QUANTITATIVE MEASUREMENT OF
FACE DETECTION ALGORITHM PERFORMANCE

Setiawan Hadi+, Adang Suwandi A, Iping Supriana S, Farid Wazdi *
+ Department of Mathematics, Universitas Padjadjaran
*STEI Institut Teknologi Bandung
email: setiawan.hadi@computer.org

ABSTRACT
Measuring accuracy of a face detection algorithm hardly ever mentioned in detail and explicitly. The standard and terminology used in each experiment reported in each paper were vary and not standardized. In this paper, a method for calculating detection accuracy empirically is proposed. It consists of two algorithms that can be used for measuring accuracy of single face image and multiple faces image detection. Experiment has been conducted by using DeWa, a multiaspect-based face detection method, to measure four facial databases. The results are presented using standard terminologies for detection algorithm such as detection rate, false positive and true negative rates.

Keywords: Face Detection, Accuracy, Bounding Box

1 INTRODUCTION
In general, successful criteria of a face detection is still in debate and has not been standardized yet. Figure 1 shows six results of a face detection algorithm which can be considered as accurate detection. However, the results will generate a question how far the accuracy can be estimated although all results show image containing face.

Due to lack of standardization and terminology about the accuracy of face detection, most of papers published in the literatures used various opinions about the accuracy. Each researcher has his own view and indicators. One of paper [2] mentioned that there are parameters that should be considered for measuring success of a face detection algorithm, those are:

- Position of bounding box where face is detected.
- The result has to be open and can be used by other face detection algorithm
- Database used for benchmarking and evaluation should be standard and has groundtruth information that can be used for measuring other algorithm

The parameters above are not exact due to it has possibilities to be discussed. This problem has not been solved up to this day.

To present the whole result of detection, most of published papers used standard terminologies in machine learning and biometrics [1]. Those terminologies are:

- Detection Rate (DR): ratio between number of faces which is (are) exactly detected by system compared with the number of face or faces determined by human expert observer or human segmented result.
- False Positive (FP): areas in image which are declared as containing face or faces but in fact it contains non-face objects
- True Negative (TN): there are undetected face or faces in image
- False Detection: False Positive + True Negative.

2 METHOD AND ALGORITHM
2.1 Measuring Accuracy Technique
Detection accuracy is measured quantitatively by calculating distance between Ground Truth Bounding Box (GTBB) and Detected Bounding Box (DBB). GTBB is a rectangular object that is bounded into face object based on expert observer knowledge. DBB is a similar object as GTBB but it bounds into face result from face detection algorithm. The detail procedures of bounding box positioning are as follows:

- Face image is marked by positioning bounding box sides as follows: (i) for the top side, the line is put exactly at the boundary between hair and forehead (ii) for the left side, the line is located on the boundary between face and left ear (iii) for the right side, the line is positioned on the boundary between face and right ear and (iv) for the bottom side, the line...
exactly is on the border of chin. The location information of the bounding box corners is recorded. The middle picture in Figure 2 illustrates this technique.

- If image is not in frontal and straight position, the bounding box can be considered as (i) exactly on the boundary between the most top hair and forehead (ii) the most left boundary between face and left ear (iii) the most right boundary between face and right ear and (iv) exactly on the most bottom of chin. The rightmost picture on Figure 2 gives illustration of this method.

Figure 2. Illustration of Bounding Box Determination (dashed line)

Figure 3 shows sample images that have been processed where the GTBB and DBB are put in each image. The solid line that is around the face is DBB resulted from a face detection algorithm implementation. The dashed line is GTBB, which is correct information of the face location. This information is obtained by careful exploration by an expert or human observer.

Figure 3. Sample of GTBB (dashed) and DBB (solid)

The ideal system will present the same position of the bounding boxes (GTBB and DBB). However, not all systems can give that promising result. To estimate the accuracy, distance between bounding boxes should be calculated. If the face detection system is ideal system, the distance must be 0 (zero).

Average distance measurement between GTBB and DBB can be estimated using simple distance such as Euclidean distance. The method is performed by calculating of distance between each top point of bounding box using these formulas:

\[
D = \frac{1}{4} (d_{TL} + d_{TR} + d_{BL} + d_{BR})
\]

where \( d_{TL}, d_{TR}, d_{BL}, \) and \( d_{BR} \) is Euclidean distance between GTBB and DBB on the four corner points, TopLeft, TopRight, BottomLeft and BottomRight.

From the formulas it can be concluded that the value of \( D \) is 0 for 100% detection accuracy. If value of \( D \) is smaller than the value of threshold \( T \), the result of detection can be categorized as True Positive (TP). On the other hand, if \( D \) is greater than \( T \), the result of detection algorithm is categorized as False Detection.

The described method above is implemented for measuring accuracy of two types of face detection system. For this single-face detection system, measurement is easy. However for multi-face detection system, measurement should be performed more carefully, by combining all detection result including the false detection.

2.2 Algorithm 1: Single-face Measurement

In this algorithm, measuring is implemented for single face image, which means there is only one face in the image.

