Model-Based Motion Tracking of Infants

Mikkel Damgaard Olsen[1(✉)], Anna Herskind[2], Jens Bo Nielsen[2,3], and Rasmus Reinhold Paulsen[1]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
mdol@dtu.dk
[2] Department of Neuroscience and Pharmacology, University of Copenhagen, Copenhagen, Denmark
[3] Department of Nutrition, Exercise and Sport, University of Copenhagen, Copenhagen, Denmark

Abstract. Even though motion tracking is a widely used technique to analyze and measure human movements, only a few studies focus on motion tracking of infants. In recent years, a number of studies have emerged focusing on analyzing the motion pattern of infants, using computer vision. Most of these studies are based on 2D images, but few are based on 3D information. In this paper, we present a model-based approach for tracking infants in 3D. The study extends a novel study on graph-based motion tracking of infants and we show that the extension improves the tracking results. A 3D model is constructed that resembles the body surface of an infant, where the model is based on simple geometric shapes and a hierarchical skeleton model.

Keywords: 3D model fitting · Infant pose estimation · Markerless motion tracking · Depth images

1 Introduction

Motion tracking of humans has attracted considerable interest in recent years, but only few studies exist in which the methods have been used for motion tracking of infants. However, the techniques can be of great benefit in the evaluation of infant development, as they can quantify the movements of infants and might be able to improve the diagnostics of different motor related diseases, namely cerebral palsy (CP). CP is the most common motor disability among children, affecting 2-2.5 out of 1000 infants [1]. It is caused by an injury to the fetal or infant brain, and the physical impairment is in many cases accompanied by disturbances of cognition and perception [2]. Among several others, preterm birth and birth asphyxia are associated with an increased risk of CP, but frequently a clear underlying pathology is not found [3,4]. During infancy symptoms are often subtle, but early warning signs include failure to meet motor milestones such as crawling and walking [5]. Due to the lack of unequivocal symptoms, in current practice, most children with CP are not diagnosed until the age

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part III, LNCS 8927, pp. 673–685, 2015.
DOI: 10.1007/978-3-319-16199-0_47

of 2 years [1]. However, studies have shown that the movement patterns of infants are influenced by CP already in the months before and after birth [6]. In fetuses, the movements can be observed and analyzed using ultrasound, while the post-birth studies are often based on analyzing videos of the infants' so-called general movements, which can be observed until the age of 5 months (corrected w.r.t term). By observing and identifying the motion patterns, cerebral palsy may thus be suspected at a much earlier age. Early identification of infants at risk of CP leads to the possibility of early intervention, which may improve the development of the infants' motor and cognitive skills. However, due to the time consuming procedure of analyzing the motion patterns, it is unrealistic to manually examine all infants born at risk of CP. The method presented in this work will thus try to move closer to an automatic system. The idea is that the system should be standard equipment, e.g. located at outpatient clinics, used for analyzing movement patterns of prematurely born infants. Because of this we cannot use a complex system that requires hours of preparation and calibration, but it should be a sort of plug-and-play system. As mentioned, studies within this field of interest are limited, but some work exists on motion tracking of infants. In [7,8] the authors use 6 sensors attached to the infants' wrists, ankles, chest and head. The sensors give temporal information about position and orientation. In [9], the authors use an optical motion system, where reflective markers are attached to the infant's limbs and multiple infra-red cameras are used to reconstruct the 3D location of the markers in space. From this, a set of parameters is extracted and used for early detection of spasticity due to cerebral palsy. In [10], the authors propose a new, optical flow-based method for quantifying the motion of infants with neonatal seizures, based on an overall quantitative measure of pixel-differences between successive frames. The optical flow-based approach is also used in [11], where the authors use color images as input to an optical flow-algorithm in order to track the position of the infants' hands and feet. Here, the authors manually initialize the position of the different body parts and adjust the positions during the tracking, in order to improve robustness of the method. In [12] the authors describe a method for tracking the 3D positions of anatomical extremities(hands, feet, head) and sub-extremities(elbow, knee) of infants based on a graph-based method equivalent to the approach in [13–15]. Based on the the work of [12], the method is extended with a model-based approach as in [16,17]. This both improves the body part localization and the tracking of the infants movements over time. The data is obtained using an affordable and easy-to-use depth sensor, Microsoft Kinect, which has revolutionized research within the field of low cost motion tracking.

