Motion Analysis for People with Cerebral Palsy: A Vision Based Approach

Mário Macedo[1*], Alexandre Candeias[1*], Manuel Marques[1]

Abstract— **We propose a methodology to classify motion of subjects with cerebral palsy based on RGB image sequences and present a new dataset with 2D facial landmark trajectories from RGB images of people with and without disabilities while performing specific types of movements. Depending on these movements, parts of the face can be occluded and we are able to recover the 3D face's shape and its motion based on the Structure from Motion framework. Using the 3D structure and the motion, we propose two different motion descriptors, one is focused on describing the spatial distribution of the motion and the other on the temporal distribution. Finally, we discuss the physical meaning of these descriptors and show that they are very informative about the degree of the subjects' disabilities. Our descriptor can classify people with and without cerebral palsy from 2D image sequences.**

I. Introduction

Despite the discoveries in the field of medicine, the percentage of youngsters with severe disabilities still reaches 0.7% [1]. Cerebral Palsy (CP), the most common cause of significant motor impairment in children [2], occurs in 2 to 3 per 1000 live births and its prevalence is declining slowly [3]. CP is associated with damages in parts of the brain that control movement, coordination, balance and posture, which impedes the normal movement [4]. However, people are affected differently by this disorder, either in the location affected, or the degree of control over that region. Therefore, classifying these people is a complex task, which still does not have a generalized result.

The Gross Motor Function Classification System aims to provide a standardized measurement of severity of the disorder [5]. However, such classification is not accurate since it relies on the subjectivity of such analysis. So, under the same circumstances, different experts might classify the same person differently. Automatic classification could circumvent this problem but its development is difficult because the publicly available data is very scarce.

Our work contributes in solving the two aforementioned problems. Firstly, we contribute with a dataset for motion analysis of people with cerebral palsy. This dataset is composed by facial landmarks trajectories (see Fig.1) extracted from RGB video sequences of people with and without cerebral palsy while executing specific types of motion.

Secondly, we contribute with two motion descriptors, named as trajectons. These trajectons are constructed using

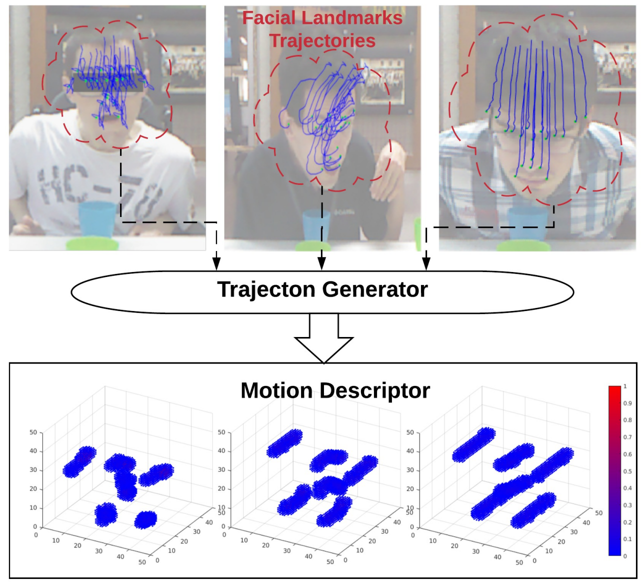


Fig. 1. Based on 2D facial landmark trajectories, we propose two motion descriptors, named as trajectons, that can classify the motion of the subjects and are informative of their disabilities. In the top, we show in blue the 2D facial landmarks trajectories, which are the input for our motion descriptor generator. From the left to the right: subject with severe difficulties, with mild difficulties and without disabilities, while moving their head forward. The 2D facial landmarks that are used in our motion descriptor are drawn in green. In the bottom, we show a visualization of one of the proposed motion descriptors for each of the subjects (in the same order as before). The trajecton corresponding to the subject without disabilities is smoother than the trajectons of the subjects with disabilities.

2D facial landmarks trajectories, acquired using simple RGB image sequences. Moreover, we can classify the subjects, using the trajectons, according to their ability to perform different movements.

