Eye Expression Recognition of Wearable View Magnification Interface for Low Vision

Hideaki Orii*a, Masafumi Aoki a, Hideaki Kawano a, and Hiroshi Maeda a

aKyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu, 804-8550, JAPAN

*Corresponding Author: orii@sys.ecs.kyutech.ac.jp

Abstract

In this paper, in order to deal with difficulties in low vision people's visual activities, we develop a user interface system with eye expression recognition. The system recognizes user's act of squinting by using a wearable camera, and displays the scaled view image to the user. We confirmed that the act of squinting was recognized relatively high accuracy through the preliminary experiments. And we implemented this recognizer, it was also confirmed that it can operate to the scaled view image.

Keywords: wearable camera, vision aid system, low vision.

1. Introduction

In recent years, the number of weak eyesight people including mild low-vision, such as presbyopia and cataracts, has been increased. There are various definitions of low vision. In general, low vision means (1) corrected both eyesight is 0.05 to 0.3, (2) though there are restrictions on daily life and learning, the people with visual impairment other than visual acuity can do visual activities. The definition (1) means that using braces, such as glasses or contact lenses, vision doesn't improve enough. Blindness means corrected both eyesight is below 0.05.

There are individual differences in appearance of low vision depending on the symptoms. For example, Fig 1 (a) shows an appearance image of sighted people. Fig 1 (b) shows appearance image of defocus, such as myopia (shortsightedness), hyperopia (farsightedness) or presbyopia. Cataract patients appear foggy even if focus is achieved. Some people have a narrow field of view in addition to defocused. So we have to think of solution according to their symptoms.

Fig. 1. Difference of appearance

(a) Sighted people, (b) Defocus.

The people with low vision have to live with inconvenience, in a lot of situations e.g. climbing stairs, walking the street, reading information including route maps and signs at public places, such as stations and airports. To solve these problems, solutions are broadly divided into two ways.

The first category is improvement of infrastructure. For example, visibility is more improved by emphasizing the color contrast to border of stairs. In reading information, difficulty is more improved by adding the voice-based guidance.

The second category is development of daily life support system by wearable computing. In recent years, because of minimization of component and performance improvement, various research and development has been proceeding.

For example, Bryant[1] developed the wearable system for low vision. The system detects obstacles adjacent to user with infrared radiation. And location information is presented to the fiber retinal scanning display.

In this paper, in order to deal with low vision's the difficulties in visual activities, we propose a user interface system with the eye recognition. The system recognizes user's act of squinting by using a wearable camera, and displays the scaled view image to the user's head mounted...
"Squinting" means eyes half closed. When a feature is seen blurred, the act is able to adjusting the focus by reducing the amount of light to eye. A similar principle can be seen in pinhole camera. The act is often done unconsciously, especially people of symptoms, such as early myopia or presbyopia. By recognizing the act of squinting, the system can present enlarged view image to the head mounted display, when the low vision finds it hard to be visible. The system is effective to appearance of defocus.

Proposed method is explained in chapter 2. Chapter 3 describes preliminary experiment which was confirmed recognition accuracy of eye expression recognition, and describes experiment which was confirmed the entire system works. Finally chapter 4 describes conclusion.

2. Proposed Method

Fig 2 shows the procedure of proposed method which is consists in “Eye expressions learning”, “Eye expression recognition”, and “View Image presentation”. Section 2.1 describes system configuration. Section 2.2 and Section 2.3 describes proposed method in detail.

2.1 System Configuration

Fig 3 shows a proposed system's overview. The system composes with two cameras and head mounted display. “Camera 1” captures eye images, and “Camera 2” captures view images.

2.2 Eye Expression Learning

Before running the system, user's learning image are prepared, are extracted by HOG feature and are learned by Gentle Boosting. In the following, we describe the method for eye expression learning.

2.2.1 Preparation of learning image

First, we take pictures of user's eye expression images, and prepare 2 class images eye area cut out. Positive images are “squinting” eye expression images. Negative images are “other” eye expression images. Fig 4 shows example of learning image.

2.2.2 Histograms of Oriented Gradient Feature

HOG feature (Histograms of Oriented Gradient Feature) [2] is one of feature value by made histograms of intensity gradients. HOG feature is widely used in face recognition and generic object recognition. Algorithm shows in below.

- **Calculation of gradient intensity and direction**
  
  Gradients $f_x(x,y)$, $f_y(x,y)$ of brightness value $L(x,y)$ are equation 1 and equation 2 in a position of image $(x,y)$.
  
  $$
  \begin{align*}
  f_x(x,y) &= L(x+1,y) - L(x-1,y) \\
  f_y(x,y) &= L(x,y+1) - L(x,y-1)
  \end{align*}
  $$
  \tag{1}

  Next, Intensity $m$ and orientation $\theta(x,y)$ are culcubrated by gradients.
  
  $$
  m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2}
  $$
  \tag{2}

  $$
  \theta(x,y) = \tan^{-1} \frac{f_y(x,y)}{f_x(x,y)}
  $$
  \tag{2}

- **Make histograms of intensity and direction each cell size**
  
  Cell size is $c_w \times c_h$ pixel (15 × 15 pixel in this paper). Histograms of gradient directions are created each cell size. Orientation of gradient is divided into 9 for each direction of 20°.

- **Normalization of histograms**
  
  Block size is $b_w \times b_h$ cell (3 × 3 cell in this paper).
paper). Histograms are normalized by the block. The normalization equation is as follows. Here, \( h_{ij} \) is a histogram in a cell, \( v \) is a feature vector in block, and position \((i,j)\) runs \(1 \leq i \leq b_w, 1 \leq j \leq b_h\) .

\[
h_{ij} = \frac{h_{ij}}{\|v\|^2 + \epsilon}
\]  
(4)

Normalization is performed by moving 1 cell. So \( h_{ij} \) is normalized repeatedly.

