**Noname manuscript No.**
(will be inserted by the editor)

# Self-Organizing Maps and Fuzzy C-means Algorithms on Gait Analysis Based on Inertial Sensors Data

**Rafael Caldas · Yabing Hu · Fernando Buarque de Lima Neto · Bernd Markert**

the date of receipt and acceptance should be inserted later

**Abstract** Human gait corresponds to the physiological way of locomotion, which can be affected by several injuries. Thus, gait analysis plays an important role in observing kinematic and kinetic parameters of the joints involved with such movement pattern. Due to the complexity of such analysis, this paper explores the performance of two adaptive methods, Fuzzy c-means (FCM) and Self-organizing maps (SOM), to simplify the interpretation of gait data, provided by a secondary dataset of 90 subjects, subdivided into six groups. Based on inertial measurement units (IMU) data, two kinematic features, average cycle time and cadence, were used as inputs to the adaptive algorithms. Considering the similarities among the subjects of such database, our experiments show that FCM presented a better performance than SOM. Despite the misplacement of subjects into unexpected clusters, this outcome implies that FCM is rather sensitive to slight differences in gait analysis. Nonetheless, further trials with the aforementioned methods are necessary, since more gait parameters and a greater sample could reveal an undercover variation within the proper walking pattern.

**Keywords** Computational Intelligence · Self-Organizing Maps Algorithm · Fuzzy Logic · Gait Analysis · Inertial Measurement Unit

## 1 Introduction

Human gait, or walking, refers to a person's way of locomotion [6]. This movement pattern has a cyclical nature with sequential movements of both legs and is described as a series of states with transitions, being divided into stance and swing phases [16]. Such pattern is usually affected by many injuries, thus an efficient gait assessment plays an important role in rehabilitation.

Institute of General Mechanics, RWTH Aachen University, Aachen, NRW, 52062, Germany.
E-mail: rafael.caldas@iam.rwth-aachen.de

Gait analysis is the systematic study of human walking pattern [9]. Such analysis involves the assessment of several biomechanical features, which allows physicians to identify causes of various injuries, enabling the observation of treatment effectiveness [1, 15, 19]. Conventional gait analysis uses laboratory-based systems and questionnaires, but these methods present several disadvantages [24]. The optoelectronic system, which corresponds to the gold standard method, is rather complicated and expensive, requiring a large space to perform the assessment [18]. On the other hand, the results obtained with specific questionnaires lack sufficient accuracy, implying a subjective conclusion [5].

In this context, over the past years, inertial measurement units (IMU) have been widely used in gait analysis, because of their easy application and low cost [8, 11]. An IMU is an electronic device that consists of accelerometers for linear acceleration and gyroscopes for angular velocity measurements in three axes [20]. IMU could dynamically acquire the signals of linear acceleration and angular velocity during the whole walking process and those quantitative signals clearly reflect motion information of human gait [4].

After acquisition and preprocessing, the sensor data is still represented in a complex way, that might difficult healthcare professionals' interpretation of the results. Considering this, adaptive algorithms, such as Self-organizing maps (SOM) and Fuzzy c-means (FCM), are efficient methods for classifying data. SOM is particularly useful for analyzing data due to its ability of projecting high-dimensional data into a simpler low-dimensional representation [13]. On this nonlinear projection, each individual is mapped to one point, where the distances among the subjects reflect their similarities [21].

Additionally, Fuzzy logic is considered an advanced method for identifying features of periodic signals [22]. Such approach uses approximate rather than exact measures, represented by the membership function. This function denotes the degree of membership, characterizing a set that extends continuously from 0 to 1 [7]. As an adaptive tool, one can use Fuzzy logic for data analysis, considering its likely smooth and continuous detection of gait phases [14]. This paper is organized so that the performance of FCM and SOM algorithms through clustering of human gait data acquired by inertial sensors can be established. Section 2 presents the methods used for data acquisition and processing. In section 3, the results are commented upon.