II. State of the Art

Human motion analysis for action recognition using 2D image point trajectories is a well known problem[6][7]. In [6], 2D point trajectories are clustered in an unsupervised way to construct a dictionary and using the description provided by this dictionary a Support Vector Machine (SVM) classifier is used to classify videos containing actions such as sitting or getting out of a car. Other works extract the basic movement of the subject and represent it as a simple 2D human stick figure and then, a predictive modular neural network time series classification algorithm is applied [8].

Proposed in [9] and [10], movement assessment of patients recovering from stroke and having Parkinson disease is based on the analysis of the 3D skeleton joints. These 3D joints

* Mário Macedo and Alexandre Candeias contributed equally.

[1] Mário Macedo, Alexandre Candeias and Manuel Marques are with the Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal. mario.macedo@tecnico.ulisboa.pt, alexandre.candeias@tecnico.ulisboa.pt, manuel@isr.tecnico.ulisboa.pt.

positions are acquired using a RGB-D Kinect camera.

Recently, different approaches using RGB-D images and deep learning are used for action recognition of movements such as walking or running [11][12].

Other methods for trajectory classification are based on time-series classification using Hidden Markov Models (HMM) [13]. Hand-Crafted features from the trajectories, such as velocity and acceleration are used as input features of SVM's to classify trajectories in [14]. Similar methods are applied in [15], the motion of stroke patients is analyzed and classified using a HMM and a logistic regression. This method acquires data from the motion of healthy and stroke patients using Inertial Measurement Units (IMU) placed in the body of the patients. An IMU provides very accurate data of the motion accelerations, but it requires calibration and it can be uncomfortable to the patients.

In this work, we rely on 2D facial landmark trajectories, computed from a simple RGB camera, to describe and classify the motion of people with cerebral palsy and we propose new descriptors for these trajectories in a similar way to the method proposed in [6].

III. Data Set

We acquired a new dataset for motion analysis of people with cerebral palsy. Our dataset contains 18 participants, 8 with cerebral palsy and 10 without any physical limitation.

The dataset is composed by 2D trajectories of facial landmarks (see Fig.1). We asked the participants to perform 4 types of movements: standing still, moving forward, moving to the left and moving to the right. We acquired 1 sequence containing each type of movement and 4 participants were recorded twice under different circumstances, giving, after segmenting the parts corresponding to the different movements, a total of 88 video sequences.

The facial landmarks were estimated using the OpenPose[16][17] and we used the 19 landmarks drawn in green on Fig.1.

All the participants signed an informed consent form, in accordance to the World Health Organization and in order to preserve their identity, only the trajectories acquired are shared.

IV. Methodology

Given a sequence of 2D images, and passing it through a facial landmark detector gives access to a data matrix W given by:

$$W = \begin{bmatrix} u_{1,1} & \cdots & u_{1,P} \\ v_{1,1} & \cdots & v_{1,P} \\ \vdots & \ddots & \vdots \\ u_{F,1} & \cdots & u_{F,P} \\ v_{F,1} & \cdots & v_{F,P} \end{bmatrix}. \tag{1}$$

The data matrix $W \in \mathbb{R}^{2F \times P}$ is composed by the 2D projections (u, v) of each P facial landmarks along F frames of each video sequence.

A. Background

Using the data matrix W, and assuming that the face is a rigid body [18], we can use Structure from Motion (SfM) framework [19] and recover the 3D structure (shape) and the motion of the face along a sequence of frames. Under the assumption of scaled orthographic projection, the data matrix W factorizes as:

$$W = MS + T\mathbf{1},$$
$$M = \begin{bmatrix} M_1 \\ \vdots \\ M_F \end{bmatrix}, \quad S = \begin{bmatrix} S_1 \ldots S_P \end{bmatrix}, \quad T = \begin{bmatrix} t_{u,1} \\ t_{v,1} \\ \vdots \\ t_{u,F} \\ t_{v,F} \end{bmatrix}, \tag{2}$$

where $M \in \mathbb{R}^{2F \times 3}$ is called the motion matrix, $S \in \mathbb{R}^{3 \times P}$ is the shape matrix, which corresponds to the 3D structure of the object, $T \in \mathbb{R}^{2F \times 1}$ is the translation vector and $\mathbf{1}$ is a vector of ones. Under this assumption, we have the metric constraint: $M_i M_i^T = \alpha_i I_2$ for $i = 1 \ldots F$, where I_2 is the 2×2 identity matrix, the factorization can be solved by [20].

