# Mask Detection and Classification

Eli Clark
*School of Engineering and Computer Science*
*Syracuse University*
Syracuse, USA
eaclark@syr.edu

Jack Guida
*School of Engineering and Computer Science*
*Syracuse University*
Syracuse, USA
jaguida@syr.edu

***Abstract*—This paper presents a proposed method to detect whether or not a person is wearing a facemask or covering from a digital image. The method described in this paper uses facial feature extraction to find regions of interest (ROI) in an image corresponding to certain facial features. The face is first passed to a canny edge detector. When a ROI is found for the eye, a Hough transform is taken on the edge image to find the most significant edges found. When a mask is being worn, a strong edge will regularly be found just below the eyes. Searching just below this ROI for a roughly horizontal Hough line is a sufficient case for the image to be classified as having a mask wearer. When only a face ROI is detected, or if the Hough transform analysis did not result in an image being found, the face is segmented into an upper and lower region. Using the average and median color values, the distances between the top and bottom regions of the segmented face are compared. A significant color difference indicates that the person is most likely wearing a mask.**

***Index Terms*—Canny edge detection, Hough transform, ROI, Viola-Jones, Mask Detection, RGB Intensity**

## I. Introduction

The COVID-19 pandemic has drastically changed the social aspects of daily life. Proven by abundant amounts of resources, wearing a mask is a simple yet effective way to slow transmission of the virus and help maintain the safety of oneself and others. A mask covering the nose and mouth are required in almost every public place. There are 3 key classifications to keep in mind in these social settings: 1) a person is wearing a mask 2) a person is not wearing a mask(including a mask being worn below the mouth) and 3) the image is a non-face image. The key features of this algorithm used to classify a mask are based on the geometry of the face, creating different regions of interest (ROI), RGB properties of the face and mask, and the physical properties of a mask being worn correctly. This algorithm implements a safety and health related system without the use of any deep learning networks.

### A. Related Work

Since the COVID-19 pandemic is relatively new, and there is not a large amount of research into mask detection. The most popular, and the only methods we have found are based on deep learning neural network approaches [7]. With a neural network based approach, training the network and learning to train the network brings the most difficulty. There needs to be copious amounts of data to successfully train the network effectively. These neural networks classify if someone is wearing a mask or not but we have not found any algorithms that aren't based on deep learning networks. There has been an abundance of research into facial recognition and extracting facial features from images. One of the methods for this is the Viola-Jones algorithm that can detect facial features. One of the disadvantages to this algorithm is that this algorithm results in a relatively high false positive rate. There are other methods of facial feature extraction that focus on splitting the face up into different regions based on the geometry of the face [1]. To accomplish any processing/recognition of the face, one of the most important steps is isolating the face from the rest of the image. With all of the research going into image processing regarding facial recognition/extraction, we have not found any significant data in the detection and classification of masks using a non-neural network approach.

## II. Proposed Method

The proposed method has three main elements. The algorithm begins by using Viola-Jones facial feature detection to find ROI's. The two types of ROI's used in this method are eye and face regions. When a ROI is detected for eyes in the image, the edges throughout the image are found and significant edges are detected using the Hough Transform. These Hough lines are used to determine if there is a mask line just below the eyes. The third stage of this algorithm consists of finding the difference in color intensity between two separate regions of a face.

### A. Facial Feature Detection

The algorithm starts by using key geometries of the face coupled with the Viola-Jones algorithm [11]. This algorithm is a trained classifier that uses Haar features to look for features specific to a face. These Haar features are shown in Figure 1. The Viola-Jones algorithm can detect many facial features including: faces, upper bodies, eye/eye pairs, noses, and mouths. This algorithm is known for having a significant false positive rate, indicating the presence of faces or facial features without there being any in a given image [11]. In the interest of identifying a face wearing a mask, this algorithm fails to recognize faces given that a person is wearing a mask correctly. Our proposed algorithm starts with recognizing the eyes. However, since Viola-Jones facial feature detection does not work perfectly, an eye region can not always be found. When this is the case, a region of interest is attempted to be found. In the data set used for the testing of this algorithm

