Chapter 3
Camera Calibration for Distance Computation

3.1 Introduction

Distance information is important and useful in the person following process. First, because the vehicle knows the position of itself by an odometer in the vehicle, the vehicle can use the distance information to compute the relative position of a human and know where the human is in the real world. Second, the vehicle can avoid striking a person in front of it by using the distance information to keep a safe distance to the person. Third, the vehicle can make suitable responses to achieve smooth and stable person following. For example, if a person is far from the vehicle, the vehicle may go forward to see more clearly for avoiding losing information of this person. So distance information is indispensable for a person-following vehicle.

But the camera is the only sensor on the vehicle, and to know if this person is far or close can only be accomplished by analyzing images captured from the camera. There is ambiguity in the inverse mapping from 2D image coordinates to a 3D world position. So for the proposed vehicle system, we adopt a method of angular-mapping camera calibration which was proposed by Wang and Tsai [3] to compute the distance between a vehicle and a person. However, when the vehicle computes the distance by this method, the system has to know the height of the person in advance. Also, the vehicle cannot follow the person when the vehicle is too close to the person. The reason is that as long as the vehicle cannot see the entire clothes of the person, the proposed method cannot work to measure the distance between the vehicle and the person.

In Section 3.2, we will review the proposed method of angular-mapping camera calibration. In Section 3.3, we will propose a method which can adapted to different people's heights so that the system does not have to know the height of the person in advance. When the vehicle is close to a person and cannot see the entire clothes of the person, we propose additionally an *area tracking* method to keep following the person. And the method is described in Section 3.4.

Before describing the above-mentioned methods, we first introduce the definitions of the coordinate systems and the directional angle of the camera used in the study. We introduce the former in Section 3.1.1 and the latter in Section 3.1.2.

3.1.1 Coordinate Systems

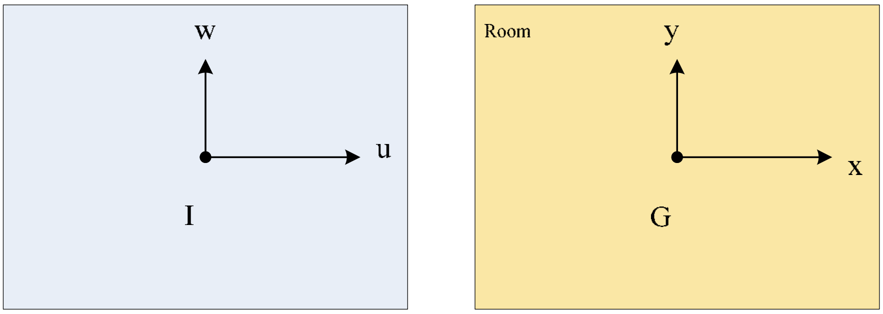
Four coordinate systems are utilized in this study which describes the relative locations between the vehicle and encountered objects. The coordinate systems are shown in Figure 3.1. The definitions of all the coordinate systems are stated in the following.

(1) Image coordinate system (ICS): denoted as (u, v). The uv-plane of the system is coincident with the image plane and the origin I of the ICS is placed at the center of the image plane.

(2) Global coordinate system (GCS): denoted as (x, y). The x-axis and the y-axis are defined to lie to on the ground, and the origin G of the global coordinate system is a pre-defined point on the ground. In this study, we define G as the starting position of the person-following process.

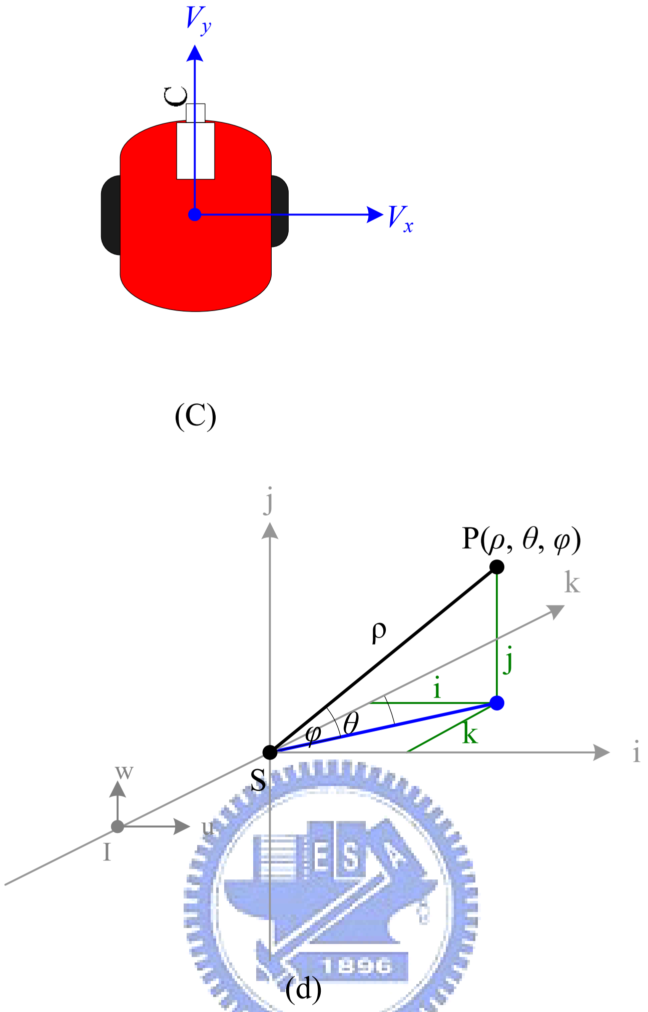
(3) Vehicle coordinate system (VCS): denoted as (V_x, V_y). The V_xV_y-plane is coincident with the ground. And the origin V is placed at the middle of the line segment that connects the two contact points of the two driving wheels with the ground. The V_x-axis of the system is parallel to the line segment joining the two driving wheels and through the origin V. The V_x-axis is perpendicular to the x-axis and goes through V.

