A Video-Based Method for Automatically Rating Ataxia

Ronnachai Jaroensri[*]   TIAM@MIT.EDU
Massachusetts Institute of Technology, Cambridge, MA

Amy Zhao[*]   XAMYZHAO@MIT.EDU
Massachusetts Institute of Technology, Cambridge, MA

Guha Balakrishnan   BALAKG@MIT.EDU
Massachusetts Institute of Technology, Cambridge, MA

Derek Lo   DEREKTLO@GMAIL.COM
Yale University, New Haven, CT

Jeremy D. Schmahmann, M.D.   JSCHMAHMANN@MGH.HARVARD.EDU
Massachusetts General Hospital, Boston, MA

Frédo Durand   FREDO@MIT.EDU
Massachusetts Institute of Technology, Cambridge, MA

John Guttag   GUTTAG@MIT.EDU
Massachusetts Institute of Technology, Cambridge, MA

Abstract

For many movement disorders, such as Parkinson's disease and ataxia, disease progression is visually assessed by a clinician using a numerical disease rating scale. These tests are subjective, time-consuming, and must be administered by a professional. This can be problematic where specialists are not available, or when a patient is not consistently evaluated by the same clinician. We present an automated method for quantifying the severity of motion impairment in patients with ataxia, using only video recordings. We consider videos of the finger-to-nose test, a common movement task used as part of the assessment of ataxia progression during the course of routine clinical checkups.

Our method uses neural network-based pose estimation and optical flow techniques to track the motion of the patient's hand in a video recording. We extract features that describe qualities of the motion such as speed and variation in performance. Using labels provided by an expert clinician, we train a supervised learning model that predicts severity according to the Brief Ataxia Rating Scale (BARS). The performance of our system is comparable to that of a group of ataxia specialists in terms of mean error and correlation, and our system's predictions were consistently within the range of inter-rater variability. This work demonstrates the feasibility of using computer vision and machine learning to produce consistent and clinically useful measures of motor impairment.

1. Introduction

Tens of millions of people are affected by movement disorders in the US and Europe alone, and this number is projected to double in the next few decades (Bach et al., 2011). Quantifying the severity of motor incapacity is useful in monitoring the progression of these diseases and measur-

---

[*]. These authors contributed equally to this work.

ing the effectiveness of treatments. Estimates of motor incapacity are most commonly made using questionnares, or visual assessments combined with numerical rating scales (Goetz et al., 2008; Schmahmann et al., 2009; Schmitz-Hübsch et al., 2006). While these rating scales have been shown to be useful, the tests must be administered by an experienced clinician and are subjective. An automated, objective measurement of severity would provide more consistent evaluations, particularly where specialists are not available.

We use computer vision and machine learning techniques to create an automated, video-based method for quantifying the severity of motor impairment in patients with ataxia. The term *ataxia* describes a heterogenous group of neurodegenerative diseases characterized by gait incoordination, involuntary extremity and eye movements, and difficulty in articulating speech (Klockgether, 2010). The severity of ataxia is typically assessed using motor function tests such as the finger-to-nose maneuver, in which a patient alternates between touching his/her nose and the clinician's outstretched finger. A neurologist observes the patient's action and rates the disease severity on a numeric scale such as the Brief Ataxia Rating Scale (BARS) (Schmahmann et al., 2009). These rating scales often consider aspects of the patient's movement such as speed, smoothness, and accuracy. This evaluation typically happens during regularly scheduled clinical visits, and is time-consuming for the neurologist. Furthermore, the rating assigned to the patient often varies from rater to rater (Weyer et al., 2007). An automated, consistent method for rating ataxia can greatly alleviate these problems.

Our work focuses on video recordings of a patient performing the finger-to-nose test; such videos might be collected during a routine clinical visit, or even in the patient's home. This poses two main challenges. First, our system must be robust to the issues raised by video quality. Clinical videos are likely to be captured with a handheld camera, and may contain camera movements such as panning and zooming. Different clinical settings could produce variations in viewing angle and lighting conditions. A more fundamental challenge in this task is that the amount of data is limited, as is common in many clinical problems. Consequently, we must rely on machine learning techniques that work with limited amounts of training data.

