Published in final edited form as:

IEEE Int Conf Rehabil Robot. 2017 July ; 2017: 841–846. doi:10.1109/ICORR.2017.8009353.

Stereo 3D Tracking of Infants in Natural Play Conditions

Shreyas S. Shivakumar¹, Helen Loeb², Daniel K. Bogen¹, Frances Shofer¹, Phillip Bryant², Laura Prosser², and Michelle J. Johnson¹

^[1]University of Pennsylvania, Philadelphia, PA 19104, USA, sshreyas@seas.upenn.edu, dan@seas.upenn.edu frances.shofer@uphs.upenn.edu and michelle.johnson2@uphs.upenn.edu.

^[2]Children's Hospital of Philadelphia, PA 19104, USA, loebh@email.chop.edu, bryantp@email.chop.edu, prosserl@email.chop.edu.

Abstract

This paper describes the design and implementation of a multiple view stereoscopic 3D vision system and a supporting infant tracker pipeline to track limb movement in natural play environments and identify potential metrics to quantify movement behavior. So far, human pose estimation and tracking with 3D cameras has been focused primarily on adults and cannot be directly extended to infants because of differences in visual features such as shapes, sizes and appearance. With rehabilitation in mind, we propose a portable, compact, marker-less, low cost and high resolution 3D vision system and a tracking algorithm that exploits infant appearance attributes and depth information. This approach achieved a mean 3D tracking error of 8.21cm and a standard deviation of 8.75cm. We also identify two potential metrics for movement behavior analysis - approximate entropy and interaction events.

I. Introduction

Cerebral palsy is the most common motor disability in infants [1]. 2 to 3 out of every 1000 infants born in the US will suffer from cerebral palsy, affecting them in different ways, primarily disrupting their motor, perceptive and cognitive abilities [2]. Early detection can facilitate timely treatment and possible reduction of disability via rehabilitation to maximize the infant's motor development during the crucial years when they have great potential for neuroplastic change. Using current methods, diagnosis usually occurs around 18–24 months of age when obvious developmental milestones are missed. Studies have shown, however, that infant movement is already affected by the age of 12 months and physicians now recommend evaluations to be made at least twice during the first year after birth [3].

Common diagnostic methods include the non-intrusive, videotape based General Movements Assessment (GMA) [4][5] and the Test of Infant Motor Performance [6], both quite successful but require a trained professional, yield only a qualitative evaluation, and cannot be used after 4 months of age. Specialists also recommend brain imaging tests such as CT scans and MRIs, but brain images are not sufficient for diagnosis and must be paired with clinical data. There is therefore a clear need for a robust, automated, non-intrusive tool that can be deployed in any part of the world to produce a quantitative understanding of an infant's motor capabilities by observing infants under natural play conditions for a reasonable duration of time.

Observing infants as they play and interact with objects provides valuable insight into their cognitive and motor developmental abilities [7][8]. The GMA describes normal infant movement as writhing movement, characterized by small, circular and elegant movements, and fidgety movements in at-risk infants as cramped, synchronized repetitive movements [4][9]. Observing infant play behavior can identify impairments such as muscle weakness, sensory loss, spasticity and muscle shortening in at risk infants [10][11].

To quantify movement characteristics, there is a need for infant pose estimation and tracking to identify the configuration of the infants body pose over the course of the observation. In terms of visual sensor technology, this data can be gathered using marker based and marker free methods. Other non-visual approaches also exist but are pseudo intrusive in the sense that they require various sensors to be mounted on the surface of the infant's skin or as part of their clothing [12]. Our goal is to provide a low cost, multiple view, marker free, 3D vision and automated body part tracking system designed specifically for infants in natural play and can be easily deployed under low resource conditions.

The main contributions of our method are:

- a compact, high resolution stereo camera design that can be designed to accommodate available camera pairs;
- a stereo method to generate accurate depth maps corresponding to the high resolution color images; and
- a 3D infant motion tracking system to accurately extract limb motion signals;

The paper is organized as follows: 1) a summary of related work in infant motion tracking; 2) methods; 3) results; and 4) discussion and conclusion.

