Can machine learning assess trunk alignment directly from raw video using a depth camera?

María B. Sánchez[a*], Ryan Cunningham[a], Penelope Butler[a] and Ian Loram[a]

[a] Research Centre for Musculoskeletal Science & Sports Medicine, Manchester Metropolitan University, Manchester, UK

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 provide more information (depth of position) than 2D video.

Research question: This study examined whether full automation of the classification of trunk alignment will be possible using deep neural network analysis of raw HD+D images.

Methods: For this test of principle, 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 a range of alignment configurations. This feasibility analysis was performed at all SATCo segmental levels. For all images, classification of alignment obtained from a trained and validated deep neural network, was compared to an expert clinician's labelling.

Results: At the optimal operating point of Receiver Operating Characteristics the neural network analysis correctly classified individual images (alignment v misaligned) with accuracy 78.6%; precision 64.7% and recall 75.6%.

Significance: This communication demonstrates, for the first time, an automated classification of trunk alignment directly from raw images (HD+D). 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. This proof of concept justifies investment of resources for clinical data collection and development of the machine learning to enhance understanding of trunk control issues in children with neurodisability such as cerebral palsy.

Keywords

Alignment; Objective measure; Machine learning; Deep neural network; Depth camera;

1. Introduction

Poor head and trunk control is a frequent consequence of neurodevelopmental conditions such as cerebral palsy, compromising a child's ability to sit independently and leading 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 with semi-automated tracking of the video sequences [7]: an automated analysis of live camera data would provide an objective and clinically-friendly tool for both assessor and patient [7-9]. The use of high-definition depth (HD+D) cameras has greater potential to eliminate use of markers and detect out of plane movement.

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. 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 90º. 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 partially simulate the test with children and to give greater range to the analysis. 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 by making 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 all the segmental levels. Figure 1 shows the process followed: i) training labels (classification of each image) for the neural network were provided by a clinician (MBS) with four years of experience doing this analysis, 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) eight identical multilayer deep neural networks were trained to provide held-out test results for all 16 participants, each network using approximately 22,000 images from 14 participants and each tested on approximately 3,000 images from two participants not used for training; iv) the neural network was trained to predict the clinical labels, from an individual image at time *t*; 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.

3. Results

At the optimal operating point of the ROC, the neural network analysis correctly classified frames (alignment v. misalignment) with an accuracy of 78.6% (Table 1). The data were biased towards alignment, therefore the precision 64.7%, and recall 75.6% provide more representative results. The F_1 score (harmonic mean of precision and recall, $F_1 = 2\left((precision \times recall) \times (precision + recall)^{-1}\right)$, 69.7% gives a balanced measure of performance, accounting for the bias in the data/labels. The F_1 score is best at 100% and worst at 0%

