

Hand Gesture Recognition Based On The Parallel Edge Finger Feature And Angular Projection

Zhou Yimin^{1,2}, Jiang Guolai¹, Xu Guoqing¹ and Lin Yaorong³

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

Email:{ym.zhou,gl.jiang,gq.xu}@siat.ac.cn

²The Chinese University of Hong Kong, Hong Kong, China

³School of Electronic and Information Engineering, South China University of Technology. Email:eyrlin@scut.edu.cn

Abstract. In this paper, a novel high-level hand feature extraction method is proposed by the aid of finger parallel edge feature and angular projection. The finger is modelled as a cylindrical object and the finger images can be extracted from the convolution with a specific operator as salient hand edge images. Hand center, hand orientation and wrist location are then determined via the analysis of finger image and hand silhouette. The angular projection of the finger images with origin point on the wrist is calculated, and five fingers are located by analyzing the angular projection. The proposed algorithm can detect extensional fingers as well as flexional fingers. It is robust to the hand rotation, side movements of fingers and disturbance from the arm. Experimental results demonstrate that the proposed method can directly estimate simple hand poses in real-time.

1 INTRODUCTION

Computer vision based hand gesture recognition has become a hot topic recently, due to its natural human-computer interaction characteristics. Hand gestures are usually composed of different hand postures and their motions. Estimating hand posture in real-time is the necessary step for hand gesture recognition. However, human hand is an articulated object with over 20 degrees of freedom (DOF) [8], and many self-occlusions can occur in its projection results. Moreover, hand motion is often too fast comparing with current computer processing speed. Therefore, real-time hand posture estimation is still a challenging work.

In recent years, many 3D sensors, such as binocular cameras, Kinect and leap motion, have been applied for hand gesture recognition with good performance. However, hand gesture recognition based on monocular camera has quite a limitation, since 3D sensors are not always available in many systems, i.e., Google Glasses.

Feature extraction is a crucial module in hand posture estimation system. Different kinds of features with low-level precision are available for detecting

and recognizing hand postures. Skin color segmentation is widely used for hand localization [10][15]. Silhouette, the outline of segmented hand area, can be used for matching predefined hand postures [16]. Edge is another common feature for model based matching [17]. Histogram of oriented gradients has been implemented in [14]. Combinations of multiple features can be taken to improve robustness of the feature extraction[5].

However, systems using low-level features can only recognize several predefined postures. High-level features (fingers, fingertips, joint locations etc) from hand image should be extracted for various hand postures estimation, since they can provide very compact representation of the input hand postures in real-time.

In [12][13], fingertips were found by fingertip masks considered the characteristics of fingertips, and the location of fingertips can be achieved based on feature matching. However, the objects which share the similar shape of fingertips, will result in a misjudgment. In [1], fingertips were located by generalized Hough Transform and probabilistic models. The detected edge segments of the black and white images is computed by Hough Transform for fingertip detection. But the light and brightness would seriously damage the quality of the images and detection.

In [3], palm and fingers were detected by skin-colored blob and ridge features. In [6][9], high-level hand features were extracted by analyzing hand contour. Hand postures can be recognized through the geometric features and external shape of the Palm and fingers [2]. It proposes a prediction model for showing the hand postures. Though the measurement error would be large due to the complexity of the hand gestures and varieties of hands.

Since flexional fingers will not appear on the hand contour, little attention has been paid on extracting high-level features of these fingers directly. In [7], a finger detection method using grayscale morphology and blob analysis is introduced, which can be used for flexional finger detection.

In this paper, a novel method for extracting and locating all the fingers based on a digital camera is proposed, where the hand postures are viewed from the orthogonal direction to the palm. Firstly, the salient hand edge image is extracted, which includes hand contour and boundaries of flexional fingers. Considering that each finger has two main "parallel" edges on its two sides, a finger image is extracted from convolution result of salient hand edge image with a specific operator G . Then hand orientation and wrist location are computed from the finger image and hand image. Finally, fingers and their angles are extracted from the local maxima of angular projection.