II. Related Work

Marker-based motion tracking is used often in rehabilitation because of its high accuracy in pose estimation. Multiple camera motion capture systems using retro-reflective markers have already been used to study preterm infant kinematics but markers can often distract the infant and skew the validity of the motion information [13][14][15][16]. Additionally, these setups usually have high cost overheads and cannot be conducted in natural environments such as a home or child care facility. We therefore believe that in the context of infant movement assessment the best approach is through non-intrusive measurement of body kinematics in natural play conditions.

Marker-free techniques can use 2D sensor setups (monochrome cameras, color cameras) or 3D sensor setups (monochrome and depth, RGB and depth cameras, stereo RGB, stereo monochrome and time of flight). 3D pose can be estimated from monocular images but making inferences about 3D kinematics from 2D images is a difficult problem and having additional depth information increases accuracy. The introduction of the Microsoft Kinect device was a breakthrough in the evolution of 3D human pose tracking algorithms and rehabilitation, whose (Kinect v1) features have been evaluated in [17][18][19] and a feasibility study for infant tracking was presented in [20]. However, limitations do exist. The

suggested working distance is between 0.5 – 3m and the device requires a wired connection. More compact setups are often desired when mounting the sensor device closer to the object of interest, such as on our *SmarToyGym*.

Stereo camera systems can be designed to overcome both limitations. By varying the stereo baseline, that is the distance between the two camera lenses, the minimum sensing distance can be reduced and accurate depth estimates can be achieved even at shallow depths. By using battery powered commercially available cameras such as the Go Pro Hero Session, the connectivity and power problem is overcome while providing higher resolution.

Algorithms trained on adult human poses are often insufficient to track infants due to the vast differences in appearance. The application of the Kinect device to monitoring infant arm and leg movements was seen in [21] but was not used during interactive play. Most relevant to our research is the tracker system by Olsen et al. [22]. A Microsoft Kinect was placed above an infant in supine position and the point cloud of the infant was obtained by planar subtraction of the floor. Localization of the infant was achieved by a color constraint on the infant's clothing. The baby pose was estimated using 3D shape primitives. These shapes are then compared against each other temporally using a point to shape distance function. The skeleton is then modeled as a hierarchy of parts and the rotations are modeled using axis angles. Our setup differs in that we use stereo cameras and an additional side view camera system to account for movements that cannot be detected by a downward camera. Additionally, Olsen et al. did not deal with infants interacting with objects during the trial and the only occlusion was inter limb self occlusion.

Another relevant approach used motion tracking to detect spontaneous movements in infants [23]. Here, feature vectors are generated from pose data and are passed to classifiers whose performance is compared and evaluated. This however required manually annotating the input data. Hesse et al. present another model based approach that leveraged synthetic data generated on a predefined infant model to estimate infant pose using a Random Fern [24]. Rahmati et al. presented a cost-effective single video camera setup for Motion Segmentation based classification [9]. Here motion features are extracted from video data and a support vector machine was used to classify the data into healthy or affected categories. None of the above methods consider infant interaction and experiments were carried out in highly constrained environments.

We present a flexible 3D vision system and what we believe to be a novel 3D infant pose estimation algorithm.

III. Methods

Our proposed approach utilizes multiple view geometry and standard image processing practices to identify and track infant limbs in 3D from 2D information (Figure 1). Currently, our method treats both stereo views independently and therefore for this study only the top view stereo setup is considered.

A. Experimental Methods

The data used in this experiment are pairs of color images extracted from high resolution (1920×1080) Go Pro Hero 4 Session videos. Four cameras were mounted on two 3D printed stereo frames and the videos were captured at 30 fps, with one setup placed right above the baby and the other positioned on the baby's right side, mounted on the frame of the *SmarToyGym* platform. The *SmarToyGym* is a platform where wireless toys are strategically hung within infant reach to elicit toy-oriented body and limb movements by means of built-in sensors and an integrated 3D vision system.

The infants that participated in this study were 3–11 months of age at the time of the study. The infants were placed in the *SmarToyGym* and 3 trials of up to 2 minutes each were collected, with the baby interacting with different toys during each trial and a single trial without infant toy interaction. The infants were provided with a blue colored onesie to wear. All infants participating in this study were accompanied by caregivers who gave written consent for their participation. For the purpose of this experiment, we analyze data collected from three healthy infants, infants A, B and C being 7.5, 7.4 months of age respectively.

