Can machine learning assess trunk alignment directly from raw video?

Abstract

Background: Current physical therapy assessments of head and trunk control status in children with cerebral palsy are subjective. Previous work has established that objective replication of an existing clinical test (Segmental Assessment of Trunk Control (SATCo)) can be made using a 2D video-based method with semi-automated tracking of the video sequences. A markerless full automation of the analysis of live camera data would provide an objective and clinically-friendly tool for both assessor and patient. The use of high-definition depth (HD+D) cameras would also address the limitations of 2D video, such as body movement out of camera plane.

Research question: This study was to examine whether HD+D analysis is suitable for the classification of the alignment of given head and trunk segments in sitting by comparing expert opinion (labelling) to machine learning classification.

Methods: Sixteen healthy male adults were recruited and a SATCo was conducted for each participant and recorded using a Kinect V2. Two different trials were collected (Control and No-Control) to simulate the physical therapy test with children. Three of the seven SATCo segmental levels were selected to perform this feasibility analysis. Classification of alignment obtained with the machine learning classification (convolutional neural networks) of all frames was compared to an expert clinician's labelling, and to a randomly selected reference aligned frame.

Results: At the optimal operating point of Receiver Operating Characteristics the neural network analysis correctly classified alignment and misalignment with an accuracy of 79.15%; with 64.66% precision and 76.79% recall.

Significance: This communication demonstrates, for the first time, an automated classification of trunk alignment directly from raw images (HD+D) and which requires minimal operator interaction. This demonstrates the potential of machine learning to provide a fully automated objective tool for the classification of the alignment component of head/trunk control in sitting that is suitable for clinical use.

Keywords

Alignment; Kinect; Clinical assessment; Objective measure; Machine learning

1. Introduction

Poor or absent head and trunk control is a frequent consequence of neurodevelopmental conditions such as cerebral palsy; it can compromise a child's ability to sit independently and lead to functional limitations [1]. The therapeutic objective is usually towards independent, unsupported sitting with the child maintaining a vertically aligned (neutral) head and trunk posture [2]. Current physical therapy assessments are generally based on tests that evaluate control status from the observation of functional abilities [3-6]. These assessments, although reliable, are subjective and most consider the head/trunk as a single unit, ignoring its multi-segmental composition [3-6].

Previous work has established that objective replication could be made of a clinician's subjective ability to identify the separate trunk segments and make a judgement of their position in space relative to a defined aligned posture. This work was based on the Segmental Assessment of Trunk Control (SATCo) [2] and used a 2D video-based method [7]. This previous study used semi-automated tracking of the video sequences: a markerless full automation of the analysis of live camera data would provide an objective and clinically-friendly tool for both the assessor and the patient [7-9]. The use of high-definition depth (HD+D) cameras has potential to address these issues while bringing the analysis closer to the clinical judgement [7-9].

The first stage towards automated analysis is for an experienced SATCo clinician to identify whether given trunk segments are aligned during a SATCo (i.e. labelling). This pilot study used a Kinect V2 (Windows®, Microsoft®) camera to record grayscale HD+D images during SATCo testing of healthy adult males that simulated the clinical test. State of the art machine learning methods were then applied to reproduce the clinical labels (classification) [10, 11]. This enabled examination of whether automated analysis of raw HD+D images is suitable for the classification of the alignment component of head and trunk control in sitting by comparing the expert's labelling to the machine learning classification.

2. Methods

Ethical approval was obtained from the Manchester Metropolitan University Ethics Committee. The participants were 16 healthy adult males (mean age 31.39 ±5.21 years, mean height 1.78m ±0.07, and weight 77.7kg ±11.1). Participants sat on a bench free of back or arm support; the height of the bench ensured that participants' feet were flat on the floor and the knees and hips were flexed at 90^o. A Kinect camera was set at a distance of 1.60m and height of 0.90m on the left side of the participants and both grayscale HD and depth images were recorded at 15Hz (asynchronous recording software was written by the authors in c++). The SATCo was conducted; a tester provided manual support to the participant's trunk to test six discrete trunk segmental levels following the published guidelines [2]. Two different trials were collected, Control and No-Control, to simulate the physical therapy test with children. For the Control trials, participants were asked to remain still for 5s in upright sitting with the arms and hands free in the air; for the No-Control trials a verbal cue was given for participants to simulate lack of trunk control; this was done by making large movements of the unsupported segments of the trunk (segments above the tester's hand support) away from the aligned position (e.g. falling forwards).

