Chapter 4
Human Detection Technique for Person Following

4.1 Introduction

Before person following, the vehicle has to detect the person and get his/her features which can be used to follow this person. There are many kinds of features which can be adopted to detect humans, such as face, shape, skin color, motion, and so on. Face and skin color are obvious characteristics of humans. Wang and Tsai [3] proposed a method which uses an elliptic skin model to detect human faces by color and shape features in images. In Section 4.3, we will review this method of face detection. But this model is suitable only in a limited range of luminance. We found that the method of detecting skin-colored ellipses is often affected by changes of luminance. In this study, we design a skin-colored model to adjust the parameters of skin-colored ellipses for different light intensities. Therefore, in Section 4.4, we propose an improved skin color model which can adjust its elliptic center to adapt to luminance changes. In addition, we use a progressive method for dealing with none-uniform luminance. In this method, we use blocks each of which consist of a square region of pixels in the image. The sizes of these blocks are all the same and vary from 320×240 to 40×30 until the system detects the human face. Before all the details of the mentioned techniques are described, we will give a brief introduction to the proposed process in Section 4.2 first.

4.2 Proposed Process of Human Detection

Following the method of human detection proposed by Wang and Tsai [3], we use two features, skin color and shape, to detect a human face. And because the followed person is a moving object, we also use motion analysis techniques to conduct motion detection. But we found that the method of detecting skin-colored ellipses is often affected by luminance changes. Therefore, we propose an improved skin-colored model to adjust the parameters of skin-colored ellipses for different light intensities. The detailed process of human detection is described in the following as an algorithm. An illustration of the process is shown in Figure 4.1 as a flowchart.

Algorithm 4.1 *Process of human detection*.

Input: Current image I_c.

Output: The result of a person's detection.

Steps:

Step 1. Apply region segmentation to I_c based on an improved skin-colored model, and motion detection by blockwise frame differencing to extract motion regions.

Step 2. Fit each extracted skin region with an ellipse to detect a possible human face.

Step 3. Apply human body detection by applying shape recognition to extracted motion regions. If detect the person, end the experiment; else go to Step 1.