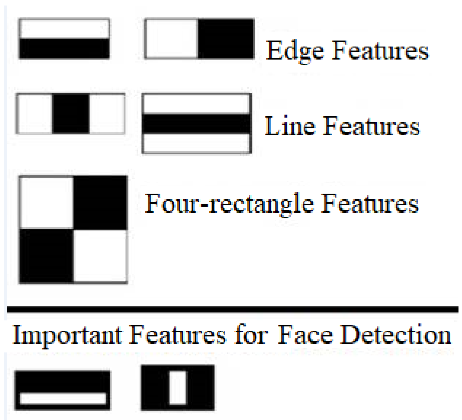


Fig. 1. Haar Features.

every image where the person was not wearing a mask at least of the two aforementioned ROIs was detected. Therefore, if no ROI was detected then the image is classified as a mask wearer. However, additional testing must be done for this conclusion to be made accurately.

### B. Edge Detection and Hough Transform

Understanding the geometry of the face when wearing a mask correctly allows for the implementation of this algorithm. If a person is wearing a mask on their face correctly, it will look similar to Figure 2. The eyes will be visible and the mask will rest right below the eyes on the person's nose. This creates two regions of interest (ROIs) - above the eyeline and below the eyeline. The region below the eyeline should be primarily composed of the mask and above will be eyes and a forehead. The mask creates a boundary on the face that should result in an edge which can be detected by the Canny edge detector. This edge detector starts by removing noise from the image using a gaussian filter. It then finds the edge strengths in both the x and y directions by convolving the output from the Gaussian filter, J, with two kernels $K_x$ and $K_y$. This gives the horizontal and vertical edge strengths at each pixel, contained in $J_x$ and $J_y$ respectively. These equations are given below.

Fig. 2. Person Wearing Mask.

$$J_x = J * k_x \quad and \quad J_y = J * k_y$$

$$where \quad k_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \quad and \quad k_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

The overall edge strength at every pixel, es, can be found by finding the magnitude of the vertical and horizontal edge strengths using the following equation.

$$e_s(i,j) = \sqrt{J_x^2(i,j) + J_y^2(i,j)}$$

This edge strength image is then run through a non-maximum suppression algorithm with the function of thinning the edges to be 1-pixel wide based on the edge orientation image, eo calculated by [12].

$$e_o(i,j) = tan^{-1}\frac{J_x(i,j)}{J_y(i,j)}$$

These edges will include some fine grained details within the image that are unnecessary and unwanted for the purposes of mask detection. Understanding the physical nature of a mask makes it clear that there will be a relatively strong edge at the face-mask boundary on the person in the image. The edge image will be binarized based on a threshold found empirically. The threshold used to sort out weak edges was determined based on the number of non zero edge strength pixels received from the non-maximum suppression algorithm. The $90^{th}$ percentile of the non-zero edge strength value was selected through empirical testing. Figure 3 shows the binarized edge image and shows how there is a clear and distinct edge going across the face.This image also shows how the mask will create a primarily horizontal edge across the face. This output is still an image only being described by its pixels without categorizing its information. This is where the algorithm will take advantage of the Hough transform. The Hough transform is a technique in image processing used to

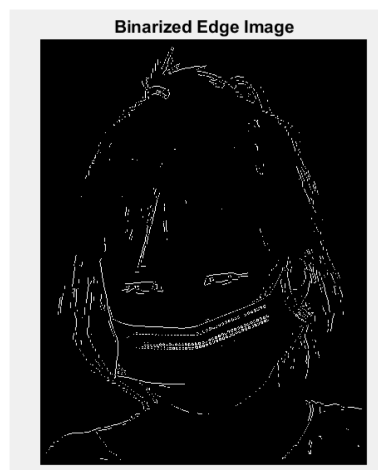


Fig. 3. Binarized Edge Image.

