Rollover Detection of Infants Using Posture Estimation Model

Ayaka Okuno
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
ayaka-okuno-kn@fuji.waseda.jp

Takaaki Ishikawa
Global Information and Telecommunication Institute
Waseda University
Tokyo, Japan
takaxp@ieee.org

Hiroshi Watanabe
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*— In predicting the posture of an infant during sleep, images taken by a surveillance camera are useful. The posture estimation of infants is designed to be used in a home environment and uses monocular camera images rather than special cameras such as distance cameras. In this paper, we compare the accuracy of infant posture estimation by two posture estimation models, OpenPose and Cascaded Pyramid Network(CPN). We also introduce a system for estimating infant's sleep turn using the posture estimation results. The experimental results show that CPN can detect infants' posture estimation with higher accuracy than OpenPose. In addition, the system is successfully used to detect the infant's turning with high accuracy.**

Keywords*— OpenPose, CPN, deep learning, Sleep Behavior in Infants

I. Introduction

Infants sleeping prone in bed is considered as a problem since it may cause sudden infant death syndrome and suffocation. The cause of sudden infant death syndrome is not known, but it is linked to sleeping on one's back. Asphyxiation can be caused by the infant's face being buried in pillows, mattresses, and quilts, or by wrapping the quilt around the infant's neck, or by falling out of bed. Even if an infant suffocates, he or she cannot turn over easily, and cannot return to the back, which may lead to death by asphyxiation. According to a survey by the Consumer Affairs Agency of Japan, suffocation accidents at bedtime are the most common cause of death due to unforeseen accidents[1].

Recently, behavioral recognition using image processing has been studied extensively. Recognition technology based on image processing can analyze the behavior of an object based on information acquired by cameras and sensors. Research is also underway on technologies that use a posture estimation model based on images from monocular cameras such as smartphones. Parents of infants and toddlers find it difficult to keep an eye on their babies and toddlers to make sure they don't sleep prone in bed at the same time as doing household chores. We propose a method for detecting tossing and turning from videos of sleeping infants without the discomfort of attaching sensors and markers to infants, and without using special cameras that are difficult to introduce into homes.

II. Rollover Detection Procedure

A. Postural estimation for infants

We use two types of posture estimation models: OpenPose[2] and Cascaded Pyramid Network(CPN)[3].

OpenPose : Realtime Multi-Person Pose Estimation is a method to obtain body coordinates from a person using deep learning proposed by Zhe Cao et al at Carnegie Mellon University. The joint positions of 15, 18, and 25 can be calculated as two-dimensional coordinates in the image. The keypoints estimated by OpenPose can be output as a JSON file with x-coordinate, y-coordinate and confidence. The coordinates of a key point that cannot be detected are displayed as 0 in both x and y coordinates value. OpenPose is a bottom-up pose estimation model in which all the key points in the image are identified and matched to each person. The bottom-up estimation model is a method that requires a lot of time to find the correct combination of keypoints because there are many combinations. In addition, false detection from the occlusion of people, clothes, and objects is also an issue. Figure 1 shows an example of an output by OpenPose.

Fig. 1. An example of the OpenPose output

CPN is a posture estimation model proposed by Yulun Chen et al. The model based on the Feature Pyramid Network is applied to detect the person, and then the posture is estimated by two stages called GlobalNet and RefineNet. GlobalNet is a pyramid network that correctly recognizes simple key points such as eyes and hands. However, it cannot accurately recognize invisible key points that are difficult to find. On the other hand, RefineNet allows for the estimation of hard-to-find keypoints by upsampling the features of each hierarchy from GlobalNet. The system can estimate the positions of 13 points, such as shoulders and elbows, and connect the joint points of each person with a line to output a human model. CPN is a top-down posture estimation model that first detects people and then estimates their postures for each person. In the top-down method, the accuracy is high because each person is estimated separately, but the computation time increases with the amount of persons in the image. The proponent of CPN, Y. Chen et al. uses Feature Pyramid Network as a person detector. However, in this paper, a single-shot multi-box detector is used as a person detector in

978-1-7281-9802-6/20/$31.00 ©2020 IEEE

order to improve the processing speed and performance. Figure 2 shows an example of an output by CPN.

The photos in Figures 1 and 2 are taken from pixabay site[4] and can be used for non-commercial purposes.

Fig. 2. An example of the CPN output

The accuracy of the two posture estimation models is compared when the infant sleeps in the upward, rightward, and leftward positions.

B. Estimating the direction in which the infact sleeps

To estimate the sleeping orientation of infants using a posture estimation model that is highly accurate with OpenPose and CPN, the coordinates of the nose and left and right hips, calculated from the posture estimation model, are used.

