Head gesture recognition for hands-free control of an intelligent wheelchair

Pei Jia and Huosheng H. Hu
Department of Computer Science, University of Essex, Colchester, UK, and
Tao Lu and Kui Yuan
Institute of Automation, Chinese Academy of Sciences, Beijing, People’s Republic of China

Abstract
Purpose – This paper presents a novel hands-free control system for intelligent wheelchairs (IWs) based on visual recognition of head gestures.
Design/methodology/approach – A robust head gesture-based interface (HGI), is designed for head gesture recognition of the RoboChair user. The recognised gestures are used to generate motion control commands to the low-level DSP motion controller so that it can control the motion of the RoboChair according to the user’s intention. Adaboost face detection algorithm and Camshift object tracking algorithm are combined in our system to achieve accurate face detection, tracking and gesture recognition in real time. It is intended to be used as a human-friendly interface for elderly and disabled people to operate our intelligent wheelchair using their head gestures rather than their hands.
Findings – This is an extremely useful system for the users who have restricted limb movements caused by some diseases such as Parkinson’s disease and quadriplegics.
Practical implications – In this paper, a novel integrated approach to real-time face detection, tracking and gesture recognition is proposed, namely HGI.
Originality/value – It is an useful human-robot interface for IWs.

Keywords Wheelchairs, Automation

Paper type Research paper

1. Introduction
To improve quality of life for the elderly and disabled people, electric-powered wheelchairs (EPWs) have been rapidly deployed over the last 20 years (Ding and Cooper, 1995; Simpson et al., 2004; Galindo et al., 2005). Up to now, most of these EPWs are controlled by users’ hands via joysticks, and are very difficult for elderly and disabled users who have restricted limb movements caused by Parkinson’s disease and quadriplegics. As cheap computers and sensors are embedded into EPWs, they becomes more intelligent, and are named intelligent wheelchairs (IWs). Various research and developments on IWs have been carried out in the last decade, such as TAO projects (Gomi and Griffith, 1998), Rolland (Röfer and Lankenau, 1998), Maid (Prassler et al., 1998), NavChair (Levine et al., 1999), UPenn Smart Wheelchair (Rao et al., 2002), SIAMO (Mazo et al., 2002).

The successful deployment of IWs requires high performance and low cost. Like all the other intelligent service robots, the main performance of IWs includes:
• the autonomous navigation capability for safety, flexibility, mobility, obstacle avoidance; and
• the intelligent interface between users and IWs, including hand-based control (joystick, keyboard, mouse, touch screen), voice-based control (audio), vision-based control (cameras), and other sensor-based control (infrared sensors, sonar sensors, pressure sensors).

As an hands-free interface, head gestures and EMG signals have already been applied in some existing IWs, such as WATSON (Matsumoto and Zelinsky, 1999; Matsumoto et al., 2001), OSAKA wheelchair (Nakanishi et al., 1999; Kuno et al., 2001), NLPRWheelchair (Wei, 2004), EMGWheelchair (Moon et al., 2005). However, these systems are not robust enough to be deployed in the real world and much improvement is necessary. The new generation of head gesture-based control of wheelchairs should be able to deal with the following uncertainty in the practical applications of IWs, the:
• background may be cluttered and dynamically changing when IWs move in the real world.
Head gesture recognition

Pei Jia, Huosheng H. Hu, Tao Lu and Kai Yuan

- user may have different facial appearances at different time, such as mustache and glasses;
- face color may change dramatically in varying illumination conditions; and
- user’s head may move around for looking rather than moving.

In this paper, a novel head gesture-based interface (HGI) is developed for the IWs, namely RoboChair, based on the integration of the Adaboost face detection algorithm (Bradski, 1998) and the Camshift object tracking algorithm (Viola and Jones, 2004). In our approach, head gesture recognition is conducted by means of real-time face detection and tracking. The developed HGI aims to solve the problems listed above and provide a useful human-robot interface for our RoboChair.