The remainder of the paper is organized as follows. Section II introduces a novel finger extraction and modelling approach. The high-level hand feature extraction is described in Section III. Simulation experiments and results are shown in Section IV. Conclusions and future works are given in Section V.

2 The procedure of finger extraction

2.1 Description of the finger model

The most notable parts of a hand are the fingers. Fingers are cylindrical objects with nearly constant diameter from root to fingertip. The proposed finger model is shown in Fig.1(A). Here, the boundary of the i^{th} finger, C_{fi} , is considered as the composition of arc edges (C_{ti} , fingertip or joint) and a pair of parallel edges ($C_{ei} + C'_{ei}$, finger body):

$$C_{fi} = (C_{ei} + C'_{ei}) + \sum C_{ti} \quad (1)$$

The finger center line (FCL) C_{FCLi} is introduced, which can be regarded as the center line of the finger parallel edges. Fingertip/joint O_{ti} locates at the end point of FCL, and it is the center of the arc curve C_{ti} as well. The distance between the parallelled finger edges are known as $2 \times d$, where d is the radius of all fingers.

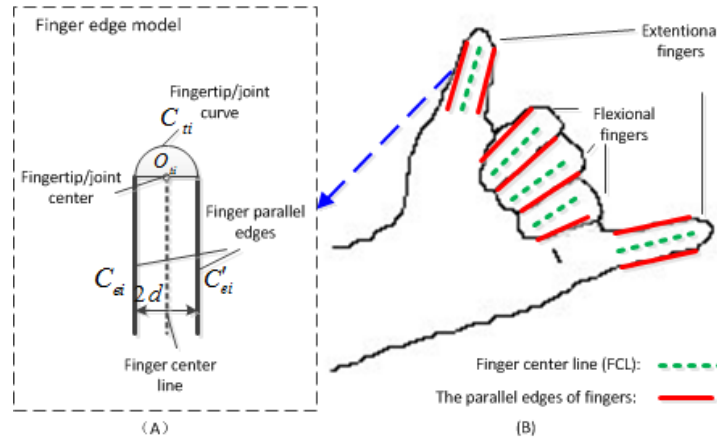


Fig. 1. Finger edge model (A) and the parallel edges (B)

As it is shown in Fig.1(B), all fingers are represented by the areas near FCLs, which can be produced from the parallelled finger edges.

2.2 Salient hand edge image extraction

The salient hand edge image $I_{edge}(x, y)$ is mainly made up of hand contour which includes boundaries of extensional fingers and arm/palm and flexional fingers. As shown in Fig.2: (a) is the grayscale hand image $I_{gray}(x, y)$; (b) is the threshold image $I_{black}(x, y)$; (c) is the canny edge image $I_{canny}(x, y)$; (d) is the contour image $I_{contour}(x, y)$; (e) is the obtained salient hand edge image $I_{edge}(x, y)$. Then the extraction of $I_{edge}(x, y)$ is summarized as follows:

1. Extract grayscale hand image $I_{gray}(x, y)$ and hand contour image $I_{contour}(x, y)$ from source color image by using skin color segmentation method in [10].
2. Extract canny edge image $I_{canny}(x, y)$ from $I_{gray}(x, y)$ [4].
3. In most hand postures, the boundaries of flexional fingers are dark edges. Applying the threshold Th_{black} (predefined) to gray-level hand image for extraction, then the obtained $I_{black}(x, y)$ is:

$$I_{black}(x, y) = \begin{cases} 1 & (I_{gray}(x, y) < Th_{black}) \\ 0 & (I_{gray}(x, y) \geq Th_{black}) \end{cases} \quad (2)$$

The boundaries of the flexional fingers are extracted by the interlapped area of $I_{canny}(x, y) \cap I_{black}(x, y)$.

4. Then the salient hand edge image $I_{edge}(x, y)$ is extracted by:

$$I_{edge}(x, y) = I_{black}(x, y) \cap I_{canny}(x, y) \cup I_{contour}(x, y) \quad (3)$$

The whole procedure is demonstrated in Fig.2.