Let θ be the angle between the vector consisting of the midpoint of the left and right hips (u_o, v_o) and the noses (b_x, b_y) and the vector of directions perpendicular to the image. We use (1) to calculate $\cos\theta$ and estimate the vertical orientation for $|\cos\theta| > \frac{1}{\sqrt{2}}$

$$\cos\theta = \frac{v_o(b_y - v_o)}{v_o\sqrt{(u_o - b_x)^2 + (v_o - b_y)^2}} \quad (1)$$

When $|\cos\theta| \leq \frac{1}{\sqrt{2}}$ holds, we estimates that it is leftward when (2) fits, and rightward when (3) fits.

$$u_o - b_x > 0 \quad (2)$$

$$u_o - b_x < 0 \quad (3)$$

Figure 3 shows the definition of joint points and vectors. The joint coordinates of the red circles are detected and the left and right hip coordinates and the coordinates of the nose are used. The photos in Figures 3 is taken from pictogram site[5] and can be used for non-commercial purposes. Figure 4 shows an image of the orientation estimation.

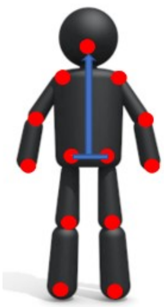

Fig. 3. Articulation points and vector diagram to be used in orientation estimation

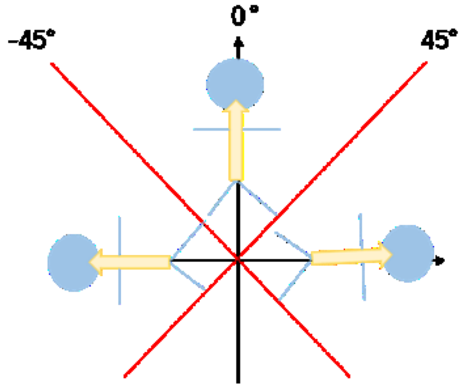

Fig. 4. Image of the orientation estimation

C. Rollover Detection Method

We use the estimated sleeping direction and the left and right shoulder and hip coordinates calculated by the posture estimation model to detect rollovers. An image of the tossing and turning detection system is shown in Figure 5.

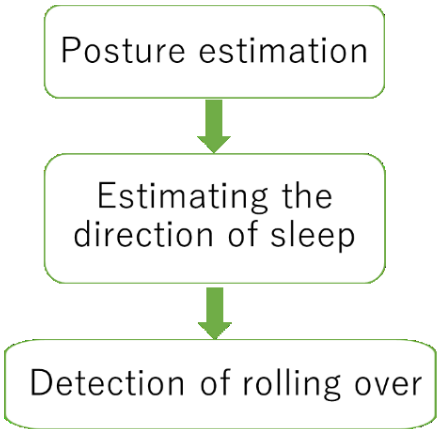

Fig. 5. Process of the tossing and turning detection system

Based on the estimated sleeping direction of the infant, the coordinates referring to the orientation are determined, and sleep-turning is detected from changes in the positions of the shoulders and hips. The detection is possible by referring to the x-coordinate in the case of vertical orientation and the y-coordinate in the case of horizontal orientation. We consider that the posture estimation model in the middle of turning over to be false, or that the person is trying to turn over but returning to a lying down posture. We detect rollover when prone state continues for 30 frames. In order to avoid detecting turning when the infant is about to turn over but returned to its back or when there were many false positives in the difficult to detect posture, we used the case where the infant is lying face down position for 30 frames after turning over. Since we use video at 60 fps, 30 frames is 0.5 seconds.

III. EXPERIMENTS

A. Postural estimation for infants

An experiment is conducted using 1101 images of infants sleeping on their backs. Sleeping infants are captured by a camera that are lying upward, left and right facing in the camera frame. Due to the assumption that there are no people around, we only use images that show only infants. To determine which model to be used in the experiment, we compared OpenPose and CPN, by recording the number of correct estimations made for the nose and shoulder and hip joint points. The results are shown in Tables 1 and 2.

TABLE I. POSTURE ESTIMATION ACCURACY OF OPENPOSE

	Upward	*Rightward*	*Leftward*	*All*
Number of photos	312	342	447	1101
Number of successful estimates	117	16	49	182
Percentage of success	37.5	4.68	11.0	16.5

TABLE II. POSTURE ESTIMATION ACCURACY OF CPN

	Upward	*Rightward*	*Leftward*	*All*
Number of photos	312	342	447	1101
Number of successful estimates	223	165	332	720
Percentage of success	71.5	48.2	74.3	65.4

From Tables 1 and 2, CPN is found to be superior to OpenPose in estimating posture in sleeping infants.

Some detection failures when using OpenPose were observed in our experiment as depicted in Figure 6 and 7. In the case of infants whose physical balance is different from that of adults, the lower body part of OpenPose is not always recognized correctly. The accuracy of OpenPose is significantly lower for leftward and rightward facing objects because it assumes that the target is facing vertical.