## 2 Methods

*Data Acquisition.* In this study, we used the OU-ISIR gait database, which was established to assist research efforts in general area of developing and testing algorithms for human identification based on gait data [17]. This dataset uses four sensors: (i) 3 IMUs, each one includes a triaxial accelerometer and a triaxial gyroscope; (ii) a triaxial KXTF9 accelerometer embedded in a smartphone ME860. They set the IMU dynamic ranges at $\pm$4g for accelerometers and $\pm$500 deg/s for gyroscopes to capture human gait signal, all the four sensors worked at 100 Hz.

Sensors were attached to a belt, pointing their x-, y- and z-axes to subject's right, upwards and backward, respectively. The centered IMU and smartphone are located at the center back waist of subjects; the left and right IMU are located at the left and right waist. During flat ground walking, 6D signal sequences were captured with each IMU. The subjects were divided into 6 groups according to their ages: Group A includes subjects under 10 years old; Group B, 10-19 years old; Group C, 20-29 years old; Group D, 30-39 years old; Group E, 40-49 years old and Group F, over 50 years old.

*Gait Temporal Features Extraction.* Based on previous works and preliminary experiments, Average cycle time (ACT) and Cadence were the kinematic features extracted [3]. The number of cycles or steps per second represents the cadence, which indicates an relevant parameter regarding the subjects' performance [10]. The minimum value of each valley was set as the start of a cycle as well as the end of the previous cycle. Thus, the average cycle time is easily obtained, providing possibly information on the symmetry of subject.

*Data Processing.* Fuzzy c-means is a clustering method that allows each data point to belong to multiple clusters with varying degrees of membership. FCM is based on the minimization of the following objective function [2]:

$$J_m = \sum_{i=1}^{D} \sum_{j=1}^{N} \mu_{ij}^m \left\| x_i - c_j \right\|^2 \tag{1}$$

herein, $D$ is the number of data points; $N$ is the number of clusters; $m$ is the fuzzy partition matrix exponent for controlling the degree of fuzzy overlap, with $m > 1$. This overlap refers to the boundaries between clusters and denotes the number of data points that have significant membership in more than one cluster; $x_i$ is the *ith* data point; $c_j$ is the center of the *jth* cluster; $\mu_{ij}$ is the degree of membership of $x_i$ in the *jth* cluster. For a given data point, $x_i$, the sum of the membership values for all clusters is 1.

FCM performs the following steps during clustering: (i) Randomly initialize the cluster membership values, $\mu_{ij}$; (ii) Calculate the cluster centers, as shown in formula (2); (iii) Update $\mu_{ij}$ according to the following, as shown in formula (3); (iv) Calculate the objective function, $J_m$.; (v) Repeat steps (ii)-(iv) until $J_m$ improves by less than a specified minimum threshold or until after a specified maximum number of iterations:

$$c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^m x_i}{\sum_{i=1}^{D} \mu_{ij}^m} \tag{2}$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{N} \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right) \frac{2}{m-1}} \tag{3}$$

Self-Organizing Maps (SOM) classify input vectors according to the pairing in the input space. At the first level, a 2D feature map is formed, where the number of output neurons is significantly larger than the desired number of clusters. Then, at the second level the output neurons, representing the subjects, are clustered such that the neurons on the map are divided into as many different regions as the desired number of clusters. During the training process, each input data point can be assigned into a cluster according to the best match unit (BMU), or its nearest output neuron [23].

In the original proposition of SOM, the Euclidean distance corresponds to the similarity criterion and the winner neuron is the one with the smallest distance [12]. Considering this process, the self-organization aims to minimize the distance, adjusting the BMU weights and its neighborhood toward the input vector using the following rule:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)h_{ci}(t)[x(t) - w_{ij}(t)] \tag{4}$$

herein, $w_{ij}$ is the weight $j$ of neuron $i$, $\alpha(t)$ is the learning rate and $h_{ci}(t)$ is the neighborhood radius function centered on winner neuron $c$ with respect to neuron $i$. Typically, both learning rate and neighborhood radius are functions, which decrease with time, in order to facilitate system convergence.

*Experiments Setup.* Initially, we set the features of different axes as inputs to the FCM algorithm, intending to explore the potential differences of the results between them after clustering. In the first experiment, two groups, A and D, were chosen. Each group comprises the kinematic features (ACT and cadence) of 30 randomly selected subjects, acquired by the three IMUs. We disregard the smartphone information because it is redundant data due to the same position of the central sensor. In the second part of FCM experiments, 15 subjects of each of the six groups (A-F) were randomly selected and their kinematic features were used as inputs to the system.

