UNIFIED PROBABILISTIC FRAMEWORK FOR SIMULTANEOUS DETECTION AND TRACKING OF MULTIPLE OBJECTS WITH APPLICATION TO BIO-IMAGE SEQUENCES

S. Karthikeyan⋆ Diana Delibaltov⋆ Utkarsh Gaur⋆ Mei Jiang† David Williams† B.S. Manjunath⋆

⋆ Department of ECE, University of California Santa Barbara
† Jules Stein Eye Institute, University of California Los Angeles
⋆{karthikeyan,diana,utkarsh,manj}@ece.ucsb.edu †{jiangm,dswilliams}@ucla.edu

ABSTRACT

We present a detection based tracking algorithm for tracking melanosomes (organelles containing melanin) in time lapse image sequences imaged using bright field microscopy. Due to heavy imaging noise detecting all the melanosomes accurately in every frame is difficult. Therefore, two sets of imperfect detections are used in a unified probabilistic approach to simultaneously perform melanosome detection and tracking. We propose a novel iterative algorithm which jointly estimates the optimal set of detections and track results in every iteration from the previous tracks and detections. Our algorithm obtains significantly better tracking results than the state of the art tracking-by-detection algorithm.

Index Terms— Simultaneous Detections and Tracking, Hungarian algorithm

1. INTRODUCTION

Melanosomes are organelle containing melanin, the most important light absorbing pigment in the animal kingdom. Melanosomes in the Retinal Pigment Epithelium (RPE) functions in the screening of light and their subcellular localization is determined by molecular motor transportation. Understanding the dynamics of melanosomes will provide general insights into organelle dynamics in eukaryotic cells. Analysis of melanosomes can give insight to the cause of genetic diseases [1]. Melanosomes are difficult to detect due to background clutter and imaging noise. Therefore, we formulate the melanosome tracking problem as a multi-object tracking problem which is robust to missing detections.

Multiple object tracking is an important and challenging problem in computer vision and bio-image analysis [2, 3, 4, 5]. The common paradigm towards multiple object tracking has shifted towards a tracking-by-detection approach. Here, tracking is achieved by associating reliable detection responses over multiple frames. The primary drawback of this approach is it assumes the existence of accurate and stable detectors, which might not be feasible in our melanosome tracking problem. It can be easily understood that detections help build effective tracks and conversely, tracking results provide valuable information about missing detections and false positive detections. In accordance with this notion, we propose to use imperfect melanosome detections to iteratively improve subsequent tracks in a novel probabilistic simultaneous detection and tracking framework.

Association based multiple object tracking (tracking by detection) is a well studied problem in computer vision. Traditional approaches such as [6, 7] tackle this problem until enough evidence is obtained to substantiate the grouping. However, task complexity limits these optimization approaches to consider only few time steps. The Hungarian algorithm [8] for associating detections is a widely used approach in several recent works. In [4], Wu et al. define an object affinity based on color, position and size and use the Hungarian algorithm to obtain the associations across neighboring frames. In [9], Stauffer et al. propose a tracklet[1] association algorithm with initialization and termination costs for modeling their source/sink analysis problem. In [3], Huang et al. propose a three level hierarchical association algorithm which uses conservative estimation of tracklets. Their algorithm also models the false alarm hypothesis of the tracklets. These approaches require robust detections which might not be always practically feasible.

To overcome this problem, some approaches make use of intermediate tracking results to estimate reliable useful detections which can enhance the final tracking performance. In this regard, Liebe et al. [10] propose a coupled detection and tracking algorithm in . They use a formulation based on a quadratic boolean program and obtain the detections and tracks using an EM style optimization algorithm. This formulation requires approximating the optimization to an EM type estimation where convergence is not always guaranteed, as opposed to the novel joint estimation framework proposed in this work. Also, the detection estimates in our approach are based on global data compared to the incremental frame-level estimates of [10]. Other approaches such as [11, 12] also tackle the problem of simultaneous detection and tracking but lack a unified optimization framework. There are a few approaches towards tracking multiple cells in bio-image sequences [5, 13], but they again rely on accurate detections.