(4) Spherical coordinate system (SCS): denoted as (ρ, θ, φ). This system is proposed by Wang and Tsai [3]. It is a 3D polar coordinate system and we explain this system in terms of the 3D Cartesian coordinate system with coordinates (i, j, k) for convenience. The origin S of the spherical system, which is also the origin of the Cartesian system, is the optical center of the camera. The ij-plane of the Cartesian system is parallel to the uv-plane in the ICS. A point P at coordinates (i, j, k) in the Cartesian space is represented by a 3-tuple (ρ, θ, φ) in the spherical space. The value ρ with $\rho \geq 0$ is the distance between the point P and the origin S. The longitude θ is the angle between the positive k-axis and the line from the origin S to the point P projected onto the ik-plane. The latitude φ is the angle between the ik-plane and the line from the origin S to the point P.



(a) (b)

Figure 3.1 The coordinate systems used in this study. (a) The image coordinate system. (b) The global coordinate system (c) The vehicle coordinate system. (d) The spherical coordinate system.



(C)

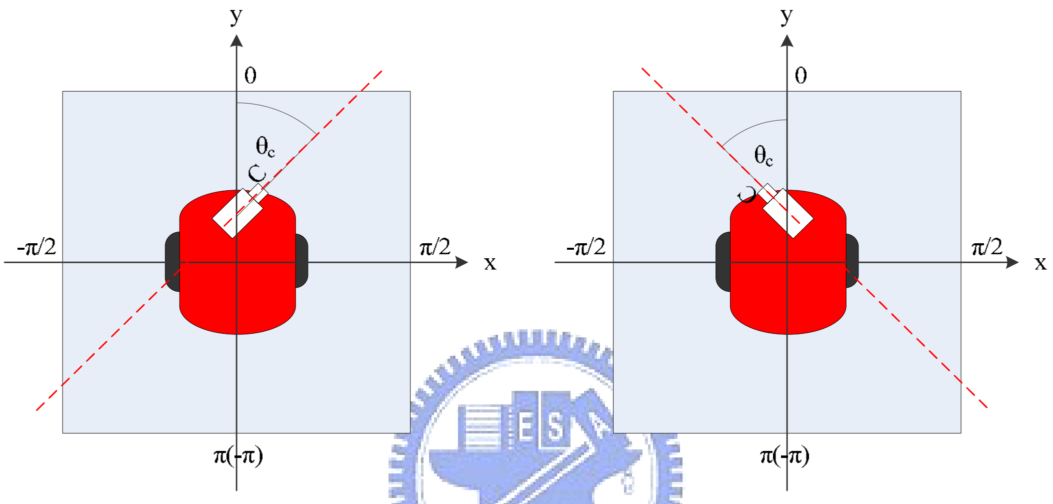
(d)

Figure 3.1 The coordinate systems used in this study. (a) The image coordinate system. (b) The global coordinate system (c) The vehicle coordinate system. (d) The spherical coordinate system. (continued)

3.1.2 Directional Angles of Camera

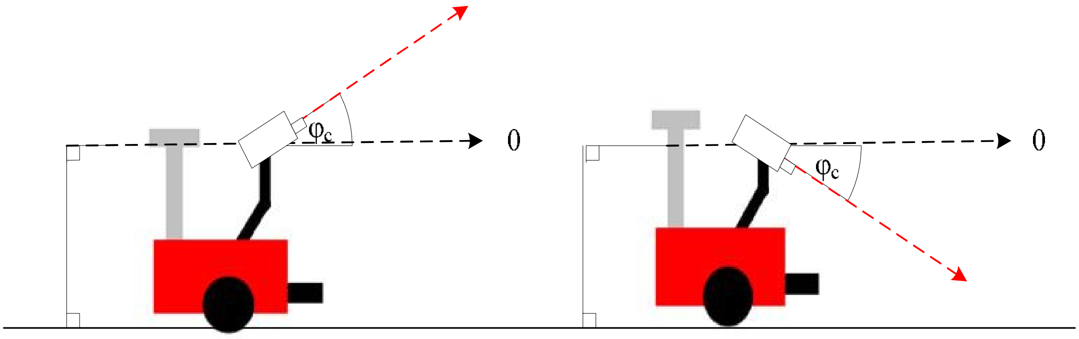
Two kinds of directional angles of a camera are used in this study. One is the pan angle and the other the tilt angle. The pan angle of the camera is defined in the VCS, denoted by θ_c. It represents the degree of horizontal rotation of the camera and is important for coordinate transformation.

We define the direction of the y-axis to be zero. The value of θ_c is exactly the angle between the camera direction and the direction of the y-axis. The range of θ_c is between 0 and π if θ_c is in the first and fourth quadrants and between 0 and $-\pi$ if θ_c is in the second and third quadrants, as shown in Figure 3.2. The tilt angle of the camera is defined as the angle between the optical axis of the camera and the ground. The angle, denoted as φ_c, represents the vertical tilting of the camera. We define the angle to be zero when the optical axis of the camera is parallel to the ground. The range of φ_c is between 0 and π/2 if the camera tilts up, and is between 0 and –π/2, else, as shown in Figure 3.3.



(a) (b)

Figure 3.2 The pan angle of the camera. (a) $0 \le \theta_c \le \pi$. (b) $0 \ge \theta_c \ge -\pi$.



(a) (b)

Figure 3.3 The tilt angle of the camera. (a) $0 \le \varphi_c \le \frac{\pi}{2}$. (b) $0 \ge \varphi_c \ge -\frac{\pi}{2}$.

3.2 Review of Proposed Distance Computation Method

Wang and Tsai [3] proposed a nonlinear angular mapping method to precisely obtain the angular transformation from the real world to the image. By the angular information of the light rays and the height of the camera, we can know the relative distances of targets in images.

