Advances in Computer Science Research, volume 88
International Conference on Computer, Network, Communication and Information Systems (CNCI 2019)

Gesture Recognition Based on Improved HOG-LBP Features

Gang Nie[1, a], Junxi Zhao[2, b]

[1]Nanjing University of Posts and Telecommunications, Nan Jing 210003，China;

[2]Nanjing University of Posts and Telecommunications, Nan Jing 210003，China;

[a]15996228023@163.com, [b]junxi_zhao@163.com

Abstract. With the development of computer technology, vision-based human gesture recognition has become an important hotspot technology in the field of human-computer interaction. However, the performance of gesture recognition is often affected by conditions such as lighting changes, background complexity, and skin color differences. In this paper, an enhanced fusion HOG feature and LBP feature algorithm are proposed for feature extraction. HOG features describe image local features and LBP features describe image texture features. Improved MB-LBP can capture large-scale structures more than LBP features. It is more able to describe more changes, that is, contain more local information. Then use the Support Vector Machine (SVM) for classification detection. The method is tested in the Jochen Triesch static hand pose data set. The results show that the recognition accuracy reaches 98.64%, which is more robust and effective than single feature extraction and traditional HOG-LBP.

Keywords: Gesture recognition, HOG-LBP, MB-LBP, SVM.

1. Introduction

With the rapid development of computer technology, human gesture recognition has become a hot topic in recent years. At present, gesture recognition plays an important role in sign language recognition, human-computer interaction, safe driving, and intelligent control.

The gesture recognition method based on computer vision usually includes three aspects: data processing, gesture analysis and recognition classification. Data processing and gesture analysis are collectively referred to as gesture detection. The detection mainly includes color, shape, texture, structure, etc. Mainly refers to the color histogram, the shape feature is mainly the gradient histogram [1]. The main problem of gesture recognition technology is that the performance of gesture recognition is easily affected by conditions such as illumination intensity, background complexity, and skin color difference [2]. Due to these challenges, it is difficult to accurately describe hand features by using single feature detection. Therefore, this paper proposes an enhanced fusion HOG-LBP feature detection and SVM classification detection method, replacing uniform LBP features with improved MB-LBP and taking RBF (radial basis function) as the core function of SVM. The experimental results show that the method is more accurate in light condition, skin color and complex background.

2. Feature Extraction

2.1 HOG.

2.1.1 Concept.

Histogram of Oriented Gradient (HOG) is a kind of gradient structure and edge information used to represent local regions of an image. It is mainly used to represent the shape features of an image. The core idea of HOG is that the detected local object shape can be described by the distribution of illumination gradient or edge direction. By dividing the whole image into small connected regions (called cells), and extracting features from each cell, generate A directional gradient histogram, a combination of these histograms used to represent the target [3].

In order to improve the accuracy, the local histogram can be normalized to the contrast in a larger range of images: the histogram is first calculated in its corresponding image range, and then all cells in the range (called a block) are used. Normalized.

2.1.2 HOG Feature Extraction Steps.

(1) First, grayscale processing the image to obtain a grayscale image;

(2) Using the gamma correction method to normalize (normalize) the color space of the input image, the purpose is to adjust the contrast of the image, reduce the influence of the change of the illumination and the shadow of the local image, and suppress the noise. The correction formula is: where gamma is generally 1/2.

(3) Calculate the gradient (including size and direction) of each pixel. The gradient calculates the gradient amplitude and phase angle. The target direction information exists in the phase angle, and the change information of the pixel gray value exists in the amplitude. Grading the gradient of the pixels in the image can not only further weaken the interference of light intensity and shadow, but also reflect the contour and texture features of the image. The gradient of the pixel in the image is calculated as follows:

$$G_x(x,y) = H(x+1,y) - H(x-1,y) \quad (1)$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1) \quad (2)$$

Where $G_x(x,y)$, $G_y(x,y)$、$H(x,y)$ represent the horizontal and vertical gradients and pixel values of the image at (x,y) pixel points, respectively.

