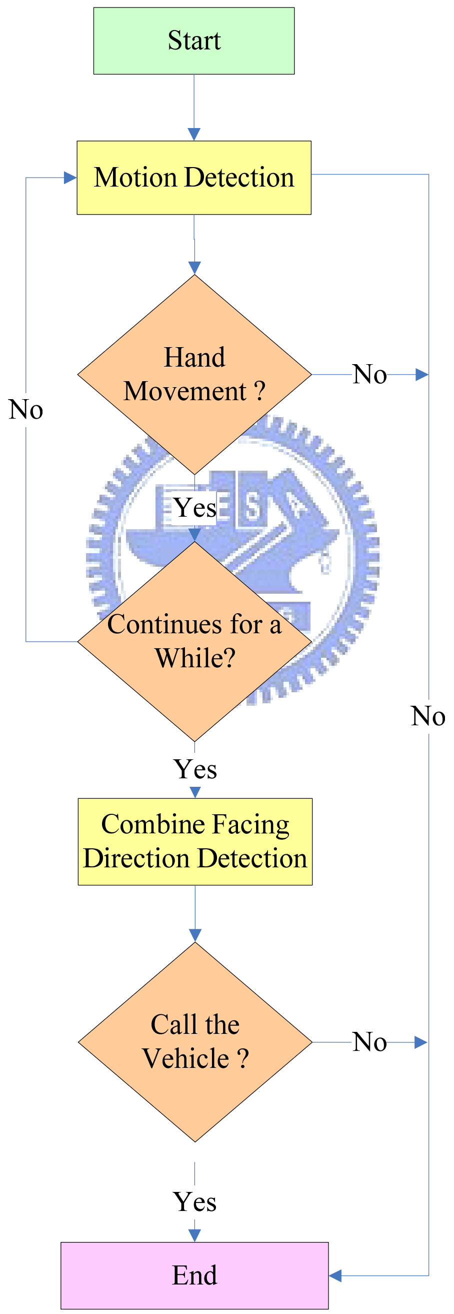


Figure 6.8 An illustration of detection of hand movement

6.3.1 Motion detection by frame differencing

For detecting the motion in the image, we adopt the method of frame differencing. Subtracting the current image from the reference image pixel by pixel is the basic idea of frame differencing. If the difference between a pixel in the current image and the corresponding pixel at the same position in the reference image is below some threshold, then it is may be considered that no motion has taken place at this pixel, and we call this pixel *stationary*. If it is not, we try to find the 'best match' pixel within a search window, denoted as *w*, in the reference image. If the best match pixel is below the threshold, we say that the target block is *stationary*; otherwise, *moving*. Repeating these steps for each pixel in the current image, we can get all the moving parts in the current image. The detail is stated in the following as an algorithm.

Algorithm 6.6 *Motion detection by frame differencing*.

Input: current image I_c, reference image I_r, thresholds T_1 and T_2, 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_1, 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_2, regard the pixel P_{ij} as *stationary*; else, *moving*.

Step 4. Repeat Step 2 for each pixel in 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.

6.3.2 Vehicle calling by combining the facing direction detection method

After detection of motion, we have to know whether or not the person waves his/her hand. The idea of detection of vehicle calling is like calling our pets. When we call our pets, we will turn to it and wave our hand in front of our body. Therefore, we measure the moving pixel inside the region of the clothes from the difference image first. Then we use the method of region growing to find the biggest region of the moving part in the region of the clothes. Finally, we combine the facing direction of the person which is mentioned in Section 6.3.1 to judge whether or not the person is facing to the vehicle and calling the vehicle for needs. The detail is stated in the following as an algorithm.

Algorithm 6.7 *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 R_{cloth}, regard this pixel to be in a moving point set M = {M

$_{00}$, M_{01}, …, M_{nn}}; else, repeat Step 1 to check the next pixel.

Step 2. Compute the moving region set R_{moving} = {$R_{m,00}$, $R_{m,01}$, …, $R_{m,mn}$} 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."