The coordinate system used in this method is shown in Figure 3.4, which includes an image coordinate system (ICS) described by image coordinates (u, v) and a spherical coordinate system (SCS) described by parameters (ρ, θ, φ). The latter is a 3D polar coordinate system which can be explained in terms of the 3D Cartesian coordinate system with coordinates (i, j, k). The ij-plane of the Cartesian system is parallel to the uv-plane in the ICS. The origin S of the SCS, which is also the origin of the Cartesian system, is the optical center of the camera. A point P at coordinates (i, j, k) in the Cartesian space is represented by a 3-tuple (ρ, θ, φ) in the SCS where ρ is the distance between the point P and the origin S. The longitude θ is the angle between the positive k-axis and the line from the origin S to the point P projected onto the ik-plane, and the latitude φ is the angle between the ik-plane and the line from the origin S to the point P.

From the mapping of the ICS to the world coordinate system, it is impossible to figure out the distance between the eye point and the point P due to the inherent ambiguity of the light ray projection. However, the projection P' of P can be represented by the longitude θ and latitude φ of P in the real world, as shown in Figure 3.5. Because camera distortion exists both horizontally and vertically, a real world data acquisition method by angular-mapping camera calibration is proposed here to compute the longitude and latitude values of each point in the image.

A grid board is used in this method. It has *m* vertical lines and *n* horizontal lines, and is attached on a wall which is perpendicular to the ground. Because the longitude and the latitude values of the intersection points in the grid have been known, the longitude and the latitude values of the other pixels in the image can be computed by an interpolation method. In this way, the longitude and the latitude values of each pixel in the image can be obtained.

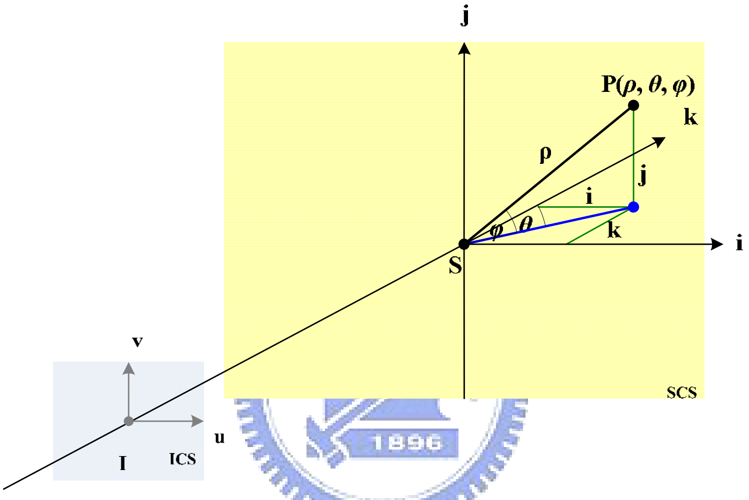


Figure 3.4 Coordinate systems used (a) Image coordinate system. (b) Spherical coordinate system.

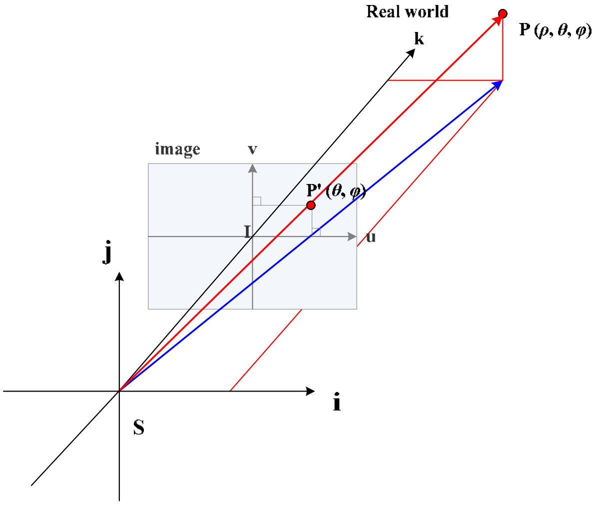


Figure 3.5 Image coordinates mapped into real world space.

However, by angular transformation only, this method cannot compute the distance between a vehicle and a person. It has to know the height of the person's clothes to compute the distance in advance. It was assumed that the height of his/her body part is around 50cm to 60cm. An illustration of the distance between the person and the vehicle is shown in Figure 3.6. Therefore, the range of distance [d_0, d_1] which is between the vehicle and the person can be computed by Eqs. (3.1) and (3.2) as follows:

$$\tan\varphi_1 - \tan\varphi_2 = \frac{h_1 - h_2}{d}; \tag{3.1}$$

$$\begin{bmatrix} d_0 \\ d_1 \end{bmatrix} = \begin{bmatrix} \dfrac{50}{\tan\varphi_1 - \tan\varphi_2} \\ \dfrac{60}{\tan\varphi_1 - \tan\varphi_2} \end{bmatrix}. \tag{3.2}$$

