Establishing Pose Based Features Using Histograms for the Detection of Abnormal Infant Movements

Kevin D. McCay[1], Edmond S. L. Ho[1], Claire Marcroft[2] and Nicholas D. Embleton[2]
[1]Northumbria University, Newcastle upon Tyne, UK and [2]Royal Victoria Infirmary, Newcastle upon Tyne, UK

***Abstract*— The pursuit of early diagnosis of cerebral palsy has been an active research area with some very promising results using tools such as the General Movements Assessment (GMA). In this paper, we conducted a pilot study on extracting important information from video sequences to classify the body movement into two categories, normal and abnormal, and compared the results provided by an independent expert reviewer based on GMA. We present two new pose-based features, Histograms of Joint Orientation 2D (HOJO2D) and Histograms of Joint Displacement 2D (HOJD2D), for the pose-based analysis and classification of infant body movement from video footage. We extract the 2D skeletal joint locations from 2D RGB images using Cao et al.'s method [1]. Using the MINI-RGBD dataset [2], we further segment the body into local regions to extract part specific features. As a result, the pose and the degree of displacement are represented by histograms of normalised data. To demonstrate the effectiveness of the proposed features, we trained several classifiers using combinations of HOJO2D and HOJD2D features and conducted a series of experiments to classify the body movement into categories. The classification algorithms used included k-Nearest Neighbour (kNN, k=1 and k=3), Linear Discriminant Analysis (LDA) and the Ensemble classifier. Encouraging results were attained, with high accuracy (91.67%) obtained using the Ensemble classifier.**

I. Introduction

The ability to recognise, analyse and reconstruct complicated motion, such as human action, has been a topic of research for several years. This research has subsequently led to the idea that human action recognition could be undertaken by computers, allowing for the automation of various activities traditionally requiring human input. Computer vision, specifically relating to human action recognition, has consequently been an active area of research for almost three decades [1]. The automated recognition of human activity has wide ranging applications including visual surveillance, content based video indexing, intelligent monitoring, human-machine learning and virtual reality [2]. This project proposes that this technology could be applied to the paediatric healthcare domain to aid with the early diagnosis of movement disorders, such as cerebral palsy.

Cerebral palsy is a condition which primarily affects movement, posture and coordination. It is an umbrella term which covers a group of lifelong neurological conditions usually caused by a brain injury occurring before, during or shortly after birth [3]. It is estimated that 1 in every 400 babies born in the UK have some form of cerebral palsy [4]. These figures suggest that there may be as many as 1,800 new cases of cerebral palsy every year, making it the most common motor disability found in children [5]. Recent advances in neonatal care have also meant that, whilst there has been a decline in infant mortality rates, there has been an increase in the incidence and severity of cerebral palsy [6].

To provide the best possible outcome, early diagnosis is seen as a key area of interest as it has the potential to allow for early intervention. Early identification also allows for the targeting of resources and for the development of parental support systems. Additionally, access to health social and educational services often rely upon a diagnosis [7].

Whilst early interventions can take a variety of forms, common challenges are found in the prediction and subsequent diagnosis of cerebral palsy [8]. The identification of those at highest risk typically involves a combination of neuroimaging results, clinical history, clinical assessments and the general experience of healthcare professionals [9], with diagnosis often not being confirmed until 18 months of age, or later in the case of those who present mild symptoms [5].

Due to the many complex factors which affect neurodevelopment predicting the likelihood of a child developing cerebral palsy from a single assessment is particularly challenging [10], as it requires techniques which exhibit both high sensitivity and specificity [9].

The pursuit of early diagnosis of cerebral palsy has been an active research area with some very promising results using tools such as the General Movements Assessment (GMA) [11]. In practice, the ability to apply these assessments is dependent upon the availability of fully trained clinicians. Not only is the training required for assessment using these tools considerable, it is also susceptible to observer fatigue, contains a degree of personal subjectivity and is reliant upon a suitable behavioural state of the infant [12].

The development of automated systems could help to significantly reduce the time and subsequent cost associated with current diagnostic practices, potentially aiding with early diagnosis. Previous studies have attempted to automate this process. However, little work exists that evaluates the viability of using the pose and joint specific movements of the infants.

