AMSE JOURNALS-AMSE IIETA publication-2017-Series: Advances B; Vol. 60; N°1; pp 25-42

Submitted Jan. 2017; Revised March 15, 2017, Accepted April 15, 2017

Human Action Recognition Based on Multiple Feature Fusion

* R.J.Ma, ** H.S. Zhang

* Department of Electronic and Information, Northwestern Polytechnical University
China, 127 West Youyi RoadXi'an Shaanxi (*1561000382@qq.com*)

** Department of Electronic and Information, Northwestern Polytechnical University
China, 127 West Youyi RoadXi'an Shaanxi (*zhanghuisheng@nwpu.edu.cn*)

Abstract

Human actions recognition generally uses geometric or statistical characteristics as training data. The geometric characteristics of the image and waveform display can be described by Pulse coupled neural network (PCNN). Experiential model decomposition (EMD) algorithm can be used for the feature extraction of waveforms. Therefore, we propose a motion feature description algorithm combined with PCNN and EMD. The experimental results show that using the method of PCNN-EMD-based feature recognition can obtain a high accuracy rate, while using the fusion feature of KPCA has a higher recognition rate.

Key words

Human action recognition, EMD, Gabor, PCNN, KPCA

1. Introduction

In order to distinguish human actions, the action must be classified on the basis of the description of eigenvalue [1]. The main problem to be solved by vision-based human motion recognition is to process and analyze the original image or image sequence data collected by sensor by computer, to learn and understand human action and behavior. Generally, based on the motion detection and feature extraction, the motion pattern of the human body is obtained through analyzing, and the mapping relation between the video content and the action type description is established so that the computer can "see" the video or "understand" the video. The vision-based human motion recognition consists of the following three steps: firstly, detecting the motion information from the image frame and extracting the underlying features; then modeling the behavior patterns or actions; finally, establishing the corresponding relationship between the underlying visual characteristics and High-level semantic information.

Back in the last century 80's, Marr proposed computer vision theory [3], the entire visual perception process is divided into the bottom, middle, high three levels. It is desirable to have the computer fully and automatically recover the three-dimensional structure information from the two-dimensional image sequence in a bottom-up manner. In recent years, more and more universities, research institutes, and commercial organizations have been engaged in the research of human action Recognition .Some international authoritative journals (such as TPAMI, IJCV, TIP, PR, CVIU) and important academic conferences (such as CVPR, ICCV, ECCV) also take human motion analysis and recognition as one of the topics.

Given that the large amount of noise, speckle and tissue texture contained in the action image, which will affect the cutting of action contour and the extraction of eigenvalue. Therefore, prior to the above-mentioned operation, the appropriate image preprocessing must be carried out to eliminate unnecessary Noise and spots, and to make the action contour more obvious to facilitate the contour of the cutting operation. In this paper, pretreatment of action image is firstly carried out, followed by feature extraction, and finally data are classified.

2. Image preprocessing

2.1 Image smoothing

Median filtering is a commonly used non-linear smoothing filter which is used to remove noise. The basic idea of it is to replace the gray value of the pixel with the median value of the gray value of the image pixel. Two-dimensional median filter can be expressed as follows:

$$y_{ij} = Med\left\{f_{ij}\right\}$$

(1)

In the formula: y_{ij} is the filter window; f_{ij} is a sequence of two- dimensional data .

2.2 Edge detection

This equation is obtained by convolution of the original image $f(x, y)$ with two convolution kernels $g_1(x, y)$ and $g_2(x, y)$. The mathematical expression is [3]:

$$S(x,y)=MAX\left[\sum_{m=1}^{M}\sum_{n=1}^{N}f(m,n)g_1(i-m,j-n),\sum_{m=1}^{M}\sum_{n=1}^{N}f(m,n)g_2(i-m,j-n)\right]$$

(2)

In fact, Sobel edge operator algorithm is used to count the weighted average, and then to carry out the differential operation. We can replace the difference with first-order partial derivative, the operational formula is as follows:

$$\begin{cases}\Delta_x f(x,y)=\left[f(x-1,y+1)+2f(x,y+1)+f(x+1,y+1)\right]-\left[f(x-1,y-1)+2f(x,y-1)+f(x+1,y-1)\right]\\ \Delta_y f(x,y)=\left[f(x-1,y-1)+2f(x-1,y)+f(x-1,y+1)\right]-\left[f(x+1,y-1)+2f(x+1,y)+f(x+1,y+1)\right]\end{cases} \quad (3)$$

The vertical and horizontal templates of the Sobel operator are shown in Figure (1).The former can detect the horizontal edge in the image, while the latter can detect the vertical edge. In practical application, each pixel in the image is convolved with these two convolution kernels, whichever is greater. The result of the calculation is an image that reflects the magnitude of the edge.

-1	-2	-1
0	0	0
1	2	1

a

-1	0	1
-2	0	2
-3	0	1

b

Fig 1. Sobel operator template

2.3 Image segmentation

Image segmentation was obtained through fuzzy c-means clustering algorithm (FCM). The well-known fuzzy ISODATA is a clustering algorithm that determines the extent to which each data point belongs to a certain cluster by membership degree. In 1973, Bezdek proposed the algorithm as an an improvement of early hard c-means clustering (HCM) method [4].

