System for Medical Mask Detection in the Operating Room Through Facial Attributes

A. Nieto-Rodríguez, M. Mucientes[✉], and V.M. Brea

Center for Research in Information Technologies (CiTIUS),
University of Santiago de Compostela, Santiago de Compostela, Spain
{adrian.nietorodriguez,manuel.mucientes,victor.brea}@usc.es

Abstract. This paper introduces a system that detects the presence or absence of the mandatory medical mask in the operating room. The overall objective is to have as few false positive face detections as possible without losing mask detections in order to trigger alarms only for healthcare personnel who do not wear the surgical mask. The medical mask detection is performed with two face detectors; one of them for the face itself, and the other one for the medical mask. Both detectors run color processing in order to enhance the true positives to false positives ratio. The proposed system renders a recall above 95 % with a false positive rate below 5 % for the detection of faces and surgical masks. The system provides real-time image processing, reaching 10 fps on VGA resolution when processing the whole image. The Mixture of Gaussians technique for background subtraction increases the performance up to 20 fps on VGA images. VGA resolution allows for face or mask detection up to 5 m from the camera.

Keywords: Face detection · Facial attributes · AdaBoost

1 Introduction

This paper introduces a system that triggers an alarm when the healthcare personnel do not wear the mandatory mask in the operating room. As apparent, the utility of the system is the detection of a breach in the medical protocol. The overall objective is to provide the least possible rate of false alarms in order not to disrupt the normal activity in the operating room, while keeping a high recall in the detection of faces with masks.

The faces classification system detects the presence of faces in an image and classifies them into two categories: (i) faces with surgical mask and (ii) faces without surgical mask. The Viola-Jones face detector is the base of this work [10].

Actual conditions in the operating room oblige to yield real-time image processing with a low false positive rate and a high recall. Figure 1 collects different challenges to the face classification system. Garment foldings and shades are

the most common source of false face detections, which in turn are very difficult
to dismiss as actual detections without further postprocessing, and thus a lower
frame rate. Mask detection poses similar challenges. Portions of clothing next to
skin areas might also give false mask detections. A worse situation occurs when
a mask prompts a false alarm of a maskless face. As will be seen throughout the
paper, this happens when its tone in the HSV color space is similar to that of the
skin. A less critical situation comes with healthcare personnel wearing surgical
mask and glasses, which makes it hard to give a detection, as either a face or
a mask. Finally, it is also difficult not to miss rotated and/or leaning faces or
masks. In this work, the former is met with synthetic rotations in the training
phase [7]. The latter is tackled with a high enough frame rate to detect the face
in a frontal position; before the person is leaned. Tackling all the above condi-
tions leads to a robust system, making its actual deployment in the operating
room acceptable by the medical staff.

This work lies within the framework of facial attributes identifiers [9], or more
generally within the context of short-term person recognition or soft biometrics
techniques [1,2]. The Viola-Jones detector is the starting point in most of the
facial attribute identifiers to find the region of interest where the trait is analyzed.
Several approaches are possible for facial attributes, namely, (i) to detect the
presence of a given attribute, e.g. beard or mustache, (ii) to provide the value
of the trait with a certain probability, e.g. hair color or facial measurements,
or (iii) to sort people, e.g. by ethnicity, age or gender. Our work classifies faces
into two categories, namely, faces with and without surgical mask. Even though
this is not the first paper on facial attributes identifiers, the lack of specific
surgical mask detectors in the literature, as well as the constraints imposed by
the conditions of the operating room causes us to introduce specific detectors
incorporating color as a property during the training phase in order to enhance
the true to false positives ratio.

The paper is organized as follows. Section 2 addresses the faces classification
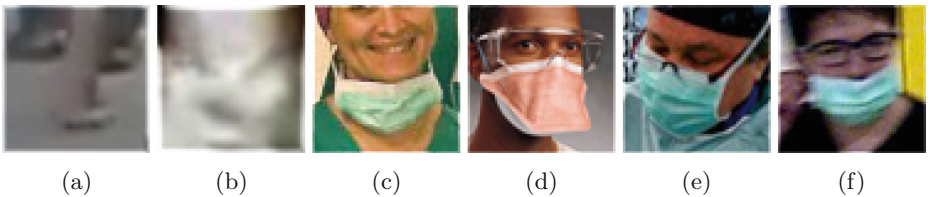system. Experimental results are given in Sect. 3. Finally, the conclusions are
outlined.