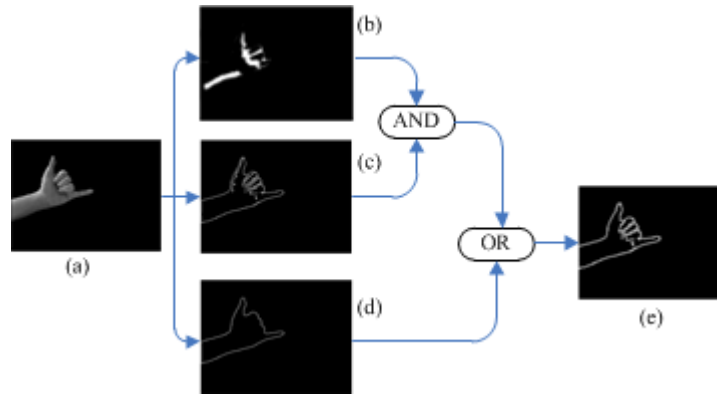


Fig. 2. Diagram of extracting hand edge procedure

2.3 Finger extraction using parallel edge feature

In order to extract fingers from their parallel edges, a rotation invariant operator G is introduced, which is given by:

$$G(x, y) = \begin{cases} -1 & 0 \leq D(x, y) \leq r_1 \\ 1 & r_2 < D(x, y) \leq r_3 \\ 0 & \text{else} \end{cases} \quad (4)$$

$$D(x, y) = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (5)$$

where (x_0, y_0) denotes the center point of G ; $D(x, y)$ is the distance from the center point; r_1, r_2 and $r_3 (r_1 < r_2 < r_3)$ are the defined parameters shown in Fig.3.

For each edge point in $I_{edge}(x, y)$, if the information of edge orientation is not considered, it can be regarded as a rough estimation of its correlative points on FCL. Here we simply assume $r_1 = r_2 - r_1 = r_3 - r_2 = N$, and $N = 0.5d$. As shown in Fig.3, the size of continuous G operator is defined as $(6N \times 6N)$, while the size of discrete G is $((6N + 1)(6N + 1))$, without loss of generality.

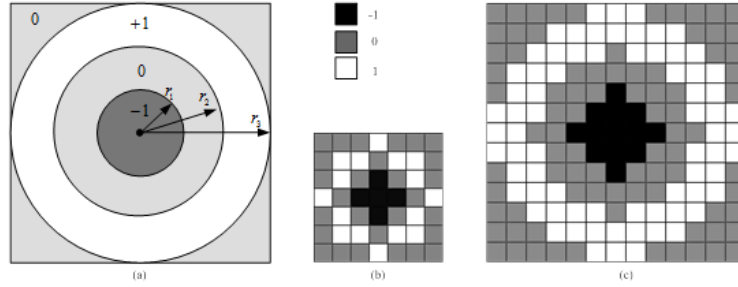


Fig. 3. (a) Continuous Operator G ; (b) Discrete operator G ($N=1$); (c) Discrete operator G ($N=2$);

The finger image $I_f(x, y)$ can be extracted by the following steps, shown in Fig.4. Fig.4(a) is the salient hand edge image $I_{edge}(x, y)$ (see Fig.2(e)) obtained from the steps in Section 2.2; Fig.4(b) is the convolution result $I_w(x, y)$ and Fig.4(c) is the finger image $I_f(x, y)$. The steps are described in details:

1. Decide the size of G with $N = 0.5d$. When the dealt hand image is too large, it should be adapted to proper size firstly;
2. Calculate the convolution $I_w(x, y) = I_{edge}(x, y) \times G$;
3. Extract rough finger image $I_{f'}(x, y)$ from $I_w(x, y)$ by,

$$I_{f'}(x, y) = \begin{cases} 1 & I_w(x, y) > T_h \\ 0 & I_w(x, y) \leq T_h \end{cases}$$

where T_h is the proper threshold in $(\sqrt{5}d, 2\sqrt{5}d)$. Due to space limit, the details of T_h definition is not given here.

