Hand Gesture Based Control Strategy for Mobile Robots

Hang Zhao, Jiangping Hu, Yuping Zhang, Hong Cheng
School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731
E-mail: zg1991ideal@163.com

Abstract: In this paper, a hand gesture based control design is proposed for mobile robots. Mobile robots can move according to the control signals encoded by hand gestures. The gesture region is segmented from complicated background and the gestures are recognized by using some techniques such as image processing, image filtering processing, morphological image processing, image contour processing, etc. Then a template matching algorithm is proposed with the help of the invariant moment matching method to recognize the hand gestures. The recognition results are decoded as feedback information to control the mobile robots. Finally, some simulation results are provided to validate the proposed control algorithm.

Key Words: Gesture recognition, Image processing, Template matching, Robot control.

1 INTRODUCTION

As rapid development of Computer Science and sensor techniques, human-computer interaction (HCI) has received great attention in recent years. HCI has provided a strong tool for the design and use of computer technology, focused on the interfaces between people (users) and computers and has become an active field of research [1]. Some trend has been changed with the introduction of techniques based on recognition of vision, sound, speech, projective displays etc. Researchers also provide a much richer and natural mode of interaction with Man-computer methods [2]. In the current day the human-computer interaction application of hand gesture is being developed vigorously. The advantage of these application is that users can control devices without touching anything such as panel, keyboard, mouse, etc. The users just control devices with facing the camera and raising the hands [3]. Among the various types of gesture, hand gestures are easy to be used and more convenient for communication. Hand gesture are basically of two types-static and dynamic hand gesture. In this paper, we most analyze and research static hand gestures recognition. static hand gestures do not involve any kind of hand movement in comparison to dynamic hand gesture, where either the entire hand moves or only the fingers move [4].

The various techniques develop in human-computer interaction (HCI) can be extended to other areas, such as surveillance, robot control, and teleconferencing [5]. Among these, sign language recognition has become an active topic of research as it provides an opportunity for the hearing impaired to communicate with the normal people without the need of an interpreter [6]. The detection and understanding of hand and body gestures is becoming an important and challenging task in computer vision.

Hand gesture recognition can be extensively applied in many fields, such as home appliance entertainment and medical system, etc. [13]. In recently years, Many application systems in many domains such as virtual environment, smart surveillance, human-computer intelligent interaction (HCII), teleconferencing, sign language translation, etc. [4]. Zeller et al. [8] presented a visual environment for very large scale biomolecular modeling application. This system permits interactive modeling of biopolymers by linking a 3D molecular graphics and molecular dynamics simulation program. Quek also presented a finger mouse application to recognized 2-D finger movement which are the input to the desktop in [9]. Berry, Wachs, et al. integrated controlling gesture into the virtual environment Battle Field in [10,11].

In this paper, a gesture based control problem is considered for mobile robots. Firstly, some static hand gestures are recognized by using the techniques, such as, mean filtering, morphological image processing, contour correspondence, with the help of HSV mode, which provides a new application of gestures recognition. Secondly, five gestures are encoded to five commands for a mobile robot, which is realized through a gesture based state-feedback control strategy. Finally some numerical simulations are presented to demonstrate the effectiveness of the proposed gesture control strategy.

2 HAND GESTURE RECOGNITION

The static posture recognition stage is composed of region of hand extraction, hand feature, extraction and static hand posture classification [14]. When we get the hand gesture image, the image is preprocessed by some way that image processing, such as, image color processing, image filtering processing, morphological image processing,

This work is supported partially by National Nature Science Foundation of China under Grants 61473061, 71503206, 61104104, and the Program for New Century Excellent Talents in University under Grant NCET-13-0091.

image contour processing, etc. To the object of gesture recognition, these process is indispensable. There are five gestures that has been defined. These gestures will be recognized by static gesture recognition system. The following is the necessary process for gesture image.

2.1 Color Models

In this section, we will realize the skin segmentation and analyze the result of the HSV model on the base of RGB color model. HSV color model include three component, Hue, Saturation, Value. When we gain the image, we will need to transform the image from the RGB model to HSV model. The method that RGB color model transform into HSV color model is showed in three equations.

$$H=\begin{cases}\dfrac{(G-B)*\pi/3}{Max(R,G,B)-Min(R,G,B)} & when\ R=Max(R,G,B)\\ \dfrac{(B-R)*\pi/3}{Max(R,G,B)-Min(R,G,B)} & When\ G=Max(R,G,B)\\ \dfrac{(B-G)*\pi/3}{Max(R,G,B)-Min(R,G,B)} & When\ B=Max(R,G,B)\end{cases} \tag{1}$$

$$S=\begin{cases}\dfrac{Max(R,G,B)-Min(R,G,B)}{Max(R,G.B)} & when\ Max(R,G,B)\neq 0\\ 0 & when\ Max(R,G,B)=0\end{cases} \tag{2}$$

$$V=Max(R,G,B) \tag{3}$$

where Max(R,G,B) is the maximum component of RGB color model.