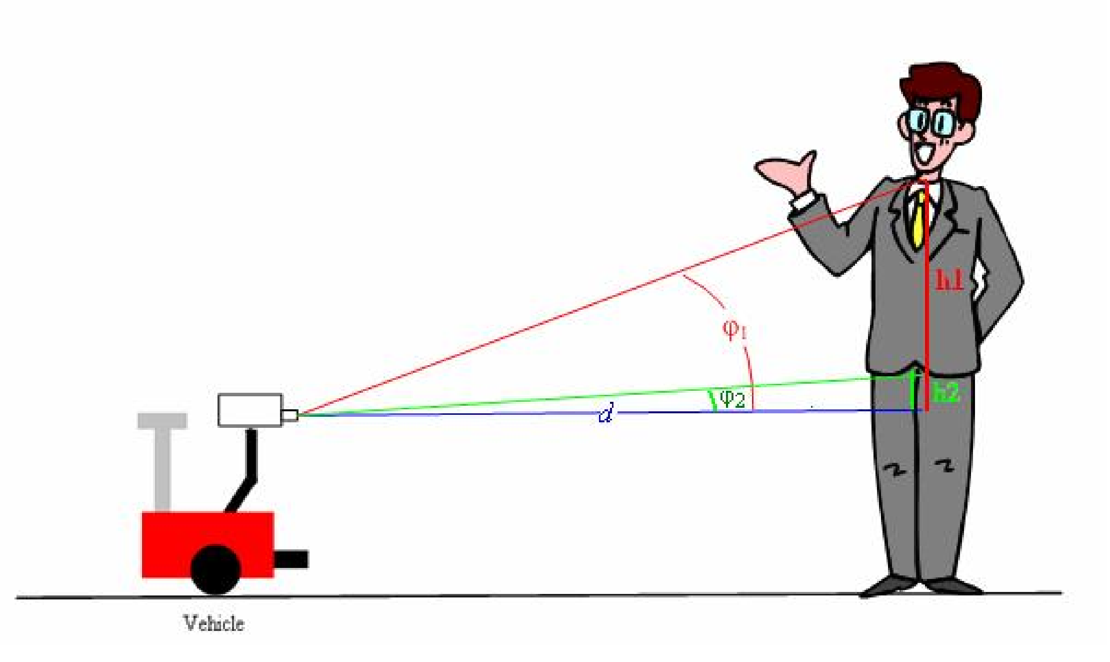


Figure 3.6 The illustration of the distance *d* between a person and the vehicle.

3.3 Adaptation to Different People's Heights

By using the distance computation method which is mentioned in Section 3.2, we can compute the distance between the vehicle and the object under the assumption that the height of the clothes of a human is in the range of 50cm to 60cm. It means that if this height is out of the above range, the system is not suitable for this person.

In Section 3.3.1, we propose a method which can detect the length of the clothes of a person and uses the reference data information to check whether for this person, the vehicle system should use the original parameter settings or adjust them. For either way, we compute the real height of this person. For distance computation, we have to let the camera see the whole length of the person's clothes. In Section 3.3.2, we use both the reference data information and the real height of the person to change the viewing distance for adapting to difference people's heights.

3.3.1 Using reference data information

To solve the problem of computing distances for different people's heights, we propose a method that uses the reference data information. The process is divided into two parts: getting the reference data and using them in the system. We define the reference data as shown in Figure 3.7 and Figure 3.8. To begin with getting the reference data, we define a *reference person* as a standard for measuring a set of parameters which are used to compute the distance between the vehicle and the person. By the reference person, we measure the length L_{refI} of his/her clothes in the real world, and that, L_{refO}, in the image. We also define the distance, $D_{standardI}$, between the original place of the reference person and that of the vehicle to be 2m. Because the vehicle has to see the whole clothes of the followed person to measure the distance, we back the vehicle, when necessary, to do so. Once the vehicle can see the whole clothes of the reference person, we stop the vehicle and measure the distance between the vehicle and the reference person. Then we define this distance to be the *standard viewing distance*, D_{refI}. The detailed process of getting the reference data information is described in the following as an algorithm.

Algorithm 3.1 *Getting the reference data information*.

Input: Current image I_c

Output: The reference data L_{refO}, L_{refI}, $D_{standardI}$. *and* D_{refI} as defined above.

Steps:

Step 1. Ask the reference person to stand in front of the vehicle.

Step 2. Back the vehicle.

Step 3. If the vehicle can see the whole clothes of the person, go to Step 4; otherwise, go to Step 2.

Step 4. Measure the distance D_{refI} between the vehicle and the person.

Step 5. Back the vehicle until the distance between the vehicle and the person is 2m.

Step 6. Compute the length of the person's clothes in the image as L_{refO}.

Step 7. Take the distance 2m as $D_{standardI}$ and the real length of the clothes of the reference person L_{refI} as 50cm.

Step 8. Record the reference data L_{refO}, L_{refI}, $D_{standardI}$ *and* D_{refI}

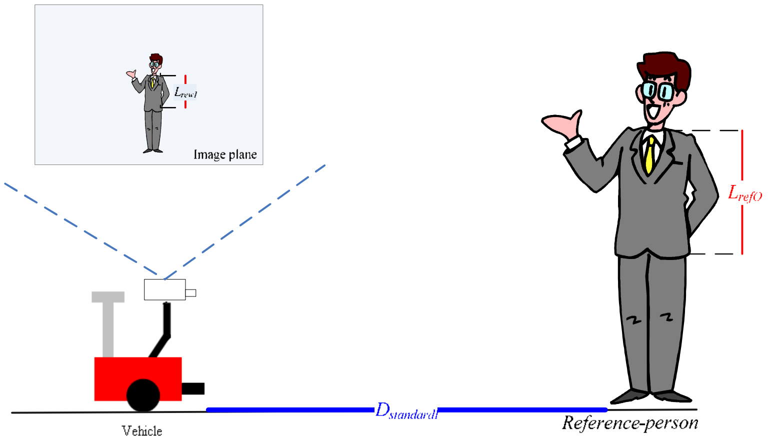
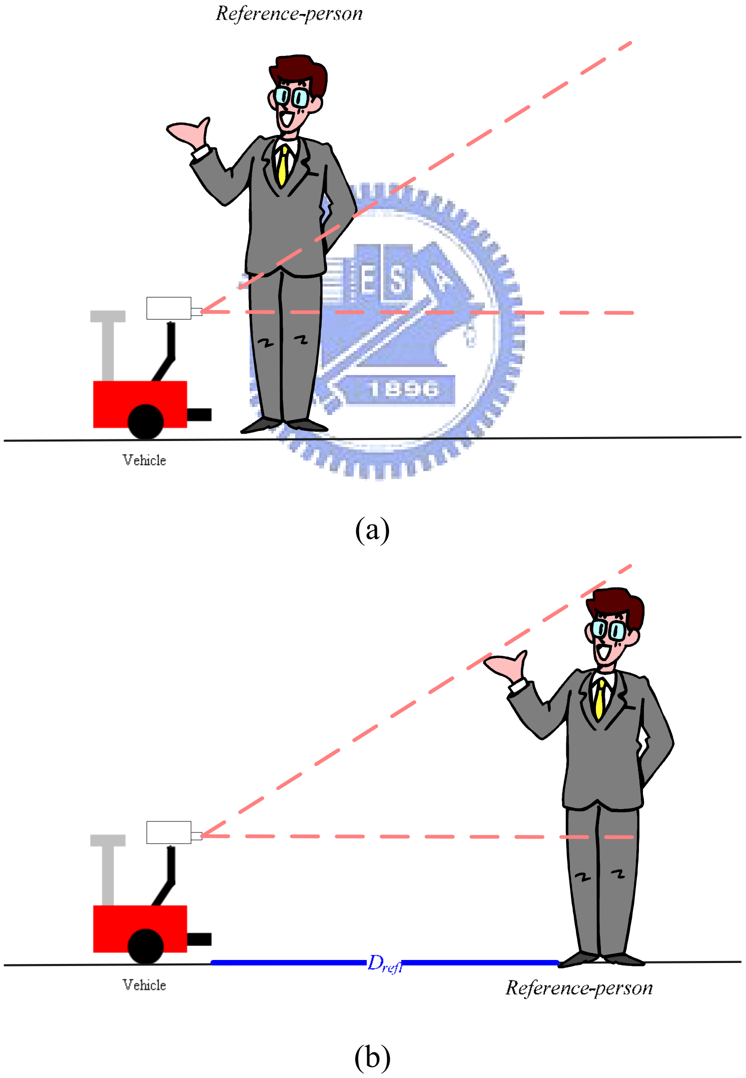