FCM divides n vectors *xi* (*i* = 1, 2, ..., *n*) into c fuzzy sets, and finds the center of clustering in each group, which minimizes the value function of non-similarity index. The main difference between FCM and HCM is that FCM uses fuzzy partitioning so that each given data point determines its degree of belonging to each group by the degree of membership between 0 and 1. Corresponding to the introduction of fuzzy partitioning, membership matrix U allows for values between 0 and 1 elements. However, combining with the normalization rule, the sum of the memberships of a dataset is always equal to 1:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,...,n$$

(4)

Then, the value function (or objective function) of FCM is:

$$J(U, c_1,...,c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j}^{n} u_{ij}^m d_{ij}^2 \quad ,$$

(5)

Here u_{ij} is between 0 and 1; c_i is the cluster center of fuzzy group I, $d_{ij}= ||c_i\text{-}x_j||$ is the Euclidean distance between the first cluster center and the jth data point; And $m\in[1,\infty)$ is a weighting index.

The necessary conditions for minimizing (5) are obtained by constructing a new objective function which is as follows:

$$\begin{aligned}\overline{J}(U, c_1,...,c_c, \lambda_1,...,\lambda_n) &= J(U, c_1,...,c_c) + \sum\nolimits_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} u_{ij} - 1) \\ &= \sum_{i=1}^{c}\sum_{j}^{n} u_{ij}^m d_{ij}^2 + \sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} u_{ij} - 1)\end{aligned}$$

(6)

Here j, j=1 to n, is the (n) constrained Lagrangian multiplier of (4). The necessary condition for minimizing (6) all input parameters is:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}$$

(7)

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}$$

(8)

From the above two necessary conditions, fuzzy c-means clustering algorithm is a simple iterative process. FCM uses the following steps to determine the center of cluster ci and the membership matrix U [4] when running in batch mode:

Step 1: initialize the membership matrix U with a random number between 0 and 1 to satisfy the constraint in (6)

Step 2: Calculate c cluster centers c_i, i=1, ..., c using Eq. (7).

Step 3: Calculate the value function according to (5). If it is less than a certain threshold, or less than a threshold relative to the previous value, then the algorithm stops.

Step 4: Calculate the new U matrix using (8). Go back to step 2.

The above algorithm can also initialize the cluster center, and then perform the iterative process. Because it can not ensure that FCM converges to an optimal solution. The performance of the algorithm depends on the initial clustering center. Therefore, we either use another fast algorithm to determine the initial cluster center, or each time to start the algorithm with a different initial cluster center, and run FCM more than once.

Let the set of objects be classified as: $X=\{x_1, x_2,...,x_N\}$, in which each object x_k has n characteristic indexes. If we want to classify X into c, each of its classification results should correspond to a CxN of fuzzy c partition space, the following is Boolean matrix $U=[u_{ik}]_{cxN}$:

M_{fc} $=\{U \subset R^{cN}$ | u_{ik}:[0,1], $\forall$ i, $\forall$ k; $\sum_{k=1}^{N} u_{ik}$ = 1, $\forall$ k;0< $\sum_{k=1}^{N} u_{ik}$, $\forall$ i},In this space, the fuzzy c-means algorithm is as follows:

Repeat for l = 1 ,2.......

Step 1: Compute the cluseter prototypes (means):

$$p_i^{(l)} = \frac{\sum_{k=1}^{N}(u_{ik}^{(l-1)})^m x_k}{\sum_{k-1}^{n}(u_{ik}^{(l-1)})^m}, 1 \le i \le c$$

(9)

Step 2: Compete the distance:

$$(d_{ik})^2 = (x_k - p_i^{(l)})^T A(x_k - p_i), 1 \le i \le c, 1 \le k \le n$$

(10)

Step 3: Update the partition matrix:

For $1 \le k \le N$

If $(d_{ik})^2 > 0$ for all i=1,2,...,c

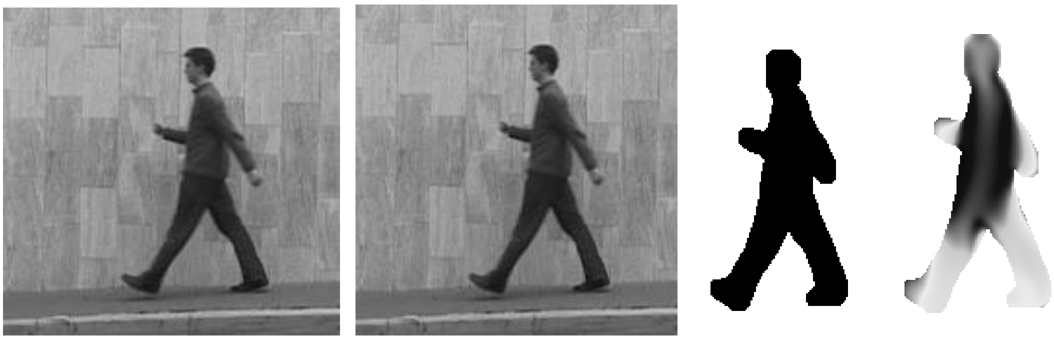
$$u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c}({}^{d_{ik\,A}}/_{d_{jk\,A}})^{2/(m-1)}}$$

