2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871700

A Comparison of Video-based Methods for Neonatal Body Motion Detection

Zheng Peng, Dennis van de Sande, Ilde Lorato, Xi Long, Rong-Hao Liang, Peter Andriessen, Ward Cottaar, Sander Stuijk, and Carola van Pul

Abstract— **Preterm infants in a neonatal intensive care unit (NICU) are continuously monitored for their vital signs, such as heart rate and oxygen saturation. Body motion patterns are documented intermittently by clinical observations. Changing motion patterns in preterm infants are associated with maturation and clinical events such as late-onset sepsis and seizures. However, continuous motion monitoring in the NICU setting is not yet performed. Video-based motion monitoring is a promising method due to its non-contact nature and therefore unobtrusiveness. This study aims to determine the feasibility of simple video-based methods for infant body motion detection. We investigated and compared four methods to detect the motion in videos of infants, using two datasets acquired with different types of cameras. The thermal dataset contains 32 hours of annotated videos from 13 infants in open beds. The RGB dataset contains 9 hours of annotated videos from 5 infants in incubators. The compared methods include background substruction (BS), sparse optical flow (SOF), dense optical flow (DOF), and oriented FAST and rotated BRIEF (ORB). The detection performance and computation time were evaluated by the area under receiver operating curves (AUC) and run time. We conducted experiments to detect motion and gross motion respectively. In the thermal dataset, the best performance of both experiments is achieved by BS with mean (standard deviation) AUCs of 0.86 (0.03) and 0.93 (0.03). In the RGB dataset, SOF outperforms the other methods in both experiments with AUCs of 0.82 (0.10) and 0.91 (0.05). All methods are efficient to be integrated into a camera system when using low-resolution thermal cameras.**

I. INTRODUCTION

Preterm birth is a leading cause of morbidity and mortality in infants. The rates of survival and survival without severe impairment drop rapidly with the decrease of gestational age. These rates are also related to active lifesaving treatment and comfort care after birth [1]. These preterm infants are often hospitalized in a neonatal intensive care unit (NICU) or a medium care unit (MCU) depending on their maturity. To provide timely and adequate care and treatment, the vital signs of the patients such as the electrocardiogram (ECG) and photoplethysmography (PPG) are continuously monitored on patient monitors. However, not all clinically relevant information is yet monitored continuously. For instance, motion patterns in infants are only intermittently observed, even though changing motion patterns in preterm infants are associated with maturation and clinical events such as late-onset sepsis and seizures [2], [3]. Therefore, continuous motion detection is significant and valuable for infant monitoring in NICUs and MCUs.

Unobtrusive methods to detect motion in neonates have been widely investigated by researchers. For instance, the ballistography obtained from a pressure-based mattress was used to detect infant motion [4]. The frequency component of motion and the instability in monitored physiological signals were extracted to represent the motion of preterm infants [5]. The long-term motion derived from PPG showed that the amount of brief motion (less than 5s) decreases with the increasing postmenstrual age (PMA) of preterm infants [2]. Additionally, camera systems were used for infant detection and tracking, which can recognize an infant's body or caregiver's appearance [6], [7]. Even though video-based methods were widely used in motion detection [8], [9], [10], the comparison of efficiency and performance of these motion detection methods for preterm infants has not been conducted.

The aim of this study is to compare the performance of four video-based methods for motion detection in preterm infants, including background substruction (BS), sparse optical flow (SOF), dense optical flow (DOF), and oriented FAST and rotated BRIEF (ORB). This paper first starts with a brief description of the datasets and methods. Afterwards, all methods are optimized and applied on two datasets. The detection performance and computation time of all methods are compared and discussed at the end of the paper.

II. DATASETS

This study uses two datasets from preterm infants admitted to the NICU and MCU in the Maxima Medical Center (MMC) in Veldhoven, the Netherlands. One dataset was acquired using an RGB camera (UI-3860LE-C-HQ, with a resolution of 1280 * 720 pixels and a frame rate of 10 fps) placed on top of the incubator to capture the head and upper body of the infant. A total of 9-hour videos from 5 infants were collected and the mean (standard deviation, SD) PMA of infants were 31.2 (2.1) weeks in this RGB dataset. The thermal dataset was from our previous study on respiration

Zheng Peng and Carola van Pul are with the Department of Applied Physics at Eindhoven University of Technology and the Department of Clinical Physics, Maxima Medical Centre, Veldhoven, The Netherlands. z.peng@tue.nl

Dennis van de Sande and Ward Cottaar are with the Department of Applied Physics at Eindhoven University of Technology, The Netherlands.