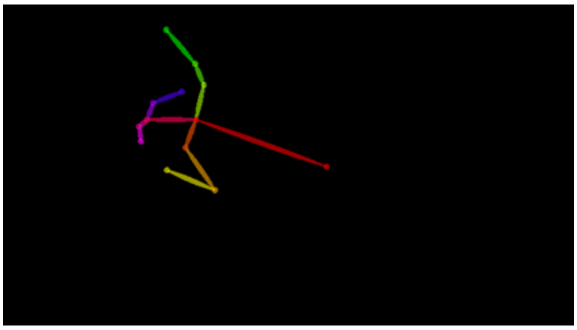

Fig. 6. An example of false detection (missing legs) by OpenPose[1]

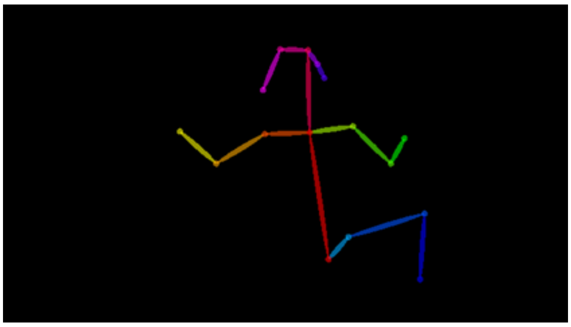

Fig. 7. An example of false detection (missing a leg) by OpenPose[2]

Some detection failures when using CPN were observed in our experiment as depicted in Figure 7 and 8. False detections are also made by CPN when the infant is in a shoulder and knee position, when there is an occlusion that assimilated with the doll, blanket, floor, or clothing, or when the joint area is difficult to see even from the human eye due to the position of the photograph. There is a blanket over the area from the waist down, a result that is clearly not a good estimation of the knee as shown in Figure 8. In contrast, Figure 9 shows that the left frame is followed by a movement to bring the shoulders and hips closer together, and the right frame is clearly a result of what appears to be a false detection.

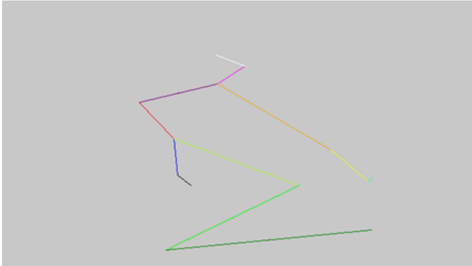

Fig. 8. An example of false detection (missing knee) by CPN[1]

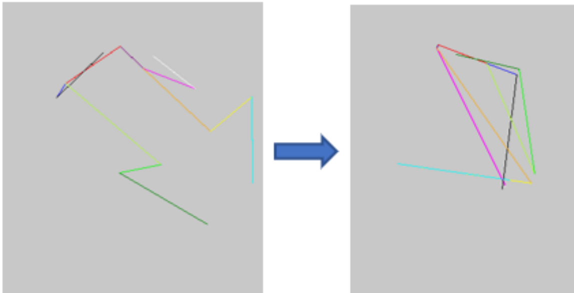

Fig. 9. An example of false detection (occlusion) by CPN[2]

The reason why the accuracy of the rightward-facing image is lower than that of the leftward-facing image is probably due to the presence of a doll close to the human skeleton in the vicinity. In addition, many of the images have similar patterns on the floor and clothes.

B. Estimating the direction in which the infant sleeps

We estimate the orientation from the first frame of 15 movies in which the infant is turning over because the orientation of the infant sleeping does not change in a short period of time. We show the experimental results in Table 3.

TABLE III. RESULT OF THE ORIENTATION ESTIMATION

Video number	*Correct Orientation*	*cosθ*	$u_o - b_x$	*Estimated orientation*
1	Leftward	0.29	+	Leftward
2	Rightward	0.28	−	Rightward
3	Leftward	0.08	+	Leftward
4	Leftward	0.03	+	Leftward
5	Upward	0.83		Upward
6	Upward	0.81		Upward
7	Upward	0.99		Upward
8	Rightward	0.37	−	Rightward
9	Leftward	0.11	+	Leftward
10	Rightward	0.94		Upward
11	Upward	0.98		Upward
12	Leftward	0.01	+	Leftward
13	Rightward	0.11	−	Rightward
14	Rightward	0.36	−	Rightward
15	Rightward	0.15	−	Rightward

978-1-7281-9802-6/20/$31.00 ©2020 IEEE

From Table 3, it can be seen that 14 out of 15 videos were successfully estimated. The video number 10, which could not be estimated correctly, was assumed to be vertical, even though it was horizontal, because the first image was mis-detected.

C. Rollover Detection Method

We detect tossing over in bed from 17 videos in which an infant is tossing about in bed. Four of the infants photographed were oriented vertically, seven were oriented to the right, and six were oriented to the left. The videos were made with the infant's body in the frame and the position in which the video was taken did not change significantly.