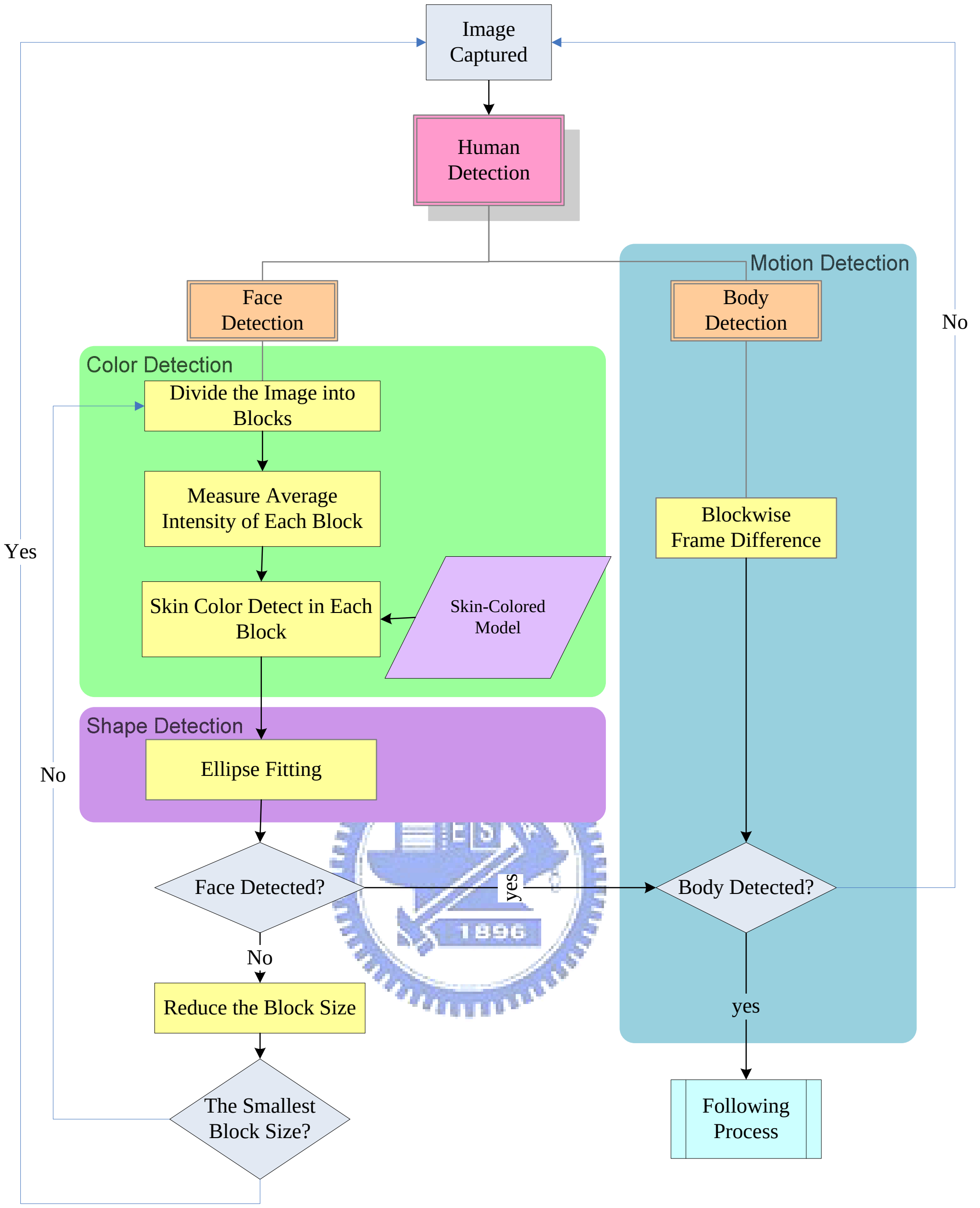


Figure 4.1 An illustration of the human detection process.

4.3 Review of a Face Detection Method

Wang and Tsai [3] proposed a method of human detection by the features of face and human body. *Skin region segmentation by color classification* and *ellipse shape fitting* are used to detect the face of a person. *Shift-tolerant blockwise frame* *differencing* is used to analyze the motion of a person. The method of human detection proposed by Wang and Tsai [3] can be divided into two parts: *human face detection* and *human body detection*.

In the first part, the rough sketch of a face is represented by an elliptical shape with the skin color. Thus, a human face in an image is detected by searching a skin-colored ellipse. After segmenting out skin regions in the image, elliptic shapes are fit to them. If a skin color region is close to an ellipse in shape, it is decided that a face is detected. They chose the YC_bC_r model to be the color space for detecting the skin color in images. The distribution of the skin color in the C_b-C_r plane is found similar to an oblique ellipse. So they define an oblique ellipse in the C_b-C_r plane to be the skin color model and the parameters of the elliptical skin model are adjusted by experiments. The center of the elliptic model is taken to have the values 103 for C_b and 158 for C_r. And the angle of rotation is set to be 145 degrees, and the lengths of major and minor axes are set to be 25.39 and 14.03, respectively. Finally, the range of luminance is taken to be from 113 to 155, as shown in Figure 4.2. Then, they determine if the region is similar to an ellipse by using rectangular and elliptic masks to measure the distribution of pixels inside or outside the ellipse.

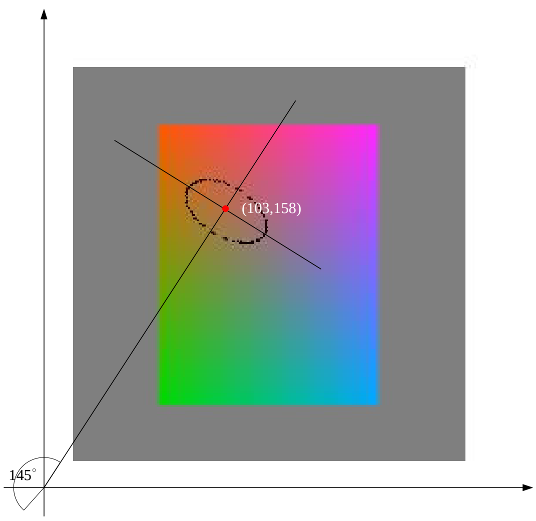


Figure 4.2 The elliptic skin model.

In the second part, they defined some terms in advance. The image captured from the camera in the current navigation cycle is called the *current image*. The image captured at the last moment is called the *reference image*. A searching window is defined to consist of a square region of pixels, whose size is larger than that of an image block. Subtracting the current image from the reference image block by block is the basic idea of blockwise frame differencing. If the difference between the target block in the current image and the candidate block at the same position in the reference image is smaller than a certain threshold T then it is possible that no motion has taken place. If it is not, it is tried to find the best *match block* with the target block within the searching window in the reference image. If the difference between the best match block and the target block is bigger than a threshold T, it is decided that the target block is moving. Then they repeat these steps for each block in the current image and get all the moving parts in the current image.

After the method of blockwise frame differencing described above is performed, the difference image shows the possible moving regions at the current moment. But the regions sometimes do not appear to have a complete shape of a human body. So the ratio of the width to the length of a moving region was used as a feature for human body detection. However, if two sequential images both include the same person but at different positions, the difference of these two images cannot form a complete shape of a human body. If the positions in these two images are close, the result of the difference image will have a thin moving region and if the positions are far away, the difference image will show a thick moving region. Thus, a range of the ratio of the width to the length of a moving region was defined as from 1/8 to 1 to consider the situation of overlapping of body regions in consecutive images.

4.4 Adaptation to Changes of Luminance

In real situations, if we use the feature of color to detect the face, we find that the result is easily influenced by the brightness in the environment. Because the skin color is not the same in different brightness, if we use the same threshold value to detect the skin color, then we cannot detect the face sometimes. So we have to change the threshold values of skin color in different lighting conditions.

