Motion Segmentation with Weak Labeling Priors

Hodjat Rahmati[1,⋆], Ralf Dragon[2], Ole Morten Aamo[1], Luc Van Gool[2], and Lars Adde[3,4]

[1]Department of Engineering Cybernetics, NTNU, Trondheim, Norway
{hodjat.rahmati, ole.morten.aamo}@itk.ntnu.no
[2]Computer Vision Lab, ETH, Zurich, Switzerland
{dragon, vangool}@vision.ee.ethz.ch
[3]Clinic for Clinical Services, St. Olavs University Hospital, Trondheim, Norway
[4]Department of Laboratory Medicine, Children and Woman's Health, Faculty of Medicine, NTNU, Trondheim, Norway
lars.adde@ntnu.no

Abstract. Motions of organs or extremities are important features for clinical diagnosis. However, tracking and segmentation of complex, quickly changing motion patterns is challenging, certainly in the presence of occlusions. Neither state-of-the-art tracking nor motion segmentation approaches are able to deal with such cases. Thus far, motion capture systems or the like were needed which are complicated to handle and which impact on the movements. We propose a solution based on a single video camera, that is not only far less intrusive, but also a lot cheaper. The limitation of tracking and motion segmentation are overcome by a new approach to integrate prior knowledge in the form of weak labeling into motion segmentation. Using the example of Cerebral Palsy detection, we segment motion patterns of infants into the different body parts by analyzing body movements. Our experimental results show that our approach outperforms current motion segmentation and tracking approaches.

1 Introduction

We aim at segmenting motion data from monocular videos into underlying body parts. Motion is an important cue for the clinical diagnosis of, for instance, cardiovascular diseases [17], Cerebral Palsy [1], or for gait analysis [9]. In this paper, we focus on the motion of infants to detect Cerebral Palsy (CP) which is a set of chronic conditions affecting body movements, posture and muscle coordination. It is caused by damage to one or more specific areas of the brain, usually occurring during foetal development or infancy. The absence of normal movement between 2 to 4 months post-term age has been shown to be a strong predictor of later Cerebral Palsy [1].

Accurate diagnosis can be achieved by analyzing the infant's bodily motion at the necessary level of detail. Recently, a number of computer-based methods have aimed to quantitatively analyze general movements in order to detect CP. However, they either use extra instruments [13] that are intrusive to the diagnosis task, or they don not provide analytic results [20]. The reciprocal relation between body parts is a strong analytic cue

⋆ Corresponding author

for CP detection. E.g., [8] proposed that a high correlation between two limbs might indicate a lack of normal behaviour. Therefore, in order to perform an analytic tool to study the disease we propose a new segmentation method to separate and track different body parts without need for any extra instrument.

The video analysis of infant motion is very challenging. Infants will often cross their limbs, twist arms, move abruptly, etc. As a result during such motions, state-of-the-art trackers, e.g. TLD [7], drift and fail at tracking the body parts at the required level of precision. Similarly, body pose trackers have difficulties due to the high motion variability. Motion segmentation, on the other hand, has no drifting problems if the body parts move distinctly – motion even simplifies the segmentation. However, an initial bounding box or an initial segmentation prior cannot be integrated conveniently and, without it, the segmentation performance is poor. Our key contribution is a new energy-minimization-based formulation of motion segmentation that allows for the integration of prior information. The result is a system that can segment long sequences reliably.

Related Work Motion segmentation is the task of grouping point trajectories from an image sequence subject to coherence of their motion over time. While earlier work focused on assigning the trajectories to subspaces, e.g. with the generalized PCA [23], subsequent work exploited sparsity [6,12] or non-negative matrix factorization [4]. Further works exploit temporal smoothness [19] or depth ordering [11]. In most recent works, the pairwise [6,4,3,5,12,19] or higher-order [15] relationships between trajectories are aggregated and a final spectral clustering [14] step of an affinity matrix $\mathbf{A}$ encoding trajectory similarity finds the association of trajectories to motion segments.