4. Apply morphological dilation operation to $I_{f'}(x, y)$ to get the finger image $I_f(x, y)$.

It should be noted that in the extracted finger images $I_f(x, y)$, noise pixels are mainly generated from edges of palm print near palm center, while faulty detection is often caused by incomplete extraction of flexional finger boundaries.

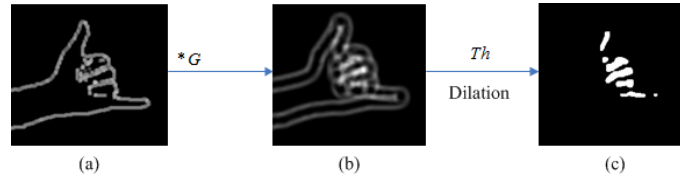


Fig. 4. Diagram of extracting fingers

3 High-level hand feature extraction

3.1 The definition of the hand center

The whole hand model is depicted in Fig.5. Here, palm center P_{hc} is regarded as the hand center since it is insensitive to various hand postures. P_{hc} is also assumed as a stable point which has the maximum distance to the closest boundary edge during estimation. Morphological erosion operation has been taken to remove the extensional fingers in the hand area [12]. Thus, the center of the hand region is defined as the mass center of the remaining region.

However, the width of a arm is quite close to wrist width, which makes it difficult to be removed through erosion operation. Large remaining arm area may cause error for the hand center determination. As a result, arm should be separated initially during this method application. In our approach, based on the finger image extracted in section II, finger center P_{fc} could be regarded as the mass center of the finger area. A circle centered on P_{fc} is drawn to remove the arm, which is the large dashed circle in Fig. 5. The diameter of the circle is set to $5R_p$, where R_p denotes the radius of the palm.

3.2 Hand orientation determination

Hand orientation is another important basic feature for hand posture recognition. It can be estimated by calculating the direction of principal axes of the silhouettes, or determined by the arm [12]. However, these methods are not always effective because the silhouettes of some postures ("fist") don't have obvious principal axes, while arm may not always appear.

In this paper, since the fingers are always located on one side of the hand palm, the vector from the hand center P_{hc} to the middle finger center P_{fc} will give a more reliable estimation of the hand orientation. The hand orientation \vec{N}_p is determined by:

$$\vec{N}_p = \frac{\overrightarrow{P_{hc}P_{fc}}}{|\overrightarrow{P_{hc}P_{fc}}|} \quad (6)$$

where $\vec{\cdot}$ is the vector denotation, $|\cdot|$ is the norm operation. The wrist position P_w could be determined through the palm radius R_p calculation,

$$P_w = P_{hc} - R_p \vec{N}_p \quad (7)$$

The diagram of finding the hand center P_{fc} , hand orientation \vec{N}_p and wrist position P_w is shown in Fig. 5.

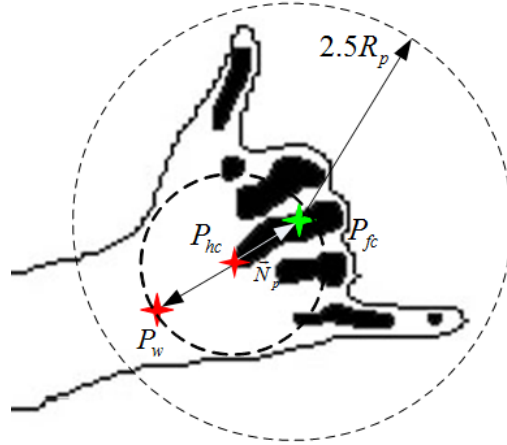


Fig. 5. The diagram of the parameter determination for hand model

3.3 Angular projection for finger localization

(A) Finger angle determination

According to the kinematic hand model [8], if there is no finger crossing, all the fingers can be regarded as radials starting from the wrist. The angular projection method described in [11] for face recognition is a feasible choice for finger angle determination via angle distribution characteristics of fingers pixels.