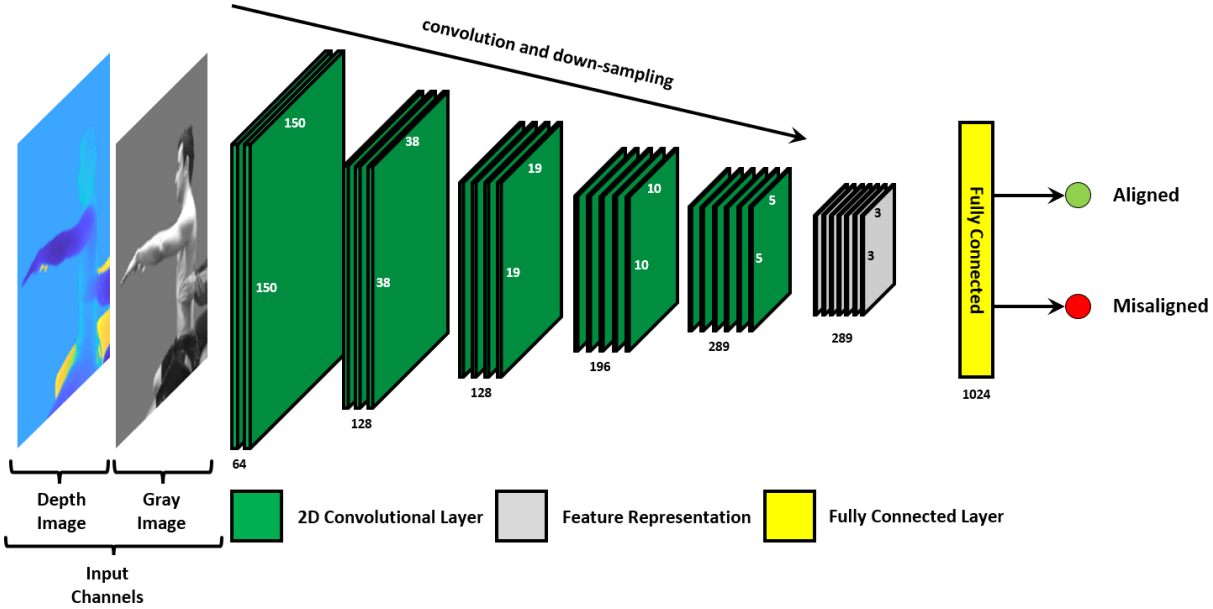

Figure 1 Network architecture.
Showing the deep neural network architecture consisting of 2 input channels (depth and grey-converted RGB image), 5 convolutional and down-sampling (max-pooling) layers, followed by a fully connected layer, and a logistic regression output layer.

4. Discussion

This method used a single Kinect camera and an analysis that is markerless. The results (Figure 2) 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 aid correct classification. At 69.7%, the F_1 score provides strong evidence for the feasibility of this method. Clinical labels were required only for training; thereafter the neural network provided fully automatic analysis of alignment without any further user interaction.

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 semi-automated 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 method reported here included depth images enabling the accurate classification of postural misalignment from raw images even in situations where motion occurred orthogonal to the sagittal plane. This method also demonstrates potential to identify a reference posture to use in conjunction with our automated tracking of specific landmarks of the head, trunk and

arms in children with cerebral palsy to identify which segments are misaligned [12].

Classification of trunk alignment in sitting is one of the components required to assess control status in the Segmental Assessment of Trunk Control (SATCo). This was a small study of a non-clinical population mimicking control/no control. Nevertheless, the results demonstrate proof of principle that deep neural networks can automate the objective classification of trunk alignment in sitting from raw HD+D images. This proof of concept justifies investment of substantial resources to collect the amount of data required from children with CP, and to develop the neural network analysis to the accuracy required for clinical use. Such a tool would enhance the understanding of the development of head and trunk control and its alterations in the presence of neuromotor disabilities.

Conflict of interest

None

.

Table 1 Summary of results.

This table presents comprehensive results for each participant, calculated over all trials, and a summary of results, calculated over all participants and trials. For each participant, the number of positively classified (*misaligned*) and negatively classified (*aligned*) images are given for the clinical expert and for the automated system, followed by measures of agreement between the clinical expert and the neural network (ANN), in the form of accuracy, precision, recall, and the harmonic mean of precision and recall, the F1 measure.

Participant #	# Positive classes (labels)	# Negative classes (labels)	# Positive classes (ANN)	# Negative classes (ANN)	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
1	688	1306	524	1470	86.76	90.46	68.90	78.22
2	418	1049	513	954	78.25	59.65	73.21	65.74
3	608	886	532	962	86.75	88.53	77.47	82.63
4	251	1237	304	1184	93.21	74.67	90.44	81.80
5	252	1160	223	1189	91.15	78.48	69.44	73.68
6	497	1035	358	1174	82.96	82.96	59.76	69.47
7	525	1044	664	905	76.10	61.30	77.52	68.46
8	546	1079	1485	140	38.52	34.75	94.51	50.81
9	1150	521	1478	193	66.73	70.09	90.09	78.84
10	404	1152	314	1242	90.36	90.45	70.30	79.11
11	401	1075	379	1097	88.48	80.47	76.06	78.21
12	439	958	381	1016	76.38	64.30	55.81	59.76
13	266	1204	745	725	58.03	26.44	74.06	38.97
14	604	1168	573	1199	85.27	79.93	75.83	77.82
15	427	1039	443	1023	69.17	47.18	48.95	48.05
16	599	843	520	922	91.05	95.19	82.64	88.47
mean	504.69	1,047.25	589.75	962.19	78.70	70.30	74.06	70.00
std	210.00	180.35	361.78	343.82	14.21	19.62	11.93	13.58
summary	8,075*	16,756*	9,436*	15,395*	78.63**	64.67**	75.57**	69.69**