B. Data Imputation and Normalization

*1) **Shape and Motion Estimation**:* The data matrix W obtained through the facial landmark detection is incomplete due to face occlusion during the video sequences. In order to complete this data matrix we used a matrix completion method [21] which imposes the motion constraints to the data. We could recover the shape (S) and parts of the motion of each subject along a video sequence. Due to the low number of landmarks detected in some parts of the video, this method is not able to estimate all the missing entries, or more specifically the motion in every frame.

We used the shape information and Lucas-Kanade method [22] to recover the motion in all frames. Using Lucas-Kanade method we have more accuracy in the detection of landmarks.

With the recovered shape (S) we recover the motion (M_i), in each frame, by solving the following problem:

$$\begin{aligned} (\alpha^*, R^*) = \underset{\alpha, R}{\operatorname{argmin}} & \quad \|\alpha RS - X\|_F^2 \\ \text{subject to} & \quad RR^T = \mathbb{I}_{2\times 2},\ \alpha > 0 \end{aligned} \tag{3}$$

where $X \in \mathbb{R}^{2 \times P}$ corresponds to the centered landmarks positions provided by Lucas-Kanade, $\alpha \in \mathbb{R}$ and $R \in \mathbb{R}^{2\times 2}$ are the motion parameters for that frame, $M_i = \alpha R$. This optimization problem has a closed form solution [23] and was integrated in each step of the Lucas-Kanade method.

Using these two methods, first estimating the shape [21] and then totally recovering the motion (3), we manage to have a full estimation of the motion (M) in each sequence and a reconstruction of the shape (S) for each subject.

*2) **Data Normalization**:* We want to design descriptors that capture the motion of each subject and are not related to the specific shape. According to the model (2), the matrix W is correlated with the shape (S).

In order to have a data matrix (W) depending only on the motion, we use the estimated motion matrices (M) and a normalized shape (S_n), the normalized data matrix is then given by: $W_n = MS_n$.

For each subject the normalized shape, $S_n = R^* S_{ref}$, is calculated by solving:

$$\begin{aligned} (R^*, D^*) = \underset{R,D}{\text{argmin}} \quad & \|RDS_{ref} - S\|_F^2 \\ \text{subject to} \quad & RR^T = \mathbb{I}_{3\times3}, \\ & D_{ii} > 0,\ D_{ij} = 0,\ i \neq j \\ & i = 1,2,3 \text{ and } j = 1,2,3, \end{aligned} \tag{4}$$

where S_{ref} is a reference shape chosen from the available shapes of each subject and S is the subject specific shape estimated using the method described in IV-B.1. This problem is an anisotropic Procrustes problem with post-scaling and can be solved according to [23].

C. Motion Descriptors

In order to describe the different motions of the subjects present in the dataset, two different descriptors, referred to as trajectons, are proposed. These trajectons are calculated using the normalized 2D landmark positions W_n and parameters α^i, which are inversely proportional to the distance of the subject along the perpendicular axis to the image plane. A trajecton is computed using a sliding window of k frames. Depending on the number of frames, we will have a different number of trajectons for each sequence.

*1) **6P Trajecton**:* This descriptor is based on creating a discretized 3D cube and is focused on measuring the spatial distribution of the landmark points over the window of frames. Each 3D point on this cube is the 2D facial landmark positions (u,v) and the value of α in each frame. While the subject is moving, the facial landmarks will occupy different voxels on this cube and each voxel will be scored accordingly, as the scheme on Fig.2 describes.

