Parkinsonism and Related Disorders xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Automated assessment of levodopa-induced dyskinesia: Evaluating the responsiveness of video-based features

Michael H. Li[a,b], Tiago A. Mestre[c,d,e,f], Susan H. Fox[c,f], Babak Taati[a,b,g,*]

[a] Toronto Rehabilitation Institute, University Health Network, 550 University Ave, Toronto, Ontario, M5G 2A2, Canada
[b] Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College St, Room 407, Toronto, Ontario, M5S 3G9, Canada
[c] Edmond J. Safra Program in Parkinson's Disease, Toronto Western Hospital, University Health Network, 399 Bathurst St, Toronto, Ontario, M5T 2S8, Canada
[d] The Ottawa Hospital Research Institute, 1053 Carling Ave, Ottawa, Ontario, K1Y 4E9, Canada
[e] Division of Neurology, Department of Medicine, 1053 Carling Ave, Ottawa, Ontario, K1Y 4E9, Canada
[f] Division of Neurology, University of Toronto, Suite RFE 3-805, 200 Elizabeth St, Toronto, Ontario, M5G 2C4, Canada
[g] Department of Computer Science, University of Toronto, 10 King's College Road, Room 3302, Toronto, Ontario, M5S 3G4, Canada

ARTICLE INFO

Keywords:
Parkinson's disease
Levodopa-induced dyskinesia
Computer vision
Clinimetric testing
Objective assessment

ABSTRACT

Introduction: Technological solutions for quantifying Parkinson's disease (PD) symptoms may provide an objective means to track response to treatment, including side effects such as levodopa-induced dyskinesia. Vision-based systems are advantageous as they do not require physical contact with the body and have minimal instrumentation compared to wearables. We have developed a vision-based system to quantify a change in dyskinesia as reported by patients using 2D videos of clinical assessments during acute levodopa infusions.

Methods: Nine participants with PD completed a total of 16 levodopa infusions, where they were asked to report important changes in dyskinesia (i.e. onset and remission). Participants were simultaneously rated using the UDysRS Part III (from video recordings analyzed post-hoc). Body joint positions and movements were tracked using a state-of-the-art deep learning pose estimation algorithm applied to the videos. 416 features (e.g. kinematics, frequency distribution) were extracted to characterize movements. The sensitivity and specificity of each feature to patient-reported changes in dyskinesia severity was computed and compared with physician-rated results.

Results: Features achieved similar or superior performance to the UDysRS for detecting the onset and remission of dyskinesia. The best AUC for detecting onset of dyskinesia was 0.822 and for remission of dyskinesia was 0.958, compared to 0.826 and 0.802 for the UDysRS.

Conclusions: Video-based features may provide an objective means of quantifying the severity of levodopa-induced dyskinesia, and have responsiveness as good or better than the clinically-rated UDysRS. The results demonstrate encouraging evidence for future integration of video-based technology into clinical research and eventually clinical practice.

1. Introduction

The current ability to measure levodopa-induced dyskinesia severity and change in response to interventions in Parkinson's disease (PD) relies on the use of clinical rating scales. Clinicians use rating scales such as the Unified Dyskinesia Rating Scale (UDysRS) that are based on characteristics including duration, anatomical distribution, and functional impact of levodopa-induced dyskinesia. However, such rating scales, although optimized and refined, are still inherently subjective and can be significantly influenced by rater experience [1]. Furthermore, interpretation of symptoms can differ widely between patients and physicians [2]. Technologies for measuring symptoms can provide an objective means of evaluating dyskinesia. Wearable sensing has been the most popular, with multiple studies demonstrating the ability to predict clinical ratings with good accuracy [3–5]. While wearables can be implemented in a discreet and wireless fashion, they require the devices to be physically attached to the user, which can be inconvenient and especially difficult to set up for individuals with PD. Monitoring of multiple limbs also requires additional instrumentation to capture all relevant movements. In contrast, video-based methods combined with computer vision algorithms can monitor several body parts using a single camera sensor, and accomplish monitoring without contact. The computer vision field has seen significant improvements in accuracy with the emergence of deep learning, allowing computational models to

