Infant Posture Assessment Based on Rotational Keypoint Detection

Xuyang Zhao*†, Shogo Takata*, Kosuke Fukumori* and Toshihisa Tanaka*†
* Tokyo University of Agriculture and Technology, Tokyo, Japan
E-mail: zhao@sip.tuat.ac.jp, s196643z@st.go.tuat.ac.jp, fukumori17@sip.tuat.ac.jp, tanakat@cc.tuat.ac.jp
† RIKEN Center for Advanced Intelligence Project, Tokyo, Japan

***Abstract*—The risk of an infant's developing neuromotor impairment is primarily diagnosis through visual examination by clinical experts. Due to the visual diagnosis is a very professional process, many infants at risk are not detected in time and develop into more severe conditions. There are some posture assessment diagnosis methods based on image/video data, but the performance will be affected by the infant's complex and diverse posture. In this paper, we propose a simple but effective method to improve infant posture assessment performance. Before the posture assessment, we first judge the direction of the infant's body in the image and calculate the inclination angle of the direction, then adjust the angle of the image to correct the infant's body orientation to upright. Last, the adjusted image will be feed into a model for posture assessment. Experimental results demonstrate that our proposed angle adjustment method can further improve the performance of the posture assessment for infants.**

I. INTRODUCTION

Neuromotor impairment affecting the health of infants and may cause lifelong disability. Usually, the disorder can affect movement, posture and manual abilities include cerebral palsy, muscular dystrophy and spina bifida [1]. Cerebral palsy is caused by damage to the immature brain, which affects the person's movements and also affects social, cognitive and perceptual skills. In current clinical diagnosis, early detection and intervention is the most important key for the treatment [2], [3]. If the diagnosis and intervention can be started within the first year of life, the treatment can improve the motor outcomes in infancy, childhood and adolescence [4]. The physical activities of infants can be quantified by observing their posture and movement. The motor delays and atypical movements are used for diagnosis the neuromotor impairment [5].

Some methods have been developed to evaluate neuromotor impairment risk; the General Movements Assessment (GMA) and the Hammersmith Infant Neurological Examination are the most commonly used methods that have a high performance [6]. The GMA is a standardized qualitative visual assessment of an infant's neuromotor impairment [7], [8] and performed by record a video to extract movement features as normal or abnormal and requires the clinical expert to observe the recorded data and give a score based on the pose and movement of the infant. The produce of the scoring stage is implemented according to the personal judgment of each clinician, which is subjective and difficult to standardize. Because of this, nowadays, the efficiency and accuracy of the GMA evaluation are low in clinical diagnosis and often inconsistencies in the diagnosis of the degree of the disease. Therefore, there is a high demand for quantitative analysis to extract disease information in the clinical diagnosis.

In recent years, with the development of computer vision technology, some motion measurement and analysis systems have been proposed that can quantitative evaluation. Background subtraction and inter-frame difference [9], [10] is used to extract the infant's movement information and quantitative evaluation. Stahl et al. [11] applied optical flow and statistical pattern recognition to extract motion information. Orlandi et al. [12] analyzed consisting of a skin model for segmentation and large displacement optical flow (LDOF) for motion tracking. With the development of deep learning and the growth of the dataset in size and quality, posture estimation methods [13], [14], [15] have shown great potential. Some methods [16], [17], [18] applied the posture estimation algorithms to a quantitative evaluation in an infant's neuromotor impairment. Usually, those methods consist of two steps: (i) posture estimation and extract the keypoints of the infant; (ii) calculate the infant's movement feature based on the information of keypoints and quantitative evaluation. The performance of those methods has a strong correlation with whether the keypoints are detected correctly or not. The current keypoint detection is based on deep learning, and the model's performance depends on the distribution of the training data. For example, Common Objects in Context (Microsoft COCO) [19] and MPII Human Pose Dataset [20] are well-known publicly available datasets of human poses. However, in those datasets, the pose images were predominantly collected from TV, sports, daily life and so on; thus, most of the subjects in the image are adult humans with an upright posture. This leads to the model's performance in the task of handling infants that cannot be fully utilized.

