Demo: Mask and Maskless Face Classification System to Detect Breach Protocols in the Operating Room

A. Nieto-Rodríguez
Centro de Investigación en Tecnoloxías da Información (CiTIUS)
University of Santiago de Compostela
Spain

M. Mucientes
Centro de Investigación en Tecnoloxías da Información (CiTIUS)
University of Santiago de Compostela
Spain

V.M. Brea
Centro de Investigación en Tecnoloxías da Información (CiTIUS)
University of Santiago de Compostela
Spain
victor.brea@usc.es

ABSTRACT

This live demo allows ICDSC participants to interact with a system to classify faces into two categories: faces with and without surgical masks. The system assigns a per-person ID through tracking in order to trigger only one alarm for a maskless face across several frames in a video. The tracking system also decreases the false positive rate. The system reaches 5 fps with several faces in VGA images on a conventional laptop. The output of our system provides confidence measures for the mask and maskless face detections, image samples of the faces, and for how many frames faces have been detected or tracked. This information is very useful for offline tests of the system. Our demo is the result of a project in cooperation with an IT company to identify breach protocols in the operating room.

CCS Concepts

•Computing methodologies → computer vision;

Keywords

Face detection; mask detection; tracking

1. SYSTEM DESCRIPTION

Fig. 1 shows the block diagram of our system. The classical Mixture-of-Gaussians (MoG) technique performs successfully for background subtraction as the first step of the system. The second step is face classification and tracking. Face classification is triggered in the foreground and in the regions estimated by the tracking block. The third step is the association of new detections, either mask or maskless faces, with the estimated position of the trackers. In this context, a tracker is the set of features associated with a face (see Section 1.2). In step 4 either new trackers are created when new face detections do not match existing trackers, or alternatively, the position of the trackers are updated to those of the existing faces. Finally, a tracker ends when the face is occluded or out of the field of view of the camera, providing the category of mask or maskless face.

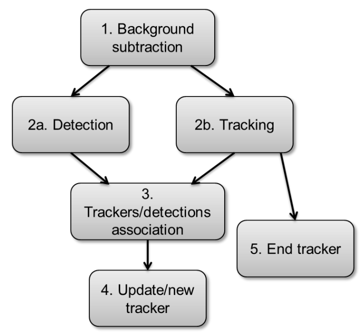

Figure 1: Block diagram of our system.

1.1 Face Classification

Fig. 2 displays the face classification system, which is the block labeled *2a. Detection* in Fig. 1. The face classification system is trained for frontal faces. Viola-Jones is the base algorithm for face and mask detection [1]. The system sorts faces out in the gray-scale space before the second stage, where color filters are used. Finally, the system makes a decision about the final class based on the highest number of detections across the scale space. The maskless face detector uses LogitBoost to learn the cascade of classifiers. The mask face detector employs Gentle AdaBoost. Also, both detectors are trained with faces randomly rotated up to $\pm 15^\circ$ to increase robustness. The number of detections across the scale space permits to balance recall and precision and to have a confidence measure of the detection. A color filter for faces in the HSV color space using the H component decreases the number of false positives. The minimization of the information class entropy in the training set yields the percentage of pixels with skin tone in the HSV color space for a maskless face to be correctly classified. Similarly, the minimization of the information class entropy causes a mask face to be successfully classified when: i) the skin pixels ratio between the upper and the 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

ICDSC '15 September 08-11, 2015, Seville, Spain

ACM ISBN 978-1-4503-3681-9/15/09.
DOI: http://dx.doi.org/10.1145/2789116.2802655

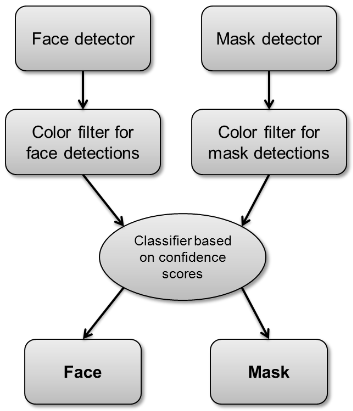

Figure 2: Face classification system.

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.

