Video-Based Early Cerebral Palsy Prediction using Motion Segmentation

Hodjat Rahmati[1], Ole Morten Aamo[1], Øyvind Stavdahl[1], Ralf Dragon[2], and Lars Adde[3,4]

***Abstract*— Analysing distinct motion patterns that occur during infancy can be a way through early prediction of cerebral palsy. This analysis can only be performed by well-trained expert clinicians, and hence can not be widespread, specially in poor countries. In order to decrease the need for experts, computer-based methods can be applied. If individual motions of different body parts are available, these methods could achieve more accurate results with better clinical insight. Thus far, motion capture systems or the like were needed in order to provide such data. However, these systems not only need laboratory and experts to set up the experiment, but they could be intrusive for the infant's motions. In this paper we build up our prediction method on a solution based on a single video camera, that is far less intrusive and a lot cheaper. First, the motions of different body parts are separated, then, motion features are extracted and used to classify infants to healthy or affected. Our experimental results show that visually obtained motion data allows cerebral palsy detection as accurate as state-of-the-art electromagnetic sensor data.**

I. Introduction

Early identification of cerebral palsy (CP), one of the major disabilities resulting from preterm birth [13], [14], facilitates early intervention and targeted follow-up of children who are likely to develop CP, and provides reassurance to the parents of children who are unlikely to develop CP. The diagnosis of CP is traditionally assured by 2 years of age, but reassessment can not provide the sub-type and clinical picture in young children with CP until 4-6 years of age [7]. Research has shown that assessment of general movements in young infants can be used as an early prognostic tool to identify CP during the fidgety movement's period (9-18 weeks post term age) [6]. However, the assessment is based upon an expert's clinical observation, which limits the widespread clinical use of general movements ([2], [3]).

Over the last years, a number of computer-based movement assessment tools have been developed aiming at performing quantitative analysis during the fidgety movement's period ([15], [12], [11]). However, they suffer from drawbacks that limit their widespread use. They must be installed in a controlled environment, use instrumentation that might affect the infant body movements, and experts are needed for interpretation and analysis of results. Recently, our research group has reported accurate results in the detection of abnormal general movements and prediction of CP using a normal monocular 2D camera and a simple frame differencing software without any need for instrumentation of the infant [3], [2], [1], [17] . While these early studies show great promise, the algorithms used for data analysis are sensitive to lighting conditions, clothing and skin color. In addition, the video stream is aggregated into movement variables or features that provide limited clinical insight.

In a clinical observation, there are many features that can be used to predict cerebral palsy. Most of these features can be employed by a computer-based method only if individual motions of different body parts are available. In addition, the relation between motions of body parts might have a strong analytic cue for prediction of CP status and CP sub-type [10]. Therefore, in order to provide an improved analytic tool to study general movements and predict CP we proposed in [16] a new motion segmentation method that tracks individual body parts separately using just a standard monocular video camera without need for extra equipment. In the present paper, we employ this methodology to the problem of predicting CP, and compare the results with those obtained by using state-of-the-art motion tracking sensors.

We hypothesized that our proposed video segmentation method had accuracy in predicting CP alike the specific motion tracking sensors when the assessments were performed during the fidgety movements period using the same motion features and classification methods. To do so, our paper provides a method to extract features from motion data of infants and classify the extracted information into one of the classes (healthy or affected) using a classification model, wherein the classification model is trained using data derived from the movements of other subjects with known category.

The extracted information relates to patterns of movement which may, or may not, be readily recognisable to a human observer. Since this approach does not involve or require these patterns to be defined or recognised as such, the method is not dependent on particular human-defined parameters. Instead, the information is extracted from the movement data in order to classify them. Thus, the present technique can take movement phenomena that are otherwise incomprehensible to humans (or they are comprehensible just for experts) into account, for example because they involve complex inter-relationships of the movements of a plurality of limbs.

The paper is structured as follows: Section II summarizes the motion segmentation method. Section III explains the feature used for classification. Section IV focuses on classification issues. Section V explains how input data are collected. In section VI, numerical results are discussed. Finally, section VII concludes the paper.