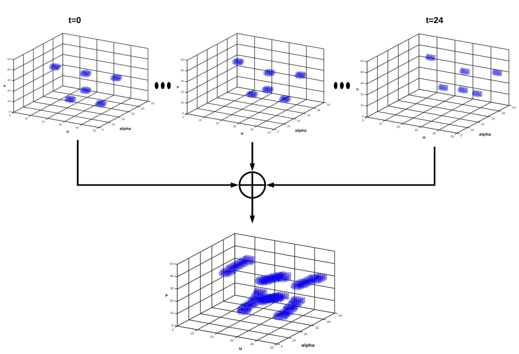


Fig. 2. The 6P trajecton is calculated by accumulating the positions (u,v), and α in a discretized cube of the facial landmarks throughout time (landmarks positions are represented in blue). The descriptor is a vectorized version of the cube which reflects the spatial occupation of the subject along a window of $k = 25$ frames.

In the end, the 6P trajecton will be composed by stacking these voxels in a vector, describing the spatial location of the facial landmarks along the window of frames.

In our experiments, we used 6 facial landmarks (the points of the eyes and corners of the nose and the mouth) in each frame. We discretized each direction (u,v and α) using 50 bins, giving a final descriptor of dimension 125000.

*2) **Shaky Trajecton**:* The Shaky trajecton captures the degree of the oscillations in the different movements. In other words, this descriptor focuses on the temporal changes of the 2D coordinates of the landmark points (u and v) and in α.

First, we compute the difference of the 2D coordinates of each landmark (u, v) and α for every pair of consecutive frames. These differences are then discretized into 9 bins along each direction. We create a 2D histogram for the differences of the coordinates in the image plane (u, v) for each landmark and a 1D histogram for the α value, where each pair of consecutive frames present in the window votes for a bin in these two histograms. In the end, the Shaky trajecton contains a 2D histogram for each of the landmark points and a 1D histogram for the α representing the whole window of frames, as shown in Fig.3.

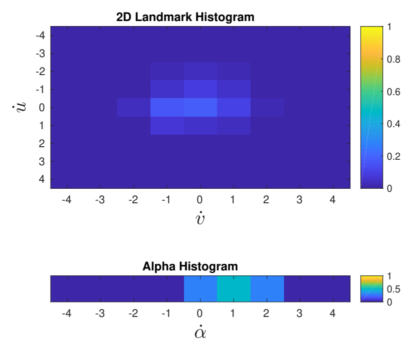


Fig. 3. Top: example of a 2D histogram for a specific landmark based on the differences of the landmark position in the image plane for every pair of consecutive frames. Bottom: α histogram is calculated by discretizing the difference of α for every pair of consecutive frames. The Shaky trajecton is a vectorized version of these two histograms for each landmark.

This descriptor is constructed by stacking the columns of the 2D landmark's histograms and the α histogram giving a vector of dimension $17 \times 81 + 9 = 1386$. In this descriptor we did not use the landmarks corresponding to the ears (see Fig.1).

V. Results

A. Descriptors Visualization

To understand the semantics of the proposed trajectons and how they can be informative of the disabilities of each subject, we show each trajecton for three different subjects, one without any physical limitation, one with mild physical limitations and one with severe physical limitations. The subjects performed two different types of movements: standing still in front of the camera, as Fig.4 shows, and moving forward, as it is shown in the top of Fig.1.

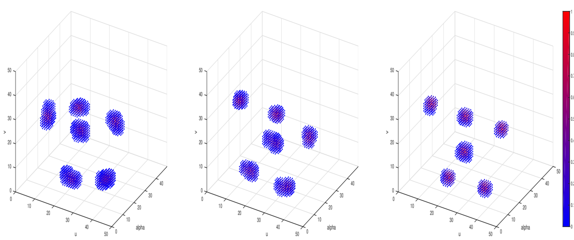
The trajectories in Figs.4 and 1 correspond to a window of 25 consecutive frames segmented while the subjects were performing the different types of movements. In this section, the trajectons correspond to the same window of frames.

(a) (b) (c)

Fig. 4. Facial landmarks trajectories (in blue) while subjects were asked to stand still in front of the camera. The trajectories are associated with subjects with severe (a), mild (b) and no (c) physical limitations.

*1) **6P trajecton**:* As previously discussed, the 6P trajecton was designed to describe the spatial variation of the movement of each subject.

Analyzing the trajecton corresponding to the movement of standing still (Fig.4), the subjects with disabilities (Figs.5(a) and 5(b)) tend to have a trajecton with the points more spread around the space, since they have difficulty to stand still. On the other hand, people without disabilities (Fig.5(c)) have no problem of staying still and their trajecton is very concentrated around a specific point in the space.



