A recurrent approach for predicting Parkinson stage from multimodal videos

John Archila[a], Antoine Manzanera[b], and Fabio Martínez[a]

[a]Biomedical Imaging, Vision and Learning Laboratory(BivL^{2}ab), Universidad Industrial de Santander (UIS). Santander, Colombia
[b]U2IS, ENSTA Paris, Institut Polytechnique de Paris, France

ABSTRACT

Parkinson's disease is a neurodegenerative disease that affects more than 6.1 million people worldwide. In the clinical routine, the main tool to diagnose and monitor disease progression is based on motor impairments, such as postural instability, bradykinesia, tremor, among others. Besides, new biomarkers based on motion patterns have emerged to describe disease findings. Nonetheless, this motor characterization has low sensitivity, especially at early stages, and is largely expert-dependent, because protocols are mainly based on visual observations. However, most of these analyses require complex and some invasive systems that additionally only bring global information of complete recordings. This work introduces a multimodal approach that integrates gait and eye motion videos to quantify and predict patient stage on-the-fly. This method starts by computing dense apparent velocity maps that represent the local displacement of the person seen from the gait in a sagittal plane and as micro-movements during the fixation experiment. Then, each frame is described as a covariance descriptor of deep feature activation maps computed over the motion field at each video time. Then, the covariance video manifold is mapped to a recurrent LSTM network to learn higher non-local dependencies and quantify a motion descriptor. Also, an end-to-end scheme allows to lately fuse both modalities (gait and fixational eye) to obtain a more sensitive Parkinson disease descriptor. In a study with 25 subjects, the proposed approach reaches an average F1-score of 0.83 with an average recall of 0.78. In a temporal prediction analysis, the approach reports major correlations with the disease considering swing phase.

Keywords: Neuro-motion patterns, Gait, Eye motion, Parkinson stages, Computer vision, LSTM.

1. INTRODUCTION

Parkinson's disease (PD) is the second most common neurological disorder in the world.[1] PD is mainly associated to the progressive loss of dopaminergic neurons in the midbrain, producing diverse cardinal signs such as rest tremor, bradykinesia, rigidity, and postural instability.[2] The diagnosis and evaluation of disease progression is typically carried out by quantifying gait motor patterns and postural configurations, or evaluating exercises related to control and coordination. The detection, evaluation, and stage characterization are however highly subjective, depending on coarse therapy rating scales. Besides, the pattern computation setup is usually based on markers, which result invasive and insufficient to capture subtle motor alterations.[3,4] In fact, some studies have reported an accuracy lower than 81% for Parkinson disease diagnosis at early stages.[5] Moreover, much of these evaluations consider isolated observations of particular alterations, being insufficient to characterize the PD. Hence, the proposal of new alternatives to capture motion patterns, as well as the design of new digital spatio-temporal biomarkers may result fundamental to carry out early diagnosis and to successfully monitor progression of the disease, and impact of different personalized treatments.

For instance, markerless approximations have been proposed from the geometrical analysis of body silhouettes over time, capturing mode motion variations, hence limited to strong motion changes.[6] Other approaches have used video analysis to recover postural configurations, emulating classical marker protocols, but again reducing dynamic study to a set of joints.[7] Recent approaches have proposed the integration of multiple motion modes to enhance disease description. For instance, typical Parkinsonian writing test are complemented with speech[8]