2.2 Filtering

We use a mean filtering (MF) to preprocess the image, where the each pixel point will be replaced by the mean value of the domain. MF is a nonlinear smoothing technique. The algorithm of computing the middle value is given as follows:

$$a_{mid}=Mid\{a_1,a_2,a_3,\ldots,a_n\}$$

$$=\begin{cases}a_{(\frac{n+1}{2})} & n \text{ is odd number}\\ \dfrac{1}{2}[a_{(\frac{n}{2})}+a_{(\frac{n+1}{2})}] & n \text{ is even number}\end{cases} \tag{4}$$

Where: $a_1,a_2,a_3,\ldots,a_n$ is a sequence. The sequence is sorted as $a_1\le a_2\le a_3\le\cdots\le a_n$. The a_{mid} is a middle value.

The mean filtering use the slidable window that is consisted by n elements to traversing all elements. These elements that window include will be sorted from smaller element to bigger element. If n is odd number, the middle value will be the value of middle element. If n is even number, the middle value will be the value of mean value. For the 2D filter, the gray value of the pixel points are set to $\{a_{ij},(i,j)\in I^2\}$. A is the slidable window. The median filter is as follows:

$$x_{mid}=mid\{a_{ij}\}=mid\{a_{(i+u),(j+v)}\in A,(u,v)\in I^2\} \tag{5}$$

2.3 Morphological Image Processing

Morphology uses simple rules to process images and employs two- or three dimensional templates, such as lines, crosses, composites, squares, circles, diamonds, etc. Morphological image processing can reduce image noise while keep the base of the original structure. The elementary operators of morphological image processing are defined as follows:

$$Erosion: A\Theta B=\{x\mid(\hat{B}_x\subseteq A\} \tag{6}$$

$$Dilation: A\oplus B=\{x\mid(\hat{B}_x\subseteq A\} \tag{7}$$

$$Opening: A\circ B=(A\Theta B)\oplus B \tag{8}$$

$$Closing: A\bullet B=(A\oplus B)\Theta B \tag{9}$$

where A is the image, B is the pixel representing of an object in the image. Then $\hat{B}$ is the set of points in B whose coordinates (x,y) have been replaced by $(-x,-y)$, i.e., $\hat{B}=\{w\mid w=-b,b\in B\}$. The translation of a set B by a point $z=z(z_1,z_2)$, denoted by $(B)_z$, it is defined as $(B)_Z=\{c\mid c=b+z,b\in B\}$ for binary image, The cvMorphologyEx structure is used to operate opening and closing on the base of erosion and dilation. The purpose is to get rid of the non-target areas and filter noises.

2.4 Contour Correspondence

Contour correspondence is a method that compares the similarity of a contour of an object with template image. In order to process the binary image, we use a cvfindContours structure [19] to get the contour image. Fig. 1 is a example that two gesture images translate to contour image through image preprocessing. We can judge the result of matching by setting the threshold value. Hu derived a set of invariant moments [16], which are widely used in image recognition. We use Hu's invariant moment matching method to get the contour image.

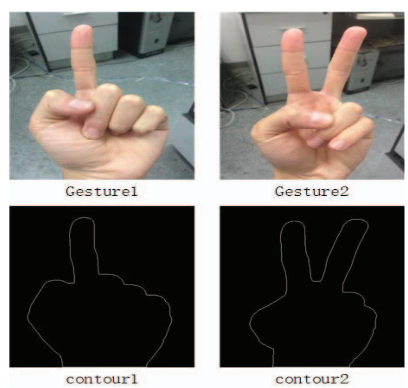

Fig 1.The example of the contour.

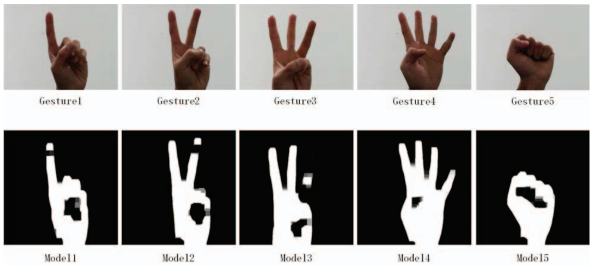

Fig 2. The example of gesture template..

In Fig. 3, we show the procedure of the gesture processing, there are some functions of the gesture processing, we use the function of cvtcolor() to transform from the RGB to the HSV, then we reduce the complexity of color to increase efficiency. GaussianBlur() and MedianBlur() are two filtering functions. In the morphological image processing, we use the function of morpologyEX() to adjust the contour of gesture and we successively set two parameters of MORPH_CLOSE and MORPH_OPEN. Finally, two functions produce the contour of the processed image. In this paper, we use Meanshift method for skin color segmentation by the function of pyrMeanShiftFiltering().