Regarding the SOM, we intended to explore the influence on changing inherent variables of the algorithm, applied to the same dataset utilized in FCM trials. Based on a previous work from our group, the y-axis was also chosen as input for the SOM [3]. In the first experiment, Ordering-phase (OP) steps or the number of iterations were set as 500, 1000, and 2000. While the subsequent parameters were fixed as following: OP learning rate = 0.9; Tuning phase (TP) learning rate = 0.02; TP Neighborhood distance = 1; and Map size = 20x20. In the second experiment, the OP steps were fixed in 500 iterations and the TP Neighborhood distances were set as 1, 10 and 20, while the other variables were not modified.

## 3 Results and Discussion

The performed experiments revealed the behavior of the two selected methods on grouping subjects due to their kinematic parameters. Fig. 1(a), (b) and (c) show the results of Fuzzy c-means according to the inputs of x, y and z axes, respectively. The subjects are grouped according to their similarities on the graph, where the solid lines connecting the samples represent the distance between them. The clusters are clearly separated and there is no relevant difference between these three cases.

As shown in Fig. 1(d), FCM outcomes, when using all axes as input, present only one subject from Group A clustered on the other group. This might indicate a small difference between such subject and Group D. Nonetheless, such graph does not present significant difference when compared to the outputs of previous experiments, considering the placement of most subjects of the Group A at left lower part of the graph. The application of more than one axis, thus, does not seem to be relevant to the algorithm. Thereupon, we used features of only one axis as input in the following experiments, conveniently y-axis was chosen.

In the second experiment, 15 subjects were randomly chosen from each of the 6 groups (A-F)and grouped into 6 clusters (Fig. 2). Since some of the clusters do not present 15 elements, the outcome seems not to corroborate with our hypothesis that the algorithm could differentiate the subjects, despite their slightly different kinematic features. For instance, only 4 individuals form clus-

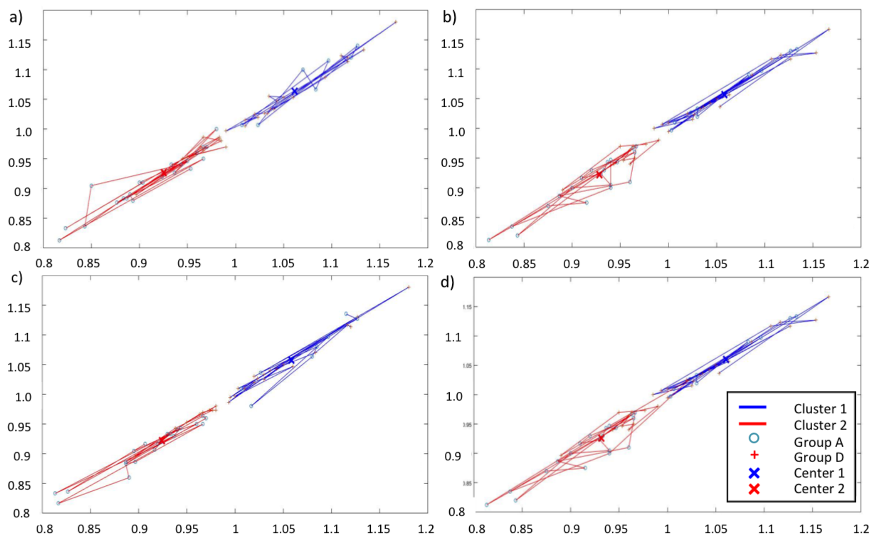


**Fig. 1** Fuzzy c-means clustering of Groups A (circle) and D (plus): (a) Features of x-axis as inputs; (b) Features of y-axis as inputs; (c) Features of z-axis as inputs; (d) Features of the three combined axes as inputs.