B. Stereo Depth Estimation

In order to extract 3D information from the pair of color images, we use a stereo method comprising of pre-trial calibration, image rectification, cost matching and filtering.

Calibration: Before the introduction of the first infant to the setup, a calibration session was recorded, where a checkerboard pattern of known dimensions was slowly translated and rotated within the field of view of the stereo camera pair. Image pairs collected from this session were used to estimate the rigid transformation between the two cameras using MATLAB's stereo calibration toolbox. To ensure that image pairs are temporally aligned, we used a synchronization LED to align image frames to the time in which the light can first be seen from both cameras. This method gave us more accurate results in comparison to manually lining up audio queues or using the Go Pro multi-camera remote, which had up-to 10 frames of mis-alignment.

Image Pair Rectification: To estimate the disparity of a 3D point across two images, a robust point matching process is required, usually a 2D search problem. To simplify the problem to 1D, we used the calibration transformations to warp the two images such that they are co-planar, resulting in a pair of rectified images.

Correspondence Matching: Once the images were rectified, a cost matching algorithm was used to estimate the disparity of a given pixel between the two rectified frames. The image generated by this process is a disparity image d_{disp} . There are many cost matching algorithms, using local windowed methods which are fast and reasonably accurate or using global methods which are slow but extremely accurate. We used the Semi Global Block Matching algorithm, which maintains a good trade-off between speed and accuracy [25].

The disparity image d_{disp} is then converted to depth values using the following formula, where f is the focal length of the camera and B is the stereo baseline obtained during calibration and d_z is the depth image. An example depth map is seen in Figure 2

$$d_z = \frac{f \times B}{d_{disp}} \quad (1)$$

To validate our depth estimates, we measured the distance between the downward facing camera and the surface of the blanket on which the infant is placed. The measurement was 55.5cm and the average estimated depth was 57cm.

Filtering: The resulting depth image usually contains noise from imperfect calibration or from depth matching discontinuities such as repeated textures or irregular surfaces. These irregularities in depth can be easily filtered. We used a Bilateral Filter with a kernel size of 7 followed by a Median Filter with a kernel size of 5 pixels.

C. Tracking

The depth image d_z is then passed to our tracking algorithm along with the the left rectified color image x . The algorithm is made of two blocks, a color based center of torso tracker and the final limb tracker.

Center of torso tracking: To identify the approximate center of the infant's torso, we followed the approach of Olsen et al. where the infant wears a single colored onesie to simplify the detection process [26]. However, we built a more robust blue color detection algorithm by using a manually extracted data-set of blue color segments of the onesie from our data-set of infant images. The algorithm extracts the mean and co-variance values from this training set and fits a multivariate gaussian distribution on the test image. A threshold was used on the output probability distribution to generate a blue color mask. This accounted for greater variations in lighting compared to a simple blue color threshold.

We then fit ellipses over the segmented blue color masks and track it's center to generate our center of torso trajectory. The depth value of the torso center is extracted by identifying the corresponding values in the depth image and passed on to the next stage of our tracker.

Limb tracking: The limb tracking algorithm was designed using dense optical flow and RGBD based segmentation methods. We wish to estimate the positions of 6 markers $m_{t,k}$: (x_b, y_b, d_b) on the head center, left hand, right hand, left leg, right leg and the center of the torso respectively, where $k = 1 \dots 6$ and $t = 1 \dots n$ where n is the total number of image frames in the trial. The center of the torso trajectory $m_{6,t}$ was already obtained in the previous module.

Initialization: Our estimates for the remaining 5 markers begin with the user selecting initial limb regions. The user clicks on points within limb regions and initial marker locations $m_{0,1..5}$ are found using our marker identification method, by first identifying skin colored regions using a segmentation technique similar to the center of torso segmentation

except here our training patches are extracted from skin patches from our data-set. The marker is then obtained by identifying the point that lies within this mask and furthest away from the center of torso location $m_{0,6}$. When the infant moves it's limbs towards it's torso, the distance metric changes to the closest point to torso location $m_{0,6}$.

