Robust Segmentation by Cutting across A Stack of Gamma Transformed Images

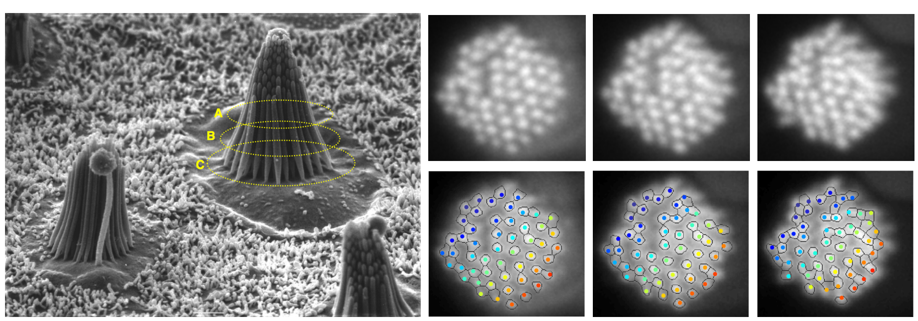
Elena Bernardis[1] and Stella X. Yu[2]

[1]Department of Computer and Information Science
University of Pennsylvania, Philadelphia, PA 19104, USA
[2]Computer Science Department
Boston College, Chestnut Hill, MA 02467, USA

Abstract. Medical image segmentation appears to be governed by the global intensity level and should be robust to local intensity fluctuation. We develop an efficient spectral graph method which seeks the best segmentation on a stack of gamma transformed versions of the original image. Each gamma image produces two types of grouping cues operating at different ranges: Short-range attraction pulls pixels towards region centers, while long-range repulsion pushes pixels away from region boundaries. With rough pixel correspondence between gamma images, we obtain an aligned cue stack for the original image. Our experimental results demonstrate that cutting across the entire gamma stack delivers more accurate segmentations than commonly used watershed algorithms.

1 Introduction

Hair cells of the inner ear transduce mechanical signals into electrical signals [1]. Each hair bundle is composed of tens of stereocilia organized in an organ-pipe-like formation of increasing height (Fig. 1). Automatic segmentation of these stereocilia in their fluorescent images is vital for medical research on hearing.



a: 3D view of hair cells **b:** 2D fluorescent slices & their segmentations

Fig. 1. Stereocilia segmentation. **a)** Hair cells are composed of tens of stereocilia organized in an organ-pipe-like formation of increasing height. **b)** Fluorescent images (Row 1) and their segmentations (our results, Row 2) at multiple heights show the cross sections (e.g. A,B,C in **a**) of individual stereocilia (marked by colored dots).