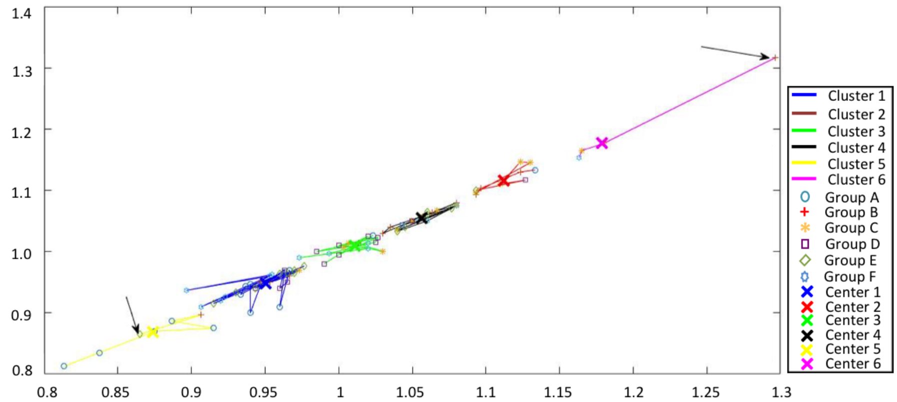


**Fig. 2** FCM clustering of Group A to F, kinematic features of y-axis set as inputs. Black arrows indicate two individuals commented upon later.

ter 6, what implies an unusual gait pattern of the misplaced individuals. In this case, either such subjects presented an unexpected performance to their age group or FCM was not sensitive enough to recognize the differences. However, we could not ensure the reason since there is no further anthropometric information, except for their age groups.

Based on the clustering outcomes, we assume that the pointed out subjects might present an abnormal gait pattern (Fig 2). The left lower one, which belongs to Group E, lies among the Group A, while the upper right subject, from Group B, was placed distant from the other subjects within cluster 6. Aside these two highlighted individuals, the subjects were generally clustered as expected, the younger subjects from Group A on the left part of the graph, in contrast to the older ones placed on the top right part of the graph.

Considering the experiments with SOM, 15 subjects were selected for each age group (A-F) and kinematic features acquired by y-axis set as input. The graphs in Fig. 3 represent SOM neighborhood weight distances, which indicate the distances between neighboring neurons. The blue hexagons represent such neurons with red lines connecting neighboring ones. The colors in the areas containing the red lines indicate the distances between neurons. For instance, darker and lighter colors represent larger and smaller distances, respectively. By varying the number of iterations, one can notice that the algorithm groups the sample only on small areas of the top corners. Even though, these clusters include few subjects, while most of the subjects are not clustered very clearly. The algorithm spread them on the map even with the increased number of iterations (Fig 3.c).

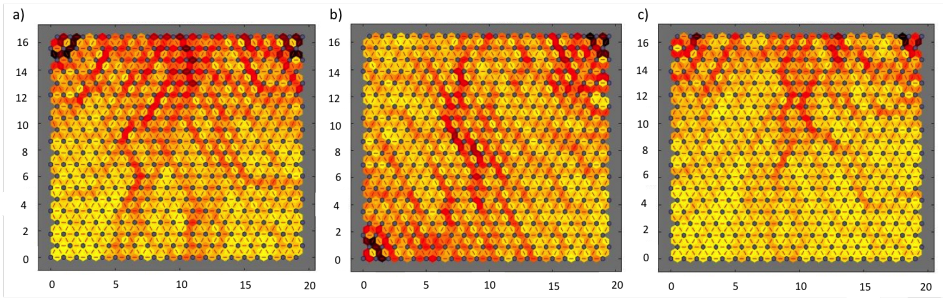


**Fig. 3** SOM Neighbor weight distance of (a) 500, (b) 1000, and (c) 2000 Ordering-phase steps or iterations.

As shown in Fig. 4, the cluster becomes clearer and more compact with the TP neighborhood distance increasing, indicated by the dark red lines. The representation of sample hits of different TP neighborhood distance provides information about data points that are associated with each neuron (Fig. 5), representing the placement of subjects due to their own gait features. These graphs show straightforward two main clusters in the upper-left and lower-right corner, where subjects with similar features are placed. Despite individuals in the middle, who have less similar features, were mapped apart from the clusters. As the neighborhood distance increases, these points in the middle are placed closer to the two main clusters (Fig 5.b).

