Palm Gesture Recognition Technology Based on Comparison of Contour Autocorrelation Functions

Anatoly Novikov [1], Evgeny Kholopov [1], Aleksey Efimov [1], Mikhail Nikiforov [1]

[1] Ryazan State Radio Engineering University named after V.F. Utkin, Gagarina st, 59/1, Ryazan, 390005, Russia

Abstract

The technology of detecting static hand gestures is considered. The technology includes four main stages of image processing: detection of palm contours; piecewise linear approximation of the palm contour; construction of the autocorrelation function (ACF) of the contour of the observed gesture; comparison of the calculated ACF with the ACF of reference gestures. Examples of identification of the main gestures with the palm using the proposed technology, as well as the resulting quantitative estimates of the similarity measure of gestures, are given. The proposed technology is characterized by high speed and lack of learning process. The presented approaches may be of interest in the construction of non-contact control systems.

Keywords

Static gesture, edge detector, piecewise linear approximation, autocorrelation function.

1. Introduction

In the modern world, there is a rapid development of personal computers. Along with this, the ways of communicating with them are also being improved. Human-machine interfaces have gone from primitive computer mice made of wood to touchscreen displays and Kinect devices. Despite this, the development of new highly efficient and convenient ways to control personal computers is an urgent task. It attracts a large number of researchers. One of the options for such interfaces can be contactless computer control with the help of a human hand.

Gestures are considered to be one of the most powerful communication channels. They are simple and understandable for people of different nationalities and cultures. This is the easiest and most natural way to communicate. Consequently, the problem of organizing computer control using hand gestures is urgent. It boils down to solving such problems as fixing a gesture by processing data from a video signal source, segmenting the necessary information on the resulting image, and determining a hand pose.

Earlier gesture recognition technologies used special gloves. They collected data on hand movements and transmitted them to a PC for further processing using special algorithms [1]. Takahashi and Kishino have developed one of the most effective gloves, capable of recognizing about 46 different gestures [2]. However, this kind of solution was not further developed, since the use of a special glove turned out to be extremely inconvenient.

At the moment, research is being carried out in directions that exclude the use of intermediate elements in gesture recognition. This task is extremely difficult and multifaceted due to both objective and subjective differences, which are associated with a large number of degrees of freedom of the hand and fingers, differences in articulation, and a different color of the skin. Moreover, gesture recognition methods must be invariant with respect to size, speed, scene illumination, background heterogeneity, and other parameters.

In [3-6], researchers suggest using special depth sensors developed by Intel to determine the hand posture. They are based on infrared sensors, which allow, in addition to the direct image of the hand, to

GraphiCon 2022: 32nd International Conference on Computer Graphics and Vision, September 19-22, 2022, Ryazan State Radio Engineering University named after V.F. Utkin, Ryazan, Russia
EMAIL: novikovanatoly@yandex.ru (A. Novikov); holopovevgu13@rambler.ru (Y. Kholopov); lexie62rus@mail.ru (A. Efimov); nikiforov.m.b@mail.ru (M. Nikiforov)
ORCID: 0000-0002-8166-8234 (A. Novikov); 0000-0002-0541-2878 (Y. Kholopov); 0000-0002-4014-8718 (A. Efimov); 0000-0002-4796-0776 (M. Nikiforov)

receive information about its position in space. This makes it possible to suppose how far the hand is from the depth sensor.

The depth camera returns 3D data the skeleton of the hand, allowing the dynamic gesture to be viewed as a time series of hand positions. After the necessary data is received from the video signal source, it remains to process it with a suitable algorithm, extract the object of interest from the background, determine its location in the scene and decide on a particular hand position.

Another direction in the recognition of the gestures of the human hand is the use of convolutional neural network [7-12]. Convolutional neural networks are based on convolution layers. Each next layer processes the previous one using filters. The weights of the convolution kernel are not known in advance and change during training depending on the input data. The transfer of information to the next layer is carried out by the activation function by converting information (numerical values) from all neurons of the previous layer into a certain value for the neuron of the current layer. The exit value shows how the neuron of the current layer activated is activated. If some signs were discovered on the previous convolution operation, then in further processing such a detailed image is no longer necessary. It decreases in dimension (condenses). Filtration already unnecessary parts reduces the retraining. This layer is called pulling. Thanks to it, the neural network becomes more resistant to changes in the input image, for example, to shifts. A complete layer (percepton) - hidden. It is connected to all the neurons of the previous layer. The last layer of a multilayer perceptron is one or more neurons. Their number is equal to the number of classes.