The rest of the paper is organised as follows. Section 2 presents the system hardware structure of our RoboChair. In section 3, we propose a new HGI for the control of our RoboChair, which is based on the combination of both Adaboost and Camshift algorithms. Experimental results are presented in section 4 to show the feasibility and performance of our new algorithm. Finally, a brief conclusion and future work are presented in section 5.

2. System hardware structure

Figure 1 shows the picture of our RoboChair that was built in 2004 and has the following components:

- six ultrasonic sensors at a height of 50 cm for obstacle avoidance (four at the front and two at the back);
- DSP TMS320LF2407-based controller for motion control of two differentially-driven wheels;
- a local joystick controller to connect to an A/D converter of the DSP-based controller;
- a Logitech 4000 Pro Webcam for recognising the user’s head gestures; and
- Intel Pentium-M 1.6G Centrino laptop with WindowsXP installed to analyze the head gestures.

The control architecture of our RoboChair is shown in Figure 2. The TI DSP chip TMS320LF2407 is used as the core processor of the motion control module. It offers excellent processing capabilities (30 MIPS) and a compact peripheral integration, so that the control system is able to achieve both real-time signal processing and high performance driving control. Note that an obstacle avoidance module is embedded in the DSP motion controller for safe operation. Our RoboChair has two control modes, namely intelligent control and manual control.

2.1 Intelligent control mode

This is currently being developed in this project. Under this control mode, our RoboChair is controlled by the proposed HGI. A Logitech webcam is used to acquire the facial images of the user. After the image data is sent to the laptop, head gesture analysis and decision-making stages are implemented. Finally, the laptop sends control decisions to the DSP motion controller that actuates two DC motors.

2.2 Manual control mode

This is already been built in our RoboChair. Under this control mode, our RoboChair is controlled by the joystick that is directly connected to an A/D converter of the DSP motion controller.

It should be noticed that no matter which control mode our RoboChair is working under, in order to deal with uncertainties in the real world, sonar readings are directly sent to the DSP motion controller for real-time obstacle avoidance and emergency handling.

3. HGI for RoboChair

3.1 Adaboost and Camshift algorithms

Adaboost is the most recent face detection method with both high accuracy and fast speed (Viola and Jone, 2004). It extracts the Haar-like features of images that contain image frequency information. This process is very fast since only integer calculation is implemented. Then a set of key features are selected from these extracted features. After being sorted according to the importance, this set of features can be used as a cascade of classifiers that are very robust and able to detect various faces under varying illumination conditions and different face colors. Also, Adaboost is able to detect profile faces.

Figure 3 shows the block diagram of Adaboost face detection algorithm. It consists of a sequence of stages:

1. data acquisition (image capturing);
2. pre-processing (filtering);
3. feature extraction (rectangular features); and
4. a parallel stage: classifiers design (boosted cascade design) and classification.

In the classifier design, supervised learning is adopted to select Adaboost features.

Camshift is a very efficient color tracking method based on image hue (Bradski, 1998), and is in fact a classical optimization algorithm. It uses a robust non-parametric technique for claiming density gradients to find the mode (peak) of probability distribution called the mean shift algorithm. Each iteration, Camshift aims to find the mean window center using a fixed window size. If either the window center or the window size is unstable, both values need to be adjusted accordingly until convergence.

Figure 4 shows the flowchart of the Camshift object tracking process, which consists of four stages:
1 initialization;
2 window size adjustment (control);
3 target search; and
4 solutions.

Note that Camshift has some limitations, it cannot:
• accurately track the face when the illumination condition changes; and
• work well under the cluttered background.

3.2 Integration of Adaboost and Camshift algorithms
Since, low-cost IWs have limited onboard computing power, Adaboost face detection algorithm cannot achieve real-time performance. On the other hand, Camshift face tracking algorithm runs very fast, but is not robust to varying illumination conditions and noisy backgrounds. In order to obtain both fast speed and high accuracy, it is necessary to integrate both algorithms in a unified framework as shown in Figure 5 (Jia and Hu, 2005).