This feasibility analysis was conducted using three segmental levels (Head, Lower-Thoracic and Free-sitting). Figure 1 shows the process followed: i) training labels (classification of each image) for the neural network were provided by an experienced clinician who identified the frames when all unsupported segments in the trunk were aligned, and the frames when one or more unsupported segments were misaligned; ii) using raw depth information from the Kinect, the background was automatically subtracted from the image to reveal the participant and tester; iii) four identical multilayer convolutional neural networks were trained to provide held-out test results for all 16 participants, each network using approximately 8,000 images from 12 participants and each tested on approximately 3,000 images from four participants not used for training; iv) the neural network was trained to predict the clinical labels, from an individual image at time *t*, and a randomly selected (at neural network training time) reference image for all images in the data set. The reference image is any image within the same trial where the person was labelled as 'aligned' by the clinician; v) after training, the neural network output for all 16 participants was compared, by receiver operating characteristics (ROC), to the clinical labels. At the optimal operating point of the ROC a neural network threshold was selected maximizing the ratio of true positive to false positive classifications.

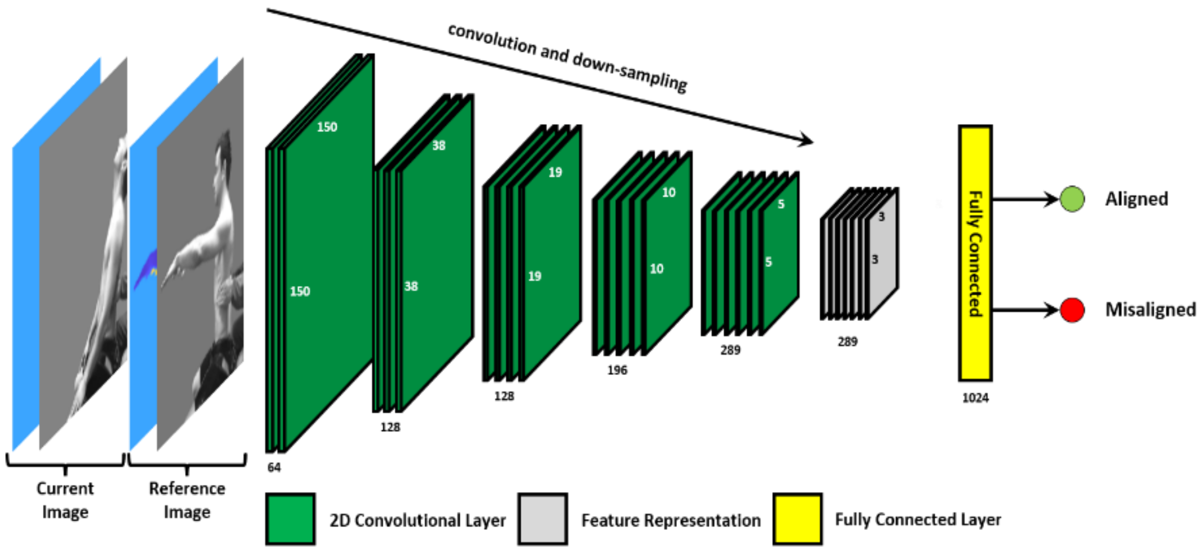

Figure 1 Network architecture.
Showing the deep neural network architecture consisting of 5 convolutional and down-sampling (max-pooling) layers, followed by a fully connected layer, and a logistic regression output layer. The current image, and reference image (far left) each represent 2 data streams, a grayscale image and a depth map.

3. Results

At the optimal operating point of the ROC, the neural network analysis correctly classified alignment and misalignment with an accuracy of 79.15% (Figure 2). The data were biased towards alignment, therefore, more representative results in the form of precision is given, 64.66%, and recall 76.79%. Using precision and recall the F_1 score was calculated, $F_1 = 2\left((precision \times recall) \times (precision + recall)^{-1}\right) = 70.21\%$, which gives balanced measure of performance, accounting for the bias in the data/labels. The F_1 score is best at 100% and worst at 0%.

