Structural-Flow Trajectories for Unravelling 3D Tubular Bundles

Katerina Fragkiadaki[1], Weiyu Zhang[1], Jianbo Shi[1], and Elena Bernardis[2]

[1] Department of Computer and Information Science, GRASP Laboratory, University of Pennsylvania, Philadelphia, PA 19104
[2] Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104
{katef,zhweiyu,jshi,elber}@seas.upenn.edu

Abstract. We cast segmentation of 3D tubular structures in a bundle as partitioning of *structural-flow trajectories*. Traditional 3D segmentation algorithms aggregate local pixel correlations incrementally along a 3D stack. In contrast, structural-flow trajectories establish *long range* pixel correspondences and their affinities propagate grouping cues across the entire volume simultaneously, from informative to non-informative places. Segmentation by trajectory clustring recovers from persistent ambiguities caused by faint boundaries or low contrast, common in medical images. Trajectories are computed by linking successive registration fields, each one registering pairs of consecutive slices of the 3D stack. We show our method effectively unravels densely packed tubular structures, without any supervision or 3D shape priors, outperforming previous 2D and 3D segmentation algorithms.

Keywords: 3D tubular structures, trajectory clustering, morphological segmentation.

1 Introduction

Automatic segmentation of tubular structures is of vital importance for various fields of medical research. An example of such tubular structures are the organ-pipe-like

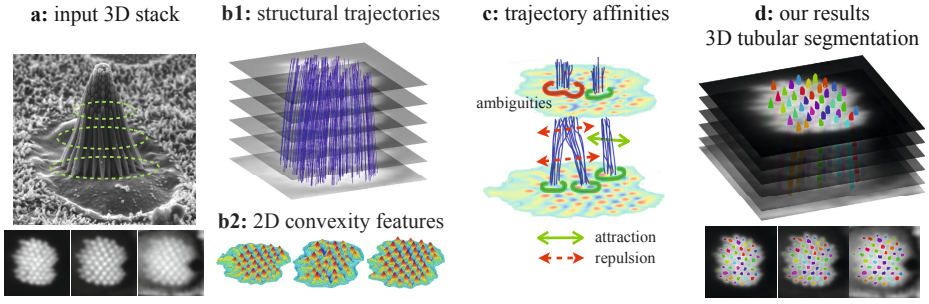

Fig. 1. Method overview. (a) A stack of 2D images of a tubular bundle. Image is courtesy of Medha Pathak and David Corey, Harvard. (b1) Structural-flow trajectories traversing the stack. (b2) 2D convexity cues. (c) Trajectory attractions and repulsions. (d) Resulting 3D segmentation.

N. Ayache et al. (Eds.): MICCAI 2012, Part III, LNCS 7512, pp. 631–638, 2012.

stereocilia bundles of the inner ear, depicted in Fig.1(a). Automatic segmentation of stereocilia in their fluorescent image stacks contributes to medical research on hearing [1].

There are two main lines of work that tackle segmentation of tubular forms: 1) Methods that compute a series of independent 2D segmentations [2–5] and then correspond them along the third dimension [6]. 2) Methods that segment directly in 3D, such as level sets, 3D watershed [4, 3, 7], region growing [8], or methods that employ 3D shape priors, often initialized via some type of user interaction [9–11]. In the former approaches, segmentation and correspondence do not interact with or benefit from each other, hence 2D segmentation mistakes often propagate to erroneous 3D correspondences. In the latter, local correlations along the third dimension are often aggregated in an incremental, feed-forward fashion. Consequently, close configurations between adjacent tubular structures that cause segmentation ambiguity to persist across multiple slices in the 3D image stack are hard to deal with.

Our main insight is that the topology of tubular structures, each with a corresponding one dimensional medial axis and a deforming continuum of 2D cross-sections along the axis direction, allows reliable registration of consecutive cross-sections. A condition for this is the medial axes directions to be non-parallel to the slicing direction. Linking of successive registration fields results in long range pixel correspondences in the 3D volume, which we call *structural-flow trajectories*. We segment densely packed tubular structures by partitioning structural-flow trajectories, as shown in Fig.1. Trajectory affinities are computed by marginalizing corresponding convexity-driven pixel affinities across trajectory lifespans (Fig.1(c)). They propagate grouping information along the 3D image stack, from informative to non-informative places and are robust to locally ambiguous grouping cues, often caused by closely attached tubular structures in a bundle. In this way, trajectory partitioning effectively unravels tubular structures automatically (Fig.1(d)), without 3D shape priors or user interactions.