(a)   (b)   (c)   (d)   (e)   (f)

Fig. 1. Challenges to the face classification system. (a) and (b) are false faces from
shades and garment foldings, (c) portions of clothing next to skin might give false
positives of mask detection, (d) is a mask with a tone similar to that of skin, (e) is
a leaning mask face with glasses, and (f) shows a face with the mask below the nose
(wrong position).

2 Faces Classification System

The faces classification system detects the presence of faces in an image and classifies them into two categories: (i) faces with surgical mask and (ii) faces without surgical mask. The final goal of the system is to generate an event whenever a person without a surgical mask enters an area in which the use of the mask is required. Figure 2 shows the components of the faces classification system. The system consists of four components: the faces detector, the faces with mask detector, and two color filters —one for each detector.
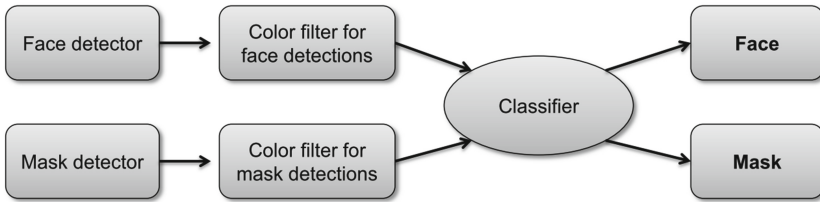

Fig. 2. Faces classification system.

The operation mode of the faces classification system is as follows. When a new image enters the system, both the faces detector and faces with mask detector generate a list of candidate detections represented with a bounding box and a score proportional to the confidence in the detection. Then, each of the lists is filtered by the corresponding color filter, eliminating false positives of the detection stage. Finally, as both lists can share detections (complete or partial overlapping), the final decision on the detected class is taken based on the confidence scores of the overlapping regions. All these steps are described in detail in the following sections.

2.1 Faces Detector

The faces detector is based on the well-known Viola and Jones faces detector [10]. This detector uses very simple rectangular features which evaluate the difference of intensity among rectangular regions. The number of possible features is huge —160,000 for a detector resolution of 24×24— due to the combination of different kinds of features, sizes and positions. The faces detector uses a cascade of classifiers, in which the first elements of the cascade discard the easiest negatives and, as it moves forward in the cascade, the classifiers are more specific and discriminate the most difficult cases. In [10], each classifier of the cascade was learned with AdaBoost.

In this paper we have used a variant of AdaBoost, called LogitBoost [4], which fits an additive logistic regression model by stagewise optimization of the log-likelihood. Moreover, we introduce random rotations in the positive training examples (faces) in order to make the learned classifier robust to small rotations

of the faces. The masks detector follows the same strategy, but we use Gentle AdaBoost [4] instead of LogitBoost. Gentle Adaboost exhibits a good generalization ability, as it prevents large updates in pure regions. This characteristic is specially important for the mask detector, as overfitting would lead the system to identify as mask other clothes of the surgical team.

Both detectors work at different image scales. This means that a face will usually be detected several times and, therefore, the number of detections can be used as a confidence in the detection. This detection threshold (ϵ_d) allows the designer to manage the balance between precision and recall (sensitivity). At this stage we look for classifiers with a high recall (we look for a low false negatives rate), as the number of false positives will be reduced by the color filters.

2.2   Color Filter for Faces

Both the faces detector and the faces with mask detector work on gray images. Nevertheless, color information is very relevant in order to identify faces and, also, to distinguish between faces with and without surgical mask. The objective of the color filters stage is to reduce the number of false positives, but trying to avoid an increase in the number of false negatives. Both filters — for faces and faces with mask— use the skin tone on the HSV color space, as HSV codifies the tone in one channel —hue channel. We have determined the skin tone range using all the faces —without mask— in the training set. The process is as follows:

1. For each face in the training set, we build its individual hue histogram.
2. The dominant hue value for each face is calculated summing up the elements of the individual histogram inside a sliding window, and selecting the highest value along the histogram. This helps to prevent the selection of isolated peaks in the histogram.
3. Each face in the training set contributes to the global histogram adding one element in the position of its selected highest value.
4. The skin tone range is the minimum interval of the global histogram to which all the values over a threshold belong.

The idea behind the faces color filter is that most of the pixels in the detection area must have a skin tone. For the faces with mask color filter, the pixels with skin tone should be in the upper part of the image, while the pixels in the lower part should have quite a different tone. Therefore, for each of the color filters, the faces classification system has to learn the features —position and size of the area in which the hue of the pixels has to be analyzed— and the threshold.