Figure 3.7 Te reference data information



(a)

(b)

Figure 3.8 Getting the reference data $D_{ref1.}$ (a) Reference person standing in front of the vehicle. (b) The vehicle can see the whole clothes of the Reference person.

Then, we can use the reference data to compute the length of clothes of any new person followed by the vehicle in the real world and in the image. We can compare these data with the reference data to check whether or not we have to adjust our parameters for measuring the distance between the vehicle and the person. To measure the length of the clothes of the person, we use a “magnification equation.” The equation relates the ratio of the image distance (D_I) to the object distance (D_0) to the ratio of the image height (L_I) to the object height (L_0). The magnification equation is as follows:

$$M = \frac{L_I}{L_0} = \frac{D_I}{D_0}. \tag{3.3}$$

Substituting the values of L_I, L_0, D_I and D_0 with L_{newI}, L_{newO}, D_{newI} and D_{newO} of the person and with L_{refI}, L_{refO}, D_{refI} and D_{refO} of the reference person, we obtain Eqs. (3.4) and (3.5) below:

$$M_{new} = \frac{L_{newI}}{L_{newO}} = \frac{D_{newI}}{D_{newO}}; \tag{3.4}$$

$$M_{ref} = \frac{L_{refI}}{L_{refO}} = \frac{D_{refI}}{D_{refO}}. \tag{3.5}$$

Because the distance between the image plane and the lens will not change in a camera, the values D_{newO} and D_{refO} are the same. The person and the reference person both stand at a distance of 2m in front of the vehicle. So we can obtain Eqs. (3.6) and (3.7) below:

$$D_{newI} = D_{refI} = 2m; \tag{3.6}$$

$$D_{newO} = D_{refO}. \tag{3.7}$$

By Eqs. (3.6) and (3.7), we can combine Eqs. (3.4) and (3.5) to get Eq. (3.8) as follows:

$$\frac{L_{refI}}{L_{refO}} = \frac{L_{newI}}{L_{newO}}. \tag{3.8}$$

Then, we can rewrite Eq. (3.8) as Eq. (3.9) below, by which we can compute the length of the clothes of the person:

$$L_{newI} = \frac{L_{newO} \cdot L_{refI}}{L_{refO}}. \tag{3.9}$$

When we use the reference data for measuring the distance, the person has to stand at a distance of 2m (denoted as $D_{standardI}$) in front of the vehicle first, and we can obtain the length of the clothes of the person in the image, L_{newO}. Then we compare this value with the reference data, L_{refO}. If the result, C_{same}, is "the same", we do not have to change the parameters in the system. Otherwise, we have to adjust them to adapt to this person with a height different from that of the reference person. The detailed process of using the reference data information is described in the following as an algorithm.

Algorithm 3.2 *Using the reference data information.*

Input: The reference data L_{refO}, and the real length of the clothes of the reference person L_{refI}.

Output: The real length of the clothes of the person L_{newI}, the length of the clothes of the person in the image L_{newO} and the comparison result C_{same}.

Steps:

Step 1. Ask the person who uses the system to stand at a distance of $D_{standardI}$ in front of the vehicle

Step 2. Compute the length of the clothes of this person in the image as L_{newO}.

Step 3. Compare the values of L_{refO} and L_{newO}.

Step 4. If L_{newO} is the same as L_{refO}, use the original parameters L_{refI} to compute the distance and D_{refI} to be the safe distance, by taking the real length of the clothes of the person L_{newI} to be L_{refI} , the length of the person in the image L_{newO} to be L_{refO} and set the compare result C_{same} as “true.”

If L_{newO} is not the same as L_{refO}, compute the real length of the clothes the person L_{newI} by Eq. (3.9). And then record L_{newI} and L_{newO}, and set the value of C_{same} as “false.”