The approach to gesture recognition based on the use of neural networks has clear advantages. These include accuracy, speed of operation and partial resistance to zooming. But there are also disadvantages to this approach. As the object moves away from the camera, the recognition accuracy decreases. Also, there are false positives on person's face. This is due to the complexity of feature extraction due to the insufficient depth of the neural network.

2. Proposed recognition technology

The proposed technology is based on the transition from a raster image of the palm to a contour version, followed by approximation of the palm contour by a closed polygon with a minimum number of vertices. At the next stage, the transition from the coordinate description of the contour to the vector one is made, the autocorrelation function (ACF) of the studied contour is calculated, which is then compared with the ACF of the reference contours. Thus, the complex processing of a raster image of the palm includes the following steps:

- highlighting the contour of the palm;
- piecewise linear approximation of the contour;
- calculation of the ACF of the approximated contour;
- comparison of the calculated ACF of the input image with the ACF of the reference image.

One of the most important stages in the technology under consideration is contour selection. The contour should not have gaps and have a minimum number of short non-informative lines. Most of all, these requirements are met by the well-known method for detecting gradient-type boundaries, the Kenny method [13]. But it also has certain disadvantages. The Kenny method has a high sensitivity, which, when processing real noisy images, leads to the appearance of false closed lines. This complicates the subsequent processing of the contour image. Therefore, within the framework of the study, the gradient-type boundary detector described in [14, 15] was used.

This algorithm provides stable estimates of partial derivatives, does not require preliminary smoothing of the image, and, unlike the Kenny method, is free from artifacts in the gradient image. Stable estimates of partial derivatives and, as a consequence, the modulus and direction of the gradient in each pixel of the image in this method are obtained by using a vector mask $(-k, -k+1, \ldots, -1, 0, 1, \ldots k-1, k)$.

It provides obtaining smoothed estimates of partial derivatives. In this case, the errors in the estimation of derivatives are the smaller, the larger the length $2k+1$ of the sliding window. Minimization of the number of detected short contour lines is achieved due to the original algorithm for finding thresholds, which are used in the process of forming contour lines [16]. The result of the operation of the described border detector is shown in Figure 1.

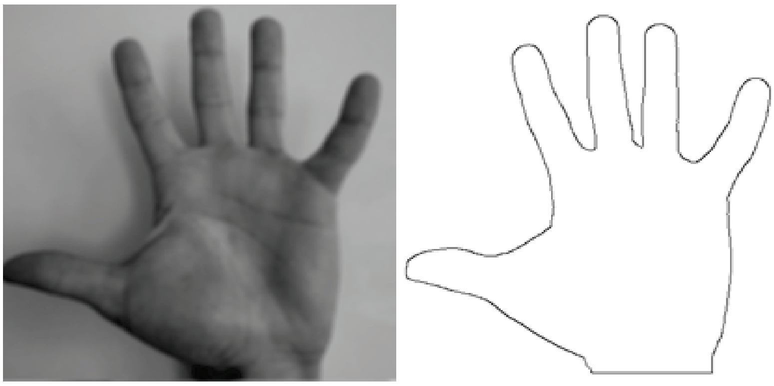
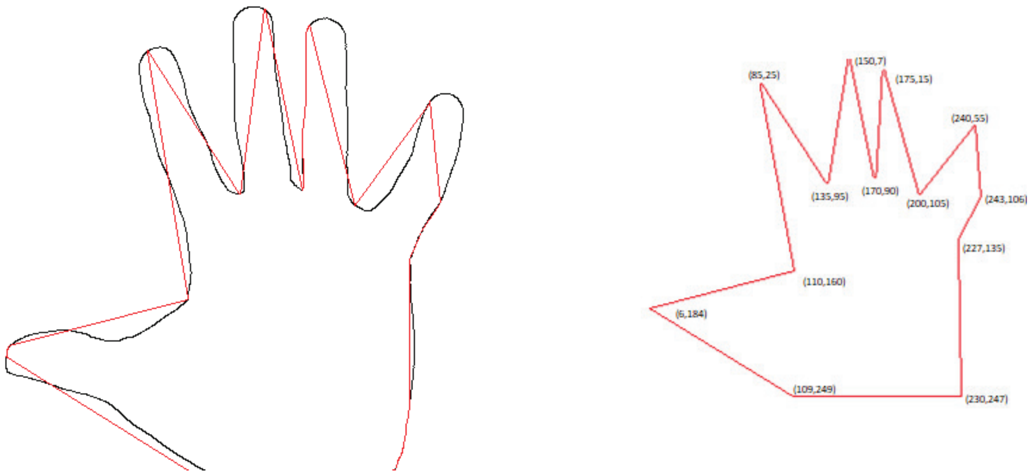