(11)

Otherwise

$u_{ik}^{(l)}$=0 if d_{ik_A} >0, and $u_{ik}^{(l)}$:[0,1] with $\sum_{i=1}^{c} u_{ik}^{(l)} = 1$

Until $\| U^{(l)} - U^{(l-1)} \| < \varepsilon$



(a) (b) (c) (d)

Fig 2. (a) The original action image; (b) is smoothed; (c) through the FCM processing; (d) through the Saliency

3. Action image feature extraction by Texture

3.1Texture feature

In this paper, we will use the texture feature value as the normalized auto correlation function (proposed by Li Xiaofeng) (Normalized auto-correlation coefficients) [6]. Suppose that in an image of size M x W, we define pixels(i, j)and the normalized autocorrelation function of pixels $(i+\Delta m, \mathrm{j}+\Delta n)$:

$$\gamma(\Delta m, \Delta n) = \frac{A(\Delta m, \Delta n)}{A(0,0)} \tag{12}$$

where

$$A(\Delta m, \Delta n) = \frac{1}{(m-\Delta m)(n-\Delta n)} \sum_{x=0}^{m-\Delta m-1} \sum_{y=0}^{n-\Delta n-1} f(\mathrm{x}, \mathrm{y})\, \mathrm{f}(\mathrm{x}+\Delta m, y+\Delta n) \tag{13}$$

In addition, in order to not influenced by the brightness of action image, we further improve the equation to exclude the average brightness of the image $\overline{f}$

$$A^1(\Delta m, \Delta n) = \frac{1}{(m-\Delta m)(n-\Delta n)} \sum_{x=0}^{m-\Delta m-1} \sum_{y=0}^{n-\Delta n-1} (f(\mathrm{x}, \mathrm{y}) - \overline{f})(\mathrm{f}(\mathrm{x}+\Delta m, y+\Delta n) - \overline{f}) \ldots.$$

Therefore, the texture features used in this paper can be defined as follows:

$$\gamma^1(\Delta m, \Delta n) = \frac{A^1(\Delta m, \Delta n)}{A^1(0,0)} \qquad (14)$$

3.2 Texture feature geometry feature

Different from the focus of the texture features, the geometric characteristics of human action image is to describe the type and edge of the action, and it is a poor variability, rather than the recognition rate caused by different scanners and image characteristics. Given that the outline of the action is usually different and that the benign and the malignant presented a different appearance, we also choose the geometric characteristics of the action as a basis for identification in this paper[5]. Totally, six geometric features are included, which is defined as follows:

$$F\mathrm{orm}\_\,factor = \frac{4\pi \cdot Area}{Perimenter^2}$$

(15)

$$Roundness = \frac{4\pi \cdot Area}{\pi \cdot Max\_Diameter^2}$$

(16)

$$Aspect\_Ratio = \frac{Max\_Diameter}{Min\_Diameter}$$

(17)

$$Convexity = \frac{Convex\_Perimenter}{Perimenter}$$

(18)

$$Solidity = \frac{Convex\_Area - Area}{\sum_{i=1}^{N} Convex\_Area_i - Area_I \,/\, N}$$

(19)

$$Extent = \frac{Area}{Bounding\_Rectangle}$$

(20)

Here Area and Perimeter are the area and perimeter of the action, Max_Diameter and Min_Diameter are the maximum and minimum diameters of the action respectively (Figure 3(a)(b)), Convex_Area and Convex_Primeter are the surrounding action. Bounding_Rectangle is the smallest rectangular area that surrounds the action (as shown in Figure 3 (d)). As for N, it is the minimum rectangle area of all the cases, as shown in Figure 3 (c) Quantity.

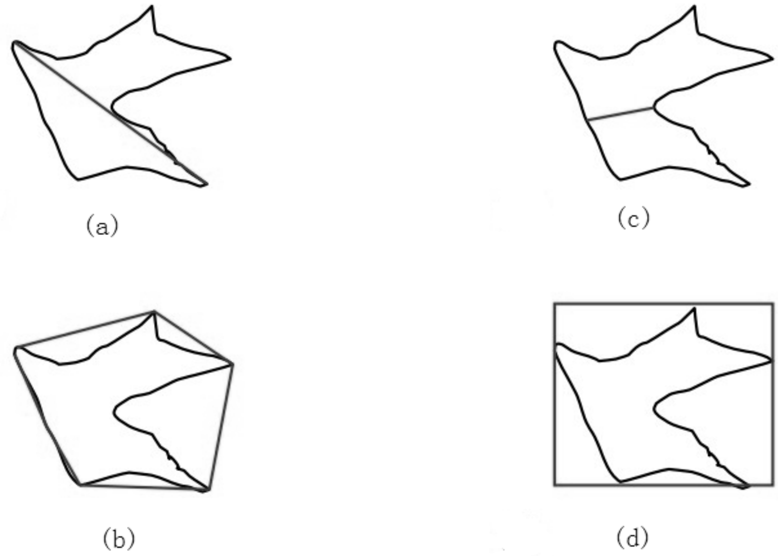


Fig. 3 Schematic diagram of action extraction:

(a) the largest diameter (b) the smallest convex polygon that surrounds the action

(c) Minimum diameter (d) The smallest rectangle surrounding the action

4 Statistical Features

4.1gabor features

Gabor wavelet transform is an exceptional case of Fourier transform. Based on the Fourier transform of region, the local information in Fourier transform is found by using a local time region function. Dimensional normalization and gray value normalization have been widely used. Since the size of the normalized dimension depends on the size of the filter, and gray-scale can reduce the normalization of the impact of light on the image, making the subsequent processing becomes easy and convenient. The Gabor filter has the following formula:

$$\mathrm{g(x)} = \frac{k^2}{\sigma^2} \cdot e^{-\frac{k^2 \cdot \|x\|^2}{2\sigma^2}} \cdot [e^{ikx} - e^{-\frac{\sigma^2}{2}}] \tag{21}$$

In the formula (21), X is the vector with directional characteristic, it is $\mathrm{X} = (\mathrm{x,y}) = \mathrm{x}\cos\theta + y\sin\theta$, the standard deviation σ^2 is the bandwidth of Gabor filter, and $\sigma = 2\pi$ is set. The expression $-e^{-\frac{\sigma^2}{2}}$ is used to reduce the illumination variation, and k represents the different frequencies in the wavelet. In other words, k is the different high frequency band in the wavelet, K is $2^{-s/2} \cdot 2^{-\pi}$, and σ is the phase angle of the Gabor wavelet. In general, the Gabor filters with different frequencies and phase angles are used in 5 different frequencies and 8 different directions. Using 40 different frequency and phase angle of the filter, the original image is operated with the convolution operation to get the filtered results, and the calculation formula is:

$$G_{s,\theta}(x,y) = f(x,y) * g_{s,\theta}(x,y)$$

(22)

In the formula (22), f(x, y) represents the original input image, while the sign bit convolution operation, $g_{s,\theta}(x,y)$ represents the frequency s, phase angle θ Gabor filter. Using the above formula, the image of different frequency characteristics can be obtained by inputting the phase angle of image, as shown below:

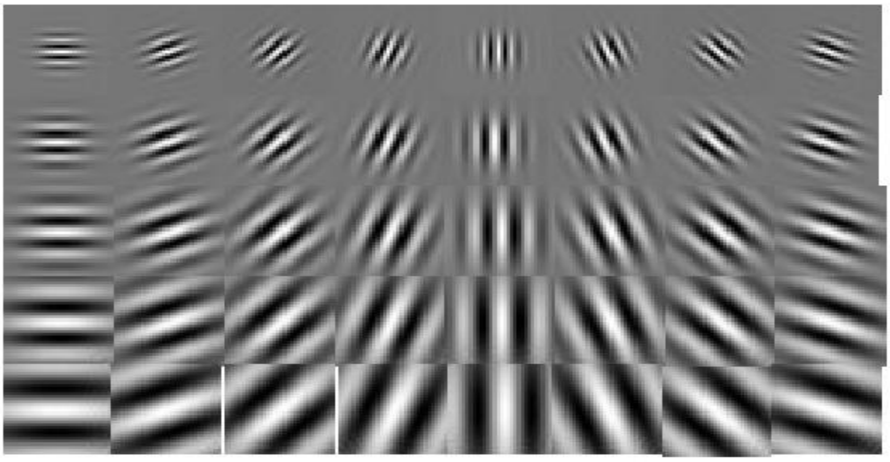


Fig. 4 Gabor filters consisting of eight different phase angles in five frequency domains

After the input image is extracted by the Gabor filter feature, it enters the stage of classification recognition.

4.2 Feature Extraction Based on PCNN and EMD

4.2.1 PCNN (Pulse Coupled Artificial Neural Network)

Since the 1990s, Eckhorn et al. (2000) has studied the phenomenon of pulse synchronous oscillations in visual cortical neurons of cats. The model of mammalian neurons was obtained, modifying the model proposed by Eckhorn and obtaining a pulse coupled neural network model[9].

The mathematical form of the model can be described by the following four equations:

$$F_{ij}(n) = \exp(-\tau_{F_{ij}})F_{ij}(n-1) + V_F \sum_{k,l} M_{ijkl} Y_{kl}(n-1) + I_{ij}$$

(23)

$$L_{ij}(n) = \exp(-\tau_{L_{ij}})L_{ij}(n-1) + V_L \sum_{k,l} M_{ijkl} Y_{kl}(n-1)$$

(24)

$$U_{ij}(n) = F_{ij}(n)(1+\beta L_{ij}(n))$$

(25)

$$\theta_{ij}(n) = \exp(-\tau_{\theta_{ij}})\theta_{ij}(n-1) + V_\theta Y_{ij}(n-1)$$

(26)

$$Y_{ij}(n) = \begin{cases} 1 & U_{ij}(n) \geq \theta_{ij}(n) \\ 0 & U_{ij}(n) < \theta_{ij}(n) \end{cases}$$

(27)

Among them:

β: connection factor of internal activity

F, V_F , τ_F : feedback input field, its amplification factor, decay time constant

L, V_L , τ_L : coupling connection domain, its amplification factor, decay time constant

θ , V_θ , τ_θ : dynamic threshold, its amplification factor, decay time constant

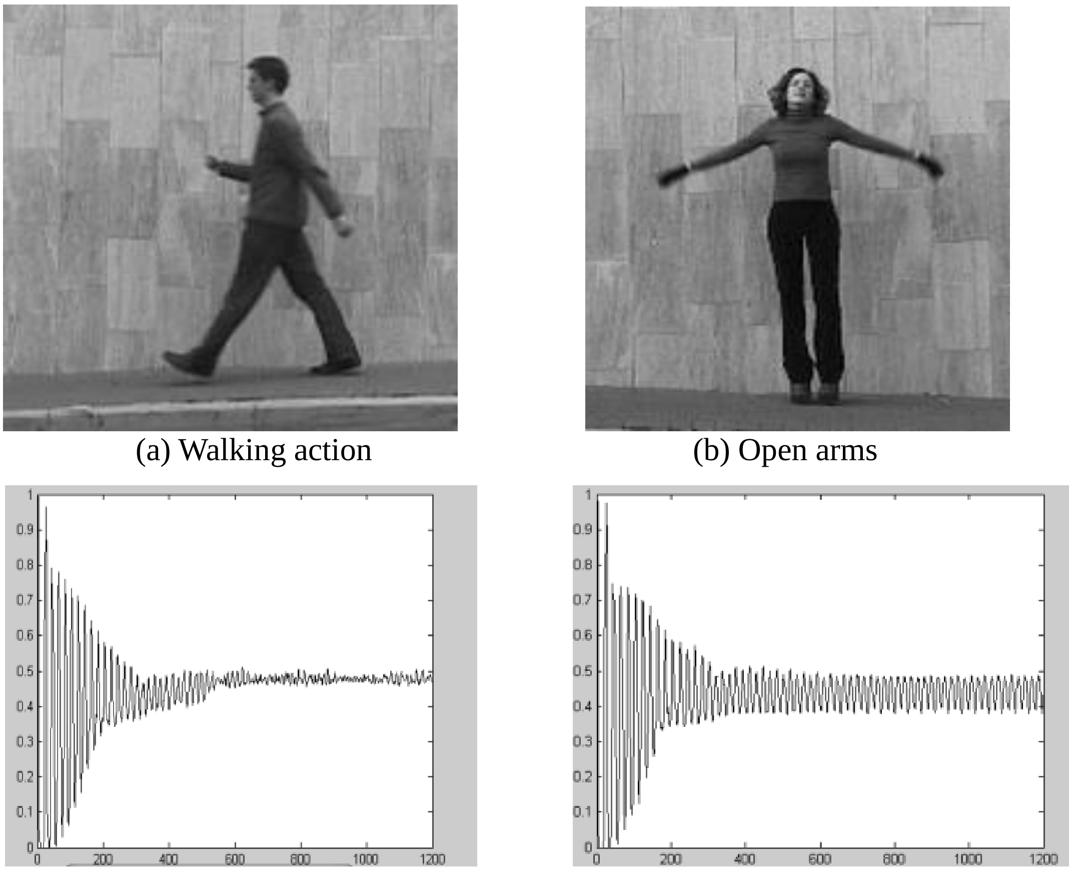
I: external excitation of neuronal forced excitation

U: Internal activity item

Y: Output pulse (0 or 1)

From the basic model it can be deduced that PCNN can be divided into two cases: uncoupled and coupled. According to the coefficient of link β, when β equals to 0, it can be considered that each neuron is a combination of independent operation. In this case, the neurons cycled, worked, and stimulated to produce pulses [10]. And the pixels with different gray values are not affected by other neurons, the independent firing frequency depends on the gray value of the pixel, and has nothing to do with the other pixels. At the same time, it can be concluded that the neurons with similar grayscale values will ignite at the same time. This conclusion also confirms the intrinsic mechanism of PCNN synchronization pulse release, which is an ideal situation.

When beta is close to zero, that is, when there is a coupling, the firing of neurons is affected not only by its own gray value but also by the neurons linked to its existence. Several adjacent neurons that meet the capture range are stimulated ahead of time to achieve ignition. This is also the mechanism in which the neurons in spatial approximation or gray-scale approximation generate bursts synchronously. From PCNN mathematical model we can know that the greater the value of connection coefficient β and the connection domain L, the greater the range of ignition. So the PCNN can be distributed in the coupling sequence of synchronous burst sequence (Neuronal oscillations).



(a) Walking action (b) Open arms

(c) Walking ignition sequence diagram (d) ignition sequence diagram of open arms

Fig 5. Sequence of PCNN for different actions

4.2.2 PCNN Feature Extraction Based on EMD Method

This section extracts action images firstly, using the EMD method according to the characteristics of the waveform data.

EMD (Empirical Mode Decomposition) method is proposed by Dr. Huang of the United States NASA. Decompose the signal Based on the time scale of the data itself to, there is no need for it to set any basis function in advance [7][8].