Therefore to overcome these deficiencies in prior approaches, we propose a novel framework for simultaneous detection and tracking. Our contributions include:

- A principled probabilistic framework for simultaneous detection and tracking which combines the modeling of detections and associated tracks together.
- A novel optimization algorithm which jointly estimates the optimal detections and associations based on the Hungarian algorithm.

2. OUR APPROACH

In this work, we propose a Bayesian formulation for simultaneous detection and tracking. Given a set of detections for all the frames in an image sequence $\mathcal{D}^i$, for the i^{th} iteration, we first build tracklets, $\mathcal{T}^i$, using a conservative tracklet building algorithm. This implies no false associations are made across the detections in this low-level association step. Given these tracklets, we want to identify the best possible combination of tracks ($\mathcal{S}^i$) and detections ($\mathcal{D}^{i+1}$) which

[1]Tracklets are high confidence object associations usually across a few frames, which can be combined to form tracks

serve as input for the next iteration. This can be mathematically written as

$$\{\mathcal{D}^i_{opt}, \mathcal{S}^i_{opt}\} = \arg\max_{\mathcal{D}^{i+1}, \mathcal{S}^i} \mathcal{P}(\mathcal{D}^{i+1}, \mathcal{S}^i | \mathcal{D}^i, \mathcal{T}^i, \mathcal{I}) \quad (1)$$

where $\mathcal{I}$ refers to the entire image sequence.

Using chain rule, we split this into two terms

$$\mathcal{P}(\mathcal{D}^{i+1}, \mathcal{S}^i | \mathcal{D}^i, \mathcal{T}^i, \mathcal{I}) = \mathcal{P}(\mathcal{S}^i | \mathcal{D}^i, \mathcal{T}^i, \mathcal{I}) \mathcal{P}(\mathcal{D}^{i+1} | \mathcal{D}^i, \mathcal{T}^i, \mathcal{I}, \mathcal{S}^i) \quad (2)$$

The first term represents the association model. Here, the information provided by the tracklets $\mathcal{T}^i$ is sufficient for the estimation of the tracks, $\mathcal{S}^i$. We note that the previous detections ($\mathcal{D}^i$) and the image information ($\mathcal{I}$) is not useful or rather redundant for the association problem. In second term we estimate the new detections for the next iteration, $\mathcal{D}^{i+1}$. We assume the $\mathcal{D}^{i+1}$ only depends on the track identity ($\mathcal{T}^i$) and the image information. Therefore (2) can be re-written as

$$\mathcal{P}(\mathcal{D}^{i+1}, \mathcal{S}^i | \mathcal{D}^i, \mathcal{T}^i, \mathcal{I}) = \mathcal{P}(\mathcal{S}^i | \mathcal{T}^i) \mathcal{P}(\mathcal{D}^{i+1} | \mathcal{I}, \mathcal{S}^i) \quad (3)$$

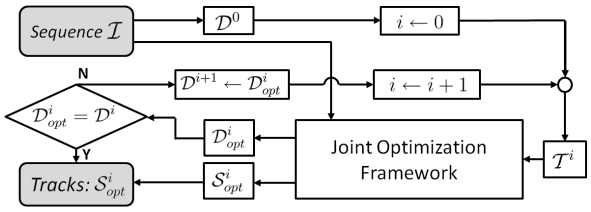

Fig. 1. Our optimization framework uses the tracklets $\mathcal{T}^i$ (obtained from current detections) and image information $\mathcal{I}$ to jointly optimize for improved detections D^i_{opt} and tracks S^i_{opt} (3). If there is no change in the detections ($\mathcal{D}^i_{opt} = \mathcal{D}^i$), we output the final tracks S^i, otherwise, the improved detections are used for the next iteration until convergence.

The overall flow of our approach is shown in Figure 1. The following section describes the models for the association and new detection estimation terms.