The data record from infants usually includes complex and diverse poses, such as lying horizontally, upright, upside down and so on. When we want to use some trained models to detect the keypoints, it is necessary to prepare a special infant dataset that includes diverse body orientations. However, to be able to collect a helpful infant dataset, we need a large enough size of the dataset that will cause a problem of privacy and the annotation is also meet a huge challenge. In order to avoid the problem of building an infant-specific dataset, we need to adapt the infant's data to the trained keypoint detection model. We propose a simple and effective method, before using the model to detect the keypoints, first, calculate the infant's body angle in the image. After obtaining the body tilt angle, adjust the image to make the infant's body recover to upright. In the experiment, we use the OpenPose [13], [14] model to detect the keypoints. The OpenPose is applied twice; in the first time, the keypoints of 1 (Neck) and 8 (Midhip) are detected, based on the two points, the body tilt angle can be calculated. Adjust the image based on the calculated angle so that the infant in the image is displayed in an upright position. The experimental results show that with the help of angle adjustment, the detection performance of the model has increased by 8%, and the result is obtained without changing the keypoint detection model and preparing a large-scale dataset.

The rest of the paper is organized as follows: Section 2 introduces the dataset used for the evaluation of the proposed method and describes the angle adjustment method in posture estimation. Section 3 describes the experimental result and the conclusions are presented in Section 4.

II. Methods

A. Dataset

In this paper, the Synthetic and Real Infant Pose (SyRIP) dataset [21] is used to evaluate the proposed method. In the SyRIP dataset, all the images are collected from YouTube videos (one of the frames) and Google Images, containing a variety of rich infant poses like crawling, lying, sitting, and so on. SyRIP dataset includes 200 real and 1000 synthetic infant images in the training part and 500 real infant images in the test part. All images with fully annotated body joints (17 keypoints with the nose, left eye, right eye, left ear, right ear, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, left hip, right hip, left knee, right knee, left ankle, right ankle). Same to the format of Microsoft COCO dataset, the groundtruth keypoints also have the form of $[x_1, y_1, v_1, ..., x_{17}, y_{17}, v_{17}]$, where x, y are the keypoint location and v is a visibility flag defined as $v = 0$: not labeled, $v = 1$: labeled but not visible, and $v = 2$: labeled and visible. Each groundtruth object also has parameter s, which define as the square root of the object segment area. Some examples (SyRIP) are shown in Fig. 1.

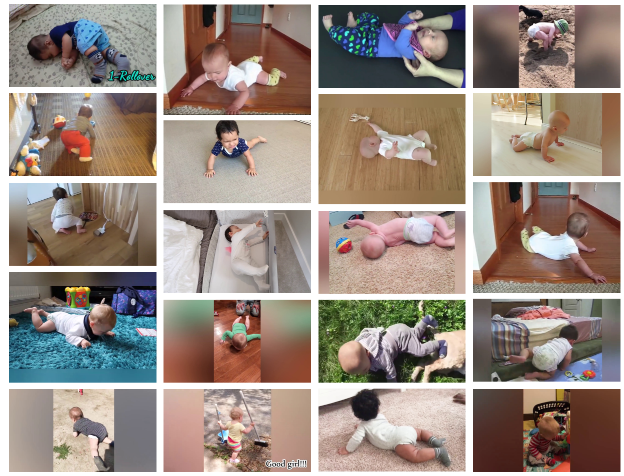

Fig. 1. Some examples in SyRIP dataset, contains various infant's daily poses.

B. Pose Estimation

Human 2D pose estimation is a problem that automatically detects the skeleton point from the body [14]. There are some methods that use person detector and pose estimation for each detection, Gkioxari et al. [22] develop a person detection and keypoints prediction model accomplished using a representation based on k-poselets. Some approaches combine the local observations on the body parts and the spatial dependencies to perform the human posture estimation. Tree-structured graphical [23] is used, which represents the object by a collection of parts arranged in a deformable configuration. Multiple tree models [24] are combined for human pose estimation, which obtains the information across different tree models. Recently, convolutional neural networks (CNN) have been widely used in posture estimation, approach [25] across all scales and consolidated to extract the spatial relationship on the body and combine the bottom-up and top-down processing to improve the performance. Pfister et al. [26] presented a new architecture that can utilize appearances across multiple frames by combining the heatmaps with spatial fusion layers and optical flow.

With the construction of the large-scale dataset and the increase in computing power, some posture estimation models based on complex neural networks have been proposed and continuously refreshing the performance records. Among these methods, OpenPose model [13], [14] shows excellent performance and computing speed. Therefore, in this paper, we apply OpenPose model to assess the infant posture. OpenPose model constructs a bottom-up system by using nonparametric representation to extract the body parts with high accuracy and also can process multi-people in the image. The keypoint association that encodes both position and orientation of human limbs are represented and the architecture is designed to learn part detection and association jointly. The greedy parsing algorithm is sufficient to produce high performance parses of body poses and preserves efficiency regardless of the number of people. Part Affinity Fields (PAFs) algorithm used to learn to associate body parts with individuals in the image. Moreover, the OpenPose model can work as a realtime system, that is widely used today for many research topics.

