Human Motion Sequence Recognition Based on Feature Selection and Support Vector Machine

Yu Yunlei¹, Wang Mei¹*, Lin Limeng¹ and Zhang Chen¹

¹College of Mechatronics Engineering and Automation, Shanghai University, Shanghai, 200444, China

*Corresponding author’s e-mail: mwang@shu.edu.cn

Abstract. Aiming at the problem of human motion sequence recognition, algorithm based on feature selection and support vector machine is proposed. Firstly, the feature extraction of human motion sequences is obtained by key frame and human joint angle calculation. Then, based on the Pearson correlation coefficient and CFS evaluation function, the algorithm of relevance feature selection is used to search the optimal feature subset from the original feature set. By reducing the dimension of the feature set, the difficulty of classification recognition is reduced. In the classification process, the support vector machine is used as the classifier to complete the recognition task of the human motion sequence. Through the recognition experiment and the contrast experiment, the effectiveness of the recognition algorithm based on feature selection and support vector machine is proved.

1. Introduction
The main content of motion recognition is to identify human body movements with specific algorithms and these body movements are detected and captured by specific devices and methods. Through the analysis, integration and extraction of the captured information, effective features can be found for efficient classification. The purpose of motion feature extraction is to extract the information related to human motion from the original image, including silhouette, optical flow, gradient, spatial and temporal characteristics, depth characteristics [1-7], etc. Motion recognition algorithms are mainly divided into three categories: probability-based methods, semantic-based methods and template-based methods. Statistical methods include hidden markov model (HMM), dynamic bayesian network (DBN), etc. Template-based methods include template matching [8], dynamic time ordering [9], etc. In this paper, the content will be introduced from the aspects of feature extraction, classification algorithm implementation. And classification experiment will be done for verification.

2. Arm motion feature extraction based on joint Angle calculation
Since the angles of human joint are corresponding to the spatial information of human body, they can reflect human body movements with reduced dimension of feature space. Then the human joint Angle was selected to construct the feature vector.
All joint data trained and tested in this paper are obtained through Kinect sensor and skeleton tracking in OpenNI open source library. According to the requirements of feature extraction, torso (P_t), left shoulder (P_l), right shoulder (P_r), left elbow (P_e) and left wrist (P_w), shown in figure 1, are selected as the key points to analyze human arm movements.
Figure 1. Kinect coordinate system and human body nodes

The coordinate system $C_w$ of Kinect sensor (Figure 1) is used as the reference to collect real-time data of human joint movement. In order to improve the efficiency of angle calculation, the shoulder coordinate system $C_S$ is defined here. For example, the shoulder coordinate system $C_S$ of the left arm is shown in Figure 2. The coordinate of joint points, detected by Kinect, can be transformed into the shoulder coordinate system $C_S$ for joint angle calculation.

Figure 2. Shoulder coordinate system and joint angle

Several joint angles are defined to reflect the arm movement. $\theta_1$ is defined as horizontally abducted of the shoulder joint. $\theta_2$ is the flexion of the shoulder joint. $\theta_3$ is opposite to the rotation of the arm for the shoulder joint. While $\theta_4$ is the flexion extension of the elbow joint (the angle between the upper arm and the lower arm). As shown in Figure 3, $\theta_3$ is calculating in elbow coordinate system.

The joint angles can be obtained:

$$\theta_1 = \arctan \frac{\frac{p_r P_c \cdot \vec{x}}{p_r P_c \cdot \vec{y}}}{\frac{p_r P_l \cdot \vec{y}}{P_r P_l \cdot \vec{y}}} = \arctan \frac{p_r P_m \cdot \vec{y}}{P_r P_m \cdot \vec{y}}$$

(1)

$$\theta_2 = \arctan \frac{\frac{p_r P_c \cdot \vec{x}}{p_r P_c \cdot \vec{y}}}{\frac{p_r P_e \cdot \vec{y}}{P_r P_e \cdot \vec{y}}} (\vec{z} = \vec{x} \times \vec{y})$$