In contrast to image segmentation approaches, those for motion segmentation are unsupervised, i.e. it is neither necessary nor possible to specify prior knowledge about the assignment of trajectories to motion segments. Modifying $\mathbf{A}$ to prevent trajectories with the same prior label to be clustered into different segments has undesired biases which leads to poor results: Since spectral clustering minimizes the inter-segment cut through $\mathbf{A}$, weighted by the sum of intra-segment weights [14], tweaking local weights creates a global bias which results in non-intuitive results. In terms of weakly supervised clustering, also called transductive learning, [24] derived a simple iterative framework in which the final k-Means-step in spectral clustering is replaced by an iterative procedure. It alternates between computing a mapping $\mathbf{F}$ from graph Laplacian to class association and regularizing $\mathbf{F}$ regarding the initial labels. While they proved that the mapping converges to a reasonable solution respecting the initial labels, our energy-based approach allows specifying additional unary terms for trajectories which are similar to initially-labeled ones. In our work, we overcome the no-prior limitation by formulating the motion clustering task as a multi-label MRF, similar to graph-cut-based image segmentation [2]. We use the generalized Potts model, thus encouraging large motion differences between segments and similar motions within segments. This may seem straightforward, but to the best of our knowledge such a scheme has not been proposed before. MRFs have been used for dense image segmentation, but only based on short-term motion obtained from optical flow, e.g. [21]. In contrast, we segment sparse motion information from trajectories which last over many frames.

The rest of the paper is organized as follows. In Sec. 2, we derive our energy formulation, in Sec. 3, *tracking by segmentation* is explained, in Sec. 4, we present experimental results, and Sec. 5 concludes the paper.

2 Integrating Prior Knowledge into Motion Segmentation

Energy Formulation This section describes the energy minimization framework used to segment trajectories into the infant's body parts. Let s be a trajectory from the set of all trajectories $\mathcal{S}$, and $\boldsymbol{L}$ be the unknown label vector assigning each trajectory s to a segment L_s. The motion segmentation is obtained through a multi-label graph-cut [10] that minimizes the energy $E(\boldsymbol{L})$ associated to a segmentation $\boldsymbol{L}$.

Similar to its use in an already extensive image segmentation literature [2],

$$E(\boldsymbol{L}) = \sum_{s \in \mathcal{S}} D(s, L_s) + \sum_{(s,r) \in \mathcal{N} \times \mathcal{N}} V(s, r, L_s, L_r). \tag{1}$$

The data term D is a penalty function that encourages high intra-segment motion similarity and V is an interaction potential that enforces low inter-segment similarity for trajectories in a local neighborhood $\mathcal{N}$ which is defined as follows,

$$\mathcal{N} = \{(s,r) \mid (s,r) \in \mathcal{S} \times \mathcal{S} \wedge d_{sp}(s,r) \leq 10 \wedge (s,r) \text{ have temporal overlap}\}, \tag{2}$$

where $d_{sp}(s,r)$ is the average spatial Euclidean distance over the common frames. Due to occlusions and fast motions, trajectories are asynchronous and span different temporal windows. Considering just trajectories that last for the whole shot lower the number of tracked points, leaving us possibly even with an empty set. So, we obtain the energy for all trajectories that have at least one frame in common. Due to transitivity, it can be expected that even trajectories that share no frames can be paired [3].

Data Term Let $\mathcal{S}_I \subset \mathcal{S}$ be the set of those trajectories s that are initially labeled with I_s. Then the data term in eq. (1) is defined as

$$D(s, L_s) = \begin{cases} 0 & \text{if } s \in \mathcal{S}_I \wedge L_s = I_s \\ K_s & \text{if } s \in \mathcal{S}_I \wedge L_s \neq I_s \\ g(s, L_s) & \text{if } s \notin \mathcal{S}_I \end{cases} \tag{3}$$

where $K_s = 1 + \sum_{(r,s) \in \mathcal{N} \times \mathcal{N}} V(s, r, L_s, L_r)$ is a large value that enforces trajectories which are initially labeled to preserve their labels during the optimization process. The first two cases in eq. (3) enforce the prior knowledge, while $g(s, L_s)$ extends prior knowledge towards initially unlabeled trajectories: $g(s, L_s)$ is a measure of dissimilarity between trajectory s and subset $\mathcal{O}_{L_s}$ – the set of trajectories that are initially assigned label L_s. We define the energy g as negative log-likelihood of the average trajectory similarity $w \in [0, 1]$ which will be given in eq. (8):

$$g(s, L_s) = -\log\big(\operatorname*{mean}_{r \in \mathcal{O}_{L_s}}((w(s,r))^{\gamma})\big). \tag{4}$$

Thus, the average similarity between a trajectory s and the set of initially-labeled trajectories $\mathcal{O}_{L_s}$ is computed using the arithmetic mean, similar to a mixture distribution.