In Section 4.4.1, we will describe the details of a skin detection model we adopt which can adapt to luminance changes. After finding out the values of C_b and C_r in every five scales with a light meter which is used to measure the light intensities, we want to apply this model in all intensities. In Section 4.4.2, we will discuss a method of curve fitting to build a smooth curve of the value of the skin color model in different light intensities. Then, we can adjust the center of the skin-colored ellipse in all intensities when the vehicle conducts the task of human detection. In addition, we also want to deal with non-uniform luminance. In Section 4.4.3, we will introduce a progressive method which uses blocks, each consisting of a square region of pixels in the image. The size of these blocks is all the same and the size is changed gradually from 320×240 to 40×30 until the system detects the human face.

4.4.1 A Skin Detection Model

Before building a skin detection model, we prepare a lot of equipment for our experiment. We use a chair, a light meter, a lamp, an IP camera, and a computer. First, a chair is fixed near a wall which is clearly lighted in the environment. Second, we fixed the position of the IP camera at a distance of 1m in front of the chair. Then, we put the lamp behind the IP camera. Third, a person is required to sit down on the chair. Fourth, we turn off all lights in the environment and turn on the lamp. Fifth, we take the light meter near the person's face and check the value of the light meter. Then, we adjust the lamp strength in order to setting the values of the light meter from 0 to 455 in every five scales. An illustration of the experiment is shown in Figure 4.3.

Under the lighting condition of each of a set of five scales, we take an image. Then, we retrieve the skin color of five parts of the face, including the forehead, the nose, the right and left cheeks, and the jaw, as shown in Figure 4.4. And we measure the average of the values of Y, Cb and Cr of these parts. The major steps of the proposed process of building a skin detection model are presented as follows. The resulting values are shown in Figure 4.5. The distribution of the results is shown in Figure 4.6.

Algorithm 4.2 *Building a skin detection model.*

Input: The scale of the intensity, denoted as *Scale*, with the value 5 and the strength of the light $L_{Strength}$ with the initial value 0.

Output: The skin detection model.

Steps:

Step 1. Fix the IP camera at a distance of 1m in front of a person.

Step 2. Put the lamp behind the IP camera.

Step 3. Put the light meter near the face of the person.

Step 4. Check the value of the light meter and adjust the value of light strengths to be $L_{Strength}$.

Step 5. Take an image.

Step 6. Retrieve the size of 16×16 in the five skin parts of the face in the image, which are the forehead, the nose, the right and left cheek, and the jaw.

Step 7. Measure the average values of Y, C_b and C_r in each of the five parts and record the values into a database.

Step 8. If the values of Y in every five parts are 234, end the experiment; otherwise, set the new value of $L_{Strength}$ as $L_{Strength}$+ $Scale$ and go to Step 4.

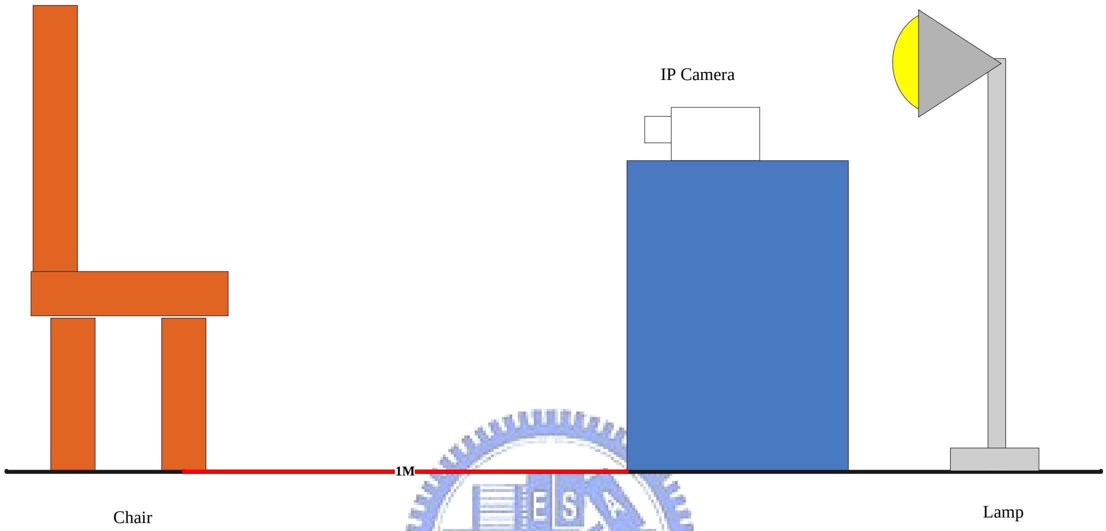


Figure 4.3 An illustration of the positions of the equipment in the experiment.