identify lines. Because the edge from the mask is primarily a straight line - or has regions that will represent a straight line - the Hough transform will be able to identify unique lines. This transforms takes a line in the form $y = mx + b$ and transforms it into the polar form by representing this line by its distance from the origin, r, and angle from the positive x axis, $\theta$ - the Hough space. A single point can be mapped in the Hough space by showing every possible line going through that point, represented by different r and $\theta$ values. If this is done for multiple points, the spot where these two lines in the Hough space intersect represents the parameters of the line that would connect these points. This is done for every pixel in the image and the Hough space is discretized by creating bins. A voting system is created and the bins containing the highest frequency of occurrences represent the strongest lines.

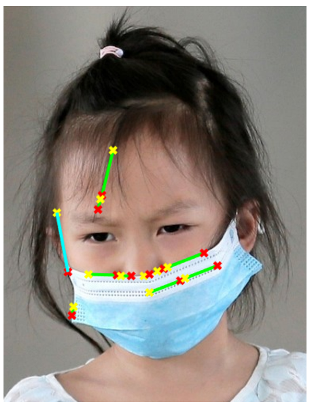

Fig. 4. Hough Lines on a Masked Person.

Fig. 5. Hough Lines on a Non-Masked Person.

Figures 4 and 5 show the distinction between the Hough lines in the previously defined ROI and will lead to the classification of a mask. These Hough lines will be checked to make sure that it overlaps/intersects with the edge detected from the Canny edge detector. As mentioned before, the edge indicating the presence of a mask will be primarily horizontal. Therefore, the algorithm will get rid of the Hough lines that fall in the range of having a theta value greater than $|45°|$

### *C. Color Intensity*

Due to lighting effects, many different types and colors of masks, or just not being able to detect an eye ROI for the image, a sufficient accuracy was not accomplished by the Hough transform method alone. To account for these differences the final stage of this algorithm separated the image into two separate regions, a lower and an upper region. And finds the difference in RGB intensities for these two areas. The median and mean values are calculated for both of the regions. The upper region from just below the eyes, to near the top of the forehead provides a good approximation of the person's skin tone. If the person is not wearing a mask the lower regions RGB intensity will regularly be much closer to the upper regions RGB intensity, than if the person is wearing a mask. Two different methods were used to get the upper and lower regions that were analyzed. If an eye region was detected, then that ROI's dimensions were used to create these two regions. And when only a face region was detected then the faces ROI was used to segment the image. Using the eye region, the upper and lower segments both had the same width as the eye ROI. The top of the lower segment and bottom of the upper segment are defined by the bottom line of the eye region detected using the Viola-Jones algorithm. The top of the upper segment is at a distance from this center line of twice the height of the eye region, and the bottom of the lower region is at this same distance below the center line. These dimensions were selected so the upper region contains the forehead to extract skin tone, and the lower contains the bottom half of the face, where a mask may be. A sample of these segments can be seen in figure 6.

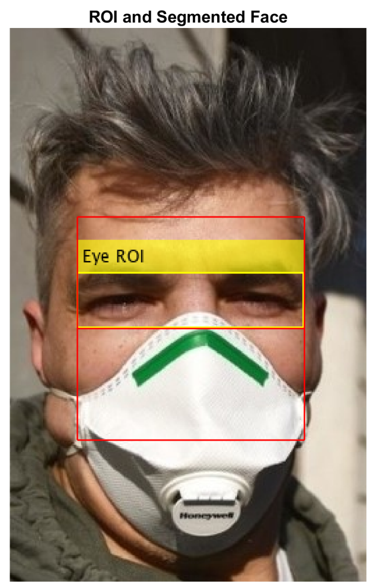


Fig. 6. Segmented Face Using Eye ROI.