2.1. Association model

Several association models have been proposed for multiple object tracking as reviewed in Section 1. In our work we use an approach similar to [3] in which the authors jointly model the tracklet associations with the false alarm hypothesis. Let the individual tracklets $\mathcal{T}^i$ in the i^{th} iteration be denoted by $\{T^i_1, T^i_2 ... T^i_{N^i}\}$. Similarly let the tracks of $\mathcal{S}^i$ be denoted by $\{S^i_1, S^i_2 ... S^i_{M^i}\}$.

Now, the association term is decomposed as

$$\begin{aligned}\mathcal{P}(\mathcal{S}^i | \mathcal{T}^i) &= \mathcal{P}(\mathcal{T}^i | \mathcal{S}^i) \mathcal{P}(\mathcal{S}^i) \\ &= \prod_{T^i_k \in \mathcal{T}^i} \mathcal{P}(T^i_k | \mathcal{S}^i) \prod_{S^i_l \in \mathcal{S}_i} \mathcal{P}(S^i_l)\end{aligned} \quad (4)$$

Here we assume the likelihoods of the input tracklets are conditionally independent given $\mathcal{S}^i$ and the tracks $\{S^i_l\}$ are independent of each other.

A Bernoulli distribution is used to model the false alarm hypothesis of the tracklet using the detector precision denoted by β. Therefore, the likelihood of a tracklet is defined as

$$\mathcal{P}(T^i_k | \mathcal{S}_i) = \begin{cases} \mathcal{P}_+(T^i_k) = \beta^{|T^i_k|} & \text{if } \exists S^i_l \in \mathcal{S}^i, T^i_k \in S^i_l \\ \mathcal{P}_-(T_k{}^i) = (1-\beta)^{|T^i_k|} & \text{if } \forall S^i_l \in \mathcal{S}^i, T^i_k \notin S^i_l \end{cases} \quad (5)$$

where $|T^i_k|$ is the number of detections in T^i_k, and $\mathcal{P}_+(T^i_k)$ and $\mathcal{P}_-(T^i_k)$ are the likelihoods of T^i_k being a true detection and a false alarm respectively.

The tracklet association priors in (4) are modeled as Markov Chains.

$$\mathcal{P}(S^i_l) = \mathcal{P}_{link}(T^i_{k_1} | T^i_{k_0}) ... \mathcal{P}_{link}(T^i_{k_{p_k}} | T^i_{k_{p_k-1}}) \quad (6)$$

where p_k refers to the number of tracklets associated to form the track S_k. Basically, the association prior is a product of transition terms representing linkage probabilities between tracklets.

We note that T^i_k cannot belong to more than one S^i_l. Thus (4) is rewritten as the following by inserting $\mathcal{P}_+(T^i_k)$ into its corresponding chain.

$$\begin{aligned}\mathcal{P}(\mathcal{S}^i | \mathcal{T}^i) = \prod_{\forall S^i_l \in \mathcal{S}^i, T^i_k \notin S^i_l} \mathcal{P}_-(T^i_k) \prod_{S^i_l \in \mathcal{S}^i} \Big[\mathcal{P}_+(T^i_{k_0}) \\ \mathcal{P}_{link}(T^i_{k_1} | T^i_{k_0}) .. \mathcal{P}_{link}(T^i_{k_{p_k}} | T^i_{k_{p_k-1}}) \mathcal{P}_+(T^i_{k_{p_k}}) \Big]\end{aligned} \quad (7)$$

2.2. Detection estimation model

In this section we propose a technique to model $P(\mathcal{D}^{i+1} | \mathcal{I}, \mathcal{S}^i)$ in (3). Our approach jointly combines both the image sequence information $\mathcal{I}$ and the track information $\mathcal{S}^i$ to obtain an optimal set of detections $\mathcal{D}^{i+1}$. Here we assume that all the image information for detection purposes is captured by the set of all detections in all frames in the time sequence, denoted by $\mathcal{D}^{all}$ and $\mathcal{D}^i \subset \mathcal{D}^{all}\ \forall i$ (refer to Figure 2). So, our aim is to estimate the best possible detections from $\mathcal{D}^{all}$ which complements the track $\mathcal{S}^i$ estimation problem. Therefore, the overall distribution is factored into different individual tracks (S^i_k). Missing detections in a track are interpolated from the other detections. Let all these interpolated detections of the k^{th} track be denoted by $D^i_{(inter,k)}$. These interpolated detections in every track assimilate all the useful information which each track provides for estimating $\mathcal{D}^{i+1}$. These steps can be mathematically written as