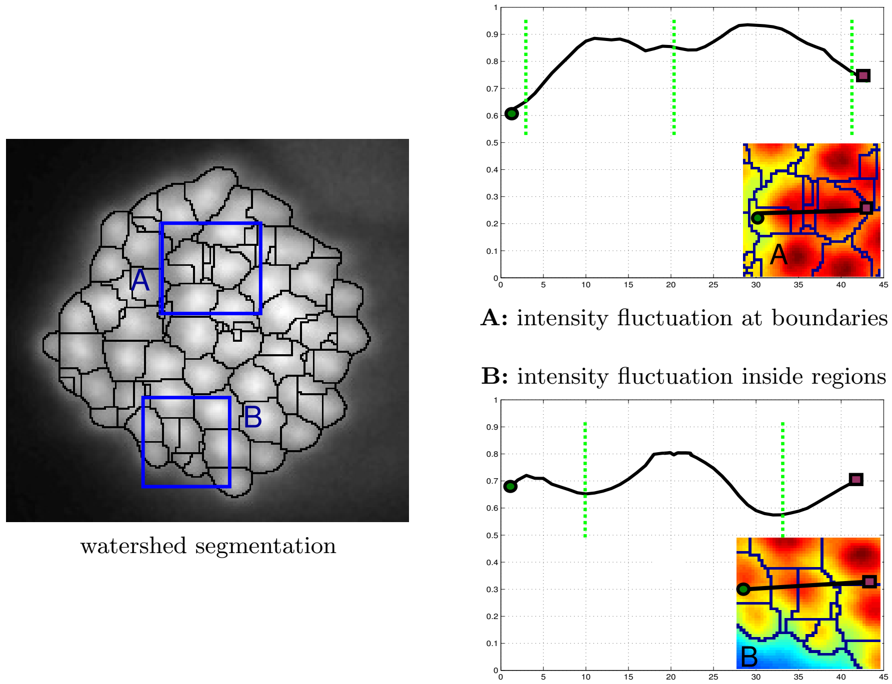


Fig. 2. Local intensity fluctuation presents considerable challenges in medical image segmentation. **A)** Fluctuation at boundaries weakens the separation between two intensity peaks. **B)** Fluctuation inside regions tends to break up an otherwise well defined intensity peak. Both cases cause oversegmentations in watershed approaches. The solid black line plots the 1D intensity profile along the line connecting the two pixels in the inset, which shows the image in a labeled window on the left. The dotted green lines mark the desired boundaries between intensity peaks.

Segmentation of such medical images often appears to be governed by global intensity levels, yet imaging noise and local intensity fluctuation presents considerable challenges. Two scenarios are illustrated in Fig. 2.

Morphological methods and energy-driven methods are widely used in medical image segmentation. While the former prescribes a local computational procedure, e.g. watershed algorithms [2, 3], the latter involves the minimization of a global energy function formulated based on either regions [4, 5] or contours, e.g. active contours [6] and level set methods [7].

While morphological methods are computationally efficient but prone to local noise, energy-driven approaches are computationally costly and critically dependent on initial seed solutions. Various techniques have been proposed to combine their benefits, e.g. watersnakes [8] and level sets for watershed [3].

Graph cuts methods have also been employed to overcome the limitations of watershed algorithms, which are essentially local segmentation methods. These

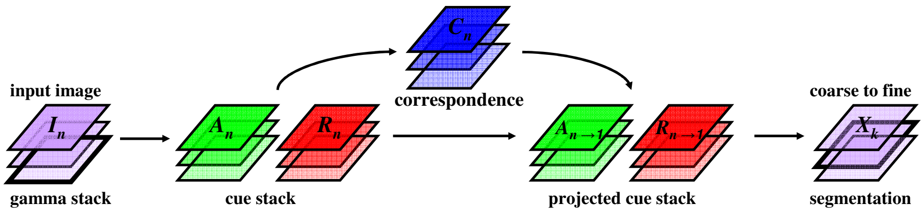


Fig. 3. Method Overview. Given an image, we build a stack of its gamma transformed versions, i.e., $I_n = I^{\gamma_n}$. For each gamma image I_n, we derive pairwise attraction A_n and repulsion R_n between pixels. We compute pixel correspondences C_n between adjacent gamma layers, and project cues at each layer to the reference layer I_1: $A_{n\to 1}$ and $R_{n\to 1}$. Cutting across the aligned cue stack produces segmentation X_k that is invariant to gamma transformations, k indicating the granularity of segmentation.

include segmenting a single connected component with isoperimetric graph partitioning [9] and thin structures with augmented banded graph cuts [10].

We present a graph cuts approach that is robust to local intensity fluctuation and can extract several regions of interest without any user initialization. We encode the impact of high and low intensities, which we will refer to as peaks and valleys, in pairwise grouping cues that encourage peak regions to stay together and valley regions to divide apart. It is the job of global integration to decide where region boundaries should be.

Our key idea is that regions of an image appear stable with respect to the gamma transformation of the image, while cues in each gamma transformed version reflect an ever changing balance between peaks and valleys, as peaks shrink and valleys expand with an increasing gamma. The desired segmentation must be the global consensus of local cues from a stack of these gamma images.

Illustrated in Fig. 3, given an image I, we first create several gamma transformed versions: $I_n = I^{\gamma_n}$. For each I_n, we define two complementary local grouping cues: a short-range attraction between nearby pixels with similar intensities and a long-range repulsion between distant pixels with similar intensities but separated by valleys. The former occurs most likely for pixels belonging to the same stereocilium and the latter for pixels belonging to adjacent stereocilia. Large repulsion demands single boundaries to occur somewhere between two distant pixels, whereas large attraction discourages the formation of boundaries between two nearby pixels, preventing the oversegmentation problems in Fig. 2. We establish rough local alignment between gamma images and project cues derived from each I_n to the original image I through pixel correspondences. We seek the optimal graph cuts across the cue stack of attraction and repulsion, producing segmentation X_k for the original image I at a granularity determined by the number of eigenvectors k.

