An Improved Representation of Junctions through Asymmetric Tensor Diffusion

Shawn Arseneau and Jeremy R. Cooperstock

Centre for Intelligent Machines
McGill University, Montreal, Canada
arseneau, jer@cim.mcgill.ca

Abstract. Junctions form critical features in motion segmentation, image enhancement, and object classification to name but a few application domains. Traditional approaches to identifying junctions include convolutional methods, which involve considerable tuning to handle non-trivial inputs and diffusion techniques that address only symmetric structure. A new approach is proposed that requires minimal tuning and can distinguish between the basic, but critically different, 'X' and 'T' junctions. This involves a multi-directional representation of gradient structure and employs asymmetric tensor diffusion to emphasize such junctions. The approach combines the desirable properties of asymmetry from convolutional methods with the robustness of local support from diffusion.

1 Introduction

Extracting high-level structure from image gradients is central to many computer vision applications such as data interpolation, 3D scene reconstruction, image enhancement, motion segmentation, and biometrics. Each requires a local description of structure that includes contours and junctions. For example, processing laser-rangefinder data or a stereo-depth map may involve interpolation of sparse data, minimizing the effects of noise, and segmenting this information into distinct objects [1]. Similar issues are found in the context of image enhancement as contours and junctions denote regions where smoothing should be inhibited [2]. In motion segmentation, identifying junctions in the spatio-temporal domain indicate points of occlusion in a video sequence [3]. Similarly, junctions are used to determine salient keypoints in fingerprints or defects in lumber [4, 5].

One approach to highlighting junctions is to apply *diffusion*, where local information is distributed to its neighbors conditioned on specific parameters and replaced with the consensus from that data. For example, isotropic diffusion applies local averaging, weighted by relative proximity, to produce a blurring effect. Most diffusion methods make use of gradient information represented by a structure tensor [6]. Although it has several benefits, the structure tensor is limited in that it may only represent gradient in a *symmetric*, or π-periodic form. This implies that diffusion using such a form also results in symmetric information, thus preventing the distinction between 'X' and 'T' junctions from being made. A method to convert this symmetric information into a richer asymmetric

form had yet to be incorporated into the diffusion framework. A novel, two-step solution is presented in Section 3, which first transforms the symmetric gradient information into a directional voting field representation and second, iteratively applies asymmetric tensor diffusion. The results of this approach are evaluated experimentally and contrasted with various competing methods in Section 4. Finally, several applications and future work are discussed in Section 5. Before describing the details of our approach, some terminology is defined and a brief review of previous convolution and diffusion approaches is provided in Section 2.

2 Background

To motivate this work, we begin with a review of convolution-based and diffusion approaches with an emphasis on junction analysis. For consistency of terminology, we refer to *direction* as the angle of a vector with respect to the x-axis, ranging between $[0, 2\pi)$ while *orientation* is π-periodic ranging between $[0, \pi)$. Symmetry in our work refers to the geometric interpretation with respect to gradient-based contours and does not refer to the concept of symmetric matrices.

2.1 Convolution-Based Approaches

Early work in the area of convolution-based, directional distribution functions (DDFs) began with the use of Gabor filters [7]. The DDF is created by convolving rotated versions of the kernel at discretely sampled angles and incrementing their respective, angular bins similar to orientation histograms [8]. (examples of DDFs are shown in figure 3) Although the Gabor uses a quadrature pair to address both even and odd-phased gradients, its form is symmetric thus preventing the distinction between 'T' and 'X' type junctions directly. Asymmetric kernels were proposed to highlight such distinctions. For example, Gaussian derivatives were used to derive logical/linear operators and one-sided filter pairs [9, 10], while later work by Simoncelli and Farid improved on the accuracy by designing a set of polar-based Gabor kernels, known as wedge-filters [11].