Pose-based analysis presents several advantages over traditional methods, as such, this paper presents an investigation into the feasibility of pose-based evaluation by attempting to establish specific features for fusion. We propose two new pose-based feature sets, annotate the MINI-RGBD dataset for use in GMA, evaluate the effectiveness of individual feature sets by conducting a series of binary classification experiments, and fuse different extracted feature sets for further evaluation and classification.

The Institution's Ethical Review Board approved all experimental procedures involving human subjects.

II. Related Works

In a preliminary piece of work, Adde et al [13] developed a prototype method of attempting to automate the GMA. This method creates uses background subtraction and identifies the difference between two frames for each pixel in a sequence of video footage. The system then assigns a point value per pixel of 0 or 1 to represent the presence of movement [14].

Stahl et al. [15] followed with a method of predicting cerebral palsy based upon statistical pattern recognition of the infant's spontaneous movements using optical flow. Wavelet frequency decomposition analysis was utilised to determine the time dependent trajectory signals in the optical flow data.

Orlandi et al [16] extracted features from video footage to classify movements as typical or atypical based upon the GMA. Using large displacement optical flow (LDOF) to track movements and obtain velocities, the displacement of each pixel was calculated every 10 frames. Background subtraction was implemented, and features were then extracted and classified using several classifiers.

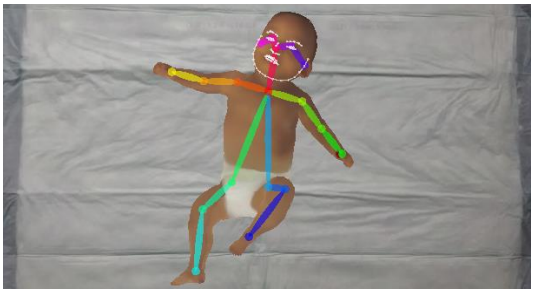

Figure 1. Example output of OpenPose's keypoint detection.

Human 2D pose estimation is an active research area with several key contributors [17]–[19]. In [18], Cao et al. present an approach to detect the 2D pose of multiple people from an image. The framework provides a representation of the keypoints which encodes both the position and orientation of human limbs, the code has also been made publicly available.

III. Methodology

A. Synthetic Dataset

One of the main challenges facing researchers attempting to automate the analysis of video data of infants is the availability of publicly available datasets. Human pose estimation generally focuses on adults, as such a dataset of infants for research purposes can be difficult to obtain. In an effort to solve this [20] proposed a method for the generation of a synthetic dataset. They have produced the Moving INfants In RGB-D (MINI-RGBD) dataset by using their previously developed Skinned Multi-Infant Linear body model (SMIL) [21]. This dataset maps real infant movements to the SMIL model to generate anonymised, synthetic footage. This paper makes use of the MINI-RGBD dataset which currently consists of twelve different sequences.

B. Data Annotation

Each of the twelve video sequences was analysed by an independent expert reviewer using the GMA [11]. The videos were classified into two categories; those who appear to demonstrate movements that we would expect to see in typically developing infants (Normal), and those who demonstrate some movements which may be of concern to clinicians (Abnormal). The reviewer noted that the age of the infant, complex movements such as fingers and facial movements, and additional assessments also play a role in the diagnosis. Since these are currently absent from this dataset a greater degree of speculation is present.

C. Pose Estimation

In this study, we explore the use of features generated from pose estimation to analyse the videos. The advantages of using poses include:

- Lower dimensionality of the features. Body movement can be represented in an abstract manner while keeping the important movement data for analysis.
- Reduced ambiguity in the classification process. Variations in the appearance of the infant caused by superfluous information, such as clothing, lighting changes and body part dimensions are not included. Additionally, external factors like parent intervention will also be disregarded, as such there is less requirement for screening the footage prior to analysis.

With the advancement of deep-learning in recent years, various robust frameworks have been proposed for estimating pose from a 2D image. In this paper, the OpenPose framework [18] is used to extract the 2D poses from the videos in the MINI-RGBD dataset. Each returned pose is represented by the 2D (x and y) coordinates of 18 landmarks on the body, which include 14 joints on the body and 4 facial landmarks. An example of the pose estimation result is illustrated in Fig 1. We used the 14 body joints for feature extraction as the facial landmarks are unavailable in some frames due to occlusion.

D. Feature Extraction

To represent the body movement, we propose two new histogram-based features. Histogram-based features, such as the Histogram of Oriented Gradients (HOG) [22], have been widely used and have demonstrated good performance in visual recognition tasks. Histogram based approaches condense the information to a lower dimensional range whilst retaining a full impression of the associated data. By using feature descriptors, a representation of an image is produced that simplifies the information, extracting useful information.

