South-Central University for Nationalities

Undergraduate Graduation Dissertation/Project

Institute: College of Computer Science

Specialty: Software Engineering **Year of Enrolment:** 2016

Title: Face detection based on skin color model and face geometric features

Student Name: Lu Liang **Student No.:** 201621234012

Instructor Name: Bo Li **Professional Title:** Associate Professor

May 2019

South-Central University for Nationalities
Undergraduate Graduation Dissertation/Project
Statement of Originality

I solemnly declare that the submitted dissertation is completed independently under the supervision of my instructor. The dissertation does not contain the results of any other individual or group that have been published except those are specifically referenced in text. I am fully aware that the legal consequences of this statement are borne by me.

Signature of Author: ________________

Date: ________________

Contents

Face detection based on skin color model and face geometric features

Abstract: Face detection plays a key role in the theory and practical application fields of face recognition, biological detection and identity identification. The purpose of face detection is to detect the position and size of the face if there is a face in the input image. At present, the face detection technology has been applied to many fields, such as image retrieval, digital video processing, access control monitoring systems etc. In this thesis, the algorithm based on the elliptical skin color model and the face geometric feature detection in YCbCr color space is used for face detection. First, the skin color region is extracted by using the skin color model, and then the initial face localization is performed by mathematical morphology processing. Then, the face candidates are screened by geometric features to achieve more accurate face localization. The images used to test the algorithm are divided into four categories, low-density faces, medium-density frontal faces, medium-density side faces and high-density faces in complex backgrounds. The experimental results and data illustrated in this paper show that the algorithm used in this paper has a low miss detection rate and can achieve more accurate face detection.

Key words：Face detection; color space; geometric features; eyes detection.

1 Introduction

1.1 Research Background and significance of research

With the advancement of society, the development of computer technology and the internet, security of information technology has become more and more important. Human face is a common and complicated visual mode. It is the most important and direct information carrier for human emotion expression and communication. The face contains a lot of valuable information, such as skin color, outline, expression and other information. People's interactions have a very important role and significance, and they are a potentially friendly human-computer interaction interface.

Identity authentication has always been a hot issue of concern. Traditional identity authentication is mainly based on the identification logo items and logos, and there are shortcomings such as loss, abandon or being forged. Biometric authentication is based on the individual's unique physiological characteristics and behavioral characteristics. Compared with traditional identity authentication technology, biometric authentication does not have the problem of losing and forgetting, and 'carry with you'. Compared to the identification of biometrics, it is necessary to rely on the cooperation of participants. and collect some of the organ characteristics of the participants, including the iris, retina or fingerprints. However, based on personal identification of frontal faces or side faces, face image can be obtained using direct natural methods rather than contact, non-mandatory methods. This method is convenient, fast and friendly, and does not violate the privacy of participants. So, this method is widely accepted.

The face recognition [1] technology includes a series of related techniques such as face detection [2] in an image or video, positioning of an eye position from a detected face, extraction of a face feature, and finally face matching. The results of face detection will have an important impact on the subsequent work, that is, the basis of face detection. Therefore, research on face detection has important research significance and use value.

1.2 Research status of face detection at home and abroad

The face detection problem originally comes from face recognition. The research of automatic face recognition technology began in the 1970s, starting with the first face recognition system developed by Kelly [3] and Canack [4]. With the progress of computer technology and e-commerce technology, the urgent need for face detection and face recognition system for production and life has attracted a lot of attention from research institutions and related personnel. The face detection and face recognition technology are successfully applied in the field of pattern recognition and image analysis processing.

1.2.1 Foreign research status

Foreign research on face detection and recognition have emerged in an endless stream. Many countries have conducted research on face detection and face recognition [5]. These studies have been highly valued and funded by the military, police and commercial companies. Well-known research institutions [6] include the Media Lab of the Massachusetts Institute of Technology in the United States, the Robotics Institute of Carnegie Mellon University, the Center for Speech Vision and Signal Processing at the University of Surrey, UK, the INRIA Institute of France, and the ART Experiment in Japan and so on. With the deepening and development of face detection and recognition technology, the number of theses on face detection and face recognition has also increased substantially in the world. Some theses with high technology content are mainly included in some important international conferences. The famous international conferences include computer vision and pattern recognition conferences, international conferences on image processing and international conferences on automatic recognition of face and gestures. The publication of these theses further promotes the development of face detection and face recognition technology.

1.2.2 Research state in China

Domestic research [7] on face detection and face recognition mainly began in the 1980s. Although it started late, it has developed rapidly. At present, there are many units in China that have researched and discussed the fields of face detection and face recognition. These units mainly include Tsinghua University, Shanghai Jiaotong University, and Institute of Automation of Chinese Academy of Sciences.

Researchers at Tsinghua University conducted a detailed systematic analysis and statistics on the skin color and human beings around the world, and proposed a face detection algorithm based on color and face features. A face matching based on multi-associated template was also proposed. The detection effect is good but the computational sufficiency is not high. Therefore, the face detection speed does not meet the requirements of real-time detection of human face. The 'Authenmatric' face recognition system

[8] and intelligent video surveillance system developed by the Institute of Automation of the Chinese Academy of Sciences is the first face detection and face recognition system with a medium distance greater than 5 meters. The system can track multiple faces quickly, accurately and stably at medium and long distances, and can keep track in the side face and back state. The algorithms have reached the international advanced level. The system has been implemented and played a role in several major domestic security departments, including the Beijing Olympics and border inspections.

1.3 Overview of face detection methods

1.3.1 Face detection based on geometric features

The face detection method based on geometric features [9] is to detect the position of different facial features and then locate the face according to the spatial geometric relationship between them. This method can be further divided into knowledge-based and local feature-based detection strategies. The knowledge-based method first locates the candidate face region, and then detects the presence of the face through the prior knowledge of the face. In the local feature-based method, local features of the face such as the eyes, nose, and lips are first examined, and then these partial combinations from a human face.

1.3.2 Face detection based on template matching

The template matching method [10] is mainly based on constructing a template of an organ of a face or a whole face, and combining the adjustment of the template parameters to calculate the difference or correlation between the target template (such as face contour, eyes, nose and mouth) and the candidate region. Through the template search and matching algorithm, the purpose of detection and positioning is achieved. This method is simple and intuitive.

This method has achieved good results in the application of a single face for the front. However, for face detection of complex backgrounds. Establishing a template that distinguishes background interference well and represents various differences in faces is the key point of the problem. In fact, templates with good distinctions of background interference are difficult to represent various face templates, and templates that can represent various faces cannot distinguish the background well. Therefore, for the method based on template matching, the accuracy rate and the missed detection rate are always contradictory.

1.3.3 Face detection based on statistics

The statistical model-based method [11] is one of the popular methods and an effective way to solve complex face detection problems. It has the following advantages: ① It does not rely on prior knowledge and parameter models, so that errors caused by inaccurate or incomplete knowledge can be effectively avoided; ② The method of instance learning is used to acquire the model. The parameters are statistically more reliable; ③ By increasing the number of learning examples, the range of detection modes can be extended, and the robustness of the detection system can be improved.

However, since this method requires an exhaustive search [12] for all possible detection windows, the computational complexity is increased, so the speed of detection needs to be improved. Such methods are mainly for frontal face detection. For rotating faces and multi-pose face detection, it is more complicated and difficult, and there are not many effective methods.

1.3.4 Face detection based on skin color model

In color images, skin color is a very distinctive feature that distinguishes between face and non-face. Skin color detection does not depend on some details of face, but also for face rotation, posture and expression changes. At the same time, it has strong robustness and stability. Therefore, it is a very natural idea to use the skin color information of the face to detect the face in the color image [13]. The method is to distinguish the face from the image background by the skin color information by utilizing the type of the skin color of the face and the difference of the background color.

Although there are many advantages in the detection of the skin color model, since the skin color detection is difficult to identify the face area of similar skin color, and the simple skin color detection cannot process the gray image, many researchers have proposed some improvements to the skin color detection. For example, face detection based on skin color, knowledge and 2DPCA combined [14], the algorithm is robust to image size changes, and the detection speed is faster, and it can also realize frontal multi-face from the complex background. Face detection based on skin color and template [15], this method combines skin color information of color image with template matching of gray image and artificial neural network classification model, which greatly improves the speed and has more strong robustness. Based on skin color segmentation and statistical analysis combined with template matching method [16], it changes the point-by-point matching method in the traditional template matching method, and further reduces the search area by statistical analysis of the segmented skin color region, greatly speed up the detection speed.

It is inaccurate to use only the skin color model to extract skin color and then to locate the face. Some skin non-human face areas, such as the neck, arms and legs, cannot be excluded. In order to more accurately locate the face region, this thesis uses the geometric features of the face to exclude the non-face region. This part will be described in detail in Chapters 3 and Chapter 4 of this thesis.

1.4 Face detection application

With the development of hardware technologies such as computer image acquisition and the deepening of various related research, the field of face detection applications has expanded to many aspects. E.g.:

(1) In the field of human-computer interaction, the computer extracts information about the user's identity, status, and intention from the image, and then responds accordingly. It allows the computer to know who you are and to serve you according to your status.

(2) In the network image transmission, the face area is given a higher priority to ensure its transmission quality, and the background part is transmitted with a lower priority to reduce the network transmission load, so that under the same bandwidth get a better-quality reconstructed image.

(3) In the content-based retrieval, the user gives an image containing the face, and using the face detection technology, a series of related images can be quickly found in the database.

(4) Identification. Face identification technology is one of the biometric identification technologies.

It has broad development and high socioeconomic benefits in the fields of social insurance, finance, customs, public security, medical care and other civil security control systems with its characteristics of directness, uniqueness and convenience.

(5) Intelligent monitoring. In traditional video surveillance systems, workers need to constantly monitor the screen to detect suspicious events and targets, so the workload is large and inefficient. In the intelligent monitoring system using the face detection technology, the computer suspiciously automatically notifies the staff when segmenting and detecting the target person, thereby reducing the workload of the staff and improving work efficiency.

1.5 Organizational structure of this thesis

This thesis is divided into six chapters, the content organization structure is as follows:

The first chapter is the introduction part, which focuses on the research significance, the research status at home and abroad and the existing face detection algorithm.

The second chapter proposes the algorithm of skin color extraction. This paper uses the face detection algorithm based on elliptical skin color model. The algorithm is implemented in Chapter 2 and the preliminary binary image after skin color segmentation is obtained.

The third chapter introduces the concept of mathematical morphology, and uses the mathematical morphology to filter the obtained binary images to eliminate the interference of non-skin areas. Then use three geometric concepts to further eliminate the influence of non-face regions, and prepare for the face geometry detection in the next chapter.

The fourth chapter is mainly about the detection of the eyes, and further accurate localization of the face by realizing the localization of the eye.

The fifth chapter is the display of the algorithm results and data analysis, but also compared other face detection algorithms.

The sixth chapter is a summary of the main work of the whole thesis and a prospect for the next step.

2 Skin color extraction

2.1 Introduction

The color has more information than the gray value, and the skin color is one of the most prominent features on the surface of the human body. The skin color does not depend on the details of the face, and can be applied to changes such as rotation and expression. The skin color is different from the color of most background objects and has relative stability. Therefore, skin color features are the most commonly used feature in face detection. Skin color features are primarily described by skin color models. The form of skin color model used is closely related to the choice of color space. Face detection based on skin color, gesture recognition, face tracking, etc. in a simple environment can be directly realized by skin color segmentation. The use of human skin color features in a complex environment can greatly reduce the range of search space and improve the detection speed.

2.2 Color spaces overview

Color space [17] is a method of defining, creating and observing colors. There are many different color spaces for color research and application, such as RGB color space, HIS color space, YIQ color space, YCgCr color space, YCbCr color space and so on.

2.2.1 RGB color space

The RGB model is a commonly used color information expression. Most of the colors seen in real life are mixed colors. The principle of RGB space is also derived from this. Under most conditions, images are described in RGB color space. It uses the brightness of the three primary colors red, green, and blue to quantify the color. This model, also known as the additive color mixing model, is a method of superimposing RGB three-color light to achieve color mixing. Therefore, it is suitable for display of illuminants such as displayers, and most of CRT monitors such as computers and televisions use this model.

The RGB color space can be illustrated by the unit cube defined by the RGB Cartesian coordinates shown in Figure 2.1. The coordinate system origin (0,0,0) represents black, the coordinate point (1,1,1) represents white, and the two vertices on the coordinate axis represent two primary colors. Therefore, the color space is a two-dimensional linear space, and any color light having a certain brightness can be represented by a point or a vector of space.

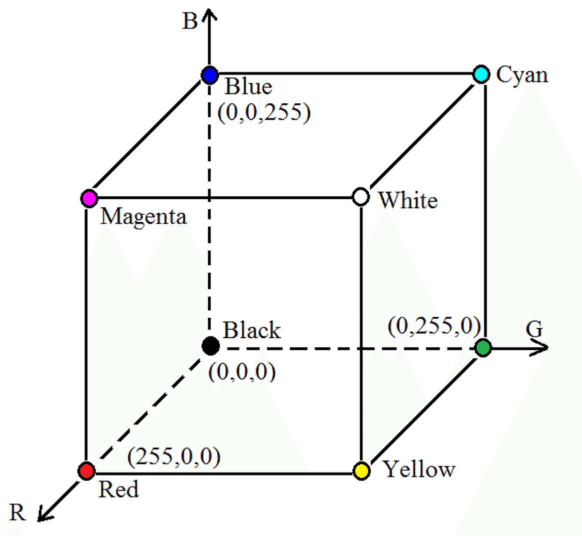


Fig. 2-1 RGB color space

The RGB color space does not have good clustering properties for skin tones. The R, G, and B values of faces shot under different lighting conditions are distributed over a wide range. Therefore, the RGB color space cannot be directly used to establish the skin color model, but the transformed color space from RGB color space has been widely used in face detection.

2.2.2 Normalized RGB color space

In the RGB color space, the three-dimensional vectors R, G, and B not only represent colors but also brightness information. If the two-pixel points [r1, g1, b1] and [r2, g2, b2] are proportional in the color space, i.e. $\frac{r1}{r2} = \frac{g1}{g2} = \frac{b1}{b2}$, then these two points have the same color, different brightness. For characterizing skin tones, the three components of RGB are not required. We can remove the luminance component in the color space and keep the "pure" color component, that is, the chrominance space (r, g) can fully represent the color, thereby obtaining a normalized space.