From this, we can get the gradient amplitude and gradient direction at the (x,y) pixel point:

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (3)$$

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \quad (4)$$

(4) Then divide the image of 128*128 into small cells of 8*8. Experiments show that when the size of the cell is 8*8, not only the higher recognition rate but also the dimension of HOG can be reduced. The gradient histogram of each cell is counted. Each pixel in the cell votes on the edge intensity based on the corresponding gradient direction. Experiments show that using the bin direction histogram can achieve the best gesture detection performance.

(5) Each 2*2 cells are combined into one block, and the division of the rectangular block has two forms of overlap and non-overlap. If the histogram of the cell area is combined in the overlap form, the same cell histogram will appear in different blocks, and at least one block feature of all the obtained blocks is complete, but the calculation amount is relatively large. The use of form non-overlap merges with a small amount of computation. The feature descriptors of all cells in a block are concatenated to obtain the description of the HOG feature of the block [4].

(6) The HOG feature of the image can be obtained by concatenating the HOG feature descriptions of all blocks in the image.

2.2 LBP and Improved LBP

2.2.1 Concept.

Local Binary Pattern (LBP) is an operator used to describe the local texture features of an image. The original LBP is defined in a 3*3 window, with the window center pixel as the threshold, and the gray value of the adjacent 8 pixels is compared with it. If the surrounding pixel value is greater than the central pixel value, the position of the pixel is marked. It is 1, otherwise it is 0. Therefore, 8 points clockwise reading can generate an 8-bit binary number. The value converted to decimal is the LBP value of the center pixel. The calculation mode is shown in Fig. 1:

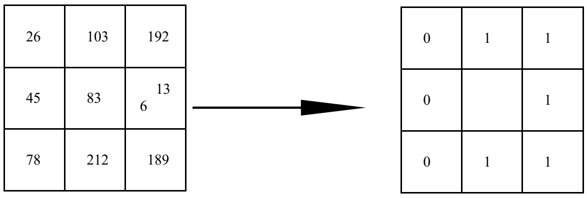

Fig. 1 LBP calculation mode

From the upper nine squares, an 8-bit binary number is formed. According to the clockwise value, it is 01111100, and then it is changed to decimal. That is, the LBP value of the pixel at this point is 104.

The LBP value of the window is calculated by the formula:

$$L(x_c, y_c)=\sum_{k=0}^{7} Z(I_k, I_c)\times 2^k \quad (5)$$

$$Z(I_k, I_c)=\begin{cases} 1, & I_k - I_c \geq 0 \\ 0, & I_k - I_c < 0 \end{cases} \quad k=0,1,\ldots,7 \quad (6)$$

Where I_c is the center of the window and the gray value is (x_c, y_c), and I_k is the surrounding pixel point around the center of the window.Since an LBP operator with p sample points will produce 2p patterns, this will result in a sharp increase in the number of binary patterns of the sample points, which is very unfavorable for texture expression. In order to solve the problem of excessive binary, Ojala proposed a "uniform mode". In this paper, the uniform mode is used to reduce the binary mode [5-7]. The uniform mode is defined as: when a loop binary number corresponding to an LBP has a maximum of two transitions from 0 to 1 or from 1 to 0, the binary corresponding to the LBP is called an equivalent mode class. By such improvement, the number of modes is reduced from the original 2p to p(p-1)+2+1, which also makes the feature vector have fewer dimensions and can reduce the influence of the high frequency noise band. As shown in Fig. 2 (b) uniform LBP map.