Further author information: famarcar@saber.uis.edu.co

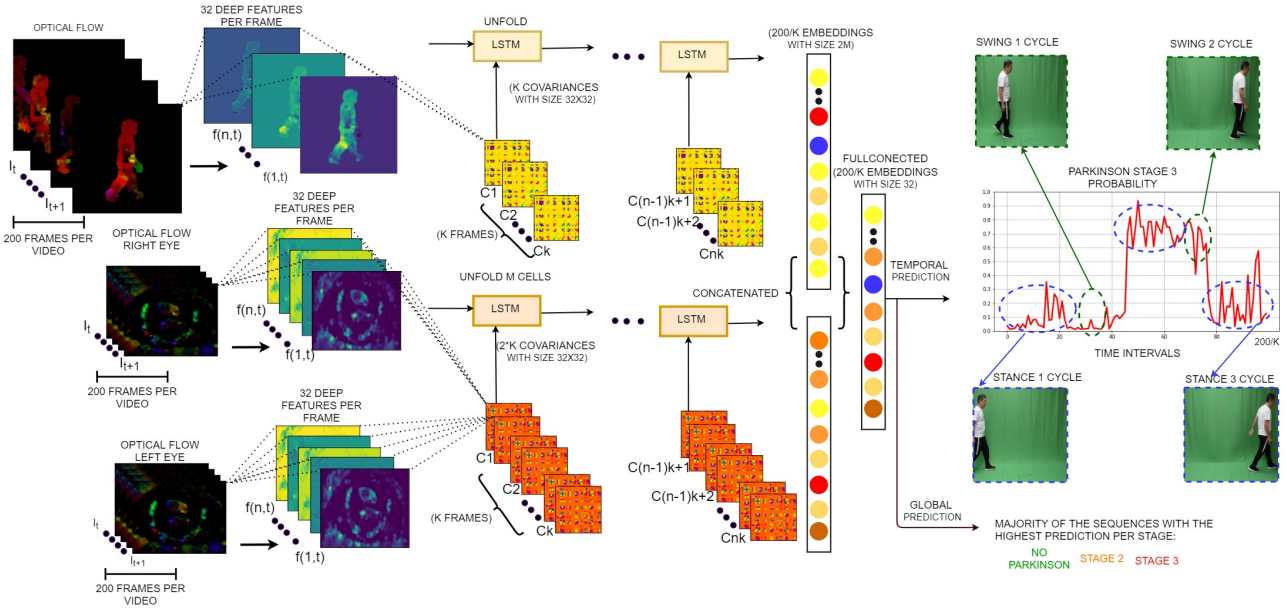

Figure 1. Proposed approach pipeline. A markerless strategy is proposed by computing a spatially dense representation for each frame along the video sequence.

and gait[9,10] patterns. These methods have a lower sensitivity and therefore the classification result effective when the tremor, the rigidity and the bradykinesia are prominent. Other works have been dedicated to explore new biomarkers such as the eye movements, that may result more sensitive to diagnose PD at early stages.[11] Some studies have found that eye fixation and smooth pursuit motor abnormalities have a strong correlation with PD.[12,13] Nevertheless, such experimental observations require sophisticated sensing devices and protocols, that limit their use in the specialist's clinical routine. Overall, these approaches summarize motion patterns to coarse index of the disease stage, losing a temporal description of the alterations related to a particular gesture.

The main contribution of this paper is a multimodal motion approach that codifies temporal patterns from gait and eyes motion, to characterize the stage of PD in the patients. The proposed approach achieves a markerless per-frame video analysis to extract temporal neuromotor patterns associated with PD, that identifies the most critical motion sequences in the patient. The spatial information from the video is compactly coded as covariance matrices of deep features based in optical flow, which thereafter feed a recurrent net to learn temporal motion dependency of this covariance sequences. Finally, a prediction performed over time on video slices stratifies Parkinson's patients into different stages. This work opens perspectives towards new tools in motor therapies of patients, following the objective of identifying which movements and gestures are more significant of each PD stage.

2. PROPOSED APPROACH

Using frame-level spatial covariance matrices to fed a recurrent neural network, the proposed approach exploit spatial and temporal motor information. Moreover, this work has the capability to integrate gait and eye fixation sequences to improve characterization and description of PD. The proposed strategy is illustrated in Figure 1. Firstly, videos that record each modality of interest, are represented as a set of frame-covariance matrices. Then, the temporal processing is performed using a Long Short Term Memory (LSTM) network. Subsequently, a latent temporal vector is learned on LSTM to capture motor patterns along sequences. Finally, a softmax function is used to predict the probabilities of the different PD stages, for each video slice of every patient.

