

대한물리치료과학회지

Journal of Korean Physical Therapy Science
2024. 12. Vol. 31, No 4, pp.116-123

Motion Estimation System for Monitoring Infant and Toddler Motor Development: A Feasibility Study

Jungae An¹, Ph.D, P.T., Kyung-Su Kang², Ph.D, Byoung-Hee Lee¹, Ph.D., P.T.

¹Dept. of Physical Therapy, Sahmyook University

²Dept. of Architecture, Sahmyook University

Abstract

Background: Infants differ from adults in their distribution of body proportions and spontaneous movements. An appropriate video frame environment is required to estimate infant and toddler movements. This study presents the feasibility of developing a deep-learning-based motion estimation system that predicts the musculoskeletal key points of infants and toddlers and recognizes motion using the Kinect system.

Design: A Feasibility study

Methods: This study investigates implementing an infant and toddler motion estimation system based on infant and toddler motion observation data captured using the Kinect system.

A total of 26 infants aged between 8 weeks and 24 months will be observed using the Kinect system over 20 days for one hour at a time, and data will be collected. In addition, YOLO Pose, which integrates RGB and depth data, is used to predict and detect information on infant musculoskeletal key points in the video. The infant movement observation data include movement observations according to the developmental period and data of postures such as prone, supine, creeping, sitting, standing, and walking. For infant motion estimation, information regarding the in-

fant motion and positional range in the image is stored in an annotation file, and deep learning is performed. The accuracy of the artificial intelligence (AI) is periodically evaluated using verification data. A PC-based software program is used to analyze the motor development of infants and toddlers. This application provides information such as image registration, infant motion analysis screens, previews, and AI analysis result screens for AI image analysis.

Conclusion: This study explored the possibility of using information from key points in infants to develop a motion estimation system that can monitor infant motor development at different stages

Key words: preterm birth, cerebral palsy, motion, dataset

Corresponding author

Prof. Byoung-Hee Lee
815, Hwarang-ro, Nowon-gu, Seoul, Republic of Korea
T: 02-3399-1634, E: 3679@syu.ac.kr

I . Introduction

According to the World Health Organization (WHO), in 2020, an estimated 13.4 million preterm infants (born before 37 weeks of gestation) were recorded and the prevalence rates ranged from 4% to 16% across different countries. Preterm infants are at a significantly increased risk of various neurodevelopmental challenges, including impairments in cognition, language, and learning as well as motor developmental disorders. Cerebral palsy is the most common motor developmental disorder affecting preterm infants (Zwicker, 2014).

The early diagnosis of motor developmental delay is challenging; however, neurodevelopmental disorders such as cerebral palsy and autism spectrum disorder can be predicted one month after birth through observations of spontaneous movements (Novak et al., 2017; Einspieler et al., 2019). Identifying children who are at risk for developmental disorders early allows for the timely initiation of appropriate therapeutic interventions, which can lead to improved outcomes and minimize long-term disabilities (Støen et al., 2017).

Various methods are available for the early diagnosis of motor developmental delays or disabilities. General Movement Assessment is a tool that evaluates abnormal movement patterns in infants through clinical observation by assigning scores based on the complexity and abilities demonstrated during specific movement milestones. However, the variability and complexity of infant movements make it challenging to apply this method without extensive expertise. Consequently, there is a growing demand for quantitative measurement methods that provide more objective assessments of infant movements (Nakajima et al., 2006).

Motion analysis systems use markers attached to the body to track major skeletal structures and joints. These markers are monitored by optical systems that provide information regarding the position and orientation of body segments. Such a system minimizes the subjective bias in the evaluation, making it effective for objectively assessing infant movements. However, limitations include the need for marker attachment, reliance on specialized laboratory settings, and high costs (Mousavi Hondori et al., 2014). In contrast, the Kinect system, which is based on depth imaging, offers a cost-effective, portable, and user-friendly solution for capturing body movements. Its application extends to the home environment, making it valuable in rehabilitation settings (Tolgyessy et al., 2021). Motion estimation involves detecting body movements, reconstructing them into three-dimensional (3D) scenes, and evaluating related movements. While traditional motion estimation systems mainly rely on two-dimensional images, the depth sensors of Kinect enable 3D skeletal tracking. This technology provides detailed information on joint positions and orientations, allowing for more precise detection of even subtle movements (Andriluka et al., 2014; Groos et al., 2022).