Our contribution is a video-based system that automatically produces ratings of motor incapacity for ataxic patients. Our system is: (1) observer-independent, (2) as accurate as human raters, who are the current gold standard for such a task, and (3) robust to the quality of videos taken in clinical settings. Our system combines optical flow and neural network-based pose estimation techniques to robustly track the location of the patient's wrist and head in each video. To facilitate training on our small dataset, we designed features based on the motion characteristics described in the BARS rating criteria. We extract these features from the wrist location signal, and use them in a learning algorithm to build a model that predicts the BARS severity rating of the patient's action. We selected a representative subset of our data to be rated by a group of experienced ataxia specialists, and found that our system performed comparably to the specialists.

This work demonstrates the feasibility of an automated method for assessing the severity of motor impairment in ataxic patients. Such a system could be useful in clinical or even home settings by allowing more frequent and more consistent assessments of the disease, or in clinical trials where consistency is required. The system would be particularly useful in areas where a neurologist specializing in ataxia is not available, which is the case in most parts of the country and the rest of the world.

2. Related Work

2.1 Ataxia Rating Scales

Ataxias are a group of neurodegenerative movement disorders characterized by incoordination of the extremities, eyes and gait (Schmahmann, 2004; Klockgether, 2010). Since their introduction, quantitative rating scales have been the norm for ataxia severity assessment. Examples of these scales include the International Cooperative Ataxia Rating scale (ICARS), and the Scale for the Assessment and Rating of Ataxia (SARA) (Trouillas et al., 1997; Schmitz-Hübsch et al., 2006). These scales require an expert clinician to visually assess the qualities of the patient's movements and determine a numerical rating, a time-consuming process. In 2009, the Brief Ataxia Rating Scale (BARS) was developed as a quantitative scale that is sufficiently fast and accurate for clinical purposes. Nonetheless, as the designer of the BARS says, an observer-independent, fine-grained method for assessing ataxia is still "sorely needed" (Schmahmann et al., 2009). This work is the first step towards such a method.

2.2 Human Motion Analysis

The use of camera systems to measure and detect pathological movement is well-studied for many diseases (Sutherland, 2002; Galna et al., 2014). These systems typically utilize multiple camera setups along with passive or active markers and other sensing methods such as electromyography to track the motion of the subject (Muro-de-la Herran et al., 2014; Kugler et al., 2013; Lizama et al., 2016). Such camera setups are expensive and complicated, must be operated by expert technicians, and often require expert interpretation. These requirements prevent the use of these systems in clinical settings. In this work, we focus on monocular consumer-quality videos, since their ubiquity could enable widespread applications of our technique in both clinical and home settings.

To the best of our knowledge, this is the first work that uses monocular video recordings to automatically assess the severity of a neurological movement disorder. Most existing approaches rely on more specialized hardware, and solve different problems: they measure movements but do not produce a rating of disease severity (Galna et al., 2014), or they focus on differentiating healthy and impaired patients (Fazio et al., 2013; Weiss et al., 2011). One exception is the work by Giuffrida et al., which produces a Unified Parkinson's Disease Rating Scale rating from inertial sensor data (Giuffrida et al., 2009; Goetz et al., 2008).

2.2.1 HUMAN MOTION ANALYSIS IN MONOCULAR VIDEOS

Human pose estimation aims to localize a human's joints in images or videos. Human analysis in monocular videos often uses pose estimation as the initial step. State-of-the-art pose estimation systems are effective at localizing joints even when they are occluded by objects or other parts of the body. We use a publicly available implementation of a convolutional neural network-based pose estimator to track the wrist of the patient in our videos (Wei et al., 2016).

Motion analysis in monocular videos often also relies on optical flow. Optical flow measures the relative motion between the pixels in one frame and another. Optical flow algorithms can be divided two categories: sparse and dense. Sparse flow methods track only a sparse set of salient points in the video, while dense flow methods measure motion at every pixel in the frame. We use sparse optical flow to track background points for video stabilization (Lucas et al., 1981), and use dense optical

flow (Brox et al., 2009) to refine the output of our pose estimator and to improve the tracking of the patient's wrist in each video.