The faces color filter uses just one feature, which is the percentage of pixels inside a rectangular area with a hue value within the skin tone range. To select the position and size of the rectangular area, and the percentage of pixels — threshold—, the system uses the average class entropy [3]:

$$E(A,\ T;\ S) = \frac{|S_1|}{|S|} Ent(S_1) + \frac{|S_2|}{|S|} Ent(S_2) \tag{1}$$

$$Ent(S) = -\sum_{i=1}^{k} P(C_i, S) log_2(P(C_i, S)) \tag{2}$$

where A is the attribute —or feature— used to classify the examples set S, and T is the threshold. Thus, for a given value of T, the examples set is divided into two sets, with S_1 being the set of examples with A values $\leq T$, and S_2 the set of examples with A values $> T$. Therefore, the classification system will look for the value of T that minimizes the average class entropy (Eq. 1), as that threshold corresponds to the value that better separates both categories (faces and non faces).

The faces with mask color filter is based on a similar approach. However, the faces with mask have two meaningful areas: (i) the upper area, which should have pixels with skin tone, and (ii) the lower area, which corresponds with the mask and should have pixels outside the skin tone range. Taken this into account, we have defined four different features:

- The quotient of the number of pixels within the skin tone range between the upper and lower part of the image.
- The percentage of pixels within the skin tone range in the two parts of the image.
- The percentage of pixels within the skin tone range in the upper part of the image.
- The percentage of pixels within the skin tone range in the lower part of the image.

This color filter is, in fact, a cascade of filters, where each stage of the cascade uses one of the described features. The cascade has the following characteristics:

- Each stage uses a single feature, and a feature can only be used in one stage.
- For each feature, the system learns the position and size of the rectangular areas —one or two areas, depending on the feature—, and a threshold —the quotient of the number of pixels or the percentage of pixels.
- For each stage, the training examples are those which passed the previous stages.
- At each stage, all the selectable features compete among each other. The classification system will pick the feature —and the positions, sizes, and threshold— that minimizes the average class entropy (Eq. 1).

2.3   Classification

All those candidate detections —generated by the detectors— that passed the color filters are classified in the corresponding category. However, there are many detections with complete or partial overlap and belonging to both classes. In order to classify these cases, the system uses the score of each detection, which corresponds with the number of detections at the different image scales. Thus, the system will select the category with the highest score.

3   Results

3.1   Face Detector

The face detector comprises two stages; one of them on gray-scale images, and a subsequent one with color. The training phase for the gray-scale case has been performed with the LFW image dataset [6]. The test phase has been run on CMU Frontal Face Test Set [8]. We have used 10,000 positive and 5,000 negative images during the training phase. In order to enhance robustness, the faces of the training phase have been subject to rotations up to $\pm 15°$. The face detector contains a cascade of 20 classifiers. Each one of them is a decision tree with two depth levels with 0.999 and 0.5 as detection and false positive thresholds, respectively. During the test the scale of the face detection goes from 20×20 up to whole size of the image with 1.2 as scaling factor between successive scales.

The color face filter has been built on the histogram of the tone in the HSV color space for more than 13,000 images from LFW [6]. The resultant histogram shows that the skin tone of faces lies within the [340°, 50°] range. The training phase has been made with 4,000 positive and negative images. The test has been run on the image dataset BAO [5]. The training phase yields the most adequate position and size of the feature within a window of 20×20 pixels, and 67 % as the best skin percentage within the feature subwindow. These values have been found by minimizing entropy (Eq. 1). Figure 3a shows the ROC curve of the face detector with and without the color filter —the ROC of the Viola-Jones detector from OpenCV is included as a reference.

3.2   Mask Detector

As in the case of face detection, the mask detector comprises two stages; one of them on gray-scale images, and a subsequent one with color. The training phase for the mask detector on gray-scale images has 4,000 positive and 15,000 negative images. A forth of the positive images have been extracted from image search engines from the Internet. The remaining percentage of positive images are the result of the same synthetic transformations used for the maskless face detector. Half of the negative set are background images. The other half comprises maskless faces.