3.3.2 Changes of viewing distance

For people with heights different from that of the reference person, because the vehicle has to see the whole clothes of them, we have to change the viewing distance. By rewriting Eq. (3.3), we can obtain Eq. (3.10) below:

$$\frac{L_I}{D_I} = \frac{L_O}{D_O}. \tag{3.10}$$

Substituting the values of L_I, L_0, D_I and D_0 with L_{newI}, L_{newO}, D_{newI} and D_{newO} of the person and with L_{refI}, L_{refO}, D_{refI} and D_{refO} of the reference person, we obtain Eqs. (3.11) and (3.12) below:

$$\frac{L_{newI}}{D_{newI}} = \frac{L_{newO}}{D_{newO}}; \tag{3.11}$$

$$\frac{L_{refI}}{D_{refI}} = \frac{L_{refO}}{D_{refO}}. \tag{3.12}$$

Because the distance between the image plane and the lens will not change in a camera, the values D_{newO} and D_{refO} are the same. And we want to see the whole clothes of the person, so we can obtain Eqs. (3.13) and (3.14) below:

$$D_{refO} = D_{newO}; \tag{3.13}$$

$$L_{refO} = L_{newO}. \tag{3.14}$$

By Eqs. (3.13) and (3.14), we can combine Eqs. (3.11) and (3.12) as Eq. (3.15) below:

$$\frac{L_{refO}}{D_{refO}} = \frac{L_{newO}}{D_{newO}} = \frac{L_{refI}}{D_{refI}} = \frac{L_{newI}}{D_{newI}}. \tag{3.15}$$

Then, we can rewrite Eq. (3.15) as Eq. (3.16) below:

$$D_{newI} = \frac{L_{newI} \cdot D_{refI}}{L_{refI}} \tag{3.16}$$

Finally, we can use Eq. (3.16) when we compute the viewing distance of the person.

When we compare the data of the new person with the reference data, if the result, C_{same}, is not the same, we have to change the viewing distance. The detailed process of changing the viewing distance is described in the following algorithm. An illustration of the distance between the person and the vehicle for the vehicle seeing the whole clothes is shown in Figure 3.9.

Algorithm 3.3 *Changes of the viewing distance*

Input: The compare result C_{same}, the reference data L_{refI} and D_{refI}, and the length of the clothes of the followed person L_{newI}.

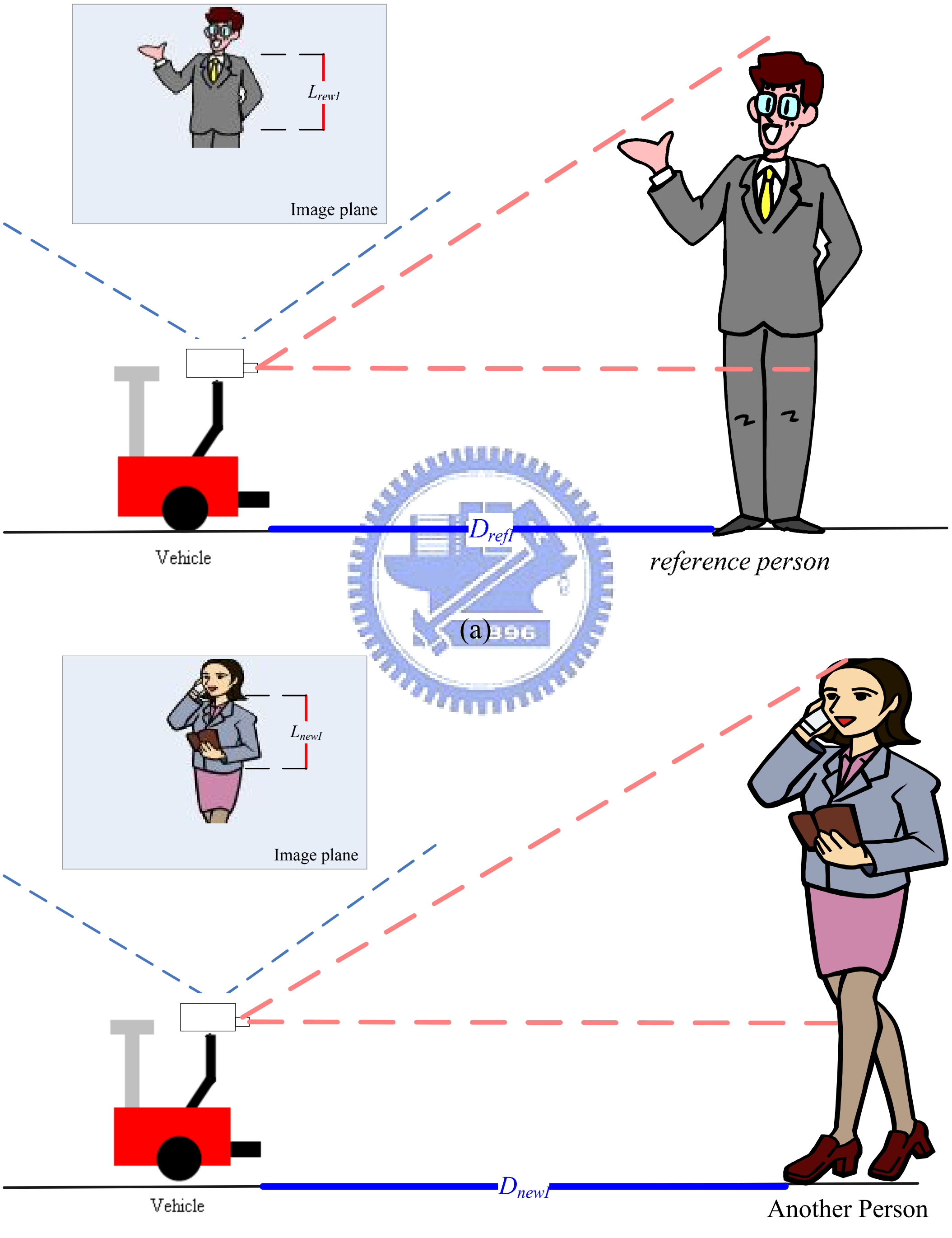
Output: An appropriate viewing distance

Steps:

Step 1. If the compare result C_{same} is true, do not change the viewing distance and exit; otherwise, go to Step 2.

Step 2. Compute the new safe distance S_{dis} by Eq. (3.16).

Step 3. Change the viewing distance with the value S_{dis}.



(a)

(b)

Figure 3.9 (a) The distance D_{refI} for the reference person. (b) The new distance D_{newI} for another person.