2 Methods

The data used in this work, are depth and color images acquired with the Kinect sensor from Microsoft. The depth images have been recorded at a resolution of 480×640 pixels and the same resolution is used for the color images. As far as the authors' knowledge, no benchmark database of dense 3D data of infants exist

Fig. 1. Color image of one of the recorded infants. The infant wears an easily recognizable (blue) bodystocking and lies on a white blanket.

and a non-public database is used, which has been created simultaneous with this study. The participating infants are 3-8 months of age. For each infant 15-30 minutes of data have been recorded, while the infants have been laying on a flat surface e.g. a mattress or a blanket (see Figure 1). It should be noted that the pictures and data are used and published with respect to an agreement signed by the participating families. The Kinect has been positioned above the infant using a tripod for ordinary cameras. Using the depth images, a 3D point cloud representation of the infant and its surroundings can be generated. In order to make tracking easier, the fact that the infant lies on a flat surface is used to differentiate between foreground (most likely the infant) and background, by fitting a least squares plane to the surface. This is simply done by solving the linear system:

$$a(\mathbf{x} - x_c) + b(\mathbf{y} - y_c) + c(\mathbf{z} - z_c) = \mathbf{d}, \tag{1}$$

where; $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ are the observed 3D points of the underlying surface, (x_c, y_c, z_c) is a 3D point on the plane, and (a, b, c) are the elements for the normalvector of the plane. $\mathbf{d}$ defines the signed distances to the plane, but is set to zero during the fitting process. It is assumed that the viewing direction of the camera is nearly perpendicular to the flat surface and thus, the normal is corrected to point towards the camera. Given the estimated plane parameters, the signed distance from every 3D point to the plane can be calculated and this is used to discard points behind the plane ($\mathbf{d} < 0$). In order to remove small deviations of the underlying surface, a non-zero value is used as threshold. In addition, the infant wears an easily recognizable bodystocking, which is used to locate the baby, using color-based pixel classification.

2.1 Body Model

In this work, we use a 3D model to describe the surface of the human body. The model is constructed from a predefined number of geometric shapes that are connected based on the underlying skeleton. In order to measure the distance from the body model to the observed data, a "point to shape"-distance function is defined for each shape [16,17]. Currently the geometric shapes are:

– Cylinder: Used for describing elongated body parts, such as arms, legs and feet. The distance between a 3D point and the cylinder can be found analytically, by projecting the point onto the medial axis of the cylinder and taking the thickness/radius of the cylinder into account. The distance function for at 3D point $\mathbf{p}$ is thus defined as:

$$d = \begin{cases} \|\mathbf{p} - \mathbf{a}\| & \text{if } \lambda \leq 0 \\ \|\mathbf{p} - \mathbf{b}\| & \text{if } \lambda \geq 1 \\ \|\mathbf{p} - (\mathbf{a} + \lambda\,(\mathbf{b} - \mathbf{a}))\,\| & \text{otherwise} \end{cases}, \tag{2}$$

where, $\mathbf{a}$ and $\mathbf{b}$ are the start- and endpoints of the cylinder and

$$\lambda = \frac{(\mathbf{p} - \mathbf{a}) \cdot (\mathbf{b} - \mathbf{a})}{\|\mathbf{b} - \mathbf{a}\|^2} \tag{3}$$

is the normalized length of the vector $(\mathbf{p} - \mathbf{a})$ projected onto the line $(\mathbf{b} - \mathbf{a})$. It should be noted, that this distance is only for calculating the distance from a 3D point to the medial axis of a cylinder and the radius of the cylinder should be included in order to calculate the distance to the surface. Moreover, the described distance function, represents the distance to a rounded cylinder, when the radius/thickness is included.