Our problem is similar to that of automatic human motion quality assessment. Pirsiavash et al. (2014) use features based on the discrete cosine transform (DCT) to estimate the judges' score for Olympic divers and figure skaters. Venkataraman et al. (2015) applied approximate entropy (ApEn) to the same task and observed slight improvements. Venkataraman et al. (2013) use shape-based dynamical features to quantify levels of impairment of stroke survivors. These methods are advantageous because they do not rely on domain-specific knowledge, but we found them to be insufficient for capturing the relevant motion characteristics in our videos.

3. The BARS Dataset

We use a dataset of videos of ataxic patients to train a machine learning model that estimates the severity of motor impairment. Our dataset consists of 90 videos showing 43 distinct subjects performing several repetitions of the finger-to-nose test with one hand at a time (Figure 1a). The videos were shot during clinic visits with handheld cameras. All videos were labeled according to a version of the BARS that uses half-point increments by the physician who created the scale. We treat these labels as the gold standard on which we train our model. The scale ranges from 0 (no impairment) to 4 (so impaired as to be unable to complete the test), and the rating criteria includes motion characteristics such as instability of the finger or disjointed movements. Figure 1b shows the distribution of the severity ratings for the 90 videos. For several patients, there are two videos for each hand. The scores were assigned separately for each hand, so a patient might have different ratings for their right and left hands.

Some videos contain panning and zooming. We adjust for this by using sparse optical flow to track points in the background, and then estimating a similarity transformation from the tracked points to stabilize each frame (Shi and Tomasi, 1994; Lucas et al., 1981). Because of the low number of salient background points in some of the videos, only 61 videos were successfully stabilized. We processed the remaining videos without stabilization.
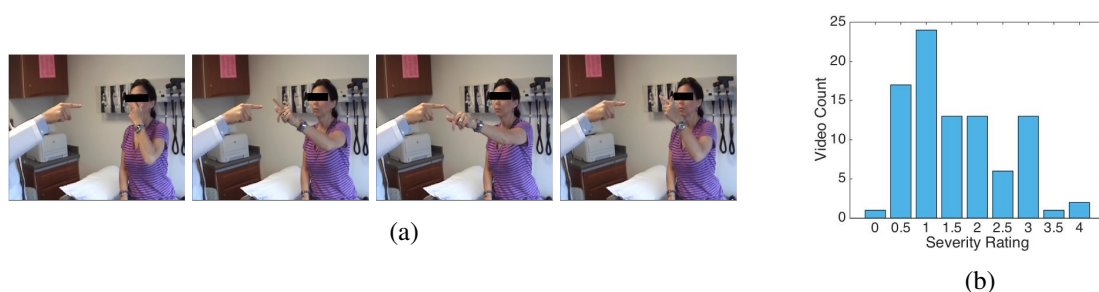

(a)

(b)

Figure 1: (a) A finger-to-nose exam. The patient alternates between touching his/her nose and the clinician's outstretched finger. Each video in our dataset contains at least two repetitions of this action. The eyes are occluded here for anonymity. (b) The distribution of severity ratings in the BARS dataset.

4. Feature Extraction and Prediction Model

We wish to extract features that quantify relevant motion characteristics, including several characteristics described in the BARS guidelines. We extract these features from the motion signal of each patient's active hand. We use pose estimation and optical flow to track the location of the patient's wrist and head, segment the motion signal into cycles, and compute features from the signal segments. This process is summarized in Figure 2, and explained in detail in the following subsections.

4.1 Head and Wrist Tracking
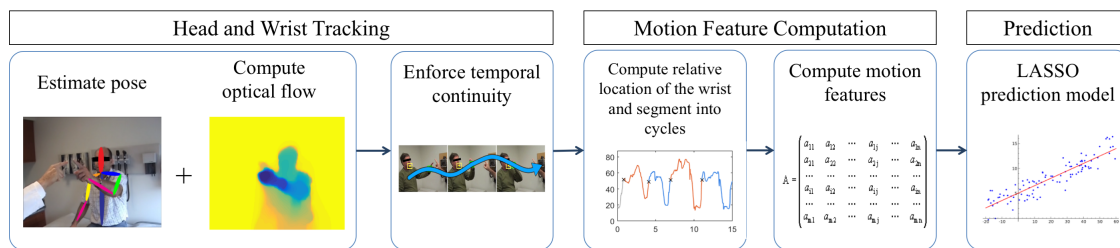
Figure 2: Our algorithm tracks the location of the patient's active wrist, segments the location signal, computes motion features, and then runs a linear regression to predict the BARS rating.