The DDF maxima for these methods do not necessarily imply gradient structure along the direction of the maxima as they are template-matching approaches at their core: implying that they are best suited to finding matches between patterns and not necessarily designed to identify gradient structure [12]. Although steerability has been explored in the use of such approaches, [10], they also require a large bank of filters to address different spatial frequencies [13]. Improved results were obtained using the rotated averaging wedge method (RWAM), which calculated average pixel values within wedge-shaped regions and generated the DDF as the 1D derivative of these values [14]. More recent work by Michelet et al. used a homogeneity function based on an asymmetric sampling grid to populate the DDF, albeit without the benefit of local support through diffusion and also requiring considerable parameter tuning [15]. Although these approaches perform well on trivial, step-edge images, they are inappropriate for estimating gradient direction on more complex data such as that in Figure 3a.

2.2 Diffusion-Based Approaches

The convolution-based approaches described in section 2.1 apply a kernel to the data at a given scale to create the DDF. An alternative is to propagate gradient information from all pixels to their corresponding neighbors. This process, known as diffusion or regularization, maintains a balance between the original information through the data consistency term, while biasing the local model using the diffusion term [2]. The gradient data is best represented using a structure tensor as it encodes not only the orientation and magnitude but the coherence, as in Equations (2-4):

$$S = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{1}$$

$$\theta = \tan^{-1}\left(\boldsymbol{e}_{1y}/\boldsymbol{e}_{1x}\right) \tag{2}$$

$$|S| = (\lambda_1 - \lambda_2) \tag{3}$$

$$\varsigma = \begin{cases} \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 & if\ \ (\lambda_1 + \lambda_2) > 0 \\ 0 & otherwise \end{cases} \tag{4}$$

where I_x is the partial derivative of image $I(x, y)$ with respect to x, ($\boldsymbol{e}_1$, $\boldsymbol{e}_2$,λ_1, λ_2) represent the eigenvectors and values from the decomposition of structure tensor S respectively and ς is the coherence measure as outlined by Jähne that provides a measure of certainty in gradient direction along $\boldsymbol{e}_1$ with respect to $\boldsymbol{e}_2$ [6].

Diffusion techniques can take many forms. For example, isotropic diffusion is a common approach where data is propagated to its neighbors based solely on relative proximity. This method reduces noise at the expense of maintaining high gradient (edge) information. Anisotropic diffusion, reviewed in detail by Tschumperlé and Deriche [2], preserves edges by restricting smoothing across high gradient regions [16].

Orientation diffusion, enforces the periodic nature of symmetric gradient information through the use of a specialized influence function [17]. Several other works refer to direction-based diffusion in the context of gradient polarity direction to bolster pixel-based feature points or in the framework of color enhancement [18, 19]. However, the focus of our work is on a phase-*independent* description of the gradient structure.

To account for more complex interaction between gradient structures, tensor voting was introduced [1, 20]. It not only diffuses based on proximity, but also on relative curvature as well as allowing for slowly-varying, orientational patterns with a biasing parameter to favor linear rather than curved contours. This approach is adept at handling both sparse and noisy data and requires minimal memory requirements by using a single tensor representation at each node. Relaxation labeling, which permits multiple representations [21], adds support to those pixel locations or *nodes* that have compatible structures based on such criteria as co-circularity, co-heliocity or the normal and tangential curvature components to the tensor fields [22, 21, 23].

3 Asymmetric Tensor Diffusion (ATD)

The previously described diffusion-based approaches directly apply symmetric information derived from the structure tensor, implying that the resulting DDFs are also symmetric. This section proposes a technique to transform the orientation of the structure tensor data into a directional-based voting field to allow for an asymmetric form at each node. Specifically, a voter distributes ballots to all of its neighbors (receivers). The ballots are collected into their respective DDFs, which are then used to seed a secondary, diffusion stage.