Despite the differences in body proportions and spontaneous movements between infants and adults, most existing human motion estimation systems have primarily focused on adult postures. Prior studies on infant motion estimation have revealed the diversity of the video frame environments used, highlighting the need for further research in this area (Adde et al., 2021). Therefore, this study aims to validate the development of a deep-learning-based motion estimation system utilizing the Kinect system to predict musculoskeletal key points and recognize movements in infants.

II. Method

1. Participants

The sample size required for this study was calculated using G*Power Version 3.1.9.7 (Franz Faul, University Kiel, Germany, 2020) through statistical evaluation. For correlation analysis (one-tail, correlation ρ H1: 0.5, α error: 0.05, power: 0.8), a total of 23 participants will be used. Considering a dropout rate of 10%–15%, 26 participants will be recruited. The study will involve 26 infants aged 8 weeks to 24 months recruited from two hospitals in Gyeonggi Province: C Women's Hospital and M Women's Hospital. Recruitment will be conducted via public announcements. Infant movement data will be collected through motion observation. The guardians of the participants will receive a detailed explanation of the research process from the principal investigator and provide written informed consent prior to participation.

The inclusion criteria for the participants in this study require obtaining written consent from the guardians of the infants. Eligible participants are full-term infants born at 37 weeks or later through normal delivery. The exclusion criteria include preterm infants born before 37 weeks of gestation, infants born with a birth weight below 2.5 kg, and infants with congenital diseases such as congenital heart defects, congenital pulmonary conditions, or musculoskeletal deformities. Participants will also be excluded if written consent from their guardians cannot be obtained.

2. Method

1) Procedure

This study aims to develop an infant motion estimation system based on the observational data of infant movements recorded using the Kinect system. Before the experiment, the general characteristics of the participating infants, including sex, age (in months), height, and weight, will be collected. Written informed consent will be obtained in advance from the guardians of the infants who meet the inclusion criteria. Motion observation data will be gathered by conducting 20 days of one-hour observation sessions using the Kinect system. To accommodate the feeding schedules, sleep patterns, and general condition of the infants, the home visit schedule for data collection will be coordinated with their guardians. During the observation, the Kinect camera will be securely mounted on the infant's crib or installed using a tripod. To ensure safety, the camera will be placed at a secure distance to prevent it from falling or becoming displaced. All data collection will be conducted by the same researcher to maintain consistency, and repeated data collection sessions will always occur in the same location. This approach will ensure uniformity and reliability in the data acquisition process.



Figure 1. Data collection using Azure Kinect camera

2) Collection of infant movement recording data

To collect observational data on infant movements, this study employs the Microsoft Azure Kinect DK (Microsoft, USA, 2020). The Azure Kinect DK is a high-performance camera equipped with a 12 MP RGB sensor, depth sensor, and 360-degree microphone array, enabling precise tracking and monitoring of motion and overall body structure.

For this study, a YOLO Pose-based model named InfantPose has been developed to integrate RGB and depth data to predict musculoskeletal key points and to detect relevant information within infant motion videos. To train and validate this system, motion data corresponding to various developmental stages will be collected in postures such as prone, supine, creeping, sitting, standing, and walking. This approach will ensure comprehensive coverage of infant movements across key developmental milestones.

Table 1. The variables of body segment during action analysis in infants

Activity	ROM	Velocity
Prone, Supine, Creeping, Sitting, Standing, Walking	Wrist	Wrist Peak Velocity
		Wrist Peak Acceleration
	Elbow	Elbow Peak Velocity
		Elbow Peak Acceleration
	Shoulder	Shoulder Peak Velocity
		Shoulder Peak Acceleration
	Hip	Hip Peak Velocity
		Hip Peak Acceleration
	Knee	Knee Peak Velocity
		Knee Peak Acceleration
	Ankle	Ankle Peak Velocity
		Ankle Peak Acceleration

3) Model validation

Using the infant movement data collected from the Microsoft Azure Kinect DK camera for supervised learning, training, validation, and test datasets are established for supervised learning. The data are collected through video recordings, from which frames are extracted for further analysis. In supervised learning, labels corresponding to the expected outcomes are crucial. For infant motion recognition, information regarding the infants' movements and their spatial range from the images will be saved in labeled files. These labeled files will indicate the specific actions and positions of the infants, which are essential for accurate motion estimation and model training.

