Facial expression recognition for multi-player online games

Ce Zhan
University of Wollongong

Recommended Citation
NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.
Facial Expression Recognition for Multi-player On-line Games

A thesis submitted in fulfillment of the requirements for the award of the degree

Master of Computer Science

from

UNIVERSITY OF WOLLONGONG

by

Ce Zhan

School of Computer Science and Software Engineering
February 2008
© Copyright 2008

by

Ce Zhan

All Rights Reserved
Dedicated to
My grandparents, Hongchao Li and Junhui Dong
Declaration

This is to certify that the work reported in this thesis was done by the author, unless specified otherwise, and that no part of it has been submitted in a thesis to any other university or similar institution.

Ce Zhan
February 22, 2008
Abstract

Multi-player on-line games (MOGs) have become increasingly popular because of the opportunity they provide for collaboration, communications and interactions. However, compared with ordinary human communication, MOG still has several limitations, especially in the communication using facial expressions. Although detailed facial animation has already been achieved in a number of MOGs, players have to use text commands to control the expressions of avatars. This thesis proposes an automatic expression recognition system that can be integrated into a MOG to control the facial expressions of avatars. To meet the specific requirements of such a system, a number of algorithms are studied, tailored and extended. In particular, Viola-Jones face detection method is modified in several aspects to detect small scale key facial components with wide shape variations. In addition a new coarse-to-fine method is proposed for extracting 20 facial landmarks from image sequences. The proposed system has been evaluated on a number of databases that are different from the training database and achieved 83% recognition rate for 4 emotional state expressions. During the real-time test, the system achieved an average frame rate of 13 fps for 320 × 240 images on a PC with 2.80 GHz Intel Pentium. Testing results have shown that the system has a practical range of working distances (from user to camera), and is robust against variations in lighting and backgrounds.
I would like to take this opportunity to express my sincere gratitude to my supervisors, Dr. Wanqing Li, Prof. Philip Ogunbona and Prof. Farzad Safaei for their invaluable guidance, advice, criticism and encouragement. I am also very grateful to my colleague Gang Zheng, for the help in programming and for numerous discussions which have given me tremendous confidence and inspiration. I wish to thank Yiyu and Xiaodong, they take care of me just like elder sister and brother. This work is partly supported by Smart Internet Technology (SIT) CRC Australia. I would like to thank SIT for providing a research scholarship.
## Contents