Here, the finger image $I_f(x, y)$ (shown in Fig.4) is transformed into polar image $I_f(r, \theta)$ (see Fig.5), with the origin point on the wrist center point P_w . Different from normal angular projection methods, considering the noise pixels in $I_f(x, y)$ located near palm center, the points with larger radius vector r should have higher probability of being a point on fingers. The adjusted angular projection in each direction is calculated as follows:

$$p_n = \int_{r_1}^{r_2} \int_{\theta_n}^{\theta_{n+1}} r I_f(r, \theta) d\theta dr, \quad n = 0, 1, 2, \dots, M-1 \quad (8)$$

where $[r_1, r_2]$ is the range of radius vector; $[\theta_n, \theta_{n+1}]$ is the n^{th} subdivision of polar angle; M is the number of subdivisions of polar angle. Then the normalized P_n is given by:

$$p_{n_norm} = \frac{p_n}{\sum_{n=0}^{M-1} p_n}, \quad n = 0, 1, 2, \dots, M-1 \quad (9)$$

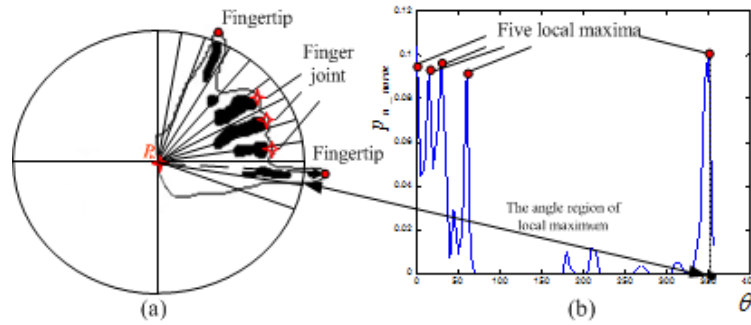


Fig. 6. (a): Angular projection of finger image; (b): The value of P_n in different angles

(B) Finger location

The polar angle subdivisions of the five fingers can be determined by the largest five local maxima of $\{p_{n_norm}\}$. In the angle subdivision of each finger, the point on the hand contour with farthest distance D_i to P_w is determined as fingertip (for extensional fingers) or finger joint (for flexional fingers). The status of the fingers (extensional or flexional) could be determined by: $D_i > \alpha_i R_p$ and $D_i < \alpha_i R_p$, where R_p is the palm radius. α_i is the constant, determined by the finger length and varied with different fingers, which is depicted in Fig.6.

4 Experiments and result analysis

The proposed hand extraction algorithm is used in a real-time hand posture estimation system for matching and tracking. The experiments are running with a PC (CPU: Pentium Core 2 Duo E7200, RAM: 2 GB) and a monocular web camera. The system could process real-time video of resolution 320*240 pixels and frame rate 15fps. The tested hand stays at a position at roughly 0.5m distance in front of the camera, keeping the palm in the direction facing the camera with freely moved hand. A video is attached for the parts of the experiments. Fig.7 shows several examples of hand estimation generated from hand features in the video. The left column is the input frames, where the red marks and blue marks indicate the estimated hand center and hand wrist center. The output hand postures are shown in the right column, where the white segments denote flexional fingers and green segments denote extensional fingers.

Table 1 lists the numerical results of correct rate for the detected fingers. It can be seen that the recognition for the extensional fingers are quite accurate. As for the flexional finger, it still has 75.2% recognition rate, which is not bad as well. Fig.8 shows the hand orientation estimation result for a gesture sequence in the experiments. The blue solid line denotes the real hand orientation, determined by the marks of the palm center and the wrist. The red dashed line (close to blue line) denotes the estimated hand orientation with the proposed method. The