(2)

$$\theta_4 = \arctan \frac{\frac{p_r P_e P_w}{p_r P_e P_w}}{\frac{p_r P_c P_e P_w}{P_r P_c P_e P_w}}$$

(3)

Figure 3. Elbow coordinate system and $\theta_3$ in the elbow coordinate system

$$\theta_3 = \arctan \frac{\frac{\vec{w} \cdot \vec{y}_e}{\vec{w} \cdot \vec{x}_e}}{\vec{w} \cdot \vec{x}_e}$$

(4)

Human motion sequence detected by Kinect are time series motion frames. Based on subsequent experimental verification, 5 frames (Start frame, end frame, and 3 intermediate frame) are enough to represent the whole motion process. Then the eigenvector obtained in this paper is:

$$A = (\theta_1^1, \theta_1^2, \theta_1^3, \theta_2^1, \theta_2^2, \theta_2^3, \theta_4^1, \theta_4^2, \theta_4^3, \theta_4^4, \theta_4^5)$$

In which, the Angle of the joint is calculated by means of spatial transformation, whose superscript indicates the number of frames $\theta_1$–$\theta_4$. The number 1 is the starting frame, number 2-4 is the middle frame, and number 5 is the end frame. Figure 4 shows a complete sequence of actions.

Figure 4. A complete sequence of actions
3. Correlation feature analysis and Feature optimization
In order to improve the efficiency of recognition and increase the calculation speed, the feature should be optimized after correlation analyzing.

The feature constructed above is normalized firstly by linear normalization:

\[ y = \frac{\theta_i - \theta_{\text{minVal}}}{\theta_{\text{maxVal}} - \theta_{\text{minVal}}} \]  

(5)

Where \( \theta_{\text{maxVal}} \) and \( \theta_{\text{minVal}} \) represent the maximum and minimum values of one-dimensional variable in original feature vector. \( \theta_i \) is the value to be normalized. And \( \theta_i \) is the result, whose value is between 0 and 1. For the convenience of analysis, the characteristic vector, which described the motion sequence, is the normalized feature \((\varphi_1, \varphi_2, \varphi_3, \ldots, \varphi_{19}, \varphi_{20})\).

To evaluate the correlation of variables in the normalized feature, Pearson correlation coefficient was selected as the basis. For example, in the correlation analysis of the characteristic vectors of \( n \) groups of motion sequences, the Pearson correlation coefficient of \( \varphi_1 \) and \( \varphi_2 \) is defined below:

\[ r_{\varphi_1 \varphi_2} = \frac{\sum^n_{i=1}(\varphi_1i - \bar{\varphi_1})(\varphi_2i - \bar{\varphi_2})}{\sqrt{\sum^n_{i=1}(\varphi_1i - \bar{\varphi_1})^2 \sum^n_{i=1}(\varphi_2i - \bar{\varphi_2})^2}} \]  

(6)

Where \( \bar{\varphi_1}, \bar{\varphi_2} \) respectively is the average of \( \varphi_1 \) and \( \varphi_2 \) of \( n \) groups.

Similarly, Pearson correlation coefficients of other variables can be obtained, and the correlation coefficient matrix of characteristic variables are shown in Table 1.

**Table 1.** Pearson correlation coefficient matrix of each characteristic variable

| \( \varphi \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 |
| 2 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 |
| 3 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 |
| 4 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 |
| 5 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 |
| 6 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 |
| 7 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 |
| 8 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 |
| 9 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 |
| 10 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 |
| 11 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 | 0.80 |
| 12 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 | 0.85 |
| 13 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 | 0.80 |
| 14 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 | 0.85 |
| 15 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 0.80 | 0.85 | 1 |

To analyze the feature relevance of each feature subset, Merit is used as the evaluation function in CFS algorithm (correlation-based feature selection), which takes into account of the relevance between features and classes.

\[ \text{Merit}_s = \frac{k \bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \]  

(7)

Where: \( k \) is the number of features in the current feature subset, \( k \in (1, 20) \). \( \bar{r}_{cf} \) is the average correlation coefficient involving all characteristics and all classes in the current feature subset. \( \bar{r}_{ff} \) is the average correlation coefficient involving only the characteristics. \( \bar{r}_{cf}, \bar{r}_{ff} \) can be drew from table 1. Merit values for various subsets of feature vector and the solution are shown in Table 2.

