Clustering Gait Data using different machine learning techniques and finding the best technique

Anubha Parashar[1]; Deepak Goyal[2]

Vaish College of Engineering, Rohtak[12];
{anubhaparashar1025[1]; deepakgoyal.vce[2]}@gmail.com

Abstract. Clustering is done in order to group the entities which are alike in one group so that grouping of more similar objects can be done. The objects placed in one group are known as clusters. In this paper we are using clustering in order to identify the human locomotion and categories the dataset making clusters. We are using two clustering techniques i.e. SOM and K-mean. So we first selected the feature and identify the principle feature then we cluster gait data and use different machine learning technique (K-mean and SOM) 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: Clustering, Biometric Identification, Feature selection, SOM, K-Mean.

1 Introduction

In order to authenticate humans we use various verification and validation technologies. To know whether a person is authenticated or not we need verification. To authenticate some person and to check his/her identity various authentications techniques are used. First is *knowledge*; in this technique a person must have particular knowledge about of his PIN, password, or any other secret code. Second is *Possession;* in this technique a person must know particular token or item like identity card, smart card. Third is *Biometric*; in this technique a person's body part is taken to identify unique identity of person like fingerprint, retina, iris, facial expression. But there are various disadvantages of above techniques [1] [2] [3]. A person may forget its password or PIN. A person may lose its identity card or smart card. Biometric authentication may not be always user friendly and is expensive too.

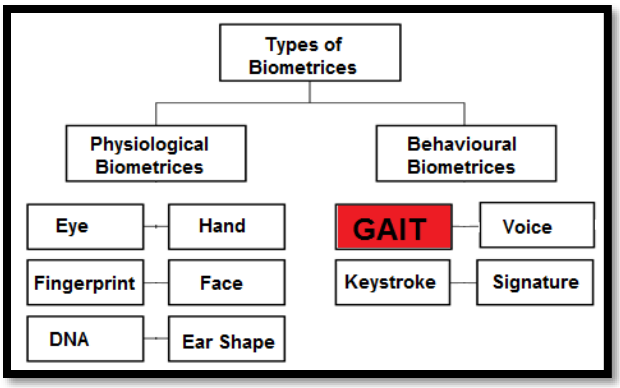

Fig. 1. Biometrics types

In contrast with such authentication techniques GAIT based biometric method outperforms all above disadvantages.

Gait means how a human move/walk and is very useful to identify person from a distance. It is very useful in detecting diseases and in surveillance system. In this paper we have taken four data sets namely croubh2, croubh3, croubh4, normal shown in figure 2.

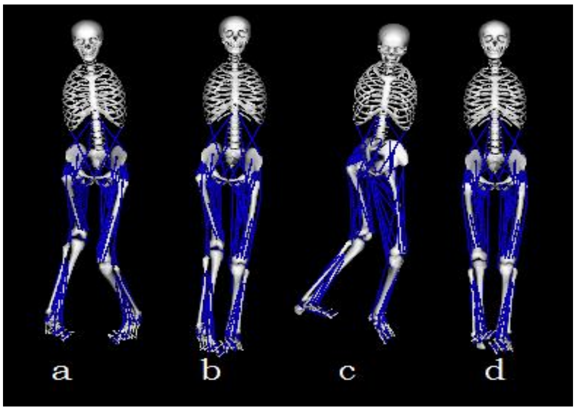

Fig. 2. Contrast enhanced images in gait sequences of Gait Database (Dataset - a (croubh2) b (croubh3) c (croubh4) d (normal)). Four samples from left to right show the gait of pathological walking (a,b,c) and normal walking (d).

2 Proposed System

Proposed system consists of eight phases. They are collection of Data Preprocessing, Gait Cycle Detection, Feature Extraction, Feature Reduction (PCA), Output Feature Vector, Clustering (K-Mean) (SOM), and output.

