Mask Wearing Detection Method Based on the Skin Color and Eyes Detection

Tengjiao He
School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China
18291261@bjtu.edu.cn

Abstract. During the period of fighting against COVID-19, wearing masks is the main measure of protection in public places. In order to obey relevant regulations, railway stations, airports, shopping malls and other public places are equipped with staff to supervise citizens to wear masks. This method not only wastes human resources, but also has the disadvantage of significantly reducing the efficiency when the flow increases. In order to improve work efficiency, save human resources, and ensure the health and safety of staff, computer vision technology can be used to develop embedded device to deal with the task of automatic mask wearing detection. This paper proposes a mask wearing detection method based on the skin color and eyes detection. This method uses the ellipse skin model, the grayscale gradient features of eyes and the geometric relationship between eyes and other parts of face to locate the face region. And by calculating the coverage of skin color in nose and mouth area, this method gives a judgement of whether a man is wearing a mask properly. This method can basically deal with the task of mask wearing detection under the specific application scenario. It also has the advantages of small volume and strong interpretability.

1. INTRODUCTION
During the period of fighting against COVID-19, wearing masks is the main measure of protection in public places. In order to obey relevant regulations, railway stations, airports, shopping malls and other public places are equipped with staff to supervise citizens to wear masks. This method not only wastes human resources, but also has the disadvantage of significantly reducing the efficiency when the flow increases. Using computer vision technology to develop embedded device can deal with the task of automatic mask wearing detection. This method improves work efficiency, saves human resources and ensures the health and safety of staff.

Mask wearing detection is a specialized task of target detection, that is, the face is automatically located in the camera shooting area and the judgment result of whether the face is wearing a mask is given. Mask wearing detection task can be decomposed into two steps: face detection and judging whether the mask is worn according to the information given by the face area, in which, face detection is a more critical step.

Nowadays, face detection methods can be roughly divided into two categories: statistical model-based methods and knowledge-based methods[1][2].

Statistical model-based face detection method, especially machine learning method, which is also the main method of current mask wearing detection field, including Eigenface algorithm[3], artificial
neural networks algorithm, etc. Among them, deep learning algorithm is the most popular research direction in current mask wearing detection field.

Among the mask wearing detection methods based on deep learning algorithm, the representative algorithms include YOLO series algorithms\cite{4}\cite{5}\cite{6}\cite{7}, SSD algorithm\cite{8}\cite{9}, Retinaface algorithm\cite{10}, MTCNN algorithm\cite{11}, DFS algorithm\cite{12}, etc. In order to prevent overfitting and enhance the ability of generalization, deep learning algorithms always require large amounts of training data and long training time, and the model is of large volume and weak interpretability. Since most terminal devices only have small memory and low computation capacity, such models are developing towards lightweight\cite{9}.

Knowledge-based face detection method is to extract the basic features of human face combining with the priori knowledge of human face to detect human face. The basic features that can be used are grayscale feature, structural feature, texture feature, contour feature and skin color feature, among which the skin color feature is more reliable and stable than other features. Skin color features do not depend on facial details and racial differences, can be easily distinguished from most background objects, and can be detected quickly. Face detection combined with skin color features can quickly eliminate the background objects and narrow the search range, thus improving the speed and accuracy of detection\cite{1}.

To get rid of the dependence of the data set, avoid the consumption of computational resources caused by model training, simplify model structure and improve the speed of detection, this paper proposes a mask wearing detection method based on the skin color and eyes detection. This method uses the ellipse skin model, the grayscale gradient features of eyes and the geometric relationship between eyes and other parts of face to locate the face region. And by calculating the coverage of skin color in nose and mouth area, this method gives a judgement of whether a man is wearing a mask properly. This method requires no training, has small volume and strong interpretability, and can be easily deployed on embedded terminal devices.