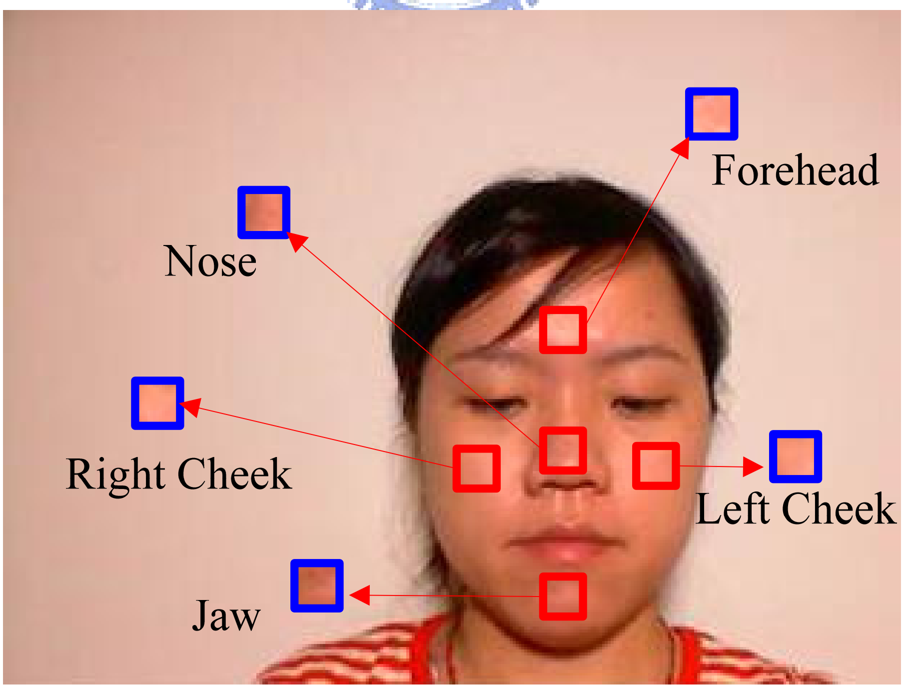


Figure 4.4 Retrieved five skin parts in the face.

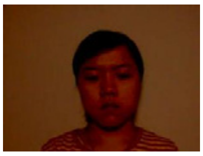
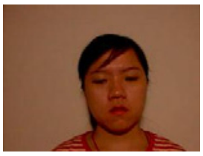
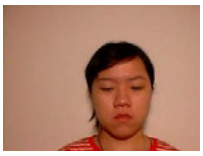
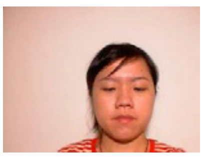
Light meter	Image	Fore-head	Right cheek	Left cheek	Jaw	Nose	Average Y	Average C_b	Average C_r
0							16.00	128.00	128.00
20							48.13	108.70	163.87
40							80.75	100.53	165.00
60							103.49	101.90	166.24
80							118.86	100.33	169.79
100							135.24	100.59	171.81
120							148.31	100.56	173.54
140							165.97	101.89	171.01

Figure 4.5 The results of the experiment of building a skin detection model. (continued)

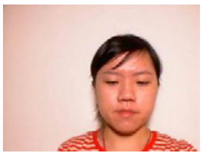
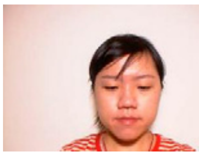
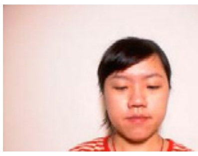
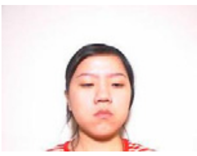
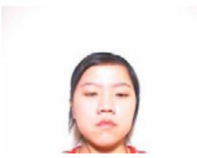
160							173.65	104.83	167.43
180							184.56	106.98	162.37
200							191.01	108.34	158.57
220							197.89	111.10	153.82
240							205.46	115.95	148.60
260							212.50	117.83	143.53
280							217.66	120.13	139.47
300							221.72	122.50	136.51

Figure 4.5 The results of the experiment of building a skin detection model. (continued)

320							225.04	123.89	133.71
340							227.66	125.23	131.91
360							229.69	125.99	130.33
380							231.38	127.00	129.13
400							232.57	127.39	128.48
420							233.37	127.81	127.95
440							233.74	127.79	127.88

Figure 4.5 The results of the experiment of building a skin detection model. (continued)