$$\begin{aligned}\mathcal{P}(\mathcal{D}^{i+1} | \mathcal{I}, \mathcal{S}^i) &= \mathcal{P}(\mathcal{D}^{i+1} | \mathcal{D}^{all}, \mathcal{S}^i) \\ &= \prod_k \mathcal{P}(D^{i+1}_k | \mathcal{D}^{all}, D^i_{(inter,k)})\end{aligned} \quad (8)$$

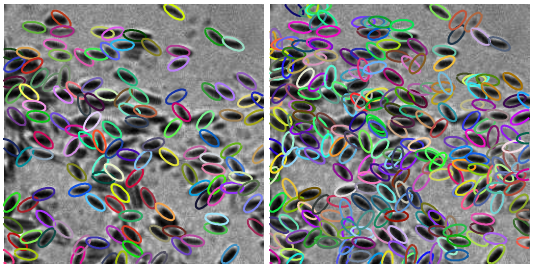

Fig. 2. Example of initial detections $\mathcal{D}^0$ (left) and $\mathcal{D}^{all}$(right) in a frame. We can observe $\mathcal{D}^{all}$ consists of several false positive detections but no false negatives (best viewed in color)

Now, the individual detections in the k^{th} track of the i^{th} iteration are denoted by $\{d^i_{k,j}\}_{j=1..n_k}$ where n_k is the total number of detections in the k^{th} track. We assume the interpolation function has the property that only the missing detections are different from $\mathcal{D}^i$. The individual interpolated detections are denoted by $d^i_{inter_{(k,j)}}$. As the detections are independently obtained we have

$$P(\mathcal{D}^{i+1}|\mathcal{I},\mathcal{S}^i) = \prod_k \prod_j P(d_{k,j}^{i+1}|\mathcal{D}^{all}, D_{(inter,k)}^i) \quad (9)$$

Our focus is to model the estimation of $d_{k,j}^{i+1} \ \ \forall j,k$ which is a subset of $\mathcal{D}^{all}$. We propose a gaussian density model for $\mathcal{P}(d_{(k,j)}^{i+1}|\mathcal{D}^{all}, D_{(inter,k)}^i)$ with the mean of the Gaussian distribution being $d_{inter_{(k,j)}}^i$. Therefore for 2-D detection information denoting the center of the detection (9) becomes,

$$P(\mathcal{D}^{i+1}|\mathcal{I},\mathcal{S}^i) = \prod_k \prod_j \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|d_{k,j}^{i+1} - d_{inter_{(k,j)}}^i\|^2}{2\sigma^2}\right) \quad (10)$$

where $d_{k,j}^{i+1} \in D^{all}$, and should be present in the same frame and $\sigma^2\mathbf{I}$ is the covariance matrix [2].

2.3. Inference

As we need to maximize (3), first we convert it into a cost function by taking negative logarithms. The overall cost which we need to optimize is the sum of negative logarithms of (7) and (10). In this Section, we propose an optimization technique which jointly estimates $\mathcal{S}^i$ and $\mathcal{D}^{i+1}$. To motivate this scheme we first need to understand how the individual cost functions can be optimized separately.

2.3.1. Track inference given detections

The cost described in (7) can be optimized by the Hungarian algorithm over tracklets similar to the one proposed in [3]. Here, a probabilistic cost is formulated to associate any two tracklets and to denote a tracklet as a false positive. In brief, to associate n tracklets a $n \times n$ cost matrix is built with the non-diagonal entries denoting the tracklet association costs and the diagonals are the false positive costs. The optimal tracklet associations and false positive set which minimize the cost globally is obtained by the Hungarian assignment on this cost matrix.