2. RELATED WORK

2.1. Illumination Compensation

In the process of image acquisition and transmission, the color information of the image can be easily affected by the color and direction of light, the image acquisition equipment and many other factors, resulting in the deviation of skin color in the image. Considering the influence of color deviation on the accuracy of skin color detection, this method adopts the illumination compensation method proposed by R.L. Hsu M.A. Mottaleb, and A.K. Jain. The steps are as follows\cite{13}:

1. Firstly, the image is converted to gray space and the gray value of each pixel in the image is calculated.
2. Sort all pixels in the image according to the gray value from high to low;
3. The first 5% pixels is extracted and if the number of these first 5% pixels is greater than a certain threshold, they are linearly amplified to 255. The coefficients used here are called the illumination compensation coefficients.
4. The R, G and B components of each pixel in the image are linearly amplified by the illumination compensation coefficients.

Considering that the color outside the iris is definitely white, it’s reasonable to set the gray value of the pixels with the largest grayscale to 255, this method has a good effect on correcting the color deviation caused by dim light and makes it easier to separate skin area from background.

2.2. Color Space Transformation

The selection of color space affects the selection of skin color model. Generally, the property of a certain color space can be learned from the following two aspects\cite{14}:

- Whether a given model can be used to describe the distribution of the "skin color" area in the color space;
- How much overlap there is between skin color and non-skin color in the color space.
RGB color space, rgb color space, YCbCr color space, YUV color space, and HSV color space are often used in Face detection\cite{15}\cite{16}.

The model of the RGB color space is based on a Cartesian cube, with three axes corresponding to the three primary colors of red, green and blue specified by the International Council of Illumination. Other color spaces can be converted by the RGB color space. A certain color can be represented by a normalized vector \((r, g, b)\), with values in the range \([0, 1]\).

In HSV color space, the \(V\) component represents lightness, \(H\) component represents chromaticity, and \(S\) component represents saturation. The following formula can be used to convert an RGB space image to an HSV space image:

\[
\begin{align*}
H &= \arccos \frac{[(R-G)+(R-B)]/2}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \\
S &= 1 - \frac{3 \min(R, G, B)}{R + G + B} \\
V &= \frac{R + G + B}{3}
\end{align*}
\]  \( (1) \)

YCbCr color space is derived from YUV color space, which is mathematically equivalent to YUV color space. Where \(Y\) component represents brightness, \(C_b\) represents blue component, and \(C_r\) represents red component. The following formula can be used to convert an RGB space image to an YCbCr space image:

\[
\begin{bmatrix}
Y \\
C_b \\
C_r
\end{bmatrix} =
\begin{bmatrix}
0.2990 & 0.5870 & 0.1140 \\
-0.1687 & -0.3313 & 0.5000 \\
0.5000 & -0.4187 & -0.0813
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix} +
\begin{bmatrix}
128 \\
128 \\
128
\end{bmatrix}
\]  \( (2) \)

Generally, images collected by cameras are RGB color space images. In order to reduce the influence of illumination on skin color detection, it is supposed to choose color space that can separate the chroma information from the brightness information\cite{15}.

Although both YCbCr color space and HSV color space can separate chromaticity from brightness, skin color clustering is not so satisfactory in the low saturation range of HSV color space\cite{15}, while YCbCr color space has the following advantages\cite{17}:

- Its construction principle is similar to the human visual perception process.
- The calculation process and the representation of spatial coordinates in YCbCr color space are simpler than those in HSI or other color space.
- Skin color clustering in YCbCr color space is satisfactory, and the overlap between skin color and non-skin color is quite small.

Considering the above advantages, the method proposed by this paper chooses to convert RGB color space images into YCbCr space images.

2.3. Skin Color Model

In order to separate skin color region from non-skin color region, the pixels in YCbCr color space need to be discriminated by skin color model first, and then the image's gray value is binarized -- the gray value of skin color region is set to 255, and the gray value of non-skin color region is set to 0.

Commonly skin color models used in face detection include: the simple threshold model, the Gaussian model\cite{15} and the ellipse skin model with nonlinear color transformation\cite{16}.

