Minimizing Respiratory Diseases using Warning from Hand-Touch-Over-Face Avoidance Based Algorithm on Deep Learning

Jitesh Kumar Bhatia
Department of Computer Engineering and Applications, GLA University, Mathura
jitesh.bhatiya@gla.ac.in

Abstract. The most moved parts of a human body are hands. They may get contaminated by touching any surface we are working with. The contaminated hands can intentionally or unintentionally reach the part of our face while eating, sneezing, setting of hair, itching a scratch, biting the nails, soothing down when stressed, etc. The touch of hands in eyes, nose and mouth may result in unhygienic symptoms and may result in various respiratory diseases. We need to minimize our touch to the parts of our face. In the presented work, we propose a system that warns the user each time his/her hand reaches his/her face. We have used CNN to identify the class feature automatically rather than manually. Also, we have used RNN so that the proposed system learns recursively. The use of such system may minimize the subsequent touch on face and will lead to minimizing the respiratory diseases due to unhygienic touch conditions.

Keywords- Hand Over Face; Respiratory Diseases; Convolutional Neural Networks; Recurrent Neural Networks; Machine Learning.

1. Introduction
The hands are the most moved organ of a human body. Also, hands are used by all the human beings to feel the touch of any surface. It results in the transfer of Virus and bacteria from a surface to our hands. This usually creates unhygienic symptoms in hands. These hands may reach our face while eating, sneezing, touching and/or setting the hair, itching in eyes, wiping tears, cleaning the mucous, to scratch an itch, to soothe down when stressed, biting nails, etc. to name a few. The contaminated hands, when touch the mouth, nose or eyes can result in infections such as influenza, coronavirus, Staphylococcus aureus (a disease carried in nasal nucosa), conjunctivitis, etc. In a nutshell, the infection in a human being is the result of self-inoculation. Self-inoculation is a type of contact transmission where a person’s contaminated hands makes subsequent contact with other body sites on oneself and introduces contaminated material to those sites [1][2]. According to a study by [3], on closely monitoring 26 students, for duration of 240 minutes, they, in an hour on an average touched their face 23 times. The average duration of mouth touching was 2 seconds whereas the average nose touching duration was 1 second. Figure 1 shows the average touch of hands on each significant part of the face. It was even notified that the urge to touch the face was increased if the movement of hands was prevented knowingly.

Many solutions to avoid the contact of hands with the face have already been given by different researchers. Some of them are identification the movement of hands with the help of sensor-based gloves, others being identified using the wearables such as a watch or a ring using IoT, whereas some of them identified through an algorithm applied to the live video through a device having camera fitted...
to it. Under the Computer Vision umbrella, hand segmentation is an emerging topic to be explored due to the various challenges like skin colour, light variations, complex backgrounds, etc. to name a few.

In the presented work, the method deals with the identification of a hand or a part of hand (finger, wrist, etc.) on the parts of the face knowingly or unknowingly which may cause a reason for contaminated eyes, nose or mouth. The motivation behind the work is the increase in the spread of respiratory diseases. One of the major reasons of the increase is the touch of contaminated hands on the major sense organs i.e. mouth, nose and eyes. We apply the Deep Learning algorithms to warn the user in the live video as and when he/she puts his/her hand over the face. This is done by producing a beep sound and highlighting the localization of the hand on the face. In this work, we first determine each pixel of the image/frame under consideration to identify whether it is a part of a hand or not. Also, by the use of Canny-Edge Detection [4] and Viola Jones Face Detection [5], we identify the edges of the face of the user. Then, if the localization of the hand is overlapping the area of the face, we may conclude that the hand is overlapping the face. However, if there is a distance in the y-z plane or the x-z plane between hand and the face, the result is negative.

2. Related work
Much of the work has been done in the identification of face in a video and segmentation of hands in image and video. However, hand over face segmentation is an area of Computer vision that still requires a plenty of efforts. To an extent, there are basically two approaches that can be followed to detect hand over a face in a 3-dimensional environment. Firstly, having sensors fitted in the hand in the form of wearable, that detects the localization of the wearable (sensors) on the face. Secondly, by having a vision based technique that identifies the presence of a hand over the face. The first technique being costly due to the involvement of hardware like sensors and hardware required for implementing IoT, the latter technique is favored by most of the researchers.

