Objective Graphical Clustering of Spatiotemporal Gait Pattern in Patients with Parkinsonism

Conference Paper *in* AIP Conference Proceedings · September 2018

DOI: 10.1063/5.0026489

7 authors, including:

Flora Ferreira
University of Minho
28 PUBLICATIONS **99** CITATIONS
SEE PROFILE

Miguel Fernandes Gago
University of Minho
42 PUBLICATIONS **431** CITATIONS
SEE PROFILE

Nafiseh Mollaei
Universidade NOVA de Lisboa
13 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Estela Bicho
University of Minho
127 PUBLICATIONS **1,498** CITATIONS
SEE PROFILE

Objective Graphical Clustering of Spatiotemporal Gait Pattern in Patients with Parkinsonism

Flora Ferreira[1,a)], Miguel Gago[2,3], Nafiseh Mollaei[4], Estela Bicho[4], Nuno Sousa[3], João Gama[1], Carlos Ferreira[1]

[1]*LIAAD, INESC TEC, Portugal*

[2]*Serviço de Neurologia, Hospital da Senhora da Oliveira, Portugal*

[3]*ICVS, Escola de Medicina, Universidade do Minho, Portugal*

[4]*Centro Algoritmi, Escola de Engenharia, Universidade do Minho, Portugal*

[a)]*Corresponding author: flora.ferreira@gmail.com*

Abstract. The goal of this study was grouping patients with parkinsonism that share similar gait characteristics based on principal component analysis (PCA). Spatiotemporal gait data during self-selected walking were obtained from 15 patients with Vascular Parkinsonism, 15 patients with Idiopathic Parkinson's Disease and 15 Controls. PCA was used to reduce the dimensionality of 12 gait characteristics for the 45 subjects. Fuzzy C-mean cluster analysis was performed plotting the first two principal components, which accounted for 84.1% of the total variability. Results indicates that it is possible to quantitatively differentiate different gait types in patients with parkinsonism using PCA. Objective graphical classification of gait patterns could assist in clinical evaluation as well as aid treatment planning.

INTRODUCTION

Gait assessment analysis is an emerging tool in clinical practice due to its potential to provide objective assessment of gait impairment [1]. Recent advances in wearable sensors have opened up a promising future for gait analysis. Although different numerous kinematic variables can be extracted from raw kinematic data, and large set of gait characteristics can be registered. Their clinical relevance is still an open controversial field [1]. With the aim to improve gait interpretation, reducing the dimensionality of gait characteristics has been addressed by multivariate statistical approaches, such as factorial analysis and principal component analysis (PCA) [2, 3]. PCA is useful to reduce the number of gait characteristics generating a new set of independent variables, the principal components (PCs), which are weighted combinations of the original variables, retaining most of the original data variability [4]. The analysis of the PCs constitutes a valued mean to interpret the most important relationships between the variables as well as the contribution of each variable to the variability in the data. PCA is considered to be a good precursor for cluster analysis [5]. Different statistical clustering techniques to classify gait types such as K-mean, fuzzy C-mean (FCM), hierarchical clustering, Self-Organizing Map based on gait analysis have been applied [6]. Carriero et al. [3] explored PCA and FCM clustering approach to develop a graphical classification method of cerebral palsy gait pattern. PCA has been applied on 26 selected kinematics variables and age from 20 healthy and 20 cerebral palsy children and FCM clustering was performed plotting the first three principal components. In classical (hard) clustering such as K-mean each data object is assigned to only one cluster and clusters do not overlap. However, an overlapping approach enable a more flexible and realistic interpretation. FCM algorithm allows the generation of partitions such that each object can belong to more than one cluster, and associated with each of the objects are membership grades which indicate the degree to which the data objects belong to the different clusters [7].