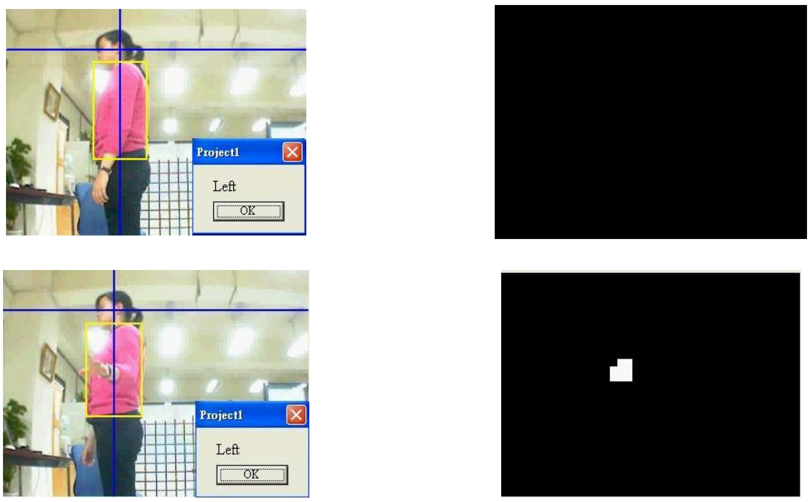
Step 4. If $R_{HandWave}$ is "true" and the facing *Direction* is "Front", set $R_{calling}$ as "true"; else, as "false".

6.3.3 Experimental results

In this section, we show some experimental results of hand movement detection in Figure 6.9. The person does not wave her hand and no motion was detected in the image at the moment t = 1. Then the person waves her hand and the movement of the hand was detected in the image at the next moments, t = 2 and t = 3. Finally, combining this result with the information of the detected facing direction, the vehicle 'knows' that the person was calling it at t = 4.

t = 1

t = 2

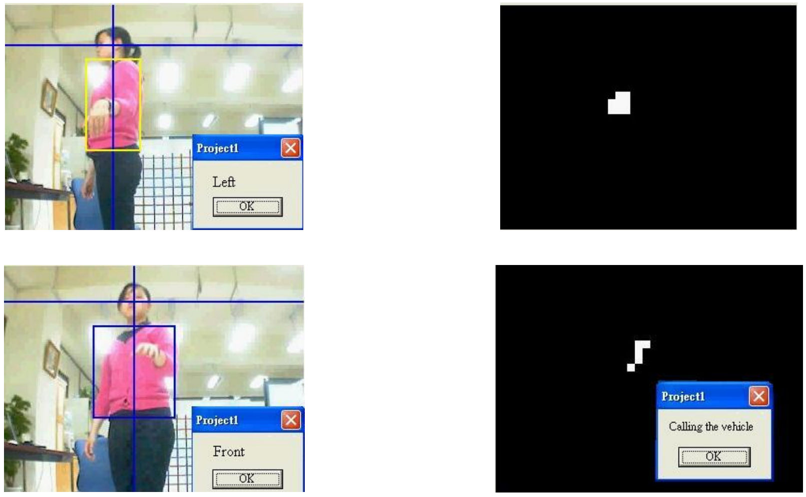


(a) (b)

Figure 6.9 An experimental result of human hand movement detection. (a) The detection of the facing direction in input images. (b) The result of the motion detection.

t = 3

t = 4



(a) (b)

Figure 6.9 An experimental result of human hand movement detection. (a) The detection of the facing direction in input images. (b) The result of the motion detection. (continued)

6.3.4 Application as a shopping cart

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 standing at an identical position and waving 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 will know whether the person is calling it and move close to the person to serve him/her. For example, in Figure 6.11, the vehicle found the person was standing in front of some milk products for a sufficiently long time and waved her hand. Then the vehicle went forward to the person and waited for a while for the person to put the milk product into the basket. Finally the vehicle backed to the original position and continues the regular ‘mission’ of following the person. The process of applying human hand movement detection in this application is shown in Figure 6.10. Some experimental results of the vehicle being used as a cart in the library are shown in Figure 6.12. The vehicle followed a person in the library at the moment t = 1. Then the person waved her hand and the movement of the hand was detected at the next moments, t = 2 and t = 3. The vehicle moved close to the person at the moment t = 4. Therefore, the person could put the book in the basket, which the person is interested in, at the moment t = 5. Finally, the vehicle backed to the position where it stood, and continued following the person at the moment t = 6.

