

Iterative Tensor Voting for Perceptual Grouping of Ill-Defined Curvilinear Structures

Leandro A. Loss*, *Member, IEEE*, George Bebis, *Member, IEEE*, and Bahram Parvin, *Senior Member, IEEE*

Abstract—In this paper, a novel approach is proposed for perceptual grouping and localization of ill-defined curvilinear structures. Our approach builds upon the tensor voting and the iterative voting frameworks. Its efficacy lies on iterative refinements of curvilinear structures by gradually shifting from an exploratory to an exploitative mode. Such a mode shifting is achieved by reducing the aperture of the tensor voting fields, which is shown to improve curve grouping and inference by enhancing the concentration of the votes over promising, salient structures. The proposed technique is validated on delineating adherens junctions that are imaged through fluorescence microscopy. However, the method is also applicable for screening other organisms based on characteristics of their cell wall structures. Adherens junctions maintain tissue structural integrity and cell–cell interactions. Visually, they exhibit fibrous patterns that may be diffused, heterogeneous in fluorescence intensity, or punctate and frequently perceptual. Besides the application to real data, the proposed method is compared to prior methods on synthetic and annotated real data, showing high precision rates.

Index Terms—Adherens junctions, curvilinear structures, iterative tensor voting (ITV), perceptual grouping.

I. INTRODUCTION

NEARLY one third of the human genome is involved in the regulation of membrane-bound macromolecules. Adherens junctions (e.g., E-cadherin) form an important subclass that maintains tissue architecture and cell–cell interactions on a multicellular model system. For example, it is well known that the loss of E-cadherin increases motility and contributes to cancer progression. However, signals associated with the adherens junctions can be sequestered along neighboring cells and form perceptual curvilinear structures. Furthermore, such a signature is widely observable in other organisms with cell wall structures that consist of both lipid and protein (e.g., algae).

Manuscript received December 08, 2010; revised March 01, 2011; accepted March 02, 2011. Date of publication March 17, 2011; date of current version August 03, 2011. This work was supported in part by 1) the Director, Office of Science, Office of Biological and Environmental Research, the Radiotracer and Imaging Technologies of the U.S. Department of Energy, and 2) a grant from the National Cancer Institute (R01CA140663) under Contract DE-AC02-05CH11231. *Asterisk indicates corresponding author.*

*L. A. Loss is with the Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94575 USA (e-mail: laloss@lbl.gov).

G. Bebis is with the Department of Computer Science and Engineering, University of Nevada, Reno, NV 89577 USA, and also with the Computer Science Department, King Saud University, Riyadh, Saudi Arabia (e-mail: bebis@cse.unr.edu).

B. Parvin are with the Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94575 USA and also with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA. (e-mail: b_parvin@lbl.gov).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2011.2129526

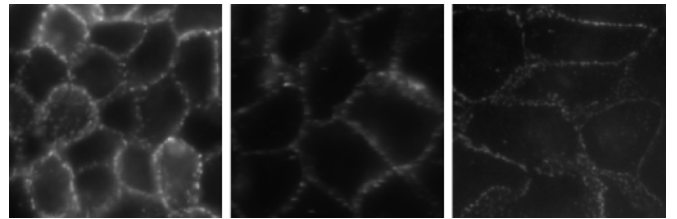


Fig. 1. Adherens junctions exhibit complex patterns. The signal is frequently diffused and sequestered along the cell membranes.

It is well known that perceptual grouping is present in the human vision system as bottom-up preattentive processes that aid in object-level delineation and recognition [1]. From when it was initially conceived by the Gestalt psychologists [2] to now, perceptual grouping has evolved from the passive observation of human behavior to its inclusion in a wide-range of computer vision applications [3]–[6]. Perceptual grouping mainly appeals to image segmentation because of its preattentive use of local cues, which reduces its complexity as well as reduces the necessity for prior knowledge when inferring structures from images. On the negative side, by relying on computed local cues from images, perceptual grouping approaches can be susceptible to scale variations and noisy measurements.

In this paper, we propose a grouping approach that explores purely preattentive cues, such as proximity, good continuity of image primitives, and minimization of the measurements' sensitivity to scale and noise by employing an iterative voting strategy to structural inference. The method proposed here is able to infer curvilinear structures from ill-defined, noisy, often incomplete signals, such as those found in the adherens junctions shown in Fig. 1.

A. Previous Work

In order to illustrate applications of perceptual grouping in the context of image structure inference and segmentation, we begin by citing the basilar works from [7]–[9], [11], [12], [18]. Marr [7] was one of the first researchers to publish the importance of preattentive cues for image understanding. Ullman [9], in a milestone for image completion, addressed the problem of image edge grouping with an optimizing cost function based on smoothness of the contour curvature. Héroult and Horaud [12] also used an optimization approach to segment oriented edges into figure and background. They utilized simulated annealing to maximize a cost function based on proximity, contrast and co-circularity. However, Lowes [8] was one of the first researchers to document utilization of perceptual grouping as part of the solution of a computer vision problem. In that context, perceptual cues such as proximity, collinearity and parallelism

were used to produce structural hypotheses for a model based matching algorithm, thereby permitting spatial correspondences between 2-D images and 3-D object models and, consequently, the recognition of real-world objects. Alternatively, Parent and Zucker [11] proposed an iterative, graph-based labeling scheme that utilized local kernels, whose topography incorporated the good continuity aspect of the Gestalt philosophy to detect local organization and infer curves from images. But it was Sarkar and Boyer [18] who combined proximity, good continuity, parallelism, and perpendicularity to establish pairwise relationships between image primitives and to populate a compatibility graph whose spectra (eigenvalues and eigenvectors) revealed curvilinear structures in an image. From all these approaches, it is important to highlight the researchers' common pursuit of the ideal set of Gestalt principles for perceptual grouping. Also, the set of perceptual cues was generally combined into a cost function that is presently called saliency. The terms *less salient* or *more salient* are, therefore, applied according to a structure's weaker or stronger response to the saliency function.

In the context of voting as a precursor for perceptual grouping, many methods have been developed. For example, Hough [25] introduced the notion of parametric clustering in terms of well-defined geometry, which was later extended to the generalized Hough transform [30]. Sarkar and Boyer [15] introduced a technique that infers structures in an image after voting for the most promising ones from a pool of structural hypotheses. Parvin *et al.* [26] developed an iterative voting system that employs funneling kernels to refine paths along low curvature regions in images. Guy and Medioni [23] proposed a general purpose approach (later revisited and formalized as the tensor voting framework [24]) that uses deformable unities to reveal perceptual structures. Tong and Tang [35] proposed an adaptive tensor voting for improved gap filling and contour closure. Their three-pass tensor voting approach promoted improvements in gap filling over the classical tensor voting. An iterative version of the tensor voting framework is described by Fischer *et al.* [32], who demonstrated how re-voting improves the orientation estimation at the input primitives and, therefore, the overall curve inference result. Loss *et al.* [33] described a scheme for figure-ground segmentation based on tensor voting that gradually eliminates background elements after multiscale voting iterations. They demonstrated the improvements caused by re-voting for figure characterization in cluttered scenarios.

In general, voting operates as a function of continuity and proximity, which can occur at multiple scales (e.g., points, lines, parallel lines, etc.). One of the main advantages of voting frameworks is their reliance on relative simple models, which considerably reduce the number of free parameters and the overall complexity.

B. Motivation

Inference of curvilinear structures from image primitives is particularly relevant to biomedical image analysis. In this field, many detection and recognition methods rely on good markers delineating the objects of interest (e.g., [20]–[22]). However, inherent technical and biological variations affect signal quality in different ways. For example, adherens macromolecules, responsible for tissue architecture, can provide morphological indices

that quantify the loss of tissue organization and cellular morphology as a result of stress conditions.

From the point of view of image segmentation, traditional perceptual grouping of curvilinear structures strongly relies on 1) good contrast between lines and background, 2) well behaved structural definition of the lines, and 3) a good assumption of the lines' scale. However, with optical resolution, adherens junctions can be ill-defined and, therefore, inconsistent with all the above assumptions. For example, they may have nonuniform intensity, be punctate (e.g., perceptual with gaps), be diffused at certain locations, and have distinct widths and lengths (Fig. 1). This significant amount of heterogeneity can happen as a result of the natural variation of chemical binding between the molecules and the staining reagent, or other technical and biological variables.