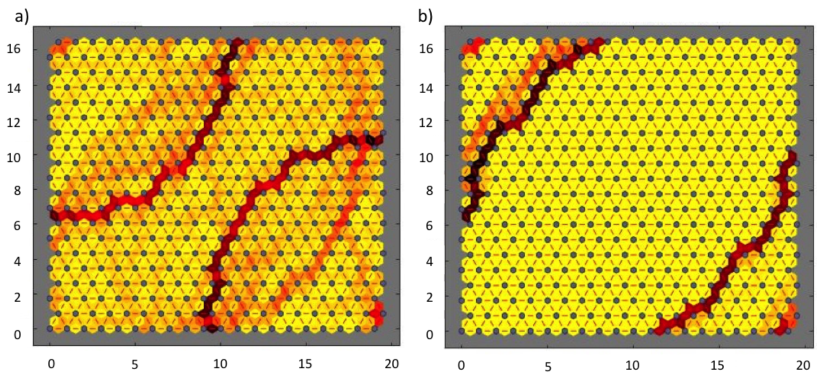


**Fig. 4** SOM Neighbor weight distance of tuning phase - Neighborhood distances (a) 10 and (b) 20.

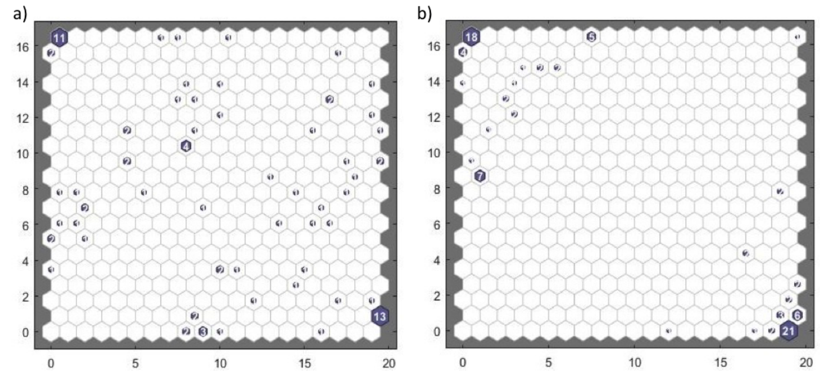


**Fig. 5** SOM Sample hits of tuning phase - Neighborhood distances (a)10 and (b) 20.

Previous research has proved that SOM could group gait data in the comparison of proper and pathological patterns [3]. FCM, nevertheless, seems to perform better when facing slight variations of healthy persons' gait pattern, inherent to different age groups. Since such algorithm allows a subject to belong to different clusters, what might be the case of healthy subjects with similar gait features.

## 4 Conclusions

In the current work, we processed human gait data acquired by inertial sensors with two adaptive algorithms, Fuzzy c-means and Self-organizing maps. The results pointed out that FCM outperformed SOM in the task of grouping healthy subjects, divided into age-related groups. Subjects in this database were healthy, implying small variations in their gait patterns, which could not be easily distinguished. In this way, Fuzzy c-means presented better results, considering that a subject could belong to several clusters. SOM, on the other hand, has the advantage of not requiring the number of clusters in advance, interesting for when no further information about the dataset is available.

Therefore, further trials with the aforementioned methods are necessary, since more gait parameters and a greater sample could reveal an undercover variation within the proper walking pattern. Thus, we intend to obtain more walking features data, provided by a system with more IMUs, enabling a global gait analysis. Moreover, the sensors might provide less information about such pattern when placed on the waist than on the lower limbs.

**Acknowledgement**

This work was supported by the Brazilian National Council for Scientific and Technological Development (CNPQ 207445/2014-1) through a doctoral scholarship granted to R.C.