The process of determining the class/category of each pixel of an image is known as semantic segmentation. The semantic segmentation basically involves either of the two methods, namely, the probabilistic approach and the deep learning technique. Though the traditional approach was the probabilistic approach, recently, the latter one is outperforming better for semantic segmentation.

Out of the many public-datasets available on the internet, the most popular ones related to the hands-over-face are VLM-HandOverFace dataset [6] and the HOF dataset [7]. The VLM-
HandOverFace dataset [6] contains more than 4000 images and is available for research for academicians. The HOF dataset [7] has been created using the random images on the internet. The dataset contains images with hands near and in front of the face. The size of the data set is 300 frames. It consists of images of people with different ethnic backgrounds, color, age and gender.

Some of the related works are described as under.

2.1. Sensor based Hand over Face Segmentation
Sudarshan B, et. al [8] designed a model for smart watches to identify the hand movement to the face by training the classifier using 2071 patterns. Out of these, 102 features were extracted to describe the movement of hands towards the face. They applied one-class SVM, iForest, Minimum Covariance Determinant and Local-Outlier Factor on the feature extracted above. For the purpose of training the classifier, they used the combination of Accelerometer data, Gyroscope data, Pressure data and Rotation Vector data. They ignored the non-hand to face movements (outliers) and achieved an F1 score of 0.93.

2.2. Computer Vision based Hand over Face Segmentation
Mahmoud M, et. al [9] detected the hand when it occluded a face. They divided each frame into 9 regions keeping face in the center, and applied LBPs or Gabor Filter for the extraction of the required feature vector. Each frame was then compared with the initial frame which included only the face. In case of any discrepancy due to a regional change in the structure of the image, they generated a probability that the change was due to the hand occluding the face. Gonzalez et al. [10] tracked the segmentation of hand over face for the identification of sign language. The authors used color and edge information for their proposed methodology. Grafsgaard et al [11] identified human gestures with the presence of two hands in the image under consideration using depth images. The authors tested their research in a computer-mediated tutoring environment. Mahmoud M et al. [12] classified various gesture descriptors such as hand shape, hand action and the region of face occluded by hand. The authors used HOG, PCA and SVM to improve the performance of the proposed approach. They concluded through their results that the multi-modal linear-SVM classifier outperformed the uni-modal linear SVM classifier. Naik N, et. al [13] used deep learning approach for the purpose of identifying gestures of face when face was occluded by hand and determined the type of emotion(s) of the human being under consideration in the image. The authors used RNN to recursively learn the features to identify a broader number of emotions categories other than being happy, sad, disgust, thinking, boredom, surprise, etc. to name a few.

3. Proposed Hand-Touch-Over-Face-Avoidance Based Method
As per the discussion in the above literature, we have seen a significant amount of work has been done in the field of Hand-Over-Face-Segmentation. However, the touch of the hand on face has not been addressed by most of the researchers. In this section, we introduce our new approach for the identification of hands over face with the hands touching the face or a part of face, which is to be avoided. We not only segment the hand over face in the x-y plane but also check that the distance between the hands and face in the x-z plane or in the y-z plane is zero or minimum. Moreover, the proposed method uses the Deep Learning Technique for more accurate feature extraction and classification of the parts of the body.

The Figure 2 illustrates the basic flow of the proposed algorithm to find the hand over part of a face with less than allowed minimum distance in the 3-dimensional plane.
Figure 2: Flow of the proposed Algorithm

Input: Dataset \( I_{\text{max}} = \{I_1, I_2, \ldots, I_n\} \) of n images and image I
Output: Beep if Hand is touching the part of Face

BEGIN

//Phase 1: Training the Classifier//
1. for i from 1 to n images
2. \( g = \text{rgb2gray}(I_i) \)
3. \( \text{face detection} = \text{Detect Face}(g) \)
4. if face_detection = 1
   \( \text{hand detection} = \text{Detect Hand}(g) \)
   if hand_detection = 1
      Apply CNN to extract advanced related features of Hand and Face simultaneously
      if 3-dimensional localization of hand and face coincide
      Apply RNN to add a category of warning

5. else
   No human face detected

//Phase 2: Image to be Tested//
6. \( g = \text{rgb2gray}(I) \)
7. \( \text{flag} = \text{Detect Face}(g) \)
8. if face_detection = 1
   \( \text{hand detection} = \text{Detect Hand}(g) \)
   if hand_detection = 1
      Apply CNN to extract advanced related features of Hand and Face simultaneously
      if 3-dimensional localization of hand and face coincide
      Apply RNN to recognize category of warning from the classified categories
      Play a beep sound of warning

9. else
   No human face detected

END

Figure 3: Pseudo Code of the proposed Algorithm
The Figure 3 shows the pseudo-code for the identification of a hand over part of the face with less than allowed minimum distance in the 3-dimensional plane between them. It also produces a warning sound with an appropriate message as soon as the detection is successful.