The mask color filter finds the definitions —position, size and threshold— of the features. Four conditions have to be met in order to classify a face as a face with surgical mask, namely: (i) the skin pixels ratio between the upper and lower part is over one; (ii) an overall percentage of pixels with skin tone above 7 % of the pixels of the upper and lower parts of the feature; (iii) a minimum percentage of 9 % of the pixels in the upper part; and (iv) less than 50 % of pixels in the lower part with skin tone. All the above numbers have been found by minimizing entropy (Eq. 1) over a set of 4,000 positive cases, 2,000 maskless faces, and 2,000 background images. The results of the mask detector with and without color are shown in Fig. 3b.
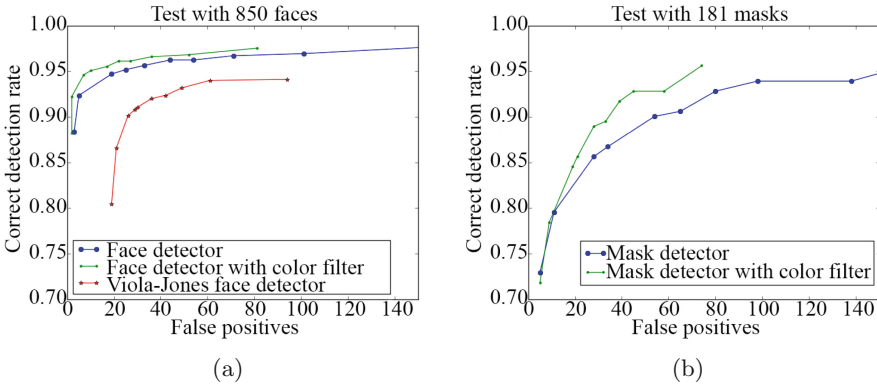
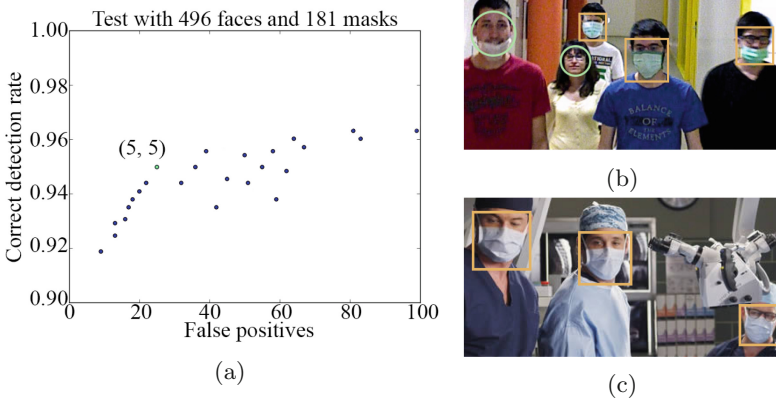**Fig. 3.** ROC curves of the face classification system for (a) face and (b) mask detectors.

Fig. 4. Face and mask detectors combined: (a) results, (b) and (c) examples of masks (in rectangles) and faces (in circles).

3.3 Face and Mask Detectors Combined

The face classifier combines the face and the mask detector. Figure shows the results with both detectors combined. The test has been run on images with maskless faces from BAO [5] and images of our own image dataset with masks. The test contains 99 images with 496 faces and 181 maskless faces. We set a minimum number of detections across all the scales for the face and the mask detector. The case with a threshold of 5 detections for both face and mask has been chosen. This results in a true positive ratio of around 95 % with less than 40 false positives. Figures 4b and c show the results with this classifier in two different scenarios. Figure 4b shows how the system works with faces and masks. Figure 4c displays an example in the operating room.

4   Conclusions

This paper has introduced a system that triggers an alarm when the healthcare personnel do not wear the mandatory medical mask in the operating room. The system combines a face and a mask detector which use the tone in the HSV color space. The positive and false positive rates are above 95 % and below 5 % on 496 faces and 181 masks from the BAO database and an own image dataset for faces with masks. The system works in real-time on a conventional PC, and it will be deployed in the operating room in the near future.

References

1. Anguelov, D., Kuang-Chih, L., Gokturk, S., Sumengen, B.: Contextual identity recognition in personal photo albums. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2007)
2. Dantcheva, A., Velardo, C., D'Angelo, A., Dugelay, J.L.: Bag of soft biometrics for person identification. Multimedia Tools Appl. **51**(2), 739–777 (2011)
3. Fayyad, U., Irani, K.: Multi-interval discretization of continuous valued attributes for classification learning. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI), Chambery, France, pp. 1022–1029 (1993)
4. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann. stat. **28**(2), 337–407 (2000)
5. Frischholz, R.: Bao face database at the face detection homepage (2012)
6. Huang, G., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts (2007)
7. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 297–304. Springer, Heidelberg (2003)
8. Rowley, H.A., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Trans. Pattern Anal. Mach. Intell. **20**(1), 23–38 (1998)
9. Vaquero, D., Feris, R., Duan, T., Brown, L., Hampapur, A., Turk, M.: Attribute-based people search in surveillance environments. In: Workshop on Applications of Computer Vision (WACV) (2009)
10. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vision **57**(2), 137–154 (2004)