Long Range Gesture Recognition in Robot Environments with a Single Camera

DoHyung Kim, Woo-han Yun, and Jaeyeon Lee
Robot Research Division, ETRI
{dhkim008, yochin, leejy}@etri.re.kr

Abstract

This paper proposes a vision-based human's arm gesture recognition method for human robot interaction particularly at a long distance where speech information is not available. We define four meaningful arm gestures for a long range interaction. The proposed method is able to recognize the defined gestures only with 320x240 pixel-sized low-resolution input images captured from a single camera at a long distance, approximately five meters distance from the camera. In addition, the system differentiates the target gestures from the users' normal actions that occur in daily life without any constraints. For human detection at a long distance, the proposed approach combines results from mean-shift color tracking, short- and long-range face detection, and omega shape detection. The system then detects arm blocks by using a background subtraction method with a background updating module and recognizes the target gestures based on the information about region, periodical motion, and shape of the arm blocks. From the experiments on a realistic and large database, a recognition rate of 97.235% is achieved, which is a sufficiently practical level for various robot applications based on human's gestures.

Keywords: gesture recognition, long-range human robot interaction

1. Introduction

Gesture recognition is not a new issue. Numerous researches on gesture recognition have been explored over the years [1]. However, there are some limitations as follows when we develop a gesture recognizer for a long-range human robot interaction with the existing methods.

- The existing gesture recognizers do not provide meaningful gestures and recognition method for human robot interaction at a long distance yet.
- Some presented studies that can recognize remotely upper body or whole body human gesture are inappropriate for low-cost robots with a single camera, because most of them use high-priced equipments including high-resolution cameras, stereo cameras, or 3D scanners.
- Other existing methods that recognize long range gestures only with a single camera assume the environments of the fixed camera. Hence, it is impossible to operate them on mobile robot.
- For reliable operations, many gesture recognizers require some constraints to users. However, it is unrealistic to expect users to cooperate for the recognition purpose in robot environments that users are freely acting.

Therefore, we need to design a novel gesture recognizer in order to satisfy the requirements in robot environments.

2. Gesture Recognition

As shown in Figure 1, we define four meaningful gestures for human robot interaction at a long distance.

- Waving is a motion gesture waving his/her right arm right and left for three seconds or so. Human can use the waving gesture in order to attract robot's attention.
- Calling is also a motion gesture waving his/her right arm up and down for three seconds or so. The calling gesture may be used to beckon a robot to come to him/her.
- Raising is a still gesture raising straight his/her right arm for three seconds or so. This gesture is defined for human to show his/her ID indirectly or say "yes" to robots.
- Stopping is also a still gesture raising his/her both arms to his/her head for three seconds or so. Human can say "stop" or "no" to robots by acting this gesture.
In order to recognize the defined target gestures for long-range interactions, the proposed system follows the flow shown in Figure 2.

When the 320x240 pixel-sized low-resolution input images are fed into the system, a human detection module is first triggered. The human detector is designed to detect a minimal of 12x12 pixel-sized faces located at a maximal range of five meters by combining the results from a mean-shift color tracker [2], short and long rage face detectors [3], and an omega shape detector effectively.

The 12x12 pixel size is too small size to include all of the main facial components such as eyes, nose, and mouth, so it is extremely difficult to detect such tiny-sized faces reliably even with moving cameras. We solve this problem by adopting a tiny face detector that is well trained enough to detect faces at a long range, and limiting its operation to only within a search region that is determined automatically. Figure 3 shows examples of detected small faces.

The proposed method then determines a gesture search region that consists of five sub-regions based on the detected face region as shown in Figure 4. Since the four target gestures are only appeared in the established search region, the method examines if the target gestures appear only in the region.

When the system detects user’s face and determines the gesture search region successfully, a capturing process of background image is triggered to detect arm blocks as shown in Figure 5 by using a background subtraction method. Since robots normally move around freely, we implemented a partial background updating module only for the gesture search region. That is, the system replaces a background image by a current image if arm blocks do not exist in the search region.

After segmenting arm blocks, the system finally recognizes the target gestures based on the information about region, periodical motion, and shape of the arm blocks. We first investigate which sub-regions have the detected arm blocks in order to remove noise blocks or users’ normal actions regarded as noise gestures. And then, for the recognition of motion gestures such as the waving and calling gesture, the system analyzes a moving direction of a right arm block. If the right arm block moves periodically, the system assumes that the moving gestures occurred. The gesture recognizer also examines a shape of the arm blocks including length,
size, and position in order to recognize the still gestures such as the raising and stopping gesture.

Figure 6. Determined 11 users’ normal actions that occur in daily life (noise gestures)

3. Experimental Results

To evaluate the performance of the proposed system, 800 video clips of the target gestures and 550 video clips of users’ normal actions were collected from ten users within three to five meters distance in general home environments. For video clips of target gestures, users were requested to look at the robot and act arm gestures for three seconds or so. We determined 11 noise gestures including scratching a head, clasping hands above a head, stretching a body with raised hands and others.

As shown in Table 1, an average recognition rate is 97.235%, which is a sufficiently practical level for various robot applications. The misclassified cases in which users’ normal actions are recognized as the target gestures represented 4.44% of the cases.

Table 1. The performance of the proposed system

<table>
<thead>
<tr>
<th>Target Gestures</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waving</td>
<td>94.50</td>
</tr>
<tr>
<td>Calling</td>
<td>97.96</td>
</tr>
<tr>
<td>Raising</td>
<td>98.50</td>
</tr>
<tr>
<td>Stopping</td>
<td>97.98</td>
</tr>
<tr>
<td>Aver.</td>
<td>97.235</td>
</tr>
</tbody>
</table>

4. Conclusion

This paper proposes a novel human's arm gesture recognition method for mobile robots to recognize four target gestures at a long range of nearly five meters. In addition, our method does not request any constraints to users and is able to distinguish the target gestures from users’ normal actions. We expect that many robot applications will make good use of the proposed method for human robot interaction at a long distance where speech information could not be available.

Acknowledgement

This work was supported by the IT R&D program of MKE & ITEP [2008-F-037-01, Development of HRI Solutions and Core Chipsets for u-Robot]

References