The simple threshold model is easy to calculate but hard to determine the value of thresholds. The histogram model doesn’t need color space transformation but requires a large number of skin color pixels and non-skin color pixels to enrich the dataset and its detection accuracy is greatly affected by the size of dataset. The Gaussian model is built on the assumption that chromaticity and brightness are uncorrelated, but the fact is chromaticity and brightness are in nonlinear relation with each other in skin
color region, which reduces the detection accuracy of the Gaussian model. The ellipse skin model with nonlinear color transformation takes the influence of brightness into account, and its accuracy is higher than that of the Gaussian model. And it doesn’t need dataset of skin color pixels and non-skin color pixels, so it is much simpler to implement than the histogram model\[16\].

To sum up, the method proposed by this paper chooses the ellipse skin model with nonlinear color transformation as skin color module. The following is the description of this model:

In order to reduce the influence of brightness on skin color clustering, Rein-Lien Hsu M.A.Mottaleb, and A.K.Jain set piecewise linear edges for skin color clustering. That’s how an YCbCr color space image is converted into an YCb’Cr’ space image\[13\]:

\[
C(Y) = \begin{cases}
(C(Y) - \overline{C}(Y)) \frac{W_c}{W_c(Y)} + \overline{C}(Y) & \text{if } Y < K_i \text{ or } Y > K_i \\
C(Y) & \text{if } Y \in [K_i, K_s]
\end{cases}
\]  

(3)

\[
W_c(Y) = \begin{cases}
W_{Cl} + \frac{(Y-Y_{max})(W_{Cr} - W_{Cl})}{K_{r} - K_{s}} & \text{if } Y < K_i \\
W_{Cl} + \frac{(Y_{min} - Y) (W_{Cr} - W_{Cl})}{Y_{min} - K_{s}} & \text{if } Y > K_s
\end{cases}
\]  

(4)

\[
\overline{C}_b(Y) = \begin{cases}
108 + \frac{(K_i - Y) (118 - 108)}{Y_{min} - K_{s}} & \text{if } Y < K_i \\
108 + \frac{(Y - K_s) (118 - 108)}{Y_{min} - K_{s}} & \text{if } Y > K_s
\end{cases}
\]  

(5)

\[
\overline{C}_r(Y) = \begin{cases}
154 + \frac{(K_i - Y) (154 - 144)}{Y_{min} - K_{s}} & \text{if } Y < K_i \\
154 + \frac{(Y - K_s) (154 - 132)}{Y_{min} - K_{s}} & \text{if } Y > K_s
\end{cases}
\]  

(6)

In Equations (3) -- (6), \(C\) is \(C_0\) or \(C_\theta\), \(\overline{C}_b(Y)\) and \(\overline{C}_r(Y)\) respectively represent the two central axes of ellipse region and \(W_c(Y)\) represents the width of skin color region. The other constants are: \(K_i = 125\), \(K_s = 188\), \(Y_{min} = 16\), \(Y_{max} = 235\), \(W_{Cb} = 46.97\), \(W_{Cr} = 38.76\), \(WL_{Cb} = 23\), \(WL_{Cr} = 20\), \(WH_{Cb} = 14\), \(WH_{Cr} = 10\).

Skin color was clustered in Cb’ and Cr’ space to form an elliptical model:

\[
\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1
\]  

(7)

\[
\begin{bmatrix}
x \\
y
\end{bmatrix} = \begin{bmatrix}
\cos \theta & \sin \theta \\
-\sin \theta & \cos \theta
\end{bmatrix} \begin{bmatrix}
C_b - c_x \\
C_r - c_y
\end{bmatrix}
\]  

(8)

Among which \(C_x = 109.38\), \(C_y = 152.02\), \(\theta = 2.53^\circ\), \(ec_x = 1.60\), \(ec_y = 2.41\), \(a = 25.39\), \(b = 14.03\).