**Table 2.** Partial Merit and their solution

| Feature set | \( k \) | \( \bar{r}_{cf} \) | Merit |
|---|---|---|---|
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |
| [1] | 1 | 0.75 | 0.87 |

To look for the optimal feature subset, the best-first-search algorithm is adopted to find the feature subset with the largest merit value. Then the optimal subset with only 17 variables (without \( \varphi_1, \varphi_7 \) and \( \varphi_13 \)) is obtained to resemble the human arm movement, and its merit is 0.89.

4. Human motion classification and recognition based on support vector machine
To classify and recognize the human motion, SVM (support vector machine) algorithm is applied here.
The basic principle of SVM is to use hyperplane to conduct sample classification. As shown in Figure 5, when faced to the problem of linear indivisibility, it solves by mapping sample data into high-dimensional space by referring to kernel functions. The radial basis kernel function \( K(x_1, x_2) = \exp(-\gamma \|x_1-x_2\|^2) \) is selected here.

**Figure 5.** Mapping of low-dimensional data to higher-dimensional data

In the process of classification and recognition, a one-to-one method is selected for SVM to get one label from 6 kinds of motion sequence. Given the training sample set \( \{(x_1, y_1), \ldots, (x_i, y_i)\} | i = 1, 2, \ldots, n; y \in \{1, 2, 3, 4, 5, 6\} \), \( x_i \) represents the sample vector, which is the optimized feature vector \( (\phi_2, \ldots, \phi_6, \phi_{12}, \phi_{14}, \ldots, \phi_{20}) \), while \( y_i \) is the label result.

The objective function and its constraints are:

\[
\min_{\frac{1}{2} \| w \|^2}, \text{s. t.}, y_i(w^T \cdot x_i + b) \geq 1, i = 1, 2, \ldots, n
\] (8)

The parameters \( w \) and \( b \) are the normal vector and intercept of the hyperplane respectively. This problem can be solved by Lagrange multiplier method. And according to Karush - Kuhn - Tucker best optimized conditions, \( w^*, b^* \) can be calculated below:

\[
\begin{cases}
    w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i \\
    b^* = -\left(\min_{y_i=+1} (w \cdot x_i) + \max_{y_i=-1} (w \cdot x_i)\right)/2
\end{cases}
\] (9)

For the unknown test sample input \( x_i \), judge the category by the following formula:

\[ f(x) = sgn(\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, X) + b^*) \]

\( sgn \) is the sign function, \( <x_i, X> \) is inner product of the vector \( x_i \) and \( X \).

In SVM, the two important parameters \((C, \gamma)\) were adjusted to get the best result. Therefore, under the training sample, the best parameters were \( C = 128 \) and \( \gamma = 0.0078125 \). Thus the prediction accuracy of the theoretical model will reach 96.25%, showing in Figure 6.

**Figure 6.** Influence of parameters c and gamma on accuracy

Training to complete the model to the characteristics of the input vector \( x_i \) to classify the classification results.

The process of classification recognition is shown figure7 below.

**Figure 7.** Flow chart of recognition algorithm
5. Experiment

5.1 Verification experiment

In the experiment, six representative movements (including lateral horizontal lift, raising hands, making a phone call, drinking water and eating, shaking hands and shaking hands) are taken as objects and numbered from 1 to 6. In the collection of training samples, the three experimental subjects performed 6 types of movements respectively, and 30 sets of data for each movement were collected, thus 180 samples were obtained to compose the training sample set.

