

Classifying Gait Data using different machine learning techniques and finding the optimum technique of classification

Anubha Parashar¹; Apoorva Parashar²; Somya Goyal³

Vaish College of Engineering, Rohtak¹; Maharshi Dayanand University, Rohtak²;
PDM College of Engineering, Bahadurgarh³.
{anubhaparashar1025¹; apoorvaparashar0000²; somyagoyal1988³}@gmail.com

Abstract. The Classification of Humanoid locomotion is a troublesome exercise because of non linearity associate with gait. The high dimension feature vector requires a high computational cost. The classification using the different machine learning technique leads for over fitting and under fitting. To select the correct feature is also the difficult task. The hand craft feature selection machine learning techniques performed poor. We have used the deep learning technique to get the trained feature and then classification we have used deep belief network based deep learning. Classification is utilized to see Gait pattern of different person and any upcoming disease can be detected earlier. So in this paper we first selected the feature and identify the principle feature then we classify gait data and use different machine learning technique (ANN, SVM, KNN, and Classifier fusion) and performance comparison is shown. Experimental result on real time datasets propose method is better than previous method as far as humanoid locomotion classification is concerned.

Keywords: GAIT, Classification, Biometric Identification, Feature selection.

1 Introduction

Human gait is supposed as unique biometric identification like thumb print. So in GAIT analysis to examine the human walking moment we collect and analyze the data [1]. Human gait can be utilized to identify people for diverse security reason and exercised before to detect many abnormalities. For disable person the classified data is used to implement the prosthetic leg.

The GAIT data can be useful to help in walking of old people and is used in pattern analysis and classification of different walk. By seeing the abnormality in GAIT

cycle, it is worth to do research for prognostication for many upcoming diseases. Differently abled and elderly people can be helped by any research regarding GAIT. During human movement, top and bottom limb of human body yield the periodic motion. Due to this particular motion human beings get a particular motion and a unique walking pattern. The biometric identity like gender, age and race based on human locomotion can be obtained by the use of GAIT analysis shown in figure 1.

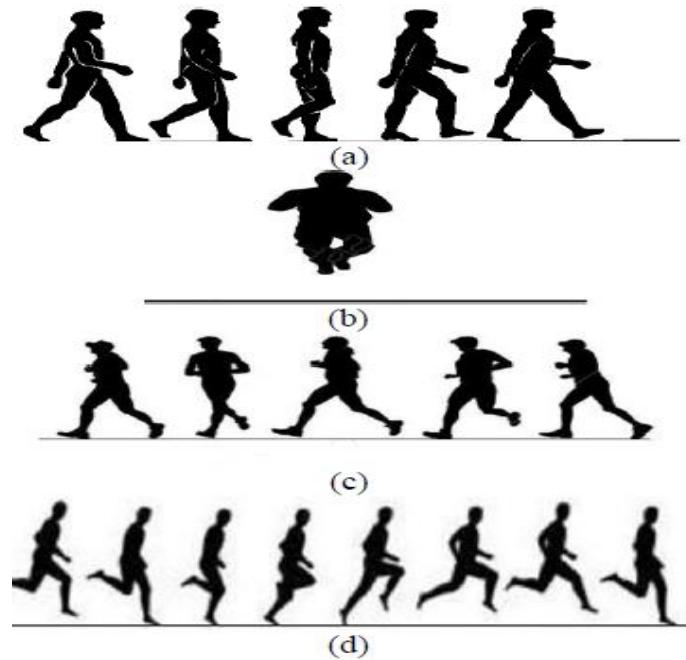


Fig. 1. Biomechanics of Locomotion: (a) Walk (b) Jump (c) Jog (d) Run

A lot of precursory research work about various machine learning techniques is gathered from the paper given [2-6] below. The classification of GAIT pattern is very accurate and unique due to rhythmic pattern. The new arena of biometric research is different category of GAIT classification. The GAIT can be helpful for security reason as it can identify from distance. In comparison with other biometric identification technique such as face reorganization i.e. iris, finger print which is considered as first generation biometric, the GAIT is divergent as it does not require any subject contact other than walking. This can be considered as behavioral authentication technique. Therefore identification using gait is becoming very popular in security. The development of highly sophisticated and more accurate humanoid can be made easy with the study of human GAIT data. So the foremost requirement is the feature reduction.

2 Proposed System

Proposed system consists of seven phases. They are collection of gait data, detection of gait data, trajectory smoothening, feature extraction, feature selection and classification. Fig. 2 tells systematic proposal. In the following subsections there are details of each phase.