For these very reasons, a requirement for grouping ill-defined curvilinear structures should not assume 1) rigid models (e.g., [25], [30]), or 2) a precomputed set of image primitives (e.g., [9], [31], [32]), or 3) segments solely extracted from boundaries (e.g., [15], [33], [36]). In addition, due to the continuous refinement required to better detect heterogeneous, low signal-to-noise contours, noniterative approaches are mostly unsuitable (e.g., [24], [35]).

C. Approach

Our work builds upon Guy and Medioni's Tensor Voting Framework [24] and the Iterative Voting Framework by Parvin *et al.* [26] in order to produce an efficient method to enhance and infer perceptually interesting curvilinear structures in images. By coupling tensor and iterative voting fundamentals, we leverage advantages of both methods to produce better results than those achieved by them individually. The main novelty of our method lies in the extension of the tensor voting framework to gradually refine curvilinear structures at different scales. Iterative *funneling* of tensor fields is shown to achieve better definition of pixel orientation and connectivity. This funneling operation gradually increases the concentration of the votes' energy over promising areas, eventually improving the inference of curvilinear structures.

In one classical tensor voting approach, thresholding is applied to the resulting saliency map, thus removing tensors with low saliency. The remaining tensors are then assumed to be part of the salient structures [29]. When voting is performed on a sparse basis (i.e., image primitives are extracted prior to tensor voting), a second pass of dense tensor voting, called densification, is sometimes used [24]. However, in the method proposed here, we impose a set of consecutive tensor voting passes, all in a dense fashion, after thresholding out tensors with low saliency in a conservative manner. Instead of densification, each iteration of the proposed approach aims at refining the previous one, where magnitude and orientation might have been disturbed by frequent low-saliency background noise. These steps resemble the ones performed in [19], where a second pass of the Hough transform voting pass is introduced to fix or refine the elements' orientation, and in [11], where a relaxation-like iterative process seeks convergence and consistence of the inferred structures. Note that the benefits of conservative elimination of tensors with

low saliency from the tensor voting application were investigated and explored in [33].

In order to assess the improvements yielded by the iterative tuning of tensor voting, we performed experiments involving synthetic configurations and real microscopic images. Synthetic data were used to help us analyze and predict the behavior of the method on structures with different degrees of punctation, width, curvature and junctions. Experiments on microscopic images aim at evaluating the method on real scenarios, providing proof of its actual potential and effectiveness.

Components of this paper have already appeared in two other versions as conference papers by the same authors: an introduction to an earlier version of the method, along with some preliminary results focusing particularly on its application from a biological point of view, is in [16]; and a slightly longer version appeared in [17]. In this paper, however, we provide a deeper and more detailed description of the improved technique, besides quantitative analysis on synthetic data, and comparisons with other approaches to the same problem.

The remaining of this paper is organized as follows: Section II describes the tensor voting and the iterative voting frameworks, along with their applications to perceptual grouping of linear structures. Section III introduces our method, extending the concepts of the tensor and the iterative voting frameworks. Experimental results are shown in Section IV, and conclusions are presented in Section V.

II. VOTING FRAMEWORKS

A. Tensor Voting Framework

In the framework proposed in [24]¹, perceptual grouping is achieved by vote casting between primitives of an image. Such primitives are represented by tensors, mathematical entities whose capability for encoding magnitude and orientation make tensor voting particularly efficient for detection of perceptually organized structures, such as edges, lines and regions. In 2-D, tensors are represented analytically as second order nonnegative definite matrices, or geometrically as ellipses, shaped by the tensors eigenvalues' magnitude and eigenvectors' directions. Initialized with an arbitrary size, shape, and orientation, input tensors are deformed due to the accumulation of votes cast by other neighboring tensors.

Votes are tensors composed of magnitude and orientation, which encode the Gestalt principles of proximity, smoothness and good continuation. The tensor's size and shape are given by its eigenvalues (λ_1, λ_2 ; $\lambda_1 \geq \lambda_2 \geq 0$), while its orientation is given by the respective eigenvectors (\vec{e}_1, \vec{e}_2). For example, consider two tensors, positioned at (x, y) and (u, v) , in the (x', y') coordinate system, as shown in Fig. 2. How can the vote from (x, y) be cast onto (u, v) subject to smoothness and proximity as stated before? The simplest way is to model smoothness and proximity as curvature and arc length, respectively. Let l be the distance between the two positions, and θ be the angle between the tangent of the osculating circle at (x, y) and a line that connects (x, y) to (u, v) . The arc length and curvature are given by $s = \theta l / \sin(\theta)$ and $\kappa = 2 \sin(\theta) / l$, respectively. Without any prior knowledge, the path defined by an os-

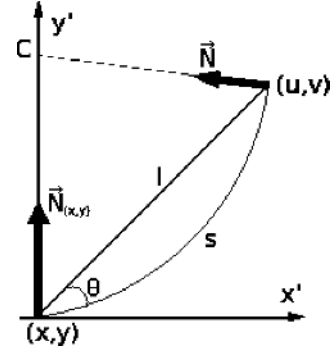


Fig. 2. Two tensors and their geometrical relationship to produce the vote $V(u, v)$ expressed by (1) (figure redrawn from [34]).

culating circle provides the minimum energy since its curvature is kept constant. The vote at position (u, v) is thus given by (1) [24]

$$V(u, v) = e^{-(s^2 + c\kappa^2)/\sigma^2} \vec{N} \vec{N}^T. \quad (1)$$

Here, \vec{N} is the vector normal to the tangent of the same osculating circle at (u, v) , which points to the center C of the circle. It can be calculated by $[-\sin(2\theta) \cos(2\theta)]^T$. The scale factor σ is the only free parameter in this expression and determines the extension of the voting neighborhood. The parameter c is a function of the scale and has been optimized at $c = -16 \ln(0.1) \times (\sigma - 1) \times \pi^{-2}$ to control the decay at high curvature areas (for instance, where two orthogonal lines meet to form a rounded corner) [24].

Depending on the nature of the input primitives, prior information about their orientation can also be used when available. The tensor voting framework was designed to offer two possible voting configurations: one that concentrates the votes according to the input orientation (Fig. 3(a)—stick field), and another one that casts votes radially (Fig. 3(b)—ball field), respectively. The voting fields are the composition of all votes that can be cast from a tensor located in the center of the field to its neighboring tensors. Their extension is controlled by σ and, for practical reasons, the fields are usually truncated past 99% decay. In addition, the stick field is limited to exist only at $|\theta| \leq 45^\circ$, as beyond this angle the osculating circle ceases to represent the smoothest path between the tensors. A somewhat similar kernel topography is also found in [9] and [13].

The tensor deformation imposed by accumulating the strength and orientation of the votes eventually reveals behavioral coherence among image primitives. The vote accumulation is simply tensor addition (e.g., summation of matrices), which can be algebraically represented by $T_{uv} = \sum_{T_{xy}} V(u, v)$, where T_{uv} is the resulting tensor at location (u, v) , after receiving the votes $V(u, v)$ from its neighboring tensors T_{xy} at locations (x, y) . Each kind of structure is expected to produce tensors of a particular shape: for example, very elongated tensors (high $\lambda_1 - \lambda_2$) for lines, and more rounded ones (low $\lambda_1 - \lambda_2$) for regions. Fig. 4 exemplifies how a set of input primitives are encoded as tensors, whose deformations resulting from accumulated votes reveal an underlying salient linear structure.

¹Tensor voting is a patented framework.

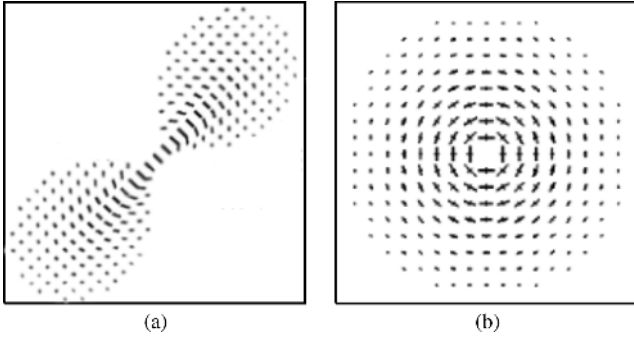


Fig. 3. Tensor voting fields. (a) Stick field: when an estimate of the initial orientation is known, and (b) ball field: when orientation information is unknown.

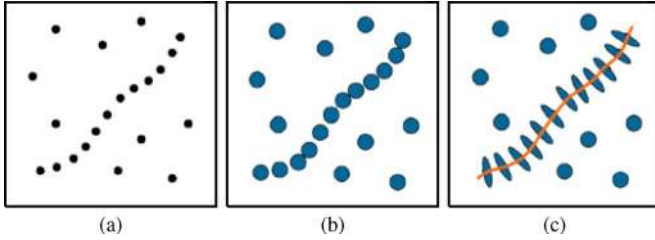


Fig. 4. Example of perceptual grouping through tensor voting. A set of (a) input primitives are (b) encoded as tensors, whose (c) resulting deformations reveal a curve.