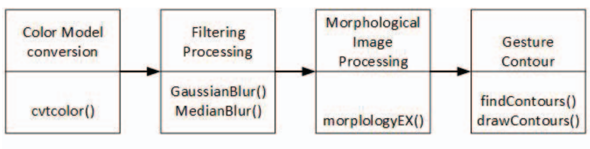

Fig 3.The procedure of the image processing

Table 1 shows the diverse recognition rate with four different algorithms. It was shown that the gesture G5 has the highest recognition rate which is 80% when we use GaussianBular(). The recognition rates of the other gestures are a little lower due to the light and the number of the gesture templates. In order to improve the recognition quality, we use MedianBlur(), then all the recognition rates have a obvious increase. The gesture G5 still has a biggest increase comparing with the rate using GaussianBular(). The morplologyEX()(CLOSE) means that parameter is MORPH_CLOSE while the morplologyEX()(OPEN) means that parameter is MORPH_OPEN. We use these two algorithms to increase the recognition rates. The recognition rates of G1, G4 and G5 also have distinct increases.

Table1. The recognition rates with different algorithms

Gesture	GaussianBlur()	MedianBlur()	MorplologyEX()(CLOSE)	MorplologyEX()(OPEN)
G_1	78%	80%	80%	85%
G_2	75%	78%	78%	78%
G_3	73%	80%	80%	80%
G_4	75%	80%	80%	85%
G_5	80%	83%	85%	88%

3 MOBILE-ROBOT GESTURE CONTROL

The recognition system provides the user with a camera to control a mobile robot through hand gestures. In this section, we use the invariant moment matching method to compute the minimal distance between the edges of the hand gestures and the edges of the gesture templates. In the case, five gestures are used to control the robot. The gesture1 means the command of going straight, the gesture2 means the command of turning right, the gesture3 means the command of turning left, the gesture4 means of going back, and the $gesture5$ means the command of stopping.

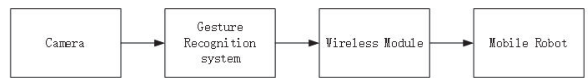

Fig 4. The realization of gesture based control

A wireless module APC220 is applied in the robot. As is shown in Algorithm 1, firstly, the system gets the initial gesture image G_t from user interface window. Secondly, G_t is translated to $\dot{G}_t$ through a series of image processing steps. Then H_k is the minimum of invariant moment by matching with gesture templates (T_K). H_u is the trained threshold value of the templates. By comparing the sizes between the H_u and $\dot{G}_t$, the system guarantees that $\dot{G}_t$ is gesture k varying from 1 to 5.

4 EXPERIMENTS

In this section, some simulation results are provided to demonstrate the gesture control of mobile robots.

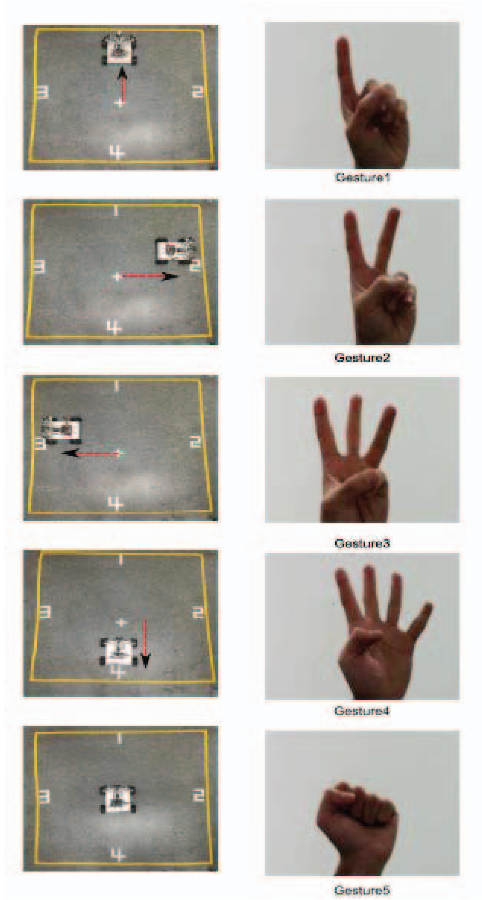

Fig 5. Gesture based control

In Fig. 5, the numbers represent the corresponding gestures and the arrows mean the moving direction of robot. For example, when the system gets the gesture1, the robot will move to the place labelled 1 from the place of cross. it means go straightly. The gesture 2 means turning right. When the system gets the gesture 2, the robot will move to the place labelled 2. The gesture 5 orders the robot to stop. Fig. 5 shows that the robot can move with the desired directions under the proposed gesture control