* Corresponding author. Toronto Rehabilitation Institute, University Health Network, 550 University Ave, Toronto, Ontario, M5G 2A2, Canada.
E-mail addresses: michaelhg.li@alum.utoronto.ca (M.H. Li), tmestre@toh.ca (T.A. Mestre), susan.fox@uhnresearch.ca (S.H. Fox), babak.taati@uhn.ca (B. Taati).

https://doi.org/10.1016/j.parkreldis.2018.04.036
Received 31 December 2017; Received in revised form 11 April 2018; Accepted 28 April 2018

be built with substantially more representational power [6]. A problem of interest in computer vision is human pose estimation, where 2D human pose can be estimated from a single image. In our previous study, a deep learning pose estimation algorithm was used to extract motions from videos of PD assessments [7]. Features of the motion trajectories were used to predict involuntary movement severity on the UDysRS Part III (Impairment) [8].

Comparing predictions to ground truth clinical ratings provides assurance that the sensor-based technology is capturing useful information. However, this method of clinimetric validation can be criticized as it does not take advantage of the improved sensitivity that sensors have over subjective human raters [9,10]. Unfortunately, a systematic review of monitoring technologies for PD was only able to recommend nine of 73 screened devices based on sufficient testing and clinimetric validation [11]. None of the recommended devices were video-based. The only study of a video-based technology that was listed for dyskinesia assessment was by Rao et al., who developed a severity score based on movement amplitude and direction [12]. While they showed moderate concurrent validity with UDysRS ratings, they did not detail any exploratory feature analysis in score development. Some of the video-based devices in the review analyzed the reliability of their systems in addition to validity; however, none of the studies attempted to measure responsiveness to treatment.

The goal of this study is to evaluate if features extracted from 2D videos of clinical assessments are sensitive to patient reported changes in dyskinesia. We have previously calculated the clinically important change (CIC) in the UDysRS Part III using a levodopa infusion protocol and an anchor-based approach [13]. In this study, the sensitivity and specificity-based approach from our prior study is replicated with objective video-based features to assess their responsiveness to treatment. Results show that video-based features are a promising complement to clinical rating scales and warrant further investigation.

2. Methods

Information on study population and protocol is identical to [13], which generated the current dataset.

In summary, inclusion criteria were: diagnosis of idiopathic PD as per United Kingdom PD Society Brain Bank criteria [14]; 30–80 years of age; stable bothersome levodopa-induced peak-dose dyskinesia for more than 25% of the day, defined as a UPDRS 4.1 rating ≥ 2 and Lang-Fahn score ≥ 1, and on stable antiparkinsonian medication for at least one month in advance of study participation.

The exclusion criteria was: a Hoehn & Yahr score of 5 in “off” state; UPDRS rating ≥ 3 for resting or action tremor when “off”; cognitive impairment, defined as a score < 24 on the Montreal Cognitive Assessment [15]; and previous surgery for PD.

Data was collected in a randomized, double-blind, placebo-controlled crossover study, with a total of 4 visits over 7 $\pm$ 2 weeks. The first visit was used for screening. Afterwards, there was a run-in, unblinded levodopa infusion visit designed to familiarize participants with changes in dyskinesia severity and to reduce placebo effect. For visits 3 and 4, subjects were randomized to receive levodopa or placebo. All three infusion visits were performed according to the classical intravenous (i.v.) levodopa infusion paradigm [16], with a washout period of 1–2 weeks. Levodopa infusion rate was between 1.0 and 1.5 mg/kg/hr depending on daily oral levodopa-equivalent dose.