4.1.1 HEAD AND WRIST LOCATION ESTIMATES USING POSE ESTIMATION

We obtain the location of the patient's active hand in each video frame. The quality of the videos, which are representative of what might be captured in practice, make this a difficult task. Most of the videos have low contrast, harsh lighting, and significant motion blur, making local appearance-based tracking techniques such as sparse point tracking (Lucas et al., 1981) or mean-shift (Comaniciu et al., 2000) ineffective. In addition, there is significant variation in viewing angle across videos, causing simple hand-recognition detectors to fail. We instead rely on a more complex pose estimation system that calculates the likely location of each body part in an image by leveraging information about the configuration of the entire body. We start with a state-of-the-art pose estimation system based on convolutional neural networks (Wei et al., 2016). The system is designed to predict the location of several body joints in an image, such as the wrist and the top of the neck (bottom of the head). We use the relative location of the wrist to the top of the neck to approximate the location of the patient's hand relative to his/her head. Note that we do not estimate the location of the doctor's hand, as it is not visible in some videos and cannot be reliably located in others due to low video quality.

Using the pose estimator "out of the box" does not produce accurate wrist location estimates. This is because the videos in our dataset contain poses that are not well-represented in the dataset that the pose estimator was trained on, the MPII Human Pose dataset (Andriluka et al., 2014). To address this, we begin with the model trained on MPII, and further train or *fine-tune* the network on a new annotated dataset of individuals performing the finger-to-nose-test, none of whom appear in the BARS dataset. Our fine-tuning dataset consists of 626 images from 26 distinct individuals; 507 images are taken from videos of 15 healthy individuals, and 119 images are downloaded from internet resources on ataxia. We annotated the locations of the joints in our dataset and fine-tuned
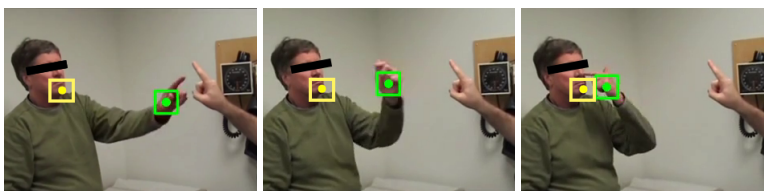
5

Figure 3: An example of our tracking results. The detected wrist position is marked in green, and the bottom of the head is marked in yellow.

the network for $6,000$ iterations with a batch size of 8. This small amount of additional training improved our wrist position estimates considerably; it reduced the Euclidean distance error by 31.9% on a 70-frame hand-annotated test set from the BARS videos.

4.1.2 TEMPORAL REGULARIZATION USING OPTICAL FLOW

The pose estimator we use is designed for images, and does not enforce temporal continuity between neighboring video frames. We enforce this continuity by temporally smoothing the joint estimates for each frame, using the estimates from neighboring frames and dense optical flow, as described in Charles et al. (2014). We further improve the smoothed wrist location estimates by constraining the estimates to fall within the fastest-moving region, which we assume to contain the patient's active hand. We determine these regions by computing dense trajectories from flow spanning multiple frames as described in Sundaram et al. (2010), and then selecting the trajectories with the highest amount of motion over the course of the video. This regularization further improves our tracking by 37.3%. Our final average tracking error is 14.3% of the range of patient's hand motion. An example of our tracking results is shown in Figure 3.
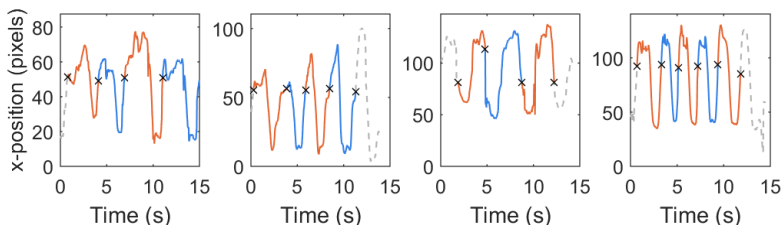

Figure 4: Wrist x-position signals (relative to the head) for four different patients that were each given a severity rating of 2. Notice that the signals differ considerably in shape and frequency. Segments are shown in alternating colors. Portions of the signal that did not fall within a full cycle (in gray) were discarded.