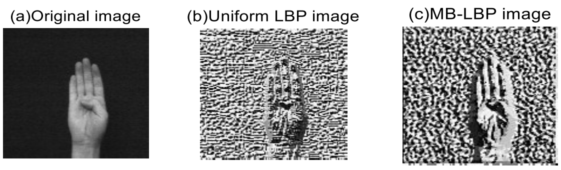

Fig. 2 Comparison of uniform LBP and MB-LBP

2.2.2 MB-LBP.

In this paper, in view of the shortcomings of traditional LBP features, a new unique feature is proposed, which is called multi-block local binary mode (MB-LBP) feature. The MB-LBP feature uses the average gray value in the sub-block to replace the pixel gray value in the LBP feature. Each sub-block contains several domain pixels (or only one pixel), and 9 sub-blocks constitute the entire MB-LBP feature. Compared to LBP, MB-LBP features can capture large-scale structures and describe more changes, that is, contain more local information. The operator with block size $S\times S$ is usually denoted by the symbol $MB_S - LBP_{P,R}^{U2}$ [8]. When S is 1, that is, there is no blocking, which is the same as the basic LBP operator; when S is 2, the average value of the adjacent 2×2, that is, the gray level of 4 pixels is taken as the value of the entire block, and later calculated as a Pixel processing. The MB-LBP diagram shown in (c) of FIG. 2, and FIG. 3 is a test example of the MB-LBP.

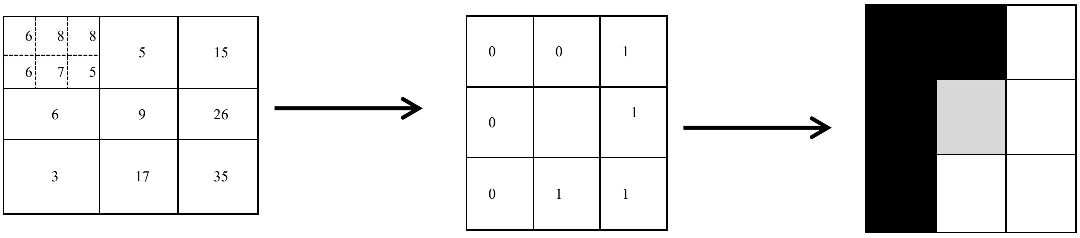

Fig. 3 MB-LBP calculation mode

2.3 SVM

Support Vector Machine (SVM) is a two-class classification model that distinguishes positive and negative samples by finding the largest interval classification plane (hyperplane). The strategy is to maximize the interval. For the linear indivisible problem, the low-dimensional space is mapped to the high-dimensional space by the kernel function, so that it is linearly separable [9]. In this paper, we use the radial basis function (RBF) as the kernel function to classify the image features to be extracted by SVM. The steps to train the SVM are as follows:

1) Extract the HOG and MB-LBP features separately and fuse them.

2) Establish training samples. The array space is allocated according to the extracted feature vector and the relevant label is pasted, the positive sample category is 1 and the negative sample category is 0.

3) Set the SVM parameters. Select the SVM type as C_SVC and the kernel function type as RBF. The maximum number of iterations in the algorithm termination condition is set to 1000 and the tolerance is FLT_EPSILON.

4) Using 5 cross-validations, the best model parameters were obtained, C=1 and =0.0478, which were used for the training set.

3. The Experimental Results and Analysis

In order to prove the superiority of the method, we use Jochen Triesch Static Hand Posture Database as the detection standard of the proposed algorithm. The gesture library contains 10 gestures of A, B, C, D, G, H, I, L, V, Y. Each of the gestures contains 24 samples collected by the researchers in three different situations: bright, dark, and complex. A total of 720 gesture motion samples are included. Among them, 510 samples are used as training samples for SVM, and the rest are tested sample. Samples of action gestures in three different backgrounds in the Jochen Triesch Static Hand Posture Database are as follows:

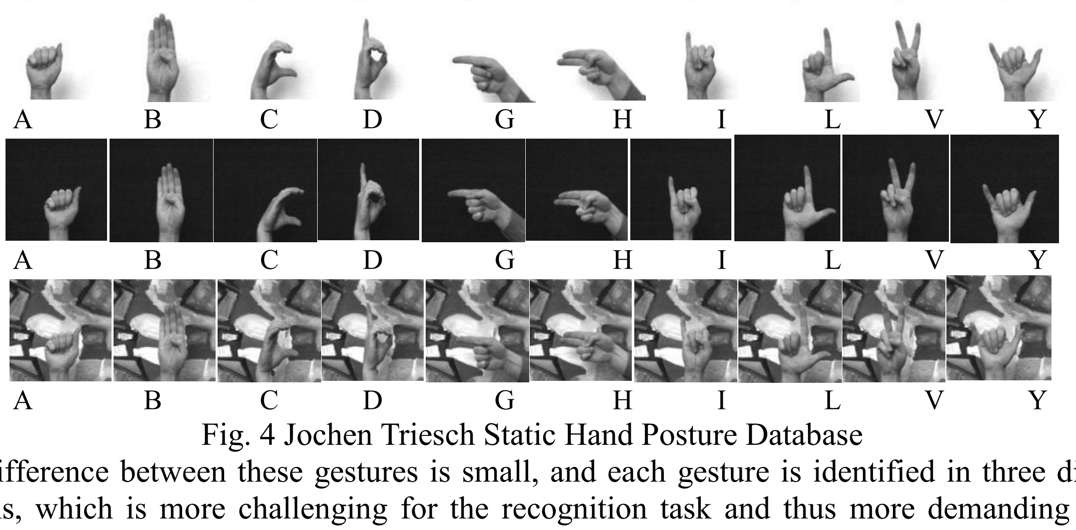

Fig. 4 Jochen Triesch Static Hand Posture Database

The difference between these gestures is small, and each gesture is identified in three different conditions, which is more challenging for the recognition task and thus more demanding on the method. The fusion feature method HOG-MBLBP is compared with other single feature extraction methods and HOG-LBP method. The experimental results are shown in Table 1:

Table 1 Test results and comparison

Method	Training Sample	Test Sample	Recognition Rate
HOG+SVM	510	210	97.86%
LBP+SVM	510	210	96.79%
HOG+LBP+SVM	510	210	98.25%
HOG+MB-LBP+SVM	510	210	98.64%

It can be seen from the table data that the recognition rate of the improved fusion HOG-MBLBP method is higher and more effective than the single feature extraction and the traditional HOG-LBP, when the training samples and the test samples of the SVM classifier are identical. The recognition rate on the Jochen Triesch Static Hand Posture Database reached 98.64%.

4. Summary

This paper proposes a gesture recognition method based on improved HOG-LBP feature and SVM classifier. It is improved on the traditional LBP method. By combining the new feature MB-LBP with HOG, more image structures can be captured. Information and more vivid performance. At the same time, the method of this paper is compared with the existing methods under the same conditions, and a higher recognition rate is obtained.

References

[1]. Cheng H, Yang L, Liu Z. Survey on 3D hand gesture recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(9): 1659-1673.

[2]. C. Zou, Y. Liu, J. Wang, et al, “Deformable part model based hand detection against complex backgrounds,” Advances in Image and Graphics Technologies, IGTA 2016 Book Series: Communications in Computer and Information Science, vol 634, pp.149-159,2016.SV

[3]. J. Žemgulys, V. Raudonis, and R. Maskeliūnas “Recognition of basketball referee signals from videos using histogram of oriented gradients (HOG) and support vector machine (SVM),” Procedia Computer Science, vol. 130, pp.953-960, 2018.M

[4]. N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", CVPR 2005, vol. 1, pp. 886-893, 2005.

[5]. J. Sun, X. Wu, “Infrared target recognition based on improved joint local ternary pattern,” Optical Engineering, vol. 55, no. 5, 2016.

[6]. A. Asthana, S. Lucey, and R. Goecke, “Regression based automatic face annotation for deformable model building,” Pattern Recognition, vol. 44, no. 10, pp.2598-2613, 2011.

[7]. H. Lategahu, S. Gross, and T, Stehle, et al, “Texture classification by modeling joint distributions of local patterns with Gaussian mixtures,” IEEE Transactions. on Image Processing A Publication of the IEEE Signal Processing Society, vol. 19, no. 6, pp.1548-1557, 2010.

[8]. Yuelin, Research on face recognition based on MB-LBP features[D], Xidian University, 2013, [online] Available: 10.7666/d.D363897.

[9]. [9].Y. T. Yang, B. Fishbain, and D. S. Hochbaum, et al, “The supervised normalized cut method for detecting, classifying and identifying special nuclear materials,” Informs Journal on Computing, vol. 26, no. 1, pp.45-58, 2013.