We will address the integration of multiple cues in Section 2, formulate our pairwise grouping cues for stereocillia images in Section 3, present experimental results in Section 4, and conclude the paper in Section 5.

2 Constrained Cuts with Attraction and Repulsion

We formulate the segmentation in a spectral graph-theoretic framework. We collect pairwise cues and seek the solution that optimizes a global criterion. We consider pairwise cues of three kinds: attraction A, repulsion R, and constraints U. These cues have been studied separately in [11–13]. We combine them for the first time in a single framework.

2.1 Graph Representation

In spectral graph methods, an image I is represented by a weighted graph $G(V, E, W)$, where V denotes the set of nodes, E the set of edges connecting the nodes, and W the weights attached to edges. A pixel then becomes a node in the graph, each pairwise grouping cue becomes a weight between two nodes, and image segmentation becomes a graph node partitioning problem: We seek k partitions of node set V such that $V = \cup_{l=1}^{k} V_l$ and $V_i \cap V_j = \emptyset$, $\forall i \neq j$.

2.2 Criterion with Attraction and Repulsion

A good segmentation should have strong within-group attraction and between-group repulsion, and weak between-group attraction and within-group repulsion.

Characterizing this intuition with *linkratio* allows us to achieve both objectives simultaneously [14]. *linkratio* L of two node sets (P, Q) measures the fraction of connections from P to Q among all the connections that P has:

$$\text{linkratio} \qquad L(P, Q; W) = \frac{C(P, Q; W)}{C(P, V; W)} \tag{1}$$

$$\text{connections} \qquad C(P, Q; W) = \sum_{i \in P, j \in Q} W(i, j) \tag{2}$$

In particular, we have $L(P, P; W) + L(P, V \setminus P; W) = 1$, i.e. maximizing a within-group *linkratio* is equivalent to minimizing its between-group *linkratio*.

We seek to maximize *linkratios* from within-group attraction and between-group repulsion, combined linearly according to their total degree of connections:

$$\max \qquad \varepsilon = \sum_{l=1}^{k} \alpha\, L(V_l, V_l; A) + (1 - \alpha)\, L(V_l, V \setminus V_l; R) \tag{3}$$

$$\text{where} \qquad \alpha = \frac{C(V_l, V; A)}{C(V_l, V; A) + C(V_l, V; R)} \tag{4}$$

α is a number between 0 and 1, indicating the total degree of attraction. $1 - \alpha$ indicates the total degree of repulsion.

2.3 Partial Grouping Constraints

We represent the partitioning by partition indicator $X = [X_1, \ldots, X_k]$, where X_l is an $N \times 1$ binary indicator for partition V_l, $X_l(i) = 1$ if pixel $i \in V_l$, and 0 otherwise, $l = 1, \ldots, k$. N is the number of pixels in the image.

We consider partial grouping constraints which require pixels a and b to belong in the same region, i.e. $X(a) = X(b)$. The collection of c such constraints can be written as $U^T X = 0$, where U is an $N \times c$ matrix, and each column of U has only two non-zero numbers, $+1$ and -1.

2.4 Optimal Solution

Our criterion ε with pairwise attraction A and pairwise repulsion R, subject to grouping constraints U can be written in a compact matrix form:

$$\text{maximize} \qquad \varepsilon(X) = \sum_{l=1}^{k} \frac{X_l^T W X_l}{X^T D X_l} \tag{5}$$

$$\text{subject to} \qquad X \in \{0,1\}^{N \times k},\ X 1_k = 1_N \tag{6}$$

$$U^T X = 0 \tag{7}$$

$$\text{where} \qquad W = A - R + D_R \tag{8}$$

$$D = D_A + D_R \tag{9}$$

1_n denotes the $n \times 1$ vector of all 1's. $D_W = \text{Diag}(W 1_N)$ is an $N \times N$ diagonal matrix, and its diagonal contain the total degree of W connections for each node.