1. For each image in face database \( R \) do
   a. Detect face using detection algorithm \( W \), save the Detected Bounding Box value
   b. Calculate the value of \( D \)
   c. If \( D \geq T \) then increase Detection Rate Accumulator \( H \) by 1

2. The Detection Rate of Face Detection Algorithm \( W \) using Face Database \( R \) is the value of Detection Rate Accumulator \( H \)

2.3 Algorithm 2: Multi-face Measurement

Performing accuracy measurement for an image that contains more than one face objects is not as easy as measuring one-face image. Although the same basic technique is used, but careful action should be considered due to result variabilities. It is assumed that the detection algorithm has successfully generated the bounding box of all faces in the image. The algorithm is as follows:
1. For each image in face database $F$ do
   a. Detect faces using detection algorithm $W$, save the Detected Bounding Box values of each face
   b. For each face in the image
      1. Find the closest Ground Truth Bounding Box $T$
      2. Calculate the value of $D$
      3. If $D \geq T$ then add one to Image Detection Rate Accumulator $G$
   c. Add $G$ to Detection Rate Accumulator $H$
2. The Detection Rate of Face Detection Algorithm $W$ using Face Database $F$ is the value of Detection Rate Accumulator $H$

3 EXPERIMENT AND RESULT

There are many face detection methods published in the literature [3]. In this paper, experiment has been conducted using a multiaspect-based face detection method called DeWa.

3.1 Face Detection Algorithm

DeWa is a framework for detecting multiple faces in multi-media environment in real-time fashion. DeWa consists of three integrated and dynamic phases that utilizes traditional yet powerful image processing and computer vision methods. The framework has been implemented for detecting multiple faces on an image, a video stream and from a realtime camera (webcam). Further explanation of this algorithm can be found in [4]

3.2 Databases Used in Experiment

Four face databases have been used for our experiment. The table shows characteristics of the databases.

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of Images</th>
<th>Number of Faces</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>FERET</td>
<td>735</td>
<td>735</td>
<td>Single Face</td>
</tr>
<tr>
<td>DWI</td>
<td>347</td>
<td>347</td>
<td>Single Face</td>
</tr>
<tr>
<td>mDWI</td>
<td>171</td>
<td>488</td>
<td>Multiple Faces</td>
</tr>
<tr>
<td>VALID</td>
<td>94</td>
<td>298</td>
<td>Multiple Faces</td>
</tr>
</tbody>
</table>

FERET and DWI are facial databases containing single face and simple complexities, which means that the background image is uniform and the face position and size are not too vary. FERET is mainly a face database for face recognition experiment developed by National Institute Standard and Technology US. DWI is a special database of Indonesian people, stands for Data Wajah Indonesia. mDWI is Indonesian people advanced face database that consists of complex facial images with non-uniform background and wide variabilities of faces.

Similar with VALID but images in VALID database consists of non-Indonesian faces.

3.3 Evaluation Mechanism

Figure 4 shows mechanism for measuring accuracy of face detection algorithm using our proposed measurement method. The four databases mentioned before are used as our image databases.

Groundtruth is information, generated by human expert, about the correct information about face in image such as position and size, including bounding box corner points. This information is known as Ground Truth Bounding Box (GTBB). DeWa, a multiaspect-based face detection algorithm, will process each database and generate information collected as Detected Bounding Box (DBB). Information from DBB and GTBB then are computed using one of measuring algorithm mentioned in section 2, depends on the characteristic of database. The result will be presented as Detection Rate, False Positive rate and True Negative rate.

The evaluation criteria mechanism can be implemented for measuring accuracy of different face detection methods or algorithms. The researcher can replace the Sidewa with his or her proposed face detection technique. By implementing our accuracy measuring methods, the performance of the face detection algorithms can be evaluated.

3.4 Result

From the experiment, shown in Table 2, we can conclude that from FERET database, the detection accuracy is 90.63% which means that 90.63% from 735 faces or images in the database is in the range of acceptance and can be said as successfully detected. 9.63% of faces are not detected. Similar interpretation can be implemented to the other database (DWI). The False Positive value for the two databases, FERET and DWI is 0.00% because simple complexity of the databases.

For mDWI and VALID databases, the value of False Positive is not zero. It means that the resulting algorithm shows incorrect detection when regions without faces are considered as faces.
Table 2. Result of Experiment

<table>
<thead>
<tr>
<th>Database</th>
<th>Images</th>
<th>Faces</th>
<th>DR</th>
<th>TN</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FERET</td>
<td>735</td>
<td>735</td>
<td>90.63%</td>
<td>9.27%</td>
<td>0.00%</td>
</tr>
<tr>
<td>DWI</td>
<td>347</td>
<td>347</td>
<td>89.78%</td>
<td>10.22%</td>
<td>0.00%</td>
</tr>
<tr>
<td>mDWI</td>
<td>171</td>
<td>488</td>
<td>92.21%</td>
<td>7.79%</td>
<td>7.96%</td>
</tr>
<tr>
<td>VALID</td>
<td>94</td>
<td>298</td>
<td>91.2%</td>
<td>8.72%</td>
<td>7.67%</td>
</tr>
</tbody>
</table>

4 CONCLUSION

In this paper a method for computing accuracy of face detection algorithm is proposed. The method is simply performed by calculating distance between groundtruth information and detected face information performed by face detection algorithm. This method is implemented by generating two algorithms for single face image measurement and multi faces image measurement.

The method is examined by using DeWa, a multiaspect-based face detection algorithm using two groups of databases, simple and complex databases, which each group consists of two facial databases. The result is presented by indicators mentioned as Detection Rate, True Negative rate and False Positive rate.

REFERENCES