Ilde Lorato and Sander Stuijk are with the Department of Electrical Engineering at Eindhoven University of Technology, The Netherlands.

Xi Long is with the Philips Research and the Department of Electrical Engineering at Eindhoven University of Technology, The Netherlands.

Rong-Hao Liang is with the Department of Industrial Design and Department of Electrical Engineering at Eindhoven University of Technology, The Netherlands.

Peter Andriessen is with the Department of Applied Physics at Eindhoven University of Technology and the Department of Neonatology, Maxima Medical Centre, Veldhoven, The Netherlands.

monitoring [11]. Three thermal cameras (FLIR Lepton 2.5, with a resolution of 60 x 80 pixels each and a frame rate of 9 fps) were positioned around the infants' beds. The video frames from three cameras were synchronized and merged to observe infants from different directions simultaneously in each frame. In total, 32-hour thermal videos from 13 infants, with mean PMA of 36.7 (3.3) weeks were used in this study.

Both datasets were retrospectively annotated based on the videos. The thermal videos were annotated by one of the authors in our previous study [11]. Another author annotated the RGB videos for this study, following the same annotation scheme described in [11]. This study focuses on three classes of event labels including gross motion, fine motion, and still. Gross motion indicates motion with the torso or chest. Fine motion involves motion from limbs, fingers, or facial expressions. The video frames corresponding to interrupting events like parent's or caregiver's hands in the view, feeding, infant out of bed, infant not in good view (e.g. very low light condition), and camera motion were excluded in this study. After exclusion, the duration of each included event and the corresponding percentage are shown in Table I.

For this study, the ethical committee of MMC provided a waiver. Informed consent was obtained from the infants' parents before the study.

III. METHODS

Four video-based methods (BS, SOF, DOF, and ORB) were implemented to measure motion between two consecutive frames for all included RGB and thermal video frames. In the preprocessing step, each frame was transformed into a grayscale image and normalized by a histogram equalization to improve efficiency and contrast. For each method, the motion was derived from video frames, taking the whole field of view of the cameras as the region of interest (ROI). Fig. 1 shows the processed frames for the different methods in both datasets.

A. Background Substruction

The BS method uses the difference between the current frame and a defined background model to detect the motion region where the model is violated [12]. The background model is first initialized with a fixed number of frames and then updated to adapt to the impact of the changing external environment such as light. There are various ways to initialize and update the background model. In this study, we applied a gaussian mixture model based on Zivkovic's method [12] to initialize and update the background model and detect the foreground region in each frame. The number of pixels that were determined as foreground by the method was used to quantify the motion in each frame.

TABLE I

EVENT DURATION BY HOURS (PERCENTAGE)

Dataset	Gross Motion	Fine Motion	Still	Total
RGB	1.9 (23%)	3.1 (37%)	3.3 (40%)	8.3 (100%)
Thermal	7.6 (32%)	11.2 (47%)	5.2 (22%)	24.0 (100%)

B. Optical Flow

The optical flow method uses the change of tracking points between consecutive frames to detect motion, assuming pixel intensities of an object do not change between consecutive frames [13]. This change is quantified by a motion vector with magnitude and direction corresponding to the detected motion. Depending on the density of pixels that are tracked by the optical flow method, the method can be categorized into SOF and DOF.

Regarding SOF, we first detected tracking points in each frame based on the Shi-Tomasi corner detection algorithm [14]. Then, we computed the optical flow of these tracking points based on Lucas-Kanade method with pyramids [15]. To prevent the mistracking of the optical flow points, we ran a 'backward check' on two consecutive frames and we only selected points in the first frame when their corresponding points calculated by backward check from the second frame were within a certain distance. The tracking points were refreshed at every tenth frame to improve robustness. Last, we calculated the average displacement (measured by Euclidean distance) of tracking points between current and previous frames to quantify the motion.

Regarding DOF, we computed the motion vectors for all the pixels between two consecutive frames based on Gunnar Farneback's algorithm [16]. The motion vectors contained the magnitude and direction of each pixel. The magnitude of all the pixels was summed to quantify the motion.