4.2 Cycle Segmentation

Figure 4 shows examples of the wrist x-position signal (relative to head) produced by our tracking algorithm. The videos in our dataset contain varying numbers of repetitions of the finger-to-nose action. To account for this, we assess the motion characteristics of each cycle independently. We first segment the wrist location signal in time. Our automated algorithm takes as input the relative position of the patient's wrist to his/her head, and attempts to segment cycles beginning when the

wrist is halfway between the endpoints of the finger-to-nose action. We approximate the endpoints of the patient's hand motion using the minimum and maximum of the wrist position signal relative to the head. Since the beginning and the end of the videos often include unrelated motions, we only compute the endpoint locations from the middle half of the signal. For robustness against noise around the midpoints of the action, we use hysteresis thresholding, a common thresholding technique used in signal processing, to detect the forward and backward portions of the position signal. We define a cycle to be finger-nose-finger or nose-finger-nose based on which designation produces the higher number of cycles in a video, and exclude any portions of the signal that do not fall within a complete cycle. The segments produced by our cycle segmentation algorithm are displayed in alternating colors in Figure 4.

4.3 Motion Features

From these wrist location signal segments, we extract features that describe characteristics of the patients' motion. To facilitate training on our relatively small dataset, we designed these features based on the motion characteristics described in the BARS guidelines for rating the finger-to-nose test in half-point increments. The total dimensionality is 14.

4.3.1 AVERAGE CYCLE DURATION

As described in the BARS guidelines, healthy patients are typically able to complete each cycle of the finger-to-nose test more rapidly than impaired patients. We compute the average time it takes for each patient to complete a cycle, as well as just the nose-to-finger and finger-to-nose portions of the cycles. We hypothesize that a difference in cycle length at low severities is more discriminative than the same difference at high severities, and so we use the logarithms of these values in our feature vector.

4.3.2 NUMBER OF DIRECTION CHANGES

We capture the amount of oscillation in the patient's movements, an important rating criteria in the BARS, by counting the number of times the wrist changes direction during the finger-to-nose action. We do this by counting the number of sign changes in the first derivative of the wrist's $x-$ and $y-$position signals. Our features include the raw counts for both signals as well as the counts normalized by the total number of cycles in the video. These features also describe the patient's degree of dysmetria, since patients who are more incoordinated have difficulty controlling the trajectory of their hand during the test.

4.3.3 VARIATION IN CYCLE DURATION

Patients with more severe ataxia are unable to perform the finger-to-nose action in a consistent manner. We capture this by computing the standard deviations of the full cycle times, and the nose-to-finger and finger-to-nose times.

4.3.4 APPROXIMATE ENTROPY (APEN)

ApEn features describe the regularity of a signal by segmenting signals into similar segments and computing the similarity of each segment in higher dimensions (Pincus, 1991). We followed

Venkataraman et al. (2015) and computed ApEn features with similarity thresholds of $r = 0.1, 0.12, 0.14,$ and 0.18. We found that an embedding dimension of $m = 3$ performed best.

4.4 Model

Using the features described in the previous section, we train a linear regression model to predict the BARS severity rating for each video. Though the BARS rating is not necessarily a linear function of our feature space, we use a simple model to avoid overfitting on our limited dataset. We use the LASSO technique, a method for linear regression that includes a regularization term to help with feature selection (Tibshirani, 1996). We use cross-validation to select the regularization parameter, and round the predicted rating to the nearest valid BARS severity (which ranges from 0 to 4 in half-point increments).