## References

1. Alaqtash M, Yu H, Brower R, Abdelgawad A, Sarkodie-Gyan T (2011) Application of wearable sensors for human gait analysis using fuzzy computational algorithm. Engineering Applications of Artificial Intelligence 24(6):1018–1025
2. Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell, MA, USA
3. Caldas R, Rativa D, Lima Neto FB (2016) An application of self-organizing maps clustering to support kinematic gait analysis and evaluation of gait symmetry. Journal of Biomedical and Health Informatics (Submitted for publication)
4. Chen Y, Hu W, Yang Y, Hou J, Wang Z (2014) A method to calibrate installation orientation errors of inertial sensors for gait analysis. In: Information and Automation (ICIA), 2014 IEEE International Conference on, pp 598–603
5. Clarke JE, Eccleston C (2009) Assessing the quality of walking in adults with chronic pain: The development and preliminary psychometric evaluation of the bath assessment of walking inventory. European Journal of Pain 13(3):305–311
6. DeCann B, Ross A, Culp M (2014) On clustering human gait patterns. In: 22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24-28, 2014, pp 1794–1799
7. Dote Y (1995) Introduction to fuzzy logic. In: Industrial Electronics, Control, and Instrumentation, 1995., Proceedings of the 1995 IEEE IECON 21st International Conference on, vol 1, pp 50–56
8. Godfrey A, Conway R, Meagher D, OLaighin G (2008) Direct measurement of human movement by accelerometry. Medical Engineering & Physics 30(10):1364–1386
9. Guha AR, Perera AM (2012) Calcaneal osteotomy in the treatment of adult acquired flatfoot deformity. Foot and Ankle Clinics 17(2):247 – 258
10. Hussain S, Xie SQ, Jamwal PK (2013) Effect of cadence regulation on muscle activation patterns during robot-assisted gait: A dynamic simulation study. IEEE Journal of Biomedical and Health Informatics 17(2):442–451
11. Kavanagh JJ, Morrison S, James DA, Barrett R (2006) Reliability of segmental accelerations measured using a new wireless gait analysis system. Journal of Biomechanics 39(15):2863–2872
12. Kohonen T (1990) The self-organizing map. Proceedings of the IEEE 78(9):1464–1480, DOI http://dx.doi.org/10.1109/5.58325

13. Kohonen T, Schroeder MR, Huang TS (eds) (2001) Self-Organizing Maps, 3rd edn. Springer-Verlag New York, Inc., Secaucus, NJ, USA
14. Kong K, Tomizuka M (2008) Smooth and continuous human gait phase detection based on foot pressure patterns. In: Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on, pp 3678–3683
15. Kse A, Cereatti A, Della Croce U (2012) Bilateral step length estimation using a single inertial measurement unit attached to the pelvis. Journal of NeuroEngineering and Rehabilitation 9(9)
16. Ma H, Liao W (2014) Human level walking gait modeling and analysis based on semi-markov process. In: 2014 IEEE International Conference on Robotics and Automation, ICRA 2014, Hong Kong, China, May 31 - June 7, 2014, pp 240–245
17. Ngo TT, Makihara Y, Nagahara H, Mukaigawa Y, Yagi Y (2014) The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. Pattern Recognition 47(1):228–237
18. Pohl MB, Patel C, Wiley JP, Ferber R (2013) Gait biomechanics and hip muscular strength in patients with patellofemoral osteoarthritis. Gait & Posture 37(3):440 – 444
19. Senanayake C, Senanayake SM (2010) A computational method for reliable gait event detection and abnormality detection for feedback in rehabilitation. Computer Methods in Biomechanics and Biomedical Engineering 14(10)
20. Tessendorf B, Gravenhorst F, Arnrich B, Tröster G (2011) An IMU-based sensor network to continuously monitor rowing technique on the water. In: Proceedings of the Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2011), IEEE press
21. Vesanto J (1999) Som-based data visualization methods. Intelligent Data Analysis 3(2):111–126
22. Vicerra RRP, David KKA, dela Cruz AR, Roxas EA, Simbulan KBC, Bandala AA, Dadios EP (2015) A multiple level mimo fuzzy logic based intelligence for multiple agent cooperative robot system. In: TENCON 2015 - 2015 IEEE Region 10 Conference, pp 1–7
23. Wu S, Chow TWS (2004) Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density. Pattern Recognition 37(2):175–188
24. Yang M, Zheng H, Wang H, McClean S, Hall J, Harris N (2012) A machine learning approach to assessing gait patterns for complex regional pain syndrome. Medical Engineering & Physics 34:740–746