Using Motion Tracking to Detect Spontaneous Movements in Infants

Mikkel D. Olsen[(✉)], Anna Herskind, Jens Bo Nielsen,
and Rasmus R. Paulsen

Department of Applied Mathematics and Computer Science,
Technical University of Denmark, Richard Petersens Plads,
Building 324, 2800 Kgs. Lyngby, Denmark
`mdol@dtu.dk`

Abstract. We study the characteristics of infants' spontaneous movements, based on data obtained from a markerless motion tracking system. From the pose data, the set of features are generated from the raw joint-angles of the infants and different classifiers are trained and evaluated using annotated data. Furthermore, we look at the importance of different features and outline the most significant features for detecting spontaneous movements of infants. Using these findings for further analysis of infants' movements, this might be used to identify infants in risk of cerebral palsy.

Keywords: Motion analysis · Motion tracking · Movement classification · Motion features

1 Introduction

In the last decades, motion tracking has become more and more popular. Whether it is marker-based or markerless, vision-based or sensor-based, the common goal is to estimate the pose and movement of people. Since the introduction of the Microsoft Kinect depth sensor in 2010, motion tracking has become a relative easy problem to solve. Without much effort, the underlying pose and motion parameters can be obtained and the next step is thus to utilize these parameters. In relation to the initial purpose of the Kinect sensor, the extracted pose parameters was used as input to the Microsoft XBox console, to control the character within a computer game. However, the list of applications is far more comprehensive. In [1,2] the pose estimation is used to extract features such as speed and step length. Features like these can be used for recognizing people, based on their gait, as shown in [3]. Other studies do not focus on recognizing a specific person, but instead recognizing different actions, such as walking, running, boxing, jumping, etc. [4]. However, common for most studies is that they focus on recognizing movements, that are easy to differentiate from each other, such as walking/jumping/punching/etc. Recently, new studies and challenges consider the concept of looking at more similar actions, such as recognizing sign language gestures, where two gestures can seem very similar to

the untrained observer [5]. In this study, we focus on movements of infants. It is known that infants in the age of 3-5 months have special movements called fidgety movements [6]. Among high-risk infants such as infants born preterm, absent or abnormal fidgity movements is a strong indicator for the motor disorder cerebral palsy. Doctors are thus able to identify a high risk of cerebral palsy, in the early months after birth, based on assessing these special movements. Small movements in the trunk, neck and limbs characterize these special movements. The movements are easiest to detect when the infant is lying on its back, unstimulated [7]. However, kicking and crying influences the infants' movements and the fidgety movements will be obscured by these larger movements. Moreover, a pacifier can completely dampen the strength of the fidgety movements. In order to be able to recognize the fidgety movements, one approach is to first detect and remove the sequences where the fidgety movements do not appear and secondly to classify the remaining movements. In this paper, we focus on the first step, where we classify sequences of motion data of awake infants, with the goal of segmenting the sequence into segments of spontaneous/non-spontaneous movements. The classification is based on features obtained from a vision based and markerless motion tracking approach. A number of previous studies focus on quantifying these spontaneous movements. In [8] the authors quantify spontaneous kicks by tethering the legs to a mobile stand. When the infant kicks, the mobile moves and this information is used for further analysis. In [9], a similar mobile system is combined with a 2D based motion tracking system. Using both the mobile-observations as well as the motion tracking results, the kicking frequency can be obtained. In this study, the goal is to;

1. Test different classifiers for segmenting spontaneous movements, based on data extracted from a markerless motion tracking system.
2. Examine the importance of different movement based features in order to classify spontaneous movements.

2 Methods

2.1 Motion Data

The data used in this study are temporal RGB-D data obtained with the Microsoft Kinect sensor. The recorded data contains both color and depth information of infants in the age of 1-6 months (corrected with respect to term). During the recording-session, the infant was positioned on a mat, while the RGB-D camera was positioned above the infant. No equipment was attached to the infant and the infant was thus able to move freely and unaffected. However, it was required that the infant only wore a short-sleeved bodystocking, in order to see the joints of the over- and under-extremities. Furthermore, the infant was in a good mood and unstimulated during the recording. The infant's parents were informed about the procedure beforehand and at any time; the parents could choose to stop the session. Unless the session was interrupted, the recording was done for minimum 5 minutes.