Position Estimation: Similar to the initial estimation, successive limb position estimations occur by identifying skin pixel groups and fitting ellipses to these regions. This works as long as there is no occlusion or limb overlap, identified by keeping track of individual skin pixel segments. When any two segments overlap or disappear they will share or lose a centroid, in which case a motion estimation and motion refinement step is used instead of a purely color based segmentation. Since limb dimensions cannot be estimated from overlapping and occluded segments, the dimensions of the estimated limb segments from $t - 1$ are used in t .

Motion Estimation: We use the Farneback dense optical flow estimation algorithm on the color image x_{t-1} and x_t to generate a displacement vector to move our set of markers from $m_{t,k}$ to the estimated marker set $\hat{m}_{t,k}$ [27] $\hat{m}_{t,k}$ is estimated by identifying the mean displacement vector in a windowed region around $m_{t-1,k}$. The window size is adaptive and depends on the size of the region of similar color and depth pixels in the marker neighborhood at $t - 1$.

We then quantize the depth image into 12 levels using K-Means. We do this to identify and restrict the estimated marker location to avoid large jumps in depth. The final depth value at the estimated marker location is obtained by identifying the skin segment to which the marker belongs and then calculating the mean depth of the maximum circular region that can fit in that skin and quantized depth segment.

Motion Refinement: To ensure that the marker doesn't jump from limb to torso or other regions of similar depth, if $\hat{m}_{t,k}$ does not correspond to a skin pixel in the skin mask, it is reassigned a new set of (x_b, y_b) co-ordinates chosen from the neighborhood of pixels such that the new pixel belongs to the skin mask. Similarly to ensure that that marker does not make drastic jumps in depth value, a similar segmentation based reassignment of marker co-ordinates is performed where the reassigned co-ordinate location is chosen from a nearby depth segment that satisfies the maximum depth variation between $\hat{d}_{t,k}$ and $d_{t-1,k}$. This resulting marker vector $\hat{m}_{t,k}$ is used to update $m_{t,k}$ once it satisfies all the above conditions. This process is repeated for the entire sequence of frames in a given trial.

IV. Results and Analysis

This section presents the results of our depth estimation and limb tracking pipeline on three different trials with three different infants. Prior to these trials, the infants were administered the Bayley Infant Neurodevelopmental Screener [28] and all our test subjects infant A, B and C were found to be healthy and do not suffer from motor learning disabilities. However, to highlight the sensitivity of our pipeline, we have chosen these infants and trials to cover three categories of limb movements - minimally active, active and very active. We first

evaluate our tracking algorithm and then propose potential metrics to quantify infant interaction behavior.

Tracking Accuracy:

For evaluation, we randomly sampled 60 images from our data-set and asked a user to click on the infant's limb extremities, head center and center of torso. The ℓ^2 norm of the 3D tracked marker estimates and the user generated 'ground truth' data was calculated; individual errors are shown in Fig 5. The mean tracking errors are **9.93 $\pm$ 8.16 cm, 8.74 $\pm$ 9.39 cm, 2.43 $\pm$ 1.66 cm, 8.75 $\pm$ 9.64 cm, 10.30 $\pm$ 12.35 cm and 9.13 $\pm$ 4.57 cm** for the right hand, left hand, head, right leg, left leg and center of torso respectively. These errors are related to the occlusion seen at the particular body part. The *Head* is usually not occluded and results in a tracking error of **2.43cm** whereas the *Left Leg* is often occluded by a toy or another limb and hence a larger error of **10.3cm** is observed. A top view of infant A's 2D displacement is seen in Fig 4.

Metrics for Infant Evaluation:

Acquiring 3D position data allowed us to test out two metrics that might elicit important information from the infant movement. Since traditional metrics such as velocity and displacement can be easily derived from our data, we hope to identify the validity of more novel metrics.