Previous studies have shown the feasibility of various statistic techniques for differentiation of gait pattern. However, the lack of a standardized method of gait differentiation, based on spatiotemporal gait variables, is still a critical issue relating to management of gait pathology in parkinsonism. In this study, the aim was to generate an objective graphical clustering that identifies subgroups of patients with parkinsonism with similar gait characteristics. Spatiotemporal variables were acquired using wearable sensors and PCA has been applied on 12 selected gait characteristics from 15 patients with Vascular Parkinsonism (VaP), 15 patients with Idiopathic Parkinson's Disease (IPD) and 15 Controls. The first few principal components with the fuzzy C-mean clustering technique was used to generate a graphical identification of different prototypes or clusters of patients with parkinsonism, according to their gait characteristics.

METHODS

Fifteen patients with VaP (6 females, median age of 79.5 years-old [73, 90]), 15 patients with IPD (3 females, median age of 78 years-old [67,83]) and 15 healthy subjects (10 females, median age of 77 years-old [65, 85]) were included. For all patients the exclusion criteria were: the presence of resting tremor, moderate-severe dementia (CDR>2), musculoskeletal disease and overt clinical progression since diagnosis, Hoehn-Yahr>3.

The participants were asked to walk a 60-meter continuous course (30 meters' corridor with one turn) in a self-selected walking speed. Two Physilog® sensors (Gait Up®, Switzerland) positioned on the two foots were used to measure the spatiotemporal variables of each stride (also known as gait cycle, that begins when the reference foot contacts the ground

and ends when the same foot contacts the ground again). To evaluate only straight walking without variations due to turns and initiation and termination of gait, two strides for initiation and termination, and the strides associated to the turn at the end of the corridor were discarded. The spatiotemporal variables considered in this study are stride time (gait cycle duration), cadence (number of steps in a minute), swing time (the time during which the foot is in the air), stride velocity (speed of one cycle), stride length (distance between successive initial contact of the same foot), % GC spent in double support (percent of gait cycle where both feet touch the ground), lift-off angle (angle between the foot and the ground when the toe leaving the ground, on a vertical plane), and strike angle (angle between the foot and the ground when the heel hits the ground, on a vertical plane). The arithmetic mean and the coefficient of variation (CV) (ratio of standard deviation to the mean) of each gait variable were calculated for all subjects' stride time series. The gait variables measurement is available for both feet. However, to preserve consistency, we refrained the results to the right side.

PCA was applied to a n x p matrix, where n = 45 is the number of participants and p=12 is the number of the gait characteristics, with prior normalization [4]. Taking into account the Kaiser's criterion, the PCs with corresponding eigenvalue greater than 1 were retained. Then, only the first few PCs that retained the most variation of data were used for further analysis. The interpretation of the PCs in this work was accomplished by examining the loading values, and only the variables with larger loading, in absolute value, were considered.

FCM cluster analysis was then performed on the first few PCs in order to classify the entire data set into groups with similar gait characteristics. FCM algorithm is based on the minimization of an objective function called c-means function that depends of the parameters *m* and *c*: the fuzzifier and the number of clusters. The cluster fuzziness parameter (or weighting exponent) $m>1$ defines the level of cluster fuzziness. While higher values of *m* result with the more fuzzy cluster, when *m* approaches to 1 give harder clusters. The optimal choice for the value of the fuzzifier is a debated matter. It has been suggested that the set [1.1,5] is a proper range of m, where $m=2$ is usually chosen [8]. The investigation carried in [8] yields that the optimal range value is [2.5, 3]. This result is in line with other works which demonstrate that larger values of *m* provide more robustness against noise and the outliers. In the present study $m=2.5$ is used. The optimal number of clusters, c, was obtained by different validation indices proposed in the literature [9].

RESULTS AND DISCUSSION

Only the first two PCs accounted 81.1% of the total variability were retained (Table 1), effectively achieving dimensionality reduction of the data. The first PC, accounting 65.6% of the variation, mainly was a function of stride velocity and length, foot elevation (mean strike angle and mean lift-off angle), percentage of gait cycle spent on double support and variability (CV of stride time, CV of swing time, CV of stride velocity and CV of stride length). The second PC, accounting 18.5% of the variation, indicated the rhythm (mean stride time, mean cadence and mean swing time). Gait characteristics are highly correlated. The composition of PC1 emphasizes the positive correlation between mean stride velocity, mean stride length, mean strike angle and mean lift-off angle and on the other side the positive correlation between the variability and the mean double support. The structure of PC2 indicates that mean stride time is highly negative correlated with mean cadence, and is positive correlated with mean swing time.