2.2 Motion Tracking

The pose estimation and motion tracking of the infants are obtained using a previously developed system [10,11]. To summarize, the system fits an articulated 3D model to the 3D data obtained from the depth sensor. The model is structured from a set of relative simple 3D structures, namely cylinders, spheres and superellipsoids. A set of parameters define the shape and orientation of these structures, which are length, radius and angle/direction with respect to their relative parent structure. The stomach/torso defines the root structure and all other structures are connected either directly or indirectly to this structure. The fitting process is done by adjusting the orientation parameters, while minimizing the error-metric between the 3D data and the 3D model. The error-metric is simply based on the Euclidean distance between the model and the data. Figure 1 illustrates an example for the resulting pose estimation.
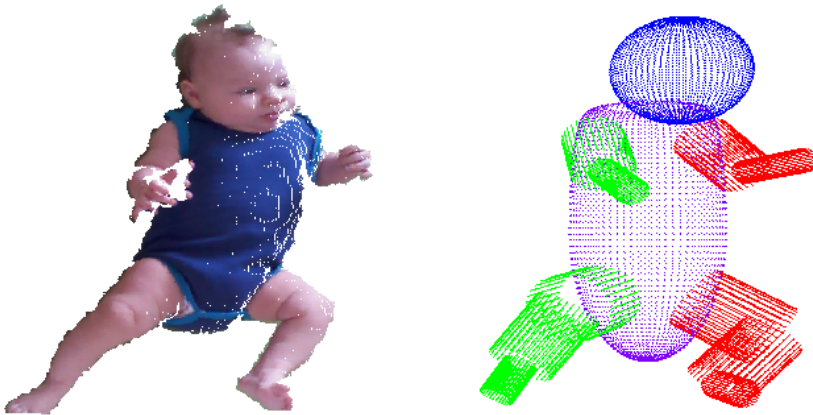

Fig. 1. Left: Colored point cloud obtained from the depth sensor. Right: 3D model fitted to the observed data.

2.3 Classification

The result from the motion tracking is a set of joint parameters, describing the pose of the recorded infant for each timestamp. The features used in this study are based on angular velocities and accelerations of the joints. For each frame, we calculate the angular velocities and accelerations, based on the joint angles in the current frame and the two previous frames. However, instead of using the raw data from a single frame, we transform the data using a sliding window approach, where we both generate mean- and median-filtered joint features. The transformations used for this approach are average, median, variance and the

Frobenius norm. Furthermore, we take the max/min values of the filtered velocities/accelerations, as we are interested in detecting frames where the infant is lying still vs. frames where the infant is doing an extreme movement with any part of the body. For classification methods we use *K-Nearest Neighbors (KNN)*, *Support Vector Machine (SVM)* and *Decision Tree*. For *KNN* we use two parameters for k, namely $k = 1$ and $k = 5$. In order to evaluate the different classifiers' performance, parts of the data have been annotated manually. The movements have been annotated either as spontaneous or calm.

3   Results

The dataset used in this study consists of 50k labelled frames taken from data recordings of 11 infants. For each frame, the frame was either labelled as being spontaneous or not. This labelling was done by one of the authors. Based on the four classification methods used in this study we examine how the training/test size influences the results. In Figure 2, the accuracy is shown, for different sizes of the training set. The test-set is simply the remaining data, when the training set has been extracted. For all four classification methods, the methods give good results, even with a small training set/large test set.
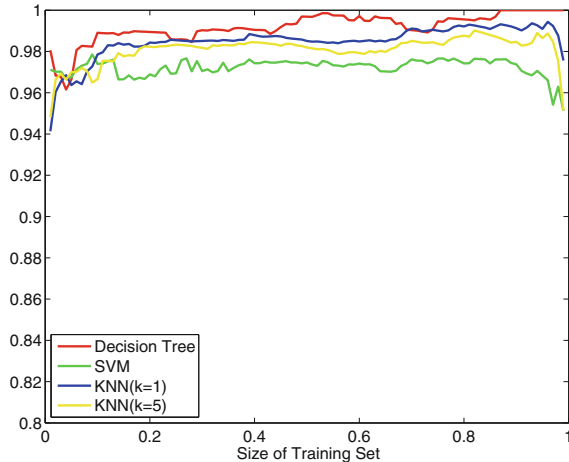

Fig. 2. The achieved accuracies for the four classifiers are illustrated, as a function of increasing size of the training set. The size is with respect to the total size of the data.