We test our algorithm on segmenting stereocilia bundles of the inner ear in their fluorescent images. We significantly outperform various baseline segmentation algorithms that do not exploit long range structural information. To the best of our knowledge, we are the first to utilize structural trajectories for capturing long range structural correspondences between pixels at different depths of a 3D volume rather than temporal correspondences between pixels of consecutive frames in a video sequence [12].

2 Long Range Structural Correspondence

Consider two consecutive slices, $I^z(x, y)$ and $I^{z+1}(x, y)$, where $z \in \mathbb{Z}^+$ denotes the slice index from bottom to top of a 3D stack. We define (u, v) to be the deformation field that registers the two slices as the one minimizing intensity and gradient pixel matching scores:

$$\underset{u,v}{\text{minimize}} \quad \begin{aligned} &|I^{z+1}(x+u, y+v) - I^z(x, y)| \\ &+ |\nabla I^{z+1}(x+u, y+v) - \nabla I^z(x, y)| + |\nabla u| + |\nabla v|. \end{aligned} \tag{1}$$

The last two robust penalization terms on gradients of the deformation field u, v encourage smoothness in registration [13]. Such smoothness constraints allow registration to

be reliably computed even in places of ambiguous grouping cues (e.g. faint boundaries), by propagating registration information from reliable (gradient rich) pixel neighbours with peaked unary matching terms. We solve for (u, v) through a coarse to fine estimation scheme with successive linearisation of the intensity and gradient constancy constraints under the assumption of small displacements, as proposed in [13]. Dense slice sampling with respect to deformation along the medial axis of the tubular structures guarantees displacements to be small from slice to slice.

We define a *structure-flow trajectory* to be a sequence of (x, y, z) points:

$$\mathrm{tr}_i = \{(x_i^z,\ y_i^z,\ z),\ z \in \mathrm{Z}_i\}, \tag{2}$$

where Z_i is the set of slice indices in which trajectory tr_i is "alive". Trajectories are dense in space and capture slice-to-slice pixel correspondences, despite illumination changes or density variations of the 2D shapes between slices across the stack. We compute structural trajectories by following per slice registration fields computed from Eq. 2 between pairs of consecutive slices. A forward-backward check determines termination of a trajectory [14]. Thus, structural trajectories can have various lifespans and adapt to the varying lengths of the 3D tubular structures (e.g. stereocilia). We visualize structural trajectories in Fig.1(b1).

The notion of a pixel trajectory has been traditionally used to describe 2D projections of a single physical point in a video sequence [14]. In our case, the notion of a structural trajectory refers to a series of physical points geometrically related via successive registrations.

3 Constrained Segmentation with Structural Flow Trajectories

In 3D segmentation, local cues are often faint and unreliable. Such ambiguities appear in batches rather than randomly scattered along a 3D stack, since the configuration of 2D cross-sections of the tubular bundle cannot change drastically from one slice to another. We address persistency of cue ambiguity by formulating 3D segmentation as spectral partitioning of structural-flow trajectories. We estimate pixel pairwise relationships at each image slice based on local convexity cues proposed in [5]. Trajectory affinities marginalize corresponding pixel relationships. Thus, grouping cues are propagated from informative to non-informative slices and provide a consistent and well-informed segmentation throughout the whole 3D volume.