2.1 PD motion modalities

Parkinson disease causes different motor and non-motor impairments, at different stages, that impact different human activities, and can be recorded using different sensors. Quantifying these impairments allows to evaluate

the disease progression or the effect of a particular treatment. In this study, we include the gait and eye fixational movements as observational descriptors of the disease. These motion modalities are described thereafter:

- **The gait** is a complex locomotion process that requires coordination between neuromotor commands and muscles to obtain optimal displacement. Regarding PD, the gait alteration is generally characterised by unbalanced postures, slow movements and rigidity, which gives a global perspective of kinematic behavior. In video analysis have been proposed strategies based on silhouettes to encode different poses of each movement, highlighting geometric patterns typical of PD.[6] Also, 2D pose patterns have allowed to measure gait cadence and limb flexion angles, but the analysis was limited to advanced stages of the disease.[7]
- **Fixational eye patterns** have been identified as strongly correlated with dopamine deficit.[14] This analysis then can provide complementary information of locomotor patterns, being sensitive even in early stages. Most studies of oculomotor patterns use the video oculography technique (VOG). This method records two-dimensional movements and performs measurements such as velocity, latency, and other kinematics,[15].[16] As an example, the latency of divergence and convergence movements is greater for PD patients than for control subjects.[17] Nevertheless, the complexity of the acquisition system and its constant calibration makes the clinical routine of the specialist more difficult.

2.2 Frame covariance representation from deep motion features

The bradykinesia that is associated with the slowness of movement[18] and the tremor described as a periodic motion[19] are the predominant motor impairments for PD. Hence, velocity fields computed from video recordings are relevant to characterize and quantify PD. In fact, this hypothesis has been successfully implemented to analyze the disease in single modality, using either gait or tremor videos.[19,20] The proposed strategy then starts by computing dynamic patterns from gait and eye video sequences based on apparent velocity fields computed on consecutive frames. The Farnebäck optical flow method was applied considering its good trade-off between speed and accuracy.[21] This method estimates a dense and regular optical flow using a quadratic polynomial approximation of each pixel's neighborhood and estimating the translation vector that locally affects each polynomial.

Then, the kinematic information provided by optical flow was projected to a bank of convolutional filters extracted from the layers of a trained deep convolutional network. This projection is carried out at each frame allowing to enrich description of motion patterns. These deep features have recently demonstrated great capability to represent complex motion patterns, by modelling non linear and large scale relations. The set of learned filters decompose kinematic information in a total of n features maps $F_t = \{f_{(1,t)}, \ldots, f_{(n,t)}; f_{(i,t)} \subset \mathbb{R}^{W \times H}\}$, enhancing nonlinear relationships. In this work, we select the n filters F_t from the first layer of a pre-trained convolutional net, as a set of learned low-level features. The deep representation was provided by an in-depth separable convolution architecture, that allows computational reduction, less redundancy in activation maps, and that requires fewer data for training.[22,23] Hence, for each frame t, a spatial covariance matrix C_t relative to the set of feature maps F_t was computed to obtain a compact embedding description of postural and dynamic performance at each time. The covariance matrix is computed as: $C_t(i,j) = \mathbb{E}\left((f_{(i,t)} - \mathbb{E}(f_{(i,t)}))(f_{(j,t)} - \mathbb{E}(f_{(j,t)}))\right)$ where the expectation $\mathbb{E}$ is calculated over the $W \times H$ points of each feature map. The dimention of gait descriptor is $n \times n$ and ocular descriptor is $2 \times n \times n$ where n is the number of deep features per frame. The compact representation of covariance gives the correlation between poses. In this way, every patients has a signature per modality. This covariative signature show different trends between deep features motion maps, enhancing discordance of movements, arrhythmic patterns and tremor.[24,25]

2.3 A continuous multimodal motion pattern quantification

During the video sequence, the proposed method quantifies temporal information of deep covariation maps, representing potential motor impairments during movement execution. Two motion modalities (gait and eye fixation) were analyzed over time by computing recurrent deep representation, quantifying the probability for each time but also recovering a global disease classification. We implemented a recurrent LSTM with the capability to

encode short and long dynamic temporal deep representations. These recurrent modules have proven successful in many different tasks to temporal modelling of events, including the analysis of other neurodegenerative diseases.[26,27]