2.4.Morphological Processing

In order to improve the connectivity of skin color region and eliminate some noise, it is necessary to carry out morphological processing on the image. Morphological processing consists of four basic operations, which are dilation, erosion, opening and closing. Dilation can fill small holes in the connected domain and small depressions at the edge of the connected domain. Erosion can eliminate small protrusions at the edge of the connected domain. Opening means eroding first and then dilating, which can eliminate some noise, separate areas at narrow gaps and eliminate protrusions. Closing means dilating first and then eroding, it can make the contour of connected domain smooth, fill small holes and eliminate narrow gaps.
2.5. Eyes Detection

Skin color detection may mark areas similar to skin color (but actually not) in the image like background objects or other parts of human body (not face). In this case, eye detection is needed to exclude skin color areas that are not human face.

Because the grayscale of iris is quite low, while the grayscale of sclera is quite high, a dramatic change of grayscale in the junction of the two parts leads to a high gray gradient value. It’s reasonable to select those points in the image that satisfy both the gray value less than a certain threshold in the original image and the gray gradient greater than a certain threshold in the gradient image as “most likely to be eyes” points[18].

Considering that the grayscale of hair, glasses, masks and other ornaments may affect the detection accuracy, it’s a must to specify eye search areas[19]: in those “likely to be face” regions, the right eye search area is a rectangle 0.08-0.36 times the width from the left boundary and 0.48-0.78 times the length from the lower boundary, and the left eye search area is a rectangle 0.64 to 0.92 times the width from the left boundary and 0.48 to 0.78 times the length from the lower boundary.

After the operations above, those “most likely to be eyes” pair may still not be “true eyes”. Since there is a certain geometric relationship between eyes and other parts of the face, the following formulas can be used to judge whether a “most likely to be eyes” pair is “true eyes” or not[16]:

\[
\frac{W_{\text{eyes}}}{W_{\text{face}}} \geq Tl \land \frac{W_{\text{eyes}}}{W_{\text{face}}} \leq Ti
\]

\[
|h_{\text{eyes}}| \leq \theta h
\]

If Equations (9) and (10) are both satisfied, this “likely to be face” region is considered to have eyes, so this region is determined to be a region with face. Otherwise, this region is considered to have no eyes, so this region is determined to be skin-like region or non-facial skin region and will be excluded.

3. MASK WEARING DETECTION MODEL BASED ON THE SKIN COLOR AND EYES DETECTION

This paper proposes a mask wearing detection model based on skin color and eyes detection. The specific steps are as follows:

(1) Firstly, illumination compensation is performed on the input image collected by the camera.

(2) Then convert the image from RGB color space to YCbCr color space.

(3) The image’s gray value is binarized by the ellipse skin model with nonlinear color transformation.

(4) After filling some small holes in the grayscale-binarized image, carry out morphological processing on the image.

(5) Each skin color area is framed with a box whose aspect ratio is 1.3 as “likely to be face” region.

(6) For each “likely to be face” region, start from the region with the largest area to perform eyes detection, excluding skin-like regions and non-facial skin regions.

(7) For regions determined to have human face, determine whether the face is properly worn with a mask according to the degree of exposure of the skin area of nose and mouth.

The flow chart of mask wearing detection model is shown below:

![Flow chart of mask wearing detection model](image)

The steps are explained in detail below.
3.1 Development Environment and Data
The computer with 8-core Intel Core i5 8250U CPU is chosen as our Hardware development environment.
MATLAB that has perfect image processing function library is chosen as our software development environment.
Image data comes from the Real-World Masked Face Dataset (RMFD), it is the fruits of the national engineering research center for multimedia software, Wuhan university sponsored activities. Its validation set contains 4015 face images of 426 people with or without masks, which are 250×250 pixels in size, with complex background and various poses. This model is designed to deal with mask wearing detection applied to entrance gates in public places. Therefore, it is supposed to choose images that meets the following criteria from the above validation set for testing: single face, simple background, eyes visible from the front, with or without masks.