– Sphere: Used to describe the head of the infant. The distance is easily computed, as the distance between the center point and the 3D point. Again, the radius of the body part should be included, in order to measure the distance to the surface, rather than the distance to the center.
– Ellipsoid/Superquadratic: The torso/stomach is modeled as a combination of two superquadratics. The upper part is modeled as a round-cornered box, in order to describe the box-like shape of the shoulders, while the lower part is modeled as a simple ellipsoid. In this case, no closed form solution exists for calculating the exact euclidean distance. Instead two differnet approaches has been used, which approximates the true distance:
  1. One solution is to use a numerical method, as explained in [18], which minimizes the distance from the 3D point to a point on the surface of the ellipsoid. The method is not generalized to superquadratics, but it is also possible to use iterative methods for approximating the distance for the upper bodypart [19].
  2. Another solution is to create a distancemap of the bodypart. Once the distancemap is created, the distance from the superquadratics to a 3D point can be approximated, by mapping the 3D point to the distancemap coordinates.

Inspired by previous work [20,21], we have chosen to model the skeleton as a hierarchical model, with the root starting from the center of the body. The identification of the body center is based on the center of mass of a set of classified pixels. As the infants wear a colored bodystocking during the recording, this can be recognized and tracked in the data. The fact that the infant lies on its back, is the reason the body center is chosen as the root joint. This is because the

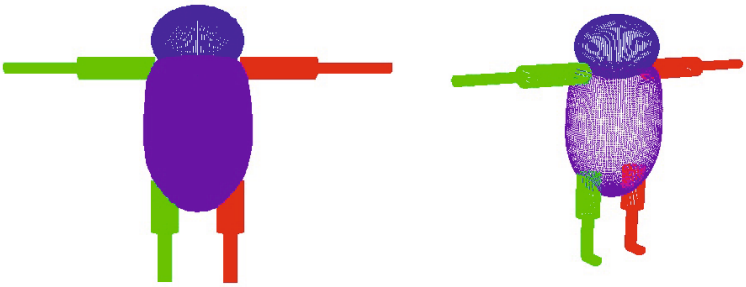

Fig. 2. Visual 2D (left) and 3D (right) representation of the model used in this study. The colors are simply to differentiate the bodyparts. Red and green are used to differentiate between left and right bodyparts, respectively.

location of the body center should be the most static part of the infant and more movement should be seen in the outer limbs. The articulated model can be seen in Figure 2, where colors are used to differentiate between left(red) and right(green) bodyparts.

In relation to the hierarchical connection between the bodyparts, the geometric shapes are oriented and positioned, using simple rotations and translations. However, instead of modeling the rotations in the global coordinate system, with euler angles, axis-angle representations are used, where rotations are limited to local coordinate systems, as the local coordinate system of each joint, changes based on the parent joints. The definition of axis-angle rotations is described widely in the literature [22,23] where the Rodriquez rotation formula can be used to construct a rotation-matrix given a axis-angle representation. Given an axis of rotation ω and a rotation angle θ, the rotation matrix can calculated as:

$$\mathbf{R} = \mathbf{I}_3 + \hat{\omega} \sin\theta + \hat{\omega}^2 (1 - \cos\theta) . \tag{4}$$

$\mathbf{I}_3$ is the 3×3 identity matrix and the $\hat{}$ operator constructs the skew symmetric matrix of the vector ω.

2.2 Fitting the Model

In order to fit the model to the observed data, the Levenberg Marquardt method is used, to iteratively refine the body parameters. The state vector and objective function will thus be defined in the following. An overview of the state parameters used in this study are listed in Table 1. Only the *Stomach* bodypart has a spatial parameter, which controls the global position of the model, while the position of the remaining bodyparts are constrained based on the hierachical model and the size parameters of the body.

The size parameters define the size and relative location of the bodyparts and are listed in Table 2. The size parameters are not part of the optimization but are either given prior to the optimization or estimated during an initialization

Table 1. Overview of the orientation parameters used for each bodypart. As can be seen, only one bodypart (*Stomach*) has a spatial parameter.