In the verification process, another three experimental subjects performed each type of movements 20 times respectively. Thus 120 human motion sequence was collected to form the test sample set and used as the input of the trained system to judge the accuracy of the algorithm.

The classification results of different action sequences are shown in Table 3:

|        | 1  | 2  | 3  | 4  | 5  | 6  | Classification accuracy |
|--------|----|----|----|----|----|----|--------------------------|
| Sample | 20 | 16 | 2  | 2  | 2  | 20 | 100%                     |
| Sample | 2  | 2  | 18 | 2  | 20 | 20 | 100%                     |
| Sample | 4  | 2  | 20 | 20 | 20 | 20 | 100%                     |
| Sample | 5  | 2  | 20 | 20 | 20 | 20 | 100%                     |
| Sample | 6  | 2  | 20 | 20 | 20 | 20 | 100%                     |

It can be seen from the above table that the average recognition rate is as high as 114/120=95%. Among them, the classification error of action 2 and action 3 is 20% and 10%, and the classification of other actions is correct all. Further analysis of the classification result shows that the main reason for misjudgment is the high similarity between action 2 and action 3.

5.2 Comparison experiment with KNN and single SVM algorithm

In order to verify the superiority of the algorithm developed in this paper, an experiment compared with KNN (k=6) and SVM is conducted. The training sample set and test set are exactly the same as the verification experiment. The result is shown in Table 4:

| Classification algorithm | Recognition rate |
|--------------------------|------------------|
| SVM                      | 90%              |
| KNN                      | 91.67%           |
| Algorithm of this paper  | 95%              |

As can be seen from the table, compared with the KNN classification algorithm and the original SVM classification algorithm, the accuracy of recognition algorithm based on feature selection and SVM was significantly improved to 95%. The verification experiment and comparison experiment prove that the algorithm developed in this paper is effective.

6. Conclusion and Discussion

A motion sequence recognition algorithm, based on the combination of relevance feature selection and SVM for classification, is proposed in this paper. It reduces the dimension of the feature vector, improves the effectiveness of the feature, and guarantees the robustness. The experiment finally proves the validity of the algorithm for human movement classification.

Acknowledgements

The work described in this paper was supported by National Key Research and Development Program Key Project (Project No. 2016YFC0302400).

References

[1] Jiang Zhihao, Li Haoxian, Zhao Chunxia. Pedestrian activity prediction algorithm based on
space-time model video monitoring [J]. Computer applications and software, 2017, 34(1):149-153.

[2] Liu Hechen. Study on accurate recognition and simulation of athletes motion 3d visual image [J], 2016, 33(8):369-372.

[3] Chen Jing. Gesture recognition technology based on Kinect and its application in teaching [D]. Shanghai jiaotong university , 2013.

[4] Yin J Q, Tian G H, Jiang H T, et al. Human action recognition oriented to family service[J]. Journal of Sichuan University, 2011, 42(4):101-107.

[5] Zhu G, Xu C, Huang Q, et al. Player action recognition in broadcast tennis video with applications to semantic analysis of sports game[C]// ACM International Conference on Multimedia. ACM, 2006:431-440.

[6] Johansson G. Visual motion perception [J]. Proceedings of the National Academy of Sciences of the United States of America, 1975, 92(7):2433-2440.

[7] Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information [M]. Henry Holt and Co. Inc. 1982.

[8] Francomano E, Tortorici A, Calderone V. Regularization of Optical Flow with M-Band Wavelet Transform. Computers & Mathematics with Applications, 2003, 45 (1/2/3) :437-452.

[9] Chen Lifen, Liao H M, Lin Jachen. Wavelet-Based Optical Flow Estimation. IEEE Trans on Circuits and Systems for Video Technology, 2002, 12 (1) :1-12.