Relaxing the binary constraints, we can solve this optimization problem [11] with the eigenvectors of $HD^{-1}WH$, where $H = I - D^{-1}U(U^T D^{-1} U)^{-1} U^T$. We then discretize the eigenvectors to obtain the final segmentation [14].

3 Pairwise Grouping Cues from Image Intensities

The success of global integration depends on the local cues that feed into it. We define a short-range attraction that pulls pixels towards region centers, a long-range repulsion that pushes pixels away from region boundaries, and partial grouping constraints that force peripheral background pixels to belong together. With pixel correspondence between gamma images, we obtain a cue stack for the original image.

3.1 Short-Range Attraction within Individual Peaks

Attraction $A(i,j)$ between pixels i and j encodes local intensity similarity. The straightforward definition

$$A(i,j) = e^{-\frac{|I_i - I_j|^2}{2\sigma_a^2}} \tag{10}$$

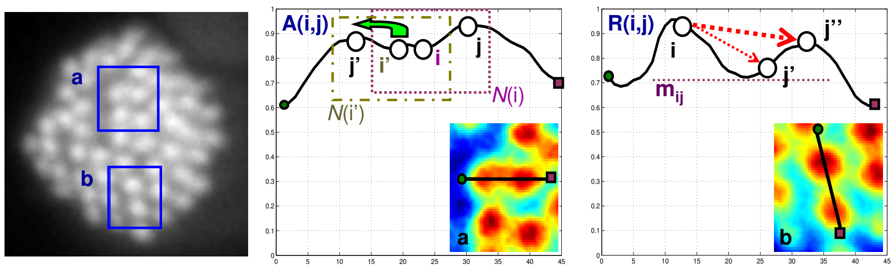


image　　**a**: attraction　　**b**: repulsion

Fig. 4. Pairwise attraction and repulsion. **a)** Our attraction is adaptive to the local intensity range within each neighborhood $\mathcal{N}(i)$, so that $A(i', j') \approx A(i, j)$, enhancing the discrimination of two nearby similar peaks. **b)** Our repulsion is strongest for nearby peaks and gets reduced as two pixels approach the inbetween valley: $R(i, j'') > R(i, j')$. m_{ij} is the minimal intensity level between pixels i and j.

requires fine parameter tuning and tends to merge nearby peaks of similar intensities. We introduce a new definition that is asymmetrical between two pixels and acts to pull pixels towards intensity peaks.

For pixels i and j, $A(i, j)$ is inversely proportional to the the maximal intensity difference between i and any pixel on the line ij, with sensitivity regulated by local intensity range δ_i in i's neighborhood $\mathcal{N}(i)$:

$$A(i, j) = e^{-\frac{\max_{t \in \text{line}(i,j)} |I_i - I_t|^2}{2\delta_i^2 \cdot \sigma_a^2}} \tag{11}$$

$$\delta_i = \max_{t \in \mathcal{N}(i)} I_t - \min_{t \in \mathcal{N}(i)} I_t \tag{12}$$

We choose $\mathcal{N}(i)$ to be slightly larger than a stereocilium so that δ_i is estimated between the peak and surrounding valleys (Fig. 4**a**). With adaptive scaling by local intensity range δ_i, $A(i, j)$ effectively enhances attraction within weak peaks and allows a single parameter setting for σ_a to work on a variety of images.

3.2 Long-Range Repulsion between Peaks

Adjacent peaks provide a strong cue as to where the boundaries should lie. This cue is encoded by long-range repulsion. Intuitively, two pixels of similar intensity should belong to different peaks if they are separated by a valley. We define repulsion $R(i, j)$ between pixels i and j to be proportional to the difference with the minimal intensity m_{ij} encountered on the line ij:

$$R(i, j) = 1 - e^{-\frac{\min(|I_i - m_{ij}|, |I_j - m_{ij}|)}{\sigma_r}} \tag{13}$$

$$m_{ij} = \min_{t \in \text{line}(i,j)} I_t. \tag{14}$$

The farther away the pixels are from the valley, the larger the intensity difference with the minimum, and the larger the repulsion (Fig. 4**b**).