Bodypart	**Parameters**
Stomach	Rotation, Position
Head	Rotation
Left Upper Arm	Rotation
Right Upper Arm	Rotation
Left Lower Arm	Rotation
Right Lower Arm	Rotation
Left Upper Leg	Rotation
Right Upper Leg	Rotation
Left Lower Leg	Rotation
Right Lower Leg	Rotation
Left Foot	Rotation
Right Foot	Rotation

Table 2. Overview of the size parameters used for each bodypart. These parameters are not part of the optimization, but are used during the creation of the 3D model.

Bodypart	**Size**	**Location**
Stomach	Extension for the three axis	Global
Head	Radius	Relative to Stomach
Left Upper Arm	Radius + Length	Relative to Stomach
Right Upper Arm	Radius + Length	Relative to Stomach
Left Lower Arm	Radius + Length	Relative to Left Upper Arm
Right Lower Arm	Radius + Length	Relative to Right Upper Arm
Left Upper Leg	Radius + Length	Relative to Stomach
Right Upper Leg	Radius + Length	Relative to Stomach
Left Lower Leg	Radius + Length	Relative to Left Upper Leg
Right Lower Leg	Radius + Length	Relative to Right Upper Leg
Left Foot	Radius + Length	Relative to Left Lower Leg
Right Foot	Radius + Length	Relative to Right Lower Leg

step. It should be noted that symmetry is utilized and it is thus assumed that symmetric bodyparts are identical, with respect to size and relative location.

Once the state parameters are defined, the next step is to optimize on these parameters using the Levenberg-Marquardt optimization scheme, in order to fit the 3D model to the observed data. By concatenating the state parameters in a state vector $\mathbf{x}$, the optimization can be written as:

$$\min_{\mathbf{x}\in\mathbb{R}} \sum_{i=1}^{N} \|\mathbf{p}_i - \mathbf{c}\left(\mathbf{p}_i, \mathbf{x}\right)\|, \tag{5}$$

where $\mathbf{c}\left(\mathbf{p}_i, \mathbf{x}\right)$ calculates the closest 3D point on the model, given the 3D data point $\mathbf{p}_i$ and the state vector. An extension to the above minimization, is that

the state vector $\mathbf{x}$ is constrained, based on the anatomical properties of the human body joints. One requirement for the Levenberg-Marquardt algorithm, is an initial starting guess. In this study, a good estimate of the starting guess is found using an existing method for detecting and locating anatomical extremities, based on graph theory [12]. Here the anatomical extremities such as head, hands and feet are located by assuming that these points are furthest away from the bodycenter, when the distance measure is based on geodesic distances over the body surface. The distance is estimated by representing the surface as a graph, where neighboring 3D points are connected by nodes. This approach is able to locate and identify both the extremities and sub-extremities such as elbows and knees. The described method is able to give an estimate on the spatial location of the extremities. In order to obtain the respective state vector, an inverse kinematics method is used.

The total body-modeling method is summarized in the following, where input is the data obtained from the Kinect sensor.

1. Apply background subtraction in order to segment the infant from the background/underlying surface.
2. Define Body Parameters either using fixed parameters or during an initialization step.
3. Use graph-based method to obtain an estimate on the location of the anatomical extremities.
4. Apply an inverse kinematics algorithm in order to obtain the state parameters, given the location of the extremities(end effectors).
5. Refine the state parameters, in order to minimize the distance between the 3D model and the observed data.

The above described approach, estimates the orientation parameters of a single frame. However, by using the optimized parameters of one frame, as starting guess for the successive frame, the human body can be tracked in time.

3 Results

In order to test and evaluate the described method, Kinect recordings of 7 infants' movements have been obtained, where each infant has been recorded for 15-30 minutes. As no ground truth data is available, a various number of frames have been manually annotated for each infant. The frames have been selected such that they cover a wide variety of poses. In Figure 3, the results from the presented method can be observed. The method is able to correctly locate and identify the different body parts and the joint angles can be extracted directly from the respective state vector. It should be noted that an offset of the 3D model has been used, to better illustrate the estimated pose.