3.1 Stage One: Directional Voting Field

Early work in the conversion of orientation data into a meaningful, directional voting field was proposed in [24]. The present work serves not only to clarify and extend this concept but also to provide a means by which to properly diffuse such information. We first determine the orientation, magnitude and coherence of the structure tensor as per Equations (2-4). Next, an *inwardly* facing *directional bin field*, which represents the initial ballots and the spatial locations of their corresponding receivers, is constructed. Each ballot points towards the axis perpendicular to the orientation of the structure tensor, as per Equations 5-6.

$$B_{ij}\left(\theta_i, \varepsilon\right) = \begin{cases} \theta_i + \varepsilon\pi & \left|\theta_i - \varphi_{ij}\right| > \frac{\pi}{2} \\ \theta_i + (\varepsilon + 1)\,\pi & otherwise \end{cases} \tag{5}$$

$$\Psi_{ij}\left(\theta_i, \varepsilon\right) = \begin{cases} 1 & -\frac{1}{2}\tau \leq x_{ij} \leq \frac{1}{2}\tau \\ 1 - \varepsilon & otherwise \end{cases} \tag{6}$$

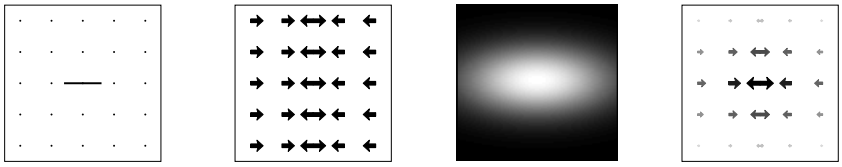
where ε=$\{0, 1\}$, B_{ij} and Ψ_{ij} denote the ballot direction and magnitude [1] respectively, θ_i is the orientation of the structure tensor at i, τ is the minimum distance between nodes, and φ_{ij} denotes the angle from voter i to receiver j with respect to the x-axis. A directional bin field created from a horizontally oriented input is illustrated in Figure 1b, where the ballots point inward towards the vertical axis. To account for the ambiguity in the original orientation, two opposing ballots are placed along the vertical axis, centered on the original data. The ballots are aligned *parallel* to the original orientation, rather than being steered toward the center point to prevent biasing at this early stage of processing.

The strength of each ballot sent from voter i to receiver j is then weighted by an anisotropic map, known as the region-of-influence (ROI) function, Λ_{ij}, aligned with the orientation of the original data:

$$\Lambda_{ij}\left(\theta_i, \varepsilon\right) = G\left(0, \sigma_x\right) \cdot \Psi_{ij}\left(\theta_i, \varepsilon\right) \cdot R\left(\theta_i\right) \tag{7}$$

where G is a 2D Gaussian with zero mean and $\sigma_y = q\sigma_x$ (Figure 1c) where q is the sigma ratio, and $R(\theta_i)$ is the rotation matrix.

[1] In essence, Ψ distinguishes between single- and double-ballot locations where the latter is assigned to points of orientational ambiguity.



(a) (b) (c) (d)

Fig. 1. Stage one: structure tensor with horizontal orientation and coherence=1 (a), B_{ij} (b), Λ_{ij} (c), and the directional voting field where both the vector magnitude and grayscale denote their relative influence (d).

Rather than summing the collected ballots at each receiver into a single value, a histogram of N directional bins is used to collect the ballots, as per Equation 8. While this requires greater memory than that of anisotropic diffusion, tensor voting, and relaxation labeling, it allows for the all-important representation of asymmetric structures.

$$DDF_j(\vartheta) = \sum_{\varepsilon=0}^{1} \sum_{i=1}^{\Omega} \tilde{S}_i \cdot \Lambda_{ij}(\theta_i, \varepsilon) \cdot m(\vartheta, B_{ij}(\theta_i, \varepsilon)) \tag{8}$$

$$m(p,q) = \begin{cases} 1 & p = q \\ 0 & otherwise \end{cases} \tag{9}$$

where ϑ denotes the directional bin, Ω the local neighborhood around j and $\tilde{S}_i$ the normalized version of the original structure tensor S_i. In brief, voter i sends a ballot $\tilde{S}_i$, weighted by Λ_{ij}, to bin $\vartheta = B_{ij}$ of receiver j.