This allows us to look at the distribution of the orientation and displacement of all the joints over a period of time instead of directly having a frame-to-frame comparison. By doing this, we can avoid finding the correspondence between two frames, which requires solving the time misalignment and variation in speed, as well as handling the difference in duration of video footage recorded of different participants. This enables us to analyse the footage holistically and examine the overall movement structure.

Inspired by the success shown in HOJ3D [23] we propose the following pose-based histogram features in 2D:

i) Histogram of Joint Orientation 2D (HOJO2D)

In this representation the 2D space is segmented into *n* bins which denote the prevalent angle of joint orientation. The joint orientation is computed by

calculating the alignment of the bone connecting a joint and parent joint:

$$bone = j_i - j_{i\text{-}parent} \quad (1)$$

where j_i and $j_{i\text{-}parent}$ are the vectors containing the 2D coordinates of the *i*-th joint and its parent joint.

We manually select the joint range to extract part specific information before a suitable bin is assigned for each joint per frame. As a result, the pose is represented by an *n* bin histogram of normalised data.

ii) Histogram of Joint Displacement 2D (HOJD2D)

In this representation the displacement of each joint is extracted and recorded every five frames. The displacements are then associated with a relevant bin, each of which represents a regular incremental increase. Again, a range of joints is selected manually for part-based analysis. In this way the displacement can be represented by a histogram of normalised data.

IV. Experimental Results

In this section, we evaluate the effectiveness of using the proposed pose-based features, HOJO2D and HOJD2D, for classifying video footage into two categories (normal and abnormal). The ground truth for comparison being the data annotation carried out by the independent expert reviewer using the GMA, as discussed in section B of the methodology. Specifically, we trained a number of classifiers using different combinations of HOJO2D and HOJD2D. The implementation is done in Matlab using built-in classification algorithms, including k-Nearest Neighbour (kNN, k=1 and k=3), Linear Discriminant Analysis (LDA) and the Ensemble classifier.

A. Experiment settings

Since the MINI-RGBD dataset has a limited number of video sequences (twelve in total), we employ leave-one-out cross-validation to evaluate the performance. We trained each classifier using the features extracted from eleven video clips and used the resultant classifier to predict the class label of the remaining video clip. This process is repeated twelve times to make sure every video clip was evaluated. The average accuracy is reported in this section.

B. Evaluation of HOJO2D

In the experiments, we evaluated the HOJO2D at different levels based on the human body hierarchy. Specifically, we first extract the HOJO2D on each limb. Then, we apply feature fusion to the limb-based features to create a wide range of features: *Arms* – fusing features from both arms; *Legs* – fusing features from both legs; *Limbs* – fusing features from all 4 limbs. We also construed a single histogram, called *Full body*, by using the joint orientation from all body parts.

We further extracted the features using both 8 and 16 bins, the results of which are listed in Table I. The best classification accuracy (91.67%) is obtained in *Arms* and *Limbs* features with 8 bins using the Ensemble classifier. In general, features or fused features which contain joint orientation computed from the arms led to better classification performance. In most cases, applying feature fusion achieved a better result than the basic limb-based features. With this small dataset, features extracted using the 8-bin setting performed better than the 16-bin setting. This could be caused by the large pose variations in the video, resulting in a diverged joint orientation distribution and therefore a reduction in the discriminative power when the number of bins increases.

C. Evaluation of HOJD2D

Similar to Section IV. B, we evaluated HOJD2D using features extracted from different limbs and applied feature fusion to generate new features for classification. The results are listed in Table II.

The best classification accuracy 100.00% is obtained in *Right Leg* and *Legs* features with 16 bins using the ensemble classifier and LDA, respectively. Again, applying feature fusion achieved a better result than the basic limb-based features in most cases, which is consistent with the results obtained from classifying HOJO2D features. Interestingly, features or fused features which contain joint orientation computed from the legs contribute to better classification performance. Also, the classification results obtained from features extracted using the 16-bin setting performed better than an 8-bin setting. This indicates the 16-bin features are more discriminative and suggested that the magnitude of joint displacement is more consistent than the joint orientation within each infant class (i.e. normal/abnormal).