The normalized space definition is as shown formula 2-1.

$$\begin{cases} r = \frac{R}{R+G+B} \\ g = \frac{G}{R+G+B} \\ b = \frac{B}{R+G+B} \end{cases} \tag{2-1}$$

2.2.3 HSI Color Space

The HSI color space is a simplified from of the Munsell [18] color space, which uses color tone and color saturation from the human visual system to describe color. The HSI color space is consistent with human perception of color and is a color space suitable for human visual characteristics. In the use of color information, the advantage of this format is that it separates the intensity (I) from the two parameters reflecting the intrinsic properties of color - hue (H) and saturation (S). The three attributes of the HSI model define a three-dimensional cone space, as shown in Figure 2-2. The grayscale shadow gradually increases in brightness from top to bottom along the axis, from the black at the bottom to the white at the top. The color on the circumference of the top of the cone has the highest brightness and maximum saturation.

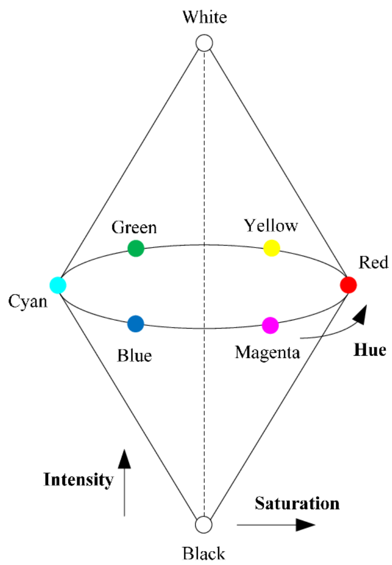


Fig. 2-2 HSI color space

The HSI color space and the RGB color space are different representations of the same physical quantity, so there is a conversion relationship between them, as shown in the equation 2-2.

$$\begin{cases} I = \frac{R+G+B}{3}; \\ S = 1 - 3 * \frac{Min(R,G,B)}{R+G+B}; \\ if\ B \leq G,\ H = arccos\left\{(R-G) + \frac{R-B}{2*\sqrt{(R-G)^2+(R-B)*(G-B)}}\right\}; \\ \qquad else,\ H = 2\pi - \left\{\frac{R-B}{2*\sqrt{(R-G)^2+(R-B)*(G-B)}}\right\}; \end{cases} \tag{2-2}$$

The HSI color space also greatly simplifies the calculated amount of image analysis and processing. Since the HSI color space is closer to human understanding and interpretation of color, the use of this color space can reduce the complexity of color image processing and increase the processing speed. Arithmetic operations or algorithms commonly used in image processing, such as edge detection or edge enhancement, can achieve good results by operating the luminance signal in the HSI color space, and it is inconvenient to do the above processing in the RGB color space. A large number of algorithms in image processing and computer vision are conveniently used in the HSI color space, which can be processed separately and independently of each other.

2.2.4 YIQ color space

Since the RGB three primary color system of the standard color television broadcasting system is subject to various conditions, the National Television Standards Committee has developed a new color space, the YIQ model. North American TV systems typically use the YIQ color space, and the YIQ color space belongs to the NTSC (National Television Standards Committee) system. Y is the luminance of the color, that is, the brightness, which is the gray value of the image we usually call, and I and Q are the chrominance, which is the properties to descript the color and saturation of image. The I component represents the color change from orange to cyan, while the Q component represents the color change from purple to yellow-green. By converting a color image from RGB space to a YIQ color space, the chrominance information in the color image can be separated from the luminance information so that the color component is not affected when processing the luminance component of the image.

This color space is derived from the RGB spatial mapping transformation, which is expressed by the following equation 2-3:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & -0.311 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2-3}$$

2.2.5 YCbCr color space

The YCbCr color representation model has the merit of separating the luminance components in

HSI, and is mainly used in digital television systems, and is the world standard for digital video signals. In the YCbCr color space, the Y component represents luminance, and the Cb, Cr components represent chrominance. The conversion between RGB space and YCbCr space is shown in equation 2-4.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.219 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{2-4}$$

2.2.6 YCgCr color space

Someone has proposed a new YCgCr color space. Also, since in the color space, the color is defined as three components of red, green, and blue, that is, each color corresponds to the three primary colors, that is, the red component Cr, the blue component Cb, and the green component Cg. Any two colors components and brightness information can form a three-dimensional space.

The YCgCr color space and RGB color space conversion formula 2-5 are as follows:

$$\begin{bmatrix} Y \\ Cg \\ Cr \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.318 & 0.439 & -0.121 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{2-5}$$

2.3 Comparison of color spaces

2.3.1 Principle of choosing color space

In the color face segmentation algorithm, in order to separate the face region from the complex background, it is necessary to use a reliable skin color model suitable for different lighting conditions. Studies have shown that although different nationalities, different ages, and different genders look different in color, this difference is mainly concentrated on brightness. If there is no brightness effect, only the skin color distribution is considered, then the skin color of different faces distribution is considered, and is concentrated in a small area, that is, the skin color distribution characteristics have certain clustering. The establishment of the skin color model is closely related to the color space selected. In general, the principle of color space selection is:

① Skin color has good clustering characteristics in this color space.

② The skin color model can fully distinguish between “skin tone” and “non-skin tone”, that is, in this color space, the overlap between “skin tone” and “non-skin tone” is as small as possible.

Most image capture devices such as color video recorders and digital cameras use a color model to represent colors, while in color space representation, chrominance information and luminance information are mixed. Due to the changes in ambient lighting and the influence of brightness parameters, the detection of human faces may become more complicated, and the result of skin color segmentation is unreliable, so RGB color space is not an ideal choice. In order to fully utilize the clustering characteristics of the skin color in the chromaticity space, it is necessary to separate the chrominance information from the luminance information in the color expression, so the color space must be converted to achieve the purpose.

2.3.2 Color space conversion result

The images below show the results of converting an image containing faces from RGB color space to HSI color space, YIQ color space and YCbCr color space.


(a)


(b)


(c)


(d)

Fig.2-3 Color space conversion result
(a) Original image
(b) Convert to HSI color space
(c) Convert to YIQ color space
(d) Convert to YCbCr color space

2.3.3 Advantages of YCbCr color space

Compared to other color spaces, the YCbCr color space has the following advantages [19]:

① The YCbCr color format is similar to the human visual perception process.

② The YCbCr color model is widely used in fields such as television display, and is often used in many video compression coding standards such as MPEG and JPEG.

③ The YCbCr color space separates the luminance components in the color space well.

④ The calculation process and spatial coordinate representation of the YCbCr color format are relatively simple, and the conversion relationship with RGB is simple, avoiding the singularity of the nonlinear space.

⑤ The experimental results show that the clustering of skin color in the YCbCr color space is better. The result of skin color distribution in YCbCr color space is shown in the figure 2-4.

Therefore, in this thesis I will perform face detection based on the YCbCr color space.

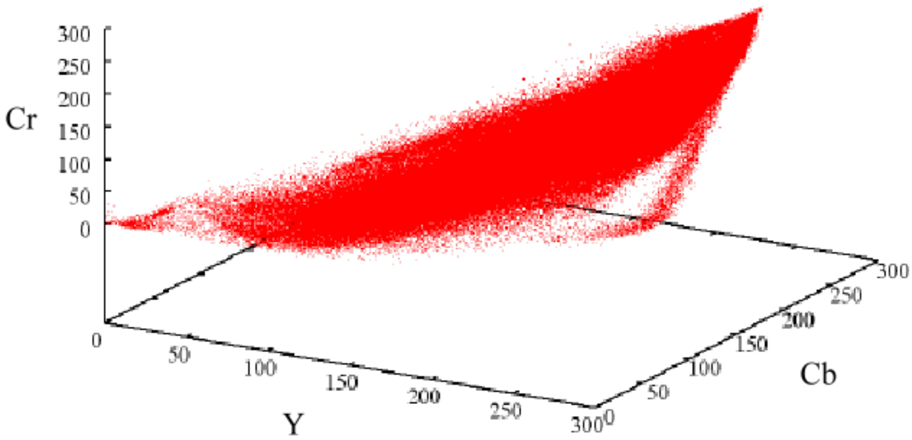


Fig.2-4 The skin color distribution in YCbCr color space.

2.4 Skin color model overview

2.4.1 Simple skin color model

The simple skin color model is the earliest and simplest skin color segmentation model. This model refers to the distribution of a large number of skin color samples in a certain space, and the distribution value range is obtained according to the distribution, and the skin color is judged by using the range. For example, a large number of skin color sample images are converted from RGB to YCbCr color space, and the distribution of skin color on Cb and Cr components is counted. By observing the distribution on the Cb and Cr components, the values of skin color in Cb and Cr are obtained. The distribution range is then used to segment the image to be detected. Pixels whose Cb and Cr values of the image to be detected fall within this distribution range are considered to be skin color points, and other pixel points are considered to be non-skin color points. A binary image of skin color segmentation is obtained by calculation to determine a skin color region.

The advantage of the simple skin color model is that the amount of calculation is simple and easy to understand. However, this model also has a major drawback, that is, using a simple threshold to determine whether it is a skin color point, it is easy to judge the non-skin color point as a skin color point, and also to determine the skin color point as a non-skin color point, thereby causing the skin color area to be detected inaccurate. It is only a matter of judging whether a pixel is a skin color point or a non-skin color point, and cannot give a probability that the point belongs to a skin color point. This model is too rigid and not flexible enough.

2.4.2 Histogram model

The histogram skin color model is a non-parametric model. The histogram statistics of skin color samples can construct a skin color probability map, that is, assign a probability value to each pixel in the discretized color space. There are two main methods used to detect skin tone pixels, the regularized lookup table and the Bayesian classifier. For a three-dimensional histogram, although the effect is good, a lot of training samples are needed and the training time is too long.

2.4.3 Gaussian distribution model

The Gaussian distribution model is one of the ten commonly used mathematical statistics-based models in pattern recognition. It describes the normal distribution of samples in space by Gaussian probability density function. The difference in skin color between people is mainly reflected in the brightness rather than the chromaticity. At the same time, the skin color of the two-dimensional chromatic space that ignores the brightness information is subject to a normal distribution.

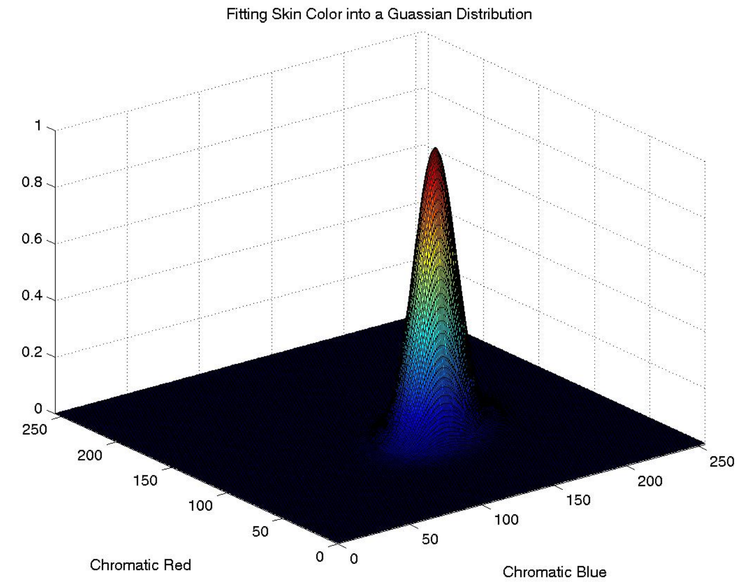


Fig.2-5 Fitting Skin color into a Gaussian Distribution

It can be seen from Fig. 2-5 that the skin color clustering characteristics in the YCbCr color space are better. And in the YCbCr color space, the skin color space has a two-dimensional Gaussian distribution. It can be assumed that the skin color distribution is represented by a Gaussian model N (m, C). X = $(Cb, Cr)^T$ represents a two-dimensional color vector, and m, C represent the mean vector and covariance matrix of the parameter, respectively. Then there are formula 2-6 and formula 2-7:

$$m = \frac{1}{n}\sum_{k=1}^{n} \bar{x}_k \tag{2-6}$$

$$C = \frac{1}{n}\sum_{k=1}^{n}(\bar{x}_k - m)(\bar{x}_k - m)^T \tag{2-7}$$

Where n is the total number of skin color samples, m is the mean of the color vectors, and C is the covariance. Calculate skin color similarity using the following formula 2-8:

$$P(Cb, Cr) = \exp[-0.5(x - m)^T C^{-1}(x - m)] \tag{2-8}$$

Combining relevant data and literature [20], the values of m and C are as follows:

$$m = [117.4316, 156.5599] \tag{2-9}$$

$$C = \begin{bmatrix} 160.1301 & 12.1430 \\ 12.1430 & 299.4574 \end{bmatrix} \tag{2-10}$$

2.4.4 Elliptical skin color model

The ellipse skin color model is proposed by Anil K. Jain [21] in the YCbCr color space. They manually selected about 850,000 skin color pixels from multiple images in the self-built image library, and plotted these skin color points in the YCbCr space and the two-dimensional projection subspace. The results are as shown in the figure 2-6:

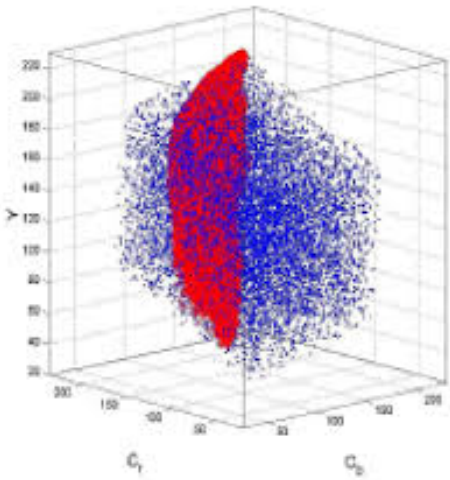


Fig.2-6 Skin color in YCbCr space and CbCr sub-plane by the Anil K.Jain experimental

As can be seen from the figures, in the YCbCr color space, the shape of the skin color cluster is a two-pointed shape, and the shape is distributed like an ellipse in the CbCr space, so Anil K. Jain uses an ellipse to approximate the skin color region. However, in the smaller and larger part of Y, the skin color clustering area will be relatively reduced. It can be seen that where the Y values are different, the distribution area is projected onto the Cb-Cr sub-plane, and the range obtained is different. According to the above, simply excluding the influence of the Y component, it is inaccurate to find the clustering region of the skin color in the two-dimensional sub-plane. We must consider the influence of the different

Y values, so that the YCbCr color space is nonlinearly divided. The skin color is tightly packed in the transformed Cb'Cr' space, and an ellipse can be used to approximate the skin color region. Its analytical expression is formula 2-11:

$$\frac{(x-ecx)^2}{a^2}+\frac{(y-ecy)^2}{b^2}=1 \tag{2-11}$$

among them,

$$\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} Cb' & -cx \\ Cr' & -cy \end{bmatrix} \tag{2-12}$$

The constants in the analytic formula 2-11 are: cx = 109.38; cy = 152.02; θ = 2.53; ecx = 1.60; ecy = 2.41; a = 25.39; b = 14.03.