3.3 Pixel Correspondence and Cue Projection

With each gamma transformation, while peaks remain peaks and valleys remain valleys, their regions of influence change: Peaks shrink and valleys expand; pixels belonging to one peak region could become part of the background. Local grouping cues derived from gamma images consequently do not completely agree with each other. We establish rough pixel correspondence and project cues on individual gamma image back to the original image.

Let $A_n(i,j)$ be the affinity (i.e. attraction) between pixels i and j at gamma image I_n. We follow the approach in [15] by computing the corresponding pixel location $C_n(i)$ as the center of mass of i's affinity field and composing them recursively to obtained the cue stack for the original image $I = I_1$:

$$A_{n\to 1}(i,j) = A_n(C_n(i), C_n(j)) \tag{15}$$

$$R_{n\to 1}(i,j) = R_n(C_n(i), C_n(j)) \tag{16}$$

$$C_n(i) = \sum_{j\in N(i)} A_n(i,j) C_{n-1}(i) \tag{17}$$

where $C_1(i)$ is pixel i's location in the original image I.

Cutting across the cue stack is equivalent to cutting a single graph with the following total attraction A and total repulsion R:

$$A = \sum_n D_{A,n}^{-1} A_{n\to 1} + A_{n\to 1} D_{A,n}^{-1} \tag{18}$$

$$R = \sum_n D_{R,n}^{-1} R_{n\to 1} + R_{n\to 1} D_{R,n}^{-1}. \tag{19}$$

where $D_{A,n}$ and $D_{R,n}$ are the degree matrices for A_n and R_n respectively.

3.4 Partial Grouping Constraints

We obtain a crude background mask by intensity thresholding on the original image. This mask is translated into our graph cuts framework as partial grouping constraints where two pixels in the peripheral background must belong together in the final segmentation. We form the constraint matrix U from the collections of these pairwise grouping constraints.

3.5 Algorithm

Given image I, we compute a segmentation using the following procedure:

1. Build a gamma image stack where $I_n = I^{\gamma_n}$ and $I_1 = I$;
2. For each gamma image I_n,
   (2.1) compute attraction A_n and repulsion R_n,
   (2.2) compute pixel correspondence C_n,
   (2.3) compute $A_{n\to 1}, R_{n\to 1}$ by projecting A_n, R_n to the original image I;
3. Compute total attraction A and repulsion R by collapsing the stack;
4. Form partial grouping constraints U from a background mask;
5. Solve the eigenvectors of weights $W = A - R + D_R$ with constraints U;
6. Obtain a discrete segmentation from the eigenvectors.

4 Experiments

We implement our algorithm in MATLAB. The same set of parameters are used for all our images ($\sim 300 \times 300$): $\gamma = \{1, 2, 4\}$, $\sigma_a = 0.3$, $\sigma_r = 2\sigma_a$, neighbourhood radius 8 and 16 for attraction and repulsion respectively. We choose the number of eigenvectors k according to the expected number of stereocilia in the image.

Fig. 5 shows that better segmentation is achieved by integrating cues over the entire gamma stack instead of an individual gamma image. Single peaks originally faint or without clear boundaries are enhanced in gamma transformed images. However, with an increasing gamma, valleys are widened and boundaries become less precise. Cutting across the gamma image stack allows segmenting out weak peaks while retaining precise boundaries throughout the image.

Fig. 6 shows our coarse to fine segmentations. When the number of eigenvectors k is small, our segmentations resemble the watershed results. However, our segmentations are not disrupted by local intensity fluctuation and do not break up salient peaks. When k increases, our segmentions locate each peak with tighter delineation. Our method successfully segments out weak peaks without utilizing the near regularity of the spatial layout of stereocilia.

Fig. 7 shows additional results on images of poor imaging quality.