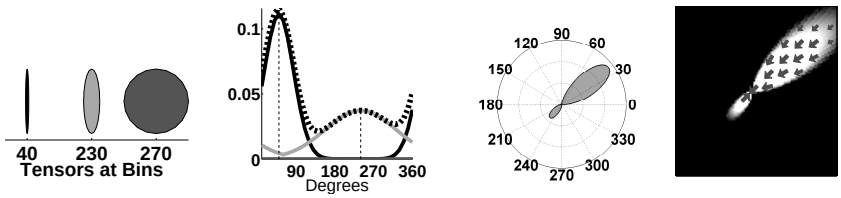
3.2 Stage Two: Iterative Diffusion of DDF

From stage one, the DDF of each receiver is represented by a single structure tensor per directional bin. This is transformed into a 1D-DDF prior to the diffusion process using a summation of 2π-periodic, normalized Gaussians, $G_{2\pi}(\mu, \sigma, x)$.

$$DDF(\vartheta) = \sum_{\beta=1}^{N} |S_\beta| \cdot G_{2\pi}\left(\theta_\beta, \sigma_{\varsigma_\beta}, \vartheta\right) \tag{10}$$

$$\sigma_\varsigma = (1 - \varsigma)(\sigma_{\max} - \sigma_{\min}) + \sigma_{\min} \tag{11}$$

Gaussians, normalized to unity area under the curve, are amplified by the magnitudes, $|S_\beta|$, where their means are centered at θ_β and variances are a function of coherence. The values of $(\sigma_{min}, \sigma_{max})$ were assigned empirically as $(0.25, 2)$ to vary between *certain* and *uncertain* estimates, where coherence is bounded between $[0, 1]$. An example of this transformation with three populated bins of successively decreasing coherence, is illustrated in Figure 2. The more elongated ellipses correspond to greater coherence values, which are reflected by Gaussians



(a) (b) (c) (d)

Fig. 2. Stage Two: Three tensors at directional bins 40^o (black), 230^o (light-gray) and 270^o (dark-gray) (a) and their corresponding Gaussians and their 1D-DDF (dashed-line) (b), polar-equivalent (c) and associated weighted, voter-facing, ballot field (d).

of lower variance. These are summed to form the 1D-DDF, which is then converted into a 2D weighting map with radial-based decay as per Equation 12 and illustrated in Figure 2d.

$$\hat{\Lambda}_{ij} = \begin{cases} 1 & \rho_{ij} < DDF\left(\varphi_{ij}\right) \\ \cos\left[\frac{\pi}{2}\left(\frac{\rho_{ij} - DDF(\varphi_{ij})}{\rho_{\max} - DDF(\varphi_{ij})}\right)\right] & DDF\left(\varphi_{ij}\right) \leq \rho_{ij} \leq DDF\left(\varphi_{ij}\right) + \rho_{\max} \\ 0 & otherwise \end{cases} \tag{12}$$

where the hat notation of $\hat{\Lambda}$ reflects the second stage, ρ_{ij} is the Euclidean distance between nodes i and j, and $\rho_{\max}$ denotes the degree of radial decay. Here, each receiver j, obtains a single ballot from voter i, corresponding to the DDF value at i pointing towards j. This ballot increments a directional bin at j that points towards i. We refer to this arrangement as the *voter-facing ballot field*. The DDF is then updated as:

$$DDF_j^t\left(\vartheta\right) = \alpha \cdot DDF_j^{t-1}\left(\vartheta\right) + (1-\alpha) \cdot \sum_{i=1}^{\Omega} \hat{\Lambda}_{ij} \cdot DDF_i^{t-1}\left(\varphi_{ij}\right) \cdot m\left(\vartheta, (\varphi_{ij} + \pi)\right) \tag{13}$$

where t refers to the iteration step and α denotes the diffusion coefficient. An example of the voter-facing ballot field is illustrated in Figure 2d.