3.4 Area Tracking

We can measure the distance between a vehicle and a person by using reference data for different people's heights. And we can control the vehicle to conduct different tasks when the distance between the vehicle and the person is different. For example, when the vehicle is far from the person, the vehicle has to go forward to see the person clearly. If the person continues standing at an identical position, the vehicle will enter the human interaction process for providing people with services. If the person is too close to the vehicle and the vehicle only 'see' some part of his/her clothes, we also hope that the vehicle can keep up with the person. Therefore, we provide an area tracking method so that the vehicle can still follow the person through getting some part of the clothes. And the vehicle will also give a warning for the person to avoid hitting the vehicle.

First, we extract the color of the clothes of the person and take it as a tracking target, which is described in Section 3.4.1. In Section 3.4.2, we will introduce a shape circumscribing method for deciding the area center of the clothes for keeping following the target. Then, we can adjust the camera orientation to keep seeing the followed target in the image. The details will be described in Section 3.4.3.

3.4.1 Human clothes color extraction

Since we already know the values of C_b and C_r of the color of the clothes after the learning process, we can define a region of the color of the clothes in the acquired image with a threshold T_1. If the values of C_b and C_r of a pixel fall into the region which is defined in advance, then this pixel is classified to be of the clothes color. At the beginning, we choose a point which belongs to the region of the clothes as the start point for region growing in a limited square window in the image. By region growing, we can get the location of the clothes. If the biggest region of the clothes is below a threshold T_2, we say that there is no matching clothes in the image; otherwise, we decide that the clothes is detected. After clothes extraction, the vehicle can follow the person by parts of clothes regions in the image sequence. An illustration is shown in Figure 3.10 and the detail of the proposed human clothes color extraction is described in the following as an algorithm.

Algorithm 3.4 *Extraction of human clothes region*

Input: Current image I_c, thresholds T_1 and T_2, the colors of clothes, $Cloth_{Cb}$ and $Cloth_{Cr}$, the clothes region set $R = \{R_{00}, R_{01}, ..., R_{mn}\}$,and a point set P = $\{P_{00}, P_{01}, ..., P_{mn}\}$ which are pixels in the images.

Output: The region of clothes R_{cloth}.

Steps:

Step 1. We assume that (P_{cb}, P_{cr}) are the values of C_b and C_r of P_{ij} and compare (P_{cb}, P_{cr}) with $Cloth_{Cb}$ and $Cloth_{Cr}$ in the following way.

$$(P_{\mathrm{cb}} - Cloth_{Cb}) < T_1; \tag{3.17}$$

$$(P_{cr} - Cloth_{Cr}) < T_1. \tag{3.18}$$

If inequalities (3.17) and (3.18) are satisfied, regard the point P_{ij} to be in the clothes point set C = $\{C_{00}, C_{01}, ..., C_{nn}\}$; else, repeat Step 1 to check the next pixel.

Step 2. Compute the clothes region set $R = \{R_{00}, R_{01}, ..., R_{mn}\}$ which have the color of the clothes by using region growing from clothes point set C with the start point C_{00}.

Step 3. Compute the number of pixels in clothes region set R, find the biggest region and denote it as R_b. If the region R_b is larger than threshold T_2, take it as the region of clothes R_{cloth}.

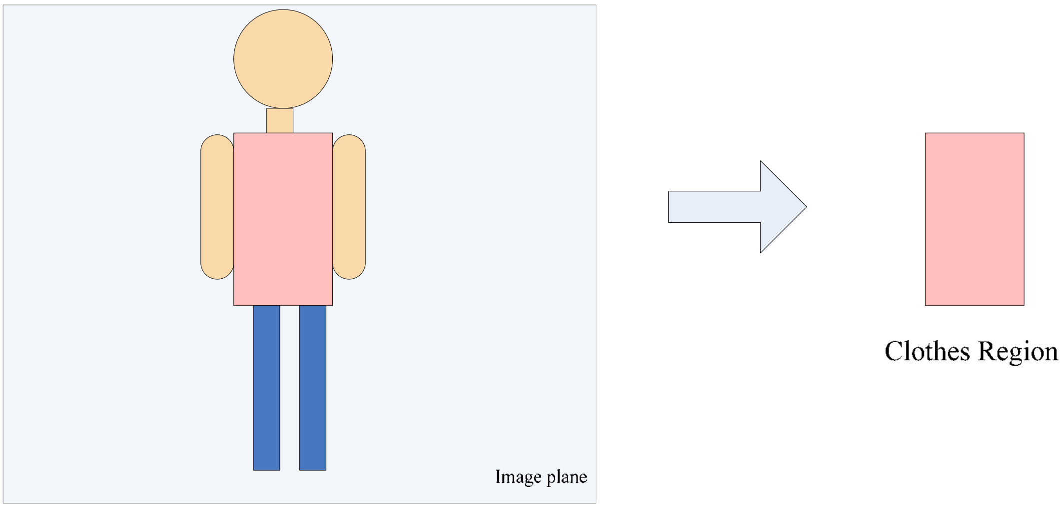


Figure 3.10 Extraction of human clothes region.

3.4.2 Deciding area centers by a shape circumscribing method

After the clothes region R_{cloth} is found, we find it to be of a rectangular shape. If the top side of the rectangle is just the upper boundary of the image, we decide that the vehicle is too close to the person so that only part of the clothes can be seen in the image. Then, we compute the four corner points of the rectangle and find the area center. An illustration is shown in Figure 3.11 and the detail of a shape circumscribing method is described in the following as an algorithm.

Algorithm 3.5 *Shape circumscribing method*

Input: The region of clothes R_{cloth}.

Output: 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 R_{cloth} and the area center $C_a(i,j)$ of the clothes region R_{cloth}.

Steps:

Step 1. Compute the position $P_{TopLeft}(i,j)$ of the top-left pixel in the clothes region R_{cloth}; the position $P_{TopRight}(i,j)$ of the top-right pixel in R_{cloth}; the position $P_{BottomLeft}(i,j)$ of the bottom-left pixel in R_{cloth}; and the position $P_{BottonRight}(i,j)$ of the bottom-right pixel in R_{cloth}.