3.1 Per Image Grouping Cues

Consider pixel p_i and its neighbourhood $N_d(\mathrm{p}_i)$ of radius r_d, as shown in Fig.2. We define a peak neighbour p_a of p_i to be a pixel in $N_d(\mathrm{p}_i)$ that can be connected to p_i by a straight line of non-decreasing image intensities, of total intensity difference $\mathrm{S}(\mathrm{p}_i, \mathrm{p}_a) = I(\mathrm{p}_a) - I(\mathrm{p}_i)$. Let $f(\mathrm{p}_i)$ be the weighted average direction from p_i to its peak neighbours:

$$f(\mathrm{p}_i) \propto \sum_{\mathrm{p}_a \in \mathcal{P}(\mathrm{p}_i)} \mathrm{S}(\mathrm{p}_i, \mathrm{p}_a)(\mathbf{p}_a - \mathbf{p}_i), \qquad ||f(\mathrm{p}_i)||_2 = 1, \tag{3}$$

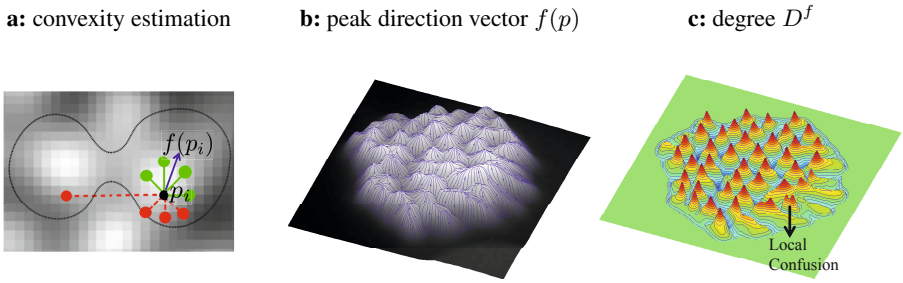

Fig. 2. 2D convexity cues. (a) Estimation of the peak vector $f(\mathrm{p}_i)$. (b) The peak vector field $f(\mathrm{p})$ points at each pixel to the closest highest intensity peak. (c) Degree image D^f. Valleys and peaks correspond to convex and non-convex regions in the original intensity image. Closely attached tubular structures in this slice, cause double valleys in D^f and confuse the corresponding pixel relationships.

where $\mathbf{p}_i$ denotes the 2D pixel coordinate of p_i and $\mathcal{P}(\mathrm{p}_i)$ the set of peak neighbors. We visualize the vector field f in Fig.2(b).

The inner product of $f(\mathrm{p}_i)$ and $f(\mathrm{p}_a)$ within the neighbourhood $N_d(\mathrm{p}_i)$, measures how much p_a's convexity center agrees with p_i's. We define $\mathrm{D}^f(\mathrm{p}_i)$ to be the sum of such inner products, indicating degree of agreement of a pixel with its surroundings:

$$\mathrm{D}^f(\mathrm{p}_i) = \sum_{\mathrm{p}_a \in N_d(\mathrm{p}_i)} f(\mathrm{p}_i)^\top f(\mathrm{p}_a). \tag{4}$$

We visualize D^f in Fig.2(c). D^f is rotationally invariant and effectively captures the rough convex shapes of the 2D cross-sections of a tubular structure. Sinks of f (dot centers) are characterized by negative values and sources of f by positive ones. In contrast to morphological charts computed straight from image intensities, D^f is robust to variations of relative intensities of the peaks and valleys in the original image [5].

Given a degree image D^f, we define repulsion $\mathbf{R}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j)$ and attraction $\mathbf{A}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j)$ between pixels p_i and p_j according to the difference of degrees $\mathrm{D}^f(\mathrm{p}_i)$, $\mathrm{D}^f(\mathrm{p}_j)$ to the minimal degree $m_{ij} = \min_{\mathrm{p}_t \in \mathrm{line}(\mathrm{p}_i, \mathrm{p}_j)} \mathrm{D}^f(\mathrm{p}_t)$ encountered on their connecting line:

$$\begin{aligned} \mathbf{R}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j) &= 1 - \exp\left(-\frac{\min(|\mathrm{D}^f(\mathrm{p}_i) - m_{ij}|, |\mathrm{D}^f(\mathrm{p}_j) - m_{ij}|)}{\sigma_r}\right) \\ \mathbf{A}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j) &= (1 - \mathbf{R}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j)) \cdot \delta(||\mathrm{p}_i - \mathrm{p}_j||_2 < r_a), \end{aligned} \tag{5}$$

where δ is the delta function. Attractions are short range, acting on pixels within r_a distance. Parameter r_a is chosen as a lower bound of the distance between adjacent structure centers. We set $r_a = 4$ pixels in all our experiments for stereocilia segmentation.