Every frame, the system is trying to keep tracking the user’s face which is always in front of the webcam in our RoboChair application:
• If the user’s face is tracked successfully, Adaboost face detection is applied in a comparatively small Camshift tracking window so that the face position and size, as well as head gesture (including frontal, left and right profiles) can be obtained rapidly. Further direction judgement is needed when frontal face is detected. Here, a simple scaled nose template matching is used to calculate the nose position in the frontal face window, through which face directions can be determined very accurately.
• If the user’s face is lost from camera view, Adaboost face detection is applied in the whole captured image frame so that the user’s face will be tracked once again.

3.3 Head gesture recognition
Intel OpenCV has already contained the trained frontal and right profile face classifiers. After simply flipping the captured image, the left profile face can be detected as well. In order to recognise the head gesture under any special situations robustly, Adaboost frontal, left profile and right profile head gesture classifiers are adopted in our research.

Since, Adaboost is an appearance-based face detection method, the left face appearance is quite similar to the right face appearance. Occasionally, it may have difficulty to decide whether a face is left profile or right profile. Therefore, a simple strategy is used to solve this situation: the profile face with bigger detection window dominates the head gesture. When both left profile and right profile detection windows are of the same size, the wheelchair will keep the status of the previous cycle.

If the profile face is detected, our RoboChair is going to turn left or right. If the frontal face is detected, further left
Head gesture recognition

Pei Jia, Huosheng H. Hu, Tao Lu and Kui Yuan

Industrial Robot: An International Journal
Volume 34 · Number 1 · 2007 · 60–68

Figure 3 Block diagram of the Adaboost face detection process

Figure 4 Flowchart of the Camshift object tracking process

frontal/right frontal/up frontal/down frontal/center frontal head gesture is to be recognised. Because of the varying distance from the face to the webcam, the detected face windows in different frames are not of the same size. Thus, it is necessary to scale the detected face windows to a standard size. Here, we use a size of $100 \times 100$ pixels as a standard face window. Then, the classical template matching method is applied in this small window to calculate the precise nose position, which will tell the exact frontal face head gestures.
3.4 Motion control commands

Owing to a big difference between the frontal face and the profile face, Adaboost detects frontal and profile head gestures separately. Also, when the user just moves his/her head to look at something, but does not want to move the wheelchair, our HGI should be able to distinguish this situation and avoid generating any unnecessary action. To achieve this, we restrict the on-board camera to focus on the face of the user who sits right in the center of the wheelchair. If the user's head and face are outside of the central position, our HGI will treat this situation as the user has no intention to control the wheelchair using head gestures. Also, we assume that useful head gestures should have a range of 45° turning angles on each side (up, down, left and right). If the turning angles of head gestures being detected are out of the range, our HGI will notify this and no motion control commands will be generated. If head gestures are detected within the specified range, RoboChair will act according to the following rules:

- Speed up, if frontal face up is recognised.
- Slow down until stop, if frontal face down is recognised.
Turn left, if left frontal/profile face is recognised.
Turn right, if right frontal/profile face is recognised.
Keep speed, if central face is recognised.

4. Experimental results

4.1 Adaboost face detection vs Camshift face tracking
In this experiment, we tested the performance of Adaboost and Camshift algorithms separately under ten special conditions:
1 normal;
2 bright;
3 darkness;
4 face color noise;
5 occlusion;
6 different face color;
7 multiple faces with different colors;
8 multiple faces with similar colors;
9 profile; and
10 profile with bright lighting.

As shown in Figure 6, the upper image in each condition shows the performance of Adaboost face detection algorithm, and the lower image shows the performance of Camshift face tracking algorithm. It is clear that the performance of Adaboost is very robust under different lighting and noise conditions and faces can be detected very accurately. However, the performance of Camshift is not robust although it is very fast. In some cases, faces cannot be detected accurately, such as in Figure 6(c), (d), (g) and (j).