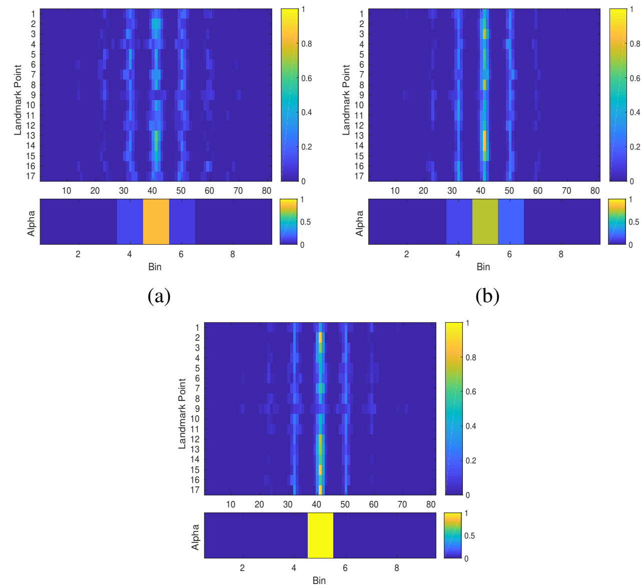
(a) (b) (c)

Fig. 5. 6P trajecton for the movement of standing still. The subject with severe disabilities Fig.5(a) has a trajectory that is more spread in space than the subject with mild disabilities (Fig.5(b)) and without disabilities (Fig.5(c)). The color of each voxel denotes its score, from 0-blue to 1-red.

For the movement of moving forward, the subject without disabilities (bottom Fig.1) shows a trajecton almost composed by straight lines, revealing a stable movement. Differently, the trajectons of the subjects with disabilities are not composed by straight lines and the lines show a shorter length. This behavior shows the difficulty of people with disabilities to perform a straight line while moving forward and a shorter amplitude of movements.

*2) **Shaky trajecton**:* The purpose of the shaky trajecton is to extract the spatial changes of the movement along time. As we explained in section IV-C, this descriptor is composed by two histograms: one with 81 bins for each of the landmark points and other with of 9 bins for alpha. Each bin contains the counts corresponding to the discretization of the time difference (velocity) of the landmark on the image plane and the time difference of alpha.

We can see clearly, in Fig.6, that the subject with severe disabilities (Fig.6(a)) has a very different descriptor from the other two subjects (Fig.6(b) and Fig.6(c)), while standing still. The velocity of the landmark points and alpha is more spread around the bins corresponding to zero velocity (bin 41 for the landmarks and 5 for alpha), while in the other subjects the descriptor shows a larger concentration around the bin corresponding to zero velocity.



(a) (b)

(c)

Fig. 6. Shaky trajecton for the standing still motion associated with the subject with severe (a), mild (b) and no difficulties(c).
For top figures, each row represents the 2D histogram of a landmark point. The bottom figure is the representation of the α histogram.

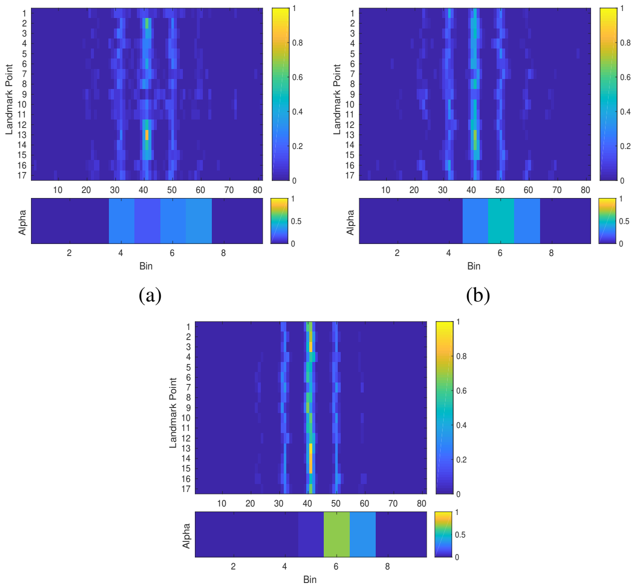
We can also see that the subject with mild disabilities (Fig.6(b)), although showing velocities of the landmark points similar to the one without disabilities, reveals his difficulties in standing still in the alpha direction, which is related with the direction perpendicular to the image plane.