The voting process can also be either *sparse* or *dense*. Sparse voting restricts tensors to cast votes only on other encoded input tensors, while dense voting extrapolates the input configuration allowing tensors to cast votes everywhere within their neighborhood. Tensor voting has been shown to be robust to considerable amounts of noise and does not depend on critical thresholds.

B. Iterative Voting

The framework proposed in [26] and [38] also uses spatial voting to detect radial symmetry and to group curvilinear patterns. Similar to the tensor voting framework, each pixel propagates its structural likelihood within its neighbors by vote casting. Here, however, the propagated information is based on either spatial curvature or gradient features and a set of pre-computed Gaussian kernels with a preferred topography to elucidate desirable saliency. The kernels' topography was devised to incorporate the Gestalt principles of proximity and continuity. The iterative voting includes a *funneling* modification of the kernels as iterations progress. The funneling process is aimed at concentrating the voting energy over emerging promising areas by gradually reducing the kernel's aperture. For each pixel at position (x, y) , its vote casts to the neighboring pixels (u, v) by

$$V(u, v) = G(u, v, \sigma) \times \begin{cases} 1, & (u, v) \in A(\sigma, \theta) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $G(u, v, \sigma) = (1/\sqrt{2\pi\sigma})e^{-((x-u)^2+(y-v)^2)/(2\sigma^2)}$, and σ is a scale parameter that regulates the extent and decay of the function's influence. $A(\sigma, \theta)$ is the limiting cone that extends longitudinally up to a radius which is a function of σ , and transversally of an angle θ . At each iteration, θ is reduced by

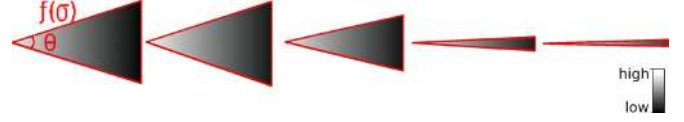


Fig. 5. The iterative voting kernel's topography (only one side is shown) as a function of θ . The voting energy is funneled as iterations progress.

$\Delta\theta$. Fig. 5 illustrates $V(u, v)$ at consecutive iterations (i.e., with different θ) bound by $A(\sigma, \theta)$.

Iterative voting is as follows: at iteration i , the kernel V , with $\theta_i = \theta_{i-1} - \Delta\theta$, is placed over a voting pixel (x, y) , and oriented along the pixel's direction. Each neighboring pixel (u, v) receives the corresponding vote $V_{x,y}(u, v)$, and accumulates its magnitude. At the first iteration, gradient or curvature is used to estimate the direction at (x, y) . At any consecutive iteration $i + 1$, the orientation at (x, y) is recomputed so that it points at its adjacent neighbor with the maximum magnitude (i.e., accumulated votes). Since a better estimation of the structural localization and local orientation are produced after each iteration, the kernels have their energy gradually funneled from initially diffused to eventually very focused. Note that θ changes linearly with i .

III. ITERATIVE TENSOR VOTING

We build upon the Tensor Voting and the Iterative Voting frameworks in order to leverage advantages from both methods, and produce a robust method to group perceptually punctate patterns (e.g., sequestered macromolecules between neighboring cells). The approach is based on progressive funneling of the stick tensor voting field (Fig. 3), which enhances the concentration of the votes over salient features, as observed in [26] for the Iterative Voting. The input to our method is the image itself (i.e., an intensity map), and the output is the inferred salient curvilinear structures. This approach is compared with two prior methods to demonstrate superior performance on images whose structures 1) may be highly punctate, 2) are surrounded by clutter, and 3) have nonuniform and diffused intensities.

A. Image Encoding

The first step is the construction of a voting space. We start by encoding every pixel in the image as an unoriented tensor. In tensor voting, an unoriented tensor has a perfect circular shape or, analytically speaking, has $\lambda_1 = \lambda_2$. This dense encoding enables the method to be completely independent of pre-computations of any sort, such as curvature or gradient, which would force the method to rely on the quality of the image's sparse representation and potentially impair the solution from the beginning. Assuming that the signal of interest has brighter intensity than its counterpart background, a tensor at location (x, y) in the image is encoded to have a size proportional to the pixel intensity I_{xy} (e.g., $\lambda_1 = \lambda_2 = I_{xy}$). If this assumption does not correspond with the images under analysis, adjustments must be made in this encoding in order to guarantee that the signal of interest is input as a larger tensor. For instance, $\lambda_1 = \lambda_2 = I_{\text{white}}/I_{xy}$, where I_{white} is the intensity of the color white in the image (usually 1 or 255), for a dark signal on bright

background. The tensor direction \vec{e}_1 and \vec{e}_2 can be chosen arbitrarily, as it does not influence the ball tensor voting in any sense. An input tensor T_{xy} at location (x, y) is thus

$$T_{xy} = \begin{bmatrix} I_{xy} & 0 \\ 0 & I_{xy} \end{bmatrix}.$$

B. Voting Iterations

After the voting space is constructed, a tensor voting pass is executed using the ball field [Fig. 3(b)]. The ball field is the only option here because no initial tensor orientation is known. Note, though, that disregarding the initial uncertainty about each pixel orientation, the ball field casts a vote with magnitude and orientation, which allows the tensors to start their characteristic structural deformation. Moreover, it should be noted that since all the pixels in the image were encoded, the voting cast is dense by nature (e.g., all tensors cast and receive votes). The deformation caused by locally accumulating all votes from the ball tensor voting pass reveals, although still inaccurately, the presence of perceptual structures in the image, or lack thereof. The resulting tensor's magnitude (λ_1, λ_2) and direction (\vec{e}_1, \vec{e}_2) are obtained by recomputing its eigen-decomposition. From the vote casting perspective, curvilinear structures are characterized by an unbalanced distribution of elements along one main direction. For this reason, in contrast to other structures, whose tensors tend to deform more evenly due to the influence from different directions, curvilinear structures produce elongated tensors. Therefore, tensors from curvilinear structures are likely to be evidenced within the stick saliency map [e.g., image formed by computing $\lambda_1 - \lambda_2$ at each location (x, y)]. Following the same reasoning, one can note that junctions (a spot in the curve where two or more curves intersect) receive votes from possibly multiple directions, deforming more evenly than curvilinear structures. In order to assure completeness of the grouped curvilinear structures, the ball saliency map [e.g., image formed by computing λ_2 at each location (x, y)] is summed to the resulting saliency map.

Our protocol then proceeds with iterative dense tensor voting, which aims at gradual refinement of the previous iterations. A thresholding step is introduced prior to each iteration so that the tensors that did not deform as expected for curvilinear structures (e.g., high $\lambda_1 - \lambda_2$) are removed. This thresholding is very conservative and aims at removing elements with extremely low saliency to enhance the processing speed. Therefore, only tensors whose saliency $\lambda_1 - \lambda_2$ is higher than a threshold value are encoded. Since the tensors are guaranteed to have a more or less accurate orientation (unoriented tensors are thresholded), consecutive iterations are performed with stick fields [Fig. 3(a)]. Note that the voting employed is still dense, so every site receives a vote, even if not initially inputted. Each iteration refines the previous one, where the tensors' magnitude and orientation seem to have been disturbed by the low-saliency but frequent background noise, as well as the often noisy properties found in ill-defined signals.

An interesting observation, inspired by the funneling progression of kernels proposed in [26] and [38], is that the stick fields are gradually modified (e.g., the field aperture is reduced) as

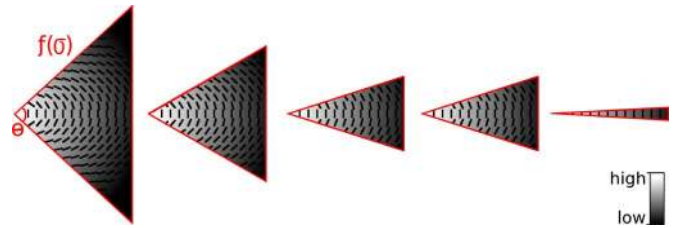


Fig. 6. The ITV kernel's topography (only one side is shown) as a function of θ . The magnitude and orientation of votes vary within the kernel, while the kernel's energy funnels as iterations progress.

the voting iterations proceed and the orientation estimations become continuously more accurate. This funneling process is a key aspect of our algorithm. It gradually reduces the diffusion of votes and concentrates the votes only over promising lines, producing better, enhanced results. Fig. 6 illustrates their topography for five exemplar iterations. Given a voting scale σ and field aperture θ , the iterative tensor vote from a pixel (x, y) at a neighboring location (u, v) can then be computed by (3).