C. Oriented FAST and Rotated BRIEF

The ORB method is a fast binary feature descriptor, combining FAST (Features from Accelerated Segment Test) [17] feature point detection and BRIEF (Binary Robust Independent Elementary Features) [18] feature descriptor. It is widely used in feature matching and object detection [9], [19]. ORB first finds pixels that are significantly different from neighbor pixels in a frame as tracking points, using the FAST algorithm. Afterwards, these tracking points in two consecutive frames are described and matched based on a BRIEF descriptor with a rotation angle [19]. To reduce the number of mismatching points, we calculated two nearest neighbors based on hamming distance for each tracking point and rejected the points whose two nearest neighbors were too close. Similar to SOF, we quantified the motion in each frame by calculating the average Euclidean distance between the matched tracking points in current and previous frames.

D. Evaluation

We tested all methods using two experiments. In the first experiment, called motion detection, we merged the labels of gross motion and fine motion as 'motion labels' and used each method to discriminate motion from still. In the second experiment, called gross motion detection, each method was used to discriminate gross motion from the other (consisting of fine motion and still). First, the motion measure calculated by each method was normalized and smoothed by a notch filter. Next, the area under receiver operating curves (AUC) was used to evaluate the performance of the methods. The

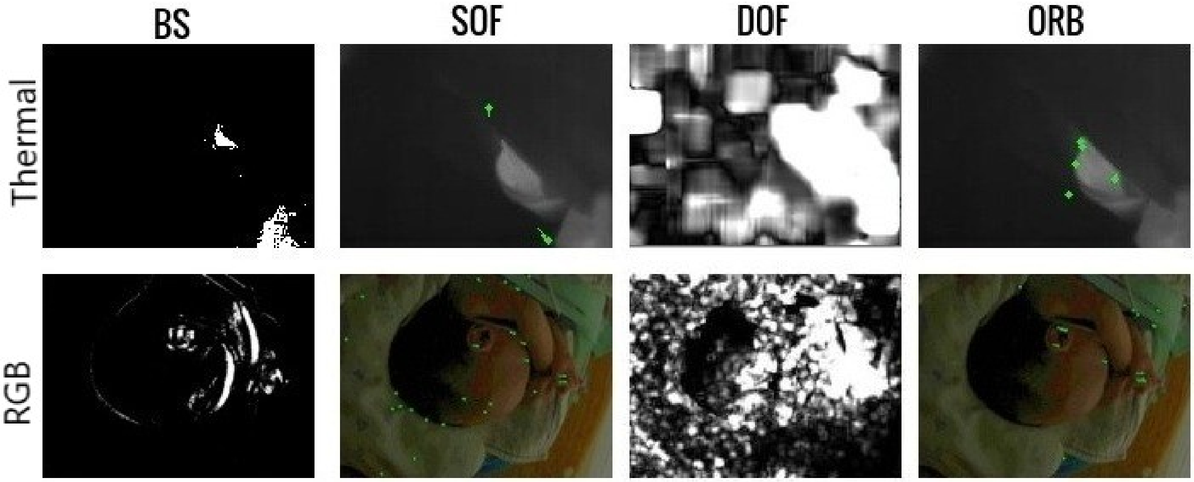

Fig. 1. Sample frames from thermal (top row) and RGB (bottom row) datasets, processed by the each method. (BS: background substruction. SOF: sparse optical flow. DOF: dense optical flow. ORB: oriented FAST and rotated BRIEF). The green dots represent the tracking points in methods (SOF and ORB).

computation time was measured by the run time on the same hardware. The mean (SD) of these metrics were used to characterize the results of all methods.

All analysis in this study was implemented using Python 3.8.3 with opencv-contrib-python 4.5.5.62 (CPU-only) on a CPU of 2.60GHz (Intel Core i7-9750H) with no architecture-specific instruction.

IV. Results

Table II shows the mean (SD) of AUC and run time of all infants corresponding to four methods and two datasets for both experiments. In the thermal dataset, BS outperforms the other methods with mean (SD) AUCs of 0.86 (0.03) and 0.93 (0.03) for both experiments. In the RGB dataset, SOF performs best for motion detection and is as good as DOF when detecting gross motion. The performance of two methods (SOF and ORB) based on sparse tracking points increases with higher video resolution, whereas the resolution has little influence on the performance of DOF. The run time does not differ per method for the thermal data, which has a low resolution. In the RGB dataset, the run time is always higher, particularly the run time of DOF increases largely at high resolution.

TABLE II
Motion detection and gross motion detection Performance in AUC and run time. Results are presented in mean (SD). The best results for RGB and Thermal are indicated in bold.