Substituting the Cb and Cr values of the pixel into the left half of the analytical expression 2-11 finds its value. If the value is greater than 1, i.e. outside the ellipse, it can be determined that the point is not a skin color point; if the value is less than 1, the point is a skin color point.

2.5 Skin color model used in this thesis

The effect of skin color segmentation depends on the choice of color space and skin color model. Different color spaces, the same skin color model, the effect of skin color segmentation is not the same. In the same color space, different skin color models are selected, and the effect of skin color segmentation is different. Therefore, choosing the right color space and the appropriate skin color model is crucial for the skin segmentation algorithm. For the selection of color space and skin color models, the following principles can be considered.

① The distribution of human skin color in this air space should have good clustering characteristics;

② Have a good skin tone model to represent the skin color part;

③ There is less overlap between the skin color and the non-skin color in this color space, which is a good way to separate the skin color from the non-skin color area.

2.5.1 Binary image

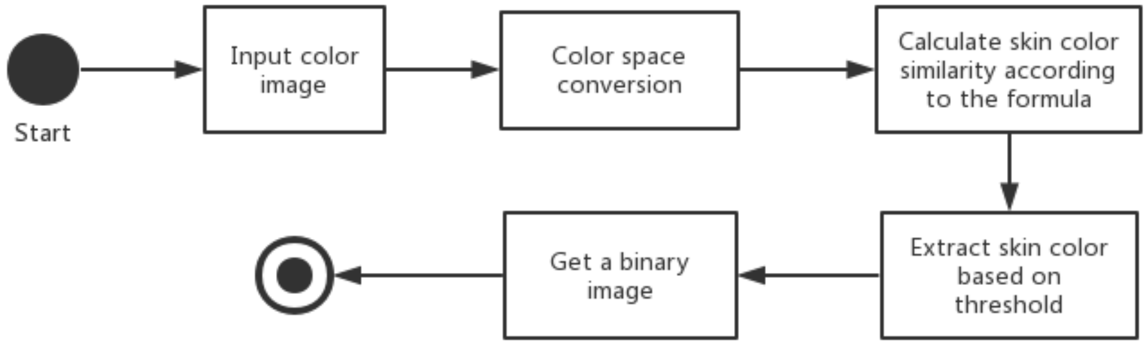


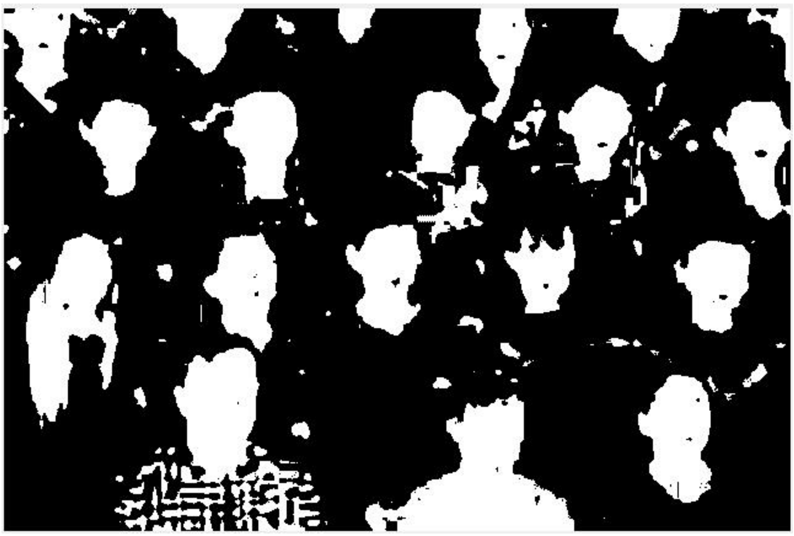
Fig.2-7 The process of extracting skin color

The process of skin color extraction is as shown in the figure2-7. After the color image to be detected is input and transformed to the desired color space, the skin color similarity is calculated, and the binary image after the skin color is extracted can be obtained.

2.5.2 The results of using different color spaces and skin color model

The following images (figure2-8(b) and figure2-8(c)) show the skin color extraction results of figure2-8 (a) based on the Gaussian distribution model and the ellipse model in the YCgCr color space. Figure 2-8(d) and figure 2-8(e) show the results of skinning extraction based on Gaussian distribution model and ellipse model in figure 2-8(a) in YCbCr color space.

(a)



(b)

(c)

(d)

(e)

Fig.2-8 Conversion effects of different skin color models in different color spaces

(a) Original image

(b) Skin color extraction results based on Gaussian distribution model in YCgCr color space

(c) Skin color extraction results based on ellipse model in YCgCr color space

(d) Skin color extraction results based on Gaussian distribution model in YCbCr color space

(e) Skin color extraction results based on ellipse model in YCbCr color space

From the results obtained above, it can be clearly obtained that in the case of the same skin color sample using the same skin color model, the effect of extracting skin color in the YCbCr color space is better than that in the YCgCr color space. At the same time, in the case of the same skin color sample, converted into the same color space, the elliptical model has better skin color extraction effect than the Gaussian distribution model. The ellipse model can better separate the non-skin area and the skin color area of the face while extracting the facial skin color.

2.6 Chapter summary

This chapter introduces the theory and knowledge involved in skin color detection, including color space and color spatial transformation, and the establishment of skin color models. The clustering characteristics of YCbCr color space make it more suitable for building skin color models. Therefore, the YCbCr space is used in this thesis. The Gaussian distribution model and the ellipse model are used to successfully extract the skin color. The binary image after skin color extraction and good experimental results are obtained.

3 Initial localization of the face region

3.1 Processing images using mathematical morphology

There are two cases where a binary image is obtained after skin color segmentation. One is the isolated point or small isolated area of the skin color that is mistaken for the introduction of background and noise. The other is the incomplete filling of the entire face due to local non-skin areas such as the eyes and eyebrows. In order to eliminate the effects of these errors, the detected image must be subjected to morphological processing. The basic idea of morphology is to use a structural element as a basic tool to detect and extract image features to see if this structural element can be properly and efficiently placed inside the image.

Mathematical morphology [22] is a discipline based on rigorous mathematical theory and analyzing the shape and framework of spatial structures. Mathematical morphology operations are based on a set of views, which means: ① its operations are defined by set operations; ② all images must be converted into sets in a reasonable way.

Different objects in the image are represented by different sets of mathematical morphology. The essence of morphological processing of images is to use the structural elements to continuously shift in the input image. By obtaining the information of the input image during the translation process and analyzing the relationship between the various parts of the image, the structural features of the whole image are analyzed. The process of selecting structural elements is critical. Depending on the structure of the research and the structure of the detected image, the selected structural elements may contain information such as size, chromaticity, grayscale, and morphology. Structural elements of different nature are formed by a collection of different points. Since it is necessary to detect different structural elements on different sides of the image, selecting structural elements suitable for human visual characteristics is a key step in processing the image. Commonly used mathematical morphology treatments include erosion, dilation, opening, closing and so on.

3.1.1 Basic concepts of mathematical morphology

(1) Reflection and translation.

The reflection of a digital image about the origin is defined as $\check{A} = \{a \mid -a \in A\}$, That is, the reflected image $\check{A}$ is obtained by taking the opposite of each point of the source image. Let A be a digital image, a is the pixel in A (that is, $a \in A$); b is a point (can be regarded as a vector b from the origin to the point), then the result of the definition is translated as $A + b = \{a+b \mid a \in A\}$, that is, the entire image moves in parallel along the direction of the vector b.

(2) Structural elements.

The selection of structural elements is closely related to what information is extracted from the target image being processed. As described above, by analyzing and verifying the correlation between the two parts of the image one by one, a set of correlations between the parts is obtained to determine the structural features of the input image. A 'structural element' is defined when analyzing the correlation between the various parts of the input image. By continuously moving structural elements in the image, you can analyze the correlation between the various parts of the image.

According to the purpose of analyzing images of different structural features, the structural elements are roughly classified into a circle, a square, a flat shape, and so on. In morphological analysis, the size and type of structural elements can vary, but in general, the size of structural elements must be significantly smaller than the size of the input image.

3.1.2 Erosion

Erosion is a process of eliminating boundary points and shrinking the boundaries to the inside, which can be used to eliminate small and insignificant noise points. Objects of different sizes can be removed by selecting structural elements of different sizes. If there is a small connection between the two objects, the two objects can usually be separated by an erosion operation.

Set A (input image) is eroded by set B (structural element), which is defined as expression 3-1:

$$A \ \Theta \ B = \{x : B + x \subset A\} \tag{3-1}$$

It can be seen that the erosion of the set A using the set B is that the set of points x included in the set A in all the set B is translated by x. As shown in the figure3-1 below, a 3×3 structure element on the left side of the image is used to perform erosion operation to the set in the middle of the image, and the result is the set on the right side of the image.

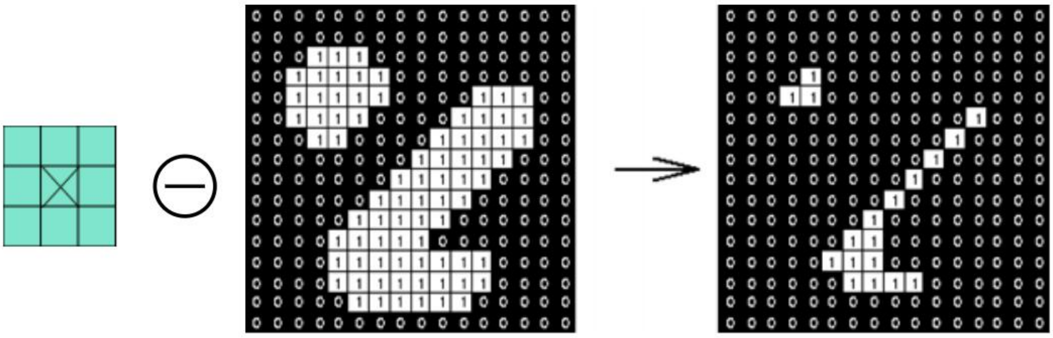


Fig.3-1 The example of erosion

3.1.3 Dilation

The effect of dilation is the opposite of erosion. It is the process of merging all the background points that the object touches into the object, and expanding the boundary to the outside, which can be used to fill the voids in the object. If the distance between two objects is relatively close, the two objects may be connected together by the dilation operation, and the dilation is useful for filling the holes in the object after the image is segmented.

From the point of view of set theory, the dilation operation and the erosion operation are mutually exclusive, that is, the dilation can be defined by the complement of the erosion. Let A^C denote the complement of set A, and $\check{B}$ denote the set of set B after the coordinate origin is rotated by 180 degrees. Then, the dilation of the set A by the set B is expressed as $A \oplus B$, which is defined as expression 3-2.

$$A \oplus B = [A^C \Theta \check{B}]^C \tag{3-2}$$

As shown in the figure3-2 below, a 3×3 structure element on the left side of the image is used to perform dilation operation to the set in the middle of the image, and the result is the set on the right side of the image.

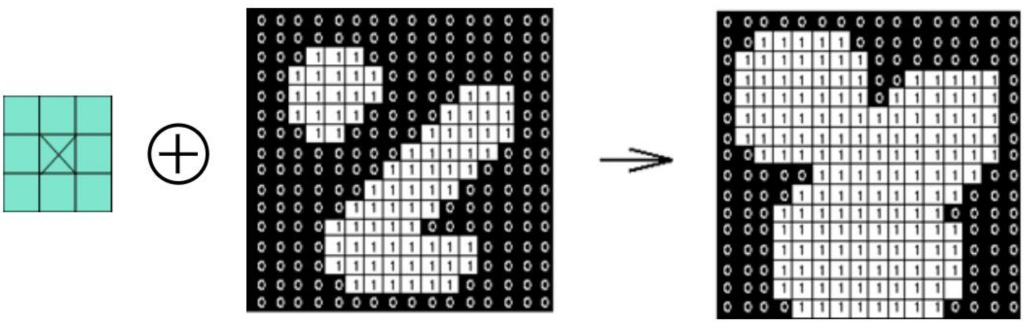


Fig.3-2 The example of dilation

3.1.4 Opening

The opening operation is the process of first erosion after dilation. This operation removes isolated small dots and small joints between two objects, eliminating small objects and smoothing the boundaries of larger objects. Suppose B is a structural element and A is an input image. Use structure element B perform opening operation to A, denoted by symbol $A \bigcirc B$, which is defined as expression 3-3.

$$A \bigcirc B = (A \Theta B) \oplus B \text{ or } A \bigcirc B = \cup (B + x : B + x \subset A) \qquad (3\text{-}3)$$

As can be seen from the above definition, the opening operation can be a combined operation of eroding and dilating: the image is first eroded by structural elements and then dilated by the use of structural elements. It is also possible to translate and combine the results of all structural elements that can be filled inside the image. When a structural element is translated inside the entire image collection, $A \bigcirc B$ is a collection of pixels that make any pixel of structural element B not exceed the boundary of image A.

As shown in the figure3-3 below, a 3×3 structure element on the left side of the image is used to perform opening operation to the set in the middle of the image, and the result is the set on the right side of the image.