[1]Department of Engineering Cybernetics, NTNU, Trondheim, Norway
[2]Computer Vision Lab, ETH, Zurich, Switzerland
[3]Department of Laboratory Medicine, Children and Woman's Health, Faculty of Medicine, NTNU, Trondheim, Norway
[4]Clinic for Clinical Services, St. Olavs University Hospital, Trondheim, Norway

978-1-4244-7929-0/14/$26.00 ©2014 Crown

II. Motion Segmentation

In this section, we explain the measurement system that extracts the motion trajectory for each body part out of a monocular video without extra instruments. Our measurement system is based on a framework proposed by [16]. This method obtains a track (trajectory) for an object in three phases. First, dense trajectories of movement are computed for the whole image. Second, a graph-based segmentation algorithm is applied to the set of trajectories in order to separate them into groups representing the individual body parts. Third, a tracker is applied to each group to compute one single trajectory for each body part. In the present case, six groups were used, representing the right hand and arm, the left hand and arm, the right foot and leg, the left foot and leg, the head, and the trunk.

A. Trajectories

Trajectories spread over the whole canvas of the video are obtained using optical flow. We use the LDOF (Large Displacement Optical Flow) algorithm proposed by [18], because it provides a very dense trajectory field which is of particular importance in our application. In detecting cerebral palsy, the key feature is fidgety movement, which can occur in any part of the infant's body. So, having a tracker which is able to cover the whole body and tracks every points might end up with a more reliable diagnosis. Furthermore, since the distribution of trajectories from optical flow is more uniform than from salient points, such as SIFT (Scale Invariant Feature Transform) [19], the segmentation, which is the second step, is more accurate.

B. Segmentation

To distinguish individual body parts from each other, trajectories are separated into groups - referred to as segments - such that similar ones are mapped to the same segment. The task of splitting is performed by a graph-cut optimization where each vertex of the graph represents a trajectory, and vertices are connected by weighted edges. The weights are measures of similarity between the trajectories. Qualitatively, the weight between two trajectories is higher the closer they are to each other, and higher the more similar their motion patterns are. The output of this phase is an assignment (labelling) of each trajectory to one of the predefined segments. As only human interaction, the user must initially label a small number of trajectories manually. The upper row of Figure 1 shows segmentation result for an infant.

C. Tracking Based on Segmentation

The final stage of the method provides one single trajectory for each body part. The algorithm iteratively computes the trajectory using the set of trajectories assigned to segment i, denoted $\mathcal{O}_i$, as follows. For each time t, we define the subset of all trajectories $s \in \mathcal{O}_i$ that are visible at t and $t+1$ as $\mathcal{S}_t$. Let $\mathbf{x}_t^s$ and $\mathbf{x}_{t+1}^s$ denote the locations of $s \in \mathcal{S}_t$ at time t and $t+1$, respectively. Initialize $\mathbf{x}_0$ from the center of mass of all trajectories assigned to $\mathcal{O}_i$ in the first frame, that is

$$\mathbf{x}_0 = \frac{1}{|\mathcal{S}_0|} \sum_{s \in \mathcal{S}_0} \mathbf{x}_0^s. \tag{1}$$

Computing the location of the trajectory by using the center of mass of $\mathcal{O}_i$ at $t > 0$ will fail due to discontinuity from partial occlusions. Instead, the following iterative procedure is used to update the tracker results

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \frac{1}{|\mathcal{S}_t|} \sum_{s \in \mathcal{S}_t} (\mathbf{x}_{t+1}^s - \mathbf{x}_t^s). \tag{2}$$

Since eq. (2) builds the update step by exploiting a large number of trajectories, it can filter out noise and unreliable trajectories, as long as their effects remain small compared to that of the majority of correctly labeled trajectories. The lower row of Figure 1 shows tracking result for the segmented body parts of an infant.

In order to overcome the illumination effect, LDOF constrains on gradient which is invariant to illumination changes. In addition, it integrates descriptor matching to the variational technique that makes the flow fields more robust. Furthermore, averaging over large set of trajectories reduces the noise effect. These make the obtained trajectories robust to illumination change and different skin color and clothing.