Method	Dataset	Motion Detection	Gross Motion Detection	Run Time (ms)
BS	RGB	0.76 (0.13)	0.89 (0.04)	30.0 (1.80)
	Thermal	**0.86 (0.03)**	**0.93 (0.03)**	16.0 (0.05)
SOF	RGB	**0.82 (0.09)**	**0.91 (0.05)**	**28.9 (2.63)**
	Thermal	0.74 (0.06)	0.87 (0.07)	15.9 (0.05)
DOF	RGB	0.79 (0.11)	**0.91 (0.05)**	234 (3.40)
	Thermal	0.78 (0.09)	0.91 (0.05)	15.8 (0.64)
ORB	RGB	0.77 (0.11)	0.84 (0.06)	36.2 (1.10)
	Thermal	0.70 (0.11)	0.81 (0.10)	**15.6 (0.02)**

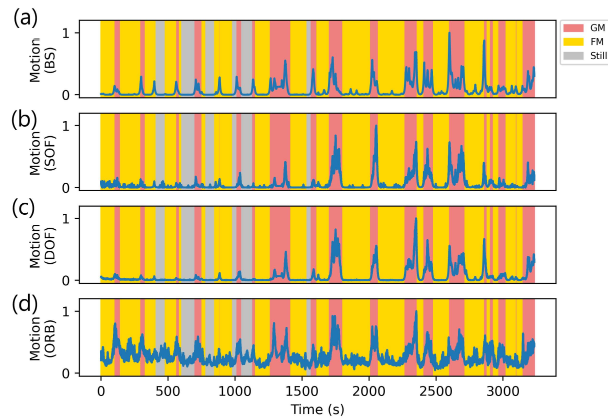

Fig. 2. Motion measure from each motion detection method over one hour in the thermal dataset, with background colored by corresponding annotations (GM: gross motion, FM: fine motion). (a) Motion measure from BS. (b) Motion measure from SOF. (c) Motion measure from DOF. (d) Motion measure from ORB.

Fig. 2 shows an example of one-hour annotations in the thermal dataset with corresponding motion measures calculated by four methods BS, SOF, DOF, and ORB. It can be observed that most gross motions are well reflected in all motion measures. BS and SOF are better at fine motion detection than DOF and ORB. ORB is most ‘noisy’ when the infant keeps still, but is more sensitive to consecutive gross motion (e.g. gross motion periods around 3000s).

V. Discussion

Our results show that all the methods can capture gross motion well based on differences between two consecutive frames. The overall lower performance in the motion detection experiment can be explained by the limited performance on fine motion detection for all methods (shown in Fig. 2).

The low performance in fine motion detection is not surprising, because the detection of quick sudden changes (e.g. eyelid twitch) is challenging for all methods. In particular, for BS, when a fine motion occurs, the changing pixels may not violate the background model over a certain threshold, leading to missing detection. Whereas tracking-point-based methods such as SOF and ORB can fail to detect fine motion since the tracking points may not locate at the changing pixels. DOF, on the other hand, assumes a slowly varying displacement field, causing the quick fine motion to be smoothed out [16].

The BS method performs best in both experiments when using thermal videos, but the performance decreases in RGB videos. This may be because the thermal videos are unaffected by light conditions, compensating the sensitivity to light conditions in BS. Moreover, BS and DOF are based on all pixels in a frame, meaning that they cannot benefit from the increased resolution of RGB videos as much as the methods based on sparse tracking points (SOF and ORB) which can be more reliable in high-resolution images. Additionally, the lower performance and the noisy motion measure of ORB may be caused by mismatched tracking points. Because the color of infants' skin is fairly uniform, it is challenging for ORB to correctly match points. ORB is more suitable in the applications of object finding with moving background [9], [19]. Interestingly, the BS method used on the thermal videos outperforms all the RGB-video results. However, further studies are needed to draw conclusions on the possible differences between the two modalities since differences in patient population and available views between the two datasets are also present.

The run time of the DOF method increases much more than other methods when running in high-resolution RGB videos. This is because of the high complexity of motion vector computation and the dense tracking points (all pixels).

One limitation of this study is that the event labels were annotated only based on visual observation of videos, which can lead to errors in the annotations when it is difficult to capture the onset and offset of fine motion events. Another limitation is that the patient population between the two datasets is different, as the thermal camera could only be used when infants are in open beds. The thermal camera cannot see through the incubator and at the time of the study, it was not allowed to use the three thermal cameras inside. The infants in the thermal dataset are, therefore, more mature with possibly different motion patterns than the preterm infants filmed with the RGB camera in the NICU.