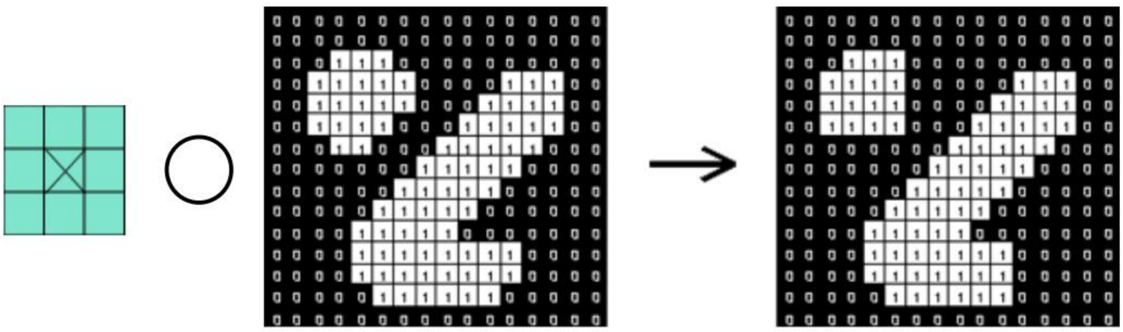


Fig.3-3 The example of opening

3.1.5 Closing

The closing operation is the process of first dilation and after erosion the object. This operation can fill the small holes inside the object, connect adjacent objects, and smooth the boundary of the object without significantly changing the area of the object. From the point of view of set theory, closing and opening operations are mutually dual operations. That is, the closing operation is first performed as a dilation operation, and then the erosion operation is performed. The binary image is performed a closing operation with a structural element, represented by the symbol $A \bullet B$, which is defined as expression 3-4:

$$A \bullet B = (A \oplus B)\ \Theta\ B \text{ or } A \bullet B = (A^C \oplus B)^C \qquad (3\text{-}4)$$

It can be seen from the above formula that performing a closing operation on the image A with the structural element B can obtain a set in which all points x in the set meet the following condition ① x in a mirrored structural element of a translation; ② structural elements and images must have some common points. It can be seen that the A●B after the closing operation contains the initial image, that is, the closed operation is an expansion operation.

As shown in the figure3-4 below, a 3×3 structure element on the left side of the image is used to perform closing operation to the set in the middle of the image, and the result is the set on the right side of the image.

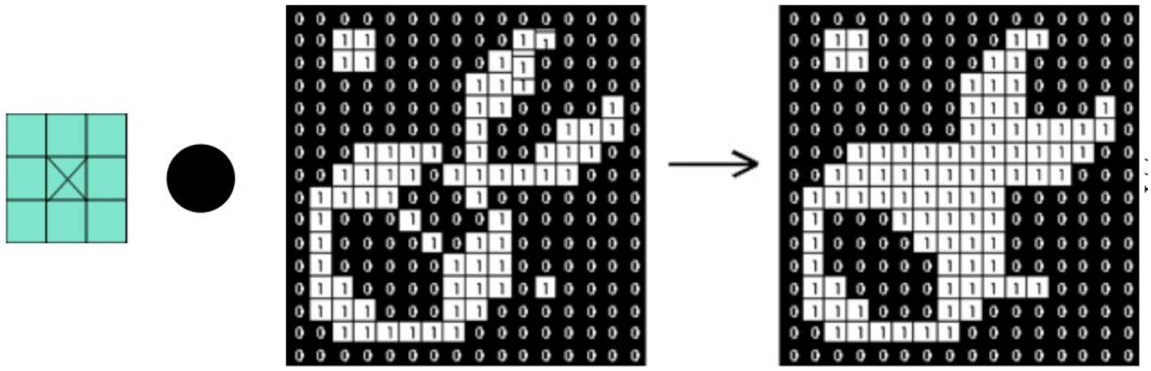


Fig.3-4 The example of closing

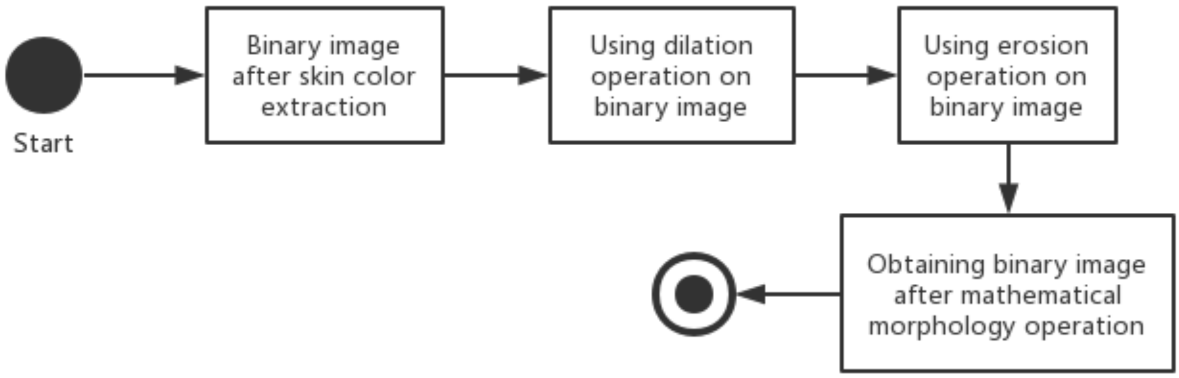


Fig.3-5 The flow of mathematical morphology processing

The flow of mathematical morphology processing is shown as above figure 3-5. The figure3-7 below shows the results of a binary image after skin color segmentation(figure3-6) after performing mathematical morphology processing. As can be seen from the figure, a lot of messy background white points have been removed, which improves the accuracy of the next face localization.

Fig.3-6 Binary image after skin color segmentation

Fig.3-7 The results of figure 3-6 after performing mathematical morphology processing.

3.2 Screening of face areas

After denoising and skin color region marking of binary images, it is necessary to screen the obtained skin color regions to improve the efficiency of later face detection and facial feature detection. This thesis uses the following three methods to perform initial screening of skin color regions.

3.2.1 Screening based on face proportion

In the binary image obtained after the optimal threshold segmentation, there are often some small white pixel dot regions, which are obtained by some regions similar to the skin color in the image background. These areas will inevitably affect subsequent face detection and facial feature detection. These interference areas generally have a common feature: the proportion of the entire image is small, so I can remove these interference factors according to the proportion of these areas.

For an image, the size of the image is usually represented by the number of pixels in the image. Through repeated experiments, 1% and 20×20 pixels of the total pixels of the image are used as thresholds for candidate face region screening. When the area of the candidate face area is smaller than any of these two thresholds, I treat it as a non-face area and remove it. Using these two thresholds can effectively eliminate some interference of non-face small areas.

3.2.2 Screening based on face length and width ratio

After the screening in the previous step, there are still some areas similar to skin color but not human faces. This step will further process the remaining skin color areas. Since some faces in the image may have rotation and tilt, it is impossible to directly use the coordinate values of the upper, lower, left, and right vertices of the region to make a judgment. However, the boundary of the region obtained when marking each region of the image can be utilized.

The detailed process is to count the coordinate values of all the pixels on the boundary of the skin color region, and record the coordinates (X_{min}, Y_1) and (X_{max}, Y_2) having the smallest and largest components in the x-axis direction. Then, the value of (X_{max} - X_{min}) is one of the length and width parameters of the face, and the value of (Y_{max} - Y_{min}) is obtained by the same method. Let the length H of the skin color region be equal to the larger number, and the width W is the smaller one, then the ratio H/W of H to W is the length and width ratio of the skin color region. For a vertical frontal face, the ratio of length to width is generally about 1, but since the face may be rotated and tilted, the face and neck are often extracted as the same face region. After repeated experiments, this thesis selects [0.8, 2.5] as the threshold of the length and width ratio of the face, and the area where the length and width ratio does not satisfy this range is considered to be the non-face area and is removed.

3.2.3 Screening based on the number of edge points in the candidate's area

After screening the above two steps, the binary image has largely ruled out the interference of the non-human face region and the background region with similar skin color. However, there are still some areas that are not non-human faces that have not been screened out by the above two steps. The screening method of the third step can further eliminate these areas.

It can be concluded that the edge of the face is much more complicated than the edge of other non-face regions that are erroneously detected, that is, the pixels at the edge of the face are more than the pixels at other non-face regions. With this, after obtaining the number of pixel points of the edge of each marked connected region, the average of the number of edge points of all the marked regions can be calculated. Because the number of marked face areas is the majority, the number of edge points in the face areas is much larger than in other areas. The area where the number of edge points is smaller than the calculated average value can be considered as a non-face area, and it can be excluded.

The result obtained by screening the skin color regions of Figures 3-7 using the above three methods is shown in Figure 3-8. As can be seen from the figure, almost all similar skin color areas that are not human faces have been removed.

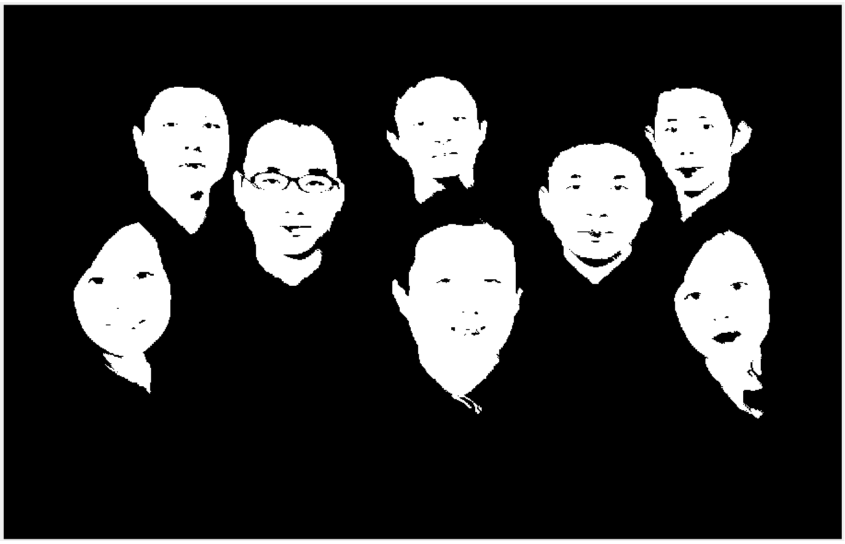


Fig.3-8 The results of screening the skin color regions of figures 3-5 using the above three methods.

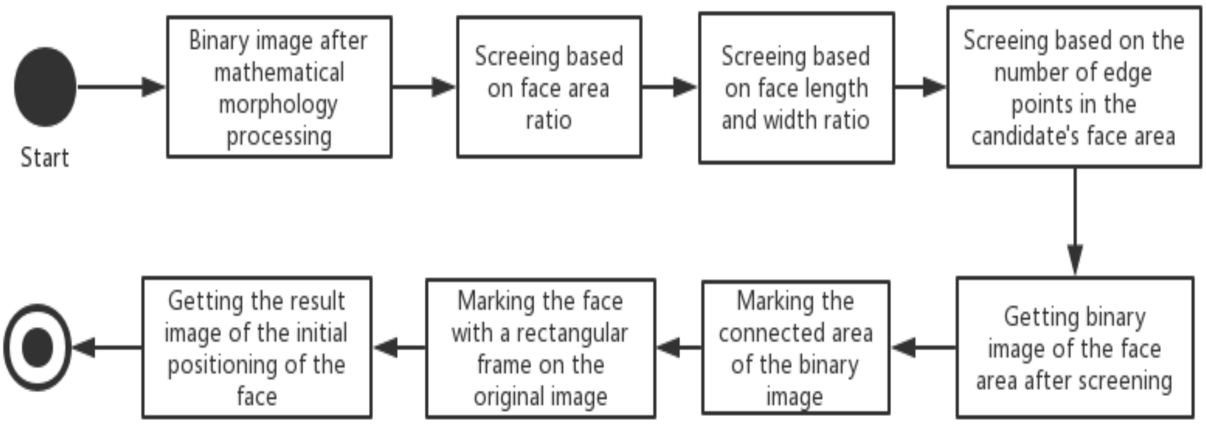


Fig.3-9 The process of initial localization of the faces.

The flow of initial localization of the faces is illustrated as figure 3-9 above. Figure 3-10 shows the result of initial positioning of the face. The figure shows that although the face area is framed by a rectangle, the skin color area of the neck is also framed, and the face cannot be accurately positioned. The next chapter will use the geometric features of the face to further accurately locate the face area.

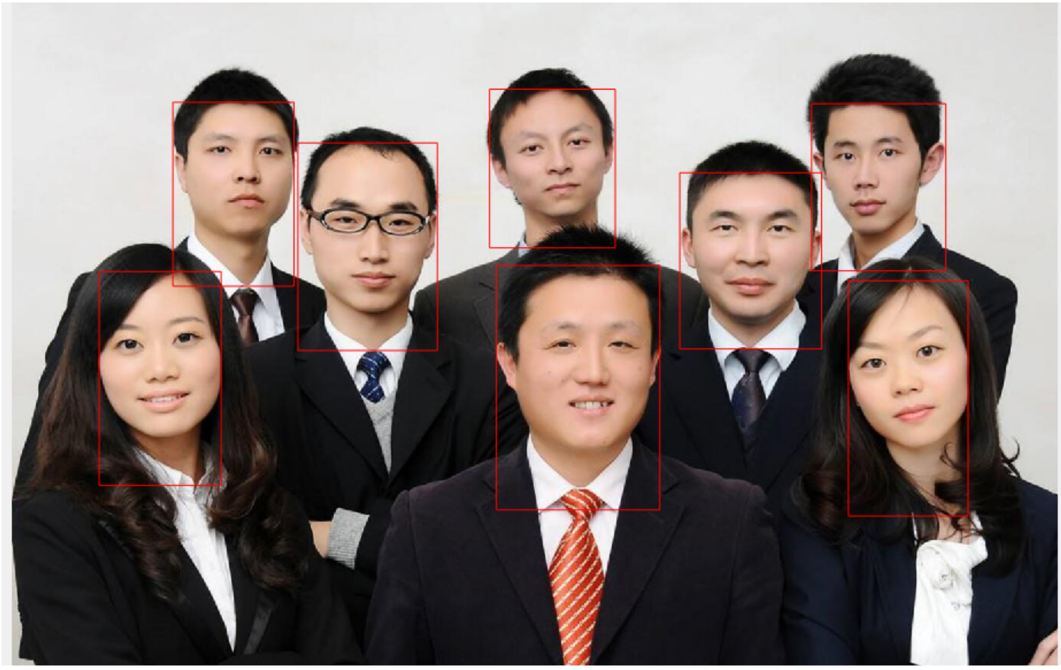


Fig.3-10 The result of face initial positioning

3.3 Chapter summary

This chapter introduces the relevant knowledge of mathematical morphology firstly, and uses the opening and closing operations in mathematical morphology to denoise the binary image after skin color segmentation. The skin color region obtained is simply screened by the proportion of the face region, the length and width ratio of the face and the number of edge points in the candidate's area. Experiments have shown that some small-area non-face areas and hand, arm and other types of face-like areas have a certain screening effect. It reduces the amount of calculation for the facial feature detection, and improves the work efficiency and accuracy of the facial feature detection.