When only a face region was detected a slightly different method is used to separate these two regions since the location of the eyes is unknown. However, this location can be found using the approximate geometry of a person's face[1]. A similar center line is chosen for between the upper and lower sections of the face, just below the eyes. This centerline is a distance of 40% of the height of the face ROI, down from the top. The height of each of the sections of face are 30% of the height of the face ROI, above and below this center line. The width of the two regions is 60% of the width of the face ROI and both boxes are centred horizontally. These values were determined empirically based on how different values cause the segments to appear on a face, and by using figure 7 as reference [1]. A sample of these segments can be seen in figure 8.

This is where the algorithm takes advantage of the RGB properties of the image. As previously explained the methods described prior will create two different ROIs with one containing primarily mask, if one is present, and the other

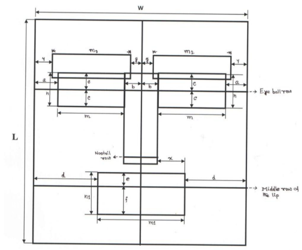

Fig. 7. Geometric Face Pattern.

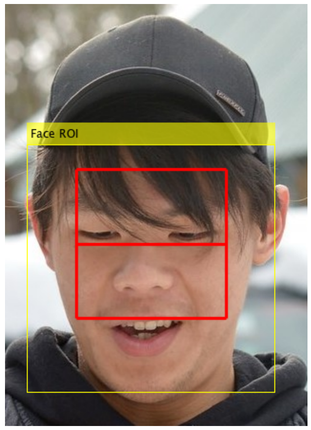


Fig. 8. Segmented Face using Face ROI.

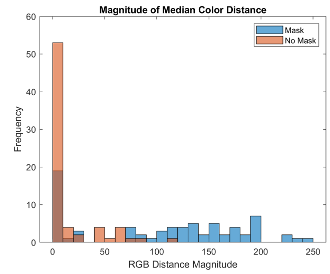


Fig. 9. Magnitude of the Distance Between Median RGB Vetors.

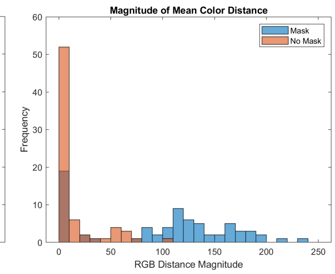


Fig. 10. Magnitude of the Distance Between Average RGB Vetors.

**Algorithm 1** Proposed Algorithm

```
for all images in set do
  if eye ROI detected then
    if Hough lines near ROI then
      Mask detected
    else
      if Average or median color distance ≥ thresh then
        Mask detected
      else
        Mask not detected
      end if
    end if
  else if face ROI detected then
    if Average or median color distance ≥ thresh then
      Mask detected
    else
      Mask not detected
    end if
  else
    Mask detected
  end if
end for
```

making up above the eyes, ie. the forehead. First the average and median RGB values are calculated for each region. These values are used to find the magnitude of the difference in RGB value between the top and bottom regions, using the following equation.

$$|RGB_{dist}| = \sqrt{(R_{top} - R_{bot})^2 + (G_{top} - G_{bot})^2 + (B_{top} - B_{bot})^2}$$

From these values, if there is a value that is greater than an empirically calculated threshold (85 for median and 75 for mean magnitude), the algorithm will mark it as a significant change due to a mask being worn. The histograms shown in figures 9 and 10 display the values calculated using the prior equations for each image in the data set tested, and were used to determine the threshold for the color (RGB intensity) distance. Therefore, if this method is implemented to a larger dataset this threshold could have to be adjusted. As expected, in the images where a mask is not being worn both the mean and median histograms show a lower difference in the two regions, indicating that they are very similar in color. And when a mask is on this value is significantly larger. This also shows that there is a very clear division between the color distance (in both the mean and median) from a mask wearer and a non-mask wearer. The advantage of using both the median and mean is having them both contribute to determine if the person is wearing a mask, that way if one classifies the image incorrectly, it can still be classified correctly overall. This technique is also a way to deal with the unique and colorful masks with patterns on them, which show up as facial features when performing extraction.