Pairwise Term We define the pairwise Energy term in eq. (1) as:

$$V\big(s, r, L_s, L_r\big) = \big(1 - \delta(L_s, L_r)\big) f\big(w(s,r)\big), \quad f\big(w\big) = -\log(1 - w^{\phi}), \tag{5}$$

where δ is the Kronecker delta function which leads to penalizing neighboring trajectories s and r from different segments by a penalty f that depends on the trajectory similarity $w(s,r)$. The more similar s and r are the higher the penalty of assigning them to different segments will be. f is defined as the negative log-likelihood of w which is weighted by ϕ analog to w in eq. (4). γ and ϕ non-linearly balances D and V in eq. (1). They are empirically set to $\gamma = 0.1$ and $\phi = 0.001$.

Trajectory Similarity The trajectory similarity $w(s,r)$ is a probabilistic measure if the two trajectories s and r belong to the same moving object. Since such trajectories usually move similarly and tend to be spatially closer than trajectories with different associations, our definition contains a motion and a distance term:

$$d_t^2(s,r) = \frac{d_{\text{sp}}(s,r)}{\delta_{\text{c}}(s,r)} \cdot d_{\text{mot}}^2(t,s,r), \quad d_{\text{mot}}^2(t,s,r) = \frac{\|\boldsymbol{v}_t^s - \boldsymbol{v}_t^r\|^2}{5\sigma_t^2(s,r)}, \tag{6}$$

where $d_{\text{sp}}(s,r)$ is the average spatial Euclidean distance over the common frames, d_{mot} is the normalized motion distance at time t, and d_t is the combined distance.

Unlike [3], we scale the spatial distance by the factor $\delta_{\text{c}}(s,r) = \log(1 + n_{\text{c}}(s,r))$ where $n_{\text{c}}(s,r)$ is the number of frames that the trajectories s and r have in common. This takes into account that the more frames two trajectory have in common the more reliable the similarity result is. The reasons are twofold: First, the short length of a trajectory indicates severe change in the neighbourhood of that trajectory (trajectories terminate in case of occlusion or fast changes, and the short length shows two of such cases happens in a short period). Thus, the optical flow is likely to be inaccurate in such a situation and that trajectory might be wrongly developed. Second, since similarity is obtained over common frames, two trajectories that show similar motion in their common frames while they have completely different motions in other frames, still get a high similarity. Therefore, it is reasonable to account less value for similarities obtained over a smaller number of frames. In the definition of d_{mot}, $\boldsymbol{v}_t^s$ is the aggregated motion of a trajectory s over 5 frames ($\boldsymbol{v}_t^s = \boldsymbol{x}_{t+5}^s - \boldsymbol{x}_t^s$). σ_t is an adaptive normalization which enables dealing with both fast and slow motions. In particular, local variations among velocities within a segment should be tolerated more if motions in the segment are changing more rapidly. Therefore, σ_t is defined, as presented in [3]:

$$\sigma_t(s,r) = \min_{a \in \{s,r\}} \sum_{t'=0}^{4} \sigma(\boldsymbol{x}_{t+t'}^a, t+t'), \tag{7}$$

where $\sigma(\boldsymbol{x}, t)$ is the local variation in the flow field at position $\boldsymbol{x}$ and frame t.

As long as two objects move next to each other, they share similar motions, and it is impossible to separate them as different objects. But as soon as they start to move differently, they can be distinguished. In order to exploit this information, the distance d between two trajectories considers the instance when they start behaving differently. So, $d(s,r) = \max_t d_t(s,r)$. Finally, the edge weight between trajectories is defined as

$$w(s,r) = \exp(-d^2(s,r)), \quad 0 \leq w(s,r) \leq 1. \tag{8}$$

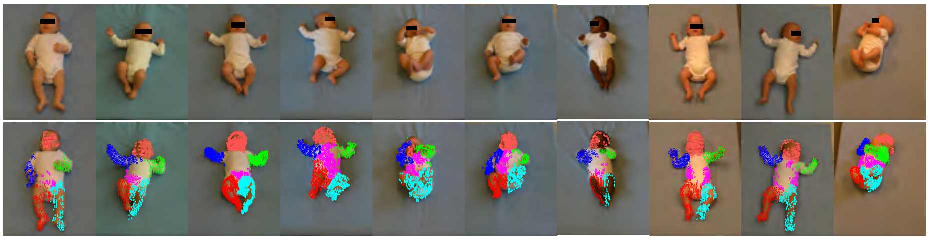

Fig. 1. Sequences 1 (left) to 10 from the experiments (upper row) and segmentation results of our proposed method (lower row) in frame 250.