Formally, the set of frame covariances $\mathbf{F} = \{f_1, f_2 \dots f_k\}$, are sequentially propagated by a set of M recurrent units. In this layer, a resulting mid-level representation captures temporal dependencies, being m_t the internal state memory that is updated from the expression: $m_t = fg_t \otimes m_{t-1} + i_t \otimes \tilde{m}_t$. Then, the resulting hidden state h_t is computed as $h_t = o_t \otimes \tanh(m_t)$, where tanh the hyperbolic tangent function (valued in $[-1, +1]$) and hidden states are initialized to zero. The fg_t is a forget gate that uses a group of k covariances Ck_t to decide how much information omit. The expression for this function is herein calculated as: $fg_t = \sigma(W_f[x_t, h_{t-1}] + b_f)$, where σ is the sigmoid function (valued in $[0, 1]$), and $[.,.]$ is the concatenation operator. The trainable parameters here are the $M \times (N + M)$ weight matrix W_f, and the bias vectors b_f.

For this recurrent module, the input i_t is learned from a vector Ck_t made by a concatenated set of covariances, and the previous state h_{t-1}. This input vector is then expressed as $i_t = \sigma(W_i[Ck_t, h_{t-1}] + b_i)$, which thereafter will be used to weight the internal memory state m_t. Then, the memory update $\tilde{m}_t = \tanh(W_m[Ck_t, h_{t-1}] + b_m)$ selects the information for the internal state. Finally, the output is computed as $o_t = \sigma(W_o[Ck_t, h_{t-1}] + b_o)$, which forms a compact descriptor of the spatio-temporal information of the covariance set. Likewise, the learned parameters are represented by the $M \times (N + M)$ weight matrices W_i, W_m, and W_o and the bias vectors b_i, b_m and b_o.

In summary, the temporal and recurrent motion description has the capability to detect and model time intervals correlated with multimodal parkinsonian patterns. For each motion modality was then built a recurrent deep model that learn complex temporal relationships from video manifold form by the set of frame-covariances. Despite the gait and eye fixational recordings are not synchronized, the proposed strategy take advantage of eye oscillatory behaviour ranging in short time from (2 ms to 20 ms) and on a long time scale (100 to 400 ms).[28] Under this premise both modalities can easily aligned and the resultant recurrent embedding descriptors can be fussed to enhance PD description. In this way, the fused model temporally quantifies the probability that depends on the execution of the patient's movement in each modality, which potentially helps to identify postures prone to correction in physiotherapeutic therapies or in medical follow-up.

2.4 Data

A total of 25 participants was included in this study: 13 control subjects (average age of 72.2 ± 6.1) and 12 PD patients (average age of 72.3 ± 7.4). The PD patients were diagnosed in second (5 patients) or third (7 patients) stage of the disease by a physician following the Hoehn-Yahr scale. This study was approved by the Ethics Committee of Universidad Industrial de Santander and a written informed consent was obtained. For eye fixational recording, the patients observed a fixed spotlight projected on a screen with a dark background, with an average duration of 4 seconds. The eye region was manually cropped (210×140 pixels) to obtain the sequences of interest. Regarding walking, markerless sagittal-plane videos were recorded with a spatial resolution of 1280×720 pixels and a temporal resolution of 60 fps. The locomotion was recorded during a 5 meter displacement, for an average duration of 4 seconds. For each participant, 6 videos for gait and 6 videos for each eye were recorded with a conventional camera, Nikon D3200 with spatial resolution of 1280×720, resulting in a total dataset of 450 videos.

2.5 Experimental setup

A leave-one-patient-out cross-validation was carried out to evaluate the proposed approach, i.e. at each iteration on the evaluation, one patient is left out to test and the remaining ones are used for training. For the whole dataset of RGB videos was computed the Farnebäck optical flow (with 5 scales and a 3×3 pixels averaging window). Then, each frame of the videos was sent to the second layer of a pre-trained MobileNet V2 that counts with a total of 32 learned filters. The resulting deep features have a spatial size of (112×112). Then, for each frame a spatial covariance matrix is computed to summarize deep feature correlations, resulting in matrices with size of 32×32. Each video is herein represented by a total of 200 frame-level covariance matrices. The LSTM outputs for the two channels are concatenated and fully connected with a softmax layer to perform the frame-level prediction. The training was performed by following a categorical cross-entropy with Adam optimizer and