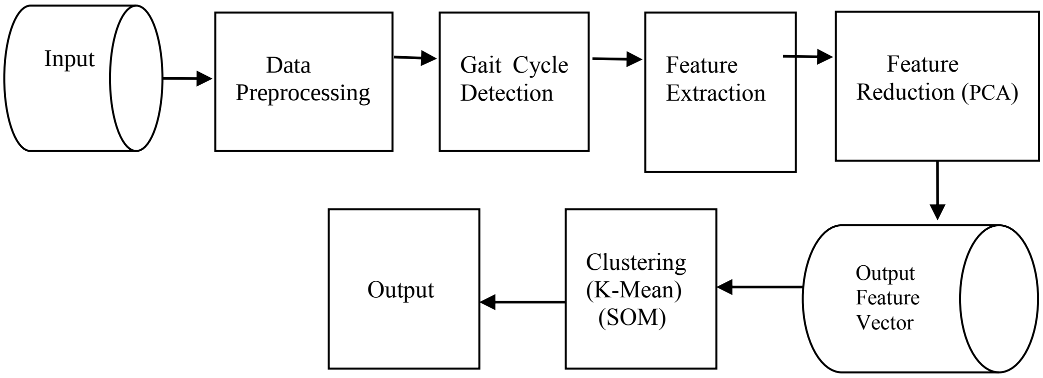

Fig. 3. Proposed System

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

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 xi 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 HMDC [8], and then features were selected, then different combinations of features are made using K-mean and SOM clustering. Then 10 fold cross-validation test also conducted to validate the statistical significance of the results.

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

3.3 Feature Selection

Here 16 features are extracted. The features are collected from different joints while locomotion [1]. In all the four datasets 16 features are extracted which is listed in table 1.

4 Clustering

Clustering is done using following machine learning techniques.

4.1 K-mean

The k-mean algorithm is generally used for compact clustering. It is very much sensitive for outlier and noise. It is applicable only for numerical data. Results of K-Mean is shown in Table 2 and Overall clustering using different no of majority vote is shown in figure 4.

Table.2. K-Mean Results

Exp. No	Training accuracy (%)	Test accuracy (%)
1	100	80

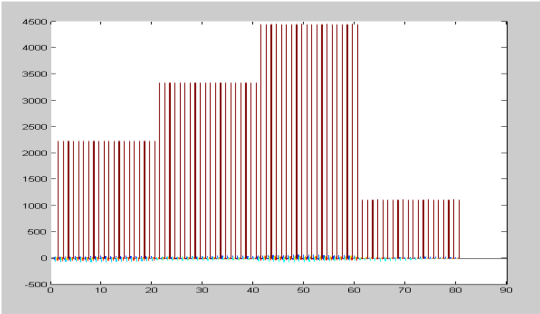

Fig. 4. Overall clustering using different no of majority vote

4.2 Clustering based on SOM

SOM is unsupervised learning method. SOM is used to map the training samples and thus is known as self-organising maps. Results of SOM classification is shown in Table 3.

Figure 5 shows the number of SOM layer and inputs and outputs. Figure 6 shows SOM neighbouring distances. Figure 7 shows Biometrics types and figure 8 shows SOM weight planes.