5. Evaluation

Because of the limited number of examples in our dataset, we use leave-one-patient-out cross validation to test our models. This approach allows us to train each model with the maximum amount of data, and evaluate the model's performance on a patient that it has not yet seen.
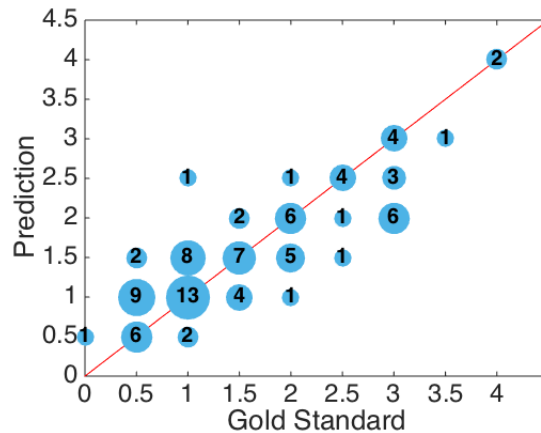

Figure 5: Our system's predictions versus the gold standard ratings on the full 90 video dataset. Dot area and the associated number represent the number of videos. Perfect prediction is represented by the red line. Our mean absolute error was 0.333 and our correlation with the gold standard labels was 0.856. This is comparable to the performance of skilled clinicians.

5.1 Prediction Error

Our model performs reasonably well in learning the gold standard ratings. Figure 5 shows our predictions versus the ground truth severity ratings. Most (87.8%) of our models' prediction errors are less than one level on the BARS. Only one video produced an absolute error greater than 1 point. In this case, tracking noise caused a cycle segmentation error that combined two cycles into one.

5.2 Comparison with Physicians Specializing in Ataxia

We have shown that our system can learn the gold standard labels with low error, but is it good enough for clinical use? We demonstrate that our system performs comparably with the current clinical state-of-the-art: human raters using a rating scale such as the BARS. We collected ratings for 32 test videos from 6 ataxia specialists, all of whom are practicing physicians as well as members of the Clinical Research Consortium for Spinocerebellar Ataxias. All of the specialists had used ataxia rating scales such as the SARA before, but none had prior experience with the half-point BARS. They were provided with textual descriptions of each BARS severity and were all shown the same $1-3$ example videos for each severity level. Each specialist rated each video independently and in randomized order. The test videos were hand-selected so that they had as uniform a severity distribution, and as few common patients with the example set as possible. No videos appeared in both the example set and the test set; however, some videos from the same patient taken during different clinical visits appear in both. While the ataxia specialist raters did not learn from as many videos as our system did, they have the advantage of learning from textual descriptions in the BARS, as well as their prior expertise in treating ataxia. This arrangement is representative of how human raters would be trained in practice.
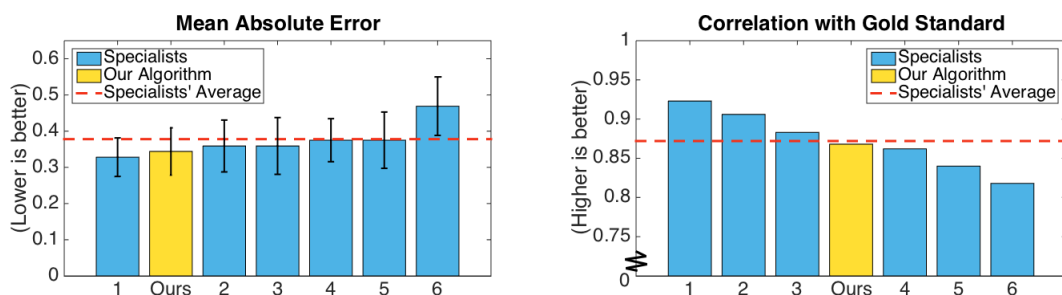

Figure 6: Comparison of our system's performance to each of the ataxia specialist raters with respect to the gold standard label. The error bars represent the standard error of the mean.

We first compare the performance of our system and the ataxia specialists on learning the gold standard labels. Figure 6 shows that our predictions fall within the range of the specialists in terms of mean absolute error and correlation with the gold standard. On average, our system achieves an absolute error of 0.344 and a correlation of 0.868 whereas the specialists achieve 0.378 mean absolute error, and 0.872 correlation. This indicates that our system can learn the BARS as well as expert physicians.