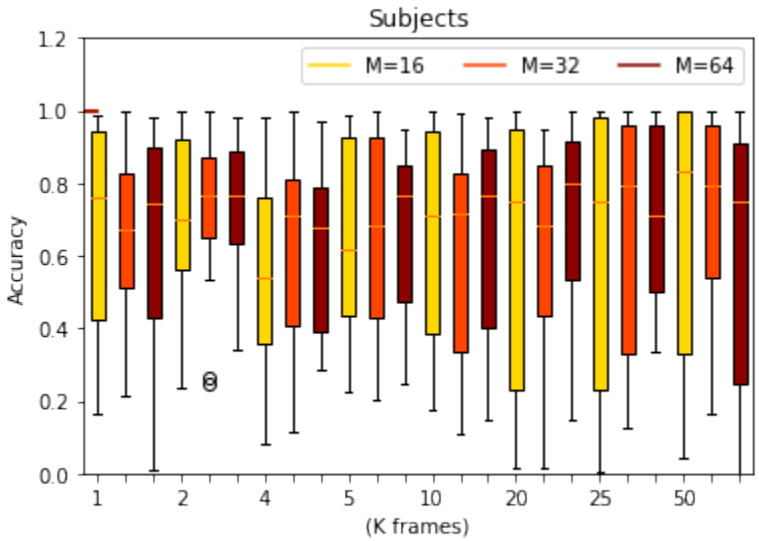

Figure 2. Mean classification accuracy as a function of the number k of covariance matrices, and the number M of cells in LSTM layer

learning rate that varies from 0.0001 to 0.001 with a step size of 0.0001, and including an early stopping strategy with categorical cross-entropy in 15 epochs.

3. EVALUATION AND RESULTS

A first evaluation was carried out to determine the best temporal interval to capture Parkinsonian motion patterns from bi-modal inputs. Figure 2 shows the performance achieved from different per-frame covariances. This plot also summarizes the performance using different hidden recurrent vector sizes of 16, 32 and 64, respectively. The window intervals with relative small input vectors (number of covariance matrices k from 1 to 5) shows less variable results, achieving a proper integration of both modalities in very short time intervals (less than 100 ms). Particularly, the integration of gait and eye fixation with $k = 2$ and 32 cells obtained the best results with an average accuracy of 0.78. This result may be related to the fact that eye fixation movements at short time scales are more persistent.[28] The following experiments will then follow the validation approach with the best configuration, *i.e,* two frame-covariances with a recurrent cell of 32 units.

Secondly, we analyze the contribution of multimodal approach with respect to isolated modalities. To obtain a global classification score for each video, a majority voting is implemented from the per-frame predictions achieved in this proposed scheme. Table 1 summarizes the classification achieved from gait and eye isolated inputs, but also, by considering a late fusion of both modalities. As expected the fusion modality has a better performance to discriminate among control and Parkinson disease at level two and three, labelled according to Hoehn and Yahr scale. In other cases, the eye fixation has a better performance to discriminate third level of Parkinson, but the fusion of modalities overcomes the discrimination of patterns.

Table 2 reports a more detailed analysis including statistical metrics such as precision (prec), recall (rec), specificity (spec), F1-score (F1-s), and Mathews correlation coefficient (MCC). The multimodal fusion alternative achieves the best performance with a predominant F1-score (average of 0.83) and MCC (average of 0.74) with respect to single modalities. The eye fixation achieves a remarkable score to discriminate the Parkinsonian population (Recall = 1) but reports a low sensitivity to differentiate between considered Parkinson stages. Accordingly, the Gait has a low classification rate between Parkinson stages and a low MCC, even for control population. In contrast, learning from both modalities the proposed approach exploits the best embedding correlations, which improves the classification performance.

Table 1. Confusion matrices for each modality and for fusion approach.