4 Face precise localization based on facial geometric features

4.1 Introduction

Faces are made up of organs such as eyes, mouth, nose, etc. It is precisely because of the differences in shape, size and structure of these organs that every face in the world varies widely. The geometric description of the shape and structure relationship of these organs can be used as an important feature of face recognition and expression recognition. Correct extraction of these features is often considered to be the focus and difficulty of pattern recognition. Geometric feature extraction is a geometric method to describe the shape, size and structure of the eyes, mouth, nose, chin and other organs, such as the position and width of the eyes, nose and mouth, the thickness of the eyebrows and the degree of bending. And the relationship between these characteristic organs. In addition to the geometric features, the face also has color features, such as pupils, black eyebrows, and red lips, which can be used.

Facial features can be divided into two categories: permanent facial features and temporary facial features. Permanent features are permanent facial features that deform as the face moves, such as eyes, eyebrows, mouth, and certain wrinkles in the face. Temporary features mainly include temporary wrinkles on the face that appear with the movement of the face, but not permanently on the face. The method of extracting facial features can be generally divided into: a global feature extraction method and a local based feature extraction method. The global-based feature extraction method is a one-time extraction of the required facial permanent features and temporary features. The local feature extraction method is to divide the face into sub-regions or decompose the facial features into some local features, and independently operate the sub-regions or local features. This paper uses local feature-based extraction methods to extract permanent facial features.

The algorithm process of face precise localization is illustrated as the flow chart (figure4-1) blow:

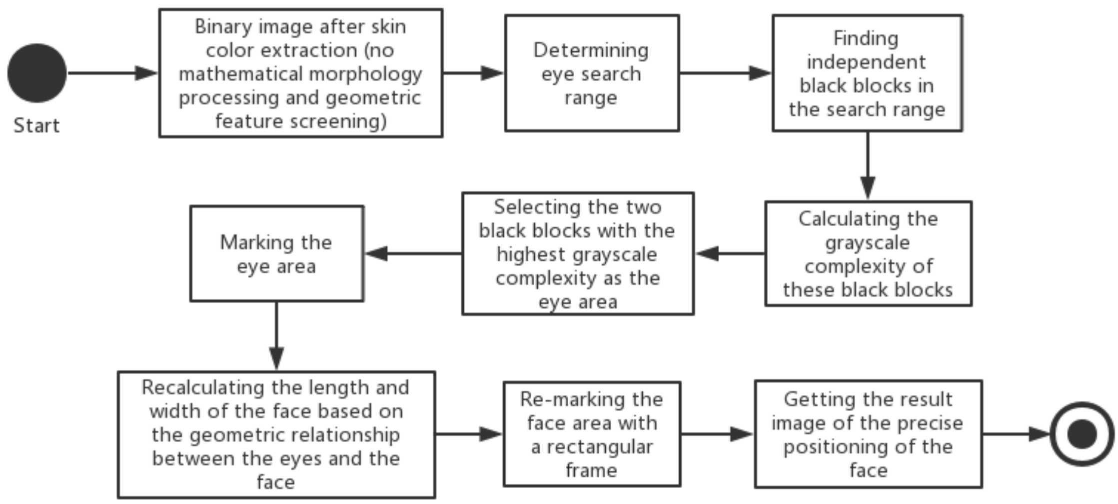


Fig.4-1 The process of face precise localization based on face geometric features

4.2 Eyes detection and localization

The eyes are the most obvious and richly moving organ in the human face, and contains many useful features. Therefore, the precise localization of the eyes plays an important role in face recognition and expression recognition. A geometric description of the eye is necessary to first find the position of the eye, that is, to localization the eye first.

4.2.1 Common eye detection algorithms

At present, there are many researches on eye positioning at home and abroad, and there are many subject areas involved, such as computer graphics, image processing, pattern recognition, statistics, biophysics and neurobiology. These research methods are summarized into the following categories:

(1) Template matching [23]

The basic principle of template matching [24] is to translate the reference template image point by point in the search area of the image, traversing each position point in the search area. At the same time, according to a similarity measure principle, the correlation value of the image area of the location point in the search area and the reference template is calculated, and then the position of the tracking point is determined according to the magnitude of the correlation value.

The template matching method mainly selects the template, and can be divided into two methods according to the dimension of the selected templates which are a two-dimensional deformable template and a three-dimensional deformable template. Then the matching eye template is used to match in the image. This method has requirements for the initial position, and the calculation amount is too large, and the real-time performance is poor. At the same time, the use of a large amount of prior knowledge has led to the detection results are not optimal, and has been greatly limited in terms of versatility.

(2) Hough transform method [25]

The Hough transform method is mainly for the study of the eyeball. Before using the Hough transform to detect the eyeball, the Canny [26] algorithm is used to extract the edge. Suppose the image space is (i, j), i and j represent rows and columns, respectively, and the three-dimensional transformation space is (i_e, j_e, R), where i_e, j_e represent the rows and columns of the eyeball, respectively, and R is the radius. For the relatively slender eyes, since the upper half of the eyeball is more covered by the eyelids, the lower semicircle is detected instead. The lower semicircle expression 4-1 is:

$$i = i_e + \sqrt{r^2 - (j - j_e)^2} \tag{4-1}$$

For each coordinate point (i_e, j_e, R) of the transformation space, there is a semicircle in the image space, and the number of edge points existing on this semicircle is the value corresponding to the coordinate point (i_e, j_e, R) on the transformation space. Experiments show that the Hough transform has the merit of strong anti-interference ability. When the contrast between the eyeball and the white of the eye is low, the extracted edge is broken or not very regular. Even so, the eyeball circle can be accurately positioned according to the peak point in the transformation space [27].

(3) Edge feature analysis method [28]

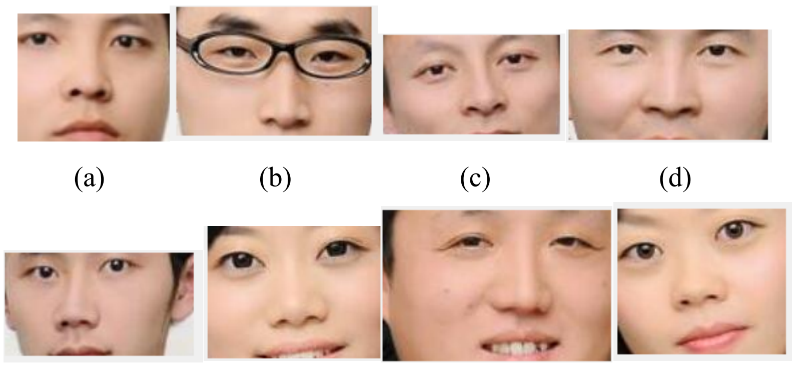
Many eye positioning methods are based on grayscale information of the image, and the edge feature analysis method is one of the typical examples. It mainly determines the left and right boundaries of the face according to the width of the peak according to the vertical gray projection curve of the face image. Then use the horizontal gray projection curve of the face area to determine the upper and lower boundaries formed by the top of the head and the middle of the nose. The prediction method is used to first determine the approximate position of the human eye in the eyebrows and the eye area, and the coordinate positions of the eyes are determined by detecting the grouping of the edges and edges of the eyebrows and the eye parts.

Before the edge detection, the eyebrows and the eye area are denoised and enhanced to smooth the image. This is a key step to ensure better edge detection in the next step. Then use the Canny algorithm to find the edge image of the eyebrow and eye area. The threshold is calculated to be high so that the edge points consist only of the strong edges of the eyebrows and the eyes and the strong edges of the pupils, and do not contain the weak edges created by the contours of the nose. Then the edge points are grouped to get several sets of separate edges. The two groups located below are the edges produced by the left and right eyes, respectively, and the centers of the two edge groups are taken as the center of the human eye. In the edge grouping algorithm, the smallest rectangle of each edge group is respectively included as the human eye, and the first two groups detected respectively correspond to the left eye and the right eye. The algorithm ends immediately after two packets appear, improving execution efficiency.

The advantage of this algorithm is that the eyebrows and eye regions are first determined by the gray projection curve, which makes the calculation relatively simple and the positioning speed is faster. However, the distribution of peaks and valleys is very sensitive to different face and attitude changes, the positioning accuracy is poor, and it is easy to fall into local minimum and cause positioning failure.

4.2.2 Determination of the eye search area

After marking the approximate position of the face, the positioning of the eye is about to take place. In order to narrow the search range of the eyes, I extracted a partial area as the search range of the eye based on the face area. According to the prior knowledge of the face and the analysis of multiple face images, I set the eye search area at the upper two-thirds of the face area. However, in actual face positioning, it is often affected by the skin area such as the neck, so I finally set the eye search area to the upper three quarters of the face area. In this way, the influence of the neck can be largely removed, and the area of the search area can be reduced to improve the efficiency. The eye search area of the figure3-8 is as shown in the figure4-2. As can be seen from the figure, the search area of the eyes is different due to the different inclinations of different faces in the image.



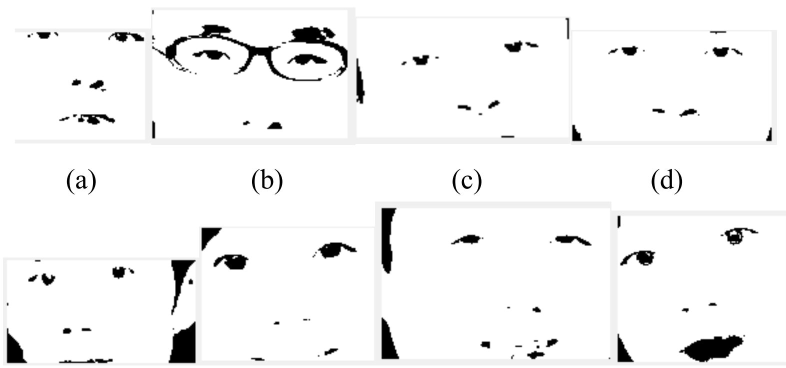
(a) (b) (c) (d)

(e) (f) (g) (h)

Fig.4-2 The searching area of eyes

(a)-(h) The part of the face region

4.2.3 Marking the eye area



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig.4-3 The binary image of eyes searching areas

(a)-(h) Difference binary images

From the binary image (figure 3-7) subjected to the denoising process in Chapter 3, it can be seen that since the eye region is small, the eye region is often considered to be noise and filtered in the filtering process. Therefore, I chose the binary image before the denoising process to search the eye area. The binary image can be seen in figure 4-3 above.

The center point of each color block in the eye search area is the candidate point of the eye. The specific step is to find isolated black blocks in the eye search area and find the coordinates of the center points of the black blocks. The concept of image complexity is then introduced to determine the candidate eye region based on the apparent grayscale variation of the white of the eye and the eyeball. The calculation formula of image grayscale complexity [29] is defined as:

$$\mathrm{Com(k)} = \sum_{i=1}^{h}\sum_{j=1}^{w-1}\left|B_{i,j+1} - B_{i,j}\right| \cdot \min(j,\ m-j) \tag{4-2}$$

Where w, h are the width and height of the acquired image, and $B_{i,j}$ is the gray value of the pixel of the i-th row and the j-th column. The formula calculates the longitudinal first-order weighted derivative of the gray value of the image block. The coefficient min(j, m-j) is the weight, and the closer the derivative is to the center of the image, the greater the weight. Experiments have shown that this definition is consistent with the calculation of the complexity of facial features. Similarly, there may be a horizontal first-order weighted derivative, but it is found through experiments that the longitudinal first-order weighted derivative of the human eye is larger, so the first-order longitudinal weighted derivative of the image block gray value is used herein.

The image block gray-scale complexity of the region is calculated by using the black block center coordinate points obtained above, and the block size is generally approximated to the size of one eye. The first two blocks with the highest gray level complexity are used as the eye area. In the experiment, it was found that there are no black spots corresponding to the eyes in the candidate face area of the detection image with multiple faces or small faces, so it is impossible to detect the eyes 100 percent. The results of eye detection are shown below in the figure4-4.

Fig.4-4 The result of eyes detection

4.3 Face area re-marking

After getting the facial features, the face area is also re-marked. In order to remove the influence of the neck, arm or other non-skin background points, the effect of accurately positioning the face is achieved. By studying the ratio of the distance between the eyes to the width of the face, the distance from the eye to the chin to the distance between the eyes, and the ratio of the distance from the eye to the top of the forehead to the distance between the eyes [30], the three ratio centers can be used to calibrate the face region. The specific implementation steps are as follows:

(1) Calculate the distance d between the eyes, assuming that the center points of the eyes are (x1, y1) and (x2, y2), respectively. The formula 4-3 for calculating the distance between the two points is as follows:

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (4\text{-}3)$$

(2) Calculate the center point of the line connecting the eyes. The midpoint formula 4-4 is as follows:

$$\bar{x} = \frac{x1+x2}{2},\ \bar{y} = \frac{y1+y2}{2} \quad (4\text{-}4)$$

(3) Based on the center point between the eyes, extend a distance between the eyes to the left and right respectively, so that the width of the face can be obtained.

(4) Still based on the center point between the eyes, upward extending a distance between the eyes and downward extending 1.7 times the distance between the eyes, the length of the face can be obtained.

Through the above four steps, the re-marking of the face area can be completed. The experimental results are shown in Figure 4-5.