Table 1: Notations and their description

| Notation  | Description of the used Notation                                      |
|-----------|-----------------------------------------------------------------------|
| N         | Number of Images in the Training Set                                  |
| I<sub>train</sub> | Set of Images used to train the classifier                         |
| I         | The Image under consideration                                        |
| Detect_Face | Algorithm to Detect Face                                           |
| Detect_Hand | Algorithm to Detect Hand                                           |

The proposed algorithm uses certain notations which are explained in Table 1. As shown in the figure, the algorithm consists of 2-phases, namely Training and Testing. In the first phase, the classifier is being trained using n number of images in the training data-set. Each time a hand over the face is found, it is stored in the set of categories with appropriate message using RNN. After the training phase is over, the testing is performed on image I. If the image I contains a human face, it checks for hands and the localization of hands. If the localization of both, face and hand coincide, the appropriate message is shown with a beep (warning sound) that was categorized using RNN in the training phase.

The steps in each phase are described as under:

3.1. Face Detection
The beginning of the proposed algorithm is with the detection of a human face in the image. The detection of a face in an image is somewhat cumbersome due to the difference in the parameters in the image such as size and color of the face, the illumination when the image was taken, the texture that was in the vicinity, the shape of the face (oval, square-shaped, etc.) and so on. So, it is very important to choose one of the best Face Detection algorithms from the literature. As per our requirements, we have chosen Viola-Jones face detector [5] due to the advantages as stated: It is faster in determining the face and can process up to 15 frames per second and uses the integrated learning algorithm known as ADABOOST to select the appropriate face features from the ones extracted from the face.

3.2. Hand Detection
Detecting Hands in an Image is one of the major challenging tasks as the shape of the hand can be changed as per the user. The detection of hands can be affected by the difference in lightning conditions in the image and the noise in the background. The skin color also adds a challenge to the detection and extraction of a hand in an image. In our method, we have used the Canny Edge Detection to identify the hand area and the localization of the hand in the image. Also, the automatic feature extraction technique known as CNN (Convolutional Neural Network) [14][15] has been used to extract the necessary features and discard the other features. CNN mainly consists of five layers namely the input layer, Convolutional Layer, Rectified Linear Unit, Pooling Layer and the Softmax
layer as shown in Figure 4. Each layer has its own function of extraction of features from the image under consideration.

3.3. Classification
The third step of the algorithm i.e. classification has been performed using RNN. Recurrent Neural Network has the tendency to categorize more categories of hand over parts of a face and gives appropriate result each time a hand is found on a part of a face. The RNN also learns the features from the existing features recursively. Thus, as more and more new images are tested using RNN, the accuracy of the algorithm goes on improving.

4. Results
We have performed our algorithm on two chosen datasets namely VLM-HandOverFace dataset and the HOF dataset. The former consists of 4300 images of faces with and without hands and the latter contains 300 images randomly selected from the Internet.

True-Positives and False-Positives are basically used to detect the number of images where the warning was issued. True positives are the number of images with hands on the face successfully identified the same. However, the number of images with face only, identified occluded with hand is termed as False positives. Some of the images with hands on the face were missed to be identified by the classifier. Such images in the dataset are termed as False Negatives. After calculation of them, the Precision, Recall and F1-score can be calculated for comparison with the other approaches. The parameters are calculated as follows:

\[
Precision, p_r = \frac{t_p}{t_p + f_p} \quad (1)
\]

\[
Recall, r_c = \frac{t_p}{t_p + f_n} \quad (2)
\]

\[
F_1 = \frac{2p_r r_c}{p_r + r_c} \quad (3)
\]

On application on the proposed algorithm in the datasets, the values calculated as per equations (1), (2) and (3) are shown in Table 2 and Table 3 respectively.