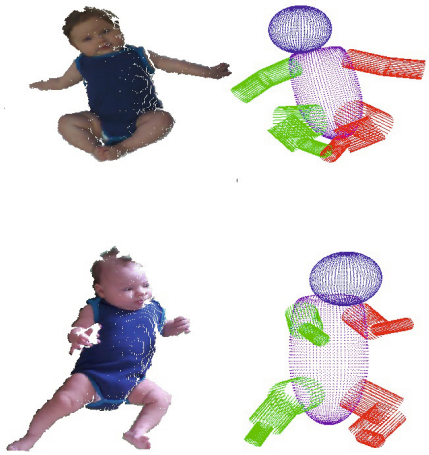

Fig. 3. Examples of the (translated) model fitted to the observed data. The point clouds are colored for visualization purposes, but only the 3D information is used during the optimization process.

In the following, the two methods (graph-based vs. model-based) are compared. The Euclidean distance between the manually positioned 3D points and the estimated joint locations are used to evaluate the tracking approach. In Figure 4 and Figure 5, the mean and standard deviation of the residuals are shown, respectively. The reader should notice that the *Stomach* and *Chest* residuals does not differ significantly, as these locations are found almost equivalently. However, the localization of the remaining joints has improved significantly, both with respect to mean and standard deviation.

The results indicate how robust the method is to locate the different body parts. As mentioned earlier, the technique can easily be extended to motion tracking, instead of human body detection. In Figure 6 the Euclidean distances between successive frames can be observed, for four different joints. Is is noteworthy to see that the graph-based tracking contains a lot of peaks/noise, while the model-based tracking gives a more smooth tracking. The reason for this, is that the model-based approach is less sensitive to deviations in data, compared to the graph-based method.

Based on the tracking results of one of the infants, Figure 7 shows the angles of the upper arms and the thighs, during a time period of 45 seconds. It is observed that the right upper arm is less active, compared to the remaining body parts.

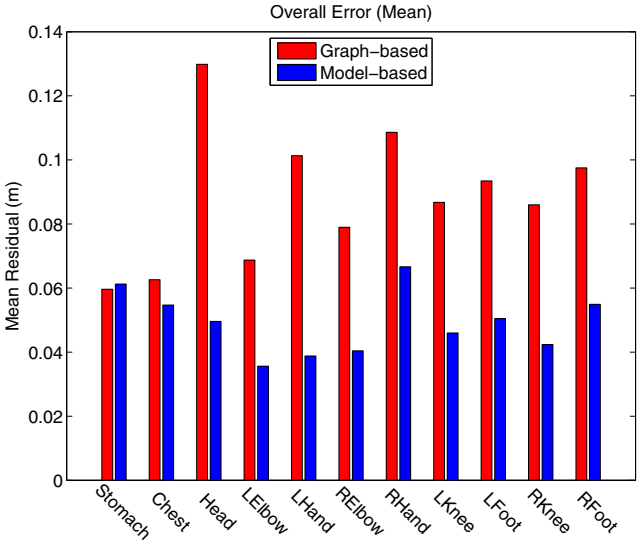

Fig. 4. The mean of the residuals listed for each body part in the model. Both the results from the graph-based approach (*red*) and the extended method (*blue*) are visualized.