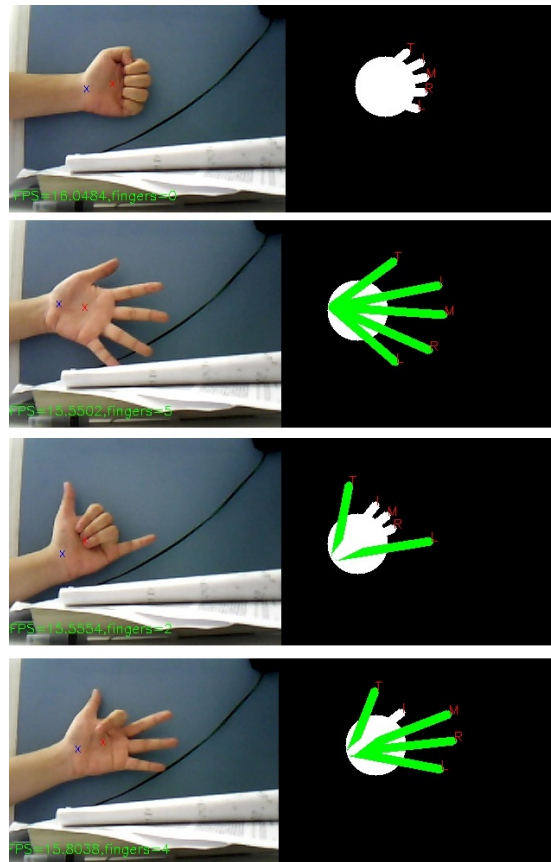


Fig. 7. Real-time 2D hand posture estimation directly using features extracted in this paper

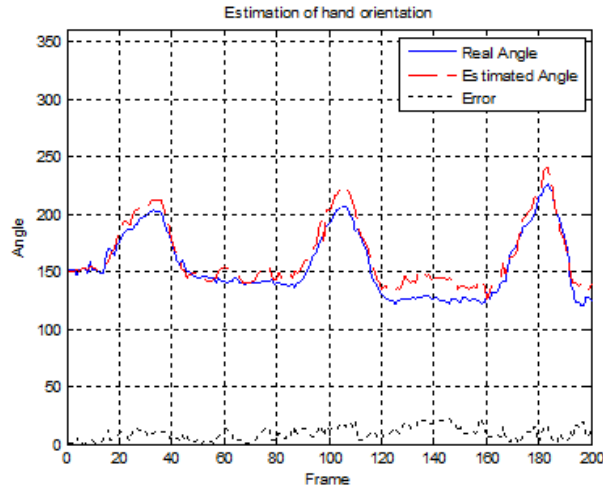
Table 1. Correct rate for detecting fingers

Total frames	Extensional fingers	Flexional fingers
5300	96.2%	75.2%

black small dashed line is the estimation error, which is calculated by:

$$Error = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (10)$$

where y_i is the actual hand orientation or final angle; \hat{y}_i is the estimation values.

**Fig. 8.** Estimation of hand orientation

In Fig.9, it shows the middle finger angle estimation result for another gesture sequence in the experiments. The blue solid line denotes the real middle finger angle determined by the marks on the middle finger and the wrist. The red dashed line denotes the estimated middle finger angle via the proposed method, and the black dashed line is the estimation error.

Experiment results demonstrate that the proposed method can give a robust estimation for the features of the hand location, hand orientation and extensional fingers. In this case, it takes the whole fingers into account and not only the finger tips. As for the flexional fingers, the error is mainly from the bad hand edge estimation or flexional thumb finger, since they do not fulfill the assumption

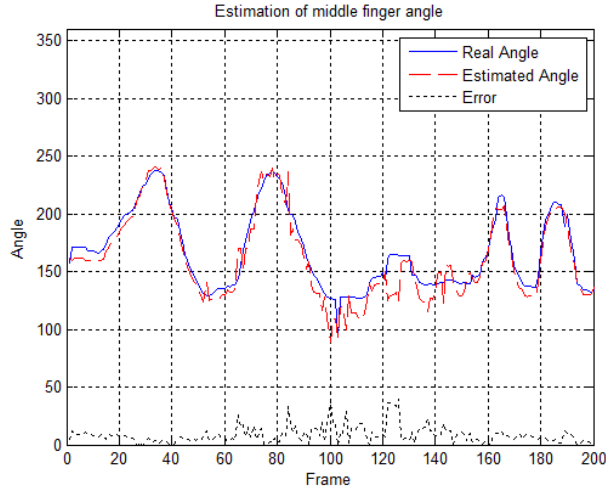


Fig. 9. Estimation of middle finger angle

that made in the angular projection. Although there are errors in the hand orientation and finger angle estimation, the results for real-time hand posture recognition has satisfied performance.

