Methodology

How did you extract features?
Trajectories of body landmarks come from running OpenPose. We use them as features together with a few simple derived signals, such as projected knee angle. We post-processed the OpenPose outputs as described in the associated tutorial.

Do you filter the time-series to compensate for the frame-by-frame prediction of the CNN?
We've found that direct measurements from OpenPose are too noisy for classical signal processing. For that reason, we decided to use a convolutional neural network (CNN). Before we feed signals to the neural network, we denoise them, interpolate missing values, center the pose to the pelvis and normalize observations. Please see the associated tutorial for some details.

Did you normalize the key point coordinates for training? Can you give some details?
Yes, we center them and scale uniformly for the whole video, and then we normalize each frame by dividing it by the visible length of the femur. Please see the paper for details. Also, visit the associated tutorial for some details.

There are several published approaches to derive the 3D pose estimations from the 2D pose (such as this https://arxiv.org/abs/1705.03098). Did you try any of these as well?
Yes, we tried those, and they didn't add enough value to our workflow. The 3D reconstruction wasn't very close to ground truth. We are hopeful that these methods will get better over time.

What is currently the processing time? Is it possible to have any real-time feedback?
We were not focusing on real-time since that wasn't that important to us for our clinical application, but there aren't any big obstacles to getting a real-time feedback. With good hardware, OpenPose runs at 30 frames per second - that's real-time. Our network is even lighter than OpenPose, so if that's the objective for a certain application, that's completely doable.

Did you compare different classification methods? What was your experience?
We also tested random forest and ridge regression models. The CNN performed as well or better than these other models, so we focused on the CNN results. More details are in the paper.

Spatial-temporal parameters are important when evaluating human gait. Did you examine any of those?
We developed models to predict walking speed, cadence, peak knee flexion, and GDI. Please check out the paper for details.

How does your approach deal with the ground reaction force (GRF) problem? How would GRF be added to the analysis?
We conjecture that some of this information can be implicitly derived through analyzing kinematics. The focus of this research is kinematics but estimating kinetics and GRFs is an interesting future area of research. It will likely be more challenging since the video is not directly measuring forces (vs. movement). If you measured GRFs separately or had pressure sensors, this could be more straightforward using current methods. We haven't tried this yet, but from the pipeline perspective there should be no problem in integrating these additional data.

Does an increase in the number of cameras improve the results?
Yes, as we found in our study, the model based on only one camera didn't manage to extract information to approximate certain joint angles, such as foot progression angle. We found that using the frontal plane, in addition to the sagittal plane, would improve our models. We expect that this is also true for other applications.

OpenPose

Can OpenPose be used for the motion of objects that are not human (animals, insects, vehicles, objects)?
OpenPose works only for human poses; however, DeepLabCut allows you to define key points on any object and requires a relatively small training dataset.

Does this model extract features only from a single object or multiple objects?
OpenPose can track multiple people in the view of the camera, but we were focusing on recordings where a single person was in view.

Utilizing Approach for Different Applications

What are some challenges with translating this approach to amputee patients?
Depending on the level of the amputation, some investigation would be needed to see whether key points (e.g., knee center, ankle center) are predicted accurately with OpenPose for amputees. Collecting images with annotations on prosthetics could allow you to extend OpenPose for this application. DeepLabCut might be a better solution for such a scenario.

Does this system work with movements other than walking, e.g., jumps, change of direction, etc.? What about non-gait-related movement conditions?
The first part of the algorithm, the key point detection, would work with any movement. We would get a time series of joints or key points, which we could process with some direct computational signal processing or deep learning. The second part – the network that we trained – was explicitly for cerebral palsy, for very controlled walking a given distance from the camera. So, this second part would need to be retrained for your application. Given a sufficiently large training dataset of consistent video recordings, though, one should be able to port our approach to other conditions, even non-gait-related movements.

Is it possible to use your pre-trained models to train the model for the disabilities?
Yes, you could start training from our pre-trained model. This could allow you to use a smaller dataset for training. If you are designing a study from scratch and are planning to use our network, then it might be worth having one camera set up in a similar way as in our study. Please see the paper for details.

Do you need to use some kind of affine transformation if, for example, the camera is filming from a different angle?
Yes – there are ways to make our models more robust, but we didn't do any of that. In order to use our models, you would need to bring your data to a format as similar as possible to what we trained our models on. Therefore, if the camera angle is different than ours, you would need some transformations and/or some retraining of algorithms for your data.

In our preprocessing step, we normalize each frame so that the length of one femur is constant – this makes our approach slightly more robust to the camera angle. Please see the associated tutorial for some details.

What do you predict as the biggest challenge in moving such laboratory-explored technology to a deployable, smart-phone application?
Getting consistent recordings from different subjects might be the most challenging. We are testing now to what extent we can rely on subjects to collect their own data.

Will the system work if the child, for example, walks along a hallway in their house rather than a standard path in the clinic?
We believe that this could work, as long as a similar view could be obtained (e.g., whole body in view), but further testing is needed. We are starting a new study to test the robustness of the approach outside of a standardized lab setting.

Any ideas on extracting additional biomechanics metrics and/or algorithm evaluation by crossing the following metrics: maxima toe-to-toe distance, minima toe/heel heights, extreme knee flexion?
We chose metrics relevant to current clinical practice in treating cerebral palsy, but definitely, other metrics are useful in other scenarios.

Resources

Is the pediatric dataset with ~1800 images available for research?
Since this is patient data that spans more than 30 years, we couldn't track down all the patients to get consent for releasing the actual videos. However, we can, and we did release the key points derived from OpenPose, which are sufficient for running our downstream analysis and potentially could be useful for your research. You can access that data here.