Figure 1: The result of selecting the contour of a static hand gesture

At the second stage of the proposed technology, a piecewise linear approximation of the palm contour is performed. In this problem, it is important that the contour is closed and contains the minimum number of vertices of the approximating polygon. In this case, all elements of the image (fingers) should be preserved. The method of piecewise linear approximation is that the two extreme points of the contour are connected by a segment. Further, the distances from each pixel of the original contour to the constructed line are found. If among the found distances there are those exceeding the specified accuracy threshold, then the most distant point is taken as a new point by the approximating polyline. The process continues until a polyline is constructed that satisfies the condition of proximity to the original contour. The result of applying a piecewise linear approximation to the contour of the palm obtained at the previous stage is shown in Figure 2.



2a: result of rough approximation, ε=25 2b: closed contour with vertex coordinates

Figure 2: Application of piecewise linear approximation

By the beginning of the third stage, we have an array of pairs of numbers $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, whose elements are the coordinates $(x_i, y_i), i = \overline{1, n}$ of the approximating polygon (Figure 2b). This is the so-called coordinate description of the approximating polygon. It is assigned a vector description $\Gamma^{(0)} = \{a_1, a_2, \dots, a_n\}$ of the contour (approximating polygon). Here $a_k = (x_{k+1} - x_k, y_{k+1} - y_k), k = \overline{1, n-1}$, a $a_n = (x_1 - x_n, y_1 - y_n)$.

The autocorrelation function of the contour $\Gamma^{(0)}$ is calculated by the formula

$$\tau_k = \frac{\left|\Gamma^{(0)}, \Gamma^{(k)}\right|}{\left(\Gamma^{(0)}, \Gamma^{(0)}\right)}, k = \overline{0, n-1} \quad (1)$$

In this formula, the contours $\Gamma^{(1)}, \Gamma^{(2)}, \dots, \Gamma^{(n-1)}$ are obtained from the contour $\Gamma^{(0)}$ by cyclic permutation of the vectors: $\Gamma^{(1)} = (a_2, a_3, \dots, a_n, a_1)$, $\Gamma^{(2)} = (a_3, a_4, \dots, a_1, a_2)$, ..., $\Gamma^{(n-1)} = (a_n, a_1, a_2, \dots, a_{n-1})$.

The scalar product of contours $(\Gamma^{(0)}, \Gamma^{(k)})$ in (1) is found by the formula:

$$(\Gamma^{(0)}, \Gamma^{(k)}) = (a_1, \bar{a}_{k+1}) + (a_2, \bar{a}_{k+2}) + \dots + (a_{n-k}, \bar{a}_n) \quad (2)$$

In (2) scalar products $(a_m, \bar{a}_{k+m})$ are products of two complex numbers, the second of which is taken as complex conjugate. The autocorrelation function of the contour $\tau_k, k = \overline{0, n-1}$ is unique for each contour and therefore it can be called a kind of "portrait" of the contour. For contours close in shape, their ACFs either coincide or differ insignificantly. The main property of the contour ACF is its invariance to scale changes, rotations and shifts of the contour.