	GAIT			OCULAR FIXATION			MERGE		
	C	**S2**	**S3**	**C**	**S2**	**S3**	**C**	**S2**	**S3**
C	**13(100%)**	0	0	**13(100%)**	0	0	**13(100%)**	0	0
S2	3(60%)	**2(40%)**	0	3 (60%)	**2(40%)**	0	0	**4(80%)**	1(20%)
S3	4(57%)	2(28%)	**1(14%)**	3(43%)	0	**4(57%)**	0	2(28%)	**5(71%)**

Modalities		prec	rec	spec	F1-s	MCC
GAIT	**C**	**1**	0.65	**1**	0.79	0.52
	S2	0.40	0.50	0.86	0.44	0.33
	S3	0.40	**1**	0.75	0.24	0.33
OCULAR FIXATION	**C**	**1**	0.68	**1**	0.81	0.58
	S2	0.40	**1**	0.87	0.57	0.65
	S3	0.57	**1**	0.86	0.72	0.70
FUSION	**C**	**1**	**1**	**1**	**1**	**1**
	S2	0.80	0.67	0.95	0.73	0.57
	S3	**1**	0.68	**1**	0.76	0.66

Table 2. Scores for two modalities and fusion approach.

A major advantage of the proposed strategy is the capability to output classification scores over time, which may enrich evaluation and analysis of disease progression. Figure 3 summarizes the temporal prediction achieved for the three different population classes (13 Control, 5 in PD stage 2, 7 in PD stage 3). In y-axis is plotted the control probability as reference of whole videos, while in x-axis is arranged information according to main gait phases, *i.e.,* first stance (1 St), second stance (2 St), first swing (1 Sw) and second swing (2 Sw). The stance phases (1,2 St) include the heel-to-toe contact sequence of the foot, while the swing phases (1,2 Sw) proceed with the foot suspended in the air.[29]

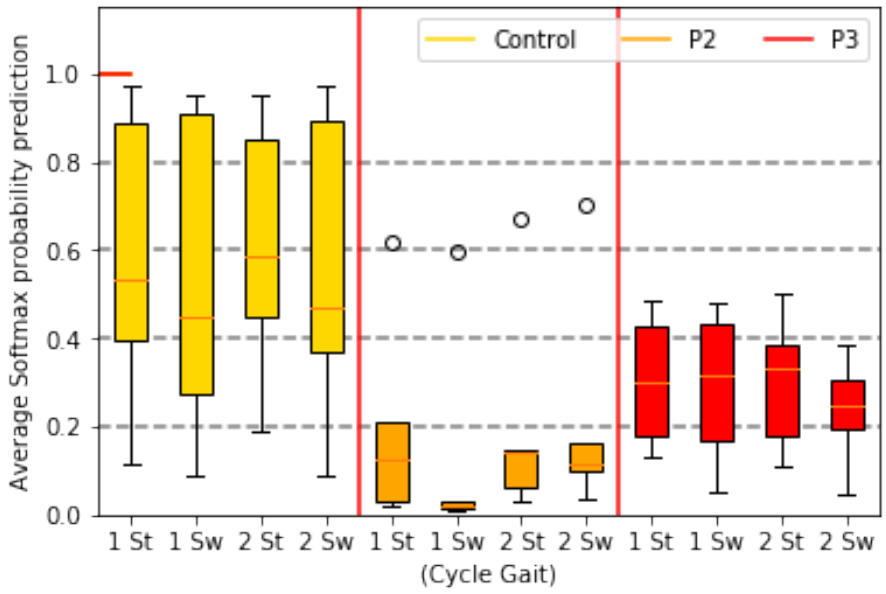

Figure 3. Average Softmax probability prediction for control reference, in function of the gait cycle.

Interestingly, the patients in stage two have lower probabilities to belong to control group, with a marked low variability in the reported predictions and reporting an outlier that corresponds to a patient with additional artifacts during the capture. It should be noted also that major discrimination in Parkinson between the two stages was in one of the swing phases, which may be related with major postural instability, a clear biomarker of the disease. For control there exist large variations for all phases, but with a clear tendency to larger probabilities. These temporal (partial) measures can be affected by a small number of observations during online predictions and therefore they should be contrasted with the final video classification.

Anyway this method has the potential to quantify and support disease diagnosis not only from a global video perspective but also discriminating among different gait phases, which are intrinsically combined with eye fixational behaviours. The identification of these poses and temporal alterations can provide tools to adapt the personalized physiotherapy technique according to the patient's needs.