TABLE 1. Gait characteristics involved in the PCA, percentage of total variation and loading values for the first two principal components.

Gait Characteristics	Median [Min, Max]	PC1 (65.6%)	PC2 (18.5%)
Mean Stride Time (seconds)	1.09 [0.89, 1.36]	0.224	**0.503**
Mean Cadence (steps/minute)	110.7 [88.7, 134.8]	-0.225	**-0.499**
Mean Swing Time (seconds)	0.41 [0.32, 0.53]	0.017	**0.632**
Mean Stride Velocity (meters/second)	0.84 [0.34, 1.57]	**-0.341**	-0.057
Mean Stride Length (meters)	1.00 [0.35, 1.39]	**-0.333**	0.105
Mean Double Support (% gait cycle)	25.00 [14.11, 35.99]	**0.289**	-0.122
Mean Strike Angle (degree)	17.00 [0.16, 31.27]	**-0.310**	0.130
Mean Lift-off Angle (degree)	52.80 [22.64, 68.97]	**-0.314**	0.114
Stride Time Variability (CV) (%)	3.00 [0.87, 8.37]	**0.310**	-0.038
Swing Time Variability (CV) (%)	4.95 [1.69, 18.42]	**0.312**	-0.064
Stride Velocity Variability (CV) (%)	5.53 [2.01, 27.36]	**0.318**	-0.072
Stride Length Variabiliy (CV) (%)	4.08 [1.34, 27.26]	**0.315**	-0.171

In the two-dimensional plot of the PCs the controls concentrated more in the left side while the patient group were spread on the graph, revealing the homogeneity of control group and the heterogeneity of patient group (Figure 1). The fuzzy clustering divided the set of observations into three subgroups with different gait characteristics. Since PCs were uncorrelated each cluster correspond to a different gait pattern. Cluster 1 covered all controls, 4 IPD and 1 VaP. The subjects of this cluster are concentrated in the negative side of PC1 which correspond the group with higher stride velocity, stride length and foot elevation. Note that the points corresponding to IPD and VaP located in this cluster are above the controls meaning that although they have some gait characteristics closer to the controls, they present higher mean stride and swing time. Cluster 2 is represented majority by VaP (8 VaP and 2 IPD) with higher double support and variability. Cluster 3 with 9 IPD and 6 VaP are between and overlapping with cluster 1 and 2, with some subjects distinguished by higher stride time and stride swing. Consistent with recently findings quantitatively synthesized in [10]

IPD and even more VaP present higher gait variability and percentage of gait cycle spent on double support (higher PC1 values), being the subjects most characterized by these gait characteristics are grouped in cluster 2.