At the last - the fourth stage - the ACF of the input circuit is compared with the set of ACF of the reference contours. To compare the ACF, a metric is introduced

$$d(\gamma_0, \gamma_m) = [2\backslash n] \cdot \sum_{k=1}^{[n\backslash 2]} \left| \tau_k^{(m)} - \tau_k^{(0)} \right|, m = \overline{1, M} \quad (3)$$

In formula (3), the following notations are introduced: γ_0 is the input circuit, $\gamma_m, m = \overline{1, M}$ are reference circuits, $\tau_k^{(0)}$ and $\tau_k^{(m)}$ $m = \overline{1, M}$ – ACF, respectively, of the input and reference circuits. Graphs of the ACF are symmetrical with respect to the point $[n\backslash 2]$, therefore in formula (3) the summation is performed not up to n, but up to $[n\backslash 2]$. Here $[\cdot]$ is the sign of the integer part of the number.

The gesture is identified by the formula

$$m^* = \arg\min d(\gamma_0, \gamma_m) \quad (4)$$

3. Experimental studies

Let us consider in more detail the algorithm for calculating the ACF of the palm contour shown in Figure 2b.

We pass from the coordinate description of the contours $\Gamma_0, \Gamma_1, \dots, \Gamma_n$ (Figure 2b) to the vector description

$\Gamma_0 = \{ (104,-24), (-25,-135), (50,70), (15,-88), (20,83), (5,-75), (25,90), (40,-50), (3,51), (-16,29), (3,112), (-121,2), (-103,-65)\}$;

$\Gamma_1 = \{(-25,-135), (50,70), (15,-88), (20,83), (5,-75), (25,90), (40,-50), (3,51), (-16,29), (3,112), (-121,2), (-103,-65), (104,-24)\}$; ...

$\Gamma_{11} = \{(104,-24), (-25,-135), (50,70), (15,-88), (20,83), (5,-75), (25,90), (40,-50), (3,51), (-16,29), (3,112), (-121,2), (-103,-65)\}$;

We calculate the scalar products of contours $\bar{\omega} = \left(\Gamma^{(0)}, \Gamma^{(k)}\right), k = \overline{1,11}$

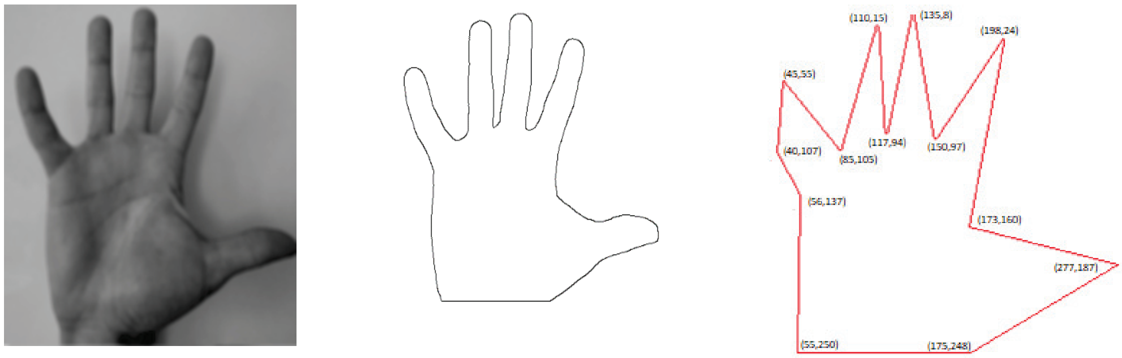
$\bar{\omega}_1 = -18961 - 50822i$; $\bar{\omega}_2 = 23752 - 31936i$; ...$\bar{\omega}_{11} = -18961 + 50822i$.

Using formula (1), we calculate the ACF values (normalized values of scalar products):

$\tau_k^{(0)} = \{0{,}3152; 0{,}5475; 0{,}3514; 0{,}1788; 0{,}2873; 0{,}1463; 0{,}1463; 0{,}2873; 0{,}1788; 0{,}3514; 0{,}5475; 0{,}3152\}$

To confirm the assumption about the invariance of the ACF to affine transformations, we will carry out the following experiment. Let's form a mirror reflection of the image of the hand, shown in Figure 2a. It is shown in Figure 3a. Let's repeat for him all the steps of the technological chain described above. The resulting ACF $\tau_k^{(1)}$ of the mirrored contour γ_1 differs only in the third decimal place from the ACF $\tau_k^{(0)}$ of the contour $\gamma_0 = \Gamma_0$.