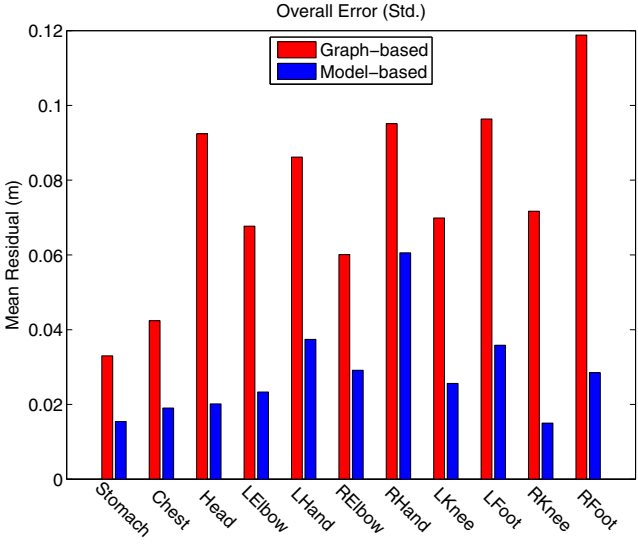

Fig. 5. The standard deviation(*right*) of the residuals listed for each body part in the model. Both the results from the graph-based approach (*red*) and the extended method (*blue*) are visualized.

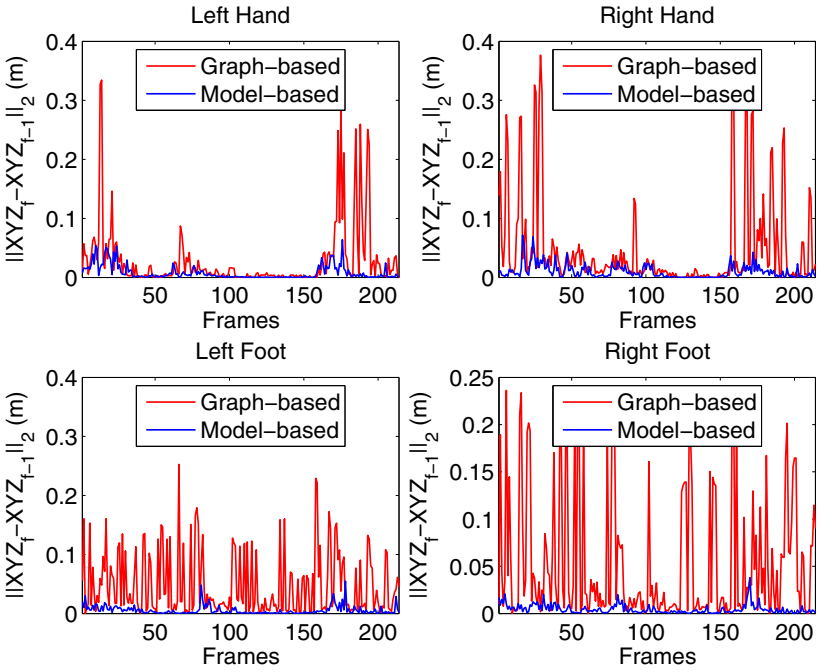

Fig. 6. Comparison of tracking results for the two methods. For each frame the Euclidean distance to the previous joint-location is calculated and visualized.

In Figure 8 the positions of the elbows and knees are visualized, which gives a better visualization of how the right elbow is much more passive, compared to the other bodyparts. This overall motion-plot might help doctors and physiotherapists to detect abnormal motion behaviors and plan the training according to these results.

Even though the study shows promising results, the tracking sometimes gets stuck in local minima, due to certain postures of the infant. One problem is e.g. when the infant rolls from supine to prone position. Here the method is unable to recover the correct orientation of some body parts and this error propagates through the tracking. This might be solved by enforcing additional temporal filtering on the body parameters. Another solution would be to create a pose library as in [16], where a number of candidate poses are evaluated and tested against the estimated pose.