Figures 10 and 11 show the changes in coordinates when the detection is successful. Figure 10 shows that the size of the left and right coordinates of the shoulder coordinate is swapped at the 28th frame. In addition, Figure 11 shows that the size of the left and right coordinates of the waist coordinate is swapped at the 26th frame. Turnover was defined as when the left and right coordinates of the shoulders and hips were swapped and 30 frames passed. Therefore, the 58th frame is detected as the turnover point.

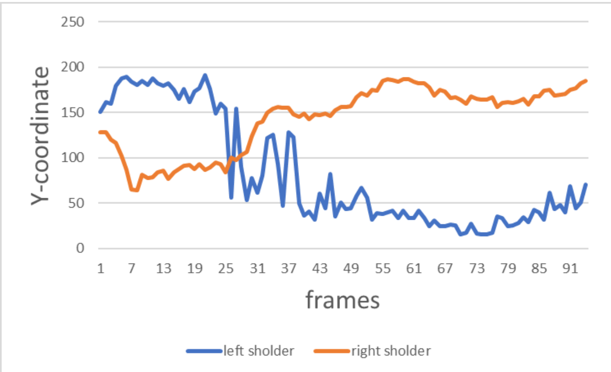

Fig. 10. Y-coordinate values for the left and right shoulders

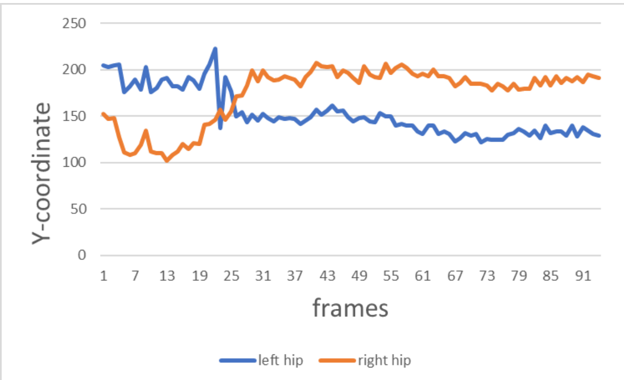

Fig. 11. Y-coordinate values for the left and right hips

Figure 12 shows the change in coordinates when the detection fails. There are two reasons for the failure of the detection. The first is that the infant's sleeping posture is difficult to estimate with their shoulders and knees close together. The second is that the posture of the infant after turning over in bed was changed from the initial orientation and the assessment axis was not appropriate.

We succeeded in detecting turnovers in bed in 14 out of 17 videos.

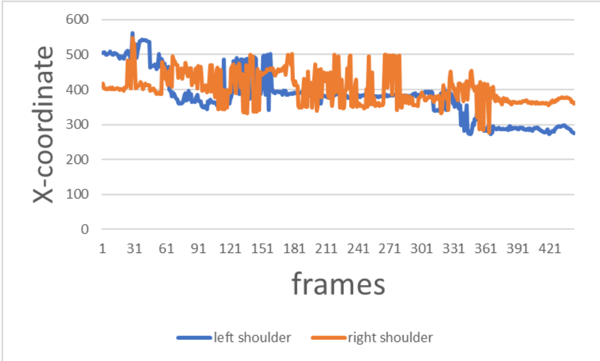

Fig. 12. X-coordinate values for the left and right shoulders

IV. Conclusion

In this paper, we compare the accuracy of two posture estimation models for infant detection and propose a method for detecting infants turning over in bed. The accuracy comparison showed that CPN is more accurate than OpenPose for infant detection. Based on the proposed method, we succeeded in detecting turnovers in 14 out of 17 videos.

Only the first frame oh the movie is used to estimate the sleeping direction. However, the results of the first frame are sometimes inaccurate. In addition, when a long video is used, the sleeping orientation may be changed from the beginning. Therefore, it is necessary to perform orientation estimation for each constant frame.

Moreover, it is revealed that frames in the midst of tossing about in bed are more likely to result in false position estimation model. The detection accuracy needs to be improved so that it can be detected correctly even in such cases.

References

[1] Consumer Affairs Agency, Government of Japan, "Beware of bedtime choking deaths of 0-year-olds! [Translated from Japanese]" pp.1-2, Oct. 2016.

[2] Z. Cao, T. Simon, S-E Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), No.121, pp.1302-1310, Jul. 2017.

[3] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded Pyramid Network for Multi-Person Pose Estimation," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp7103-7112, Jun. 2018.

[4] pixabay, Retrieved Feb 4, 2020, from https://pixabay.com/ja/images/search/baby/?pagi=4

[5] free-pictograms, Retrieved Feb 4, 2020, from https://free-pictograms.com/

978-1-7281-9802-6/20/$31.00 ©2020 IEEE