3.2 Trajectory Partitioning

We compute trajectory pairwise relationships by marginalizing pixel relationships across trajectory lifespans. We define repulsion $\mathbf{R}_T(\mathrm{tr}_i, \mathrm{tr}_j)$ between trajectories tr_i and tr_j

to be the maximum of corresponding pixel repulsions and attraction $\mathbf{A}_T(\mathrm{tr}_i, \mathrm{tr}_j)$ to be the minimum of corresponding pixel attractions as follows:

$$\begin{aligned}\mathbf{R}_T(\mathrm{tr}_i, \mathrm{tr}_j) &= \max_{z \in \mathrm{Z}_i \cap \mathrm{Z}_j} \mathbf{R}_\mathrm{p}(\mathrm{p}_i^z, \mathrm{p}_j^z) \cdot \delta(|\mathrm{Z}_i \cap \mathrm{Z}_j| > 0) \\ \mathbf{A}_T(\mathrm{tr}_i, \mathrm{tr}_j) &= \min_{z \in \mathrm{Z}_i \cap \mathrm{Z}_j} \mathbf{A}_\mathrm{p}(\mathrm{p}_i^z, \mathrm{p}_j^z) \cdot \delta(|\mathrm{Z}_i \cap \mathrm{Z}_j| > 0),\end{aligned} \tag{6}$$

where superscript z indicates the slice index of a pixel. The above cue marginalization reflects the nature of tubular structure bundles: in some slices, tubular structures attached to each other confuse corresponding degree fields as shown in Fig.2(c), causing leakage in segmentation. On the contrary, over-segmentation of 2D cross sections is highly *unlikely* under our convexity cues. As such, trajectory affinities essentially try to detect the informative slice where attached structures separate.

We classify trajectories as foreground or background by thresholding their average degrees $\mathrm{D}^f(\mathrm{tr}_i) = \underset{z \in \mathrm{Z}_i}{\mathrm{mean}}\, \mathrm{D}^f(\mathrm{p}_i^z)$ at 0. Let $X \in \{0,1\}^{|\mathcal{T}| \times K}$ be the matrix whose columns are the indicator vectors of K clusters. We cluster foreground trajectories by maximizing intra-cluster attractions $\mathbf{A}_T$ and inter-cluster repulsions $\mathbf{R}_T$ [15]:

$$\begin{aligned}&\text{maximize} && \varepsilon(X) = \sum_{k=1}^{K} \frac{X_k^\top (\mathbf{A}_T - \mathbf{R}_T + \mathbf{D}_R) X_k}{X_k^\top (\mathbf{D}_A + \mathbf{D}_R) X_k} \\ &\text{subject to} && X\mathbf{1}_{|\mathcal{T}|} = \mathbf{1}_{|\mathcal{T}|},\end{aligned} \tag{7}$$

where $\mathbf{D}_A = \mathrm{Diag}(\mathbf{A}_T \mathbf{1}_{|\mathcal{T}|})$, $\mathbf{D}_R = \mathrm{Diag}(\mathbf{R}_T \mathbf{1}_{|\mathcal{T}|})$ are degree matrices and $\mathbf{1}_{|\mathcal{T}|}$ is the $|\mathcal{T}| \times 1$ vector of 1. We choose K to be a rough upper-bound of the total number of tubular structures present in the stack, estimated from the per frame degree fields. We obtain the near-global optimal continuous solution of Eq.7 from the top K generalized eigenvectors of $(\mathbf{A}_T - \mathbf{R}_T + \mathbf{D}_R, \mathbf{D}_A + \mathbf{D}_R)$. We discretize the eigenvectors by rotation [16] and obtain K clusters. We repeatedly merge clusters with no repulsion between their trajectories. Structure bifurcation can be accommodated by a hierarchical segmentation scheme, where cluster merging probabilities depend on ratios of cluster attractions versus repulsions. We summarize our method in Algorithm 1.