C. Angle Adjustment Method

Usually, the infant's data includes complex and diverse poses, such as lying horizontally, upright, upside down, which is different from the public dataset. That will have a certain impact on the performance when applying trained models to the posture estimation. In order to use the trained model by the public dataset, we need to adjust the infant's data to fit the training dataset. Before the posture assessment, we first use OpenPose to detect the keypoints of 1 (Neck) and 8 (Midhip), and based on the two points to calculate the body tilt angle.

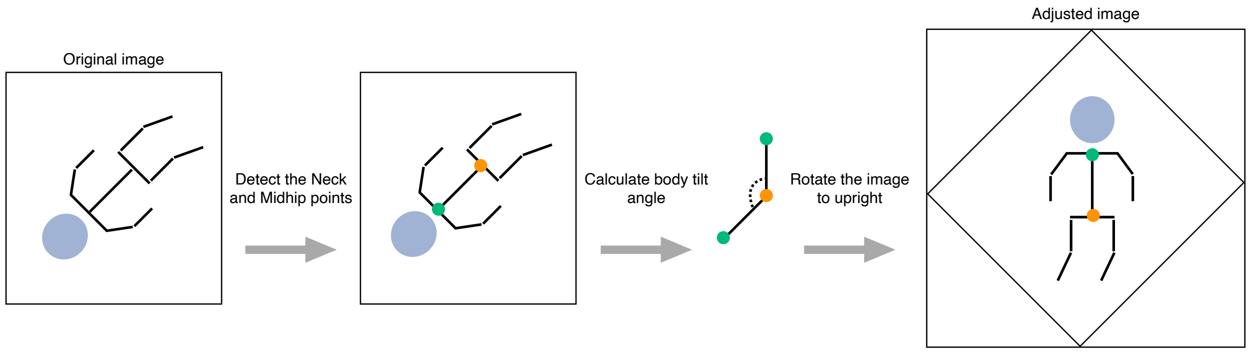

Fig. 2. The overall architecture of the image rotation process. The image is rotated to upright by using the neck and midhip keypoints.

Next, rotate the image based on the body tilt angle to make sure the infant in the image is in an upright position. After the adjustment, the image will feed into OpenPose again for posture estimation.

In the procedure of body tilt angle estimation, the body tilt angle is calculated based on the detected points of the neck and the midhip. However, this will bring a disadvantage, if the OpenPose model can not detect the neck and the midhip keypoints rightly, there will be errors in image rotation. After rotating the image, the keypoints of the neck and midhip will estimate again by using OpenPose and repeat rotation until the body tilt angle is small than a threshold (3 degrees in this paper). In order to make the rotation process more stable, a step by step adjustment method is also used. After calculating the angle of the image, it will first rotate by half of the angle and then perform the angle calculation based on OpenPose again, repeat the rotation and angle calculation until the body tilt angle is small than a threshold. With the rotation of the image, the size of the image will become a square (the side length is the diagonal length of the original image). The image rotation process of the image is shown in Fig 2.

III. Experimental Results and Discussion

In the experiment, the SyRIP dataset (500 real infant images in the test part) is used to evaluate the proposed method. During the rotation, five different adjust rates are used, adjust rate means rotate the image by $1/r$ $(r = 1, 2, 3, 4, 5)$ of the calculated body tilt angle. Microsoft COCO evaluation method is used, which defines the object keypoint similarity (OKS) and mean average precision (AP) over ten thresholds as the evaluation value. The d_i is the euclidean distances between each corresponding detected keypoint and groundtruth, s is define as the square root of the object segment area, k_i is substantially for different keypoints, keypoints on a person's body (shoulders, knees, hips, etc.) tend to have a much larger value than on a person's head (eyes, nose, ears). The OKS is defined as:

$$OKS = \frac{\sum_i \left[\exp(-\frac{d_i^2}{2s^2k_i^2}) \cdot \delta(v_i > 0)\right]}{\sum_i \left[\delta(v_i > 0)\right]} \tag{1}$$

TABLE I
Results of the OpenPose model and angle adjustment method on the SyRIP dataset, AP and AP at OKS = 0.5 & 0.75 are evaluated separately.