3.2 Illumination Compensation
The input image collected by the camera is performed with illumination compensation method mentioned in 0. The results are shown in the following figures, in which the left is the original picture, and the right is after illumination compensation. It can be seen that illumination compensation solves the problem that skin color looks dark caused by dim light, making skin color easier to be recognized by skin color model:

![Figure 2 Sample of illumination compensation.](image)

3.3 Color Space Transformation
Equation (2) is used to convert the image from RGB color space to YCbCr color space. The results are shown in the figures below:

![Figure 3 Sample of color space transformation.](image)

3.4 Binarization
Equations (3) -- (6) are used to perform nonlinear transformation on YCbCr image and traverse each pixel on the image to calculate whether it conforms to the ellipse skin model. The gray level of pixels that conform to the model is set to 255, and the gray level of pixels that do not conform to the model is set to 0. The results are shown in the figures below, with the white areas as the skin color regions and the black areas as the background:

![Figure 4 Sample of binarization.](image)
3.5. Morphological Processing
In order to enhance the connectivity of the skin color regions and separate the skin color regions from the background, small holes were filled, closing and erosion operations were performed on the grayscale-binarized image. The results are shown in the figures below. From top to bottom, they are holes filling, closing and erosion. It can be seen that some noise is eliminated and the connectivity of large skin color regions is enhanced:

![Morphological processing](image)

Figure 5 Sample of Morphological processing.

3.6. Frame Out “Likely to be Face” Regions
Finding the smallest bounding box for each skin color region. Considering that wearing a mask would cause the loss of skin color area, the “most likely to be face” region is formed by taking 1.3 times the width of the bounding box as the height of the “most likely to be face” region and the width of the bounding box as the width of the “most likely to be face” region. The results are shown in the figure below. The green boxes that frame out skin color regions are the “most likely to be face” regions:

![Likely to be face regions](image)

Figure 6 Sample of “likely to be face” regions.

3.7. Eyes Detection
For each “most likely to be face” region, eyes detection is performed from the region with the largest area by using the method mentioned in 0, excluding skin-like regions and non-facial skin regions. The results are shown in the figures below. The blue boxes represent eye search areas, and the locations of both eyes are marked by red crosses:
3.8. Mask Wearing Detection
For regions determined to have human face, wearing the mask properly means that both mouth and nose are covered. Therefore, it can be judged whether the face is completely wearing a mask according to the degree of exposure of the skin area of nose and mouth. The nose area is 47% to 67% times the height from the upper boundary of the face box and is the same width as that of face box, and the mouth area is 67% to 95% times the height from the upper boundary of the face box and is the same width as that of face box. Traversing all the pixels in the nose area and mouth area, calculate the percentage of skin color contained in the corresponding areas. When the percentage of skin color contained in the nose area exceeds 30% or the percentage of skin color contained in the mouth area exceeds 25%, this part can be considered to be exposed to the air\(^4\). The results are shown in the figures below, in which the green box means wearing a mask properly, while the red box means not wearing a mask or wearing a mask improperly:

![Figure 8 Results of mask wearing detection.](image)

3.9. Analyses of Experimental Results
According to the results of the experiment above, this method has been proved to be able to deal with the task of mask wearing detection under the specific application scenario. This method can also be used in images with glasses or large non-facial skin color areas. As shown in the figures below, the left is the detection result of wearing a mask improperly, and the right is the detection result of wearing glasses with large non-facial skin color area:

![Figure 9 Results under special condition.](image)

But images with uneven illumination (strong light shines on one side or the upper half of the face in shadow) or forehead obscured are prone to lose target, give wrong judgement or wrong location of face. As shown in the figures below, Fig. 10 shows the detection result under the condition of uneven illumination, and Fig. 11 shows the detection result under the condition of forehead obscured.
4. CONCLUSION
In order to help the epidemic prevention and control and deal with the task of automatic mask wearing detection in public places, this paper proposes a mask wearing detection method based on the skin color and eyes detection. According to the analysis, the advantages of this model are as follows:

(1) This method has been proved to be able to deal with the task of mask wearing detection under the specific application scenario;

(2) There is no model training process in this method, which means it gets rid of the dependence of data set and avoids the consumption of computational resources caused by model training;

(3) Compared with deep learning model, the total number of parameters used in this model is very small, so it occupies less memory, has a faster detection speed, and is easier to be deployed on embedded terminal devices.