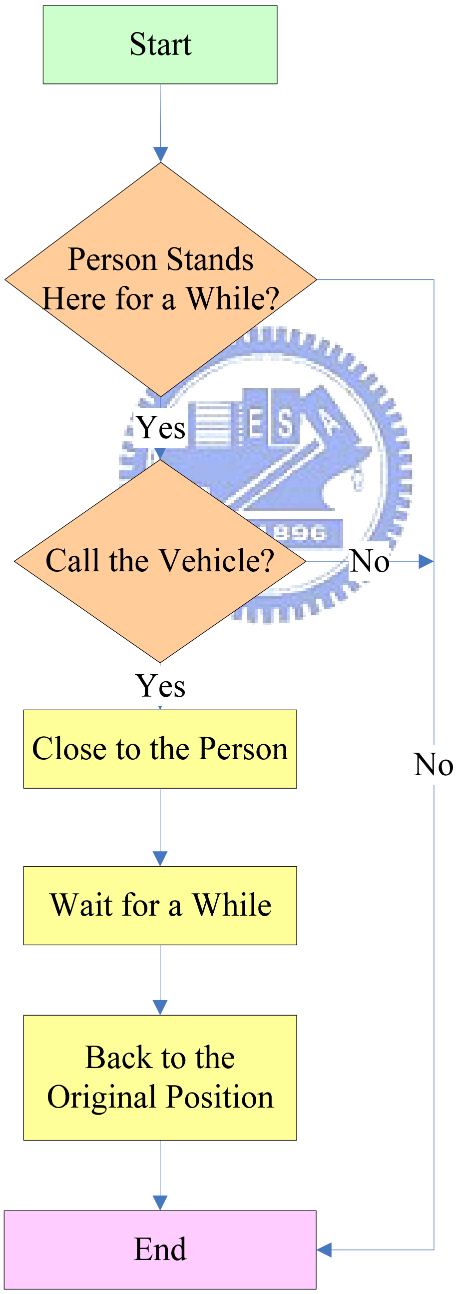


Figure 6.10 The process as a shopping cart.

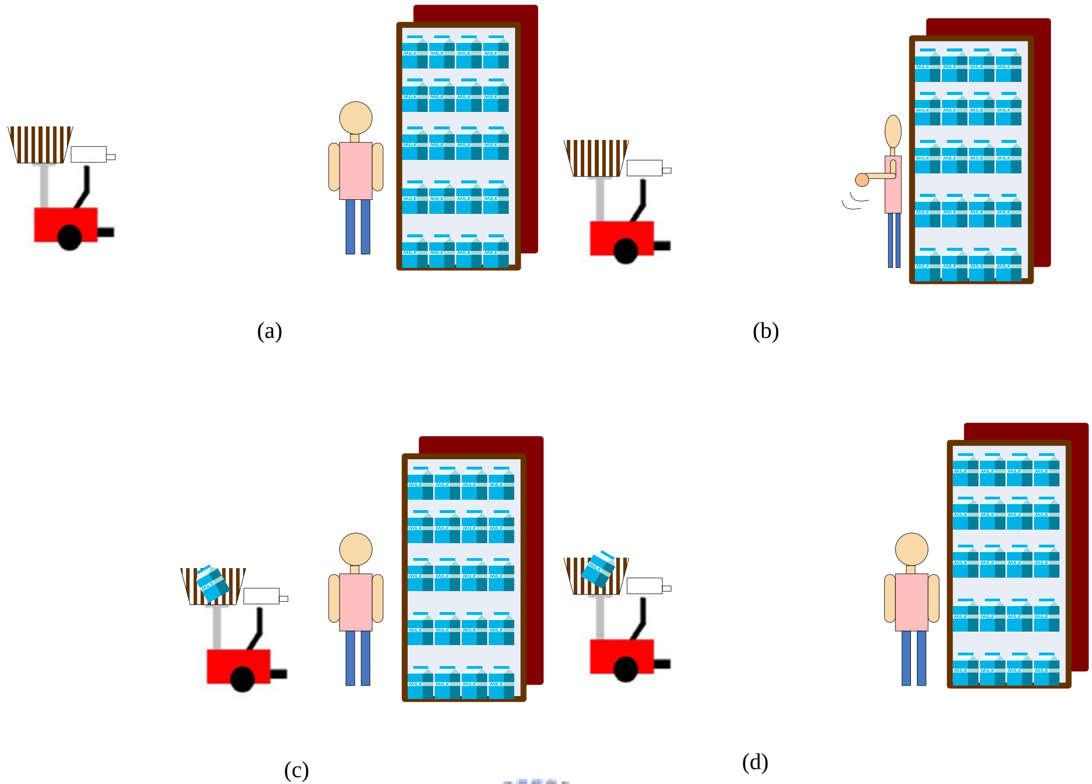


Figure 6.11 The illustration of calling the vehicle as a shopping cart.

$t = 1$

$t = 2$

Figure 6.12 An experimental result of the vehicle being a cart in the library.

t = 3

t = 4

t = 5

t = 6

Figure 6.12 An experimental result of the vehicle being a cart in the library. (continued)