D. Evaluation of Fusing HOJO2D and HOJD2D

In the final experiment, we apply fusion on HOJO2D and HOJD2D. Here, we focused on evaluating 3 features, including *Arms*, *Legs* and *Limbs*, with the 8-bin and 16-bin settings. The results are listed in Table III.

In most cases, the results indicate that fusing HOJO2D and HOJD2D achieved better classification accuracy compared with using HOJO2D or HOJD2D individually. The best performance is obtained in all features with the 8-bin setting and the Legs feature with the 16-bin setting using the ensemble classifier. This suggests a more robust classifier can be obtained by training with the fused features.

V. Conclusion

In this project, we proposed two new pose-based features, HOJO2D and HOJD2D, for analysing and classifying body movement of an infant from video footage. To evaluate the effectiveness of the proposed features, we conducted a series of experiments to classify the body movement into two categories, normal and abnormal. Encouraging results are obtained as high accuracy (91.67%) can be obtained in a lot of different settings with the ensemble classifier. A few specific part-based settings can obtain 100% classification accuracy.

The proposed features demonstrate the benefits of pose-based analysis by dealing with information which can traditionally affect classification results, such as loose clothing, illumination changes, body part variation, background clutter and other people being present in shot.

The video sequences used here are synthetic which means that the appearance of the infant is slightly different to that of real data. We would like to test our proposed method on real video footage of infants, as additional details may also be included. Also, the quantity of the video sequences included in the MINI-RGBD dataset is small. We will look to extend this work by classifying a larger dataset.

TABLE I. HOJO2D CLASSIFICATION ACCURACY

features			*classification accuracy (%)*		
Type	**bins**	**kNN(k=1)**	**kNN(k=3)**	**LDA**	**Ensemble**
Left Arm	8	50.00	66.67	66.67	83.33
Right Arm		58.33	16.67	25.00	25.00
Left Leg		33.33	66.67	58.33	33.33
Right Leg		50.00	50.00	33.33	83.33
Left Arm	16	75.00	75.00	83.33	83.33
Right Arm		50.00	16.67	33.33	33.33
Left Leg		16.67	41.67	58.33	33.33
Right Leg		33.33	41.67	66.67	33.33
Arms	8	75.00	75.00	66.67	**91.67**
Legs		25.00	41.67	50.00	41.67
Limbs		41.67	58.33	58.33	**91.67**
Full body		66.67	50.00	50.00	58.33
Arms	16	66.67	66.67	75.00	66.67
Legs		25.00	8.33	41.67	33.33
Limbs		66.67	66.67	83.33	66.67
Full body		83.33	50.00	66.67	75.00

TABLE II. HOJD2D CLASSIFICATION ACCURACY

features			*classification accuracy (%)*		
Type	**bins**	**kNN(k=1)**	**kNN(k=3)**	**LDA**	**Ensemble**
Left Arm	8	83.33	50.00	66.67	75.00
Right Arm		50.00	66.67	66.67	83.33
Left Leg		41.67	66.67	66.67	50.00
Right Leg		83.33	58.33	58.33	66.67
Left Arm	16	66.67	50.00	66.67	75.00
Right Arm		58.33	33.33	33.33	33.33
Left Leg		66.67	75.00	5000	75.00
Right Leg		83.33	58.33	83.33	**100.00**
Arms	8	66.67	50.00	33.33	50.00
Legs		50.00	58.33	58.33	91.67
Limbs		75.00	75.00	75.00	91.67
Full body		58.33	58.33	58.33	66.67
Arms	16	58.33	50.00	50.00	50.00
Legs		66.67	58.33	**100.00**	91.67
Limbs		58.33	58.33	83.33	91.67
Full body		58.33	58.33	66.67	75.00

TABLE III. HOJO2D + HOJD2D CLASSIFICATION ACCURACY

features			*classification accuracy (%)*		
Type	**bins**	**kNN(k=1)**	**kNN(k=3)**	**LDA**	**Ensemble**
Arms	8	75.00	50.00	66.67	**91.67**
Legs		50.00	58.33	50.00	**91.67**
Limbs		66.67	58.33	83.33	**91.67**
Arms	16	75.00	58.33	75.00	66.67
Legs		41.67	50.00	58.33	**91.67**
Limbs		66.67	66.67	83.33	66.67

REFERENCES

[1] J. K. Aggarwal and L. Xia, "Human activity recognition from 3D data: A review," *Pattern Recognit. Lett.*, vol. 48, pp. 70–80, 2014.