Again, from Fig. 2 OP 's distance l and angle θ , s is the arc length and κ is the curvature. \vec{N} is the vote direction and $A(\sigma, \theta)$ is the limiting cone that, as in the Iterative Voting, extends longitudinally up to a radius which is a function of σ , and transversally of the angle θ . c is the same as in the tensor voting framework

$$V(u, v) = e^{-(s^2 + c\kappa^2)/\sigma^2} \vec{N} \vec{N}^T \times \begin{cases} 1, & \text{if } (u, v) \in A(\sigma, \theta) \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

The iterations can be stopped after 1) a predetermined number of times, 2) the quantized aperture of the voting field is small enough, producing the same field as in a previous iteration, or 3) $\theta \leq 0$.

C. Curve Inference

Iterative voting produces a saliency map where curvilinear structures are highly enhanced. Lines can be segmented from the saliency maps using any standard morphological thinning or nonmaxima suppression technique, exploring magnitude and direction of the resulting saliency map. The outcome of this process is an image whose curvilinear structures are inferred. Fig. 7 depicts the entire process.

D. Application to Segmentation of Adherens Junctions

Fig. 8 shows an example of the adherens junctions imaged at $40 \times$ magnification and the intermediate results of tensor voting. It is clear that the proposed method regularizes and enhances punctate signals iteratively. In this context, the funneling process groups pixels belonging to the cellular structure with improved precision. The first iteration, produced by the ball voting, results in a diffused pattern. However, this is the initial condition where oriented stick tensor fields brings focus to the promising parts of the signal.

Figs. 9 and 10 show in more detail the nature of the imaged cell adherens junctions and sample results from the ITV iterations.

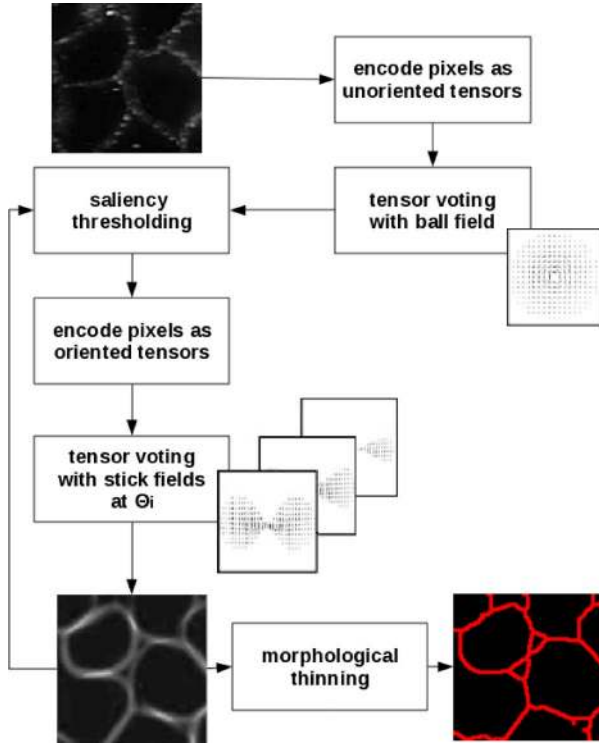


Fig. 7. Inference of curvilinear structures by ITV. The primary theme is the feedback loop for refinement of the voting aperture.

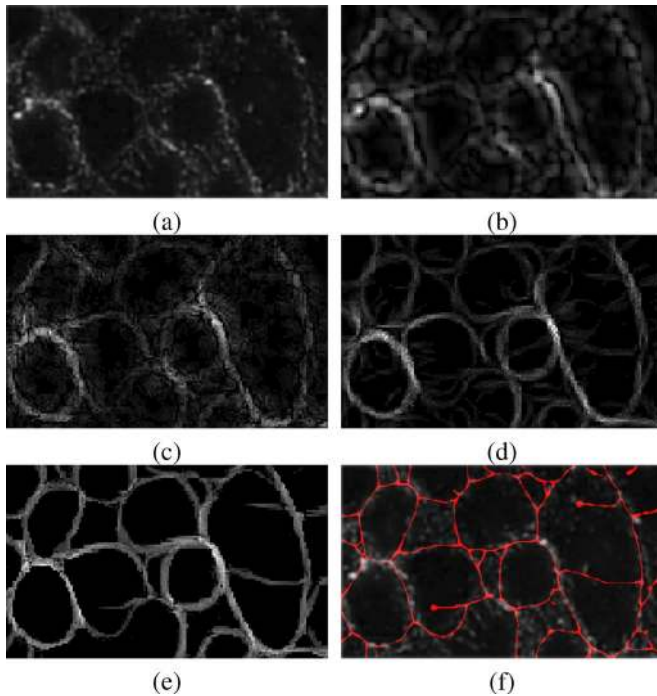


Fig. 8. Grouping of adherens junctions through ITV. (a) Original signal. (b) Result of first iteration. (c) and (d) Examples of intermediate iterations. (e) Grouped curvilinear structures. (f) Inferred structures.

IV. EVALUATION

In this section, we demonstrate the performance of the iterative tensor voting (ITV) for grouping of perceptual curvilinear structures, where validation has been performed on synthetic

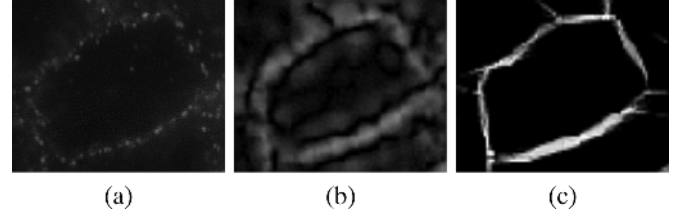


Fig. 9. Close view of grouping through ITV. (a) Original signal. (b) First iteration (ball voting). (c) Final iteration (at $\theta = 5^\circ$).

and annotated real data. The annotated data correspond to samples that have been imaged through fluorescence microscopy.

A. Synthetic Data

1) *Experimental Design*: In order to test ITV's applicability and limitations, a set of synthetic configurations were created. The synthetic data (top row) were generated to emulate a variety of spatially distributed signals, which are found in real data. Such a variety corresponds to the geometry of signal formation, density of punctation, signal width, and signal intensity. Figs. 11 and 12 show samples of synthetic data that have been used for this study. The first column of Fig. 11 shows the prototype geometry and signal composition through a 1) straight line, 2) circle, 3) figure eight-shaped object with an X-junction, in addition to a 4) figure B-shaped object with a T-junction and a Y-junction. The data set has 48 images: four figures \times 3 widths \times 4 punctation densities. All images are 256×256 pixels with the intensity in the range of $[0, 1]$. The processing parameters are $\sigma_{\text{ball}} = 10$, $\text{Threshold}_{\text{ball}} = 40\%$, $\sigma_{\text{stick}} = 3$, $\text{Threshold}_{\text{stick}} = 5\%$, $\Delta\theta = 5^\circ$.

2) *Qualitative Analysis*: Fig. 13 shows the performance of ITV on three synthetic configurations. The first and last rows show the original configuration and the inferred curvilinear structure, respectively. The intermediate rows demonstrate stages of ITV. Rows 2–5 show saliency maps at different iterations, namely, ball voting and stick voting at $\theta = 45^\circ$, $\theta = 30^\circ$ and $\theta = 15^\circ$. The results combine highly salient pixels from ITV's stick or ball saliency maps. This is important for better preservation of junctions, whose conflicting orientations locally reduce the stick saliency (i.e., $\lambda_1 - \lambda_2$) and increase the ball saliency (i.e., λ_2), as mentioned in Section III. It is clear that ITV can evolve from a punctate pattern to a strong filament. In this context, the dense voting allows pixels to be interpolated, thus, playing an important role in gap filling.

The synthetic data also allowed us to analyze ITV's sensitivity to punctation. Fig. 14 shows ITV's inferred structure as a function of the punctation density. In general, one can observe that ITV performs fairly well in all cases where the punctate density is higher or equal to 0.05 points/pixel² [Fig. 14(b)–(d)]. The highly punctate patterns of Fig. 14(a) can cause noisy inference of contours with potential gaps.