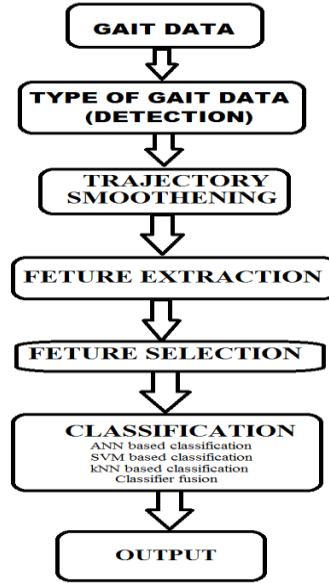


Fig. 2. Architecture of the proposed system

$$w_i = \frac{1}{\sum_{k \neq i} \|x_k - x_i\|_2}$$

According to the equation, if a feature point x_i is far from the other feature points then it has a small weight. Alternately, it is near to most of the other feature points then it will have a larger weight. After normalizing the weights as $\sum_i w_i = 1$, the centroid is calculated as:

$$x_c = \sum_i w_i x_i$$

The final step of compact criteria for a feature point x_i is expressed as $\|x_k - x_i\|_2 < r_o$, where r_o is a distance threshold. This criterion helps to eliminate the redundant feature points and ensures that the feature points are concentrated around the centroid.

3 Methodology

Firstly, we have collected data of different gait using accelerometer [8]. Secondly, features were selected on the basis of ANOVA and IFA (Incremental Feature

Analysis) to get optimal features from the existing features. Then different combinations of features are made by IFA using ANN, SVM, and KNN classifier. Lastly, classifier fusion technique is used.

3.1 Trajectory smoothening

After smoothening, we detect and remove the self co-articulated strokes. And as these strokes are not part of gesture therefore they need to be removed at this stage as they are just hand movements.

3.2 Feature Extraction

The features have been collected from [7].

Table.1. Crouch dataset description

Feature Category	Feature Name	Feature Category	Feature Name
F1	Pelvis_tx	F10	knee_angle_r
F2	Pelvis_ty	F11	ankle_angle_r
F3	Pelvis_tz	F12	hip_flexion_l
F4	Pelvis_tilt	F13	hip_adduction_l
F5	Pelvis_list	F14	hip_rotation_l
F6	Pelvis_rotation	F15	knee_angle_l
F7	hip_flexion_r	F16	ankle_angle_l
F8	hip_adduction_r	F17	lumbar_extension
F9	hip_rotation_r		

3.3 Feature Selection

For selecting the final optimal features here we are using 2-stage process. In first stage we perform the ANOVA test to obtain feature [9], [10], [11]. In second stage after the significant features were obtained from Stage 1, the IFS technique is used to determine the optimal features.

4 Classification

Classification is done using following machine learning techniques.

4.1 Classification based on ANN

The network is trained many times till root mean squared error reaches an optimum level. To adjust the network weights, we train the system through back-propagation algorithm. Results of ANN classification is shown in Table 2.

Table.2. ANN Results

Exp. No	Hidden units	Network structure	Iterations	Training accuracy (%)	Test accuracy (%)
1	50	33-50-40	100	84	80
			500	86	80
2	52	33-52-40	100	86	80
			500	90	82
3	54	33-54-40	100	86	82

4.2 Classification based on SVM

SVM is supervised learning method. We optimize class separation hyper plane so we get the maximum distance between hyper plane and pattern separating the classes. Now dataset is trained with different kernel functions such as polynomial, quadratic, linear and radial basis function. Results of SVM classification is shown in Table 3.

Table.3. SVM Results

Exp No.	Function Kernel	Training accuracy(%)	Test accuracy (%)
1.	Linear function	68	71
2.	Quadratic function	70	76
3.	Polynomial function	84	84
4.	Radial Basis Function	88	84

4.3 Classification based on KNN

KNN classifier is based on relating the anonymous instances to the known according to some distance function. To train *KNN* we use different *k* values as it is an essential parameter. In training phase we used different number of *k* values such as 3, 5, 7 and 9. Results of KNN classification is shown in Table 4.

Table.4. KNN Results

Exp No.	K	Training accuracy (%)	Testing accuracy (%)
1	3	74	70
2	5	79	80
3	7	78	74
4	9	84	81

4.4 Classifier fusion

Classifier fusion combines many classifiers and provides the result by combining the results of the individual classifiers. In this paper, we have combined the results of ANN, SVM and KNN models to achieve the classifier fusion result. The experiment of classifier fusion was done by combining the other classifiers. Thus, the performance of classifier fusion is likely superior than the individual classifiers used in isolation.

Once the training is done, testing is done by 5-fold cross validation shown in table 5.

Table.5. 5-fold cross validation result

Expt#	ANN classifier Accuracy (%)	SVM classifier Accuracy (%)	kNN with k=5 Accuracy (%)	Classifier fusion Accuracy (%)
1	89.67	91.62	80.76	89.67
2	92.91	91.92	78.89	90.75
3	87.82	88.52	91.82	87.82
4	93.60	85.60	90.40	89.60
5	91.01	89.21	92.21	86.75

5 Experimental results and discussions

5.1 Data Set Used

In order to evaluate our proposed approach of human activity recognition, we have used crouch datasets. The data set have four classes. Each has three different data base so total 12 dataset for all four classes. It has total 17 features (Times series data). The dimension of data is 17*300 for a particular data set, in which we choose one gait cycle for our experiment. So our dataset has 17*100 elements each. Hence total 17*4*3*100 element.