In order to reduce the model aliasing problem of EMD, Wu et al. proposed the EEMD method based on the EMD decomposition characteristic of white noise, based on the statistical mean of uncorrelated random sequences Not equal to zero . Gaussian white noise of a certain amplitude is added to the signal $x(t)$ to be decomposed each time,and the signal $x(t)$ is subjected to EMD decomposition. Then the IMFs components obtained by multiple decomposition are averaged, and the mean of the IMFs is regarded as the final IMFs [11]. Specific steps are as follows:

Step 1 The EMD was used to decompose the original signal, and the IMFs were screened. Finally, the residual IMFs and the residual signals decomposed by EMD were recombined to obtain $x(t)$;

Step 2 Make M the number of EMD decompositions, k the coefficient of white noise added, initialize m = 1;

Step 3 And the m-th EMD decomposition is performed on the reconstructed signal $x(t)$ in the first step

(1) Each time a different Gaussian white noise $n_m(t)$ is added to the original signal $x(t)$, the white noise signal is obtained $x_m(t)$.

$$x_m(t)= x(t)+ kn_m(t) \tag{28}$$

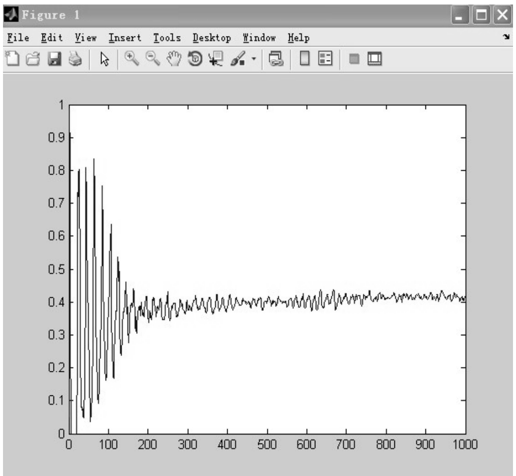
(2) $x_m(t)$ is decomposed with EMD, L components of eigenfunction modal are obtained, where $IMF_{i,m}$ is the m-th intrinsic mode function of the m-th EMD decomposition.

If $m<M$, set $m=m+1$,return (1) to continue；

Step 4 The mean values of IMFs obtained by EMD decomposition are calculated.

$$\overline{IMF_i} = \frac{\sum_{m=1}^{M} IMF_{i,m}}{M} (i=1,2,\cdots L, m=1,2,\cdots M) \tag{29}$$

And taking $\overline{IMF_i}$ as the i-th eigenmode component IMF_i obtained by EEMD-decomposition. The following figure shows the decomposition of PCNN waveforms using the EMD algorithm.



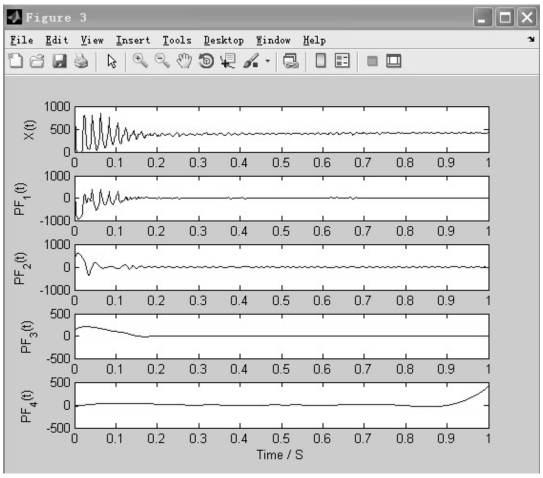


Fig 6 (a). PCNN extracted waveforms

Fig 6(b). LMD processing of the waveform

4.2.3 Feature description of waveform signal based on fractal dimension

As shown in the PCNN ignition sequence in Figure6(a) , we use the LMD method to process the waveform data as shown in Figure 6(b) . And it is used as the new expression feature combined with the eigenvector extracted by gabor.

In order to further describe the decomposed waveform, and for the convenience of EGG classification, the algorithm of fractal dimension that proposed by Katz is used for obtaining the dimension of each wavelet [12].

$$FD = \frac{\log n}{\log n + \log \frac{D}{L}}$$

(30)

The definition of fractal dimension is as follows, where L is the total length of the signal waveform curve, i.e. the sum of the distances between two successive points; d is the distance between the first point and another point in the sequence, and the specific point is the furthest point away from the first point, i.e. the expression of d is

$d=\max(\mathrm{dist}(s_1-s_2))$

(31)

n is the number of the steps in the waveform and n=L/a, where a is the average distance between two successive points.

Calculating the corresponding fractal dimension to each of the wavelets, and we can obtain five eigenvalues for each electrogastric signal sequence. At the same time, fractal dimension for the original signals are also calculated, hence, a six dimensional eigenvector can be acquired as $X=[x_1,x_2\ldots x_6]$.

5. Experiment analysis

5.1 Experimental data

What we are using now is Olympic Sports Dataset which contains videos of athletes practicing different sports. We have obtained all video sequences from YouTube and annotated their class label with the help of Amazon Mechanical Turk [13].