Additionally, we compare our predictions with inter-rater variability among the specialists. Figure 7 compares our prediction with the specialists' ratings for individual videos. In this result, we make no distinction between the gold standard rater and the other specialists, to more accurately capture inter-rater variability among ataxia specialists. Our predictions were consistently within the range of inter-rater variability. Our prediction falls within the range of specialist ratings 78% of the time, and within 0.5 points of the range for all but one video, where a low severity patient moved their finger slowly, causing us to overestimate their severity rating by 1.

Our results indicate that our system performs comparably to neurologists specializing in ataxia. In areas where ataxia specialists are not available, our system may provide a more accurate evalua-
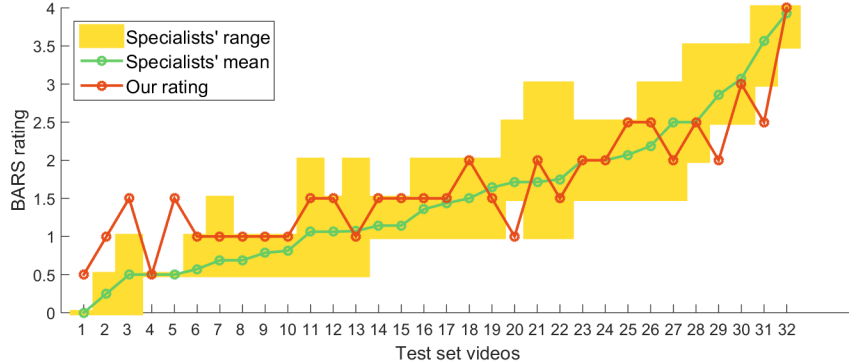
Figure 7: Range of ataxia specialists' ratings for each test videos. Our rating (red) was almost always within the range of the ataxia specialists' ratings (yellow).

tion of ataxia than human clinicians. Our system could also be useful for producing more consistent ratings of ataxia between different clinical visits, or between different clinicians.

6. Discussion and Conclusion

We described an automatic, video-based system for quantifying the severity of motion impairment of ataxic patients performing the finger-to-nose test described in the Brief Ataxia Rating Scale (BARS). The system is designed with consideration for the low video quality that one might expect from clinical settings; it is robust to variations in camera angle, harsh lighting conditions and motion blur. Our system uses convolutional neural network-based pose estimation and optical flow to track the location of a patient's wrist and head. From the wrist motion signals, we extract features such as average cycle time and amount of oscillation. We use these features to build a linear regression model to predict the severity rating of the patient's action. We show that our system can predict BARS ratings with comparable accuracy to experienced ataxia specialists. More importantly, most of our system's predictions fall within the range of ratings supplied by these specialists. This suggests that our system might be a suitable observer-independent alternative to traditional human-administered rating scales.

Our system is limited to using a simple linear model that may not capture all of the relevant characteristics of the motion. It tends to overestimate the rating for lower severity videos, and underestimate the rating on higher severity videos. A larger training set would permit the use of a more complex model than linear regression, which might help capture these more complicated cases.

We hope that this work will lead to a system that can be used in the management of movement disorders in clinical or home settings, and to reduce observer bias in clinical trials. The automatic and consistent ratings provided by such a system can improve the monitoring and treatment of these disorders. Such a system would be particularly beneficial in many parts of the country that do not have access to specialists.

10

Acknowledgments

References

Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3686–3693. IEEE, 2014.

Jan-Philipp Bach, Uta Ziegler, Günther Deuschl, Richard Dodel, and Gabriele Doblhammer-Reiter. Projected numbers of people with movement disorders in the years 2030 and 2050. *Movement Disorders*, 26(12):2286–2290, 2011.

Thomas Brox, Christoph Bregler, and Jagannath Malik. Large displacement optical flow. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 41–48. IEEE, 2009.

James Charles, Tomas Pfister, Derek Magee, David Hogg, and Andrew Zisserman. Upper body pose estimation with temporal sequential forests. In *Proceedings of the British Machine Vision Conference 2014*, pages 1–12. BMVA Press, 2014.

Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Real-time tracking of non-rigid objects using mean shift. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 142–149. IEEE, 2000.