Methods	AP	$AP^{OKS=.50}$	$AP^{OKS=.75}$
OpenPose	80.60	81.73	80.96
Adjust (rate=1)	87.04	88.70	87.63
Adjust (rate=2)	88.07	89.56	88.62
Adjust (rate=3)	87.81	89.39	88.33
Adjust (rate=4)	88.08	89.57	88.59
Adjust (rate=5)	87.76	89.27	88.32

The results with different adjust rate are shown in Table I and some output samples also shown in Fig 3. By using angle adjustment method, the posture assessment performance has been well improved, especially when the infant's posture in the image is horizontal.

IV. Conclusions

Due to the diagnosis through visual examination is professional in the infant's developing neuromotor impairment, some infants are not intervened in time and develop into serious conditions. Although some computer vision-based assistance systems have been proposed. However, the infant's posture has different characteristics from the adult dataset, and various body directions often affect the performance of the posture estimation models. In order to adapt the infant's data to trained models, the direction of the infant's body is adjusted before detecting the keypoints. According to experimental results, the angle adjustment method shows its effectiveness in posture estimation. Furthermore, the angle adjustment method does not require modifications to the posture estimation model nor does it rely on reconstructing a large number of the infant-specific dataset. With the angle adjustment, the performance of the posture estimation model can be significantly improved.

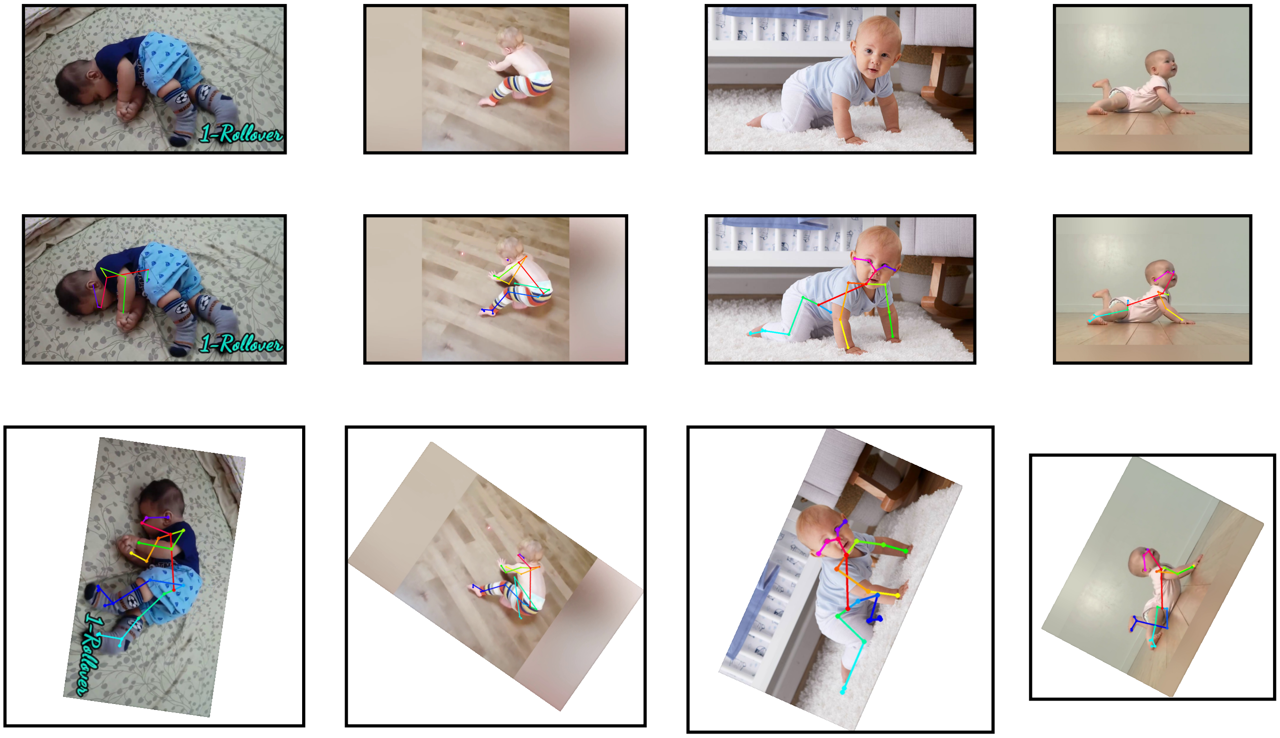

Fig. 3. Some posture assessment results from the SyRIP dataset, the first line is the original image, the second line is the evaluation result of OpenPose model and the third line is the OpenPose model evaluation result with the angle adjustment images.