4.2 Speed issues for Adaboost and Camshift
Table I shows that, under the same conditions, the proposed method is much faster than applying Adaboost alone, without taking the image capturing time into consideration. The experiments are finished in WindowXP on the Intel Pentium-M 1.6G Centrino laptop, with the image resolution 640 × 480 and the minimum face size 20 × 20 or 40 × 40.

As shown in Figure 6, the upper image in each condition shows the performance of Adaboost face detection algorithm, and the lower image shows the performance of Camshift face tracking algorithm. It is clear that the performance of Adaboost is very robust under different lighting and noise conditions and faces can be detected very accurately. However, the performance of Camshift is not robust although it is very fast. In some cases, faces cannot be detected accurately, such as in Figure 6(c), (d), (g) and (j).

4.4 RoboChair demonstration
Finally, RoboChair demonstrations are presented here to show the feasibility and performance of the proposed HGI framework. Figure 8 shows a sequence of images that our RoboChair is entering and passing a gate:
• start moving forward;
• turning right;

Table I Speed of Adaboost and Camshift

<table>
<thead>
<tr>
<th>Method</th>
<th>Minimum face size</th>
<th>Time cost per frame (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaboost</td>
<td>20 × 20</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>40 × 40</td>
<td>0.29</td>
</tr>
<tr>
<td>Integration</td>
<td>20 × 20</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>40 × 40</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The results are statistically calculated during a period of five minutes.

4.3 Nose template matching for recognition of frontal face posture
Unlike left profile and right profile head gestures, further investigation is conducted to recognise frontal head gestures. Figure 7 shows clearly that our proposed head gesture recognition method is fairly feasible. There are five frontal head gestures to be recognised, namely:
1 center frontal;
2 up frontal;
3 down frontal;
4 left frontal; and
5 right frontal.

Each head gesture has two images: the upper image shows the detected face and the lower image shows the recognised head gesture. As can be seen, the small rectangle indicates the face posture based on its relation with the big rectangular box.

Figure 6 Robustness of Adaboost face detection (row 1, 3) and inaccuracy of Camshift face tracking (row 2, 4)
It is clear that the proposed HGI is able to recognise the user’s head gesture for hands-free control of our RoboChair. No manual operation is needed, which will make the user’s life much easier and comfortable.

Figure 9 shows a sequence of images demonstrated by our RoboChair under head gesture control:
• turn right;
• continue turning right with head posture “right up”;
• turn left;
• turn left with hand color noise;
• continue turning left with hand color noise; and
• forward.

In this demonstration, some uncertainty such as hand color noise has been deliberately added into images. When head gestures are not in normal postures, i.e. the non-vertical head gestures in Figure 9(d) and (e), HGI is still able to identify the user’s intention and control the RoboChair very well, which is very robust. However, HGI will ignore head gestures if the user’s head is not located in the center of images or is looking around the surroundings without the intention of moving.

5. Conclusion and future work

This paper describes the design and implementation of a novel hands-free control system for IWs. The developed system provides enhanced mobility for the elderly and disabled people who have very restricted limb movements or severe handicaps. A robust HGI, is designed for vision-based head gesture recognition of the RoboChair user. The recognised gestures are used to generate motion control commands so that the RoboChair can be controlled according to the user’s intention. To avoid unnecessary movements caused by the user looking around randomly, our HGI is focused on the central position of the wheelchair to identify useful head gestures.

Our future research will be focused on some extensive experiments and evaluation of our HGI in both indoor and outdoor environments where cluttered backgrounds, changing lighting conditions, sunshine and shadows may bring complications to head gesture recognition.
References


Institute of Automation, Chinese Academic of Sciences, Beijing.

Further reading


Corresponding author
Huosheng H. Hu can be contacted at: hhu@essex.ac.uk