We measure the goodness of segmentation by scoring it with respect to the ground-truth center locations of stereocilia. Let $\texttt{disk}(i)$ denote a disk of some fixed radius throughout the haircell bundles, located at stereocilium center i. Let $\texttt{segment}(i)$ denote the segment of maximal overlap with $\texttt{disk}(i)$. Our score is a number between 0 and 1, measuring the extent of overlap between $\texttt{disk}(i)$ and $\texttt{segment}(i)$:

$$\texttt{score}(i) = \frac{\texttt{disk}(i) \cap \texttt{segment}(i)}{\texttt{disk}(i) \cup \texttt{segment}(i)}. \tag{20}$$

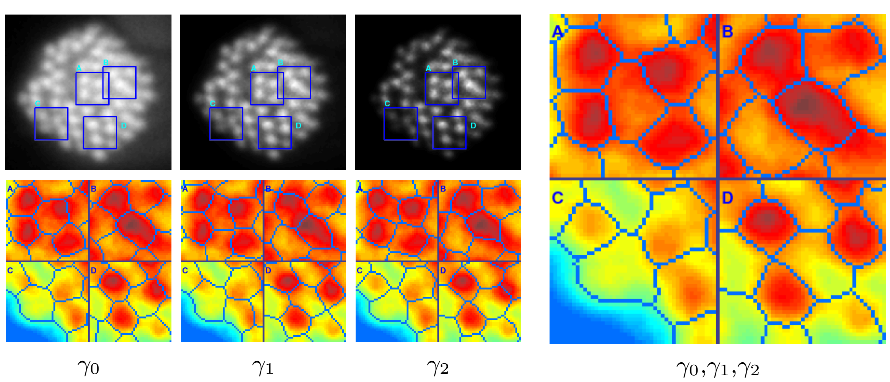


Fig. 5. Better segmentation is obtained by cutting across the gamma stack instead of a single gamma image. Left shows 3 individual gamma images and their segmentations in 4 labeled windows. Right shows the segmentations based on all 3 γ images.

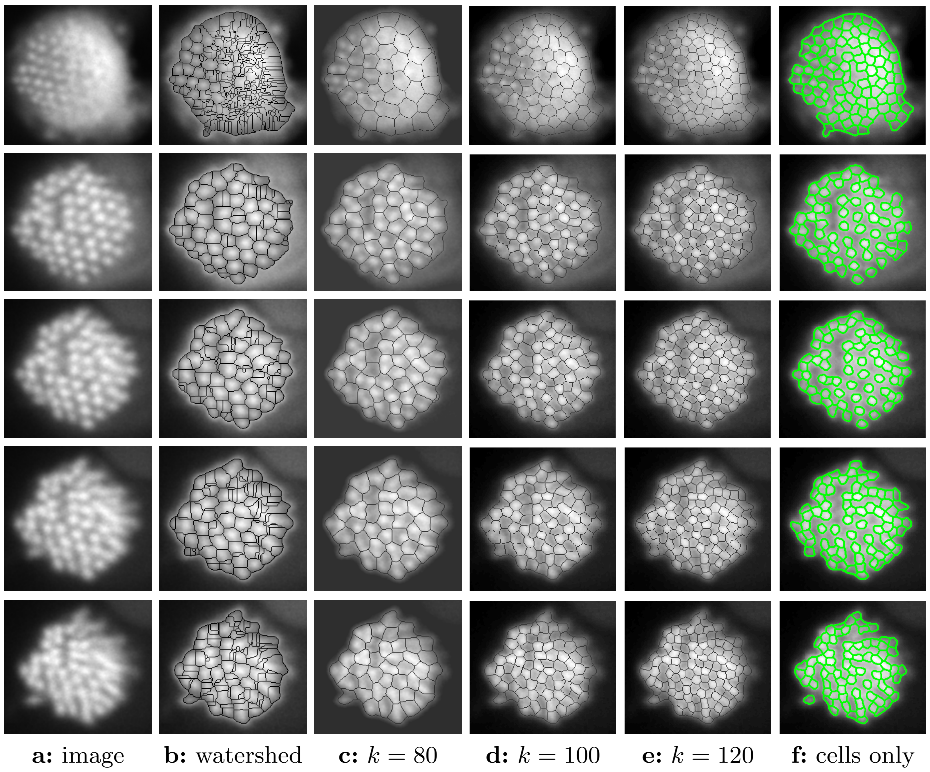


Fig. 6. Coarse-to-fine stereocilia segmentations. For each image (Column **a**), we show watershed segmentations (Column **b**) and our results (Columns **c-e**) as the number of eigenvectors k increases. Extracted stereocilia (Column **f**) show that our method is robust to local intensity fluctuation, can discover weak peaks and precise boundaries.

