Gesture Recognition Method with Illumination Robustness for Wheelchairs

Bin Zhang, Xianyun Luo and Qiang Gao*
School of Mechanical and Electrical Engineering, Soochow University, China

*Corresponding author email: gaoqiang@suda.edu.cn

Abstract. Disabled elderly people have to rely on wheelchair life. And the wheelchairs are used in daily life with different illumination conditions. This paper proposes a hand gesture recognition method to control wheelchairs based on dynamic colour clustering. As dynamic clustering skin detection algorithm is used, the gesture recognition accuracy is improved. Using skin detection, arm removal and polygon fitting methods, the gesture images collected by a Kinect V2, are recognized as control commands to drive the wheelchair. Experiments show that the proposed method can be applied to different illumination conditions with better recognition accuracy.

1. Introduction

At present, the global population is aging, and about 10% of the population is aged 60 or over. It is expected that by 2050, this number will increase to 20% [1], approaching 1.5 billion people and entering the ultra-aged society. Also, more than one billion people in the world suffer from a disability, and about one-third of them are physically disabled [2]. The medical and nursing care of this special group not only puts a huge burden on families but also becomes a huge pressure on economic and social development. With the increase of age, the physical fitness of the elderly is declining, and their mobility is gradually weakening. Many elderly people with mobility problems choose wheelchairs as a means of mobility. According to World Health Organization (WHO) statistics, about 75 million people worldwide need wheelchairs to assist in their daily lives [3].

The joystick control method is the most commonly used mechanism for navigation of a wheelchair. However, due to the different strengths of the hands, not all wheelchair users, including the elderly, physically disabled, stroke patients and patients with spinal cord injuries, can use the joystick [4]. In addition to the joystick, the currently common control methods are touch screen [5], voice [6], eye-tracking [7], brain wave [8], limb movements [9] and hand gesture control [10, 11]. For the elderly with weak eyesight and slow movement, the operation of touch screen is more complicated, and observing the road conditions and screen will lead to distraction and reduced safety; Voice recognition relies on clear and accurate commands and a quiet environment; Eye-tracking control requires users to concentrate for a long time during the control process, which is prone to eye fatigue, and gazing at the camera will also cause safety problems caused by distraction; Brain wave control relies on various sensors affixed to the head, which results in poor comfort. The hand is the most flexible part of the human body. The intuitiveness and flexibility of human hands contain a lot of non-verbal information. It is not only an important medium for communication between people but also an important choice for human-computer interaction.

Previous research based on gesture control often have problems such as insufficient driving angle information, complex gesture instructions that cause memory burden, and gesture instructions that are
too standard. A Kinect depth camera is used to obtain color and depth images of the hand area. After hand area segmentation and polygon fitting, the gestures are simplified to achieve precise control commands for the wheelchair.

The use environment of the wheelchair includes indoor and outdoor. The lighting conditions vary greatly in different environment. The change of lighting directly affects the segmentation effect of the hand area, and thus affects the accuracy of gesture recognition. The segmentation of the hand area depends on the skin detection algorithm. It is difficult to adapt the skin detection using traditional thresholding methods to different lighting conditions. In this paper, the dynamic clustering skin detection algorithm is applied [12], which makes the gesture recognition algorithm less sensitive to light. The experiments show that the skin detection algorithm implemented effectively improves the accuracy of gesture recognition and has better illumination robustness.

2. Gesture Recognition Method with Illumination Robustness

Kinect V2 is an image acquisition device in the wheelchair designed in this article (figure 1). The hand region should be extracted and the invalid background and arm region should be removed because the algorithm proposed relies on hand-fitted polygon images. First, the depth camera will take a depth image of a specified depth range of 0.5m to 0.7m. To further eliminate irrelevant information in the image, a pre-processing operation is performed to obtain the hand image with suitable size and distinctive and reliable features (figure 2). The image preprocessing operations include Gaussian filtering [13] and median filtering [14].