3) *Quantitative Analysis*: We compared ITV's performance quantitatively on the synthetic data set and compared it with its two precursory approaches: tensor voting (TV) and iterative voting (IV). For TV, $\sigma = 10$ and $\text{Threshold} = 40\%$. For IV,

²Threshold values are computed from the maximum saliency at each iteration to remove tensors with lower saliency.

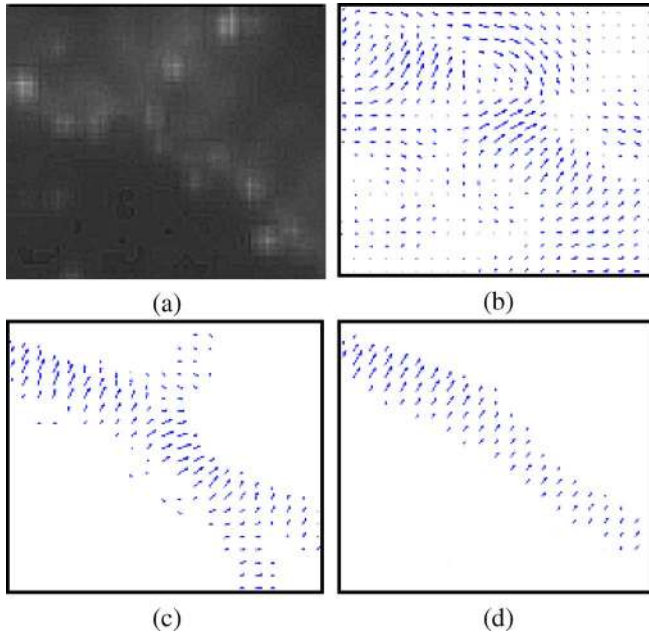


Fig. 10. Progressive refinement of the voting landscape (tensors' largest components) by ITV. (a) original image. (b) First iteration (ball voting). (c) Four iterations (at $\theta = 30^\circ$). (d) Final iteration (at $\theta = 5^\circ$).

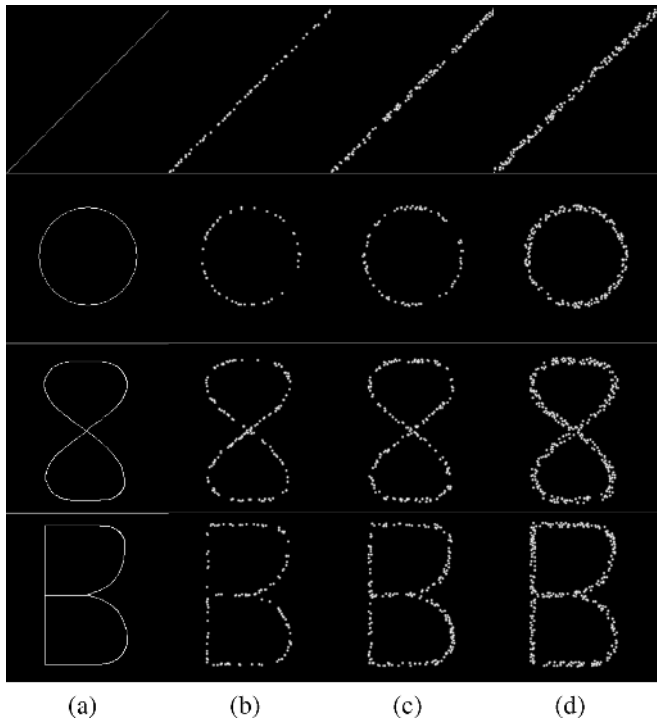


Fig. 11. Synthetic data with fixed punctate density (0.09 points/pixel²) changing as a function of line width. (a) Prototype signal. (b) Three pixel-wide punctate signal. (c) Five pixel-wide punctate signal. (d) Seven pixel-wide punctate signal.

$\sigma = 10$, $\delta_{\text{initial}} = 30^\circ$, $\Delta\theta = 10^\circ$, Threshold = 1%; All parameters, including ITV's, were defined by visual inspection of sample results, and used across all images in the data set and throughout the experiments in this paper. The percentage of mismatches between the methods' inferred structure and the

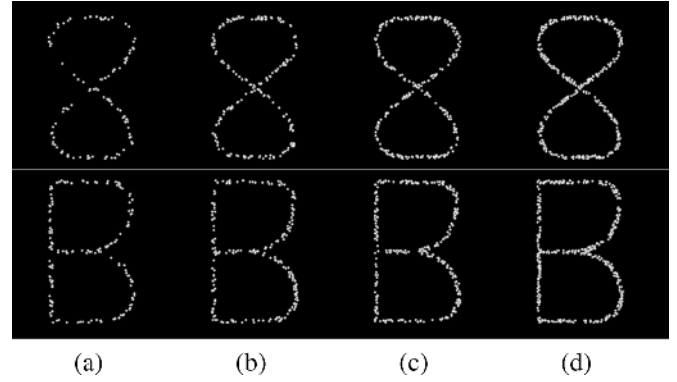


Fig. 12. Synthetic data with fixed width (five pixels) changing as a function of punctation density. (a) 0.03 points/pixel². (b) 0.05 points/pixel². (c) 0.07 points/pixel². (d) 0.09 points/pixel².

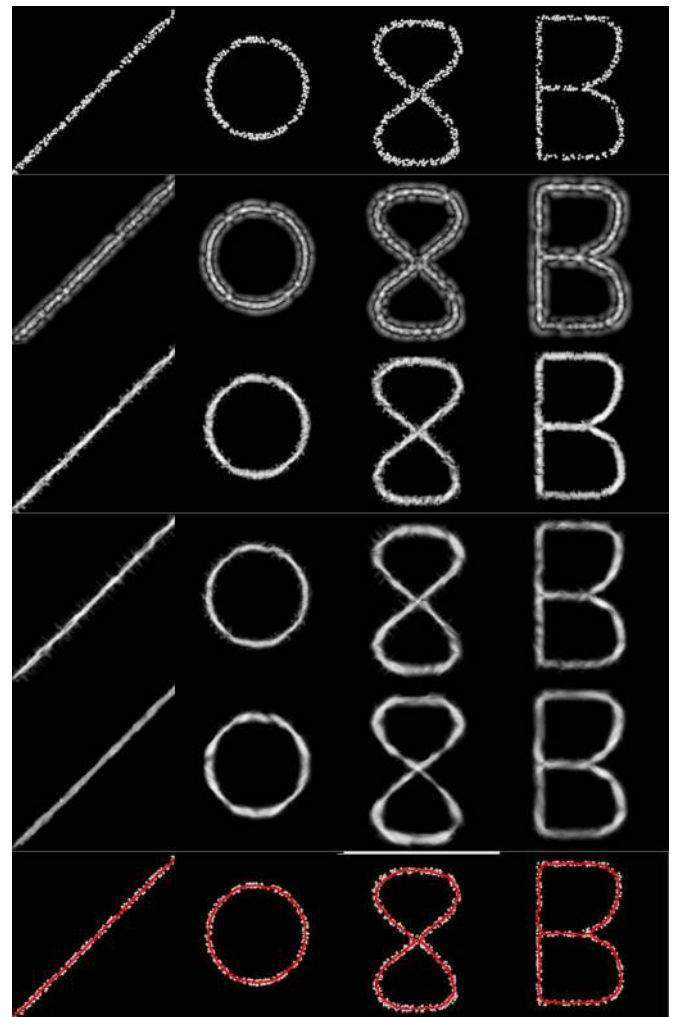


Fig. 13. Progressive refinement and final delineation of the synthetic data by ITV. First row: punctate signals. Second row: ball voting results. Third to fifth rows: intermediate results with stick fields. Last row: inferred structures over initial signal.

prototype signal at the correspondent width is used for quality assessment of the results.