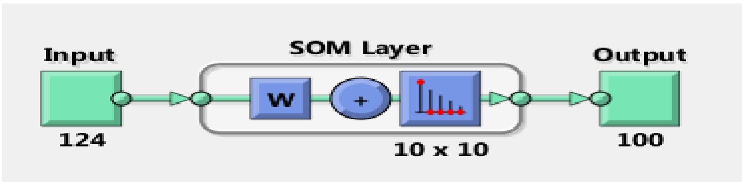

Fig. 5. SOM network architecture

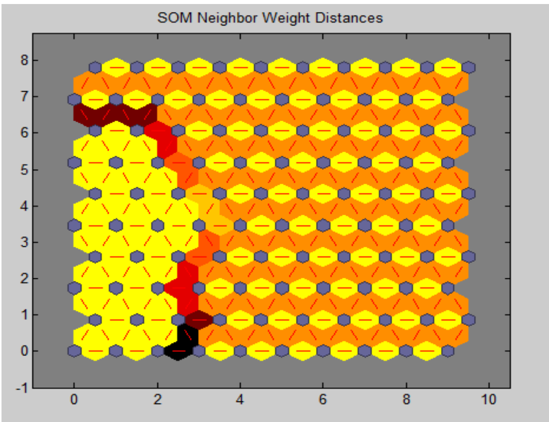

Fig. 6. SOM neighbouring distances

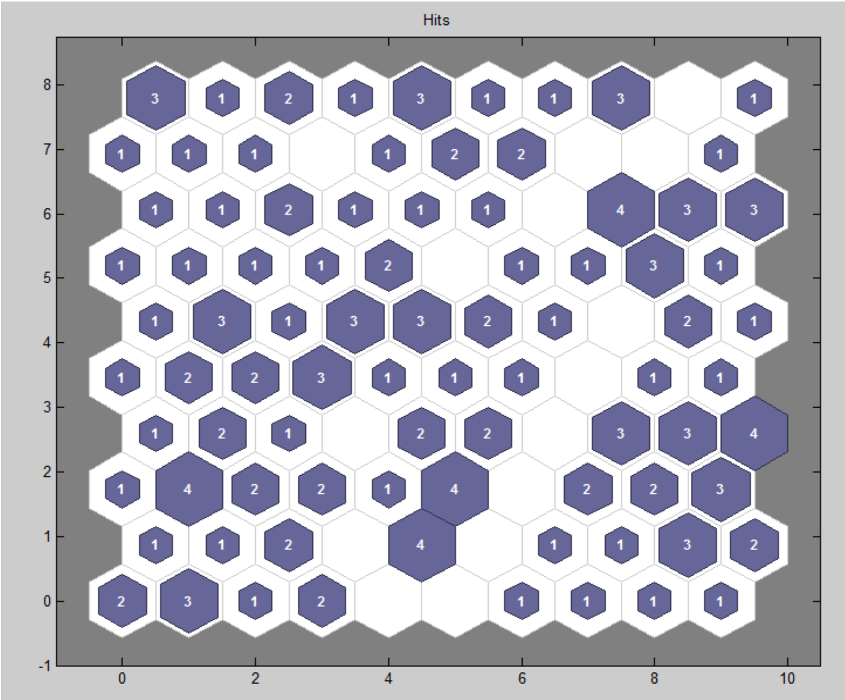

Fig. 7. Biometrics types (in this paper GAIT biometric clustering is used)

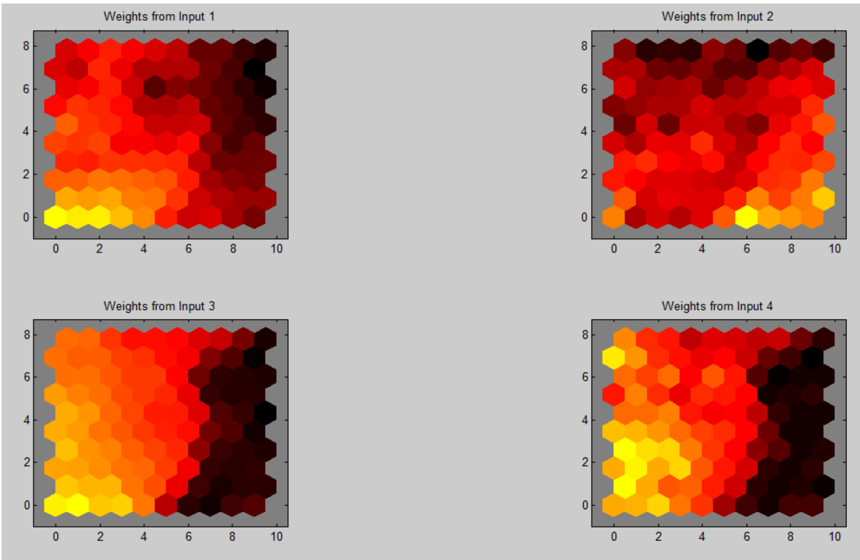

Fig. 8. SOM weight Planes

Table.3. SVM Results

Exp No.	Training accuracy (%)	Test accuracy (%)
1.	100	71

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 16 features (Times series data). The dimension of data is 16*300 for a particular data set, in which we choose one gait cycle for our experiment. So our dataset has 16*100 elements each. Hence total 16*4*3*100 element.

5.2 Performance of the classifiers

10-fold cross validation is used to determine the performance and validity of the model. Results of cross validation process are shown in Table 6.

Table.4. Comparison of success rate by different classifier using 5–fold cross validation

Classifier	Rate of Success (%)
K-Mean	80
SOM	71

5.3 Validation

Correlation coefficient	0.9955
Mean absolute error	8.2394
Root mean squared error	117.3938
Relative absolute error	0.741 %
Root relative squared error	9.4381 %
Total Number of Instances	203

It is noted that from all the other individual classifiers for various feature set performed the performance of the K-Mean is best. Therefore we get the highest overall accuracy of 90%.

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 K-Mean and SOM respectively. After this, 10-fold cross validation was used to provide overall accuracy of the system. Overall accuracy was observed to be 80, 71.1%, using the K-Mean and SOM 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 Knowlwdge 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.
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 Sahjalan: 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.