**summation over rows*
***scores calculated over all data combined*

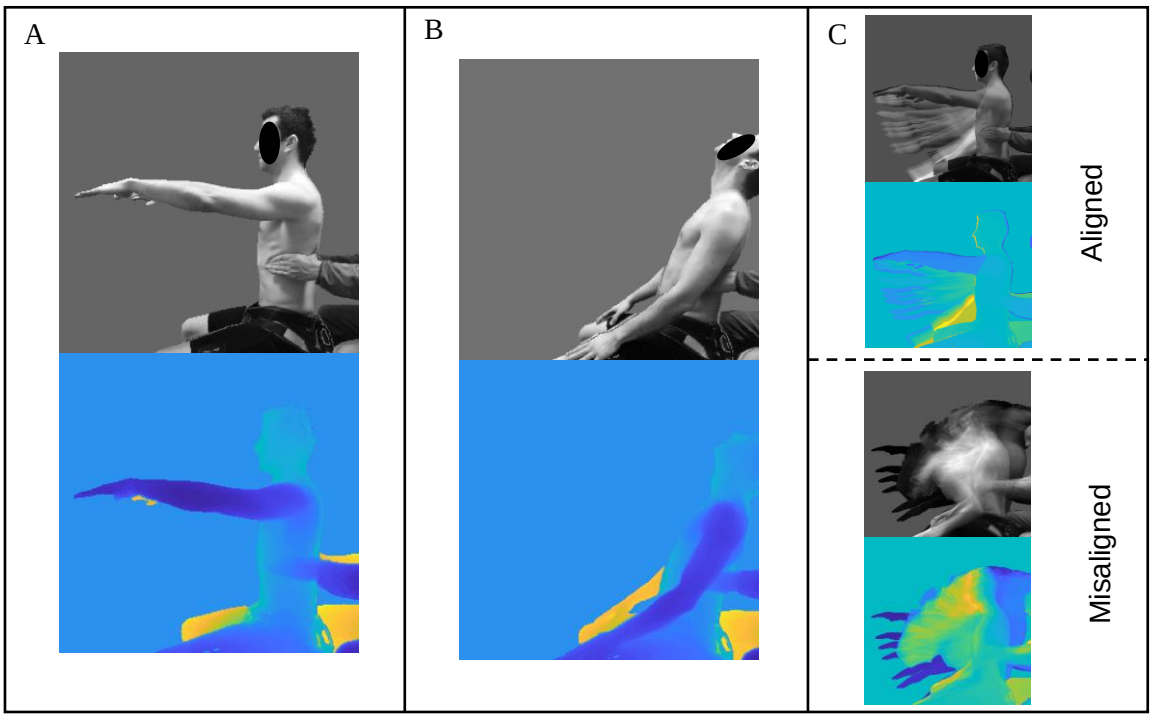

Figure 2 Representative trial example of Static control when testing at the Lower-Thoracic segmental level.

In all grayscale and depth images, the background was subtracted using the information in the depth image. Showing **A)** an example of 'aligned' posture (classified by the expert) presented in the form of grayscale (top) and depth image (bottom); **B)** an example of 'misaligned' posture (classified by the expert) presented in the form of grayscale and depth image; and **C)** distribution of grayscale and 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. Gait Posture. 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. 2528-35.

[12] Cunningham R, Sánchez MB, Butler PB, Southgate MJ, Loram ID. Fully Automated Video-based Estimation of Postural Point-features in Children with Cerebral Palsy Using Deep Learning. 2018.