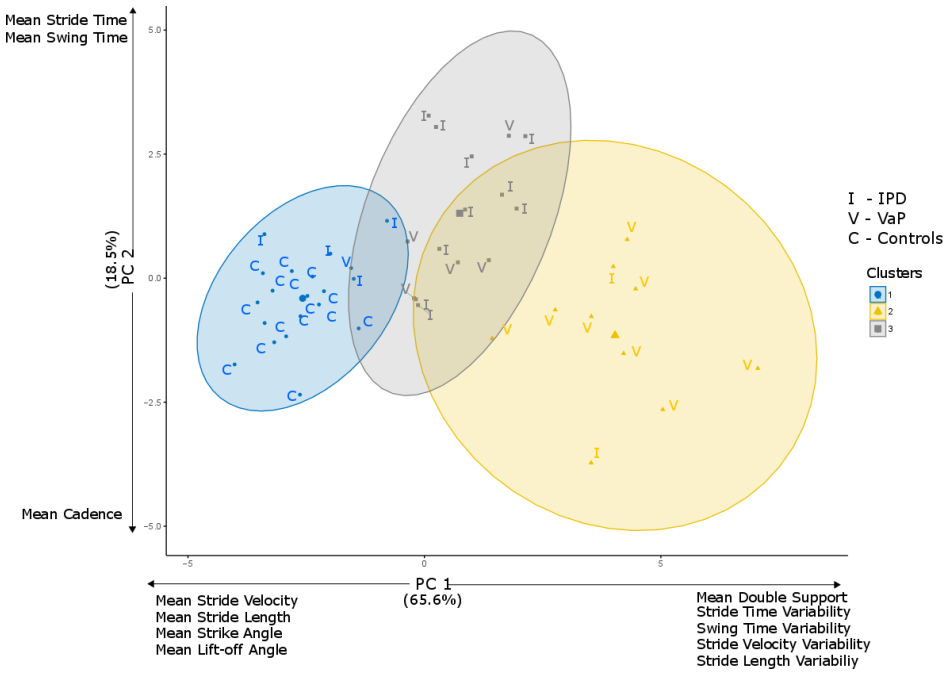

FIGURE 1. Clustering representation of different subgroups with different gait types by using PC1 vs PC2. As a PC1 increases, the gait characteristics values on the right also increase, while the gait characteristics on the left decrease. As a PC2 increases, the gait characteristics values on the top also increase, while the gait characteristics on the bottom decrease.

CONCLUSION

This study shows that the use of PCA and fuzzy c-means cluster analysis allow an objective graphical cluster of subgroups of subjects with similar gait characteristics. Graphical representation as Figure 1 facilitates the read out of the gait pattern of each cluster and each individual. Therefore, graphical clustering could aid in clinical evaluation and management of the disease. Further studies with larger sample size are required to enhance our understanding of the new gait variables provides by PCA based on different gait characteristics and their utility as mirrors of gait performance differentiation.

ACKNOWLEDGMENTS

This work is partially funded by the ERDF through the COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT as part of project UID/EEA/50014/2013.

REFERENCES

1. S. Lord, B. Galna, and L. Rochester, “Moving forward on gait measurement: toward a more refined approach”, in *Movement Disorders*, 28(11), (2013), pp.1534-1543.
2. J. H. Hollman, E.M. McDade, and R.C. Petersen, “Normative spatiotemporal gait parameters in older adults”, in *Gait & posture*, 34(1), (2011), pp. 111-118.
3. A. Carriero, A. Zavatsky, J. Stebbins, T. Theologis, and S.J. Shefelbine, S. J., “Determination of gait patterns in children with spastic diplegic cerebral palsy using principal components”, in *Gait & posture*, 29(1), (2009), pp. 71-75.
4. I.T. Jolliffe, I. T., “Principal component analysis and factor analysis”, in *Principal component analysis*, Springer, New York, NY, (1986), pp. 115-128
5. J. Abonyi, B. Feil, “Cluster analysis for data mining and system identification”, in Birkhäuser Verlag AG, Berlin (2007)

6. C. Prakash, R. Kumar, and N. Mittal, “Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges”, in *Artificial Intelligence Review*, *49*(1), (2018), pp. 1-40.
7. J. C. Bezdek, E. Robert, and F. William Full. “FCM: The fuzzy c-means clustering algorithm”, in *Computers & Geosciences* 10, no. 2-3 (1984), pp.191-203.
8. K. Zhou, C. Fu and S. Yang. “Fuzziness parameter selection in fuzzy c-means: the perspective of cluster validation,” in *Science China Information Sciences 57,* no. 11 (2014), pp.1-8.
9. Malika Charrad, Nadia Ghazzali, Veronique Boiteau, Azam Niknafs (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, 61(6), 1-36.
10. Y. Moon, J. Sung, R. An, M. E. Hernandez, J. J. Sosnoff, “Gait variability in people with neurological disorders: a systematic review and meta-analysis”, in *Human movement science* 47 (2016), pp.197–208.