Approximate Entropy (ApEn): is used to distinguish signals based on their unpredictability and complexity and has been used as a metric for comparing pre term and full term infants based on the ApEn of their center of pressure [29]. Healthy infants will exhibit less predictability in their movements when compared to at risk infants. However, in our study we are comparing healthy infants and are therefore only trying to quantitatively characterize the complexity of their movements. Infant A exhibited less movement compared to infant B and C and fell into our category of minimally active. Infants B and C were exposed to toys and exhibited more movement with their limbs, with infant C being more active than infant B. We calculated the ApEn value of the magnitude of the 3D displacement of each limb and compared the sum of ApEn for all the limbs. Infant A had an ApEn value of **0.59** whereas infant B and C were **0.75** and **0.91** respectively. Our expectation is that at-risk infants will exhibit ApEn values much smaller than those seen with Infant B and C, and in the range or lower than Infant A, similar to the center of pressure ApEn variation in [29].

Toy Interactions: Infants B and C were exposed to two different toys, a hanging lion toy and a floor monkey toy. We believe that healthy babies will more easily and frequently interact with toys when compared to at-risk infants. The number of attempts made by the infant to interact with a toy could be a good feature to distinguish between infants and therefore we calculated the number of times an infant's limb came in close proximity to the toys. Fig 7 shows a plot of the distance of the infant's right leg from the toys. We then identified the number of occasions in which the infant's leg came within 22cm of the toy (in red). The right leg was chosen because the infants attempted to play with the toy with his/her right leg. Infant B's leg came within 22cm of the toy **4** times, whereas infant C came within

22cm of the toy only **once**. This threshold was selected arbitrarily for illustration purposes only. A note to be made is that in infant C's trial, the tracker fails at around the 2200 mark (fig 7) as the infant's right leg completely occludes the right hand.

V. Discussion and Conclusion

We present a novel approach to depth estimation and tracking of infants based on stereo 3D vision. The 3D information generated by this setup is a suitable alternative to the Kinect as it is more robust, portable and easily altered to accommodate various setups. However, since the Kinect SDK does not have limb or kinematic tracking algorithms for infants, a quantitative comparison was not possible. The resulting depth images from our setup are at the same resolution as the input RGB images and therefore have a higher depth image resolution. Our tracking pipeline can be used with any RGBD sensor data, even from a Kinect. While our approach does require additional setup time and a calibration process, we believe that the pros of the setup outweigh the cons and hence we believe that rehabilitative systems would benefit greatly from a portable, flexible, high-resolution 3D vision system.

The system is *robust* as it does not need to be tethered to a computer. Another feature is that any available camera pair can be used to generate depth estimates. This is a great advantage to low resource environment where acquiring a MoCap system is not be feasible. The system also performs at a constant frame rate unlike the Kinect devices, whose frame rate depends on the processing capabilities of the computer to which it is tethered. It can also be more *portable* as the GoPro cameras used did not need to be plugged into the wall for power. The setup is *flexible* because depending on the distance to the object of interest, the stereo baseline can be altered to ensure accurate depth estimation even under 0.5m, unlike the Kinect. This is one of the biggest benefits to the *SmarToyGym* as it allowed us to easily integrate this system into the gym structure. The higher *resolution* comes from the fact that each pixel in the color image is assigned a depth estimate value. Therefore the resolution of the depth map is limited only by the resolution of the color camera.

We also present a tracking algorithm to work alongside our vision setup to track infant body parts that is robust to varying degrees of partial occlusion. Some of the motivations moving forward are to fix the problem of nearly-complete occlusion, where a marker moves from one limb to another. Infant movement, in comparison to adult movement is more unorganized due to visually complex poses but we believe that this could be solved by enforcing a strong kinematic model to the tracker for better occlusion handling.

For the purpose of automating the diagnosis of motor control impairments however, we believe that our method can be applied to successfully extract raw limb motion information from infants. We present this work as a preliminary validation that accurate movement data can be collected from infants in natural play. While we do not validate the broader hypothesis in this work, we plan to use this setup in an extensive study with at-risk infants to validate our hypothesis that data obtained from this setup can be used to classify healthy and at-risk infants using various metrics extracted from this data such as the two metrics proposed here.

Acknowledgements

We gratefully acknowledge the support of the National Institutes of Health (NIH). We also thank Roshan Rai, Kevin Bui, the SmarToyGym team, the Rehabilitation Robotics Laboratory and the Children's Hospital of Philadelphia.