Fig. 15 provides comparative results of each method on a dataset. Our analysis indicates that ITV provides more consistent results than its precursory approaches. A more detailed

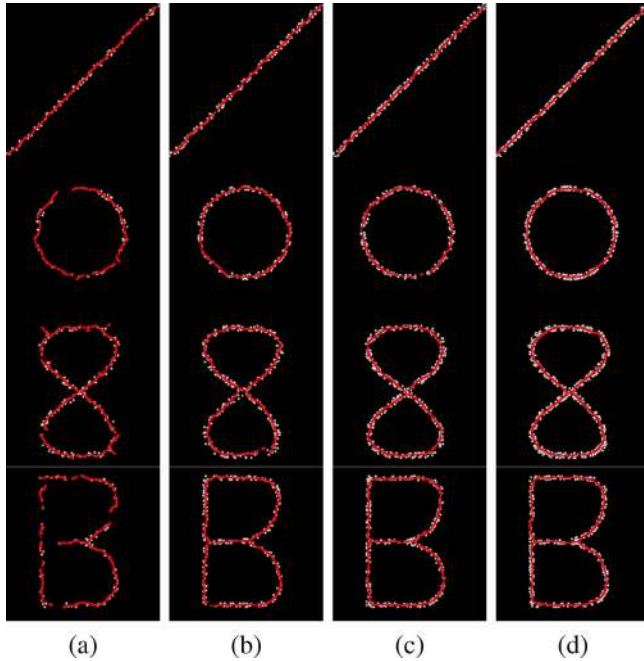


Fig. 14. Qualitative analysis of ITV's sensitivity to signal punctuation. (a) 0.03 points/pixel². (b) 0.05 points/pixel². (c) 0.07 points/pixel². (d) 0.09 points/pixel². ITV performs fairly well at 0.05 points/pixel² and above (b)–(d).

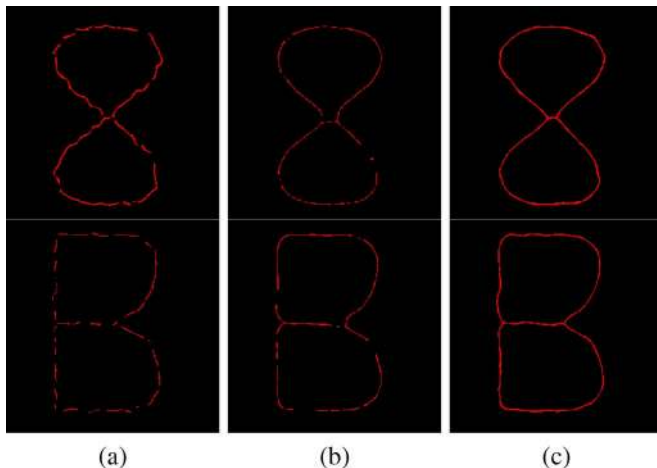


Fig. 15. Qualitative comparisons between methods. Sample results produced by (a) IV, (b) TV, and (c) ITV. ITV produces smoother and more consistent curves than its precursor methods.

analysis of the visual results indicate that TV was incapable of closing big gaps along the contour. Although increasing the scale σ for tensor voting could potentially facilitate gap closure, it would also compromise the high frequency features, such as sharp corners and junctions. On the other hand, each iteration of ITV allows the curve to grow progressively and eventually close the gap. Furthermore, due to the highly punctate aspect of the signals, IV seems to fail to update the kernel's orientation, producing misaligned fragments along the figure's contour. Whereas, ITV's iterations allow the pixels' orientation to converge slowly (i.e., with low $\Delta\theta$) and robustly towards the correct solution, producing smoother curves. Table I summarizes the results obtained by each method according to the signal width and punctuation density.

TABLE I
QUANTITATIVE RESULTS FOR SYNTHETIC DATA

Method	Line Width (pixels)	Punctuation Density (points/pixel ²)			
		0.03	0.05	0.07	0.09
IV	3	54%	65%	78%	82%
	5	59%	68%	79%	83%
	7	61%	73%	81%	86%
TV	3	64%	70%	82%	90%
	5	64%	73%	88%	90%
	7	65%	78%	89%	91%
ITV	3	91%	97%	99%	100%
	5	93%	98%	100%	100%
	7	93%	98%	100%	100%

B. Annotated Images

In this section, we evaluated the performance of ITV against manually annotated images. In this context, samples were stained for their adherens junctions, and then imaged with a $40\times$ magnification objective.³ The data set consisted of 274 1344×1024 images, 14 of which were annotated by two cell biologists. These images were selected for annotation for their diversity and not their similarity, in order to capture important biological and technical heterogeneity.^{4 5} The annotations served as representatives in human performance and as a baseline for algorithmic comparisons. Subsequently, the remainder of this section is dedicated to the evaluation of the performance of ITV, both qualitatively and quantitatively.

From a qualitative perspective, ITV generates contours with high agreement with human annotations. However, it does not properly represent nor detect junctions. Fig. 17 illustrates a sample of results, where the left and right columns, respectively, show the original images and human annotations that are overlaid with computed representation. The red and green contours represent the two annotations, while the white contours are the ITV results. Fig. 18 shows substandard performance of ITV at junctions. The rationale is that tensors, at the junction, do not deform in an elongated fashion.

Next, a quantitative analysis is performed where precision and recall rates are computed from the (mis)matches between human annotations and automatic detected contours. In this context, precision measures the probability of a detected contour to determine a true adherens junction. Recall measures the probability of an adherens junction to be correctly detected. In order to account for digital displacement of the contours, matches were evaluated at different versions of the annotation, with each of them being dilated by a factor that varied between 1 to 20 pixels.⁶ Fig. 19 shows the precision and recall rates obtained from averaging matching rates between the two annotations. In this section, HA refers to the Human Annotation, the performance achieved by the specialists, which was computed by averaging reciprocal (mis)matches between the annotations. Ad-

³Detailed information about the experimental design and imaging protocol can be found in [38].

⁴Statistically speaking, each image consists of over 150 cells, and every cell can be considered as an independent test tube.

⁵The only instruction given to the biologists was to trace adherens junctions they could perceive from the image. No constraints regarding closed contours or any other higher level inference was imposed.

⁶20 pixels served as the upper bound for being the actual average width observed for the signals analyzed.

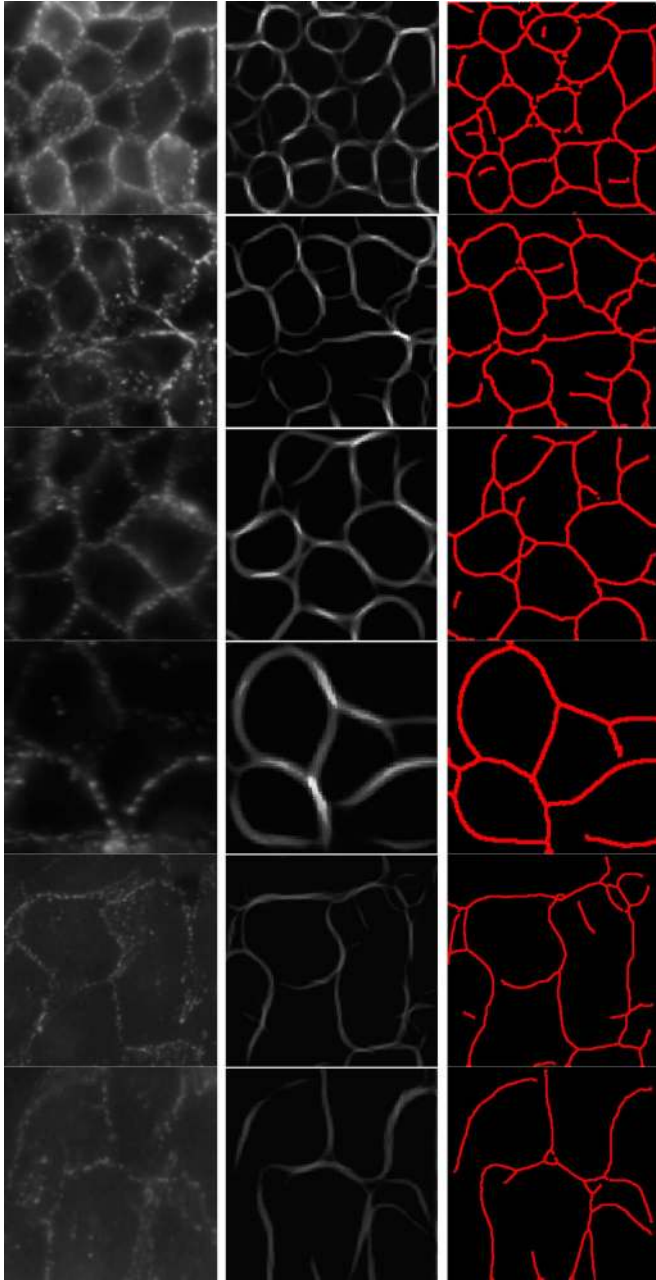


Fig. 16. Grouping and delineation of adherens junctions by ITV. Left: original signal. Center: grouped signal. Right: inferred adherens junctions.

ditionally, NC indicates the performance for negative curvature maxima (i.e., negative value of the maximum principal curvature). NCIV stands for iterative voting initialized with NC, and NCTV for tensor voting initialized with NC. Lastly, TV refers to tensor voting (ball voting only).