2.1. Skin Detection

![Figure 1. The proposed wheelchair [10].](image1)

![Figure 2. Preprocessed image.](image2)

![Figure 3. Clustering regions in YCr and YCb subspace.](image3)
In this paper, 4000 hand images under different lighting conditions were collected. We calibrated the skin regions and obtained clusters of skin pixels in the YCr and YCb subspace, which is shown in figure 3. The two clustering regions can be approximated by two trapezia. Figure 4 shows a simplified model, where $C_b_{\text{min}} = 77$, $C_b_{\text{max}} = 130$, $C_r_{\text{min}} = 131$ and $C_r_{\text{max}} = 183$.

For a new input image, the number of pixels $N(C_r)$ ($i = 0, 1, \ldots, 255$) denote to different $C_r$ values, and record the total number of pixels as $N$. Let $C_r = C_r_{\text{max}}$ for at least $N_0 = 0.1N$ pixels in the image $[C_r_{\text{min}}, C_r_{\text{max}}]$. Similarly, find the minimum value of $C_b$ in the range $[C_b_{\text{min}}, C_b_{\text{max}}]$ that is related to at least 10% of the pixels in the image. The number $N(Y)$ of different brightness values $Y$ of the pixels of $C_r = C_r_{\text{max}}$, smf take $Y_0$ and $Y_1$ at 5% and 95%. Calculate $Y_2$ and $Y_3$ in the same way. Other parameters are calculated as $h_{C_r} = C_r_{\text{max}} - C_r_{\text{min}}$, $h_{C_b} = C_b_{\text{max}} - C_b_{\text{min}}$. The trapezoidal boundaries $M_{C_r}(Y_p)$ and $M_{C_b}(Y_p)$ of the model are calculated as follows:

$$M_{C_r}(Y_p) = \begin{cases} C_r_{\text{min}} + h_{C_r} \left( \frac{Y - Y_{\text{min}}}{Y_0 - Y_{\text{min}}} \right), & Y \in [Y_{\text{min}}, Y_0] \\ C_r_{\text{max}}, & Y \in [Y_0, Y_1] \\ C_r_{\text{min}} + h_{C_r} \left( \frac{Y - Y_{\text{max}}}{Y_1 - Y_{\text{max}}} \right), & Y \in [Y_1, Y_{\text{max}}] \end{cases}$$  

$$M_{C_b}(Y_p) = \begin{cases} C_b_{\text{min}} + h_{C_b} \left( \frac{Y - Y_{\text{min}}}{Y_0 - Y_{\text{min}}} \right), & Y \in [Y_{\text{min}}, Y_0] \\ C_b_{\text{max}}, & Y \in [Y_0, Y_1] \\ C_b_{\text{min}} + h_{C_b} \left( \frac{Y - Y_{\text{max}}}{Y_1 - Y_{\text{max}}} \right), & Y \in [Y_1, Y_{\text{max}}] \end{cases}$$  

For the pixels to be detected, set:

$$dC_r = C_r - C_r_{\text{min}}$$  

$$\Delta C_r(Y_p) = M_{C_r}(Y_p) - C_r_{\text{min}}$$  

$$\Delta C_b(Y_p) = C_b_{\text{max}} - M_{C_b}(Y_p)$$  

Normalize $\Delta C_r(Y_p)$ and $\Delta C_b(Y_p)$ according to equation (4) and equation (5):
\[ \Delta'Cr(Y_p) = \begin{cases} \frac{\Delta Cr(Y_p)}{A_{YC}}, & A_{YC} \geq A_{YCB} \\ \Delta Cr(Y_p), & \text{otherwise} \end{cases} \]  

(4)

\[ \Delta'Cb(Y_p) = \begin{cases} \frac{\Delta Cb(Y_p)}{A_{YC}}, & A_{YC} \geq A_{YCB} \\ \Delta Cb(Y_p), & \text{otherwise} \end{cases} \]  

(5)

where \( A_{YC} \) and \( A_{YCB} \) are the areas of the trapezia. Take the ratio of \( \Delta'Cb(Y_p) \) and \( \Delta'Cr(Y_p) \) as \( \alpha \). Calculate the threshold as below:

\[ \begin{align*}
I_p &= \beta \cdot \left[ \Delta'Cr(Y_p) - dCr_p \right] + \left( \Delta'Cb(Y_p) - dCb_p \right) \\
J_p &= \alpha \Delta Cb(Y_p) + dCb_p
\end{align*} \]  

(6)

where \( \beta = \min((Y_1 - Y_0), (Y_3 - Y_2)) / \max((Y_1 - Y_0), (Y_3 - Y_2)) \).