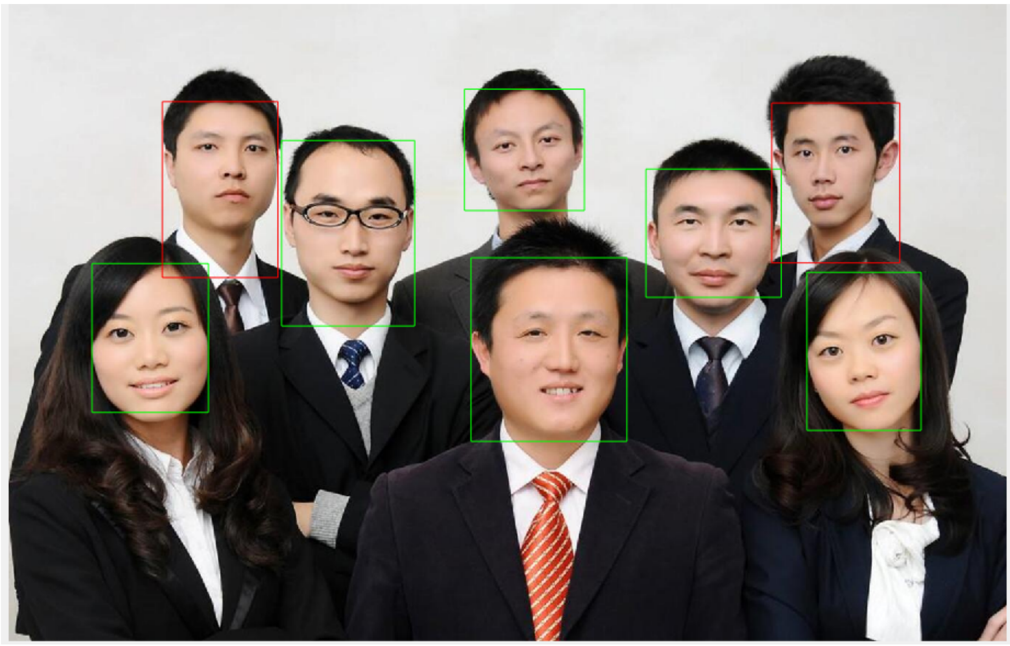


Fig.4-5 The result of face area re-marking

As can be seen from the figure, the precise positioning result of the face is affected by the eye detection result. Face detection results that eyes are not correctly detected are inferior to those of other faces that eyes are correctly detected.

4.4 Count the number of faces

After the face is marked, the number of the face is counted, and the rectangular frame marking the face is Sequential marked. The results obtained are shown in Figure 4-6. Because the order in which the images are traversed is a vertical traversal, so the order of the labels is also vertical.

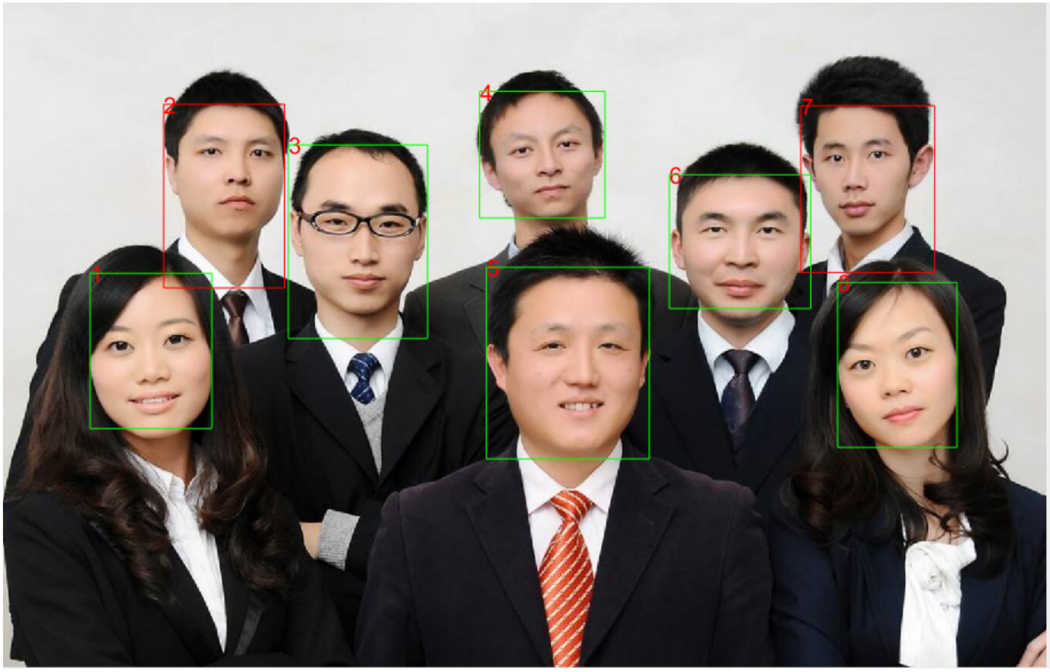


Fig.4-6 Sequential labeling of rectangular boxes that mark faces

4.5 Chapter summary

This chapter uses color features and geometric features to independently detect the eye area, and then uses the structural relationship between the face and the eye to accurately locate the face. Such an algorithm can greatly eliminate the interference of the non-face area of the skin color and the background area of the similar skin color, thereby improving the accuracy of the face feature detection.

5 Experimental results and data analysis

The previous four chapters detail the extraction and segmentation of the skin color region of the image to be detected using the elliptic skin color model in the YCbCr color space, as well as the screening of candidate face regions and the detection of facial geometric features in the region. At the same time, the knowledge about image processing involved in these steps is also introduced in detail. This chapter will describe the entire algorithm flow, show some experimental results, compare the method and other methods, and analyze the experimental results.

The experimental environment is mainly composed of development tools and software platforms. The development tool used in this thesis is MATLABR2018a, and the software platform is Windows10.

The test images are mainly composed of 50 pictures downloaded from the Internet. In order to better test the detection effect of the algorithm under various complex backgrounds, the test pictures selected in this thesis contain different lighting conditions, different backgrounds and different ages. I divided these pictures into four categories, namely low-density face images (the number of faces is less than 5), medium-density frontal face images, medium-density side face images, and high-density face images in complex backgrounds.

5.1 Algorithm flow overview

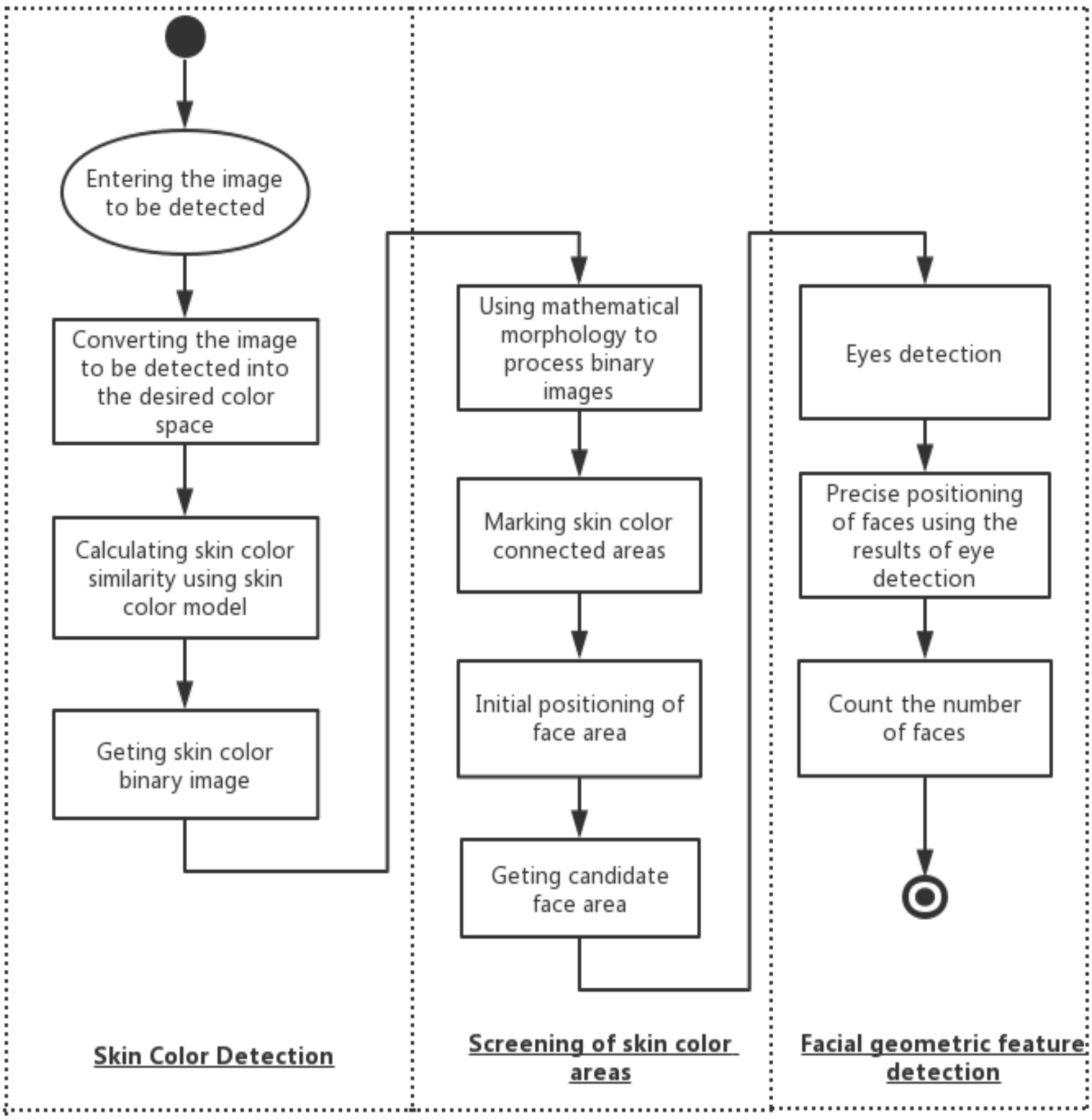


Fig.5-1 The flow chart of face detection algorithm

Figure 5-1 depicts the entire algorithm flow. It can be seen from the figure that the algorithm flow is detected by stepwise screening and detection, and then the geometric features of the face are detected. In general, the entire algorithm flow consists of three modules, the first module is the detection of skin color, the second module is to filter the skin color region, and the third module is the detection of the face geometric feature.

The first module consists mainly of three parts:

(1) Converting the input image to be detected to the desired color space.

The original image of the RGB color space is converted to a grayscale image to obtain the length and width of the image. The original image is then converted from the RGB color space to the YCbCr color space to prepare for subsequent calculation of skin color similarity.

(2) Calculating skin color similarity.

The elliptic model is used to calculate the similarity of the skin color of each pixel in the image, and the skin color segmentation is performed according to the obtained calculation result.

(3) Getting the skin color binary image.

In the previous step of calculating the skin color similarity, if the calculated result is greater than 1, then it is not the skin color point, and the chromaticity of the point is adjusted to 0, that is, it is displayed as black. On the other hand, if the calculation result is less than 1, it means that the point is a skin color point, and the chromaticity of the point is adjusted to 255, that is, it is displayed as white. A binary image of the skin tone can be obtained after traversing all the pixels of the entire image.

The main function of the second module is to perform a preliminary screening of the skin color region obtained by the first module, and finally obtain the candidate face region. The module mainly includes the following two parts:

(1) Processing binary images using mathematical morphology.

Erosion and dilation operations are performed on the skin color binary image obtained by the first module to remove the influence of noise and reduce the calculation amount of subsequent operations.

(2) Initial positioning of the face area.

Using the rules of the proportion, length and width ratio of the face and the number of face edge points, the candidate face area is initially screened to eliminate the interference of the non-face area.

The third module is mainly to detect and locate facial features, and finally get the position of the eyes and the precise position of the face. The module is mainly composed of the following three parts:

(1) Eyes detection.

First, determine the search range of the eye, search for independent black blocks in the search area, and perform eye detection by calculating the grayscale complexity of the image of these black blocks to obtain an eye region.

(2) Precise positioning of the face.

Re-mark the face area using the ratio of the distance between the eyes and the chin and the top of the forehead.

(3) Counting the number of faces.

The number of face regions that are filtered out is counted, and each rectangular frame that circles the face is marked with numbers in order.

5.2 Algorithm implementation principle

In the process of implementing this detection algorithm, I followed the following principles:

(1) The output rectangle should contain as many of the faces as possible in the image to be detected. Not reducing the correct rate is the premise of the whole skin segmentation process, and it is also the primary premise of the face detection algorithm.

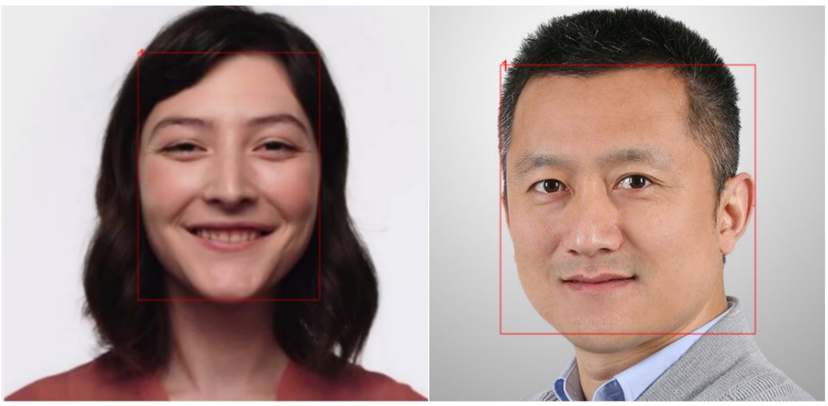
(2) The output rectangle should contain as few non-skin areas as possible. In other words, it is necessary to eliminate the interference of the non-face area of the skin color or the background area of the like-skin color.

(3) The detection speed should be controlled within a certain range. However, due to the different size and complexity of the pictures used, this thesis does not make a detailed design of the detection speed.

Among the above three principles, the first principle means that the correct rate is high and the missed detection rate is low. The second principle means that the false detection rate is low, and the third principle means that there is a certain detection speed. But these three principles are interrelated and mutually restrictive. Therefore, comprehensive consideration, I will try my best to meet the above three principles as much as possible. The method used in this paper satisfies the first principle as much as possible to improve the correct rate. On this basis, the second principle is satisfied as much as possible to reduce the false detection rate.

5.3 Face detection results display

The images below show the face detection results for different images to be detected.



