Improved AdaBoost Algorithm for Face Detection and Its Application
Xin-chao ZHAO¹, Jia-zheng YUAN²*, Xian-kai HUANG³ and Hong-zhe LIU¹

¹Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China
²Institute Institute of Computer Technology, Beijing Union University, Beijing 100101, China
³Beijing Open University, Beijing 100101, China
*Corresponding author

Keywords: Face detection, Improved adaBoost algorithm, Real-time, Multi-face.

Abstract. Face detection plays a crucial role in developing human-robot interaction (HRI) for Intelligent Educational Robot (IER) to recognize users or speakers. In this paper, we introduce an intelligent vision algorithm that is able to detect human face from complex scene and filter out all the non-face but face-like images. The human face is detected in real time using the approach called AdaBoost-based Haar-Cascade Classifier[1-2], and the real human face detection is improved to implement from single-face detection to multi-face detection. Furthermore, variable head pose is taken into account, such as pitch, roll, yaw, etc. The proposed robot vision algorithm for human detection is tested to be effective and robust through real-time experiments.

Introduction
It is essential for intelligent robots to have the ability of recognizing people during communication and cooperation between robots and people [3-4]. In this paper, we focus on one of the significant human-robot interaction technologies, the intelligent facial vision algorithm to robustly detect human face from a variety of complex scenes. Facial vision algorithm has attracted considerable attention in numerous practical applications [5-7] in multiple areas, since it is a vital part in the intelligent human-robot interaction and directly leads to whether robots are capable of interactive communication with people in the educational environment. Imprecise human detection will lead to the insufficient interactions between humans and IER. Therefore, it is indispensable for the IER to detect human faces with a high accuracy.

Related Work
In the recently literatures, many experts research in the intelligent interaction robot. Bernhard Froba et al address in developing intelligent system of face tracking of mobile robot using Kalman filter[8]. Kwang Ho An et al. pay attention to use relatively small amounts of critical rectangle features selected and trained by AdaBoost learning algorithm, and they detect the initial position, size and view of a face correctly [9]. Paul Viola et al present the robust real-time face detection with the integral image method and AdaBoost training algorithm [10]. The Xangdong Xie et al present a real-time tracking algorithm on eye feature tracking [11].

Compared with the previous approaches, we modify and improve the AdaBoost face algorithm for rapidly multi-face detection in the sequence image frames[16-19], and propose a scheme that is effective and robust for the problems of variation of scene and head pose. This approach not only improves the face detection accuracy, but retains the real-time detection speed at the same time.
**Proposed Method**

**Original Method**

Various algorithms are proposed for face detection in the real-time application such as robot system, the existing deficiencies which often limit the application are usually discussed [12].

- **Scene problem**
  How does the detector response when a new object enters the field of view scene and therefore the object should be detected.

- **Head pose problem**
  There are two conditions which make the tracker lost track. One is that the object left the scene, and the other is losing tracking due to the failure of the tracker.

According to the deficiencies mentioned above, we proposed the algorithm of face detection to resolve those drawbacks. The whole system flow chart is illustrated schematically in Figure 1.

![Figure 1. The whole system flow chart.](image1)

**Improved Method**

The AdaBoost algorithm [13] is adopted to perform the face detection in image sequences. We denote it as Global AdaBoost Face Detection algorithm (GAFD), since its huge time consuming and running loading to the system, it is executed in low frequency, especially for intelligent robot system running in real-time. The proposed diagram is shown in Figure 2.

![Figure 2. The algorithm flow chart.](image2)
For fast multi-face detection, we improve the AdaBoost algorithm in order to reduce its running time by utilizing the state of the tracked face. We call the algorithm as Local AdaBoost Face Detection algorithm (LAFD). The LAFD presumes that there is single face in the image and its previous state of the tracked face is presented. The tracker tracks and predicts the new state of the face between the sequential image frames using the Kalman Filter [14, 15]. Partial image which is called region of interest (ROI) is obtained for LAFD, according to the prediction of the tracker. The LAFD is launched in high frequency up to real-time. The tracker controller controls the running timing both GAFD and LAFD and maintains the states of tracked faces.

Some improvements are proposed in this paper, aiming at the defects of the traditional AdaBoost algorithm. The improved algorithm is mainly to provide face relevance, which is used to the ROI selection, the approach reduces the redundant information to minimum, that is unnecessary for face detection, among selected ROI. A certain number of samples, labeled positive or negative, are selected as training set, then the algorithm is used for feature selection in ROI. The weak classifiers are boosted into a stronger classifier by the following steps:

1. The given sample images of \((x_i, y_i)_{i=1}^N\) as input, where \(N\) is number of samples, \(y_i = 0, 1\) represent negative and positive samples, respectively.
2. Initialize the weights \(\omega_{1,i} = \frac{1}{2m} \cdot \frac{1}{2l}\) for \(y_i = 0, 1\), where \(m\) and \(l\) are the number of the negative and positive samples, respectively.
3. Normalize the weights \(\omega_t\) for \(t = 1, \ldots, T\), so that it is a probability distribution.