However, the classifiers are trained and tested without taking into account, that data trained from one infant is used to classify data from the same infant. We therefore train the classifiers on data from one infant, while the testing is done on data from the remaining infants. This is considered the worst case, as one could increase the size of the training set, by using more than one infant

for training. In Figure 3, the result can be seen, where cross-validation is used in order to consider training with all infants. Again, the size of the training set is varied, but in this case, the size of the total data set is related to the particular infant. This new choice of training/test sets yields an overall decrease in accuracy, as expected, but we are still able to obtain satisfactory results.
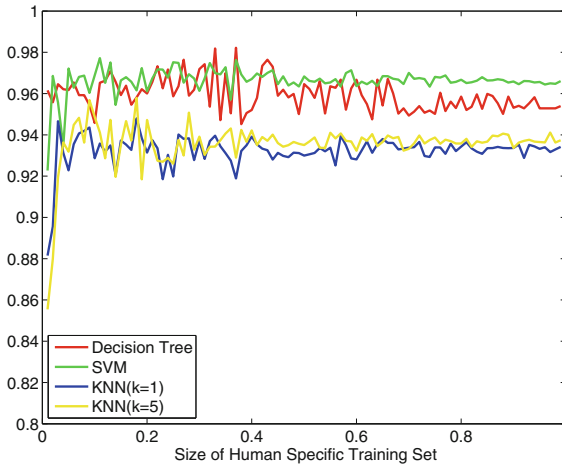

Fig. 3. For a more fair result, the training set is only based on data from one infant and the test set is based on data not belonging to the same infant. Cross-validation is used to train the classifiers on each infant.

Using a leave-one-out approach, we train a decision tree (the classifier that obtained the best results above) on data from all but one infant and segment the data from the leaved-out infant. The segmentation can be seen in Figure 4 where the ground truth segmentation can be compared with the estimated segmentation. Doing this for all infants, i.e. training the model on all but one infant and test the model on the leaved out infant, we are able to estimate the mean and standard deviation of the accuracy, which shows how good the methods are to generalize to an unknown infant. This has been done for all four methods and the results can be seen in Figure 5. We observe that we are still able to obtain good results for all four methods.

3.1   Parameter Importance

In order to point out the most important features used for detecting spontaneous movements, we use a leave-one-out approach. By removing one feature and training the classifiers, we compare the accuracy with the result obtained with the full set of features. This is done using 10-fold cross-validation. Figure 6 shows the results for the four types of classifiers. It should be noted that the importance-quantity has been normalized. It can be seen that the most important features
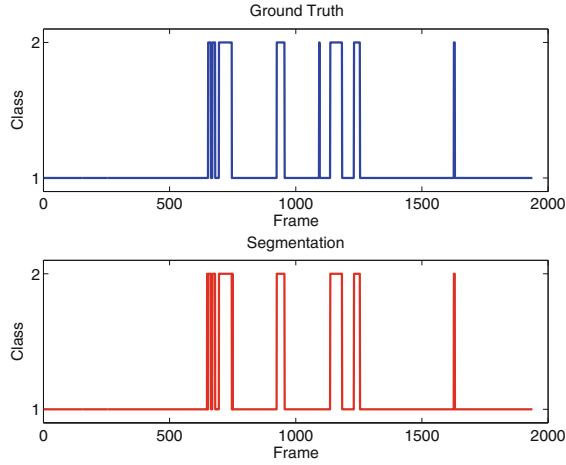
Fig. 4. Comparison between ground truth and estimated segmentation. Class 1 is spontaneous movements and Class 2 is non-spontaneous movements. Top: The ground truth segmentation of one infant's spontaneous movements. Bottom: The estimated segmentation of the infant's spontaneous movements.
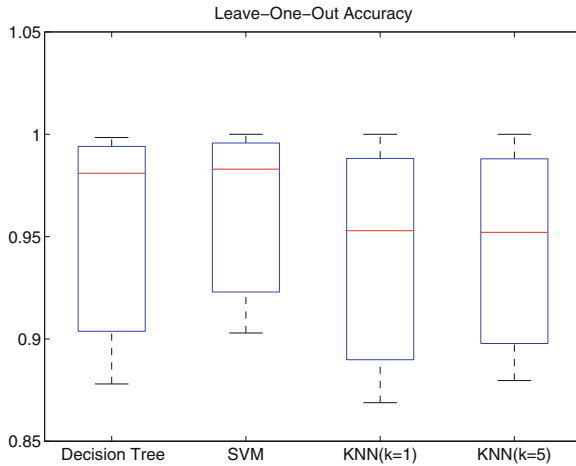

