Deep Convolutional Neural Networks for Motion Instability Identification using Kinect

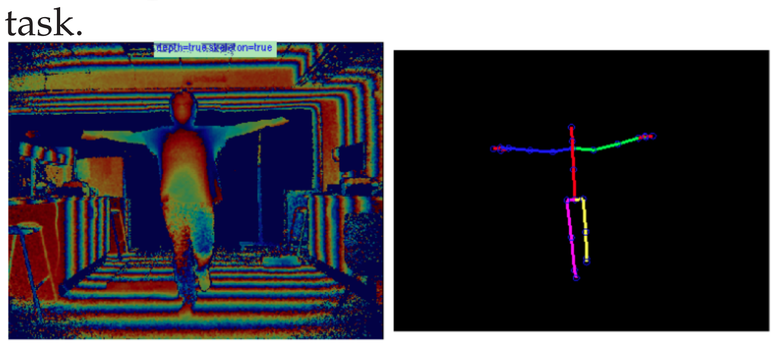
Daniel Leightley, Subhas C. Mukhopadhyay, Hemant Ghayvat, Moi Hoon Yap
dleightley@ieee.org

Manchester Metropolitan University

King's College London

Problem Statement

Evaluating the execution style of human motion can give insight into the performance and behaviour exhibited by the participant. This could enable support in developing personalised rehabilitation programmes by providing better understanding of motion mechanics and contextual behaviour [Ye et al., 2013]. However, performing analyses, generating statistical representations and models which are **free from external bias, repeatable and robust** is a difficult task.



The figure above shows an example of the data collected during a clinical trial assessing human balance. The output is obtained via the Microsoft Kinect.

In this work, we propose a framework which evaluates clinically valid motions to identify unstable behaviour during performance using Deep Convolutional Neural Networks.

Framework Proposal

The framework is composed of two parts;

1) Instead of using the whole skeleton as input, we divide the human skeleton into five joint groups. For each group, feature encoding is used to represent spatial and temporal domains to permit high-level abstraction and to remove noise these are then represented using distance matrices.

2) The encoded representations are labelled using an automatic labelling method and evaluated using deep learning.

Normalisation

Skeletal coordinate system: Data obtained via a MoCap system is captured within a predefined action space. To undertaken action analysis and classification it is important to place the participant skeletal structure at the centre the coordinate system to become view-invariant. To achieve this, we perform the following: where the root joint (*Hip Centre*) for each frame is subtracted from all other joints of the frame.

Acknowledgement

This work was supported by Royal Society International Exchanges Scheme (grant number: IE150436)

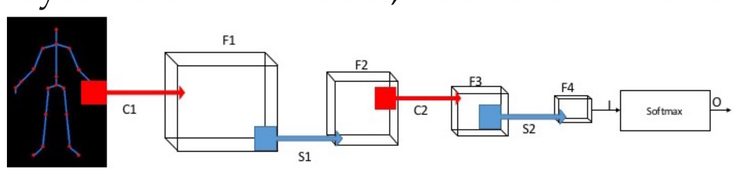
Deep Learning

To enable effective, efficient and representative motion classification, a computational model must be detailed and complex to provide true representation. Researchers have started to adopt deep, complex and highly representative models e.g. Deep Convolutional Neural Networks (DCNNs) [Li et al., 2015].

There is clear interest in utilising DCNNs for recognition and classification. A DCNN is capable of recognising patterns which contain varying degrees of shift, distortion and noise. We utilise this unique characteristic of DCNN to classify unstable motions from the patterns; the structure of the DCNN is as follows:

Layer: Configuration		Feature map dimensions	
C1:	4x4 templates	F1:	24x24
S1:	3x3 templates	F2:	12x12
C2:	3x3 templates	F3:	10x10
S2:	3x4 templates	F4:	5x5
I:	300 vector	O:	2 outputs

To generate the model, we follow the proposal in Ijjina *et al.* [Ijjina and Mohan, 2014] and create a 4-layer DCNN model, as shown below:



Features

Recognising the context and behaviour of human motion is not a straightforward task, more so when identifying instability from a diverse range of participants, environments and modalities. The feature used in this work as listed below:

Joint Group	Features	Kinect Joints
$F_{LeftArm}$	Left arm Euler Angle, Euclidean distance between the left shoulder and left hand, x and y axis vectors. $Length = \{1\ldots,12\}$	*LeftShoulder*, *LeftElbow*, *LeftWrist*, *LeftHand*
$F_{LeftLeg}$	Left leg Euler Angle, Euclidean distance between the left hip and left foot, x and y axis vectors. $Length = \{1\ldots,12\}$	*LeftHip*, *LeftKnee*, *LeftAnkle*, *LeftFoot*
$F_{RightArm}$	Right arm Euler Angle, Euclidean distance between the right shoulder and right hand, x and y axis vectors. $Length = \{1\ldots,12\}$	*RightShoulder*, *RightElbow*, *RightWrist*, *RightHand*
$F_{RightLeg}$	Right leg Euler Angle, Euclidean distance between the right hip and right foot, x and y axis vectors. $Length = \{1\ldots,12\}$	*RightHip*, *RightKnee*, *RightAnkle*, *RightFoot*
F_{Torso}	Torso Euler Angle relative to the body, Euclidean distance between the spine base and head, Body Movement Zone, Body lean angle (relative to the floor with torso as a reference), Centre-of-Mass (between left shoulder, right shoulder, spine mid), x and y axis vectors. $Length = \{1\ldots,16\}$	*SpineBase*, *SpineMid*, *Neck*, *Head*, *SpineShoulder*

There are several pose-based features and measurements which can be extracted from the skeletal stream, as shown above. However, there are difficulties in identifying the variables which are capable of describing the motion efficiently. To contribute these, we include the Body Movement Zone (BMZ).

Body movement zone: We encode the normalised total space volume occupied by the participant. This is computed by identifying the total space covered by the skeleton per frame using standard volume meter-squared calculations.

Distance matrices representation: We represent each encoded frame as a Euclidean distance matrices. This results in a set of distance matrices representing each encoded frame.

Results & Discussion

The K3Da Dataset [Leightley et al., 2015] was used in this study, which is a clinically validated dataset. The following motions were extracted from the K3Da Dataset: *Chair Rise*, *One-leg Balance (Eyes Open)*, *One-leg Balance (Eyes Closed)* and *Tandem Balance*.

The performance of DCNNs in identifying unstable motions compared to a range of classical machine learning techniques that are popular amongst researchers.

For each, a 10-fold cross-validation using the random 'leave-one-out' technique was implemented. The results obtained overall are presented below:

Iteration:	Median	Mean
SVM	90.95	91.10
RF	90.09	89.66
AdaBoost	82.87	84.26
LPBoost	71.03	72.22
RUSBoost	62.35	63.39
Total Boost	78.92	80.25
Bagging	72.61	73.83
SubSpace KNN	81.29	82.65
GRBM	82.50	81.27
DCNNs	**96.32**	**96.20**

We found that DCNNs performs consistently high when compared to other machine learning approaches. The average classification accuracy is 96.20% (with median 96.85%) when compared to ground truth labelling. Amongst the machine learning approaches, SVM has the closest result to DCNNs, with the accuracy of 91.10% (median 90.95%). The poorest performance was obtained by RUSBoost, with accuracy of 63.39% (median 62.35%). These results suggest that DCNNs is capable of identifying unstable motions with minimum error.

Conclusion

It appears that the experimental results demonstrate our proposed feature set combined with deep learning provides a high classification accuracy and could provide greater insights for a clinician in developing rehabilitation strategies or to aid in confidence boosting. The ability of DCNNs to recognise motions over other standard machine learning techniques is apparent for our experiments.

References

[Ijjina and Mohan, 2014] Ijjina, E. P. and Mohan, C. K. (2014). Human action recognition based on mocap information using convolution neural networks. In *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*, pages 159–164.

[Leightley et al., 2015] Leightley, D., Yap, M. H., J. Coulson, Y. B., and Mcphee, J. S. (2015). Benchmarking human motion analysis using kinect one: an open source dataset. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–7.

[Li et al., 2015] Li, S., Zhang, W., and Chan, A. B. (2015). Maximum-margin structured learning with deep networks for 3d human pose estimation. In *The IEEE International Conference on Computer Vision (ICCV)*.

[Ye et al., 2013] Ye, M., Zhang, Q., Wang, L., Zhu, J., Yang, R., and Gall, J. (2013). A survey on human motion analysis from depth data. In *Lecturer Notes in Computer Science*, pages 149–187.