Detecting Visually Observable Disease Symptoms from Faces

Kuan Wang* and Jiebo Luo

University of Rochester

Abstract
This paper presents a generalized solution to detect visually observable symptoms on faces using semi-supervised learning combined with machine vision algorithms. The proposed algorithm relies on the disease-related statistical facts, and uses the results to detect abnormalities. This approach is in contrast with most existing methods that are based on supervised learning, which is limited by the availability of labeled training data, and therefore offers the major advantage of flagging any unusual and visually observable symptoms.

Keywords: Computer Vision, Imbalanced Dataset, Anomaly Detection, Semi-supervised Learning, Outlier Detection, Clinical Informatics.

*Correspondence: ekuanwang@gmail.com

1 Problem Statement
Currently existing recognition algorithms are usually based on assumptions and are usually trained for specific symptoms, therefore the performance and utility are constrained by lacking training data of unusual symptoms. We propose to adopt semi-supervised outlier detection, combined with computer vision features extracted from normal faces datasets and their statistical models, to produce a reliable mechanism of detecting abnormal symptoms that are visually observable from faces.

2 Methodology
We collected 8278 pictures of normal frontal faces images [4] following gender, age, and race distributions of the adult U.S. population [4] as our training dataset; we collected images for 20 different symptoms [10-11] and expanded it into 237 pictures of faces, paired with 237 pictures randomly picked from normal faces datasets [5-9] as our testing dataset.

We labeled the pictures of the training dataset and the testing dataset to suggest the polygons that bound all related pixels for certain face components polygon using ASMs [1] with 194 landmarks [3]; we also hand labeled pictures missing important facial features. For each set of labels, a binary feature can be obtained to represent its face component by highlighting all pixels bounded inside the polygon. We further transformed the original picture from the RGB color space into the CIELAB color space.

We adopted outlier detection to detect anomalies; an outlier is one observation contains at least one variant that appears to deviate markedly from other members of the sample in which it occurs. Table 1 illustrates the variants we used in the outlier detection and their statistical summarization obtained. For abbreviation, \( \alpha \) represents the aggregate value of the CIELAB alpha channel; \( \beta \) represents the aggregate value of the CIELAB beta channel; \( H \) is the process of applying the well-known Hough Transform on the CIELAB feature of the skin area, and then further applying a counting function to count how many circular structures we found; the mechanism is based on Size Invariant Circle Detection [2].

The values of those variants listed in Table 1 can be easily computed by investigating binary features and the corresponding CIELAB feature, and we further summarized the mean values \( (\mu) \) and standard deviations \( (\delta) \) of the data in training dataset. An outlier is hence defined as a variant whose value is not in \( \mu \pm t \times \delta \), where \( t \) is the threshold. By choosing different \( t \) value, we defined the degree of normality; plotting the values of variants and comparing them with the normal interval will eventually split the testing dataset into flagged group and unflagged group with respect to different \( t \) values.

<table>
<thead>
<tr>
<th>Variant</th>
<th>( \mu )</th>
<th>( \delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ( \alpha(\text{Eye})/\Sigma(\text{Eye}) )</td>
<td>138.426</td>
<td>12.412</td>
</tr>
<tr>
<td>2 ( \beta(\text{Eye})/\Sigma(\text{Eye}) )</td>
<td>138.214</td>
<td>13.345</td>
</tr>
<tr>
<td>3 ( \alpha(\text{Lip})/\Sigma(\text{Lip}) )</td>
<td>150.725</td>
<td>9.752</td>
</tr>
<tr>
<td>4 ( \Sigma(\text{LFace})/\Sigma(\text{RFace}) )</td>
<td>0.962</td>
<td>0.186</td>
</tr>
<tr>
<td>5 ( \Sigma(\text{LEye})/\Sigma(\text{REye}) )</td>
<td>0.958</td>
<td>0.071</td>
</tr>
<tr>
<td>6 ( H(\text{Face}) )</td>
<td>2.233</td>
<td>3.141</td>
</tr>
</tbody>
</table>

Table 1: Variants for the outlier detection algorithm, with their mean values and corresponding standard deviation.

3 Summary of Experiment
In this study, we picked the threshold \( t \) from \( t=0.0 \) to \( t=3.0 \), with the interval of 0.1, 31 sets of experiments in total.

Figure 1: The ROC curve computed by 31 sets of experiments, using the maximum likelihood fit of a binomial model [12]
References