3 Tracking Based on Segmentation

Along side segmentation, tracking is another important issue. Although there are many tracking algorithms already providing astonishing results on the type of sequences for which they have been designed, we experimentally found none of them to perform sufficiently well on Cerebral Palsy problem. The reasons are manifold: fast motions, high nonrigidity, frequent changes in appearance, etc. For example TLD as proposed in [7], despite being fast and reliable tracking for many applications, fails to track the limbs (upper row of Fig. 9). Therefore, we propose a motion segmentation based tracker. It tracks a specific point $\boldsymbol{x}$ using the motion from segment $\mathcal{O}_i$. We could initialize $\boldsymbol{x}$ manually, or from the center of mass at a labeled frame of all trajectories in $\mathcal{O}_i$. Tracking $\boldsymbol{x}$ using the motion of the center of mass of $\mathcal{O}_i$ would fail due to discontinuity from partial occlusions. Instead, an iterative procedure is used to update the tracking results, as follows. For each segment $\mathcal{O}_i$ and each time step 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 $\boldsymbol{x}_t^s$ and $\boldsymbol{x}_{t+1}^s$ denote the respective locations of s. Then, $\boldsymbol{x}$ is updated iteratively using the average motion of the trajectories:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \frac{1}{|\mathcal{S}_t|} \sum_{s \in \mathcal{S}_t} (\boldsymbol{x}_{t+1}^s - \boldsymbol{x}_t^s). \tag{9}$$

Since eq. (9) 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.

4 Experimental Results and Discussion

In this section the performance of our proposed method is analyzed on two different data sets. First, several videos from infants that have been taken to study Cerebral Palsy are used to analyze the motion of body parts. The second part investigates the generality of the proposed motion segmentation algorithm by applying it on standard data sets.

4.1 Performance on Videos of Infants

In all experiments in this section, we used the experimental set-up that was used in *(St. Olavs Hospital, Trondheim, Norway)*. During the experiments, we analyzed the first

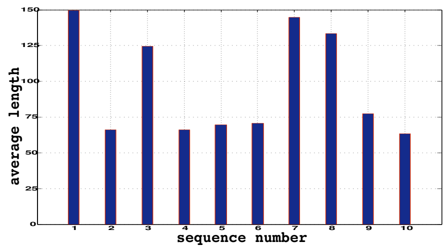

Fig. 2. Average length (in number of frames) of the trajectories in number of frames for different sequences.

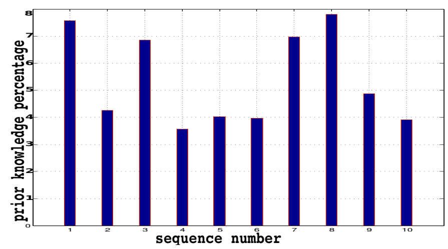

Fig. 3. Percentage (%) of trajectories used as prior knowledge with respect to the total trajectories number for different sequences.

Fig. 4. Seq. 1 ground-truth segmentation for frames 1, 50, 200, 300 from left to right.

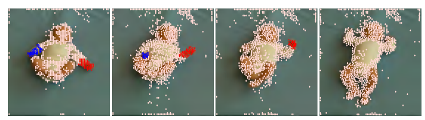

Fig. 5. Seq. 1 segmentation results of [3] for frames 1, 50, 200, 300 from left to right.

1000 frames of 10 sequences showing different infants carrying out different motions (Fig. 1). It is worth mentioning that these sequences are a magnitude longer than the Hopkins 155 [22] and the Freiburg-Berkeley [16] dataset with an average length of 30 and 245 frames, respectively. As ground truth, we manually annotated a dense segmentation of every 250th frame as displayed in Fig. 4. Trajectories are obtained as proposed in [3]. Fig. 2 shows the average length of the trajectories for 10 sequences used in this study. As it can be seen, due to occlusions, fast and complicated motion patterns, the trajectories last just for 96.5 frames in average.