Let \( Cb_{ps} = Cb_{max} - dCb_{ps} \) and a pixel is classified as a skin pixel if it satisfies equation (7).

\[ \begin{align*}
C_p - Cb_p &\geq I_p \\
|Cb_p - Cb_{ps}| &\leq J_p
\end{align*} \]  

(7)

2.2. Arm Region Removal and Polygon Fitting

To further adapt to situations where the arms are bare or the color of the coat worn is similar to the skin color, the arm region should be removed. Figure 5 shows the process of arm region removal and polygon fitting. First, Euclidean distance transformation is performed on the image. The grey value of each pixel after the distance transformation represents the minimum distance from the boundary. Since the palm width is larger than the wrist and arm, the gray value at the center of the palm is generally the largest, and the gray value is the largest. The interference area of the arm is eliminated according to the palm position and the inscribed circle of the palm. The removal result is shown in figure 5 (c).

To make the input image features of gesture recognition more prominent, binarization processing and polygon fitting are performed on the image after arm removal to reduce the number of contour points of the image, while maximally retaining the shape features of the gesture. In this paper, the polygon fitting algorithm is implemented by Douglas-Peucker algorithm [15]. Figure 5(d) shows the gesture image after polygon fitting.

\[ \text{Figure 5. Arm region removal and polygon fitting: (a) Skin detection, (b) Distance transformation, (c) Arm region removal, (d) Polygon fitting} \]

This case simply justify the caption so that it is as the same width as the graphic.

2.3. Classification of the Hand Gestures

The elderly are the primary users of wheelchairs. The definition of gestures must consider their behavioral habits and psychological characteristics. Therefore, gestures must be easy to operate and remember. In particular, complicated gesture instructions should be avoided. This article defines three categories of natural gestures for controlling wheelchair movement: pointing gestures, fist gestures, and 5-finger open gestures. The pointing gestures are subdivided into two types according to the different ways of extracting the pointing direction. The directions of Point 1 and Point 2 are defined as the left and right contour of
the index finger respectively. The definitions and functions of the four types of gestures are shown in table 1.

Table 1. Gesture classification and function definition.

| Polygon images | Classification | Function |
|----------------|----------------|----------|
|                | Point 1        | Navigation |
|                | Point 2        | Navigation |
|                | 5-finger open  | Stop      |
|                | Fist           | Back      |

3. Experiments

3.1. Skin Detection Results Comparison

Figure 6. Skin detection images ranged in input image, threshold method and proposed method: (a) outdoor (cloudy), (b) outdoor (sunlight), (c) outdoor (shade), (d) indoor (light), (e) indoor (no light), (f) indoor (high light).

As shown in figure 6, in several typical use scenarios of wheelchairs, skin detection is performed using the YCrCb threshold method and the proposed skin detection algorithm respectively. Experiments show that compared with the threshold method, the skin detection algorithm in this paper has less noise and better detection effect.

3.2. Classification Results Comparison

Model training is performed by support vector machine (SVM). The distance from the contour point to the palm after the polygon fitting and Hu invariant moments are used as feature vectors. A total of 1,000 pictures are selected as the data set, and the ratio of the training set to the test set is 7:3. Table 2 shows the test results of the test set using the threshold method and the proposed skin detection algorithm. The algorithm effectively improves the accuracy of SVM classification.
Table 2. SVM model test results.