The training set for maskless faces has been the LFW data base [2], with 5,000 positive and 10,000 negative images for the gray-scale space, and 4,000 positive and negative images for the color space. The test has been run on the CMU data base for the gray-scale space [3], and on the BAO data base for the color space [4]. The training set for mask faces has been built from 4,000 positive and 15,000 negatives found in the Internet. Interested readers can probe further on our face classification system through [5].

1.2 Tracking System

The tracking is implemented with the optical flow. In so doing, a regularly distributed set of points in a grid, what we call a tracker, is created in the face region when the face is detected for the first time. The optical flow gives the estimated location of the faces during the next frame. Incorrect assignments among points across frames are fixed with the search of the right geometrical transformation through RANSAC. The trajectory of the face is depicted through the different geometrical transformations throughout the frames of a video. Trackers and faces have associated bounding boxes. When the overlap between both exceeds 50%, the position of the tracker is updated to the position of the face. The tracker ends when the system does not find a right geometrical transformation. This might occur because the face is occluded or out of the field of view of the camera. Finally, apart from the tracking itself, the condition of having a face across several frames helps decrease the false positive rate.

2. EXPERIMENTAL RESULTS

The system has been tested on a PC with an IntelÂl' CoreTM i7-3770 CPU @3.40 GHz processor and two different cameras, namely, RGB-D Asus Xtion Pro camera, and the IP camera D-Link DC-2230. The performance on VGA images has reached up to 5 fps with several people. This setup permits to detect faces up to 5 m from the camera. The frame rate can be increased by changing the scale space, although at the cost of missing faces at longer distances. Five detections across the scale space produce a recall above 95%. The scale space goes from 20 $\times$ 20 pixels up to the whole size of the image with 1.2 as scaling factor between successive scales. Fig. 3 shows the ROC curve and two sample images from the Internet processed by the system. Mask faces are shown with a square; maskless faces with a circle. The face detected on the bottom image is classified as maskless face because the mask is in a wrong position (under the nose). ICDSC attendants will have the chance to run live tests wearing surgical masks in front of a laptop with a camera, and examine the output of the system.

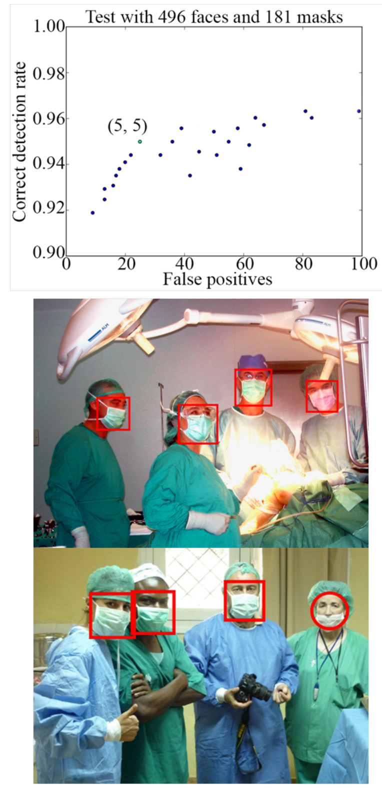

Figure 3: ROC curve of the system and two sample images from the Internet.

3. REFERENCES

[1] P. Viola and M. Jones, "Robust Real-Time Face Detection", Int. J. Comput. Vision 57(2), 137-154, 2004.
[2] G. Huang et al., "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments", Technical Report 07-49, University of Massachusetts.
[3] H. Rowley et al.,"Neural Network-based Face Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, 1998.
[4] R. Frischholz, "BAO Face Database at the Face Detection", 2012.
[5] A. Nieto-Rodríguez et al., "System for Medical Mask Detection in the Operating Room Through Facial Attributes", P. Paredes et al. (Eds.). IbPRIA 2015, LNCS 9117, pp. 138-145, 2015.

4. ACKNOWLEDGMENTS

TEC2012-38921-C02-02 MINECO (FEDER), EM2014/012 (FEDER), AE CITIUS (CN2012/151, (FEDER)), GPC2013/040 (FEDER), and TIN2014-56633-C3-1-R (FEDER).