Segmentation Fast and complicated motions render the segmentation and tracking difficult. To illustrate this, consider Fig. 5, that shows the results obtained with the unsupervised segmentation of Brox and Malik [3]. They follow a similar procedure for obtaining trajectories, but with unsupervised spectral clustering on an affinity matrix in order to group trajectories. Compared to the ground truth (Fig. 4), it can be seen that the overall segmentation is not reliable and only very distinct motions could be separated.

Their approach can hardly be blamed for such failure. Babies exhibit rather erratic limb motions, such that points on the same limb nevertheless move quite differently. Motion *per se*, i.e. without further prior knowledge, is hardly strong enough a cue to support a correct segmentation. To supply the segmentation algorithm with such prior information, we label some trajectories for each segment in frames 1 and 500, frames 250 and 750 are used for the evaluation. As it can be observed from Fig. 3, on average only 5% of the trajectories are initially labeled. Compared to the full annotation needed in current practice, this effort is negligible. As it is visible in Fig. 6 our approach infers the remaining 95% of the labels. A qualitative comparison between our segmentation result and the ground truth segmentation (Fig. 4) indicates a high precision.

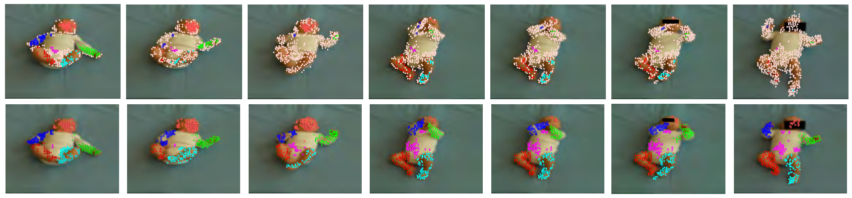

Fig. 6. Seq. 1 segmentation results for frames 1, 50, 200, 300, 650, 700, 800 and 950 from left to right. The upper row shows the results for the baseline where no segmentation method is applied, and the lower row is the results for the proposed method. Frames 800 and 950 have been anonymized after the segmentation.

Fig. 7 shows a quantitative evaluation where we calculate the ratio of correctly labeled trajectories for three cases: in the first one we manually labeled just the first frame, in the second case two frames (1 and 500). In addition, the result when using the prior labels without transduction are provided as baseline. As it can be observed, there is a substantial gain, which is necessary for our application. With the additional prior knowledge, for most of the sequences the segmentation is very precise, except for Seq. 5 where we deal with very complicated motions, complete occlusions and the body rolling onto the sides. Since we focus on high precision, for the following tracking experiment we used two labeled frames.

Occlusion is a longstanding problem in motion segmentation. Frames 650–800 in Fig. 6 show a case of severe occlusion where the head is occluded by both hands. As it can be seen, the segmentation remains correct and in frame 950, new trajectories in the occluded area on the head are labeled correctly. Partial occlusion is less of a problem for our proposed method: there are some trajectories left that can still stand in for the terminated ones. These are joined by novel trajectories upon the reappearance of the previously occluded region. In case of a complete occlusion, trajectories could be linked to each other again by providing further manual labeling or by high dissimilarity to all other motion segments.

Tracking Although we are tracking body parts and therefore a comparison to pose estimation methods seems reasonable, we compare our tracker performance with a tracking algorithm because pose estimation methods have skeleton constraints as additional prior while our method has the same input as a tracker. In order to study the performance of our tracking method, it is compared with the performance of TLD [7] as a representative state-of-the-art tracker. It tracks an object in a 3-step procedure, of tracking, learning, and detection. The tracker follows the object frame by frame, the detector localizes all appearances that have been observed so far, and finally, their *P-N learning* method is applied to estimate the detector error and update it for future use.