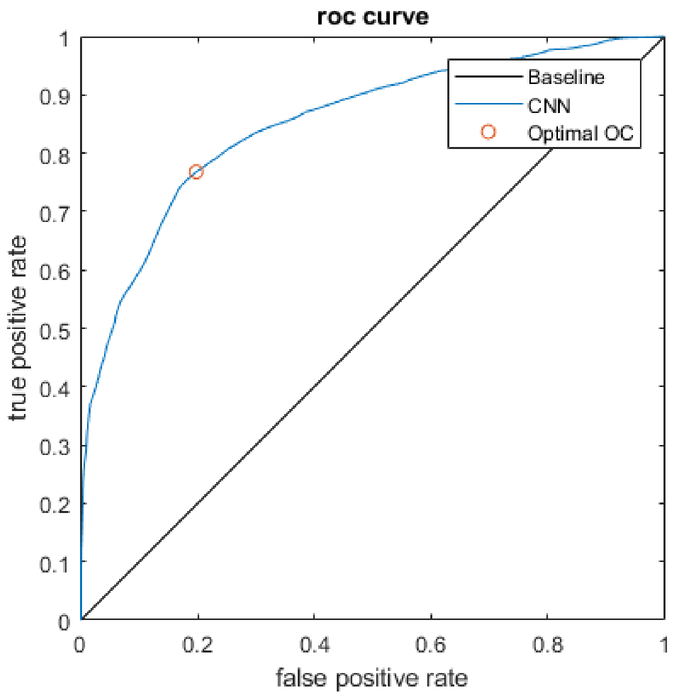

Figure 2 Receiver operating characteristics (ROC).
This figure shows the true positive classification rate versus false positive classification rate for neural network (CNN) output thresholds between 0 and 1, compared with the clinical labels of alignment. At the optimal operating point (Optimal OC) this network correctly classifies 79.15% of all frames (precision: 64.66%, recall: 76.79%, F_1: 70.21%).

4. Discussion

This method used a single Kinect camera and an analysis that is markerless. The results (Figure 3) depict the quality of data that can be collected with this method, the ability to automatically remove background to reveal the participant showing that there is very little noise in both images, and to differentiate the left and right limbs. These features are essential for correct classification and contributed to the strong evidence, shown by the F_1 score, for the feasibility of the method presented in this study. The proposed method only requires an operator to identify one aligned frame, and thereafter the neural network can provide fully automatic analysis of alignment. Participants simulated lack of trunk control by moving away from the aligned posture; movement displacement occurred in all three planes of motion. Previous studies used a 2D video-based method to track markers in the sagittal plane [7, 9]; participant's movement in planes other than the sagittal would generate movement artefacts reducing the accuracy of the method. The introduction of depth images to this analysis, unlike 2D feature tracking, enables the accurate classification of postural misalignment, even in situations where motion occurs orthogonal to the sagittal plane.

This preliminary work affirms that HD+D analysis is suitable for the classification of the alignment component of head and trunk control in sitting. Beyond this study, accuracy and robustness will almost certainly improve with additional data, additional cameras, more detailed labelling, and as the neural network architecture is tuned and refined. This work demonstrates the potential of machine learning to provide an objective and clinically-friendly assessment of seated head and trunk control.