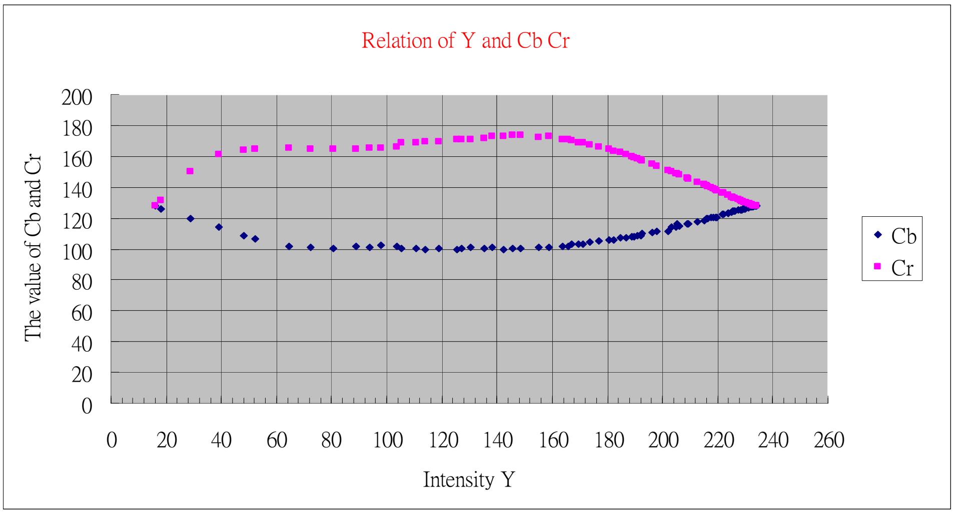


Figure 4.6 The distribution of results of C_b and C_r.

4.4.2 Curve fitting

After measuring the values of C_b and C_r of each set of five scales of the light meter, we can find out that the distribution of data has a trend described as follows. The distribution of C_b is almost a curve of the second order with intensity values from 16 to 80 and from 146 to 234. And from the intensity value of 81 to 145, the distribution of C_b is almost a line. Besides, the distribution of C_r has the similar situation. The distribution of C_r is almost a curve of the second order from the value of intensity 16 to 64 and 160 to 234. And from the value of intensity 65 to 159, the distribution of C_b is almost a line. Therefore, we use a method of least squares error (LSE) fitting to conduct curve fitting. And we then can use the resulting curves for detection of the skin color from the value of intensity 0 to 234. The adopted LSE fitting method of using polynomial curves is explained as follows.

First, we generalize a straight line to a curve of the *k*th-degree polynomial:

$$y = a_0 + a_1 x + \ldots + a_k x^k, \tag{4.1}$$

with the residual given by

$$R^2 \equiv \sum_{i=1}^{n}\left[y_i - \left(a_0 + a_1 x_i + \ldots + a_k x_i{}^k\right)\right]^2 . \qquad (4.2)$$

The partial derivatives are

$$\frac{\partial\left(R^2\right)}{\partial a_0} = -2\sum_{i=1}^{n}\left[y - \left(a_0 + a_1 x + \ldots + a_k x^k\right)\right] = 0\,; \qquad (4.3)$$

$$\frac{\partial\left(R^2\right)}{\partial a_1} = -2\sum_{i=1}^{n}\left[y - \left(a_0 + a_1 x + \ldots + a_k x^k\right)\right]x = 0\,; \qquad (4.4)$$

$$\frac{\partial\left(R^2\right)}{\partial a_k} = -2\sum_{i=1}^{n}\left[y - \left(a_0 + a_1 x + \ldots + a_k x^k\right)\right]x^k = 0\,. \qquad (4.5)$$

These lead to the following equations:

$$a_0 n + a_1 \sum_{i=1}^{n} x_i + \ldots + a_k \sum_{i=1}^{n} x_i^k = \sum_{i=1}^{n} y_i\,; \qquad (4.6)$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + \ldots + a_k \sum_{i=1}^{n} x_i^{k+1} = \sum_{i=1}^{n} x_i y_i\,; \qquad (4.7)$$

$$\cdot\, a_0 \sum_{i=1}^{n} x_i^k + a_1 \sum_{i=1}^{n} x_i^{k+1} + \ldots + a_k \sum_{i=1}^{n} x_i^{2k} = \sum_{i=1}^{n} x_i^k y\,, \qquad (4.8)$$

or, in matrix form,

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \cdots & \sum_{i=1}^{n} x_i^k \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots & \sum_{i=1}^{n} x_i^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_i^k & \sum_{i=1}^{n} x_i^{k+1} & \cdots & \sum_{i=1}^{n} x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \vdots \\ \sum_{i=1}^{n} x_i^k y \end{bmatrix} . \qquad (4.9)$$

This is a Vandermonde matrix. We can also obtain the matrix for a least squares fit by writing:

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} . \qquad (4.10)$$

Pre-multiplying both sides by the transpose of the first matrix then gives

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \tag{4.11}$$

So, we have

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \cdots & \sum_{i=1}^{n} x_i^k \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots & \sum_{i=1}^{n} x_i^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_i^k & \sum_{i=1}^{n} x_i^{k+1} & \cdots & \sum_{i=1}^{n} x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \vdots \\ \sum_{i=1}^{n} x_i^k y \end{bmatrix}. \tag{4.12}$$

As before, given n points and fitting with polynomial coefficients a_0, ..., a_k gives

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix}. \tag{4.13}$$

In matrix notations, the equation for a polynomial fit is given by

$$Y = XA. \tag{4.14}$$

This can be solved by pre-multiplying by the matrix transpose as follows:

$$X^T Y = X^T XA. \tag{4.15}$$

This matrix equation can be solved numerically, or can be inverted directly if it is well formed, to yield the solution vector:

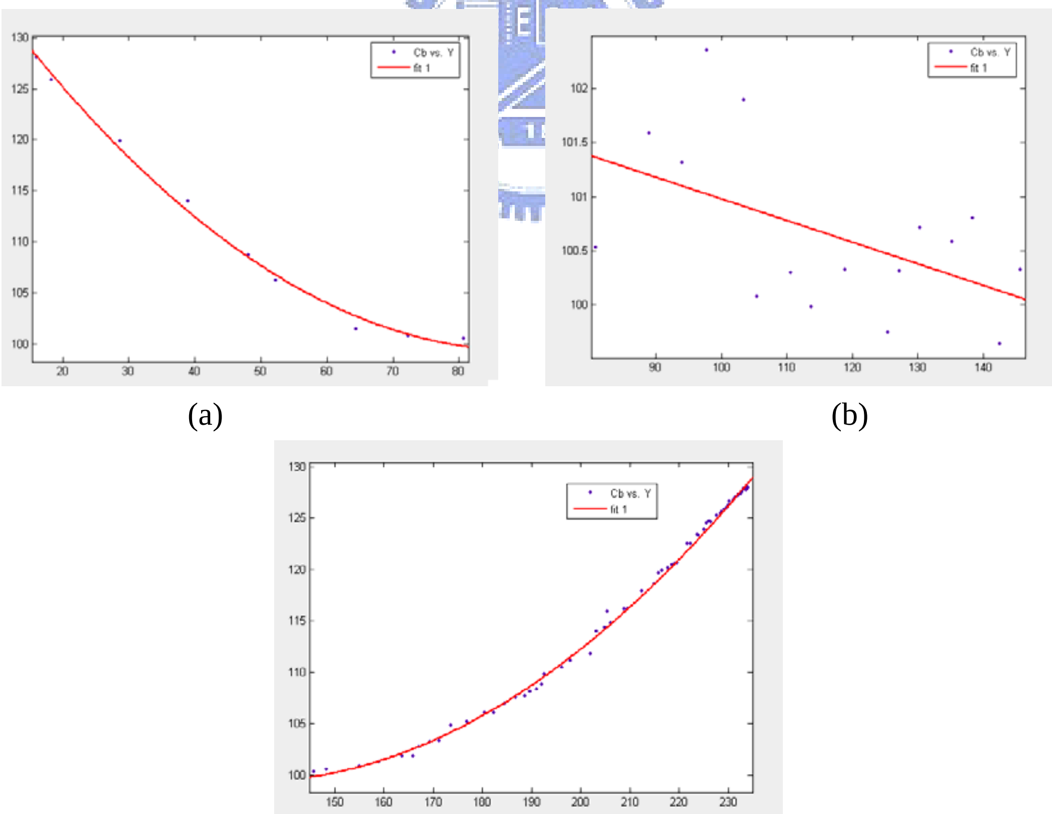
$$A = (X^T X)^{-1} X^T Y. \tag{4.16}$$

The result of curve-fitting of C_b from 16 to 80 is as Eq. (4.17), from 81 to 145 is as Eq. (4.18) and from 146 to 234 is as Eq. (4.19) as below and the results are shown in Figure 4.7 and Figure 4.8.

$$f(x) = 0.005344 \times x^2 - 0.9564x + 142.2 \;; \tag{4.17}$$

$$f(x) = -0.01999x + 103 \;; \tag{4.18}$$

$$f(x) = 0.002797 \times x^2 - 0.7378x + 148 \;. \tag{4.19}$$



(a) (b) (c)

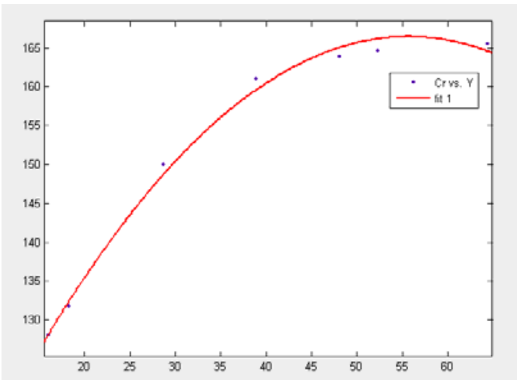
Figure 4.7 The results of line fitting of C_b. The intensity Y is from (a)16~80 (b) 81~145 (c)146~234.