Since TLD is a single object tracker, we run it for each body part separately. In some frames TLD falsely does not report a location. To penalize this, we assign these frames the highest error of this sequence. In Fig. 8, the sequence-averaged tracking error of the different body parts is plotted. We evaluate the tracking errors in terms of the Euclidean distance to the ground truth points that are determined manually. As can be observed,

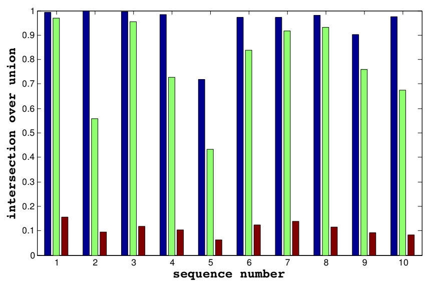

Fig. 7. The GT intersection over union for different sequences. Given are the results for the segmentation of trajectories with one manually labeled frame in green, two manually labeled frames in blue and for the baseline in red.

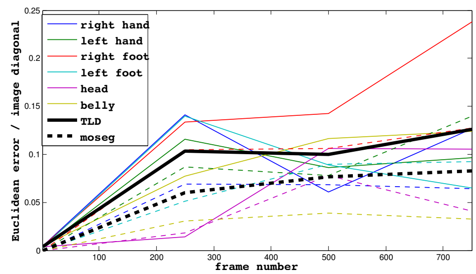

Fig. 8. The Euclidean tracking error of TLD (straight) and our motion-segmentation approach (dashed) over the frames in a sequence, averaged over all 10 sequences. The colored thin curves denote the errors of the different limbs. The averages over all limbs are displayed in black and bold.

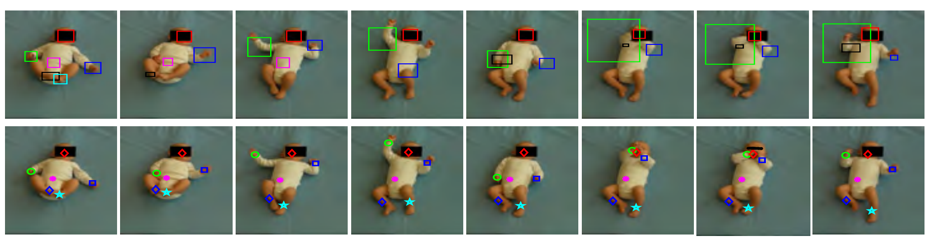

Fig. 9. Seq. 1 tracking results of TLD (upper row) and of proposed method (lower row) for frames 1, 50, 250, 350, 450, 650, 750 and 950 from left to right.

our approach is considerably more precise for this task. The main problem with TLD is that it cannot find the object for many frames. The average number of frames without tracking results over all sequences and all body parts for TLD is 12.46%, while it is 0.38% for our proposed method which only once looses a body part: in Seq. 5, where the baby rolls onto its right side and the right foot is completely occluded.

The upper row of Fig. 9 shows a qualitative result of TLD. The tracker lost the left foot in frame 50. A bit later, the same happens to the right foot and in frame 450 TLD redetects it wrongly at the right hand. Similar problems regularly occur for the other body parts, hence the performance is insufficient for our task. The results of our proposed tracker are shown in the bottom row of Fig. 9. All body parts are tracked effectively in all the frames, without drifting or part loss. Occlusions (frames 650–750) were dealt with well.

4.2 Comparison with Standard Benchmarks

In this section we challenge our segmentation method with different subjects in order to investigate its generality. To do so, the three video sequences *cats02*, *cats04* and *ducks01* of the *Freiburg-Berkeley* data set [3,16] are considered, in which we deal a)

b)

c)