During the training of the deep learning model, the validation dataset will be used to evaluate the performance of the model periodically in terms of its accuracy and the selected hyperparameters. The model architecture and hyperparameters will be adjusted and modified iteratively to enhance the model performance. The model will undergo multiple iterations, each time adjusting the hyperparameters, to generate the final deep learning model with high validation accuracy. Once the model is optimized, the learned weights and relevant hyperparameter information, along with other artifacts, will be saved for future use. The developed infant motion estimation system is implemented using PyTorch and CUDA version 11.8, and the model will undergo 150 training repetitions to ensure sufficient learning and performance evaluation.

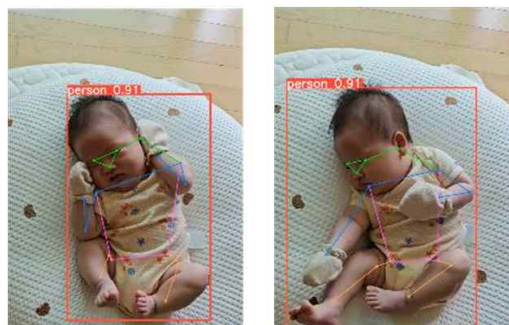


Figure 2. Infants skeletal key points in motion estimation system

4) Implementaion of motion analysis software

To analyze and monitor the normal motor development of infants based on their age, the extracted musculoskeletal key point data will be utilized with a PC-based application. This program will provide various functionalities for AI-assisted video analysis, including video registration, motion analysis screens, preview options, and AI-generated results.



Figure 3. Screen of AI motion analysis

3. Data Analysis

The Azure Kinect DK Software Development Kit (SDK) 1.4.1 was used to develop the camera software for data collection in this study. The data preprocessing software and deep-learning-based computer vision models were implemented using Python 3.10 and the PyTorch 2.3.1 (CUDA 11.8) deep learning framework. The statistical hypothesis testing will utilize SPSS version 18.0 to calculate the means and standard deviations. Descriptive statistics will be applied to analyze the general characteristics of the study participants. The sample group consists of four types of video input: (1) RGB video, (2) RGB + depth video, (3) RGB + IR video, and (4) RGB + depth + IR video. These video samples train two deep learning-based models for infant pose estimation and behavior recognition using feature extractors based on CNN and Transformer architectures.

A comparison of these models will be conducted using evaluation metrics such as the mean average precision and intersection over union with a threshold of 0.5. The performance differences between the models will be analyzed using a custom-built validation dataset, and the statistical differences between the models will be tested using a t-test. To analyze the intra- and inter-rater reliabilities of the motion recognition results obtained from the experimental and home cameras, the intra-class correlation (ICC) will be calculated. The ICC values will be interpreted as follows: 0.0–0.40 indicates "poor" reliability, 0.40–0.59 denotes "fair" reliability, 0.60–0.74 is considered "good," and 0.75–1.0 is categorized as "excellent" reliability.

III. Discussion

This study proposes the feasibility of developing a deep-learning-based computer vision model for predicting musculoskeletal key points and recognizing motions in infants using smart home cameras. This study aims to construct a learning dataset, design a learning and inference pipeline, and develop a motion estimation system to support the validity of these approaches.

Infants have distinct physical characteristics and spontaneous movement patterns that differ from those of adults. However, the existing benchmark datasets are based on adults, making them unsuitable for accurately capturing the

physical traits of infants. The RGB-depth camera, which is based on depth imaging technology, is highly reliable for capturing body movements, portable, and easy to use. The depth sensor of the camera provides critical 3D surface distance information, enabling skeleton estimation. This allows for precise tracking of joint positions and orientations, making it easier to access and monitor small movements. Using a depth camera in addition to RGB imaging is particularly beneficial in the context of infant monitoring. Most baby surveillance cameras are positioned above infant beds, and analyzing the height differences in the crib can be essential for applications such as detecting falls. To train the deep learning model, data are collected using devices that capture both RGB and depth information. After preprocessing, the dataset is divided and verified using both existing RGB-only models and the newly developed models. Further performance analyses will be conducted through model optimization and lightweight techniques to ensure the effectiveness of the system.

IV. Conclusion

This study aims to develop a robust infant pose estimation model that can handle various environmental changes, such as night mode, by constructing a comprehensive dataset beyond the limitations of existing models that rely solely on RGB data. The model development incorporates different algorithms, such as those for detecting inverted positions, with the potential to contribute to the prevention of infant-related accidents, including suffocation and falls. In addition, the information obtained from key infant points can be utilized to monitor developmental milestones, providing a valuable tool for tracking normal growth and ensuring that infants progress appropriately at various stages. These findings can contribute to the development of systems that support early detection and intervention in developmental monitoring.