III. Features Extraction

Feature extraction involves reducing the amount of information required to describe a large set of data accurately and problem-specific. When analyzing complex data, one of the major problems arises from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

As described in [9] the lack of a fluent character in the movement and the rigidity of the body could be a predictor for CP. A nonfluent motion pattern and rigidity of the body can be observed in the motion signal. For visual comparison the components of typical motion trajectories of an impaired and an unimpaired infant are shown in Figures 2 and 3. In addition, a high correlation between two limbs might indicate a lack of normal behaviour [10]. Among extensive features that could be employed for the task of CP detection, three types of features are extracted as proposed in [15]: *area out of standard deviation (STD) from moving-average*, *periodicity* and *correlation between trajectories*. The first two represent fluency and monotony character of the motion, and the last is a measure of mutual dependencies between the limbs motions.

A. Area Out of STD from moving-average

The movement of an unimpaired infant is smoother compared to impaired one. A parameter that could represent smoothness of a trajectory is its deviation from an average

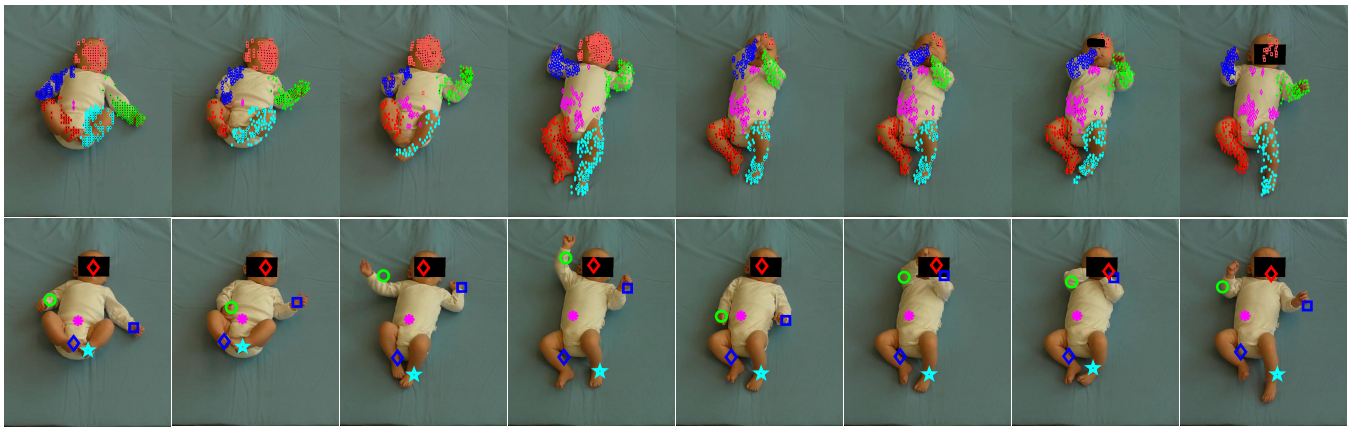

Fig. 1. Overview of segmentation (upper row) and tracking (lower row) results of [16].

version of its motion. This deviation is calculated by measuring the area out of STD from moving-average. To measure this feature, first, the moving average for each point $\mathbf{x}_t$ of the trajectory is calculated as the arithmetic-mean of a window of the trajectory centered on that point as follows,

$$\bar{\mathbf{x}}_t = \frac{1}{n_w} \sum_{k \in \mathbf{w}_t} \mathbf{x}_k, \tag{3}$$

where $\mathbf{w}_t$ is a subset of the trajectory centered on $\mathbf{x}_t$, and n_w is the size of $\mathbf{w}_t$. n_w should be so large that neither averaging is meaningless nor it is diffused. In our experiments $n_w = 25$, which it is equal to number of samples in $1s$. Then, the standard deviation for each point is obtained:

$$\sigma_t^2 = \frac{1}{n_w} \sum_{k \in \mathbf{w}_t} (\mathbf{x}_k - \bar{\mathbf{x}}_k)^2 \tag{4}$$

Finally, the area of the trajectory that exceeds $\bar{\mathbf{x}}_t \pm \sigma_t$ is summed up and normalized by the length of the signal in order to calculate the area out of STD from moving-average for that trajectory. The red area in Figure 2 shows this feature.