5.2 Performance of the classifiers

The train and test accuracies are evaluated for different parameters of ANN such as iterations and hidden units. The highest train and test accuracy was observed for network structure 40L54N40L (L, N corresponds to linear and nonlinear layers), iterations = 500. This network parameters were used for performing 5-fold cross validation to determine the performance and validity of the model. Results of cross validation process are shown in Table 6.

Table.6. Comparison of success rate by different classifier using 5-fold cross validation

Classifier	Rate of Success (%)
ANN	82.58
SVM	80.11
KNN	87.72
Classifier fusion	93.56

5.3 Comparative analysis

In our proposed system we perform comparative analysis by selecting the optimal features. Now the comparison is done with various set of gestures as seen in Fig. 3.

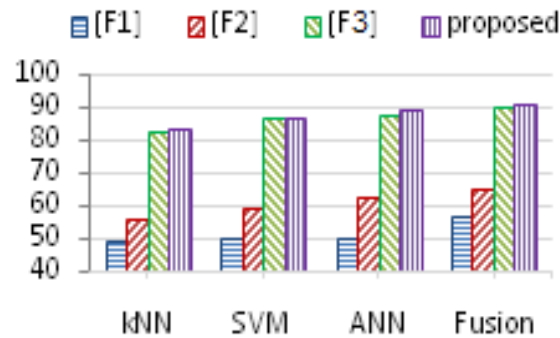


Fig. 3. Classification of Different GAIT pattern: comparison of proposed optimal features using ANN, SVM, KNN, and classifier fusion: (a) Crouch_2 (b) Crouch_3 (c) Crouch_4 (d) Normal [10]

It was noted that from all the other individual classifiers for various feature set performed the performance of the classifier fusion is best. Therefore we get the highest overall accuracy of 93.56% by using the combination of optimal features and classifier fusion as compared to other feature and classifier combinations.

5.4 Conclusion

The proposed system can be used for developing a gesture controlled hexadecimal keyboard making human-computer interaction easier. From the existing literatures, a total of 10 features were selected. To check the statistical significance of the 10 features ANOVA test was performed. The 10 significant features were then arranged in the descending order of the F-static value which was then fed to the IFS to select

the optimal features. Total number of features or the optimal features was observed to be 6, 7 and 8 for ANN, SVM and KNN respectively. The results of the three individual classifiers were combining to provide the classifier fusion results. After this, 5-fold cross validation was used to provide overall accuracy of the system. Overall accuracy was observed to be 82.58, 8.11%, 87.72% and 93.56% using the ANN, SVM, KNN and classifier fusion respectively.

References

1. Rattiya Thanasoontornerk, Chatkaew Pongmala, Areerat Suputtiada: Tree Induction for Diagnosis on Movement Disorders Using Gait Data. In: 2013 5th International Conference on Knowledge and smart technology (KST)
2. Kiichi_Katoh,, *Makoto Katoh*.: Using FS in Tuning a Biped Locomotion Surrogate Robot for Walking Training. In: 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)
3. Mitsuru Yoneyama, Yosuke Kurihara, Kajiro Watanabe,: Accelerometry-Based Gait Analysis and Its Application to Parkinson's Disease Assessment—Part 2: IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 21, NO. 6, NOVEMBER 2013
4. M. Ogata¹, T. Karato¹, T. Tsuji¹, H. Kobayashi¹ and K. Irie².: Development of Active Walker by using Hart Walker. In: SICE-ICASE International Joint Conference 2006 Oct. 18-21, 2006 in Bexco, Busan, Korea
5. Hiroshi Kobayashi, Takeo Karato, and Toshiaki Tsuji.: Development of an Active Walker as a New Orthosis. Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation August 5 - 8, 2007, Harbin, China
6. Daniel T.H. Lai, Ahsan Khandoker, Rezaul K. Begg and M.Palaniswami.: A hybrid Support Vector Machine and autoregressive model for detecting gait disorders in the elderly. In Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007.
7. Daniel T.H. Lai, Ahsan Khandoker, Rezaul K. Begg and M.Palaniswami: A hybrid Support Vector Machine and autoregressive model for detecting gait disorders in the elderly. Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007.
8. Anubha Parashar, Somya Goyal, Apoorva Parashar, Bharat Sahjlan: Push Recovery for Humanoid Robot in Dynamic Environment and Classifying the Data Using K-Mean. ICTCS 2016 - CSI-Udaipur Chapter.
9. Semwal, Vijay Bhaskar, et al : Biped model based on human Gait pattern parameters for sagittal plane movement. Control, Automation, Robotics and Embedded Systems (CARE), 2013 International Conference on. IEEE, 2013.
10. Semwal, Vijay Bhaskar, et al.: Biologically-inspired push recovery capable bipedal locomotion modeling through hybrid automata. Robotics and Autonomous Systems 70 (2015): 181-190.
11. Semwal, Vijay Bhaskar, Manish Raj, and Gora Chand Nandi. :Biometric gait identification based on a multilayer perceptron. Robotics and Autonomous Systems 65 (2015): 65-75.