ACKNOWLEDGMENT

This work was supported by JST CREST Grant Number JPMJCR1784 including AIP challenge program, Japan.

REFERENCES

[1] B. Klein, C. P. Society, M. Health, and D. D. Committee, "Mental health problems in children with neuromotor disabilities," *Paediatrics & child health*, vol. 21, no. 2, pp. 93–96, 2016.

[2] I. Novak, C. Morgan, L. Adde, J. Blackman, R. N. Boyd, J. Brunstrom-Hernandez, G. Cioni, D. Damiano, J. Darrah, A.-C. Eliasson *et al.*, "Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment," *JAMA pediatrics*, vol. 171, no. 9, pp. 897–907, 2017.

[3] A. Herskind, G. Greisen, and J. B. Nielsen, "Early identification and intervention in cerebral palsy," *Developmental Medicine & Child Neurology*, vol. 57, no. 1, pp. 29–36, 2015.

[4] A. Spittle, J. Orton, P. J. Anderson, R. Boyd, and L. W. Doyle, "Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants," *Cochrane database of systematic reviews*, no. 11, 2015.

[5] L. Zwaigenbaum, S. Bryson, and N. Garon, "Early identification of autism spectrum disorders," *Behavioural brain research*, vol. 251, pp. 133–146, 2013.

[6] D. M. Romeo, D. Ricci, C. Brogna, and E. Mercuri, "Use of the hammersmith infant neurological examination in infants with cerebral palsy: a critical review of the literature," *Developmental Medicine & Child Neurology*, vol. 58, no. 3, pp. 240–245, 2016.

[7] M. Hadders-Algra, "General movements: a window for early identification of children at high risk for developmental disorders," *The Journal of pediatrics*, vol. 145, no. 2, pp. S12–S18, 2004.

[8] C. Einspieler and H. F. 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, pp. 61–67, 2005.

[9] 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 human development*, vol. 85, no. 9, pp. 541–547, 2009.

[10] Y. Osawa, K. Shima, N. Bu, T. Tsuji, T. Tsuji, I. Ishii, H. Matsuda, K. Orito, T. Ikeda, and S. Noda, "A motion-based system to evaluate infant movements using real-time video analysis," in *13th international conference on biomedical engineering*. Springer, 2009, pp. 2043–2047.

[11] 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 Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 4, pp. 605–614, 2012.

[12] 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," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 3598–3601.

[13] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *CVPR*, 2017.

[14] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[15] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[16] C. Chambers, N. Seethapathi, R. Saluja, H. Loeb, S. R. Pierce, D. K. Bogen, L. Prosser, M. J. Johnson, and K. P. Kording, "Computer vision to automatically assess infant neuromotor risk," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 11, pp. 2431–2442, 2020.

[17] M. D. Olsen, "Motion tracking of infants in risk of cerebral palsy," Ph.D. dissertation, Technical University of Denmark, 2016.

[18] V. Marchi, A. Hakala, A. Knight, F. D'Acunto, M. L. Scattoni, A. Guzzetta, and S. Vanhatalo, "Automated pose estimation captures key

aspects of general movements at eight to 17 weeks from conventional videos," *Acta Paediatrica*, vol. 108, no. 10, pp. 1817–1824, 2019.

[19] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[20] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," in *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, 2014, pp. 3686–3693.

[21] X. Huang, N. Fu, S. Liu, and S. Ostadabbas, "Invariant representation learning for infant pose estimation with small data," *arXiv preprint arXiv:2010.06100*, 2021.

[22] G. Gkioxari, B. Hariharan, R. Girshick, and J. Malik, "Using k-poselets for detecting people and localizing their keypoints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3582–3589.

[23] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International journal of computer vision*, vol. 61, no. 1, pp. 55–79, 2005.

[24] Y. Wang and G. Mori, "Multiple tree models for occlusion and spatial constraints in human pose estimation," in *European Conference on Computer Vision*. Springer, 2008, pp. 710–724.

[25] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European conference on computer vision*. Springer, 2016, pp. 483–499.

[26] T. Pfister, J. Charles, and A. Zisserman, "Flowing convnets for human pose estimation in videos," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1913–1921.