B. Periodicity

As Figure 3 indicates, the movement of an affected infant is more stationary and periodic while a healthy infant show more complicated, time-dependant characters in their motion. The periodicity P of the signal can be seen as a measure for the occurring frequencies in the movement patterns. We measure P by counting the number of intersections between the signal and a local mean-value. To do so, we divide the signal into three subsets with equal length and calculate the mean value for each subset, black solid lines in Figure 3, then calculate the time-spans between consecutive intersections between these lines and the signal. Then, the arithmetic mean μ and standard deviation σ of these time-spans are derived. Finally, periodicity is defined as

$$P = \frac{1}{\mu + \sigma}. \tag{5}$$

Since μ and σ are independent of the signal length, we don't need to normalize the periodicity with the signal length.

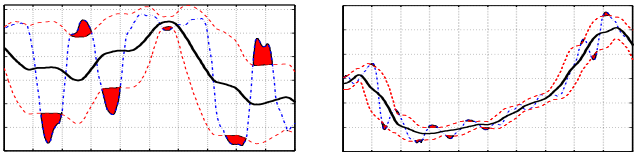

Fig. 2. Horizontal position over time (pixel vs frame) for the right hand. Area out of STD of moving-average for an impaired (left) and an unimpaired infant (right). Solid black, dash-dot blue, and dash red lines are the moving-average, main trajectory, and borders of STD, respectively. The red area show area out of STD from moving-average.

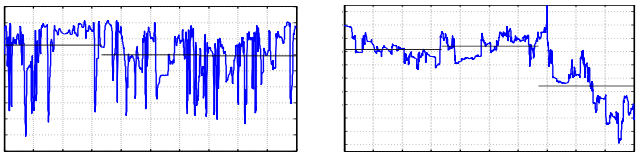

Fig. 3. Horizontal position over time (pixel vs frame) for the right foot. Periodicity for an impaired (left) and an unimpaired infant (right). Solid black lines show the local mean-values, and blue lines stand for the trajectories.

C. Correlations Coefficient

Affected infants tend to show more mutual dependencies on their body motion with respect to healthy ones where each body limbs moves more independently. To measure these dependencies we calculate correlation coefficients between body limbs as well as the head and the trunk.

The aforementioned features are calculated separately for each trajectory (two hands, two feet, head, and trunk) and combined to a feature vector for a whole sequence. In the next step, the feature vector of different infants are used as input to train the classifier. In the classification phase the trained classifier assigns one of the two classes (*affected* and *healthy*) to each member of the test set.

IV. Classification

Due to promising performance of *support vector machine* classifiers (SVM) as well as having properties such as simultaneously minimizing the empirical classification error and maximizing the geometric margin between different

classes, we consider this classifier for our classification task. SVM is a supervised learning method that analyzes data and recognizes patterns used for classification. Given a set of training data, each labeled as belonging to one of two classes, an SVM training algorithm builds a model that assigns new data into one class or the other.

A support vector machine constructs set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, and tries to find the hyperplane with largest distance from the nearest points of any training data classes. The original idea of optimal hyperplane was a linear classifier and it is generalized to nonlinear case by introducing kernels that map the space to a higher dimensional such that it is linearly separable in the new space [4]. The current standard incarnation was proposed by [8], it uses soft margin to choose a hyperplane that splits the training set as cleanly as possible in situations where there is no hyperplane that can classify all the data correctly.

The effectiveness of SVM depends on the selection of the kernel functions and soft margin parameter C. In our application any value $C \geq 1$ led to the same results, and we choose the kernel function to be linear and the margin parameter $C = 1$.

V. Input Data

The data set we studied consists of 78 infants with an age of 10-18 weeks post-term, from which 12 were confirmed diagnosed with CP at five years age and two at two years age. Therefore, altogether 14 of the 78 infants were confirmed diagnosed with CP. Two sets of data were captured from each infant: first, 6 sensors placed on the body, one for each wrist, one for each ankle, one for head, and one for chest, to measure the infant's movement. As the second data set, infants were filmed by a normal monocular camera at the same time when sensors were capturing the motions.

A. Video Data

A Sony DCR-PC 100E camera recorded the videos of the infants with a frame rate of 25 frames/s. The infant was placed on a standard mattress with rigid, transparent walls and a stationary digital video camera was installed at a distance of about 110 cm above the infant. This resulted in an experimental set-up where a similar camera position was assured for all recordings.