2.3.2. Detection inference given tracks

To obtain the optimal set of detections (10) given the tracks, we first observe that for all detections, if $d_{inter_{(k,j)}}^i = d_{k,j}^i$ then $d_{k,j}^{i+1} = d_{k,j}^i$. This implies the detections in the previous iteration which are not false positives are preserved in the next iteration. So, only the missing detections have to be optimally chosen from $\mathcal{D}^{all}$. Now, we intent to optimally map the interpolated detections (from the missing positions in the tracks) to detections in $\mathcal{D}^{all}$. This can be optimized by constructing a cost matrix $\mathbf{C}_f$ at every frame f with the missing detections as the rows and the elements in $\mathcal{D}^{all} \setminus \mathcal{D}^i$ in frame f as the columns. Let the detection interpolation corresponding to the i^{th} row be d_i^{fr} and the detection in $\mathcal{D}^{all}$ corresponding to j^{th} column be d_j^{fc}, now the cost is written as

$$\mathbf{C}_f(i,j) = \frac{\|d_i^{fr} - d_i^{fc}\|^2}{\sigma^2} \quad (11)$$

Subsequently, we use the Hungarian algorithm on $\mathbf{C}_f$ to map the interpolated detections to elements in $\mathcal{D}^{all}$ at every frame. Thus, we obtain the optimal set of missing detections which minimizes (10).

[2]The detection responses we typically work with are the center coordinates of the detections and therefore assuming the covariance matrix as $\sigma^2\mathbf{I}$ is justified.

2.3.3. Joint track and detection inference

One of the possible ways to optimize (1) is an EM style approach to iteratively infer the detections given tracks (Section 2.3.2) and tracks given detections (Section 2.3.1). But, this does not guarantee the cost function in (1) is globally optimized. Our aim is to propose an inference framework which jointly optimizes both the tracks and detections and hence globally optimizes (1).

For our algorithm to work, we have an additional constraint on interpolation function (to obtain $D_{(inter,k)}^i$), that it only depends on the head and tail tracklet between which it is present. As noted earlier, [3] solves the tracklet association problem by a Hungarian algorithm over all possible tracklet associations. Under this interpolation assumption, the modified tracklet association cost which optimizes (3) can be obtained as the sum of the tracklet association cost in Section 2.3.1 and the missing detection cost similar to Section 2.3.2. We note that the missing detections cost is affiliated to every possible tracklet pair association and is denoted by a $n \times n$ matrix $\mathbf{C}_{Det}$ where n is the total number of tracklets.

Here, we do not have the track information as in Section 2.3.2. For every missing detection in frame f (while associating any two tracklets), we initially tentatively assign it to the nearest detection in $\mathcal{D}_{all} \setminus \mathcal{D}^i$ in f, where every interpolated detection incurs a cost as in (11). Let this interpolated detection cost across any two tracklets T_p^i and T_q^i be denoted by $\mathbf{C}_{Det}(p,q)$ which is the sum of all the interpolation costs when associating the tracklets. Let the total number of tracklets be n. The overall cost to associate any two tracklets is a sum of the tracklet association cost and overall interpolated detections cost. Therefore, the joint cost matrix $\mathbf{C}_J$ of dimensions $n \times n$ to associate any two tracklets T_p^i and T_q^i is expressed as

$$\mathbf{C}_J(p,q) = \begin{cases} \ln \mathcal{P}_-(T_p^i) & \text{if } p = q \leq n \\ \ln \mathcal{P}_{link}(T_q^i|T_p^i) + 0.5[\ln \mathcal{P}_+(T_p^i) + \ln \mathcal{P}_+(T_q^i)] & \\ +\mathbf{C}_{Det}(p,q) & \text{if } p,q \leq n \text{ and } p \neq q \\ -\infty & \text{otherwise} \end{cases} \quad (12)$$