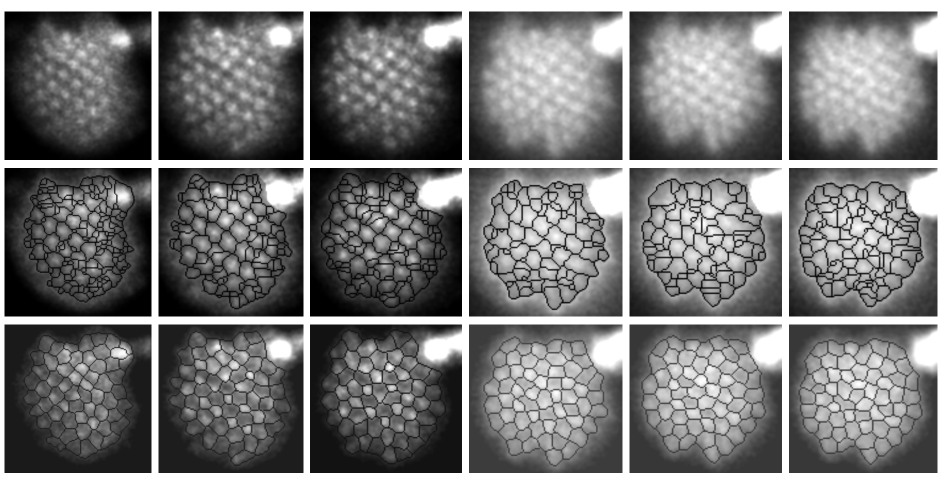
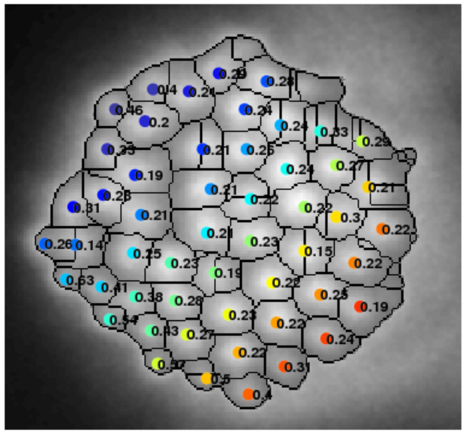
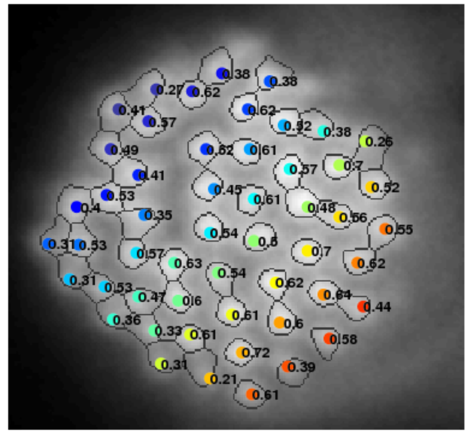


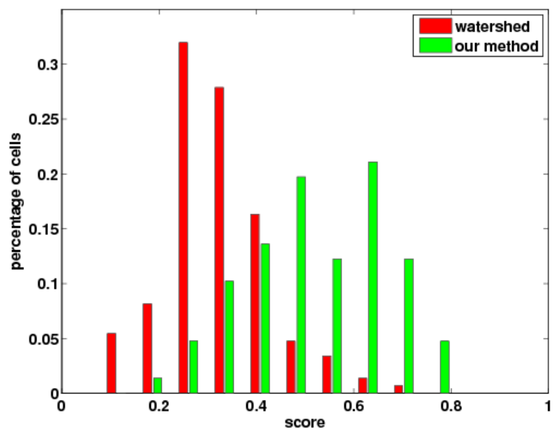
Fig. 7. Our method works equally well on noisy and low-contrast images. $k = 40$. Rows 1-3 show images, watershed results and our results respectively.



a: watershed



b: our method at $k = 120$



c: distribution of scores

Fig. 8. Segmentation scores with respect to ground-truth stereocilia centers. These center locations are marked by colored dots. Each number indicates the score of a particular segment that contains a stereocilium center. **a** and **b** show a score example for watershed and our method. **c** shows the distribution of scores from all the images. Our method has a higher score than watershed overall.