B. Sensor Data

Movements of the infants in terms of x, y, and z coordinates were captured simultaneously with 25 Hz using six miniBird motion sensors. One sensor was placed on each of the infant's wrists, one on each ankle, one on the sternum and one on the forehead as described in Table 4 and Figure 5.

Each infant's movement data is extracted from one of several recordings. The interesting and relevant temporal parts of the selected recording, also referred to as *Region Of Interest* (ROI), has been analysed and selected by a physiotherapist. This means that every record contains one or more ROIs. As an example, Figure 6 displays measured

Fig. 4. Placement of the sensors [5].

Sensor nr.	Placement
Sensor 1	Left ankle
Sensor 2	Chest
Sensor 3	Right wrist
Sensor 4	Left wrist
Sensor 5	Right ankle
Sensor 6	Head

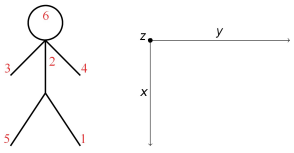

Fig. 5. Coordinate system [5].

values for all 6 sensors in xy-plane for a simple ROI. It can be concluded that all the data accessed and utilized during this paper, is composed of ROIs from infants with normal and abnormal movements. A ROI contains time series data from all sensors in x, y and z directions.

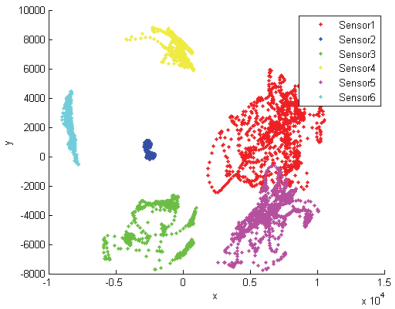

Fig. 6. Measurements from different sensors [5].

VI. Experimental Results and Discussions

In this section, we study the classification performance on data sets from both sensors and the motion-segmentation method. To do so, first, we obtain track of the body parts as explained in section II. Then, the aforementioned feature set is extracted from both sensor data and trajectories obtained by the motion-segmentation method. Finally, the same classifier is applied on the both feature sets to separate affected infants from healthy ones. We used cross-validation to assess how the results of our analysis will generalize to an independent data set. To do so, an iterative procedure is considered such that each time one of the subjects is used for testing the classification model built by training the classifier on the rest of the data set. This is repeated until all the data set has been used for testing once. To measure the prediction performance, we use three measurements as follows,

$$\textit{Sensitivity} = \frac{TP}{TP + FN},$$

$$\textit{Specificity} = \frac{TN}{TN + FP},$$

$$\textit{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP},$$

where TP is the number of correctly labeled as affected, FN is the number of wrongly diagnosed as healthy, TN is the number of correctly detected as healthy, and finally, FP is the number of wrongly labeled as affected. Table I

shows these results for both data sets. As it can be seen, the data provided by motion segmentation method is as rich as sensor data with respect to classification results. The reason for slightly better results of motion segmentation is that this method obtains a track for the body limbs by using information of a large set of trajectories tracking different particles on each part. Therefore, it captures more informations. In addition, considering the fact that motion segmentation method uses only 2D motions of the videos, while results for sensors are based on 3D motions, makes our results more valuable.

By further notice to the results, it can be seen that the achieved sensitivity is relatively low. The reasons for that could be twofold: first, although we have the most comprehensive data set reported so far, the number of affected infants is just 14 and this is a small number to derive a general model for affected infants. Second, we used the features proposed by [15] which might not be sensitive enough.

TABLE I

CLASSIFICATION RESULTS FOR BOTH SENSOR AND THE MOTION-SEGMENTATION DATA SET.

Data set	Sensitivity	Specificity	Accuracy
Motion Segmentation	50%	95%	87%
Sensor Data	50%	92%	85%

Considering the fact that our method is based on a simple monocular camera, and doesn't need any extra equipment or clinician experts, makes it economically cheap to predict CP and it could be used in all clinics, as it is our final goal. These facts make our method even more valuable. The proposed method is the first video based approach that treats different body parts separately. Since the track for each part is obtained from a large set of points on that part, the results are more robust to noise. In addition, it could capture every small motion of the body parts. Finally, having motions of different parts separately has the potential of distinguishing CP sub-types.