The same analysis can be done for the shaky trajecton corresponding to the forward movement (Fig.7). We see that the velocities of the landmark points and alpha are more spread in the histograms of the subjects with disabilities. More interesting, we can see that the subject with severe disabilities while moving forward show negative velocities in alpha (Fig.7(a) bin 4), this means that this subject while moving forward, sometimes, moves in the opposite direction.

B. Trajecton Clustering

Previously, we discussed how the visualization of the proposed trajectons is informative and descriptive of the subject's disabilities for three different subjects.

To better understand how the trajectons can be informative in the whole dataset and have a description for each subject's sequence, we constructed a bag of words(BOW) model using k-means clustering. This model was trained using trajectons computed from windows from the sequence of each subject. After learning the centroids using k-means, each of the trajectons votes to a specific centroid, creating a description for the whole sequence of the subject that is independent of



(a) (b) (c)

Fig. 7. Shaky trajecton visualization for the moving forward motion. Figure (a) is associated with the subject with severe difficulties, (b) with mild difficulties and (c) with no difficulties.

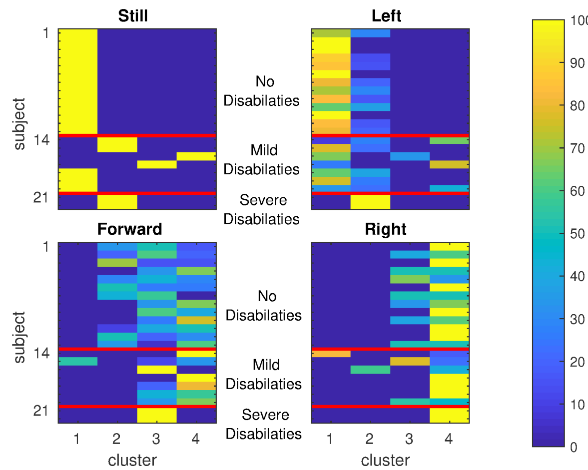


Fig. 8. Clustering using k-means for the 6P trajecton. The color denotes the votes percentage in each cluster. We have many trajectons, each trajecton votes for a cluster, for each subject since the trajectons are computed using a time-window of 25 frames.

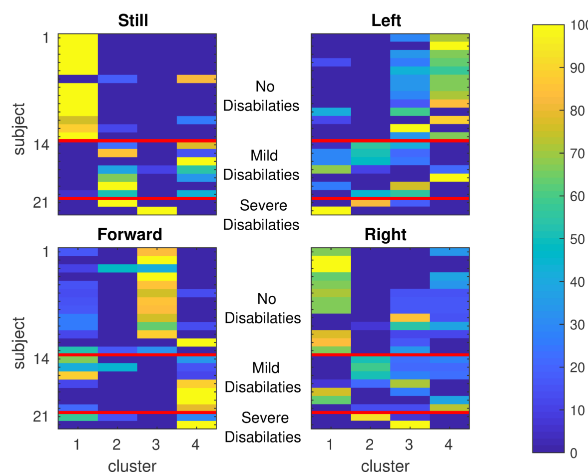


Fig. 9. Clustering using k-means for the Shaky trajecton. The color denotes the votes percentage in each cluster. We have many trajectons, and each trajecton votes for a cluster, for each subject since the trajectons are computed using a time-window of 25 frames.

the number of frames present in each sequence. We created a separate BOW model for each of the different movements, namely forward, standing still, going left and going right. In our experiments, we found that using 4 clusters and training the model separately in the different types of movement yield better results.