The higher the score, the more precise the segmentation. As the number of eigenvectors increases, our segmentation captures a more precise shape of individual stereocilia. Fig. 8 shows with both image examples and statistics that our method overall scores higher than watershed.

Our method segments the background into multiple valley regions, which are of little interest to medical researchers. By requiring the mean intensity in the region center to be higher than the periphery, we get rid of valleys and automatically extract the stereocilia, as shown in Fig. 6**f**.

5 Conclusions

The segmentation of medical images appears to be governed by the global intensity level, yet local intensity fluctuation poses considerable challenges to both local methods such as watershed and global methods such as level sets.

We develop a spectral graph-theoretic method which finds the best segmentation on a stack of gamma transformed versions of the original image. Each gamma image produces two kinds of local grouping cues: short-range intensity similarity cues that pull pixels towards stereocilia centers, and long-range intensity difference cues that push pixels away from stereocilia boundaries. We obtain a cue stack for the original image using pixel correpondences between gamma images. We then seek the optimal graph cuts across the aligned cue stack which maximize within-group attraction and between-group repulsion. The near-global optimal solution can be found efficiently using eigendecomposition.

Our method has only a few parameters and requires little tuning. We obtain accurate and robust results on a variety of images with the same set of parameters, demonstrating the advantage of cutting across the entire gamma stack instead of the original image or any gamma image alone, and achieving better performance than watershed algorithms.

The segmentation issues we investigate in this paper are not restricted to stereocilia images. Our approach of making a global decision based on two types of local cues operating at different spatial ranges and from multiple gamma images provides a robust and efficient alternative to watershed or level set methods in many medical image applications.

References

1. Vollrath, M.A., Kwan, K.Y., Corey, D.P.: The micromachinery of mechanotransduction in hair cells. Annual Review of Neuroscience **30** (2007) 339–65
2. Meyer, F.: Topographic distance and watershed lines. Signal Process. **38**(1) (1994) 113–125
3. Tai, X., Hodneland, E., Weickert, J., Bukoreshtliev, N., Lundervold, A., Gerdes, H.: Level set methods for watershed image segmentation. In: SSVM07. (2007) 178–90
4. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Math. (1989) 577–684
5. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. (1990) 452–72
6. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing (1998)
7. Malladi, R., Sethian, J.A.: Level set methods for curvature flow, image enchancement, and shape recovery in medical images. In: In Proc. of Conf. on Visualization and Mathematics, Springer-Verlag (1997) 329–45
8. Nguyen, H.T., Worring, M., van den Boomgaard, R., Worring, H.T.N.M.: Watersnakes: Energy-driven watershed segmentation (2003)
9. Grady, L.: Fast, quality, segmentation of large volumes - isoperimetric distance trees. In: ECCV (3). (2006) 449–62

10. Sinop, A.K., Grady, L.: Accurate banded graph cut segmentation of thin structures using laplacian pyramids. In: MICCAI (2). (2006) 896–903
11. Yu, S.X., Shi, J.: Segmentation given partial grouping constraints. PAMI **26**(2) (2004) 173–83
12. Yu, S.X., Shi, J.: Understanding popout through repulsion. In: CVPR, Kauai Marriott, Hawaii, USA (9-15 Dec 2001)
13. Yu, S.X., Shi, J.: Segmentation with pairwise attraction and repulsion. In: ICCV, Vancouver, Canada (9-12 July 2001)
14. Yu, S.X., Shi, J.: Multiclass spectral clustering. In: ICCV, Nice, France (11-17 Oct 2003)
15. Yu, S.X.: Segmentation induced by scale invariance. In: CVPR, San Diego (20-26 June 2005)