Acknowledgment

This work was carried out with the support of "National Research Foundation of Korea." Assignment number: NRF-2022R1F1A1071626

Reference

- World Health Organization, <https://www.who.int/news/item/09-05-2023-152-million-babies-born-preterm-in-the-last-decade>
- J.G. Zwicker, "Motor impairment in very preterm infants: implications for clinical practice and research", *Developmental Medicine Child Neurology*, vol. 56, no. 6, pp. 514-515, 2014.)
- Adde L., Brown A., Van Den Broeck C., DeCoen K., et al. (2021) The In-Motion-App for remote general movement assessment: a multi-site observational study. *BMJ Open*, 11(3), e042147.
- Andriluka M., Pishchulin L., Gehler P., Schiele B. (2014) 2D human pose estimation: New benchmark and state of the art analysis. *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, 3686-3693.
-

- Bisny T. Joseph (2023). Stress In Babies: Symptoms, Causes, And Prevention. https://www.momjunction.com/articles/unexpected-causes-and-signs-of-stress-in-babies_00351741/
- Campbell SK., Palisano RJ., Orlin M. (2014). Physical therapy for children, 4th ed. St. Louis, MO, Elsevier Saunders.
- Einspieler C., Bos AF., Kriebler-Tomantschger M., Alvarado E., et al. (2019). Cerebral palsy: early markers of clinical phenotype and functional outcome. *J. Clin. Med*, 8, 1616.
- Groos D., Adde L., Stoen R., Ramampiaro H., et al. (2022) Towards human-level performance on automatic pose estimation of infant spontaneous movements, *Comput Med Imaging Graph*, 95,102012.
- Hesse N., Schroder A.S., Muller-Felbe W., Bodensteiner C., et al. (2017). Body Pose Estimation in Depth Images for Infant Motion Analysis. *Annu Int Conf IEEE Eng Med Biol Soc*, Jul,1909-1912.
- Huang X., Fu N., Liu S., & Ostadabbas, S. (2021). Invariant representation learning for infant pose estimation with small data. 16th IEEE international conference on automatic face and gesture recognition.
- Kahn-D'Angelo L., Blanchard Y., & McManus B. (2012). The special care nursery, 4th ed, Elsevier Saunders, 903-943
- Li M., Wei F., Li Y., Zhang S., et al. (2020). Three-dimensional pose estimation of infants lying supine using data from a kinetic sensor with low training cost. *IEEE Sensors Journal*, 21(5), 6904-6913
- Marchi V., Hakala A., Knight A., et al. (2019). Automated pose estimation captures key aspects of General Movements at eight to 17 weeks from conventional videos. *ACTA Paediatrica*, 108(10), 1733-1925
- Mousavi Hondori H and Khademi Met. (2014). A reviews technical and clinical impact of the Microsoft Kinect in physical therapy and rehabilitation. *J Med Eng*, 2014;2014:846514. doi: 10.1155/2014/846514. Epub 2014 Dec 10
- Nakajma Y., Einspieler C., Marschik PB., Bros A., & Prechtl H. (2006). Does a detailed assessment of poor repertoire general movements help to identify those infants who will develop normally? *Early Hum Dev*. 82(1):53-9.
- Novak I., Morgan C., Adde L., Blackman J., et al. (2017). Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment. *JAMA pediatrics*, 171(9), 897-907.
- Støen R., Songstad NT., Silberg IE., Fjørtoft T., et al. (2017). Computer-based video analysis identifies infants with absence of fidgety movements. *Pediatric Research*, 82(4), 665-70.
- Schmidtke L., Vlontzos A., Ellershaw S., Lukens A., et al. (2021). Unsupervised human pose estimation through transforming shape templates. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2484-2494
- Shin HI., Shin H-I., Bang MS., Kim D-Y., et al. (2022). Deep learning-based quantitative analyses of spontaneous movements and their association with early neurological development in preterm infants. *Scientific Reports*, 12, 3138
- Tolgyessy M., Dekan M., & Chovanec L. (2021) Skeleton Tracking Accuracy and Precision Evaluation of Kinect V1, Kinect V2, and the Azure Kinect. *Appl. Sci*, 11(12), 5756
-