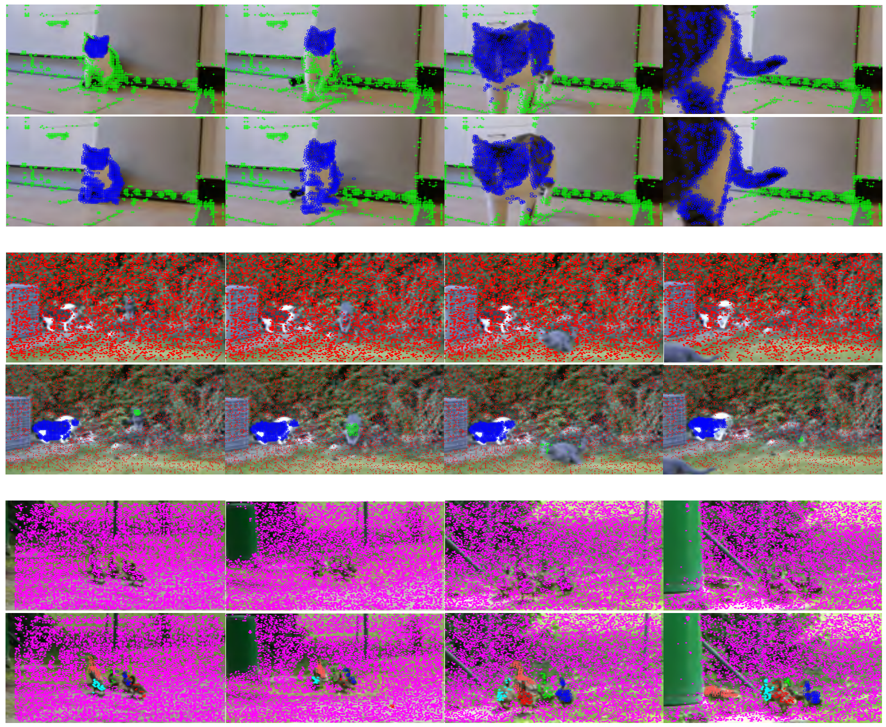

Fig. 10. Segmentation results for [3] (upper rows) and our proposed method (lower rows) for frames (a) 1, 70, 90, 110 of *cats02*, (b) 1, 40, 50, 70 of *cats04*, and (c) 1, 100, 300, 380 of *ducks01*. For the sake of visibility, the background trajectories are thinned out in *cats04*.

with occlusion, disocclusion, camera motion, fast motion and low texture objects. The segmentation results of our proposed method (initial labels in frame 50 in *cats02* and *cats04*, and 1 and 200 in *ducks01*) as well as those of Brox and Malik [3] are displayed in Fig. 10. As the results show, [3] only distinguishes very different motions from each other. This is why no object is detected in Seq. *cats04* and *ducks01*. On the other hand, our segmentation method performs reliably: in Seq. *cats02* except very small parts of the legs in a short period of the video, the cat is correctly segmented from the background. Seq. *cats04* shows a case where one of the cats has very low texture as well as fast motions, however the segmentation results are mostly correct. Finally, Seq. *ducks01* shows multiple similar objects that move next to each other, the segmentation task has become even more challenging because of occlusion, disocclusion and exit of one of the ducks. Despite all these, our method managed to segment all the objects correctly through the whole shot.

Fig. 11 shows a quantitative comparison between the performance of our proposed method and results of the baseline from the initial labels, where no segmentation algorithm is applied. As it can be seen, the proposed method has managed to segment the objects with high precision (96% on average).

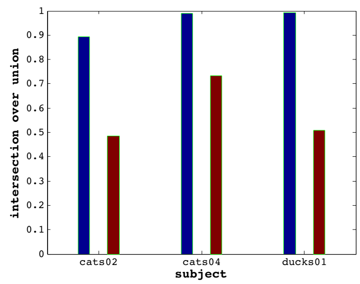

Fig. 11. The GT intersection over union for the proposed method in blue and the baseline in red with no segmentation applied on trajectories.

Although our method performs reliably, it could suffer from some points. First, the segmentation task depends on the distance between trajectories, and since our distance measure (6) only allows for the verification of translational model, we might have problems with segmenting other models of motion. For example, the legs of the cat in Seq. *cats02* are wrongly segmented because we deal with fast scaling. The second problem could arise from optical flow or trajectory inaccuracy. Although we used one of the promising optical flow methods, we could still have problem in situations with fast motion and low texture. This could be visualized in the last image of Seq. *cats04* where the low texture cat dose a fast jump and trajectories are wrongly developed.

4.3 Performance on Cerebral Palsy detection

In [18], the proposed method is applied on a set of 82 videos of infants with the same set-up as mentioned in early this section. For each of the infants six trajectories representing motions of different body parts are developed, then a set of features are extracted, and finally, the classification results are compared with those obtained by electromagnetic sensors. The comparison shows the advantage of our method.