Algorithm 1 (Gesture based control)

Require: G_t //initial gesture image

Step1: Hand Image Preprocessing

$G_t \Rightarrow \dot{G}_t$ // $\dot{G}_t$ is contour image

Step2: Contour Correspondence

For k =1 to 5do

$$H_k = Min(H_u(\dot{G}_t, T_k))$$

If $H_k \leq H_u$ then

$\dot{G}_t$ = gesture k

End if

End for

Ensure :

Switch($\dot{G}_t$)

Case gesture 1:

$$U = K(X_i - X_1) \to 0$$

Case gesture 2:

$$U = K(X_i - X_2) \to 0$$

Case gesture 3:

$$U = K(X_i - X_3) \to 0$$

Case gesture 4:

$$U = K(X_i - X_4) \to 0$$

Case gesture 5:

$$U = K(X_i - X_5) \to 0$$

End switch

5 CONCLUSION

In this paper, a gesture based control problem was considered for mobile robots. Firstly, gesture templates were built for different gestures with complex features and a gesture recognition algorithm was proposed to deal with the static posture recognition problem by using image processing and patter recognition methods. Then, a gesture based control algorithm was developed for a mobile robot by encoding the recognized gesture information. In the future work, we will continue to investigate the dynamic gesture based control for mobile robots, where the dynamic gesture recognition will be considered, and the robot will take a series of motions according to dynamic gestures

REFERENCES

[1] A. Just, Two-handed gestures for human-computer interaction, Japanese Studies, vol. 35, No. 1, pp. 124-125, 2007.

[2] D. Z. Cheng, Matrix and Polynomial Approach to Dynamic Control Systems, Science Press, Beijing, China, 2002.

[3] C. Hardenberg, F. Berard, Bare-hand human-computer interaction, In Proceedings of the ACM Workshop on Perceptive User Interfaces, Orlando, FL, USA, pp. 1-8, 2002.

[4] C. Hsieh, D. Liou, D. Lee, A real time hand gesture recognition system using motion history image, In: The 2nd International Conference on Signal Processing Systems(ICSPS), pp.394-439, 2010.

[5] G. Plouffe, A.M. Cretu, P. Payeur, Natural human-computer interaction using static and dynamic hand gestures, in: 2015 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE), 2015.

[6] Y. Wu, T. S. Huang, Vision-based gesture recognition: a review, in Gesture-Based Communication in Human-Computer Interaction, Lecture Notes in Computer Science, vol. 1739, pp. 103-115, 2001.

[7] A. Choudhury, A. K. Talukdar, and K. K. Sarma, A review on vision-based hand gesture recognition and applications, Intelligent Applications for Heterogeneous System Modeling and Design, pp. 261-286, 2015.

[8] M. Zeller, J. C. Phillips, A. Dalke, et al., A visual computing environment for very large scale biomolecular modeling, IEEE 24th International Conference on Application-Specific Systems, Architectures and Processors, pp. 3-12, 1997.

[9] F. K. H. Quek, M. Zhao, Inductive learning in hand pose recognition, International Conference on Automatic Face

and Gesture Recognition, IEEE Computer Society, pp. 78-83, 1997.

[10] G. A. Berry, Small-wall: A Multimodal Human Computer Intelligent Interaction Test Bed with Applications, Dept. of ECE, University of Illinois at Urbana-Champaign, MS thesis ,1998.

[11] J. P Wachs, M. Kolsch, H. Stern, Vision-based hand-gesture applications, Commun. ACM, 54(2), pp. 60-71, 2011.

[12] C. H. Wu, W. L. Chen, C. H. Lin, Depth-based hand gesture recognition, Multimedia Tools and Applications, 75(12), pp 7065–7086, 2016.

[13] J. H. Li , Q Pan, PL Cui, Image recognition based on invariant moment in the projection space, International Conference on Machine Learning and Cybernetics, pp. 3606-3610, 2004.

[14] M. K. Hu, Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, 8(2), pp. 179-187, 1962.

[15] B. G. Batchelor, F. M. Waltz, Morphological Image Processing, Machine Vision Handbook, pp. 801-870, 2012.

[16] R. Lukac, B. Smolka, K. Martin, et al. Vector filtering for color imaging. IEEE Signal Processing Magazine, 22(1), pp. 74-86, 2005.

[17] R. Laganière, OpenCV 2 computer vision application programming cookbook, Packt Publishing, Limited, 2011.

[18] X. Lv, Y. Xu, M. Wang, Real-Time Hand gesture rRecognition for robot hand interface, Communications in Computer and Information Science, 461, pp. 209-214, 2014.