### Abstract

vi

### Acknowledgments

vi

### 1 Introduction

1

1.1 Motivation and Objectives ........................................... 1

1.2 Contributions .......................................................... 3

1.3 Publication List .......................................................... 3

1.4 Organization of Thesis ................................................. 4

### 2 Facial Expression Recognition: Literature Review

6

2.1 Overview ............................................................. 6

2.2 Face Detection ........................................................ 7

2.2.1 Knowledge-based Methods ........................................... 8

2.2.2 Learning-based Methods ............................................. 9

2.2.3 Discussion ............................................................ 12

2.3 Facial Feature Extraction ............................................. 13

2.3.1 Geometric Feature Extraction ................................. 13

2.3.2 Appearance Features Extraction ............................... 15

2.3.3 3D Modeling ......................................................... 18

2.3.4 Discussion ........................................................... 18

2.4 Facial Expression Classification ................................. 19
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Miscellaneous Actions [21]</td>
<td>21</td>
</tr>
<tr>
<td>4.1</td>
<td>The approximate relationship between distance of user to camera and facial component resolution</td>
<td>35</td>
</tr>
<tr>
<td>4.2</td>
<td>Mouth detection results for different face resolutions by using cascade classifiers with different stages</td>
<td>39</td>
</tr>
<tr>
<td>4.3</td>
<td>Mouth detection results on resized faces and corresponding original faces</td>
<td>40</td>
</tr>
<tr>
<td>8.1</td>
<td>The performances of face detector on different databases</td>
<td>62</td>
</tr>
<tr>
<td>8.2</td>
<td>Coding scheme for combination of frequencies</td>
<td>66</td>
</tr>
<tr>
<td>8.3</td>
<td>The best recognition rates achieved by using different SVM kernels</td>
<td>68</td>
</tr>
<tr>
<td>8.4</td>
<td>The performance of facial landmark localization module on BioID database</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>(see Figure 8.9 as a reference for landmark locations)</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>Recognition results for 7 expressions</td>
<td>74</td>
</tr>
<tr>
<td>8.6</td>
<td>Recognition results for 4 expressions</td>
<td>75</td>
</tr>
<tr>
<td>8.7</td>
<td>Failure samples with corresponding expressions in training database</td>
<td>76</td>
</tr>
<tr>
<td>8.8</td>
<td>The average processing CPU time for each module</td>
<td>78</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.1</td>
<td>General processing stages of facial expression recognition</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>The shape model used in [42]</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>Outline of the facial feature detection method proposed in [89]</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>Haar-like rectangle features used in [90]</td>
<td>17</td>
</tr>
<tr>
<td>2.5</td>
<td>Selected Haar-like rectangle features for different expressions: (a) neutral, (b) happiness, (c) anger, (d) sadness, (e) surprise, (f) disgust, (g) fear.</td>
<td>17</td>
</tr>
<tr>
<td>2.6</td>
<td>Six prototypic emotional expressions: anger, surprise, sadness, disgust, fear, and happiness</td>
<td>19</td>
</tr>
<tr>
<td>2.7</td>
<td>Action units (AUs) [21]</td>
<td>20</td>
</tr>
<tr>
<td>3.1</td>
<td>The architecture of the proposed system</td>
<td>30</td>
</tr>
<tr>
<td>4.1</td>
<td>The Haar-like feature set used in Viola-Jones face detection method [88]</td>
<td>35</td>
</tr>
<tr>
<td>4.2</td>
<td>The extended Haar-like feature set [54]</td>
<td>35</td>
</tr>
<tr>
<td>4.3</td>
<td>Rotated Summed Area Table (RSAT)</td>
<td>36</td>
</tr>
<tr>
<td>4.4</td>
<td>Calculation scheme of the pixel sum of rotated rectangle</td>
<td>36</td>
</tr>
<tr>
<td>4.5</td>
<td>Calculation scheme for rotated features</td>
<td>37</td>
</tr>
<tr>
<td>5.1</td>
<td>The coarse-to-fine facial landmark localization process</td>
<td>42</td>
</tr>
<tr>
<td>5.2</td>
<td>Facial landmark estimation</td>
<td>43</td>
</tr>
<tr>
<td>5.3</td>
<td>Illustration of the corner refining method</td>
<td>45</td>
</tr>
<tr>
<td>6.1</td>
<td>Gabor filters with the orientation of $0, \frac{\pi}{4}, \frac{\pi}{2}$ and $\frac{3\pi}{4}$</td>
<td>49</td>
</tr>
</tbody>
</table>
6.2 Gabor filters with the frequency of 0.25, 0.5, 1 and 1.5

6.3 Gabor filters with $\sigma$ value of $\pi$, $2\pi$, $3\pi$ and $4\pi$

6.4 A family of Gabor kernels with three different frequencies and six different orientations

6.5 A face image after convolution with Gabor filters shown in Figure 6.4

7.1 Separating hyperplanes. The left is a random one, right one maximizes the margin of separability

8.1 Sample images from JAFFE database

8.2 Sample images from AR Face database

8.3 Sample images from AT&T Face database

8.4 Sample images from BioID Face database

8.5 Sample images from BioID Face database

8.6 Sample images from CIT Face database

8.7 Sample images from FG-NET Face database

8.8 Face detection results from different databases

8.9 34 facial landmark points which represent the facial geometry

8.10 The recognition performance when one of the 34 landmarks is removed

8.11 The recognition rate for different frequencies

8.12 The recognition rate for different number of orientations

8.13 Facial component detection results

8.14 Facial component detection results from BioID database

8.15 Facial component detection results from FG-NET database

8.16 Real-time facial component detection results

8.17 Average Facial landmark detection rates for different face resolutions

8.18 Facial landmark localization results from BioID database

8.19 Facial landmark localization results from FG-NET database

8.20 Real-time Facial landmark localization results

8.21 Recognition rates at different distances
A.1 Haar-like feature computation with the integral image. The feature value is: $S_1 - S_2$, with $S_1 = E - B - D + A$ and $S_2 = F - C - E + B$. 85