This work is supported by 1-R21-HD084327-01 under the National Institute of Health (NIH).

References

- [1]. "Cerebral Palsy data and statistics," <https://www.cdc.gov/ncbddd/cp/data.html>, accessed: 2016-12-10.
- [2]. Christensen D, Van Naarden Braun K, Doernberg NS, Maenner MJ, Arneson CL, Durkin MS, Benedict RE, Kirby RS, Wingate MS, Fitzgerald R, et al., "Prevalence of cerebral palsy, co-occurring autism spectrum disorders, and motor functioning- autism and developmental disabilities monitoring network, usa, 2008," *Developmental Medicine & Child Neurology*, vol. 56, no. 1, pp. 59-65, 2014. [PubMed: 24117446]
- [3]. Spittle AJ, Brown NC, Doyle LW, Boyd RN, Hunt RW, Bear M, and Inder TE, "Quality of general movements is related to white matter pathology in very preterm infants," *Pediatrics*, vol. 121, no. 5, pp. e1184-e1189, 2008. [PubMed: 18390959]
- [4]. Prechtl H, Fargel J, Weinmann H, and Bakker H, "Postures, motility and respiration of low-risk pre-term infants," *Developmental Medicine & Child Neurology*, vol. 21, no. 1, pp. 3-27, 1979. [PubMed: 437383]
- [5]. Darsaklis V, Snider LM, Majnemer A, and Mazer B, "Predictive validity of prechtl's method on the qualitative assessment of general movements: a systematic review of the evidence," *Developmental Medicine & Child Neurology*, vol. 53, no. 10, pp. 896-906, 2011. [PubMed: 21679361]
- [6]. Campbell SK and Hedeker D, "Validity of the test of infant motor performance for discriminating among infants with varying risk for poor motor outcome," *The Journal of pediatrics*, vol. 139, no. 4, pp. 546-551, 2001. [PubMed: 11598602]
- [7]. Bhat A and Galloway J, "Toy-oriented changes during early arm movements: hand kinematics," *Infant Behavior and Development*, vol. 29, no. 3, pp. 358-372, 2006. [PubMed: 17138291]
- [8]. Alexander GM, Wilcox T, and Woods R, "Sex differences in infants visual interest in toys," *Archives of sexual behavior*, vol. 38, no. 3, pp. 427-433, 2009. [PubMed: 19016318]
- [9]. Rahmati H, Aamo OM, Stavdahl Ø, Dragon R, and Adde L, "Video-based early cerebral palsy prediction using motion segmentation," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE*, 2014, pp. 3779- 3783.
- [10]. Boyd R and Graham HK, "Botulinum toxin a in the management of children with cerebral palsy: indications and outcome," in *European Journal of Neurology*, vol. 4, no. Suppl. 2, 1997, pp. S15-S21.
- [11]. Fehlings D, Rang M, Glazier J, and Steele C, "Botulinum toxin type a injections in the spastic upper extremity of children with hemiplegia: child characteristics that predict a positive outcome," *European Journal of Neurology*, vol. 8, no. s5, pp. 145-149, 2001. [PubMed: 11851743]
- [12]. Rogers E, Polygerinos P, Walsh C, and Goldfield E, "Smart and connected actuated mobile and sensing suit to encourage motion in developmentally delayed infants," *Journal of Medical Devices*, vol. 9, no. 3, p. 030914, 2015.
- [13]. Fallang B, Saugstad OD, Grøgaard J, and Hadders-Algra M, "Kinematic quality of reaching movements in preterm infants," *Pediatric research*, vol. 53, no. 5, pp. 836-842, 2003. [PubMed: 12612201]
- [14]. Toledo A and Tudella E, "The development of reaching behavior in low-risk preterm infants," *Infant Behavior and Development*, vol. 31, no. 3, pp. 398-407, 2008. [PubMed: 18289691]
- [15]. Gonçalves RV, Figueiredo EM, Mourão CB, Colosimo EA, Fonseca ST, and Mancini MC, "Development of infant reaching behaviors: Kinematic changes in touching and hitting," *Infant Behavior and Development*, vol. 36, no. 4, pp. 825-832, 2013. [PubMed: 24140840]