*1) **6P trajecton**:* We show in Fig.8 the results of the clustering for the 6P trajecton. We can see that for the still movement there is a separation between the subjects with disabilities and the subjects without disabilities. Subjects without disabilities tend to be assigned to the 1st cluster and the subjects with disabilities are spread across other clusters. In the other types of movements(forward, left and right) is difficult to have a clear separation between the subjects. Nevertheless, it is visible that the pattern of the distribution of the votes by the clusters is different in subjects with disabilities and without disabilities.

*2) **Shaky trajecton**:* In Fig.9, we present the same clustering analysis for the Shaky trajecton. Using this type of trajectons, we can see more clearly that the subjects without disabilities tend to be grouped together in the same cluster (e.g cluster 1 in still movement, cluster 3 in forward , cluster 4 in left movement and cluster 1 in right movement) and subjects with disabilities tend to be spread in the other clusters. This spreading of the subjects with disabilities in different clusters can be explained by the variability of the motion of these subjects, making it difficult to have examples of subjects with the same kind of motion difficulties.

C. Subject Classification

In order to test the performance of the trajectons in a classification task, we used a simple NN(nearest neighbour) classifier using the description for the whole sequence provided by k-means, basically each subject's sequence is described by the rows present in Figs.8 and 9. The k-means dictionary was trained using trajectons from the subjects in the training set, using the k-means centroids each sequence was described. In testing time each subject's sequence was assigned with the label(with or without disabilities) of the nearest neighbor from the training set, using the k-means description of the sequence. We divided our dataset in a training set and a test set, the test set contains 4 subjects, 2 without any disabilities and 2 with disabilities. In table I we show the accuracy values of the average classification using NN for 20 different combinations of subjects in the training/test set.

Due to the small amount of data available, we found that when removing the whole sequences from the test subjects of the training set while learning the k-means dictionary led to very low accuracy levels (see ”Complete Remove” column in table I). To solve this problem we used part of the sequences of the subjects from the test set to learn the k-means dictionary, however, these parts were not classified in test time to avoid information repeated in the training and test set (see ”Partial Remove” column in table I).

	Complete Remove	Partial Remove
6P trajecton	65.3%	71.5%
Shaky trajecton	**66.5%**	**93.1%**

TABLE I

AVERAGE CLASSIFICATION ACCURACY FOR 20 DIFFERENT COMBINATIONS OF SUBJECTS USING NN(NEAREST NEIGHBOUR). COMPLETE REMOVE DENOTES THE DATASET WHERE THE COMPLETE SEQUENCE FROM THE SUBJECTS IN THE TEST SET WAS REMOVED IN TRAINING TIME. PARTIAL REMOVE DENOTES THE DATASET WHERE SOME FRAMES FROM THE SEQUENCES OF THE TEST SUBJECTS WERE USED IN TRAINING TIME (HOWEVER THESE FRAMES WERE NOT USED IN TEST TIME).

Analyzing the results present in table I, we can take the same conclusions as in the previous section, we see that the Shaky trajecton discriminates better the subjects with disabilities than the 6P trajecton. This means that the differences in the motion between subjects with and without disabilities are more related to the differences in the velocities of the motion than with the spatial location and occupancy of the motion.

VI. CONCLUSIONS

In this work, we focused on creating a methodology to classify different subjects suffering from cerebral palsy based on RGB image sequences.

As there was no type of data set available, one had to be constructed for this work. This data set consisted on labelled 2D trajectories of facial landmarks of different subjects (with and without disabilities) while performing predefined movements.

Moreover, two different types of descriptors were developed, which were able to describe the trajectory of the subjects while they were performing different motions. One focusing on the spatial information whereas the other focused on the temporal information. Analyzing these descriptors, we found that the one based on the temporal information(Shaky trajecton) is more informative of the subjects disabilities. Due to their semantics, these descriptors could be used as an analysis tool for the classification of subjects with disabilities.

In the end, we were able to automatically classify the motion of healthy and disabled subjects using the proposed descriptors showing promising results. In future work, we plan to use the proposed descriptors in more challenging motion classification tasks.

ACKNOWLEDGMENT

The authors specially thank to users, staff and the board Centro de Reabilitação de Paralisia Cerebral Calouste Gulbenkian for their availability. This work was funded by a FCT project, under the CMU-Portugal program, [CMU/ECE/0005/2017] and by FCT grant [UID/EEA/50009/2019].