Our approach offers many advantages. First, the only free parameters to tune are the sigma ratio q for G (empirically set to 1/2) and the scale that dictates the range over Ω. Second, by diffusing non π-periodic DDF information, local support is enforced at the pixel level. Finally, this method can also represent endpoints as well as curved structures using the DDF form.

4 Experimental Evaluation

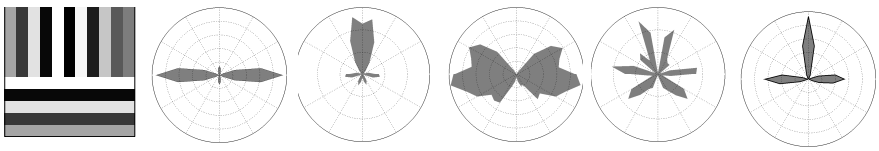
The ATD method is first compared to the convolution approaches against a T-junction image. Next, it is contrasted against the diffusion methods for two tensor field layouts. Finally, ATD is applied to two real-world applications. For all trials and algorithms, N=36.

4.1 Results Against Convolution Approaches

A test-image having asymmetric gradient information derived from several spatial frequencies was used, as shown in Figure 3a. All methods used a scale to match the image of 11x11 pixels. Parameters were tuned for best results against a step-type corner swatch: exemplifying the need for a bank of filters for convolution approaches. The Gabor identified the horizontal orientation; however, was incapable of distinguishing the lack of a downward gradient direction as the filters themselves are symmetric. The one-sided, wedge filter and RWAM fall victim to the presence of several pixel-values along orientations that do not radiate from the center of the image while the ATD properly depicts the three gradient directions.

4.2 Results Against Diffusion Approaches

The test sets used to compare with the previous diffusion methods depict an 'X' and 'T'-shaped structure tensor layout (Figures 4(a,g) respectively). Note that the DDFs shown are from the annotated points 'P' and 'Q'. For the isotropic, anisotropic and tensor voting methods, the DDF is represented by a single structure tensor and visualized as an ellipse oriented along $\boldsymbol{e_1}$ with major and minor radii corresponding to (λ_1, λ_2). From their results, it is not obvious whether it resulted from two, perpendicular tensors, or a single tensor with less certainty without further processing. While relaxation labeling was able to distinguish the two orientations, only the ATD could disambiguate between the two cases.



(a) (b) (c) (d) (e) (f)

Fig. 3. (a) Input image and resulting DDF's for (b) Gabor, (c) one-sided [10], (d) wedge filters [11], (e) RWAM [14] and (f) ATD

4.3 Results with Real-World Data

Junction structures are key to occlusion detection in video sequences as occlusion is denoted as splitting or merging of contours [3]. Using the *flower garden* sequence, a spatio-temporal slice was extracted and ATD applied. Close-ups of the annotated locations of Figure (5d) correspond to occlusion, disocclusion and no occlusion respectively. The scale used was 9x9 and DDFs reflect three iterations for this experiment.

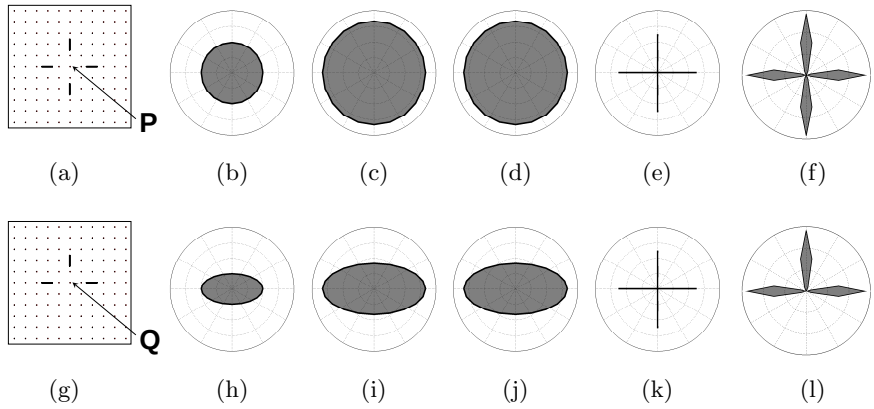


Fig. 4. Top row: X-junction test case and bottom row: T-junction test case. 1st column: initial test layouts, 2nd column: isotropic, 3rd column: anisotropic, 4th column: Tensor Voting [1], 5th column: Relaxation Labeling [22] and 6th column: ATD.