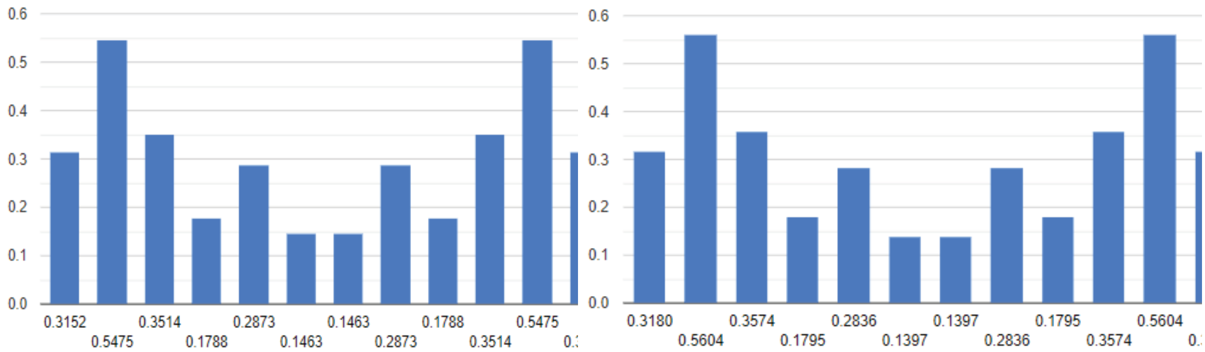
$\tau_k^{(1)} = \{0{,}3180; 0{,}5604; 0{,}3574; 0{,}1795; 0{,}2836; 0{,}1397; 0{,}1397; 0{,}2836; 0{,}1795; 0{,}3574; 0{,}5604; 0{,}3180\}$



3a: original image 3b: highlighted contour 3c: approximating polygon with vertex coordinates

Figure 3: Processing steps for a mirrored hand gesture

A graphical representation of the ACF of the original "Five" gesture (Figure 1) and its mirror image (Figure 3a) are shown in Figures 4a and 4b, respectively.



4a: reference gesture 4b: specular reflection

Figure 4: ACF graphs

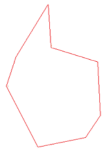
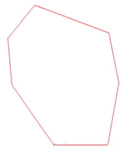
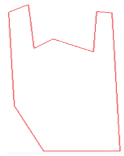
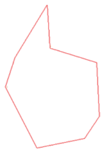
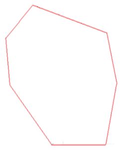
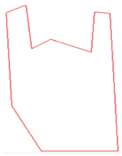
Visual comparison of the ACF of the compared images of the hand, shown in Figure 4, indicates their undoubted closeness. This conclusion is also confirmed by the value of the criterion (4) for the proximity of the contours: $d(\gamma_0, \gamma_1) = 0{,}0054$. This means that the contours are very close and the actually compared objects are identical up to a mirror image. This experiment confirms the conclusion that the ACF of the contour is its original "portrait". Taking into account the individual properties of the contour in its autocorrelation function makes it possible to compare different objects with each other.

The next experiment consisted in comparing 4 reference hand gestures. For each of them, all stages of the proposed technology were completed and the ACF was calculated.

Next, the ACF of each of the 4 reference contours was compared with the ACF of other contours using metric (3). The results of this comparative analysis across the 4 main gestures are shown in Table 1.

The main conclusion of this study is as follows: ACF contours of two different gestures do not match. Metric values $d(\gamma_0, \gamma_m)$ for them are significantly higher than the established threshold $\delta = 0{,}01$.

Table 1
Values of the criterion $d(\gamma_0, \gamma_m)$ of contour proximity able title

				
	0	1.07	2.16	1.69
	1.07	0	0.96	2.82
	2.16	0.96	0	2.34
	1.69	2.82	2.34	0

4. Conclusion

A technology for recognizing static hand gestures is presented, which is based on the idea of comparing the autocorrelation function of the contour of the studied gesture with the ACF of reference contours. Brief descriptions of the algorithms for extracting the contours of the palm and piecewise linear approximation of the selected contour are given. The algorithm for calculating the ACF of the contour is described in detail. An example of calculating the ACF of a contour is given. A metric is proposed for estimating the degree of proximity of the computed ACF contours.