4. CONCLUSIONS

This work introduces an online fusion strategy that integrates Gait and eye fixational patterns to classify Parkinson disease at two different stages and with respect to control population. The proposed approach uses a compact encoding of frame-level covariance of deep features calculated on dense optical flow images, that allows a rich primary representation of postural configurations, as well as, local motion velocity patterns. This statistical representation is then mapped to a recurrent LSTM network to learn motion temporal patterns, with the major advantage to produce online predictions of the disease. The gait patterns bring global descriptors while the eye fixational patterns can contribute to a major sensitivity to discriminate among disease stages. The results show a robust performance for global video prediction in the considered study of 25 patients and a total of 450 video sequences. The frame-level prediction shows remarkable correlation with the disease during swing phases. Nonetheless, some mistakes were reported on Parkinson stages, which may be associated to insufficient sensitivity of the descriptors, but also to expert annotation biases. The proposed approach results very compact, which brings opportunities to support observational analysis on-the-fly, allowing to suggest (spatial) regions, and (temporal) segments with major association to the disease. Reciprocally, the physiotherapist can design customized routines for the improvement of the most critical gait movements. Future works include a study that involves additional expert observations, larger number of patients and a more detailed analysis about time prediction. Also, in future works will be included additional description of side dependent (e.g. unilateral) symptom patterns, and any other conditions that may be explained into the developed study.

ACKNOWLEDGMENTS

The authors acknowledge the Vicerrectoriá de Investigación y Extensión (VIE) of the Universidad Industrial de Santander for supporting this research work by the project: *Cuantificación de patrones locomotores para el diagnóstico y seguimiento remoto en zonas de dificil acceso*, with SIVIE code 2697.

REFERENCES

[1] Feigin, V. L., Nichols, E., Alam, T., Bannick, M. S., Beghi, E., Blake, N., Culpepper, W. J., Dorsey, E. R., Elbaz, A., Ellenbogen, R. G., et al., "Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016," *The Lancet Neurology* **18**(5), 459–480 (2019).

[2] Moustafa, A. A., Chakravarthy, S., Phillips, J. R., Gupta, A., Keri, S., Polner, B., Frank, M. J., and Jahanshahi, M., "Motor symptoms in parkinson's disease: A unified framework," *Neuroscience & Biobehavioral Reviews* **68**, 727–740 (2016).

[3] Post, B., Merkus, M. P., de Bie, R. M., de Haan, R. J., and Speelman, J. D., "Unified parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?," *Movement disorders: official journal of the Movement Disorder Society* **20**(12), 1577–1584 (2005).

[4] Bennett, D. A., Shannon, K., Beckett, L. A., Goetz, C., and Wilson, R., "Metric properties of nurses' ratings of parkinsonian signs with a modified unified parkinson's disease rating scale," *Neurology* **49**(6), 1580–1587 (1997).

[5] Rizzo, G., Copetti, M., Arcuti, S., Martino, D., Fontana, A., and Logroscino, G., "Accuracy of clinical diagnosis of parkinson disease: a systematic review and meta-analysis," *Neurology* **86**(6), 566–576 (2016).

[6] Ortells, J., Herrero-Ezquerro, M. T., and Mollineda, R. A., "Vision-based gait impairment analysis for aided diagnosis," *Medical & biological engineering & computing* **56**(9), 1553–1564 (2018).

[7] Sato, K., Nagashima, Y., Mano, T., Iwata, A., and Toda, T., "Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning–based 2d pose estimator," *PloS one* **14**(11), e0223549 (2019).

[8] Pham, H. N., Do, T. T., Chan, K. Y. J., Sen, G., Han, A. Y., Lim, P., Cheng, T. S. L., Nguyen, Q. H., Nguyen, B. P., and Chua, M. C., "Multimodal detection of parkinson disease based on vocal and improved spiral test," in [*2019 International Conference on System Science and Engineering (ICSSE)*], 279–284, IEEE (2019).

[9] Vasquez, J., Bocklet, T., Orozco, J. R., and Noeth, E., "Comparison of user models based on gmm-ubm and i-vectors for speech, handwriting, and gait assessment of parkinson's disease patients," 6544–6548 (05 2020).

[10] Vásquez-Correa, J. C., Arias-Vergara, T., Orozco-Arroyave, J. R., Eskofier, B., Klucken, J., and Nöth, E., "Multimodal assessment of parkinson's disease: A deep learning approach," *IEEE Journal of Biomedical and Health Informatics* **23**(4), 1618–1630 (2019).