A video protocol, including items required to assess UDysRS Part III, was administered at regular intervals during each infusion visit. An exception was the dressing task, which was excluded due to the infusion tubing. Participants omitted their PD medication for 12 h in advance of the visit such that they would begin the protocol in the “practically-defined off” state. The levodopa/placebo infusion lasted for 2 h, followed by up to 2 h of post-infusion observation. Three clinically important events were established as anchors: onset, maximum intensity, and remission of dyskinesia. Participants were queried every 15 min to determine if they were experiencing one of the anchors (“patient-reported”). The video protocol was performed every 30 min and when anchors were reported. UDysRS Part III ratings were performed post-hoc by three neurologists blinded to the visit and infusion time elapsed (“physician-rated”). Ratings were averaged to produce a single score.

Videos were recorded with a consumer grade video camera at a resolution of 480 $\times$ 640 or 540 $\times$ 960 and 30 frames per second. The study protocol was approved by the University Health Network Research Ethics Board and written informed consent was provided by all participants.

The UDysRS includes tasks such as communication, drinking from a cup, and ambulation. Previous analysis of the dataset indicated that video-based features from the communication task were most predictive of dyskinesia severity [7,8]. Therefore, only features computed from the communication task were used. Joint positions were extracted from assessment videos using Convolutional Pose Machines (CPM), a state-of-the-art human pose estimation algorithm [17]. CPM produced a 14-point skeleton annotating the location of the head, neck, shoulders (Lsho, Rsho), elbows (Lelb, Relb), wrists (Lwri, Rwri), hips (Lhip, Rhip), knees (Lkne, Rkne), and ankles (Lank, Rank) for each frame. Frame-by-frame pose estimates were combined over time to form joint trajectories.

Two pre-processing steps were taken to remove noise from trajectories before feature extraction. Camera movement was determined by detecting and tracking distinctive points in the background of the video. Joint trajectories were stabilized by subtracting the camera movement signal. Since pose estimates were produced individually for each frame, large discontinuities could occur in joint trajectories if estimation was temporarily poor. To identify discontinuities, frame-to-frame distance was thresholded, and only segments containing poses CPM estimated with high confidence (where the confidence was an output of CPM) were kept. Gaps between segments were linearly interpolated.

The skeleton produced by CPM included a head annotation, but it was not suitable for tracking head turning as it was positioned at the top of the head. Instead, the head and neck annotations were used to initialize a bounding box on the face, which was tracked using MEEM, a general purpose object tracker [18]. The face trajectory (Face) replaced the head and neck trajectories, leaving a total of 13 trajectories. Additional information about preprocessing can be found in Ref. [8].

A total of 416 features were extracted, with 32 features extracted for each joint trajectory. To ensure that features were comparable across different participants and camera distances, all trajectories were normalized by dividing by the respective participants' head length (i.e. distance between the head and neck annotations). There were 15 kinematic features, which were summary statistics, i.e. maximum, median, mean, standard deviation (stdev), interquartile range (IQR), of the speed, acceleration, and jerk. There were also 16 spectral features that were computed from the power spectral density (PSD) of the displacement (Disp) and velocity (Vel) signals, where the PSD described the distribution of frequencies in the signal and was estimated using Welch's periodogram. The spectral features were the peak magnitude, entropy, total power, half point (the frequency dividing the PSD into equal halves), and relative power in 0.5–1 Hz, > 2 Hz, > 4 Hz, and > 6 Hz power bands for both PSDs. The last feature was the convex hull, which was a measure of the area that the joint moved within. As the communication task was made of multiple parts (e.g. counting backwards, conversation), the video was split, resulting in 2–4 shorter clips. Features were computed in each clip and averaged to produce the video's final feature set. The average length of a video was 31 s.

Differences in feature values were computed between time points. To assess responsiveness, it must be determined whether features can detect changes in patient-reported dyskinesia severity. Therefore, two key differences were calculated: change between the start of the infusion (“off” state) and reported onset of dyskinesia, and change between reported onset and remission of dyskinesia. The difference between non-anchor time points and the preceding anchor were considered “no