The results of experimental studies of the proposed technology on pairwise comparison of the ACF of reference circuits are presented. The conducted studies confirm the possibility of using complex contour analysis to recognize static palm gestures.

5. References

[1] Bulygin D.A., Mamonova T.E. Real-time hand gesture recognition // Science Bulletin of the NSTU 2020, v.78, no. 1, p. 25–40.

[2] Smedt Q. De, Wannous H., Vandeborre J.-P. Heterogeneous hand gesture recognition using 3D dynamic skeletal data // Computer Vision and Image Understanding, 2019, vol. 181, pp. 60–72.

[3] Smedt, Q. De, Wannous, H., Vandeborre, J.P. Skeleton-based dynamic hand gesture recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2016, pp. 1–9.

[4] Wang, C., Liu, Z., Chan, S.C. Superpixel-based hand gesture recognition with kinect depth camera. IEEE Transactions on Multimedia 17, 2015, pp. 29–39.

[5] Ren, Z., Yuan, J., Meng, J., Zhang, Z., Robust part-based hand gesture recognition using kinect sensor. IEEE Transactions on Multimedia 15, 2013, pp. 1110–1120.

[6] Wu D., Zhu F., Shao L. One shot learning gesture recognition from rgbd images. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2012, pp. 7-12.
[7] D. Zh. Satybaldina, G. V. Ovechkin, K. A. Kalymova, A system for recognizing static hand gestures using a depth camera // Vestnik RGRTU. 2020. No. 72 Vestnik of RSREU. 2020. No 72, p. 93-105.
[8] Nguyen T.T. Algorithmic and software for figure recognition using Fourier descriptors and neural networks. Bulletin of the Tomsk Polytechnic University. Tomsk: From TPU, 2010. T. 317, No.5. WITH. 122-125.
[9] Dong J., Xia Z., Yan W., Zhao Q. Dynamic gesture recognition by directional pulse coupled neural networks for human-robot interaction in real time // Journal of Visual Communication and Image Representation, 2019, vol. 63, p. 102583.
[10] E. R. Muratov, M. B. Nikiforov, and A. M. Skachkov. Artificial Intelligence in Dynamic Gesture Recognition. doi: 10.25559/ SITITO.16.202004.883-892 // Modern Information Technologies and IT Education. - 2020. T. 16, No. 4. S. 883-892
[11] Liao B., Li J., Ju Z., Ouyang G. Hand Gesture Recognition with Generalized Hough Transform and DC-CNN Using Realsense // Eighth International Conference on Information Science and Technology (ICIST). Cordoba, Granada, and Seville, Spain; 2018. p. 84-90.
[12] Liu P., Li X., Cui H., Li S., Yuan Y. Hand Gesture Recognition Based on Single-Shot Multibox Detector Deep Learning // Mobile Information Systems. 2019; 2019:3410348. (In Eng.) doi: https://doi.org/10.1155/2019/3410348
[13] John Canny. A Computational Approach to Edge Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. PAMI-8 NO 6, 1986. pp. 679–698.
[14] Novikov A.I., Pronkin A.V. Detector of boundaries of the gradient type for images of the underlying surface // Vestnik RGRTU. 2019, No. 68. - S. 68-76. doi: 10.21667/1995-4565-2019-68-2-68-76.
[15] Novikov A.I., Pronkin A.V. Gradient method and program for detecting the boundaries of the brightness difference // Proceedings of ITNT-2020, May 26-29, in 4 volumes - Samara, Samara University, 2020. T. 2: Imaging and remote sensing of the Earth. S. 111-119.
[16] Loginov A.A., Novikov A.I., Sablina V.A., Shcherbakova O.V. Study of the possibility of using complex contour analysis in the problems of classification and combination of contours // Bulletin of the Russian State Technical University, No. 1 (issue 43). Ryazan, 2013. p. 20-24.