An analysis of the matching rates in Fig. 19 revealed that ITV developed the closest recall-precision curve to HA. It also showed that ITV had similar improvement rates to those produced by the specialists, presenting similar decay across the dilation factor. Furthermore, comparing the rate of agreement (RA) between the methods and HA, we can conclude that ITV had the highest agreement with the specialists, showing 91% precision and 94% recall rates of agreement (Table II). RA was obtained from averaging the sum of differences of results—precision (P) or recall (R)—from each dilated version ($P(i)$ or

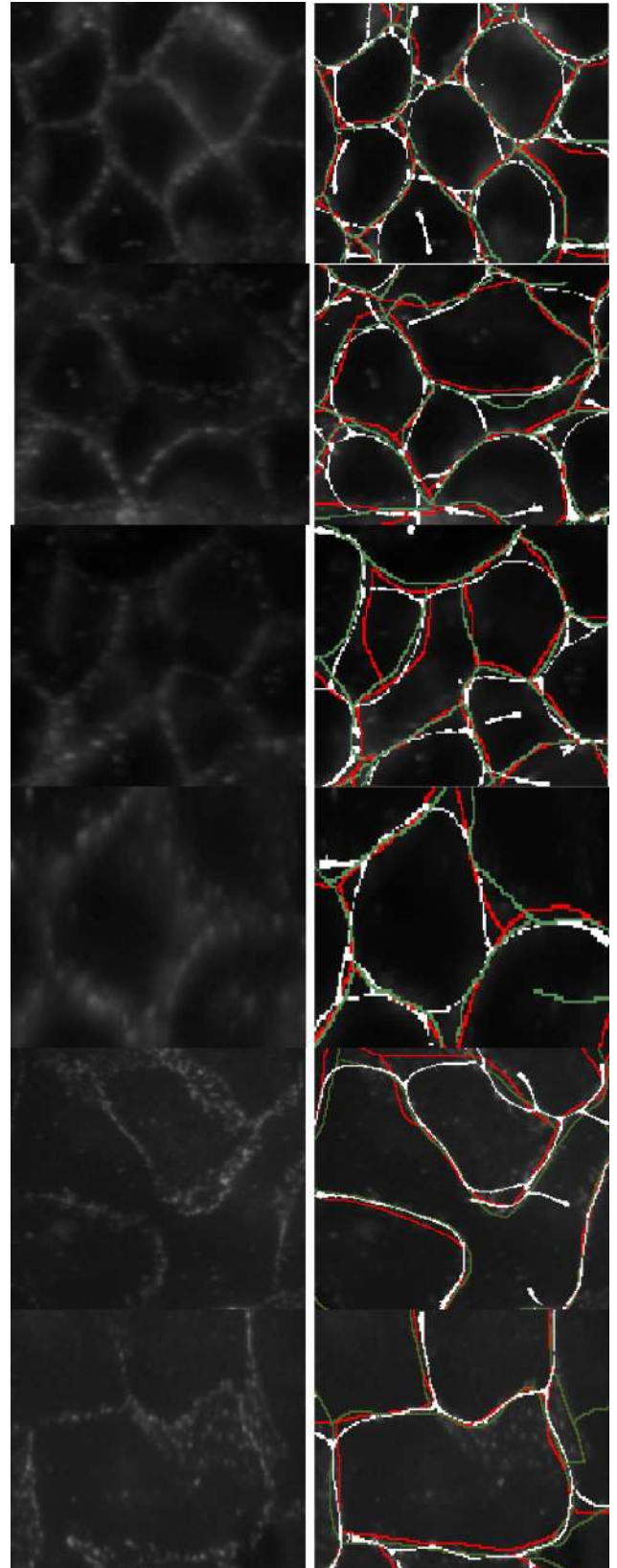


Fig. 17. Human delineation of adherens junctions versus ITV. Left: patches of the original images. Right: delineation by ITV (white) and the specialists (red/green).

$R(i)$, where i is the dilation factor) of the annotation and the method being evaluated. Equation (4) shows how RA is computed for P . For its counterpart R , $P(i)$ is replaced by $R(i)$.

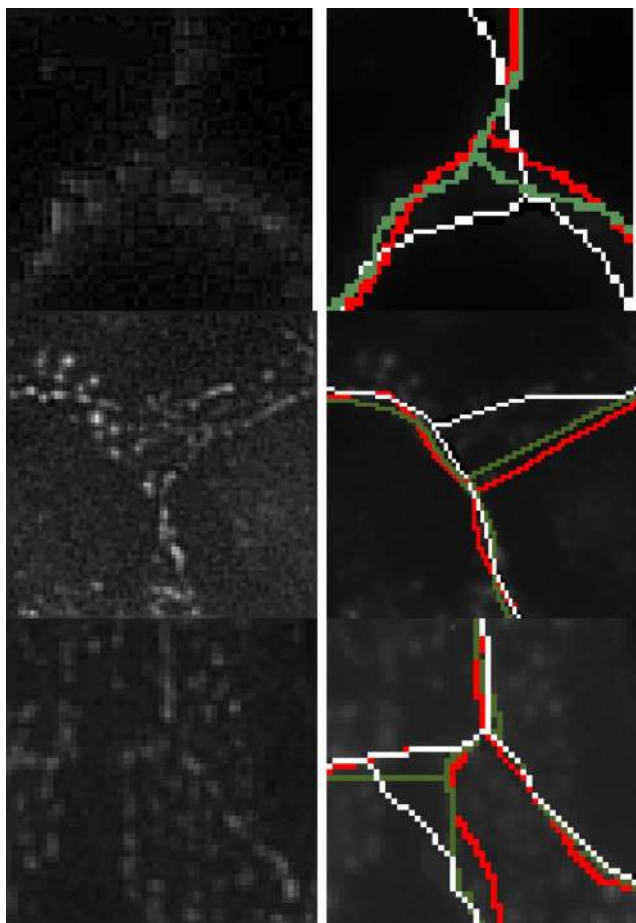


Fig. 18. Contour junctions are the main sources for misdetection. Left: location near junctions evidenced. Right: delineation by ITV (white) and the specialists (red/green).

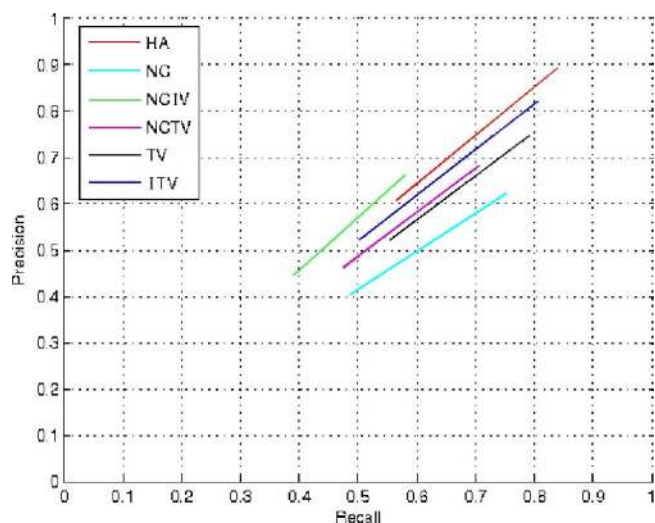


Fig. 19. Recall versus precision plot of the overall performance as a function of the dilation factor. Comparison between human annotation (HA), negative curvature maxima (NC), iterative voting initialized with NC (NCIV), tensor voting initialized with NC (NCTV), tensor voting (TV), and the ITV are shown. ITV not only performs better than the other methods but also produces a 91% rate of agreement with HA in precision and 94% in recall.