(a) (b)

Fig.5-2 Single face detection result

(a), (b) Different single face detection result

Fig.5-3 Double face detection result

Fig.5-4 Medium density frontal face detection results

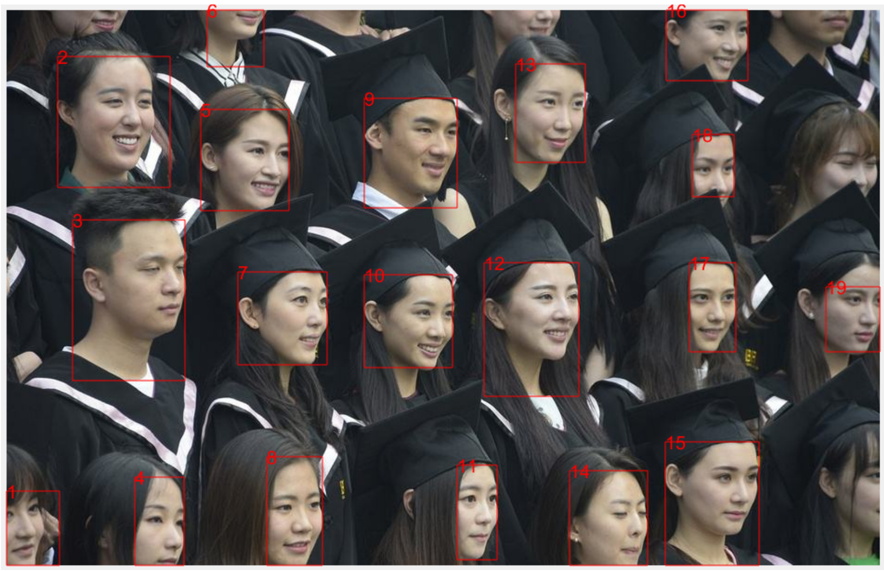


Fig.5-5 Medium density side face detection results

Fig.5-6 Face detection results with higher density and complex background

It can be seen from the figures that for images with small face density, the detection results are better and the false detection rate is lower. However, for images with high face density in complex backgrounds or images with oblique faces, the false detection rate and missed detection rate increase. The next section will do data analysis on the detection results of these pictures.

5.4 Data statistics and analysis

This section uses the classification model evaluation indicators for the face detection results of different categories pictures for data analysis.

5.4.1 Basic concept introduction

(1) Meaning of the four signs.

TP (True Positive): The real is a positive example, predicted as a positive example, and the prediction is correct. For example, the true value is 1, and the prediction is also 1.

FN (False Negative): The real is a positive example, predicted as a negative example, predicting errors. For example, the true is 1, and the prediction is 0.

TN (True Negative): The real is a negative example, predicted to be a negative example, and the prediction is correct. For example, the true value is 0 and the prediction is also 0.

FP (False Positive): The real is a negative example, predicted as a positive example, predicting errors. For example, the true value is 0 and the prediction is 1.

(2) Confusion Matrix [31].

The confusion matrix, also called the error matrix, is a standard format for the accuracy evaluation, and is represented by a matrix of n rows and n columns. The specific evaluation indicators have overall accuracy, drawing accuracy, user accuracy, etc. These precision indicators reflect the accuracy of image classification from different aspects. In artificial intelligence, the confusion matrix is a visualization tool, especially for supervised learning. In unsupervised learning, it is generally called a matching matrix. In the image accuracy evaluation, it is mainly used to compare the classification result with the actual measured value, and the accuracy of the classification result can be displayed in a confusion matrix.

The example table of the confusion matrix is as follows:

Table5-1 The format of confusion matrix

		Real value	
		Positive	Negative
Detection value	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

(3) Accuracy [32].

Accuracy is defined as the ratio of the number of samples correctly classified by the classifier to the total number of samples for a given test data set. The formula is:

$$\text{Accuracy} = \frac{TN+TP}{TP+FP+FN+TN} \qquad (5\text{-}1)$$

(4) Precision

The precision is calculated as: The ratio of the predicted results to the actual value can be understood as the case without “false positives”. The formula is:

$$\text{Precision} = \frac{TP}{FP+TP} \qquad (5\text{-}2)$$

(5) Recall

The recall rate is calculated as the ratio of the number of correct classifications to the number of all should be correctly classified (in accordance with the target label), which can be understood as the absence of “missing” in the recall rate. The formula is:

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (5\text{-}3)$$

5.4.2 Data demonstration

I use the total pixels of the image to describe the size of the entire image. The size of a face is represented by the number of pixels of its circumscribed rectangle. The size of the non-face area is represented by the sum of the pixels of the image minus the pixels of the face area.

(1) Data of single face detection.

The pixels of image are: 500×750.

Table 5-2 The confusion matrix of detection result of single face

		Real value	
		Face	Non-face
Detection value	Face	111,283.2	0
	Non-face	0	263,716.8

The three evaluation indicators can be calculated based on the formula 5-1, formula 5-2, formula5-3. The results are show below:

$$\text{Accuracy} = \frac{111{,}283.2+263{,}716.8}{111{,}283.2+0+0+263{,}716.8} = 100\%$$

$$\text{Precision} = \frac{111{,}283.2}{111{,}283.2+0} = 100\%$$

$$\text{Recall} = \frac{111{,}283.2}{111{,}283.2+0} = 100\%$$

(2) Data of low density faces detection.

The pixels of image are: 498×715.

Table 5-3 The confusion matrix of detection result of low density faces image

		Real value	
		Face	Non-face
Detection value	Face	16,326.4	0
	Non-face	0	339,743.6

The three evaluation indicators can be calculated based on the formula 5-1, formula 5-2, formula5-3. The results are show below:

$$\text{Accuracy} = \frac{16{,}326.4+339{,}743.6}{16{,}326.4+0+0+339{,}743.6} = 100\%$$

$$\text{Precision} = \frac{16{,}326.4}{16{,}326.4+0} = 100\%$$

$$\text{Recall} = \frac{16{,}326.4}{16{,}326.4+0} = 100\%$$

(3) Data analysis of medium density frontal faces detection.

The pixels of image are: 425×640.

Table 5-4 The confusion matrix of detection result of medium density frontal faces image

		Real value	
		Face	Non-face
Detection value	Face	53,576.46	5952.94
	Non-face	1215	211,255.6

The three evaluation indicators can be calculated based on the formula 5-1, formula 5-2, formula5-3. The results are show below:

$$\text{Accuracy} = \frac{53{,}576.46+211{,}255.6}{53{,}576.46+5952.94+1215+211{,}255.6} = 97.4\%$$

$$\text{Precision} = \frac{53{,}576.46}{53{,}576.46+5952.94} = 90\%$$

$$\text{Recall} = \frac{53{,}576.46}{53{,}576.46+1215} = 97.8\%$$

(4) Data of medium density side faces detection.

The pixels of image are: 598×940.

Table 5-5 The confusion matrix of detection result of medium density side faces image

		Real value	
		Face	Non-face
Detection value	Face	118,694.17	20,946.03
	Non-face	13,453.5	409,026.3

The three evaluation indicators can be calculated based on the formula 5-1, formula 5-2, formula5-3. The results are show below:

$$\text{Accuracy} = \frac{118{,}694.17+409{,}026.3}{118{,}694.17+20{,}946.03+13{,}453.5+409{,}026.3} = 93.9\%$$

$$\text{Precision} = \frac{118{,}694.17}{118{,}694.17+20{,}946.03} = 85\%$$

$$\text{Recall} = \frac{118{,}694.17}{118{,}694.17+13{,}453.5} = 89.8\%$$

(5) Data of high density faces in complex background detection.

The pixels of image are: 586×1117.

Table 5-6 The confusion matrix of detection result of high density faces image

		Real value	
		Face	Non-face
Detection value	Face	77,614.4	32,804
	Non-face	32,648.616	511,494.984

The three evaluation indicators can be calculated based on the formula 5-1, formula 5-2, formula5-3. The results are show below:

$$\text{Accuracy} = \frac{77{,}614.4+511{,}494.984}{77{,}614.4+32{,}804+32{,}648.616+511{,}494.984} = 90\%$$

$$\text{Precision} = \frac{77{,}614.4}{77{,}614.4+32{,}804} = 70.3\%$$

$$\text{Recall} = \frac{77{,}614.4}{77{,}614.4+32{,}648.616} = 70.4\%$$

The number of face detection results for all test images is shown in the table 5-7 below.

Table 5-7 The result of face detection

Categories	True number of faces	Detection number of faces	Mistake detection number of faces	Missing rate(%)	False rate(%)
Low-density faces	23	23	1	0	4.3
Medium-density frontal faces	246	240	25	2.4	10.2
Medium-density side faces	211	179	32	15.2	15.1
High-density faces	405	382	120	5.6	29.6

5.4.3 Conclusion

Through the above calculations, the following conclusions can be drawn:

(1) The adaptability of the algorithm is good, and it has a good effect on the detection of rotation or side face, and it can basically detect the face accurately.

(2) When the image resolution is high or low-density face (single face or double face), the algorithm used in this paper can easily separate the face area from the background and accurately detect the eye area. However, in the case where the background is complicated, the number of faces is relatively large, or the resolution of the image is low, the detection of the eyes is likely to fail, and thus the face cannot be accurately positioned.

(3) It can be seen from the above data that the precision rate has been maintained at a high value, indicating that the method used in this paper satisfies the principle of not false detection.

(4) The downside is that the algorithm used in this paper can't distinguish the skin-like area from the real face area, and there is a phenomenon of missing face, especially when the face has a certain angle of rotation and the face is incomplete.

5.5 Comparison and analysis of other face detection algorithms

This section uses a face detection algorithm based on a Gaussian distribution model in YCbCr color space, which uses the same images to be detected for face detection. And compare the experimental results with the methods used in this thesis.

5.5.1 Experimental result display

The image to be detected used here is identical to the image to be detected used in Section 5.3, and the experimental results are shown in the following figures.

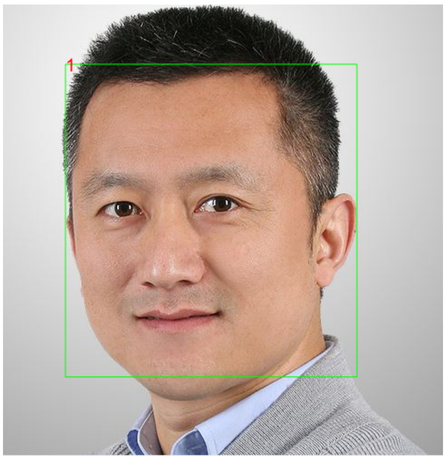


Fig.5-7 Single face detection result based on Gaussian distribution model

Fig.5-8 Double face detection result based on Gaussian distribution model

Fig.5-9 Medium density frontal face detection results based on Gaussian distribution model

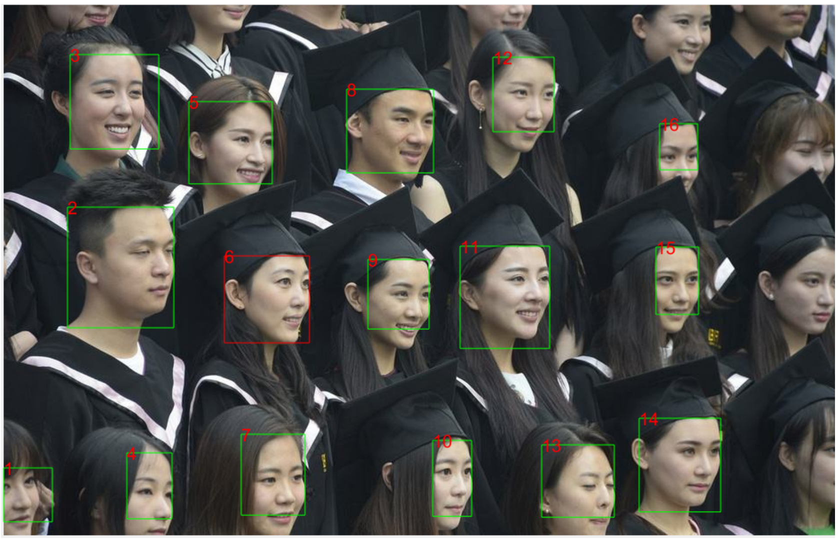


Fig.5-10 Medium density side face detection results based on Gaussian distribution model

Fig.5-11 Higher density face detection results in complex background based on Gaussian distribution model

5.5.2 Data comparison

Data analysis based on the Gaussian model-based face detection algorithm is also performed using the same three evaluation indicators as in Section 5.3. The following is a comparison of the evaluation indicators of the algorithm used in this paper and the face detection algorithm based on Gaussian skin color model. The first table compares the detection of low-density face images, including single-face or double-face images.

Table5-8 Data comparison of face detection results for low-density face images

	Accuracy	Precision	Recall
The algorithm used in this thesis	100%	100%	100%
Face Detection Algorithm Based on Gaussian Distribution Model in YCbCr Space	100%	100%	100%

Table5-9 Data comparison of face detection results of medium density frontal face images

	Accuracy	Precision	Recall
The algorithm used in this thesis	97.4%	90%	97.8%
Face Detection Algorithm Based on Gaussian Distribution Model in YCbCr Space	93.7%	76.1%	97.9%

Table5-10 Data comparison of face detection results of medium-density side face images

	Accuracy	Precision	Recall
The algorithm used in this thesis	93.9%	85%	89.8%
Face Detection Algorithm Based on Gaussian Distribution Model in YCbCr Space	91.1%	85%	80.2%

Table5-11 Data comparison of face detection results of higher density faces in complex background

	Accuracy	Precision	Recall
The algorithm used in this thesis	90%	70.3%	70.4%
Face Detection Algorithm Based on Gaussian Distribution Model in YCbCr Space	85.5%	59%	63.1%

The line charts below show the changes in the accuracy, precision, and recall of the two algorithms as the number of faces and the background complexity of the image increase. Let algorithm A represent the algorithm used in this paper, and algorithm B represents face detection algorithm based on Gaussian distribution model in YCbCr space. The four categories represent low-density face detection, medium-density frontal face detection, medium-density side face detection, and high-density face detection in complex backgrounds, respectively.