Table 2. Experiment Results by the proposed approach on [6].

| Parameters | Value |
|------------|-------|

Figure 4: Architecture of CNN [15]
Table 3. Experiment Results by the proposed approach on [7].

| Parameters                                      | Value |
|-------------------------------------------------|-------|
| Total Number of Images in the dataset           | 300   |
| True Positive, $t_p$                           | 231   |
| False Positive, $f_p$                          | 34    |
| True Negatives, $t_n$                          | 27    |
| Precision, $p_r$                               | .8716 |
| Recall, $r_c$                                  | .8953 |
| F1-measure                                     | .8833 |

The data can also be depicted in the graphical representation in the form of a bar graph shown in Figure 5.
The above results depict that the proposed method outperforms in the case of both the chosen data-sets with the F1-score of .89 and .88 respectively.

5. Conclusion
In the presented paper, we have proposed an algorithm which can continuously keep a check on a person who is in front of a camera such as web-cam. The person is getting continuously monitored and as soon as he/she puts his/her hand on the part of his/her face (intentionally or un-intentionally), the proposed algorithm warns the user with a beep sound. As the number of warning increases, the person may reduce his/her habit and shall start avoiding touching his/her face. Also, our proposed algorithm gets improved while testing due to the Recurrent Neural network. Both, the datasets used and the algorithm being public, can be reproduced by any other researcher to solve the problem of avoidance further.

References
[1] Nicas M and Best D 2008 A study quantifying the hand-to-face contact rate and its potential application to predicting respiratory tract infection J. Occup. Environ. Hyg. 5 347–52
[2] G D G, R A, L A, P M and I A 2008 A survey of knowledge, attitudes and practices towards avian influenza in an adult population of Italy BMC Infect Dis 8
[3] Kwok Y L A, Gralton J and McLaws M-L 2015 Face touching: a frequent habit that has implications for hand hygiene Am. J. Infect. Control 43 112–4
[4] Rong W, Li Z, Zhang W and Sun L 2014 An improved Canny edge detection algorithm 2014 IEEE International Conference on Mechatronics and Automation (IEEE)
[5] Viola P and Jones M J 2004 Robust real-time face detection Int. J. Comput. Vis. 57 137–54
[6] Ghanem S, Imran A and Athitsos V 2019 Analysis of hand segmentation on challenging hand over face scenario Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (New York, NY, USA: ACM) (Link to Dataset: https://sites.google.com/view/sghanem/vlm-handoverface)
[7] Khan A U and Borji A 2018 Analysis of hand segmentation in the wild 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (IEEE) (Link to Dataset: https://drive.google.com/open?id=1hHUvINGICvOGcaDgA5zMbzAIUv7ewDd3)
[8] Sudharsan B, Sundaram D, Breslin J G and Ali M I 2020 Avoid Touching Your Face: A Hand-to-face 3D Motion Dataset (COVID-away) and Trained Models for Smartwatches pp 1–9
[9] Mahmoud M, El-Kaliouby R and Goneid A 2009 Towards communicative face occlusions: Machine detection of hand-over-face gestures Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 481–90
[10] Mahrishi, M., Hiran, K. K., Meena, G., & Sharma, P. (2020). Machine Learning and Deep
Learning in Real-Time Applications. IGI Global. http://doi:10.4018/978-1-7998-3095-5

[11] Grafsgaard J F, Fulton R M, Boyer K E, Wiebe E N and Lester J C 2012 Multimodal analysis of the implicit affective channel in computer-mediated textual communication Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI ’12 (New York, New York, USA: ACM Press)

[12] Mahmoud M M, Baltrušaitis T and Robinson P 2014 Automatic detection of naturalistic hand-over-face gesture descriptors Proceedings of the 16th International Conference on Multimodal Interaction - ICMI ’14 (New York, New York, USA: ACM Press)

[13] Naik N and Mehta M A 2018 Hand-over-Face Gesture based Facial Emotion Recognition using Deep Learning 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET) (IEEE)

[14] Kim P 2017 Convolutional Neural Network MATLAB Deep Learning (Berkeley, CA: Apress) pp 121–47

[15] Albelwi S and Mahmood A 2017 A framework for designing the architectures of deep convolutional neural networks Entropy (Basel) 19 242