The pseudo code for the algorithm described in this section can be seen in algorithm 1.

## III. Results

In order to verify the robustness of the results, a database from Kaggle was used as the input set of images. 100 images from this database, a combination of mask wearers and non-mask wearers, were chosen to represent the data. The algorithm would classify the image and compare it to the ground truth, a given input file to the algorithm. From the results and the ground truth, this led to an accuracy of 84%. This is a drastic increase from the original accuracy of 30% when only searching for the horizontal Hough transform lines below the eye ROI and without segmenting the face into finer detailed ROIs. As seen in figure 11 even if the mask is worn very low on the nose, which is still acceptable, when the eye ROI is extended down to twice its initial height, the strong horizontal lines from the mask to face transition are detected. However, the accuracy rating was very low because many masks were not being detected due to patterns in the mask, or the strongest edges in an image showing up due to the background. Therefore, an additional step of color intensity was measured for the images where an eye or face ROI was detected but no mask was detected from Hough lines. A perfect example of this can be seen in figure 12 where there is a detailed pattern on the mask of this person, in addition to the person's hair being dark. This results in the Hough lines from the image not appearing at the upper masks edge. Also the mask wearer has pretty light skin and is wearing an overall light colored mask, which makes this case especially hard to determine correctly. However, as seen in the title of the image the magnitude of the distance between the mean RGB values is 80.88, which is just above the mean threshold of 75, and is classified as a mask wearer. Another detail that presented some difficulties was when a person was wearing glasses or sunglasses. In this case, the Viola-Jones feature detection algorithm had a difficult time detecting eyes correctly in the image. However, it is still regularly able to detect a face in these images. This is where the image segmentation using a face ROI proved to be especially robust. This process is shown in figure 14, where the subject is wearing glasses, and therefore only a face ROI can be detected. By segmenting the persons face, a split can be made just along the top of the persons mask, revealing a magnitude of the difference between the two regions, for both mean and median, to be much larger than the thresholds specified in this report.

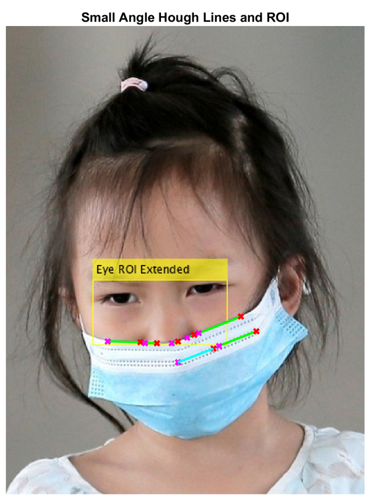


Fig. 11. Extended Eye ROI and Roughly Horizontal Hough Lines.

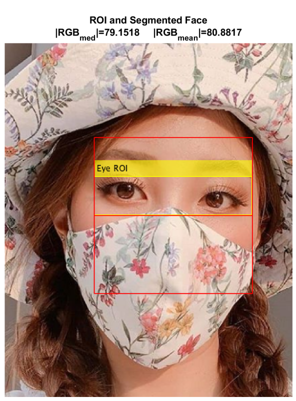


Fig. 12. Mask Detection Using Color Intensity Base on Eye ROI, for a Pattern Mask.


Fig. 13. Non-Mask Detection Using Color Intensity Base on Eye ROI.


Fig. 14. Mask Detection Using Color Intensity Base on Face ROI.

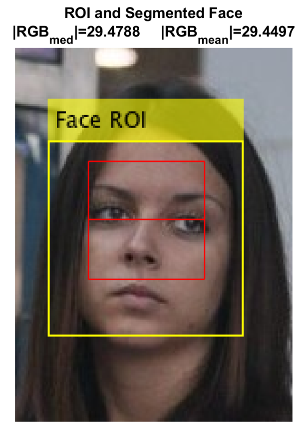


Fig. 15. Non-Mask Detection Using Color Intensity Base on Face ROI.