Patrik Fazio, Gino Granieri, Ilaria Casetta, Edward Cesnik, Sante Mazzacane, Pietro Caliandro, Francesco Pedrielli, and Enrico Granieri. Gait measures with a triaxial accelerometer among patients with neurological impairment. *Neurological Sciences*, 34(4):435–440, 2013.

Brook Galna, Gillian Barry, Dan Jackson, Dadirayi Mhiripiri, Patrick Olivier, and Lynn Rochester. Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease. *Gait & posture*, 39(4):1062–1068, 2014.

Joseph P Giuffrida, David E Riley, Brian N Maddux, and Dustin A Heldman. Clinically deployable kinesia technology for automated tremor assessment. *Movement Disorders*, 24(5):723–730, 2009.

Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results. *Movement disorders*, 23(15): 2129–2170, 2008.

Thomas Klockgether. Sporadic ataxia with adult onset: classification and diagnostic criteria. *The Lancet Neurology*, 9(1):94–104, 2010.

Patrick Kugler, Christian Jaremenko, Johannes Schlachetzki, Juergen Winkler, Jochen Klucken, and Bjoern Eskofier. Automatic recognition of parkinson's disease using surface electromyography during standardized gait tests. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5781–5784. IEEE, 2013.

L Eduardo Cofré Lizama, Fary Khan, Peter VS Lee, and Mary P Galea. The use of laboratory gait analysis for understanding gait deterioration in people with multiple sclerosis. *Multiple Sclerosis Journal*, page 1352458516658137, 2016.

Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.

Alvaro Muro-de-la Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors*, 14(2):3362–3394, 2014.

Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

Hamed Pirsiavash, Carl Vondrick, and Antonio Torralba. Assessing the quality of actions. In *European Conference on Computer Vision*, pages 556–571. Springer, 2014.

Jeremy D Schmahmann. Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *The Journal of neuropsychiatry and clinical neurosciences*, 2004.

Jeremy D Schmahmann, Raquel Gardner, Jason MacMore, and Mark G Vangel. Development of a brief ataxia rating scale (bars) based on a modified form of the icars. *Movement Disorders*, 24 (12):1820–1828, 2009.

T Schmitz-Hübsch, S Tezenas Du Montcel, L Baliko, J Berciano, S Boesch, Chantal Depondt, P Giunti, C Globas, J Infante, J-S Kang, et al. Scale for the assessment and rating of ataxia development of a new clinical scale. *Neurology*, 66(11):1717–1720, 2006.

Jianbo Shi and Carlo Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994.

Narayanan Sundaram, Thomas Brox, and Kurt Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. In *European conference on computer vision*, pages 438–451. Springer, 2010.

David H Sutherland. The evolution of clinical gait analysis: Part ii kinematics. *Gait & posture*, 16 (2):159–179, 2002.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

P Trouillas, T Takayanagi, M Hallett, RD Currier, SH Subramony, K Wessel, A Bryer, HC Diener, S Massaquoi, CM Gomez, et al. International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. *Journal of the neurological sciences*, 145(2):205–211, 1997.

Vinay Venkataraman, Pavan Turaga, Nicole Lehrer, Michael Baran, Thanassis Rikakis, and Steven Wolf. Attractor-shape for dynamical analysis of human movement: Applications in stroke rehabilitation and action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 514–520, 2013.

Vinay Venkataraman, Ioannis Vlachos, and Pavan K Turaga. Dynamical regularity for action analysis. In *BMVC*, pages 67–1, 2015.

Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. *arXiv preprint arXiv:1602.00134*, 2016.

Aner Weiss, Sarvi Sharifi, Meir Plotnik, Jeroen PP van Vugt, Nir Giladi, and Jeffrey M Hausdorff. Toward automated, at-home assessment of mobility among patients with parkinson disease, using a body-worn accelerometer. *Neurorehabilitation and neural repair*, 25(9):810–818, 2011.

Anja Weyer, Michael Abele, Tanja Schmitz-Hübsch, Beate Schoch, Markus Frings, Dagmar Timmann, and Thomas Klockgether. Reliability and validity of the scale for the assessment and rating of ataxia: a study in 64 ataxia patients. *Movement disorders*, 22(11):1633–1637, 2007.