The optimal tracks are obtained by the Hungarian algorithm on $\mathbf{C}_J$ which assigns every row to a unique column. If a tracklet is assigned to itself, it is a false positive tracklet and will be removed from the successive iterations. We also note that every tracklet pair selected by the Hungarian also could pick a set of detections from $\mathcal{D}^{all}$ which are the missing detections. This solution is optimal if there are no missing detections common to different tracks. In case such conflict exists, we use the track information and recompute the detection assignment by solving a frame level Hungarian according to Section 2.3.2. Subsequently we build a new $\mathbf{C}_{Det}$ after resolving conflict. This is followed by updating $\mathbf{C}_J$ according to (12). When this approach converges and no such conflict exists the solution obtained $(\mathcal{D}^{i+1}, \mathcal{S}^i)$ globally optimizes the product of (7) and (10) which is (1).

In the next iteration we use D^{i+1} after removing the false positive tracklets, to obtain $(\mathcal{D}^{i+2}, \mathcal{S}^{i+1})$. We also note that the overall cost in every iteration decreases in practice as we remove the false positive tracklets and missing detections help in building longer tracklets thereby reducing the number of tracklets and association costs involved.

3. EXPERIMENTS

Dataset: Melanosomes are imaged by Bright-Field microscopy because of their dark pigment. Our datasets were obtained from mouse retina. In our experiment we examined two sequences, Seq1 ($\sim$120 melanosomes) and Seq2 ($\sim$150 melanosomes) of 50 frames each.

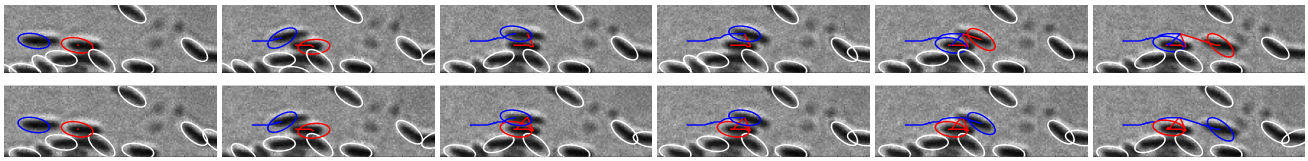

Fig. 3. Our approach (bottom row) tracks better when some detections are missed compared to the baseline approach (top row). We show tracking results on frames 2,14,21,24,25,37 for both the approaches. The missing detections in frames 21-24 causes incorrect association in the baseline approach

The melanosomes predominantly had an ellipital shape and move in rapid bursts.The dataset also contained severe background clutter.

3.1. Implementation details

Our aim is to track the melanosomes in the image sequences. Each melanosome is detected by a elliptical template based detector at different angles. An example detection to initialize the tracking D^0 and D^{all} is shown in Figure 2. We use the ellipse center coordinates to indicate the detections as the angles are not always reliable.

Tracklet Construction: After the detections, conservative tracklets are obtained similar to [3] by associations over successive frames. Here, we first define linkage between detections (d_i) and (d_j) in successive frames using affinity based on ellipse centers, $A(d_i, d_j) = e^{-\frac{||d_i - d_j||^2}{t}}$. To prevent unsafe tracklets, we associate only if the affinity between a pair of detections is significantly higher than other conflicting pairs.

Track and Detection Estimation parameters: We associate pairs of tracklets (T_p, T_q) to obtain $\mathbf{C}_J(p, q)$ and our associations are valid only if the tail (end frame) of tracklet T_p is greater than the head (start frame) of tracklet T_q. The tracklet linkage probability across these two tracklets $P_{link}(T_q|T_p)$ is defined as a product of two terms, motion $P_m(\mathcal{T}_q|\mathcal{T}_p)$ and time, $P_{time}(\mathcal{T}_q|\mathcal{T}_p)$. Here, the motion term uses a linear model to predict the position of the head of the tracklet T_q from the tail of T_p. This prediction error (e) is used to model the motion affinity as gaussian distribution, $P(T_q|T_p) \sim \mathcal{N}(0, \sigma_e^2 \mathbf{I}_{2\times 2})|_e$. To model $P_{time}(\mathcal{T}_q|\mathcal{T}_p)$ we calculate the number of frames between the two tracklets (n_t) (tail of T_p and head of T_q). If $n_t < n_0$, $P_{time}(\mathcal{T}_q|\mathcal{T}_p) = \lambda e^{-\lambda n_t}$, otherwise it is 0. In our experiments we set $\sigma_e^2 = 12$, $\lambda = 5$. The only parameter in the detection estimation part is $\sigma = 5$ in (11).