This study analyzes the motion detection performance using video frames collected on infants, corresponding to (gross) motion and still. However, interrupting events (e.g. caregiver takes infant in/out of the bed) are quite common in the daily routine in NICUs and MCUs. To achieve continuous motion monitoring, future work will investigate infant's presence detection (to detect infants in/out of bed) and infants segmentation (to automatically select ROI) methods using videos [6], [7], [10]. In addition, the motion measures calculated by all methods will be further processed into binary or trinary signals explicitly indicating the motion status of infants to clinicians.

VI. CONCLUSIONS

This study compares (gross) motion detection performance and computation time for BS, SOF, DOF, and ORB using low-resolution thermal videos and high-resolution RGB videos of infants. Our findings suggest that using BS with low-resolution thermal videos to detect (gross) motion in preterm infants is more suitable, SOF is a good alternative when using high-resolution videos. This study is a first step towards the use of videos to continuously monitor motion in neonatal wards, which could lead to prediction and detection of clinical deteriorations and events.

REFERENCES

[1] M. A. Rysavy et al., "Between-Hospital Variation in Treatment and Outcomes in Extremely Preterm Infants," N. Engl. J. Med., vol. 372, no. 19, pp. 1801–1811, May 2015.

[2] I. Zuzarte, P. Indic, D. Sternad, and D. Paydarfar, "Quantifying Movement in Preterm Infants Using Photoplethysmography," Ann. Biomed. Eng., vol. 47, no. 2, pp. 646–658, Feb. 2019.

[3] J. Bekhof, J. B. Reitsma, J. H. Kok, and I. H. L. M. Van Straaten, "Clinical signs to identify late-onset sepsis in preterm infants," Eur. J. Pediatr., vol. 172, no. 4, pp. 501–508, Apr. 2013.

[4] R. Joshi et al., "A ballistographic approach for continuous and non-obtrusive monitoring of movement in neonates," IEEE J. Transl. Eng. Heal. Med., vol. 6, 2018.

[5] Z. Peng et al., "Body Motion Detection in Neonates Based on Motion Artifacts in Physiological Signals from a Clinical Patient Monitor," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Nov. 2021, pp. 416–419.

[6] C. H. Antink et al., "Fast body part segmentation and tracking of neonatal video data using deep learning," Med. Biol. Eng. Comput., vol. 58, no. 12, pp. 3049–3061, Dec. 2020.

[7] R. Weber, S. Cabon, A. Simon, F. Poree, and G. Carrault, "Preterm Newborn Presence Detection in Incubator and Open Bed Using Deep Transfer Learning," IEEE J. Biomed. Heal. Informatics, vol. 25, no. 5, pp. 1419–1428, May 2021.

[8] R. S. Rakibe and B. D. Patil, "Background Subtraction Algorithm Based Human Motion Detection," Int. J. Sci. Res. Publ., vol. 3, no. 5, 2013.

[9] S. Xie, W. Zhang, W. Ying, and K. Zakim, "Fast detecting moving objects in moving background using ORB feature matching," in Proceedings of the 2013 International Conference on Intelligent Control and Information Processing, ICICIP 2013, 2013, pp. 304–309.

[10] X. Long et al., "An efficient heuristic method for infant in/out of bed detection using video-derived motion estimates," Biomed. Phys. Eng. Express, vol. 4, no. 3, Apr. 2018.

[11] I. Lorato et al., "Towards continuous camera-based respiration monitoring in infants," Sensors, vol. 21, no. 7, Apr. 2021.

[12] Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," Pattern Recognit. Lett., vol. 27, no. 7, pp. 773–780, May 2006.

[13] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," Artifical Intell., vol. 17, no. 1–3, pp. 185–203, 1981.

[14] J. Shi and C. Tomasi, "Good features to track," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994, pp. 593–600.

[15] J.Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm," in technical report, Intel Microprocessor Research Labs, 1999.

[16] G. Farneb, "Two-Frame Motion Estimation Based on Polynomial Expansion," in Scandinavian Conference on Image Analysis (SCIA), 2003, vol. 2749, no. 1, pp. 363–370.

[17] E. Rosten and T. Drummond, "Machine Learning for High-Speed Corner Detection," in European conference on computer vision (ECCV), 2006, pp. 430–443.

[18] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in European conference on computer vision (ECCV), 2010, pp. 778–792.

[19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in Proceedings of the IEEE International Conference on Computer Vision, 2011, pp. 2564–2571.