(4) Based on the priori knowledge of skin color detection and eyes detection, the parameters in this model have clear meanings, so this model is highly interpretable;

In view of the problems caused by uneven illumination, the illumination compensation method needs to be optimized, to put it specifically, algorithms like adaptive illumination compensation or Retinex that can reduce the influence of uneven illumination can be taken.

REFERENCES
[1] E Li. Overview of Face Detection Methods [J]. Information Technology & Informatization, 2018(4):24-26.
[2] Yi Ning, Jinlong Chen. Overview of Face Detection Based on Internet Videos [J]. Computer Knowledge and Technology, 2018(3):166-167.
[3] Yaru Zhu, Jingjie Wang, Wenqing Li, Fang Zhao, Qing Ye. An Intelligent Monitoring System for Mask Wearing Based on Eigenface Algorithm [J]. China Science and Technology Information, 2021(2):109-110,112.
[4] Junjie Xiao. Masked Face Detection and Standard Wearing Mask Recognition Based on YOLOv3 and YCrCb [J]. computer engineering & Software, 2020(7):164-169.
[5] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society, 2021(65).
[6] Junlin Guan, Xin Zhi. Mask Wearing Detection Method Based on YOLOv4 Convolutional Neural Network [J]. Modern Information Technology, 2020(11):9-12.
[7] Feng Wang. Improved yolov5 Mask and Helmet Wearing Artificial Intelligence Detection and Recognition Algorithm [J]. Construction and Budget, 2020(11):67-69.
[8] Preeti Nagrath, Rachna Jain, Agam Madan, Rohan Arora, Piyush Kataria, Jude Hemanth. SSDMNV2: A real-time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustainable Cities and Society, 2021(66).
[9] Baihan Zou, Yating Qin, Xiaobing Yuan, Xuesong Zhang. Research on Mask Face Detection Based on Lightweight CNN [J]. Computer Engineering & Software, 2020(8):186-188.
[10] Zuodong Niu, Tao Qin, Handong Li, Jinjun Chen. Improved Algorithm of RetinaFace for Natural Scene Mask Wear Detection [J]. Computer Engineering and Applications, 2020(12): 1-7.
[11] Yuanda Wang. UCloud Open Face Mask Detection Service with AI Algorithm to Accelerate Epidemic Prevention and Control [J]. Communications World, 2020(5):33-34.
[12] Xiubao Zhang, Ziyuan Lin, Wanxin Tian, Yan Wang, Haifeng Shen, Jieping Ye. Mask-wearing recognition in the wild. Science in China (Information Sciences), 2020(7):1110-1120.
[13] R.L. Hsu, M.A. Mottaleb, and A.K. Jain, Face detection in color images. IEEE Trans. Pattern Analysis Machine Intelligence, 2002, 24(5); 696—706.
[14] Luhong Liang, Haizhou Ai, Guangyou Xu, Bo Zhang. A Survey of Human Face Detection [J]. CHINESE JOURNAL OF COMPUTERS, 2002(5):449-458.
[15] Jing Zhao. Research on Face Recognition Technology Based on Skin Color [D]. Tianjin: Tianjin University of Technology, 2013.
[16] Yongguang Cheng. Research on Face Detection Algorithm Based on Skin Color and Facial Geometric Features [D]. Beijing: Beijing Jiaotong University, 2008.
[17] S.L. Phung, D. Clmi, A. Bouzcrdounm. A novel skin color model in YCbCr color space and its application to human face detection. IEEE International Conference on Image Processing, 2002.
[18] Xuewu Wang. Face Detection Based on Eye Gradient Feature and Application of Fisher Face Recognition Technology [D]. Hunan: Xiangtan University, 2005.
[19] Ja Wei. Research on the Human Face Detection Based on Skin-Modeling and Eye Brightness Detection [J]. Software Guide, 2012(11):127-128.