3.2. Results

We test our joint detection and tracking approach on two sequences, Seq1 ($\sim$120 melanosomes) and Seq2 ($\sim$150 melanosomes) of 50 frames each. We have the ground truth (center positions) for 10 prominently moving melanosomes in each sequence. We compare our joint detection and tracking approach to the baseline approach [3] which only infers tracks from the detections. In [3] the missing detections are interpolated only using the track information. Our approach outperforms the baseline approach [3] as shown in Figure 4 where the errors in individual melanosomes are highlighted. Also, Figure 3 visually shows instances where the baseline approach can fail when some detections are missed. However, our approach jointly identifies the missing detections and tracks and therefore achieves better performance.

Acknowledgments We would like to thank NSF award III-0808772 and NSF OIA 0941717 for funding our research.

4. CONCLUSION

In this paper we proposed a probabilistic framework for simultaneous detection and tracking of multiple objects. This approach jointly optimizes for the detections and tracks in every iteration. We obtain promising results on melanosome tracking problem. Our algorithm is fast and it converges in about 26 seconds on a 2.4 GHz machine to track about 120 objects simultaneously after the detections are obtained in all frames. We outperform the baseline method in [3] which only optimizes the track associations given the detections.

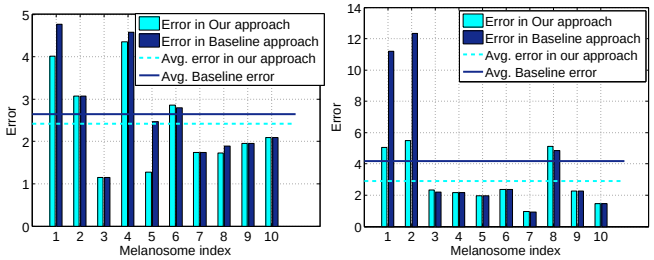

Fig. 4. Comparison of the root mean square error of our algorithm with the baseline approach of both Sequence1 (left) and Sequence2 (right).

5. REFERENCES

[1] D. Gibbs et al., "Retinal pigment epithelium defects in humans and mice with mutations in myo7a: imaging melanosome-specific autofluorescence," *Investigative ophthalmology & visual science*, 2009.

[2] M. Andriluka et al., "People-tracking-by-detection and people-detection-by-tracking," in *CVPR 2008*.

[3] C. Huang et al., "Robust object tracking by hierarchical association of detection responses," in *ECCV 2008*.

[4] B. Wu et al., "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors," *IJCV 2007*.

[5] K. Li et al., "Cell population tracking and lineage construction with spatiotemporal context," *Medical Image Analysis, 2008*.

[6] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control, 1979*.

[7] T. Fortmann et al., "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE Journal of Ocean Engineering, 1983*.

[8] H.W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly, 1955*.

[9] C. Stauffer, "Estimating tracking sources and sinks," in *CVPRW 2003*.

[10] B. Leibe et al., "Coupled detection and trajectory estimation for multi-object tracking," in *ICCV 2007*.

[11] R. Ilin and R.W. Deming, "Simultaneous detection and tracking of multiple objects in noisy and cluttered environment using maximum likelihood estimation framework," in *IEEE OCEANS 2010*.

[12] Y. Huang and X. Luo, "Simultaneous detection and tracking in airborne video," in *ICCTD 2009*. IEEE.

[13] R. Bise et al., "Reliable cell tracking by global data association," in *ISBI 2011*.