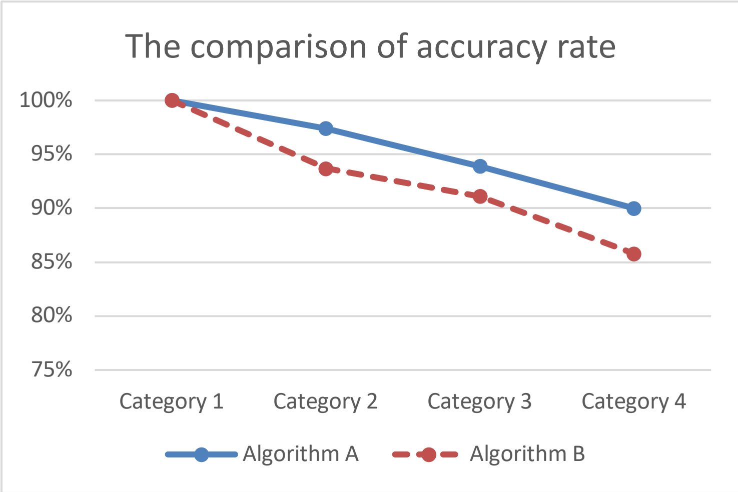


Fig.5-12 The line chart of the accuracy rate of the tow algorithm

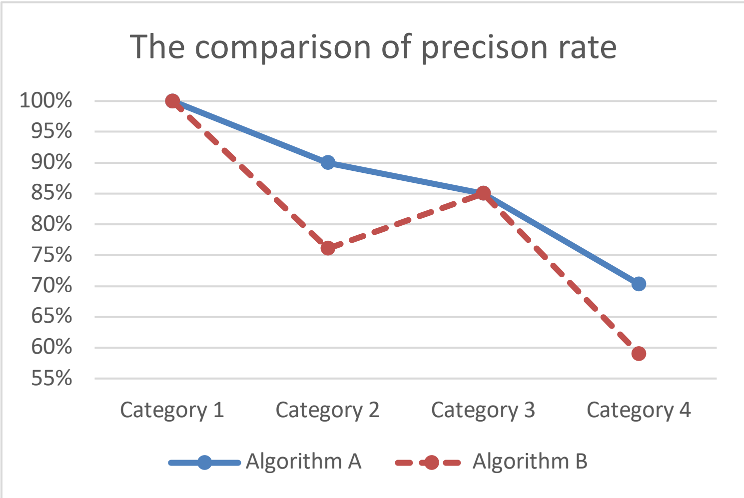


Fig.5-13 The line chart of the precision rate of the tow algorithm

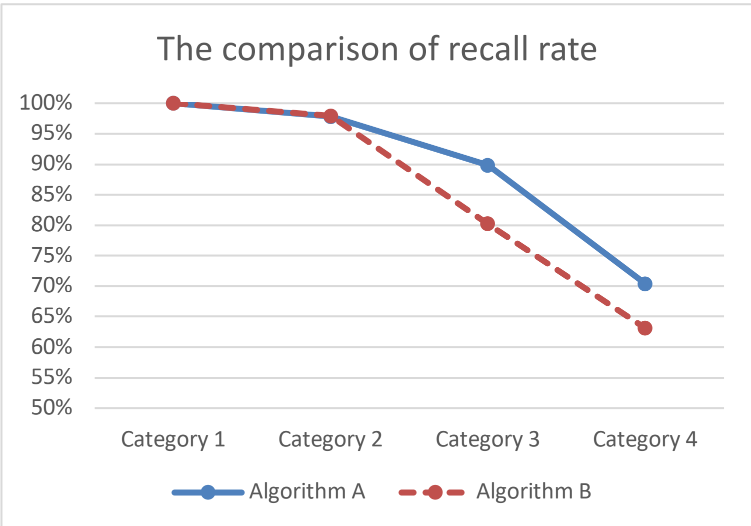


Fig.5-14 The line chart of the recall rate of the tow algorithm

5.5.3 Conclusion

Through comparison of image results and data comparison, the following conclusions can be drawn:

(1) It can be seen from the picture that the face detection algorithm based on the Gaussian distribution model cannot detect the face well. Therefore, the neck area cannot be separated from the face area, and the rectangular frame encloses the neck area and the face area together, and accurate face positioning cannot be achieved.

(2) For low-density face images, the algorithm used in this paper and the face detection algorithm based on Gaussian distribution model have good effects, and can accurately locate the face without the phenomenon of missed detection and false detection.

(3) When the face density starts to increase, the detection effect of the face detection algorithm based on the Gaussian distribution model begins to deteriorate. It can be seen from the comparison results of medium-density positive face detection that the algorithm used in this paper has better adaptability.

(4) For the detection of side faces, the algorithm used in this paper is better than the face detection algorithm based on Gaussian distribution model in terms of accuracy rate and recall rate. It can be seen that the face detection algorithm based on Gaussian distribution model lacks flexibility, and the missed detection rate of side face images is high.

(5) In the high-density face detection in complex background, it can be seen from the data that the accuracy, precision and recall rate of the face detection algorithm based on Gaussian distribution model are lower than the algorithm used in this paper. That is to say, the face detection algorithm of the Gaussian distribution model has a higher false detection rate and missed detection rate, and the similar skin color area cannot be well excluded.

5.6 Chapter summary

This chapter first introduces the entire algorithm flow used in this paper, and briefly describes the main parts of each module implemented by the algorithm. Then the experimental results are presented, and the experimental results are statistically analyzed using a confusion matrix, and the experimental data is described using the accuracy rate, precision, and recall rate. The detection results of the face detection algorithm based on the Gaussian distribution model, which is different from the algorithm used in this paper, are shown. Then the detection results are described using the same evaluation indicators, and the two algorithms are compared. The conclusion is drawn by comparing the results.

6 Summary and Future Prospect

6.1 Summary

Based on careful analysis and summarization of existing related work, this thesis designs and implements a face detection algorithm based on skin color and face geometric features. The main work of this thesis is:

(1) Before performing face detection using skin color, first convert the image to be detected from

the RGB color space to the desired color space.

(2) This thesis uses the elliptical skin color model to extract the skin color region, and then uses the erosion, dilation, opening, closing operations in mathematical morphology to eliminate the interference of some non-skin regions.

(3) After mathematical morphology processing, there are still some background spots similar to skin color. In this thesis, the interference of non-face areas is further eliminated by using the face proportion, the ratio of the length and width of the face, and the pixel points on the edge of the face, which improves the efficiency for subsequent work.

(4) When detecting the geometric features of the face, that is, the detection of the eye, the search range of the eye is first determined according to the distribution rule of the face organ. Then the detection and positioning of the eye is performed within the reduced search range, which greatly improves the detection speed.

(5) In this thesis, through the geometric relationship between face and eyes, the length and width of the face are recalculated to achieve accurate positioning of the face.

6.2 Future prospect

Due to the limited level, the thesis still has some shortcomings, which require further research and improvement.

(1) When the light is heavily affected and the part of the face is extremely dark or extremely bright, the skin color of the face changes, and the candidate face area extracted during the skin color detection is incomplete. Or when the face area is similar to the background area color, the extracted candidate face area is merged with the similar skin color background. Caused the failure of the precise positioning of the face.

(2) When the face is blocked by accessories such as hair and glasses, the extracted face area of the candidate is not standardized, which may easily lead to false detection and missed detection of the face.

(3) When the angle of deflection of the face in the image is large, the geometric relationship of the facial features cannot be used to detect and locate the facial organs such as the eye, which may easily cause false detection and missed detection, resulting in failure of facial feature positioning. Multi-angle face detection is not yet possible.

(4) Most of the experimental samples in this experiment were based on yellow races. This thesis does not do further research on other color races, so the face detection effect for other color races may not be very good.

Although the algorithm proposed in this paper has certain detection accuracy, due to the above problems, the results of face detection still have the phenomenon of false detection and missed detection. In the next step, I will continue to conduct in-depth research on these issues and strive to solve them as soon as possible.

References

[1] M.H.Yang, D.Krigman, N.Ahuja. Detecting Faces in Images: A survey [J], IEEE Trans. On Pattern Analysis and Machine Intelligence, 2002, 24(1):34-58.

[2] Chellappa R, wilson C L, Sir ohey S. Human and machine recognition of faces [J]. Proceedings

of IEEE, 1995,83(5):705-740.

[3] M. D. Kelly. Visual identification of people by computer. Tech report [J], AI-130, Stanford AI Project, Stanford, CA, USA,2004,36(12),105-112.

[4] T. Canada, Picture Processing by Computer Complex and Recognition of Human Faces [J]: PhD thesis edition Kyoto University, 1973,36(18),109-112.

[5] Jing Huang, Yunyi Shang, Hai Chen. Improved Viola-Jones face detection algorithm based on HoloLens[J]. EURASIP Journal on Image and Video Processing, 2019(1).

[6] Yan Ke. Research on Face Detection Algorithm Based on Skin Color Model [D], Wu Han: Huazhong University of Science and Technology, 2008. (In Chinese)

[7] Peng Huo. Research on the Status and Development of Face Detection Technology [J]. Science and Technology Information,2012(24):137. (In Chinese)

[8] Liangmin Wang, Jianming Zhang, Yongzhao Zhan, etc. Research Status and Development of Face Detection Research[J]. Journal of Jiangsu University: Natural Science Edition,2003,(03):75-79. (In Chinese)

[9] Verma, A, Raj, S.A, Midya, A, Chakraborty, J. Face detection using skin color modeling and geometric feature[R]. Informatics, Electronics & Vision (ICIEV), 2014 International Conference on,2014.

[10] Teja, M.H. Real-time live face detection using face template matching and DCT energy analysis[P]. Soft Computing and Pattern Recognition (SoCPaR), 2011 International Conference of,2011.

[11] Li Chunming*a,b , Li Yushana, Wu Ruihongb, Li Qiumingc, Zhuang Qingdeb, Zhang Zhanb a Institute of Electronic CAD, Xidian University, Xian, 710071,China, b Hebei University of Science and Technology, Shijiazhuang 050054, China, cXingtai Electronic Instrument Factory, 054000, Xingtai, China. A Higher-Order Statistics-Based Approach to Face Detection. Proceedings of Photonics Asia 2004:Information Processing and Data Storage——5637volume[C]. 2004:8.

[12] Zhangqi Bian, Xuegong Zhang, etc. Pattern Recognition (Second Edition) [M], Beijing: Tsinghua University Press, 2000: 207-208. (In Chinese)

[13] Shi Lei. A Face Detection Algorithm in Color Image Based on Skin Color Segmentation. Information Engineering Research Institute, USA. Proceedings of 2013 2nd International Conference on Mechanical Properties of Materials and Information Technology (ICMPMIT 2013) [C]. Information Engineering Research Institute, USA: Intelligent Information Application Technology Association, 2013:5.

[14] Jizeng Wang, Hongmei Yang. Face Detection Based on Template Matching and 2DPCA Algorithm[P]. Image and Signal Processing, 2008. CISP '08. Congress on,2008.

[15] Kherchaoui S, Houacine A. Face detection based on a model of the skin color with constraints and template matching[P]. Machine and Web Intelligence (ICMWI), 2010 International Conference on,2010.

[16] Congqin Yi, Li Li. Face detection based on skin color segmentation and color images [J]. Journal of Guizhou University of Technology, 2003, 32(4): 65-68. (In Chinese)

[17] Samruddhi Y. Kahu, Rajesh B. Raut, Kishor M. Bhurchandi. Review and evaluation of color spaces for image/video compression [J]. Color Research & Application,2019,44(1).

[18] Ruiqi Gong, Jianping Sheng. Conversion from Munsell color space to RGB [J]. Industrial Control Computer, 2019, 32 (01): 85-86.

[19] Honglin Zhang. Visual C++ digital image pattern recognition technology and engineering practice [M]. Beijing: People's Posts and Telecommunications Press, 2003. (In Chinese)

[20] Chi Lu. Research and implementation of face detection algorithm based on skin color segmentation [D]. Northeastern University, 2012. (In Chinese)

[21] Wei Li, Qinghua Yang, Xianbo He. Face Detection Algorithm based on Double Ellipse Skin Model. IEEE Beijing Section、IEEE China Council, Beijing University of Technology, Beijing University of Posts and Telecommunications, The Hong Kong Polytechnic University, Andhra University. Proceedings of 2011 IEEE 2nd International Conference on Software Engineering and Service Science (ICSESS 2011) [C]. IEEE Beijing Section、IEEE China Council、Beijing University of Technology、Beijing University of Posts and Telecommunications、The Hong Kong Polytechnic University、Andhra University: IEEE BEIJING SECTION,2011:5.

[22] Laurent Najman, Hugues Talbot (Eds). Mathematical morphology: from theory to applications[M]. ISTE-Wiley. ISBN 978-1-84821-215-2. (520 pp.) June 2010.

[23] Lu JW, Tan YP. Improved discriminant locality preserving projections for face and palm print recognition [J]. NEUROCOMPUTING, 2011, 74(18): 3760-3767.

[24] Wang Jian, Zhao Honglian. Eye detection based on multi-angle template matching[R]. Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on,2009.

[25] Lei Wang, Yulong Mo, Feihu Qi. Eye feature extraction based on Hough transform and eyelid elastic template [J], Journal of Infrared and Millimeter Waves. 1999, 18(1): 53-60. (In Chinese)

[26] Yueyuan Zhang, QingHua Zeng, Jianye Liu, Yineng Li, etc. Improved image edge detection algorithm based on Canny[J]. Navigation and Control,2019,18(01):84-90.

[27] Jianbo Su, Bo Xu. Introduction to Applied Pattern Recognition Technology: Face Recognition and Speech Recognition [M], 2001.

[28] Xiajun Zhu. Face Recognition Research and Eye Localization Algorithm [D], Zhejiang University of Technology, 2004. (In Chinese)

[29] Sanliang Hong, Qingxiang Wu, Binshu Chen, etc. Eye Localization Based on Maximal Complexity of Rotating Scan Lines[J]. Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education. ISCI 2015.

[30] Min Luo. Research on face detection and facial feature localization based on skin color[D]. Jiangsu University,2008. (In Chinese)

[31] Amalia Luque, Alejandro Carrasco, Alejandro Martín, etc. The impact of class imbalance in classification performance metrics based on the binary confusion matrix [J]. Pattern Recognition, 2019, 91: 216-231.

[32] Junker M., Hoch R., Dengel A.. On the evaluation of document analysis components by recall, precision, and accuracy [M]. IEEE, 1999.