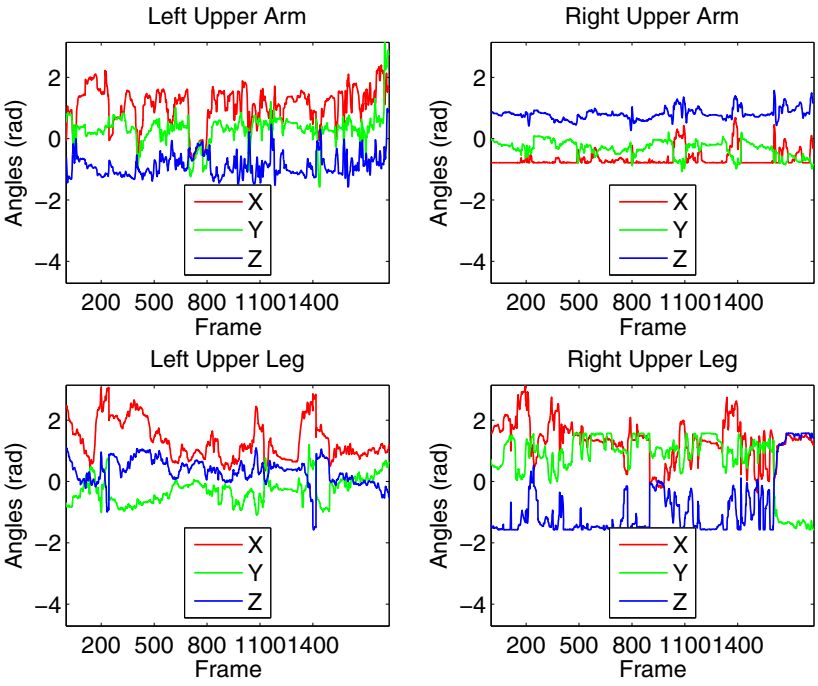

Fig. 7. The angles with respect to the local x-, y- and z-axis, for four bodyparts. It is noteworthy that the right upper arm is less active, compared to the three other bodyparts.

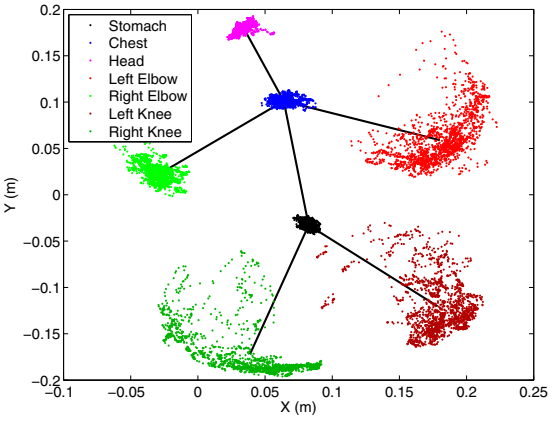

Fig. 8. Visualization of the location of 7 bodyjoints, namely; stomhach, chest, head, left/right elbows and left/right knee. The lower variation of the right elbow, shows that less movement has occurred for this joint.

4 Conclusion

We have described a method for 3D motion tracking of infants, based on a model-based approach. We show how this method gives good results, with respect to both accuracy and precision, compared to a novel study on motion tracking of infants. The method is based on fitting a 3D model to the observed 3D data, obtained with the Microsoft Kinect sensor. The model is combined by a number of geometric shapes that are connected based on a hierarchical skeleton model. We show how the results can be used to assess the motion pattern of infants by evaluating the raw motion parameters or the spatial 3D motion paths. This is a step closer to an automatic system that can help doctors assess the development of infants' motor control in order to detect motor impairing diseases such as cerebral palsy. In future work, we will focus on making the tracking more robust, such that the method is able to recover from difficult poses e.g. when the infants rolls over from supine to prone position.

Acknowledgments. The authors would like to thank the Helene Elsass Center and the Ludvig and Sara Elsass Foundation for funding the project as well as all of the infants and their families for participating in this project. Furthermore, the authors would like to thank the organization APA, for helping make contact with the participating families.