2.2.3 Gentle Boosting

Based on the HOG feature, eye expression images are learned by Gentle Boosting. Gentle Boosting is derived from AdaBoost [3] which makes a decision on weighted vote from a number of weak classifiers. Gentle Boosting is more robust than AdaBoost against outliers. Two class classifiers, positive and negative, are created. The algorithm has been published in [4].

2.3 Eye Expression Recognition and View Image Presentation

“Eye expression recognition” is performed in the following steps.

- **Input eye image**
  
  When user wears wearable camera and the system is launched, eye image taken in one eye is captured from camera 1.

- **Cut out the eye area**
  
  As pretreatment, Eye image captured from camera 1 is cut out the eye area by using template matching. Template image is used as a learning image. Template matching is performed to estimate the regions of high similarity in eye image. In this study, a normalized cross-correlation function is used. Degree of similarity \( R \) is as follows. Input image's brightness value \( I(i,j) \) is in position\((i,j)\). Output image's brightness value \( T(i,j) \) is in position\((i,j)\).

  \[
  R = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} I(i,j)T(i,j)}{\left( \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} I(i,j)^2 \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} T(i,j)^2 \right)^{1/2}}
  \]  
(5)

High degree of similarity of this value is closer to 1. Effect of misalignment during installation is reduced by doing this process.

- **Extract HOG feature**
  
  Feature values are extracted from the eye image which is cut out by HOG feature. The algorithm is the same with 2.2.2.

- **Recognize the eye expression**
  
  The eye image is recognized positive or negative from extracting HOG feature.

“View Image Presentation” is performed in the following steps.

- **Input view image**
  
  If the eye image is recognized as “Squinting”, view image is captured from camera 2 at the same time.

- **Present view image**
  
  The view image is gradually enlarged toward the center, and presented to user's head mounted display. If user stops acting squinting, the system judges visible and stops the spread. Consequently, the system can be enlarged to magnification of any user.

3. Experiment

To confirm the effectiveness of proposal method, preliminary experiment was performed. In preliminary experiments, we confirmed recognition accuracy of eye expression. And the system was implemented, we also confirmed the entire system working.

3.1 Preliminary Experiment

In preliminary experiments, eye expression images for learning and evaluation were captured, when subject intentionally turned squinting. And we investigated the recognition rate. The detail is described below.

3.1.1 Experimental Method

First, we prepared learning images. Wearable camera attached to the subject. And Image sequences were captured, when the subject repeated “normal” expression and “squinting” expression every few seconds. Input image resolution was \(320 \times 240 \) pixel and frame rate was 30 frame/sec. From image sequences, positive and negative images were created each 500 by cut out the eye area. Positive images were “squinting”, and Negative images were “other”. Size of Image is \(260 \times 153 \) pixel.

Next, we prepared evaluation images. Image sequences were the same subject, but taken in the situation that installation of wearable camera was slightly misaligned.
Evaluation images were also prepared each 500. In extracting HOG feature, all image sizes were resized to 120 \times 75 pixel. Dimension of whole HOG feature was 1458. Learning by Gentle Boosting was 15 times.

3.1.2 Experimental Result

Table 1 shows identification result learning images and evaluation images. The rate of positive in evaluation was relatively low. However, it has no problem in practical use.

One cause of it is accuracy of detection of the eye area. Template image used in this experiment was a negative image with open eyes. Because input image was seen only one eye, template matching could broadly detect eye area. But misalignment of positive examples of incorrect were little larger than learning images. When misalignment occurs, HOG feature is tends to occurring false positives, because of focusing on the edges of the structure of local area. It should be considered for the feature of this environment.

3.2 Experiment

The entire system was implemented, and we confirmed eye expression recognition and view image scaled. The PC with implementation was OS:Windows 7 Professional, CPU:Intel Core i7 2.93GHz, and RAM:4GB.

3.2.1 Experimental Result

The same subject at the preliminary was worn wearable camera, and the system was launched. Resolutions of two cameras were 320 \times 240 pixel. When the subject was acting eye expression squinted at few seconds, we confirmed that view image was enlarged. To make sure that each process can be executed successfully, results were displayed on the PC monitor. Fig. 5 shows results of view image. The left was input view image, and the right was output view image. Eye expression recognition accuracy was the same with preliminary, because same learning images were used. And the processing time per 1 frame was measured. The average time was 30 msec to 50 msec. It has no problem in practical use.

4. Conclusions

In this paper, we proposed a user interface system with wearable camera for low vision people. Eye expression images of squinting were learned by Gentle Boosting. In preliminary experiment, it was confirmed that eye expression recognition was relatively high accuracy. And we implemented this method, it was also confirmed that it can operate to the scaled view image.

In the experiment, we defined that the user's target to see was center, when user was acting squinting. So, scaling operations in the view image was toward to the center of the image. But the target may be located outside of the center in fact. To solve this problem, analyzing the user's gaze is needed for identify the target to see.

In the future, low vision are participated in the experiment, we will make a detailed study of this.

Acknowledgment

A part of this work was supported by Grant-in-Aid for JSPS Fellows Grant Number 24·9839.

References


<table>
<thead>
<tr>
<th>Table 1. Identification result.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Positive (Learning)</td>
</tr>
<tr>
<td>Negative (Learning)</td>
</tr>
<tr>
<td>Positive (Evaluation)</td>
</tr>
<tr>
<td>Negative (Evaluation)</td>
</tr>
</tbody>
</table>

Fig. 5. Results of view image