### A. *Limitations*

The algorithm presented has some limitations in terms of the input image. This algorithm assumes that given there is a face, this face image will be cropped/zoomed to fit only the face. It also assumes that this person is looking head on into the camera. The performance declines as the head is tilted or it is more of a profile image. This algorithm also assumes there is a relatively neutral background. Ideal circumstances would be the face image cropped to only the face region with a white background. A detailed background can cause the Viola-Jones method of facial feature extraction to pick up facial features that do not exist. Given these limitations, it is reasonable to say that the results are effective because of the extensive research into isolating people/faces - if there are any - from the rest of a larger image.

## IV. Conclusion

Our new algorithm, which uses facial features, geometry of the face, and physical properties of wearing a mask, gives effective results in detecting and classifying mask wearers and non-mask wearers. This implementation requires no previously trained models and strongly differs from the neural network based implementations of a mask detection system. For a data set of 100 images combined with mask wearers and non-mask wearings, our algorithm achieved an accuracy of 81%. Future work in this area of mask detection without deep learning networks will be focused on more accurately extracting facial features from images with people wearing masks.

## References


[1] S. Pramanik and D. Bhattacharjee, "Geometric feature based face-sketch recognition," International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), Salem, Tamilnadu, 2012, pp. 409-415, doi: 10.1109/ICPRIME.2012.6208381.

[2] P. R. Police Patil, R. Pramod and S. Sandhya, "A General Approach on Facial Feature Extraction and Face Attributes," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 151-155, doi: 10.1109/CSITSS.2018.8768504.

[3] Liya Ding and A. M. Martinez, ”Precise detailed detection of faces and facial features,” 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1-7, doi: 10.1109/CVPR.2008.4587812.

[4] Anjali , Avinash Kumar , Mr. Ravi Ranjan, “An Algorithm for Face Detection and Feature Extraction,” International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 4, April 2014

[5] Saman Cooray, Noel O’Connor, “Facial Feature Extraction and Principal Component Analysis for Face Detection in Color Images”

[6] A. Majumder, L. Behera and V. K. Subramanian, ”Automatic and Robust Detection of Facial Features in Frontal Face Images,” 2011 UkSim 13th International Conference on Computer Modelling and Simulation, Cambridge, 2011, pp. 331-336, doi: 10.1109/UKSIM.2011.69.

[7] Kulkarni, A., Vishwanath, A., Shah, C., Praveen, V., Wang, Y., Docca, A. and Dinea, C., 2020. Implementing A Real-Time, AI-Based, Face Mask Detector Application For COVID-19 — NVIDIA Developer Blog. [online] NVIDIA Developer Blog. Available at: [Accessed 29 October 2020].

[8] R. Sikarwar, A. Agrawal and R. S. Kushwah, ”An Edge Based Efficient Method of Face Detection and Feature Extraction,” 2015 Fifth International Conference on Communication Systems and Network Technologies, Gwalior, 2015, pp. 1147-1151, doi: 10.1109/CSNT.2015.167.

[9] S. R. Benedict and J. S. Kumar, ”Geometric shaped facial feature extraction for face recognition,” 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, 2016, pp. 275-278, doi: 10.1109/ICACA.2016.7887965.

[10] W. M. K. Wan Mohd Khairosfaizal and A. J. Nor’aini, ”Eyes detection in facial images using Circular Hough Transform,” 2009 5th International Colloquium on Signal Processing Its Applications, Kuala Lumpur, 2009, pp. 238-242, doi: 10.1109/CSPA.2009.5069224.

[11] W. LU and M. YANG, ”Face Detection Based on Viola-Jones Algorithm Applying Composite Features,” 2019 International Conference on Robots Intelligent System (ICRIS), Haikou, China, 2019, pp. 82-85, doi: 10.1109/ICRIS.2019.00029.