N is the maximum dilation factor 20. A detailed analysis of these performances shows that any postprocessing of NC and

TABLE II
RATE OF AGREEMENT FOR PRECISION (RA_P) AND RECALL (RA_R)
BETWEEN THE METHODS AND THE HUMAN ANNOTATION

	RA_P	RA_R
NC	75%	88%
NCIV	80%	78%
NCTV	81%	84%
TV	86%	95%
ITV	91%	94%

TV improved their results (compare, for instance, NCIV and NCTV with NC, and ITV with TV). However, one can notice that the initialization played an important role in all methods. TV had a better recall rate than NC, meaning it better resembled the human annotation. Therefore, another conclusion is that TV served as better input than NC

$$RA_P = 1 - \left[\frac{1}{N} \sum_{i=1}^N |P_{HA}(i) - P_{method}(i)| \right]. \quad (4)$$

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced the ITV method to group ill-defined visual signals along a curvilinear structure. This method coupled the tensor and iterative voting frameworks to leverage advantages from both methods. As a result, complex patterns along a curvilinear path could be perceptually grouped and delineated. Results on inference of adherens junctions (i.e., membrane-bound macromolecules) were validated through comparison with synthetic and annotated real data by expert cell biologists. Quantitatively, the method was shown to produce 1) superior results when compared with prior techniques, and 2) delineations comparable to annotations produced by specialists. In particular, our method achieved precision and recall rates of 91% and 94%, respectively. One limitation is that junctions (e.g., T-junctions, X-junctions) are not well characterized, which will be the subject of future efforts. An *ad hoc* junction detector, for example, can be feasibly added through the analysis of the ball saliency map produced by tensor voting. In addition, we plan to extend our method to confocal microscopy to infer this class of signals in 3-D.

REFERENCES

- [1] I. Biederman, *Human Image Understanding: Recent Research and Theory*. New York: Academic, 1985.
- [2] M. Wertheimer, "Laws of organization in perceptual forms (English title)," *II. Psychologische Forschung*, vol. 4, pp. 301–305, 1923.
- [3] Y. Hirogaki, T. Sohmura, H. Satoh, J. Takahashi, and K. Takada, "Complete 3-D reconstruction of dental cast shape using perceptual grouping," *IEEE Trans. Med. Imag.*, vol. 20, no. 10, pp. 1093–1101, Oct. 2001.
- [4] J. Jia and C. K. Tang, "Inference of segmented color and texture description by tensor voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 771–786, Jun. 2004.
- [5] S. Bonneau, M. Dahan, and L. Cohen, "Single quantum dot tracking based on perceptual grouping using minimal paths in a spatiotemporal volume," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1384–1395, Sep. 2005.
- [6] G. Harit and S. Chaudhury, "Video shot characterization using principles of perceptual prominence and perceptual grouping in spatio-temporal domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 12, pp. 1728–1741, Dec. 2007.
- [7] D. Marr, "Early processing of visual information," *Phil. Trans. R. Soc. London*, vol. 275, no. 942, pp. 483–519, 1975.

- [8] D. G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artif. Intell.*, vol. 31, pp. 355–395, 1987.
- [9] S. Ullman, "Filling-in the gaps: The shape of subjective contours and a model for their generation," *Biol. Cybern.*, vol. 25, 1976.
- [10] J. Weickert, "Coherence-enhancing diffusion filtering," *Int. J. Comput. Vis.*, vol. 31, no. 2, pp. 111–127, 1999.
- [11] P. Parent and S. W. Zucker, "Trace inference, curvature consistency, and curve detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 8, pp. 823–839, Aug. 1989.
- [12] L. Héroult and R. Horaud, "Figure-ground discrimination: A combinatorial optimization approach," *IEEE Trans. Pattern Recognit. Mach. Intell.*, vol. 15, no. 9, pp. 899–914, Sep. 1993.
- [13] F. Heitger and R. von der Heydt, "A computational model of neural contour processing: Figure-ground segregation and illusory contours," in *Proc. Int. Conf. Comput. Vis.*, 1993, pp. 32–40.
- [14] S. X. Li, H. X. Chang, and C. F. Zhu, "Fast curvilinear structure extraction and delineation using density estimation," *Comput. Vis. Image Understand.*, vol. 113, no. 6, pp. 763–775, 2009.
- [15] S. Sarkar and K. Boyer, "A computational framework for preattentive perceptual organization: Graphical enumeration and voting methods," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 2, pp. 246–267, Feb. 1994.
- [16] L. Loss, G. Bebis, and B. Parvin, "Tunable tensor voting for regularizing punctate patterns of membrane-bound protein signals," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 1378–1381.
- [17] L. Loss, G. Bebis, and B. Parvin, "Tunable tensor voting improves grouping of membrane-bound macromolecules," in *Proc. IEEE Math. Methods Biomed. Image Anal.*, 2009, pp. 72–78.
- [18] S. Sarkar and K. Boyer, "Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1996, pp. 478–483.
- [19] G. Gerig and F. Klein, "Fast contour identification through efficient Hough transform and simplified interpretation strategy," in *Proc. Int. J. Conf. Pattern Recognit.*, 1986, pp. 498–500.
- [20] E. Escolar, G. Weigold, A. Fuisz, and N. Weissman, "New imaging techniques for diagnosing coronary artery disease," *Can. Med. Assoc. J.*, vol. 174, no. 4, pp. 487–495, 2006.
- [21] N. L. Prigozhina, L. Zhong, E. A. Hunter, I. Mikic, S. Callaway, D. R. Roop, M. A. Mancini, D. A. Zachariasa, J. H. Price, and P. M. McDonough, "Plasma membrane assays and three-compartment image cytometry for high content screening," *Assay Drug Dev. Technol.*, vol. 5, no. 1, pp. 29–48, 2007.
- [22] P. Diagaradjane, J. Cardona, N. Casasnovas, A. Deorukhkar, S. Shentu, N. Kuno, D. Schwartz, J. Gelovani, and S. Krishnan, "Imaging epidermal growth factor receptor expression in vivo: Pharmacokinetic and biodistribution characterization of a bioconjugated quantum dot nanoprobe," *Clin. Cancer Res.*, vol. 14, no. 3, pp. 731–741, 2008.
- [23] G. Guy and G. Medioni, "Inferring global perceptual contours from local features," *Int. J. Comput. Vis.*, vol. 20, no. 1, pp. 113–133, 1996.
- [24] G. Medioni, M. S. Lee, and C. K. Tang, *A Computational Framework for Feature Extraction and Segmentation*. New York: Elsevier, 2000.
- [25] P. V. C. Hough, "Machine analysis of bubble chamber pictures," in *Proc. Int. Conf. High Energy Accelerators Instrumentat.*, 1959.
- [26] B. Parvin, Q. Yang, J. Han, H. Chang, B. Rydberg, and M. H. Barcellos-Hoff, "Iterative voting for inference of structural saliency and characterization of subcellular events," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 615–623, Mar. 2007.
- [27] H. Chang, K. L. Andarawewa, J. Han, M. H. Barcellos-Hoff, and B. Parvin, "Perceptual grouping of membrane signals in cell-based assays," in *IEEE Int. Symp. Biomed. Imag.*, 2007, pp. 532–535.
- [28] G. Medioni and S. B. Kang, *Emerging Topics in Computer Vision*. Upper Saddle River, NJ: Prentice Hall, 2004, ch. 5.
- [29] J. Lim, J. Park, and G. Medioni, "Text segmentation in color images using tensor voting," *Image Vis. Comput.*, vol. 25, no. 5, pp. 671–685, 2007.
- [30] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM*, vol. 15, pp. 11–15, 1972.
- [31] L. Williams and D. Jacobs, "Local parallel computation of stochastic completion fields," *Neural Computat.*, vol. 9, pp. 859–881, 1997.
- [32] S. Fischer, P. Bayerl, H. Neumann, R. Redondo, and G. Cristobal, "Iterated tensor voting and curvature improvement," *Signal Process.*, vol. 87, no. 11, pp. 2503–2515, 2007.
- [33] L. Loss, G. Bebis, M. Nicolescu, and A. Skurikhin, "An iterative multi-scale tensor voting scheme for perceptual grouping of natural shapes in cluttered backgrounds," *Comput. Vis. Image Understand.*, vol. 113, no. 1, pp. 126–149, 2009.
- [34] W. S. Tong, C. K. Tang, P. Mordohai, and G. Medioni, "First order augmentation to tensor voting for boundary inference and multiscale analysis in 3-D," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 594–611, May 2004.
- [35] W. S. Tong and C. K. Tang, "Robust estimation of adaptive tensors of curvature by tensor voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 434–449, Mar. 2005.
- [36] S. Mahamud, L. Williams, K. Thornber, and K. Xu, "Segmentation of multiple salient contours from real images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 433–444, Apr. 2003.
- [37] L. Risser, F. Plouraboue, and X. Descombes, "Gap filling of 3-D microvascular networks by tensor voting," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 674–687, May 2008.
- [38] J. Han, H. Chang, K. Andarawewa, P. Yaswen, M. H. Barcellos-Hoff, and B. Parvin, "Multidimensional profiling of cell surface proteins and nuclear markers," *IEEE Trans. Computat. Biol. Bioinformat.*, vol. 7, no. 1, pp. 80–90, Jan./Mar. 2010.