The result of curve-fitting of C_r from 16 to 80 is as Eq. (4.20), from 81 to 145 is as Eq. (4.21), and from 146 to 234 is as Eq. (4.22) below:

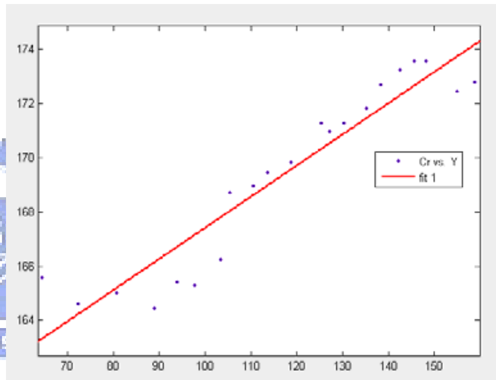
$$f(x) = -0.02437 \times x^2 + 2.711\, x + 91.03; \tag{4.20}$$

$$f(x) = 0.1148\, x + 155.9; \tag{4.21}$$

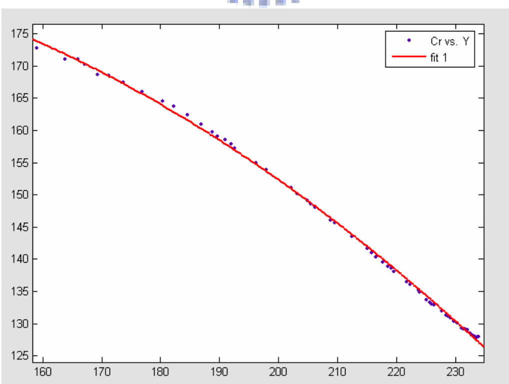
$$f(x) = -0.002908 \times x^2 + 0.5187\, x + 164.9. \tag{4.22}$$



(a)



(b)



(c)

Figure 4.8 The results of line fitting of C_r. The intensity Y is from (a)16~64 (b) 65~159 (c)160~234.

4.4.3 Human detection by a progressive method

The intensity is non-uniform in the real environment. If we use one value of intensity to detect the face of the person in an image, we find out that we will detect other things frequently and consider them erroneously to be the skin color. The intensities of each pixel in the image are not the same. And if we measure each pixel by different intensities which is mentioned in Section 4.4.2, it will be inefficient. Therefore, we use a progressive method for dealing with this situation. In this method, we use blocks which are all the same and the size is from 320×240 to 40×30 until the system detects the human face.

First, we define some terms for use in the proposed method.

(1) *Current image*: The image captured from the camera at the current moment.

(2) *Block*: A block consists of a square region of pixels, which is the unit of the image.

(3) *Intensity set*: A set consists of the value of the average intensity values in each block.

(4) *Center set*: A set consists of the value of the center of elliptic skin model which is mentioned in Section 4.3 in each block.

The detailed process of human detection by a progressive method is described in the following as an algorithm.

Algorithm 4.3 *Human detection by a progressive method*

Input: Current image I_c , threshold T, a block set $B = \{b_{11}, b_{12}, \ldots, b_{mn}\}$, a center set $C_{elliptic} = \{C_{e\,11}, C_{12}, \ldots, C_{mn}\}$ and a skin region set $R = \{R_2, R_2, \ldots, R_n\}$

Output: The detection result of a person's face D_{result}.

Steps:

Step 1. Set the value of W and H as 320 and 240.

Step 2. Set the width and height of each block as W and H.

Step 3. Average the values of intensity of each pixel in each block.

Step 4. Find the values of centers, C_b and C_r, for each block by Eq. (4.17) to Eq.

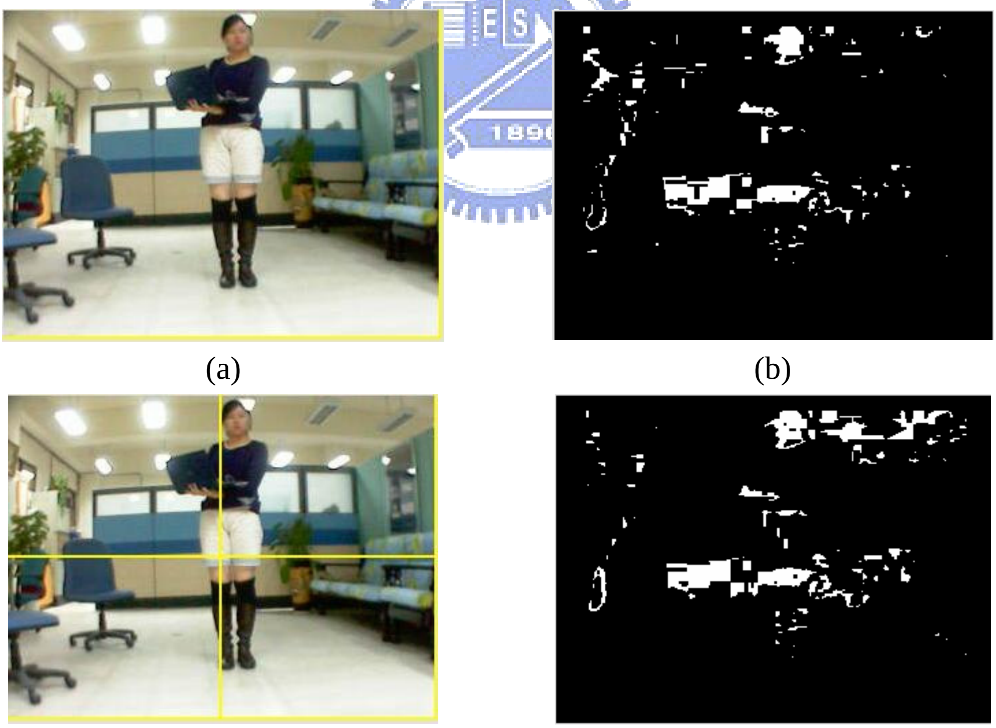
(4.22). Then put the values into the center set $C_{elliptic}$.