| Classification     | Image samples | Accuracy (threshold method) (%) | Accuracy (proposed method) (%) |
|--------------------|---------------|---------------------------------|-------------------------------|
| Point 1            |               | 92.65                           | 98.66                         |
| Point 2            |               | 93.84                           | 99.64                         |
| 5-finger open      |               | 98.03                           | 99.67                         |
| Fist               |               | 100                             | 100                           |
| Total              |               | 96.13                           | 99.95                         |

4. Conclusion
This paper proposes a wheelchair gesture recognition algorithm with illumination robustness. The algorithm includes image acquisition, hand region segmentation, polygon fitting, and gesture classification. Wheelchairs are used in complex scenes with different illumination conditions. The accuracy of gesture recognition is affected by the performance of skin detection. By introducing a dynamic clustering skin detection algorithm, the influence of illumination on the detection effect is reduced, and the accuracy of gesture recognition is improved. The effectiveness of the hand gesture algorithm is proved.

References
[1] 2013 Functional decline and dependence in ageing populations. Anel side event at 66th World Health Assembly https://www.who.int/ageing/events/wha66/en/.
[2] World Health O 2011 World report on disability 2011: World Health Organization
[3] 2018 Assistive technology https://www.who.int/en/news-room/fact-sheets/detail/assistive-technology.
[4] Trujillo-Leon A, Bachta W and Vidal-Verdu F 2018 Tactile Sensor-Based Steering as a Substitute of the Attendant Joystick in Powered Wheelchairs IEEE Transactions on Neural Systems and Rehabilitation Engineering 26 1381-90
[5] Filipescu A, Solea R, Filipescu A, Stamatescu G and Ciubucciu G 2018 Trajectory-Tracking Sliding-Mode Control of the Autonomous Wheelchair Modeled as a Nonholonomic WMR. In: 2018 IEEE 14th International Conference on Control and Automation (ICCA), 12-15 June 2018, (Piscataway, NJ, USA: IEEE) pp 1168-73
[6] Kolozcaronyari A, Stojanovicacute R, Zupan A, Semenkin E, Stanovov V, Kofjaccaron D and Skraba A 2019 Speech-recognition cloud harvesting for improving the navigation of cyber-physical wheelchairs for disabled persons Microprocessors and Microsystems 69 179-87
[7] Li T, Yang J, Bai D and Wang Y 2018 A New Directional Intention Identification Approach for Intelligent Wheelchair Based on Fusion of EOG Signal and Eye Movement Signal. In: 2018 International Conference on Intelligence and Safety for Robotics, ISR 2018, August 24, 2018 - August 27, 2018, (Shenyang, China: Institute of Electrical and Electronics Engineers Inc.) pp 470-4
[8] Lin Y-T and Kuo C-H 2016 Development of SSVEP-based intelligent wheelchair brain computer interface assisted by reactive obstacle avoidance. In: 2016 IEEE International Conference on Industrial Technology (ICIT): IEEE) pp 1572-7
[9] Dey P, Hasan M M, Mostofa S and Rana A I 2019 Smart wheelchair integrating head gesture navigation. In: 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST): IEEE) pp 329-34
[10] Jiayi Z 2018 Research on gesture interaction and autonomous navigation of the intelligent wheelchair in unknown environment
[11] Kawarazaki N and Diaz A I B Gesture recognition system for wheelchair control using a depth sensor. In: 2013 IEEE Symposium on Computational Intelligence in Rehabilitation and Assistive Technologies (CIRAT): IEEE pp 48-53
[12] Brancati N, De Pietro G, Frucci M and Gallo L 2017 Human skin detection through correlation rules between the YCb and YCr subspaces based on dynamic color clustering Computer Vision and Image Understanding 155 33-42
[13] Deng G and Cahill L W 1993 An adaptive Gaussian filter for noise reduction and edge detection. In: 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, pp 1615-9 vol.3
[14] Sun T and Neuvo Y 1994 Detail-preserving median based filters in image processing Pattern Recognition Letters 15 341-7
[15] Prasad D K, Leung M K H, Quek C and Cho S-Y 2012 A novel framework for making dominant point detection methods non-parametric Image and Vision Computing vol 30 843-59