1. Geometric characteristics

Bone and malignant actions were extracted from the above mentioned texture features and six geometric features. Each action image was divided into 30 feature values, including 24 texture features and 6 geometric features according to the previous geometric feature formula.

$$\gamma^{1}(\Delta m,\Delta n)=\frac{A^{1}(\Delta m,\Delta n)}{A^{1}(0,0)}(\Delta m,\Delta n=0\sim 4,\Delta m{=}\Delta n=0) \quad (32)$$

24 features, $F\text{orm\_}\, factor$, $Roundness$, $Aspect\_Ratio$ $\;Convexity$, $Solidity$, $Extent$ -6 features.

Therefore, the action totally eigenvalues 30 properties, of which there are 24 texture features, 6 geometric features.

2. Statistical characteristics

The program takes 32 Gabor kernel functions in four directions (v = 0, 1, ..., 3) and 8 directions (ie K = 8, u = 0, 1, ..., 7).

At the same time, fractal dimension for the original signals are also calculated. Hence, a six dimensional eigenvector can be acquired as X=[x_1,x_2...x_6].

So there are 38 statistical features, composed of 6 fractal features (PCNN-EMD) and 32 gabor characteristics.

5.2 Feature dimensionality

In summary, a 68-demensional eigenvector can be obtained by using the features combined by sampling entropy and fractal dimension. If conducting decomposition in sections for a set of collected data (the data are divided into N sections), then a vector of N*68 dimensions can be gained. In this case, a dimension reduction is necessary, and the PCA algorithm is employed in terms of features [10][11].

PCA is to find the direction of the projections that may best represent the original data in the sense of the least mean square deviation, and the projections are orthogonal. The process can be simply described as: assuming that there are M samples of n dimensions x_1, $x_{1,\dots}x_M$.

Calculate the n-dimensional mean vector m and the covariance matrix Σ of n×n rank:

$$\Sigma = \frac{1}{M}\sum_{i=1}^{M}\left(x_i - m\right)\left(x_i - m\right)^T \tag{33}$$

Calculate the eigenvalue of Σ and the eigenvector, and then select the eigenvector of the first k maximum eigenvalue as the direction of the principal component according to the variance contribution, and it is generally considered that the rest n-k directions are contributed by noise.

A matrix A of n×k can be comprised with k eigenvalues. The principal component representation can be obtained by projecting the original data to a k-dimensional space, and it can be expressed as:

$$y = A^t\left(x - m\right) \tag{34}$$

PCA is a linear method, so it cannot represent the nonlinear structure in the data correctly, which limits the capacity of handling complex problems for PCA model. Scholkopf extended PCA to a non-linear area, which is mainly to make PCA transformations in a high dimensional space. Such method transforms the non-linear problem into normal problem of eigenvalue using kernel skills, which is called kernel PCA method (KPCA)[16]. KPCA method can be widely applied in the field of face recognition, handwritten digit recognition, etc. and displays a better performance compared to the PCA algorithm. In astronomy, the structure of the acquired spectrum data collection cannot be expressed because of its classification, redshift coupling, and high dimensional features. Therefore, we tried to use KPCA algorithm for non-linear structural feature extraction. Differences in data structure extraction of the PCA and KPCA algorithms are shown in Fig.7

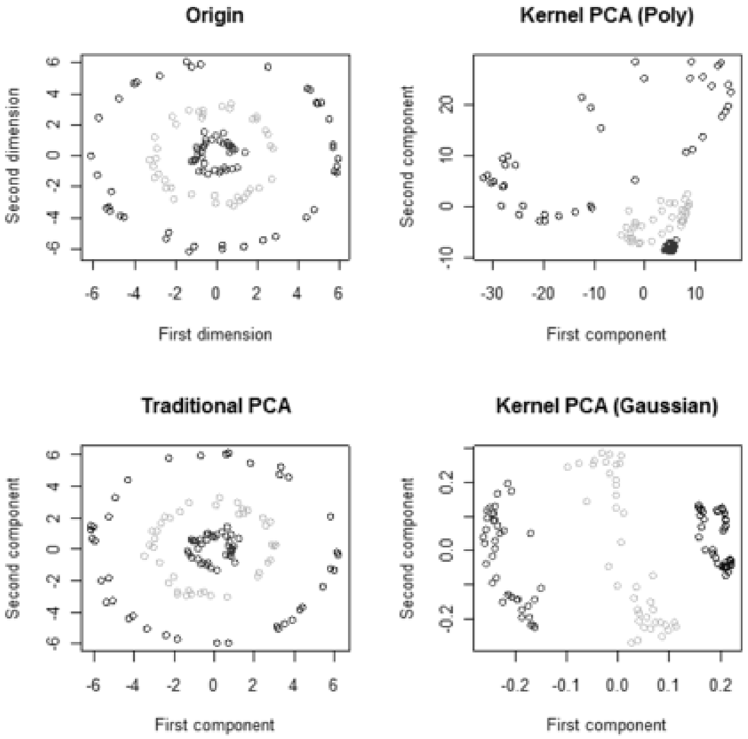


Fig 7. Feature extraction when using PCA and KPCA

Only when appropriate non-linear mapping Φ is selected, will the proper direction in the feature space be found to represent the direction of highest variance for the data set. The key of KPCA is the selection of the kernel function and its relevant parameters. Different kernel functions and parameters represent different non-linear mapping, and this is the common problem that exists in the algorithms based on kernel skills. This is a N-P C problem and yet there is no perfect solutions to it. In the experiments, different kernel widths and number of the principal components will be taken and the influences of the values will be discussed.