Step 5. Use (C_{cb}, C_{cr}) to denote the values of C_b and C_r of $C_{e\,ij}$. Adjust the center of elliptic skin model by C_{cb} and C_{cr}. Detect the skin color in the block b_{ij}. And repeat Step 4 until we detect the skin color in all blocks.

Step 6. Detect the face of the human by a method of *ellipse shape fitting* which is mentioned in Section 4.3. If a face is detected, set the D_{result} as true; else go to Step 7.

Step 7. Divided W and H by 2. If the value of W is smaller than 40 and the value of H is smaller than 30, set the D_{result} as false; else go to Step 2.

By the progressive method as described above, we can detect the face in images, as shown by the example in Figure 4.9 and Figure 4.10.



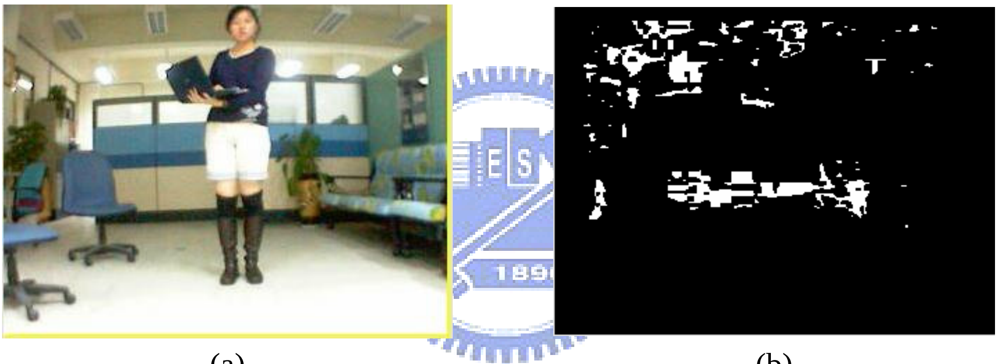
(a) (b)

(c) (d)

Figure 4.9 The detection of human face by the progressive method. (a) The block size is 320×240. (b) Skin color region of (a). (c)The block size is 160×120. (d) Skin color region of (c). (e) The block size is 80×60 and the face of the person is detected successfully. (f) Skin color region of (e).

(e) (f)

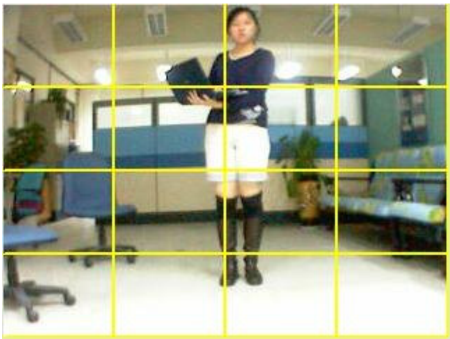
Figure 4.9 The detection of human face by the progressive method. (a) The block size is 320×240. (b) Skin color region of (a). (c)The block size is 160×120. (d) Skin color region of (c). (e) The block size is 80×60 and the face of the person is detected successfully. (f) Skin color region of (e).(continued)



(a) (b)

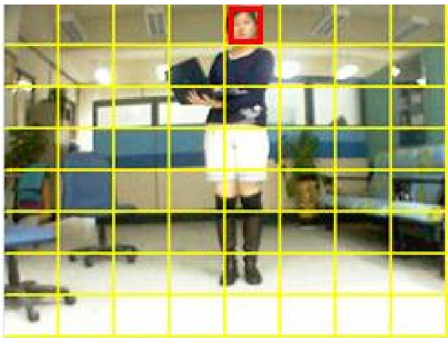
(c) (d)

Figure 4.10 The detection of human face by progressive method. (a) The block size is 320×240. (b) Skin color region of (a). (c)The block size is 160×120. (d) Skin color region of (c). (e) The block size is 80×60 and detect the face of the person. (f) Skin color region of (e). (g) The block size is 40×30 and the face of the person is detected successfully. (h) Skin color region of (g)



(e)

(f)



(g)

(h)

Figure 4.10 The detection of human face by progressive method. (a) The block size is 320×240. (b) Skin color region of (a). (c)The block size is 160×120. (d) Skin color region of (c). (e) The block size is 80×60 and detect the face of the person. (f) Skin color region of (e). (g) The block size is 40×30 and the face of the person is detected successfully. (h) Skin color region of (g). (continued)