REFERENCES

[1] “World Report on Disability - Summary,” *World Report on Disability*, 2011.

[2] M. J. Maenner, S. J. Blumberg, M. D. Kogan, D. Christensen, M. Yeargin-Allsopp, and L. A. Schieve, “Prevalence of cerebral palsy and intellectual disability among children identified in two U.S. National Surveys, 2011-2013,” *Annals of Epidemiology*, 2016.

[3] E. Sellier, M. J. Platt, G. L. Andersen, I. Krägeloh-Mann, J. De La Cruz, and C. Cans, “Decreasing prevalence in cerebral palsy: a multi-site European population-based study, 1980 to 2003,” *Developmental Medicine & Child Neurology*, 2016.

[4] S. Gulati and V. Sondhi, “Cerebral Palsy: An Overview,” *The Indian Journal of Pediatrics*, 2017.

[5] C. Morris and D. Bartlett, “Gross motor function classification system: Impact and utility,” *Developmental Medicine and Child Neurology*, 2004.

[6] P. Matikainen, M. Hebert, and R. Sukthankar, “Trajectons: Action Recognition Through the Motion Analysis of Tracked Features,” *IEEE 12th ICCV Workshops*, 2009.

[7] A. Gaidon and C. Schmid, “Recognizing activities with cluster-trees of tracklets,” *BMVC2012*, 2012.

[8] V. Petridis, B. Deb, and V. Syrris, “Detection and identification of human actions using Predictive Modular Neural Networks,” in *17th Mediterranean Conference on Control and Automation*, 2009.

[9] S. Spasojević, A. Rodić, and J. Santos-Victor, “Kinect-based approach for upper body movement assessment in stroke,” in *Mechanisms and Machine Science*, 2019.

[10] S. Spasojević, J. Santos-Victor, T. Ilić, S. Milanović, V. Potkonjak, and A. Rodić, “A vision-based system for movement analysis in medical applications: The example of parkinson disease,” in *Lecture Notes in Computer Science*, 2015.

[11] P. Wang, W. Li, P. Ogunbona, J. Wan, and S. Escalera, “RGB-D-based human motion recognition with deep learning: A survey,” *Computer Vision and Image Understanding*, 2018.

[12] P. Wang, W. Li, C. Li, and Y. Hou, “Action Recognition Based on Joint Trajectory Maps with Convolutional Neural Networks,” Tech. Rep., 2016. [Online]. Available: https://arxiv.org/pdf/1612.09401.pdf

[13] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, “Object Trajectory-Based Activity Classification and Recognition Using Hidden Markov Models,” *IEEE TIP*, 2007.

[14] X. Xiao, H. Hu, and W. Wang, “Trajectories-based motion neighborhood feature for human action recognition,” in *IEEE ICIP*, 2017.

[15] J. Guerra, J. Uddin, D. Nilsen, J. McLnerney, A. Fadoo, I. B. Omofuma, S. Hughes, S. Agrawal, P. Allen, and H. M. Schambra, “Capture, learning, and classification of upper extremity movement primitives in healthy controls and stroke patients,” in *IEEE ICORR*, 2017.

[16] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, “Convolutional pose machines,” in *IEEE CVPR*, 2016.

[17] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” *IEEE CVPR*, 2017.

[18] M. Marques and J. Costeira, “3D face recognition from multiple images: A shape-from-motion approach,” in *IEEE FG*, 2008.

[19] C. Tomasi and T. Kanade, “Shape and motion from image streams: a factorization method.” *Proceedings of the National Academy of Sciences*, 1993.

[20] D. Weinshall and C. Tomasi, “Linear and Incremental Acquisition of Invariant Shape Models from Image Sequences,” *IEEE TPAMI*, 1995.

[21] M. Marques and J. Costeira, “Estimating 3D shape from degenerate sequences with missing data,” *Computer Vision and Image Understanding*, 2009.

[22] S. Baker and I. Matthews, “Lucas-Kanade 20 Years On: A Unifying Framework,” *IJCV*, 2004.

[23] J. C. Gower and G. B. Dijksterhuis, *Procrustes Problems*, 2004.