Algorithm 1. Unraveling Tubular Structures

1: Let $\{I^z,\ z = 1 \cdots T\}$ denote an ordered sequence of T images in a 3D stack.
2: **for all** $I^z, z = 1 \cdots T$ **do**
3: Compute peak vector field $f^z(\mathrm{p}_i)$ and degree field $\mathrm{D}_z^f(\mathrm{p}_i)$ using Eq. 3 and Eq. 4.
4: Compute pixel attractions $\mathbf{A}_\mathrm{p}$ and repulsions $\mathbf{R}_\mathrm{p}$ using Eq. 5.
5: **end for**
6: Compute structural trajectories $\mathrm{tr}_i, i = 1 \cdots |\mathcal{T}|$ using method of [14].
7: Compute trajectory degrees $\mathrm{D}^f(\mathrm{tr}_i),\ \ i = 1 \cdots |\mathcal{T}|$.
8: Classify trajectories as foreground $\mathcal{T}^F = \{\mathrm{tr}_i | \mathrm{D}^f(\mathrm{tr}_i) > 0\}$.
9: Compute foreground trajectory attractions $\mathbf{A}_T$ and repulsions $\mathbf{R}_T$ using Eq. 6.
10: Compute the top K generalised eigenvectors V of $(\mathbf{A}_T - \mathbf{R}_T + \mathbf{D}_R, \mathbf{D}_A + \mathbf{D}_R)$.
11: Discretize eigenvectors V by rotation [16] to obtain K trajectory clusters $G_i, i = 1 \cdots K$.
12: **while** $\exists\, G_i, G_j, \mathbf{R}_T(G_i, G_j) = 0$ **do**
13: Merge G_i, G_j
14: **end while**

4 Experiments

We test our method on segmenting stereocilia of the hair cells in the inner ear from their fluorescent image stacks. Each stack is 7 to 20 slices long and contains 50 to 70 stereocilia. 3D ground-truth stereocilia centers are marked in each image stack. Ground-truth samples are illustrated in the first column of Fig. 3. We compare our method against three baseline approaches: 1) 3D k-means on pixel intensities. Number of centers k is chosen to achieve best performance. The resulting clusters are pruned based on their size and aspect ratio. 2) 3D watershed (MATLAB built-in implementation) 3) Dot finding [5] using code provided by the authors. Given the 2D output dots of [5], we produce the 3D segmentation by linking segmented dots between consecutive slices via Hungarian matching. We evaluate performance with the following metrics:

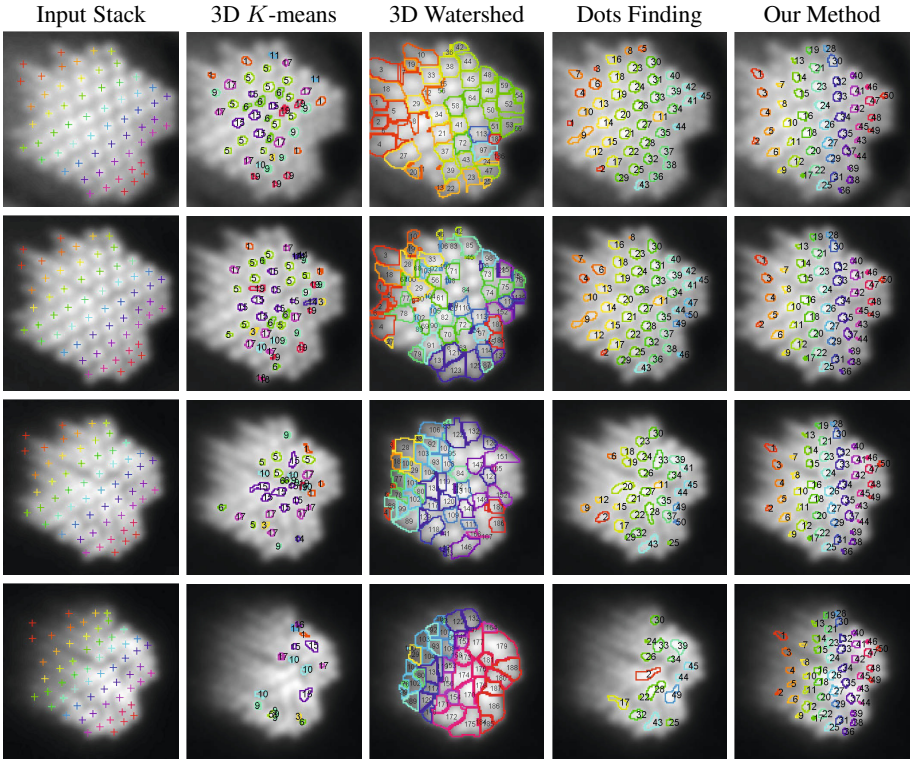

Fig. 3. Segmenting a stereocilia stack (best viewed in color). First column shows 4 (out of 22) images of a stereocilia stack with corresponding 3D ground-truth tubular structure centers. Depth decreases from top to bottom. Column 2-5 show 3D segmentation using 3D k-means, 3D watershed, dot finding [5] and our method respectively. Numbers and colours indicate tubular structure identities. In 3D watershed, tubular structures leak across faint boundaries and break arbitrarily between slices. In dot finding, notice the leaking segments of numbers 9, 20, 38, 43, etc. Our method provides consistent 3D segmentations, correcting leakages and miss-detections.

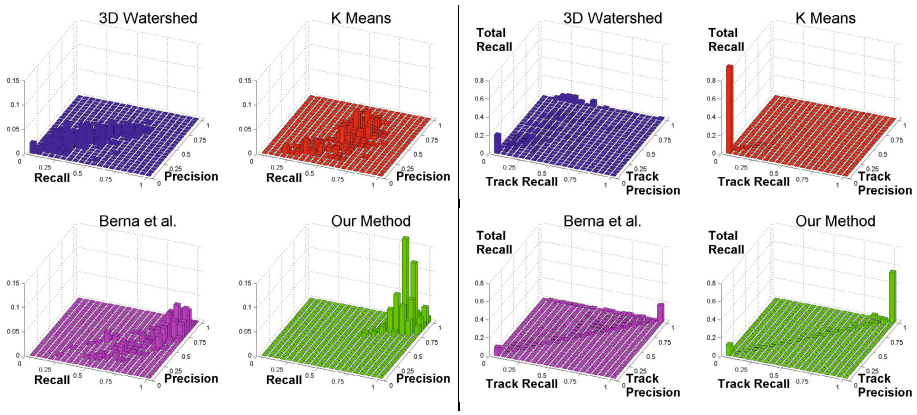

Fig. 4. *Left:* Precision-recall for 2D slice segmentation. We histogram $(\mathrm{prec}_z,\ \mathrm{rec}_z)$ values for all slices z in our stacks. *Right:* Precision-recall for 3D tubular structure identification. We histogram $(\text{track-rec}_i,\ \text{track-prec}_i)$ of all labelled tubular structures in our stacks. Best performance is achieved when the histogram is concentrated at the **right top corner** (precision=1, recall=1). Our method (in green) has significantly higher precision and recall in both tasks.

1) **Goodness of 2D segmentation.** For each slice z, given m_z ground-truth structure centers and n_z segment centers hypotheses, let d_{ij}^z be the Euclidean distance between structure center i and segment center j in slice z. We use the following measures:

$$\mathrm{prec}_z = \frac{\#\{j\colon \min_{i=1}^{m_z} d_{ij}^z \leq \tau\}}{n_z}, \qquad \mathrm{rec}_z = \frac{\#\{j\colon \min_{i=1}^{m_z} d_{ij}^z \leq \tau\}}{m_z}. \tag{8}$$

We visualize the histogram of $(\mathrm{prec}_z, \mathrm{rec}_z)$ over all slices in Fig.4 *left*. Same evaluation metrics were used in [5].