VII. CONCLUSIONS

Early detection of cerebral palsy is important in order to establish training programs to reduce the functional consequences of the brain damage. This is possible by analyzing motion patterns that occur during young infancy. Previous computer based approaches suffer either from being intrusive to the infant's motion pattern by using extra instrument, or lack of precise analytic explanation for their results weakening the construct validity of the methods. In this paper, we employed the motion-segmentation method to extract motion data out of video. By this, we strongly indicate the possibility of bringing such assessments into clinical settings without need for clinical experts. It is expectable to increase possibilities to make predictions of also CP sub-types by analyzing different body parts separately. The experimental results indicate that our motion-segmentation method could replace motion capture system.

REFERENCES

[1] L. Adde, J. Helbostad, A. R. Jensenius, M. Langaas, and R. Støen. Identification of fidgety movements and prediction of cp by the use of computer-based video analysis is more accurate when based on two video recordings. *Physiotherapy theory and practice*, 29(6):469–475, 2013.

[2] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, K. H. Grunewaldt, and R. Støen. Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study. *Developmental Medicine & Child Neurology*, 52(8):773–778, 2010.

[3] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, and R. Støen. Using computer-based video analysis in the study of fidgety movements. *Early human development*, 85(9):541–547, 2009.

[4] A. Aizerman, E. M. Braverman, and L. Rozoner. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control*, 25:821–837, 1964.

[5] A. Berg. Modellbasert klassifisering av spedbarns bevegelser. 2008.

[6] M. Burger and Q. A. Louw. The predictive validity of general movements–a systematic review. *European Journal of Paediatric Neurology*, 13(5):408–420, 2009.

[7] C. Cans. Surveillance of cerebral palsy in europe: a collaboration of cerebral palsy surveys and registers. *Developmental Medicine & Child Neurology*, 42(12):816–824, 2000.

[8] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[9] C. Einspieler and H. F. Prechtl. 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*, 11(1):61–67, 2005.

[10] N. Kanemaru, H. Watanabe, H. Kihara, H. Nakano, R. Takaya, T. Nakamura, J. Nakano, G. Taga, and Y. Konishi. Specific characteristics of spontaneous movements in preterm infants at term age are associated with developmental delays at age 3 years. *Developmental Medicine & Child Neurology*, 2013.

[11] D. Karch, K.-S. Kang, K. Wochner, H. Philippi, M. Hadders-Algra, J. Pietz, and H. Dickhaus. Kinematic assessment of stereotypy in spontaneous movements in infants. *Gait & posture*, 36(2):307–311, 2012.

[12] D. Karch, K. Wochner, K. Kim, H. Philippi, M. Hadders-Algra, J. Pietz, and H. Dickhaus. Quantitative score for the evaluation of kinematic recordings in neuropediatric diagnostics. *Methods Inf Med*, 51(4):341–347, 2012.

[13] B. Larroque, P.-Y. Ancel, S. Marret, L. Marchand, M. André, C. Arnaud, V. Pierrat, J.-C. Rozé, J. Messer, G. Thiriez, et al. Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the epipage study): a longitudinal cohort study. *The Lancet*, 371(9615):813–820, 2008.

[14] N. Marlow, D. Wolke, M. A. Bracewell, and M. Samara. Neurologic and developmental disability at six years of age after extremely preterm birth. *New England Journal of Medicine*, 352(1):9–19, 2005.

[15] L. Meinecke, N. Breitbach-Faller, C. Bartz, R. Damen, G. Rau, and C. Disselhorst-Klug. 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.

[16] H. Rahmati, R. Dragon, O. M. Aamo, L. Van Gool, and L. Adde. Motion segmentation with weak labeling priors. In *Submitted to GCPR*, 2014.

[17] A. Stahl, C. Schellewald, Ø. Stavdahl, O. M. Aamo, L. Adde, and H. Kirkerod. An optical flow-based method to predict infantile cerebral palsy. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 20(4):605–614, 2012.

[18] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. In *Computer Vision–ECCV 2010*, pages 438–451. Springer, 2010.

[19] H. Zhou, Y. Yuan, and C. Shi. Object tracking using sift features and mean shift. *Computer Vision and Image Understanding*, 113(3):345–352, 2009.