5 Summary and future work

Feature extraction is a crucial module in computer vision based hand posture recognition system. The implementation of this module has a considerable effect on the robustness and processing speed of the system. High level hand features are desirable because they can provide very compact representation of the input hand in real-time operational mode.

In this paper, a high level hand feature extraction method has been proposed for recognizing hands with whose palm facing to the camera. Hand edges have been extracted for detecting fingers via the parallel edge feature. Angular projection centered on wrist point is then used for getting the angle and length of each finger. A simple 2D hand model can be produced directly from these features. The proposed method can make robust estimation for features of hand location, hand orientation and extensional fingers. Hence, the estimated hand model can be used for controlling mechanical hands or interacting with Google Glasses.

However, if the salient hand edge is not well detected, false or miss detection for flexional fingers could occur. Future work will focus on improving salient hand edge detection algorithm with tracking modules.

ACKNOWLEDGMENT

This work is partially supported under the Shenzhen Science and Technology Innovation Commission Project Grant Ref. JCYJ20120615125931560 and partially supported by Introduced Innovative R&D Team of Guangdong Province (201001D0104648280).

References

1. J. Barrho, et al.: Finger Localization and Classification in Images based on Generalized Hough Transform and Probabilistic Models. The 9th International Conference on Control, Automation, Robotics and Vision, 2006, pp.1–6.
2. M. K. Bhuyan, et al: Hand pose recognition using geometric features. National Conference on Communications (NCC), 2011, pp. 1–5.
3. L. Bretzner, et al.: Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. 5th IEEE International Conference on Proceedings, 2002.
4. J. Canny: A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, **8** (1986) 679–698.
5. C. Cao, et al.: Real-time multi-hand posture recognition. International Conference on Computer Design and Applications (ICCD), 2010.
6. C.-C. Chang, et al.: Feature alignment approach for hand posture recognition based on curvature scale space. Neurocomputing, **71** 2008 1947–1953.
7. N. Dung Duc, et al.: Finger extraction from scene with grayscale morphology and BLOB analysis. IEEE International Conference on Robotics and Biomimetics, 2008, pp. 324–329.
8. A. Erol, et al.: Vision-based hand pose estimation: A review. Computer Vision and Image Understanding. **108** 2007 52–73.
9. Z. Feng, et al.: Features extraction from hand images based on new detection operators. Pattern Recognition. **44** 2011 1089–1105.
10. M. J. Jones and J. M. Rehg: Statistical color models with application to skin detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **1** 1999 274–280.
11. H. I. Kim, et al.: Rotation-invariant face detection using angular projections. Electronics Letters. **40**, 2004 726–727.
12. K. Oka, et al.: Real-time fingertip tracking and gesture recognition. IEEE Computer Graphics and Applications. **22** 2002 64–71.
13. J. R. Parker and M. Baumbach: Finger recognition for hand pose determination. IEEE International Conference on Systems, Man and Cybernetics, 2009, pp. 2492–2497.
14. J. Romero, et al.: Monocular real-time 3D articulated hand pose estimation. 9th IEEE-RAS International Conference on Humanoids, 2009.
15. L. Sha, et al.: Hand posture recognition in video using multiple cues. IEEE International Conference on Multimedia and Expo, 2009.
16. N. Shimada, et al.: Real-Time 3-D Hand Posture Estimation Based on 2-D Appearance Retrieval Using Monocular Camera. The proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01), 2001.
17. W. Ying, et al.: Capturing natural hand articulation. 8th IEEE International Conference on Computer Vision, **2**, 2001, 426–432.