5 Conclusions

In this paper, we dealt with segmenting and tracking the body limbs of infants in order to provide an analytical tool for clinical diagnosis. The sequences pose multiple problems such as parts having similar appearances, moving in complex ways, and being regularly occluded. Moreover, these parts need to be segmented and tracked with high precision. In our evaluation, the state-of-the-art trackers and motion segmentation algorithms had severe problems with these videos. The manual introduction of prior knowledge appeared to be mandatory at that point, but of course the overhead needed to be kept at a minimum. Therefore, a framework was designed that allows prior knowledge to be integrated into motion segmentation. We proposed a novel energy-minimization-based motion segmentation algorithm. Weak manual annotation came out to suffice to thereupon handle most of the videos automatically. A simple tracker, built on top of the motion segmentation yielded results with sufficient quality. Our experiments showed that our new approach outperforms current tracking and motion segmentation approaches.

References

1. Adde, L., Helbostad, J.L., Jensenius, A.R., Taraldsen, G., Grunewaldt, K.H., Støen, R.: 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) 1
2. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. Intl. Journal of Computer Vision 70(2), 109–131 (2006) 2, 3
3. Brox, T., Malik, J.: Object segmentation by long term analysis of point trajectories. In: ECCV. pp. 282–295 (Sep 2010) 2, 3, 4, 6, 8, 9
4. Cheriyadat, A.M., Radke, R.J.: Non-negative matrix factorization of partial track data for motion segmentation. In: ICCV. pp. 865–872 (Oct 2009) 2
5. Dragon, R., Rosenhahn, B., Ostermann, J.: Multi-scale clustering of frame-to-frame correspondences for motion segmentation. In: ECCV (Oct 2012) 2
6. Elhamifar, E., Vidal, R.: Sparse subspace clustering. In: CVPR. pp. 2790–2797 (2009) 2
7. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. TPAMI 34(7), 1409–1422 (2012) 2, 5, 7
8. Kanemaru, N., Watanabe, H., Kihara, H., Nakano, H., Takaya, R., Nakamura, T., Nakano, J., Taga, G., Konishi, Y.: 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) 2
9. Köktas, N.S., Duin, R.P.W.: Statistical analysis of gait data to assist clinical decision making. In: Medical Content-Based Retrieval for Clinical Decision Support. pp. 61–68 (2010) 1
10. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. Pattern Analysis and Machine Intelligence, IEEE Transactions on 28(10), 1568–1583 (2006) 3
11. Lezama, J., Alahari, K., Sivic, J., Laptev, I.: Track to the future: Spatio-temporal video segmentation with long-range motion cues. In: CVPR. pp. 3369–3376 (2011) 2
12. Li, Z., Guo, J., Cheong, L.F., Zhou, S.Z.: Perspective motion segmentation via collaborative clustering. In: ICCV (2013) 2
13. 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) 1
14. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: NIPS. vol. 14, pp. 849–856 (2002) 2
15. Ochs, P., Brox, T.: Higher order models and spectral clustering. In: CVPR (2012) 2
16. Ochs, P., Malik, J., Brox, T.: Segmentation of moving objects by long term video analysis. TPAMI (2013) 6, 8
17. Punithakumar, K., Ayed, I.B., Islam, A., Goela, A., Li, S.: Regional heart motion abnormality detection via multiview fusion. In: MICCAI. pp. 527–534 (2012) 1
18. Rahmati, H., Aamo, O.M., Stavdahl, Ø., Dragon, R., Adde, L.: Video-based early cerebral palsy prediction using motion segmentation. In: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE (2014) 10
19. Shi, F., Zhou, Z., Xiao, J., Wu, W.: Robust trajectory clustering for motion segmentation. In: ICCV (2013) 2
20. Stahl, A., Schellewald, C., Stavdahl, Ø., Aamo, O.M., Adde, L., Kirkerod, H.: An optical flow-based method to predict infantile cerebral palsy. Neural Systems and Rehabilitation Engineering, IEEE Transactions on 20(4), 605–614 (2012) 1
21. Sun, D., Sudderth, E.B., Black, M.J.: Layered segmentation and optical flow estimation over time. In: CVPR (2012) 2

22. Tron, R., Vidal, R.: A benchmark for the comparison of 3D motion segmentation algorithms. In: CVPR (2007) 6
23. Vidal, R., Hartley, R.: Motion segmentation with missing data using powerfactorization and GPCA. In: CVPR. pp. 310–316 (2004) 2
24. Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. Advances in Neural Information Processing Systems 16 (2004) 2