Conflict of interest

None

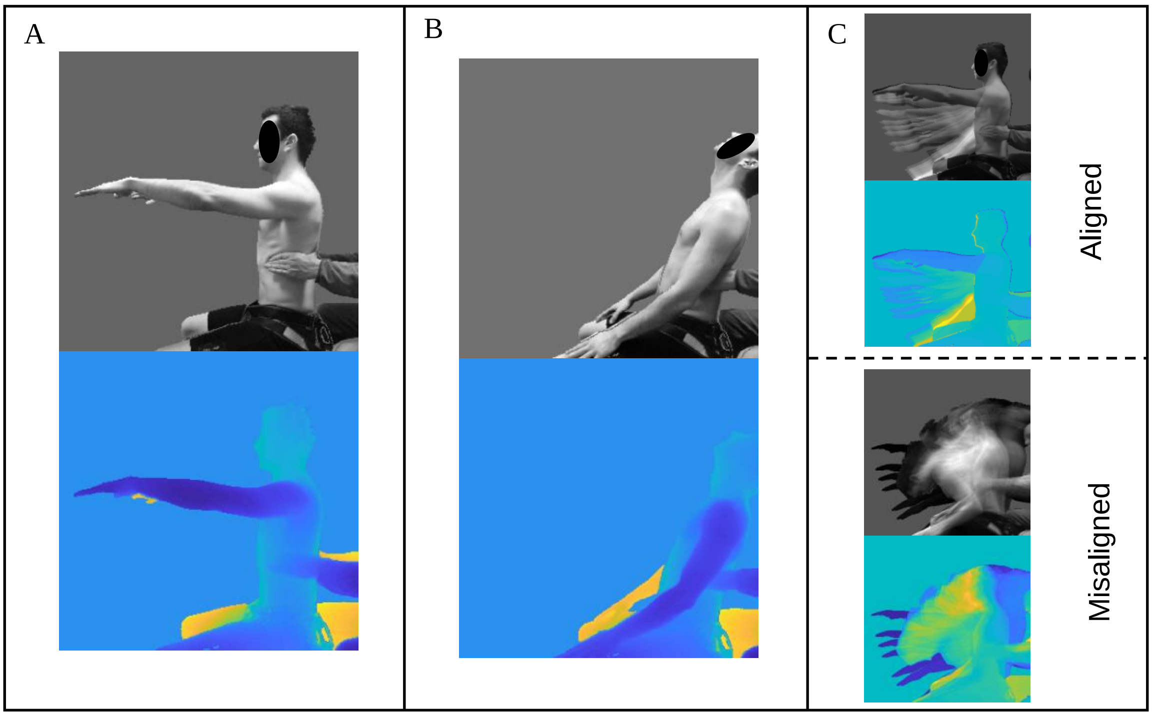

Figure 3 Representative trial example of Static control when testing at the Lower-Thoracic segmental level.

In all grayscale and depth images, the background was subtracted using the information in the depth image. Showing **A)** Reference grayscale (top) + depth (bottom) image classified by the expert as 'aligned'; **B)** a grayscale (top) + depth (bottom) image to be classified by the neural network; and **C)** distribution of grayscale + depth images, correctly classified by the neural network as aligned (top) or misaligned (bottom).

References

[1] Jensen J, van Zandwijk R. Biomechanical aspects of the development of postural control. In: Korff T, De Ste Croix M, editors. Paediatric biomechanics and motor control: theory and application. Abingdon, Oxon: Routledge research in sport and exercise science; 2012. p. 139-59.

[2] Butler P, Saavedra SL, Sofranc M, Jarvis S, Woollacott MH. Refinement, reliability, and validity of the Segmental Assessment of Trunk Control. Pediatr Phys Ther. 2010;22:246-57.

[3] Heyrman L, Molenaers G, Desloovere K, Verheyden G, De Cat J, Monbaliu E, et al. A clinical tool to measure trunk control in children with cerebral palsy: the Trunk Control Measurement Scale. Res Dev Disabil. 2011;32:2624-35.

[4] Pountney TE, Cheek L, Green E, Mulcahy C, Nelham R. Content and Criterion Validation of the Chailey Levels of Ability. Physiotherapy. 1999;85:410-6.

[5] Reid DT. Sitting Assessment for Children with Neuromotor Dysfunction SACND. San Antonio, Texas: Therapy Skill Builders; 1997.

[6] Russell DJ, Rosenbaum PL, Avery LM, Lane M. Gross motor function measure (GMFM-66 and GMFM-88) user's manual. First ed. London: Mac Keith Press; 2002.

[7] Sánchez MB, Loram I, Darby J, Holmes P, Butler PB. A video based method to quantify posture of the head and trunk in sitting. Gait Posture. 2017;51:181-7.

[8] Sánchez MB, Loram I, Darby J, Holmes P, Butler PB. The potential of an automated system to identify the upper limb component of a controlled sitting posture. Gait Posture. 2017;58:223-8.

[9] Sánchez MB, Loram I, Holmes P, Darby J, Butler PB. Working towards an objective segmental assessment of trunk control in children with cerebral palsy. 2018;65:45-50.

[10] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 2015;abs/1511.06434.

[11] Zeiler MD, Krishnan D, Taylor GW, Fergus R. Deconvolutional networks. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition2010. p.