[11] Jones, R., "Biomarkers: casting the net wide," *Nature* **466**(7310), S11–S12 (2010).

[12] Anderson, T. J. and MacAskill, M. R., "Eye movements in patients with neurodegenerative disorders," *Nature Reviews Neurology* **9**(2), 74–85 (2013).

[13] Ekker, M. S., Janssen, S., Seppi, K., Poewe, W., de Vries, N. M., Theelen, T., Nonnekes, J., and Bloem, B. R., "Ocular and visual disorders in parkinson's disease: common but frequently overlooked," *Parkinsonism & related disorders* **40**, 1–10 (2017).

[14] Ekker, M. S., Janssen, S., Seppi, K., Poewe, W., de Vries, N. M., Theelen, T., Nonnekes, J., and Bloem, B. R., "Ocular and visual disorders in Parkinson's disease: Common but frequently overlooked," *Parkinsonism & Related Disorders* **40**, 1–10 (2017).

[15] Clarke, A. H., "Laboratory testing of the vestibular system," *Current opinion in otolaryngology & head and neck surgery* **18**(5), 425–430 (2010).

[16] Khosla, A. and Kim, D., [*Optical Imaging Devices: New Technologies and Applications*], CRC Press (2017).

[17] Hanuška, J., Bonnet, C., Rusz, J., Sieger, T., Jech, R., Rivaud-Péchoux, S., Vidailhet, M., Gaymard, B., and Ržička, E., "Fast vergence eye movements are disrupted in parkinson's disease: a video-oculography study," *Parkinsonism & Related Disorders* **21**(7), 797–799 (2015).

[18] Váradi, C., "Clinical features of parkinson's disease: The evolution of critical symptoms," *Biology* **9**(5) (2020).

[19] Alper, M. A., Goudreau, J., and Daniel, M., "Pose and optical flow fusion (poff) for accurate tremor detection and quantification," *Biocybernetics and Biomedical Engineering* **40**(1), 468–481 (2020).

[20] Bashir, K., Xiang, T., Gong, S., and Mary, Q., "Gait representation using flow fields.," in [*BMVC*], 1–11 (2009).

[21] Farnebäck, G., "Two-frame motion estimation based on polynomial expansion," in [*Scandinavian conference on Image analysis*], 363–370, Springer (2003).

[22] Sifre, L. and Mallat, S., "Rigid-motion scattering for image classification," *Ph. D. thesis* (2014).

[23] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C., "Mobilenetv2: Inverted residuals and linear bottlenecks," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 4510–4520 (2018).

[24] [*Towards Data-Driven Modeling of Pathological Tremors*], *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* **Volume 2: 16th International Conference on Multibody Systems, Nonlinear Dynamics, and Control (MSNDC)** (08 2020). V002T02A030.

[25] Liu, X., Li, W., Liu, Z., Du, F., and Zou, Q., "A dual-branch model for diagnosis of parkinson's disease based on the independent and joint features of the left and right gait," *Applied Intelligence* , 1–12 (2021).

[26] Tabarestani, S., Aghili, M., Shojaie, M., Freytes, C., Cabrerizo, M., Barreto, A., Rishe, N., Curiel, R. E., Loewenstein, D., Duara, R., et al., "Longitudinal prediction modeling of alzheimer disease using recurrent neural networks," in [*2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*], 1–4, IEEE (2019).

[27] Zhao, A., Qi, L., Dong, J., and Yu, H., "Dual channel lstm based multi-feature extraction in gait for diagnosis of neurodegenerative diseases," *Knowledge-Based Systems* **145**, 91–97 (2018).

[28] Herrmann, C. J., Metzler, R., and Engbert, R., "A self-avoiding walk with neural delays as a model of fixational eye movements," *Scientific reports* **7**(1), 1–17 (2017).

[29] Cicirelli, G., Impedovo, D., Dentamaro, V., Marani, R., Pirlo, G., and D'Orazio, T., "Human gait analysis in neurodegenerative diseases: a review," *IEEE Journal of Biomedical and Health Informatics* (2021).