Step 2. Compute the area center $C_a(i,j)$ by

$$C_a(i) = \frac{P_{TopLeft}(i) + P_{TopRight}(i)}{2}$$

$$C_a(j) = \frac{P_{TopLeft}(j) + P_{BottonLeft}(j)}{2}$$

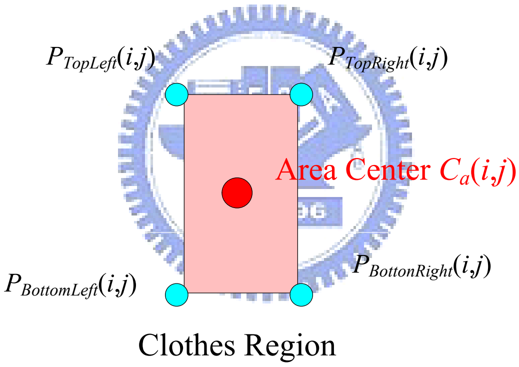


Figure 3.11 Deciding the area center.

3.4.3 Adjustment of Camera orientations for area monitoring

After the area center is found, we can adjust the orientation of the camera for keep following the person. The vehicle may have some quakes because the floor is not flat. We defined a threshold T_3 to decide whether the followed person is to the right or left of the vehicle and whether the person would be out of the image. In the above situation, we have to adjust the orientation of the camera to make the part of the clothes of the person to be in the image center. Therefore, we measure the longitude θ and latitude φ values of the area center $C_a(i,j)$ by the *angular mapping* method which is mentioned in Section 3.2. Because the height of the person does not change in the following process, we only have to adjust the horizontal orientation of the camera with the longitude θ by adjusting the orientation of the vehicle. An illustration is shown in Figure 3.12 and the detail of the adjustment of the camera orientation is described in the following as an algorithm.

Algorithm 3.6 *Adjustment of the camera orientation*

Input: The center $C_{image}(i, j)$ in the image ,the area center $C_a(i, j)$ and threshold T_3

Output: Adjust the orientation of the camera which is held by a mechanical arm.

Steps:

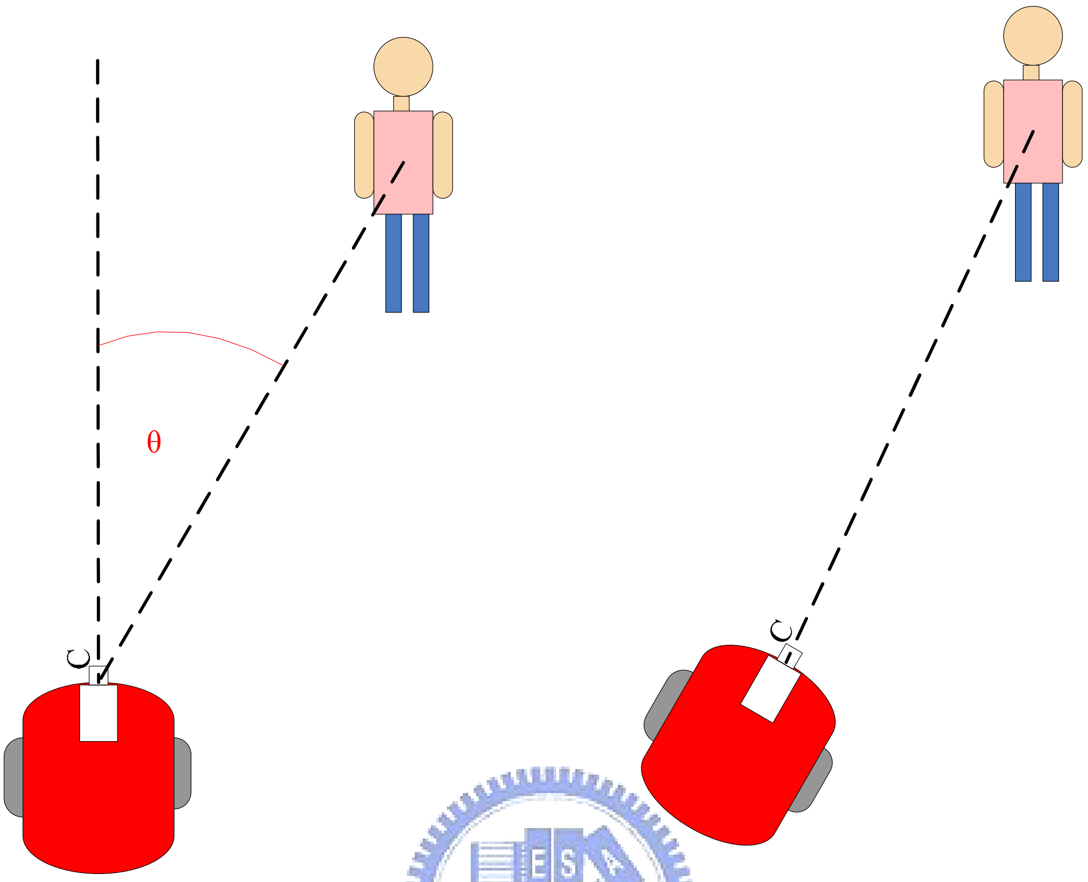
Step 1. Compute the differences of image center $C_{image}(i, j)$ with the area center $C_a(i, j)$ in the following way:

$$(C_{image}(i) - C_a(i)) < T_3; \tag{3.19}$$

$$(C_{image}(j) - C_a(j)) < T_3. \tag{3.20}$$

Step 2. If inequalities (3.19) and (3.20) are satisfied, we do not have to adjust the orientation of the camera ; else, go to Step 3

Step 3. Compute the longitude θ of the $C_a(i, j)$ and adjust the camera orientation with the value θ.



(a) (b)

Figure 3.12 (a) The person is too right to the vehicle. (b) The result of the adjustment.