2) **Goodness of 3D identification.** Given m 3D ground-truth tubular structures and n 3D tubular structures hypotheses, let ℓ_i^g denote the length of ground-truth structure i and ℓ_j^d denote the length of segment structure hypothesis j. We use the following measures:

$$\text{track-rec}_i = \max_{j=1}^{n} \frac{\#\{z\colon d_{ij}^z \leq \tau\}}{\ell_i^g}, \qquad \text{track-prec}_i = \max_{j=1}^{n} \frac{\#\{z\colon d_{ij}^z \leq \tau\}}{\ell_j^d}. \tag{9}$$

Tracking precision and recall together quantify how consistently the 3D segmentation hypotheses capture the 3D ground-truth structures [17]. We visualize the histogram of $(\text{track-rec}_i, \text{track-prec}_i)$ over all labelled tubular structures in Fig.4 *right*. We set $\tau = 3$.

Our method outperforms all baseline approaches. Low contrast and faint boundaries make stereocilia segmentation challenging. Our gain in performance comes from corrections of leakages and miss-detections by propagating separations or detections from informative to ambiguous places in the 3D volume, as shown in Fig.4. Miss-detections in our method are often due to localization errors: segment hypotheses centers are a bit more than 3 pixels away from the corresponding ground-truth. A local gradient descent for discovering the intensity peak in the local neighbourhood could alleviate from such localization mistakes. We did not add this step to keep the method clean.

5 Conclusion

We presented an algorithm for unravelling 3D tubular structures in a tight bundle by propagating grouping information across multiple cross-sections of their 3D volume simultaneously via spectral partitioning of structural-flow trajectories. Our qualitative and quantitative results show our method effectively integrates local grouping cues for accurate segmentation and identification of densely packed structures, outperforming 3D and 2D baseline segmentation algorithms. We are currently exploring ways of applying our algorithm to 4D cell tracking, where both temporal and structural correspondences would mediate cues for segmenting spatio-temporal cell structures.

References

1. Ciuman, R.R.: Auditory and vestibular hair cell stereocilia: relationship between functionality and inner ear disease. Journal of Laryngology Otology 125, 991–1003 (2011)
2. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
3. Beucher, S.: Mathmatique, C.D.M.: The watershed transformation applied to image segmentation. Scanning Microscopy International 6, 1–26 (1991)
4. Meyer, F.: Morphological segmentation revisited. Space, Structure and Randomness 183, 315–347 (2005)
5. Bernardis, E., Yu, S.X.: Finding dots: Segmentation as poping out regions from boundaries. In: CVPR (2010)
6. Bernardis, E., Yu, S.X.: Structural correspondence as a contour grouping problem. In: Proceedings of IEEE Workshop on MMBIA (2010)
7. Wahlby, C., Sintorn, I.M., Erlandsson, F., Borgefors, G., Bengtsson, E.: Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. Journal of Microscopy 215, 67–76 (2004)
8. Haralick, R., Shapiro, L.: Image segmentation techniques. Computer Vision, Graphics and Image Processing 29(1), 100–132 (1985)
9. de Bruijne, M., van Ginneken, B., Viergever, M.A., Niessen, W.J.: Adapting Active Shape Models for 3D Segmentation of Tubular Structures in Medical Images. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 136–147. Springer, Heidelberg (2003)
10. Lorigo, L.M., Grimson, W.E.L., Faugeras, O., Keriven, R., Kikinis, R., Nabavi, A., Westin, C.F.: Codimension - two geodesic active contours for the segmentation of tubular structures. In: CVPR (2000)
11. Dorin, H.T., Tek, H., Comaniciu, D., Williams, J.P.: Vessel detection by mean shift based ray propagation. In: Proceedings of IEEE Workshop on MMBIA (2001)
12. Fragkiadaki, K., Shi, J.: Exploiting motion and topology for segmenting and tracking under entanglement. In: CVPR (2011)
13. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
14. Sundaram, N., Brox, T., Keutzer, K.: Dense Point Trajectories by GPU-Accelerated Large Displacement Optical Flow. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 438–451. Springer, Heidelberg (2010)
15. Yu, S.X., Shi, J.: Understanding popout through repulsion. In: CVPR (2001)
16. Yu, S.X., Shi, J.: Multiclass spectral clustering. In: ICCV (2003)
17. Smith, K., Gatica-perez, D., Odobez, J.-M., Ba, S.: Evaluating multi-object tracking. In: Workshop on Empirical Evaluation Methods in Computer Vision (2005)