Then select the best weak classifier with respect to the weighted error:

\[
e_t = \min_{f,p,\theta} \sum_i \omega_i |h(x_i, f, p, \theta) - y_i|
\]

where a weak classifier \(h(x, f, p, \theta)\) consists of a feature \(f\), a threshold \(\theta\) and a polarity \(p\) indicate the direction of the inequality:

\[
h(x, f, p, \theta) = \begin{cases} 1, & pf(x) < p\theta \\ 0, & \text{other} \end{cases}
\]

where \(x\) is a 24 \(\times\) 24 pixel sub-window of an image. After that, choose the classifier:

\[
h_t(x) = h(x, f_t, p_t, \theta_t)
\]

where \(f_t, p_t\) and \(\theta_t\) are the minimizers of \(e_t\).

Defining \(\beta_t = \frac{1}{1 - e_t}\), and update the weights:

\[
\omega_{t+1,i} = \omega_{t,i} \beta_t^{1-e_i}
\]

where \(e_i = 0\) if sample \(x_i\) is classified correctly and \(e_i = 1\) otherwise.

4. By defining \(\alpha_t = \log \frac{1}{\beta_t}\), the final strong classifier is:

\[
C(x) = \begin{cases} 1, & \sum_{t=1}^T \alpha_n h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0, & \text{other} \end{cases}
\]

Experimental Results

To evaluate the efficiency and accuracy of the algorithm proposed in this paper, the experiments adopt two training data set. One is MIT+CMU frontal face dataset and CMU profile face dataset that is same as Viola and Jones, and the other is our training set. The approach is tested using 6500 560\(\times\)420 pixel color images, which were collected from photos, videos on the Internet. One or more
faces are contained in each image, with different lighting conditions and complex scenes. The faces are different in size, pose, location and facial expression. Especially, most of the faces are multi-pose face which include rotated frontal and profile face.

Figure 3 shows a face detection result for single face image, we mark the face area detected using a green rectangle bounding box. Figure 4 shows a result for multi-face detection in the images with complex and confusing background. Figure 5 shows a result for multi-face detection, the head pose of the people in the candidate images are various and different.

![Figure 3. Face detection result for single face image.](image3.png) ![Figure 4. Face detection result for multi-face image with confusing scene.](image4.png)

![Figure 5. Face detection result for multi-face image with varied pose.](image5.png)

Comparison results of face detection are shown in Table 1. Experiments demonstrate that the proposed algorithm is feasible and robust. Furthermore, it performs at 26 frames/s speed, the frame's size is 560 x 420 pixel, and the configuration of the testing computer is Intel Core 2 Dual 2.8 GHz and 4G RAM. Therefore, this approach can provide good inputs for face recognition and facial expression recognition.

<table>
<thead>
<tr>
<th>Detection Algorithm</th>
<th>Testing Dataset</th>
<th>Missing Rate(%)</th>
<th>Correct Rate(%)</th>
<th>Average Speed(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viola Jones</td>
<td>MIT+CMU</td>
<td>31.3</td>
<td>72.3</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>Our Dataset</td>
<td>35.6</td>
<td>68.2</td>
<td>135</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>MIT+CMU</td>
<td>21.8</td>
<td>87.5</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Our Dataset</td>
<td>24.1</td>
<td>84.6</td>
<td>84</td>
</tr>
<tr>
<td>Our Method</td>
<td>MIT+CMU</td>
<td>6.9</td>
<td>93.8</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Our Dataset</td>
<td>5.2</td>
<td>96.3</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 1. This is the comparison results of face detection.
Conclusion and Future Work

In this paper we have implemented a novel methodology for robust multi-face detection in image sequences, intended for HRI in IER applications. Experimental results have confirmed the effectiveness and the increased computational efficiency of the proposed approach, proving that the characteristic advantages are maintained, leading to merits that combine efficiency, accuracy and robustness simultaneously.

We intend to use the proposed approach to support natural interaction with autonomously navigating robots that guide visitors in exhibition centers and museums. To be more specific, the proposed method will provide input for the recognition and analysis of facial expressions that people utilize when engaged in various conversational states.

Future work includes extension of the face detection algorithm to handle stereo vision by exploiting epipolar constraints. Moreover, the approach presented in the paper will be widely employed in an integrated system for naturalistic human-robot interaction.

Acknowledgments

This paper is supported by the following projects: The National Natural Science Foundation of China (No. 61271369, No. 61372148, No. 61571045); Beijing Natural Science Foundation (No. 4152016, No. 4152018); Beijing Advanced Innovation Center for Imaging Technology (No. BAICIT-2016002); The National Key Technology R&D Program (No. 2014BAK08B02, No. 2015BAH55F03).

The authors would like to thank CASIA for providing the databases used in the experiments related in this paper, and the anonymous reviewers for their constructive comments.

References