[2] L. Yao, Y. Liu, and S. Huang, "Spatio-temporal information for human action recognition," *Eurasip J. Image Video Process.*, vol. 2016, no. 1, p. 39, Dec. 2016.

[3] NHS, "Cerebral palsy - NHSUK," *15/03/2017*, 2017. [Online]. Available: https://www.nhs.uk/conditions/cerebral-palsy/. [Accessed: 14-May-2018].

[4] "Introduction to cerebral palsy - CP - Disability Charity - Scope UK." [Online]. Available: https://www.scope.org.uk/support/families/diagnosis/cerebral-palsy. [Accessed: 14-May-2018].

[5] CerebralPalsy.org.uk, "Cerebral Palsy." [Online]. Available: http://www.cerebralpalsy.org.uk/. [Accessed: 14-May-2018].

[6] C. P. Panteliadis, "Cerebral palsy : a multidisciplinary approach," in *Cerebral palsy : a multidisciplinary approach*, 2018, pp. 1–349.

[7] A. H. Shevell and M. Shevell, "Doing the 'talk': Disclosure of a diagnosis of cerebral palsy," *J. Child Neurol.*, vol. 28, no. 2, pp. 230–235, 2013.

[8] A. P. Basu and G. Clowry, "Improving outcomes in cerebral palsy with early intervention: new translational approaches," *Frontiers in Neurology*, vol. 6:24. 2015.

[9] C. Marcroft, A. Khan, N. D. Embleton, M. Trenell, and T. Plötz, "Movement recognition technology as a method of assessing spontaneous general movements in high risk infants," *Front. Neurol.*, vol. 6, no. JAN, p. 284, 2015.

[10] A. Spittle, "How do we use the assessment of general movements in clinical practice?," *Dev. Med. Child Neurol.*, vol. 53, no. 8, pp. 681–682, 2011.

[11] C. Einspieler and H. F. R. Prechtl, "Prechtl's assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 11, no. 1. Wiley-Blackwell, pp. 61–67, 01-Feb-2005.

[12] T. Fjørtoft, C. Einspieler, L. Adde, and L. I. Strand, "Inter-observer reliability of the 'Assessment of Motor Repertoire - 3 to 5 Months' based on video recordings of infants," *Early Hum. Dev.*, vol. 85, no. 5, pp. 297–302, 2009.

[13] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, and R. Støen, "Using computer-based video analysis in the study of fidgety movements," *Early Hum. Dev.*, vol. 85, no. 9, pp. 541–547, 2009.

[14] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, K. H. Grunewaldt, and R. StØen, "Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study," *Dev. Med. Child Neurol.*, vol. 52, no. 8, pp. 773–778, 2010.

[15] A. Stahl, C. Schellewald, Ø. Stavdahl, O. M. Aamo, L. Adde, and H. Kirkerod, "An optical flow-based method to predict infantile cerebral palsy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 605–614, 2012.

[16] S. Orlandi, K. Raghuram, C. R. Smith, D. Mansueto, P. Church, V. Shah, M. Luther, and T. Chau, "Detection of Atypical and Typical Infant Movements using Computer-based Video Analysis," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2018, pp. 3598–3601, 2018.

[17] R. A. Güler, N. Neverova, and I. Kokkinos, "DensePose: Dense Human Pose Estimation In The Wild," 2018.

[18] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," *Cvpr*, vol. XXX, no. Xxx, pp. 1302–1310, 2017.

[19] H. S. Fang, S. Xie, Y. W. Tai, and C. Lu, "RMPE: Regional Multi-person Pose Estimation," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017–Octob, pp. 2353–2362, 2017.

[20] N. Hesse, C. Bodensteiner, U. G. Hofmann, and A. S. Schroeder, "Computer Vision for Medical Infant Motion Analysis : State of the Art and RGB-D Data Set."

[21] N. Hesse, S. Pujades, J. Romero, M. J. Black, C. Bodensteiner, M. Arens, U. G. Hofmann, U. Tacke, M. Hadders-Algra, R. Weinberger, W. Müller-Felber, and A. S. Schroeder, "Learning an Infant Body Model from RGB-D Data for Accurate Full Body Motion Analysis."

[22] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.

[23] L. Xia, C. C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 20–27.