References

1. Himmelmann, K.: Epidemiology of cerebral palsy. Handbook of Clinical Neurology **111**, 163–167 (2013)
2. Bax, M., Goldstein, M., Rosenbaum, P., Leviton, A., Paneth, N., Dan, B., Jacobsson, B., Damiano, D.: Proposed definition and classification of cerebral palsy. Developmental Medicine & Child Neurology **47**(8), 571–576 (2005)
3. Goldsmith, S., Badawi, N., Blair, E., Taitz, D., Keogh, J., McIntyre, S.: A systematic review of risk factors for cerebral palsy in children born at term in developed countries. Developmental Medicine and Child Neurology **55**(6), 499–508 (2013)
4. McIntyre, S., Morgan, C., Walker, K., Novak, I.: Cerebral palsy-don't delay. Dev Disabil Res Rev **17**(2), 114–29 (2011)
5. Murphy, N., Such-Neibar, T.: Current problems in pediatric and adolescent health care. In: Cerebral Palsy Diagnosis and Management: The State of the Art, pp. 149–69 (2003)
6. Einspieler, C., Prechtl, H., Bos, A., Ferrari, F., Cioni, G.: Prechtl's Method on the Qualitative Assessment of General Movements in Preterm. Wiley, Term and Young Infants. Clinics in Developmental Medicine (2008)
7. Berg, A.: Modellbasert klassifisering av spedbarns bevegelser (2008)
8. Rahmanpour, P.: Features for movement based prediction of cerebral palsy (2009)
9. Meinecke, L., Breitbach-Faller, N., Bartz, C., Damen, R., Rau, G., Disselhorst-Klug, C.: Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy. Human Movement Science **25**(2), 125–144 (2006)

10. Karayiannis, N.B., Varughese, B., Tao Jr., G.: J.D.F., Wise, M.S., Mizrahi, E.M.: Quantifying motion in video recordings of neonatal seizures by regularized optical flow methods. IEEE Transactions on Image Processing **14**(7), 890–903 (2005)
11. Stahl, A., Schellewald, C., Stavdahl, O., Aamo, O.M., Adde, L., Kirkerod, H.: An optical flow-based method to predict infantile cerebral palsy. IEEE Transactions on Neural Systems and Rehabilitation Engineering **20**(4), 605–614 (2012)
12. Olsen, M.D., Herskind, A., Nielsen, J.B., Paulsen, R.R.: Motion tracking of infants. In: 22nd International Conference on Pattern Recognition (ICPR). (2014, to appear)
13. Plagemann, C., Ganapathi, V., Koller, D., Thrun, S.: Real-time identification and localization of body parts from depth images. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 3108–3113 (2010)
14. Baak, A., Müller, M., Bharaj, G., Seidel, H.P., Theobalt, C.: A data-driven approach for real-time full body pose reconstruction from a depth camera. In: IEEE 13th International Conference on Computer Vision (ICCV), pp. 1092–1099. IEEE (November 2011)
15. Schwarz, L.A., Mkhitaryan, A., Mateus, D., Navab, N.: Estimating human 3d pose from time-of-flight images based on geodesic distances and optical flow. In: FG, pp. 700–706. IEEE (2011)
16. Ganapathi, V., Plagemann, C., Koller, D., Thrun, S.: Real-time human pose tracking from range data. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part VI. LNCS, vol. 7577, pp. 738–751. Springer, Heidelberg (2012)
17. Droeschel, D., Behnke, S.: 3d body pose estimation using an adaptive person model for articulated icp. In: Jeschke, S., Liu, H., Schilberg, D. (eds.) ICIRA 2011, Part II. LNCS, vol. 7102, pp. 157–167. Springer, Heidelberg (2011)
18. Eberly, D.: Distance from a point to an ellipse, an ellipsoid, or a hyperellipsoid (2013)
19. Portal, R., Dias, J., de Sousa, L.: Contact detection between convex superquadric surfaces. Archive of Mechanical Engineering LVII **2**, 165–186 (2010)
20. Mnier, C., Boyer, E., Raffin, B.: 3d skeleton-based body pose recovery. In: 3DPVT, pp. 389–396. IEEE Computer Society (2006)
21. Shen, S., Tong, M., Deng, H., Liu, Y., Wu, X., Wakabayashi, K., Koike, H.: Model based human motion tracking using probability evolutionary algorithm. Pattern Recognition Letters **29**(13), 1877–1886 (2008)
22. Moeslund, T.B., Hilton, A., Krger, V., Sigal, L., eds.: Visual Analysis of Humans - Looking at People. Springer (2011)
23. Pons-Moll, G., Rosenhahn, B.: Ball joints for marker-less human motion capture. In: IEEE Workshop on Applications of Computer Vision (WACV) (2009)