Fig. 5. Results for doing leave-one-infant-out experiment.

are based on velocities and that the maximum and summed velocity in a time windows significantly characterizes the spontaneous movements.
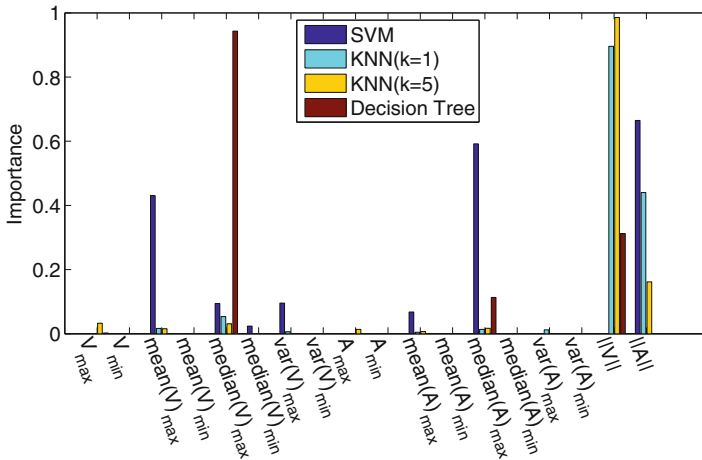

Fig. 6. The importance of the different parameters are illustrated for the four classifiers, based on the leave-one-out approach

4   Conclusion

Using annotated motion tracking data of moving infants in the age of 1-6 months, we have been able to segment sequences of spontaneous movements in infants. This was done using four different classifiers, which all proved to obtain similar results, ranging from $92-98\%$ accuracy, based on different classifiers and different sizes of the training/test data. In addition, we evaluated the importance of the different features used in this study, where the maximum velocity and summed velocity over time are important features for the spontaneous movements.

References

1. Jensen, R.R., Paulsen, R.R., Larsen, R.: Analysis of gait using a treadmill and a time-of-flight camera. In: Kolb, A., Koch, R. (eds.) Dyn3D 2009. LNCS, vol. 5742, pp. 154–166. Springer, Heidelberg (2009)
2. Gabel, M., Renshaw, E., Schuster, A., Gilad-Bachrach, R.: Full body gait analysis with kinect. In: Proceedings of EMBC 2012 (2012)
3. Little, J., Boyd, J.: Recognizing People by Their Gait: The Shape of Motion (1996)

4. Zhou, F., De la Torre, F., Hodgins, J.K.: Hierarchical Aligned Cluster Analysis for Temporal Clustering of Human Motion. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2013)
5. Agarwal, A., Thakur, M. K.: Sign language recognition using Microsoft Kinect (2013)
6. Einspieler, C., Prechtl, H.F.: Prechtl's assessment of general movements: a diagnostic tool for the functional assessment of the young nervous system. Mental Retardation and Developmental Disabilities Research Reviews (2005)
7. Einspieler, C., Prechtl, H.F.R., Bos, A.F., Ferrari, F., Cioni, G.: Prechtl's Method on the Qualitative Assessment of General Movements in Preterm. Term and Yong Infants (2005)
8. Heathcock, J.C., Bhat, A.N., Lobo, M.A., Galloway, J.C.: The relative kicking frequency of infants born full-term and preterm during learning and short-term and long-term memory periods of the mobile paradigm. Physical Therapy (2005)
9. Landgraf, J.F., Tudella, E.: Effects of external load on spontaneous kicking by one and two-month-old infants. Brazilian Journal of Physical Therapy (2008)
10. Olsen, M.D., Herskind, A., Nielsen, J.B., Paulsen, R.R.: Body-part tracking of infants. In: 22nd International Conference on Pattern Recognition (2014)
11. Olsen, M.D., Herskind, A., Nielsen, J.B., Paulsen, R.R.: Model-based motion tracking of infants. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014 Workshops. LNCS, vol. 8927, pp. 673–685. Springer, Heidelberg (2015)