- [16]. Rönqvist L and Domellöf E, "Quantitative assessment of right and left reaching movements in infants: A longitudinal study from 6 to 36 months," *Developmental psychobiology*, vol. 48, no. 6, pp. 444–459, 2006. [PubMed: 16886181]
- [17]. Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, Cook M, and Moore R, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.
- [18]. Berger K, Meister S, Nair R, and Kondermann D, "A state of the art report on kinect sensor setups in computer vision," in *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*. Springer, 2013, pp. 257–272.
- [19]. Mousavi Hondori H and Khademi M, "A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation," *Journal of medical engineering*, vol. 2014, 2014.
- [20]. Olsen MD, "Motion tracking of infants in risk of cerebral palsy," 2016.
- [21]. Friedman H, Bar-Yosef O, Gordon G, Forkosh O, and Schneidman E, "Analysis of infant neuromotor development using a computer-based approach," *Neural Plasticity and Cognitive Modifiability*, p. 35, 2013.
- [22]. Olsen MD, Herskind A, Nielsen JB, and Paulsen RR, "Model-based motion tracking of infants," in *European Conference on Computer Vision*. Springer, 2014, pp. 673–685.
- [23]. Olsen MD, Herskind A, Nielsen JB, and Paulsen RR, "Using motion tracking to detect spontaneous movements in infants," in *Scandinavian Conference on Image Analysis*. Springer, 2015, pp. 410–417.
- [24]. Hesse N, Stachowiak G, Breuer T, and Arens M, "Estimating body pose of infants in depth images using random ferns," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 35–43.
- [25]. Hirschmuller H, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 328–341, 2008. [PubMed: 18084062]
- [26]. Olsen MD, Herskind A, Nielsen JB, and Paulsen RR, "Body part tracking of infants," in *Pattern Recognition (ICPR), 2014 22nd International Conference on IEEE*, 2014, pp. 2167–2172.
- [27]. Farnebäck G, "Two-frame motion estimation based on polynomial expansion," in *Scandinavian conference on Image analysis*. Springer, 2003, pp. 363–370.
- [28]. Aylward G, "Bayley infant neurodevelopmental screener," San Antonio, 1995.
- [29]. Dusing SC, Kyvelidou A, Mercer VS, and Stergiou N, "Infants born preterm exhibit different patterns of center-of-pressure movement than infants born at full term," *Physical therapy*, vol. 89, no. 12, p. 1354, 2009. [PubMed: 19815647]

Fig. 1:

Overview of the depth estimation and pose tracking pipeline. An example from both stereo views is illustrated, however since both views were not integrated we only considered the top view pair for this study. The entire pipeline can be thought of as three distinct blocks, the **stereo method** for depth estimation, the **center of torso (COT) trajectory estimation** for infant localization and the **limb trajectory tracker** for complete infant body part tracking.

Fig. 2:

Top: A pair of temporally aligned top camera rectified color frames **Bottom:** Corresponding depth map generated by our stereo depth estimation pipeline. Colorbar indicates depth in cm. Depth values were capped at 67cm. Depth of floor from camera is 65cm.

Fig. 3:

Left: Cropped region of baby from rectified frame. **Right:** After segmentation of skin pixels and basic morphological filtering.

Fig. 4:

2D displacement during the selected trial of Infant A. (0,0) denotes the centre of the blanket and all units are in cm.

Fig. 5:
3D tracking error (in cm) for the Right Hand, Left Hand, Head, Right Leg, Left Leg and Center of Torso.

Fig. 6:

Results from the tracking algorithm. Images 1–4 show correctly estimated poses and tracked extremities. Images 5–6 depict some of the poses for which the tracker fails. The tracker fails in Image 5 due to the high degree of inter limb occlusion in the legs. Image 6 is unsuccessful in accurately estimating the orientation of the detected right arm.

Fig. 7:

Top: Infant B made 4 attempts at playing with the hanging toy seen in Image 2 of Fig 6.

Bottom: Infant C made 1 attempt at playing with the floor toy. 22cm threshold was selected arbitrarily.