5.3 Result

The current release of the Olympic Sports Dataset contains 16 sports. We divided the data into five groups, and selected one group to test, the other four groups are used to train SVM. SVM parameters C and γ are in the form of random numbers, C is between 1 and 1000, and γ is between 0.001 and 0.1. The final classification accuracy is shown in Table 1 below.

Table 1 Results contrast

Sport class	Laptevet al[15]	Laptev et al[13]	Our Method
high-jump	52.4%	68.9%	70.3%

long-jump	66.8%	74.8%	72.9%
triple-jump	36.1%	52.3%	68.2%
pole-vault	47.8%	82.0%	85.7%
gymnastics-vault	88.6%	86.1%	85.3%
shot-put	56.2%	62.1%	77.4%
snatch	41.8%	69.2%	78.1%
clean-jerk	83.2%	84.1%	91.3%
javelin-throw	61.1%	74.6%	85.7%
hammer-throw	65.1%	77.5%	78.9%
discus-throw	37.4%	58.5%	70.6%
diving-platform	91.5%	87.2%	91.5%
diving-springboard	80.7%	77.2%	79.1%
basketball-layup	75.8%	77.9%	85.2%
bowling	66.7%	72.7%	87.3%
tennis-serve	39.6%	49.1%	73.1%
Average	62.0%	72.1%	80.04

6. Conclusion

This shows that using PCNN-EMD-based feature recognition method can obtain a high accuracy rate, while the fusion feature using KPCA has a higher recognition rate. The experimental results show that the proposed method is effective [13][15].

References

1. P.A Dhulekar, S.T Gandhe, H Chitte, K Pardeshi, *Human Action Recognition: An Overview*, 2016, Proceedings of the International Conference on Data Engineering and Communication Technology, 2016, Lavasa, India, pp 481-488.
2. A. Sadhu, An Application of Multivariate Empirical Mode Decomposition Towards Structural Modal Identification, 2016, Rotating Machinery, Hybrid Test Methods, Vibro-Acoustics & Laser Vibrometry, vol. 8, no.1, pp 303-309.
3. Y. Wang, G Mori, Human action recognition by semilatent topic models, 2009, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31 , pp.1762-1774
4. R.M.G Tello, S.M.T Müller, T.F Bastosfilho, A Ferreira, Comparison of new techniques based on EMD for control of a SSVEP-BCI, 2014, IEEE International Symposium on Industrial Electronics ISIE IEEE, pp.992-997.
5. Y. Juan, G. Stevenson, S. Dobson, KCAR: A knowledge-driven approach for concurrent activity recognition, 2015, Pervasive & Mobile Computing, vol. 19, pp 47-70.

6. H.B Pan, J Li, Online human action recognition based on improved dynamic time warping, 2016, IEEE International Conference on Big Data Analysis, 2016, Washington DC, USA,pp 1-5.
7. D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, 2005, Mechanical Systems & Signal Processing, vol. 19, no. 2, pp 259-270.
8. J.J Seo, H.I Kim, W.D Neve, M.R Yong, Effective and efficient human action recognition using dynamic frame skipping and trajectory rejection ☆, 2016, Image & Vision Computing , vol.58, pp 76-85.
9. Z.B Wang, X.G Sun, Y.N Zhang, Z Ying, Y.D Ma, Leaf recognition based on PCNN, 2016,Neural Computing and Applications ,vol. 27, no. 4, pp 899-908.
10. A.K Helmy, G.S El-Taweel, Image segmentation scheme based on SOM–PCNN in frequency domain, 2016, Applied Soft Computing, vol.40, pp 405-415.
11. M Van, H.J Kang, Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection, 2015, Iet Science Measurement Technology, vol. 9, no.6, pp 671-680.
12. M. Kartz, Fractals and the analysis of waveforms, 1988, Computers in Biology & Medicine, vol. 18, no. 3, pp.145-156.
13. J.C Niebles, C.W Chen, F.F Li. Modeling Temporal Structure of Decomposable Motion Segments for Activity Classification, 2010, 11th European Conference on Computer Vision (ECCV), 2010, Heraklion, Crete, Greece, pp.\392-405.
14. C.C Chang, C.J Lin, LIBSVM: A Library for Support Vector Machines, 2011, ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 1-27.
15. I. Laptev, M Marszalek, C Schmid, B Rozenfeld, Learning realistic human actions from movies, 2008, computer vision and pattern recognition, 2008, Anchorage, AK, USA, pp.1-8.
16. S.B Ginsburg, G Lee, S Ali, A Madabhushi, Feature Importance in Nonlinear Embeddings (FINE): Applications in Digital Pathology, 2015, IEEE Transactions on Medical Imaging, vol. 35, no.1, pp.76-88.