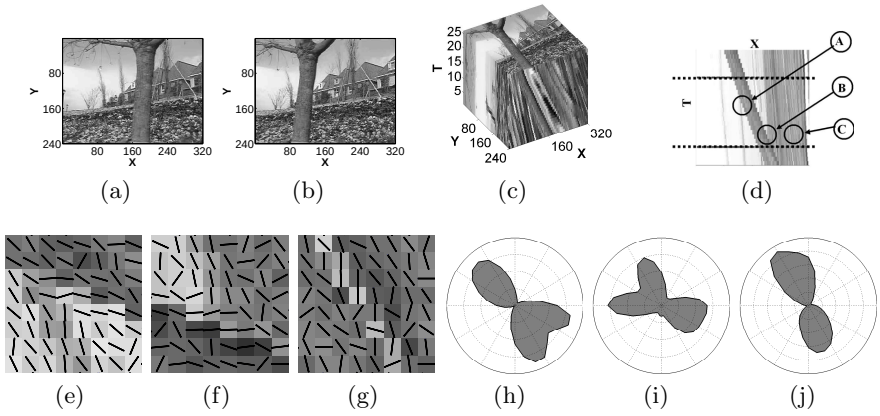


Fig. 5. Two images from the *flower garden* sequence (a,b) with the corresponding spatio-temporal volume (c), where a 2D slice is taken from y=80 (d), indicating three sample locations:'A','B','C' denoting occlusion, disocclusion and no occlusion respectively. The initial gradient orientations (e,f,g) and resulting DDFs from proposed approach (h,i,j) taken from the center of the sample regions

Junction are also important in fingerprint analysis. A trivial junction model was implemented as a proof-of-concept. For a given DDF, lobe-based features were depicted at DDF maxima with an associated saliency equal to the area of said lobe (integral between the maxima's left and right-wise local minima). Lobes having a saliency of less than 10% of the maximum saliency per DDF were trimmed. Figures (6(b-e)) depict the identification of nodes with a 1,2,3 and 4 lobes respectively.

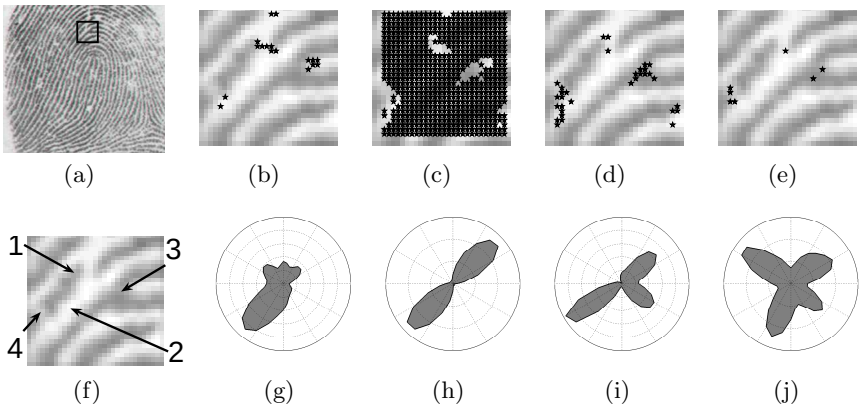


Fig. 6. Fingerprint (a) and its sub-image (f), junction classification for 1,2,3 & 4 lobes (b-e) respectively with DDF close-ups corresponding to labels (1-4)from (f) in (g-j).

5 Conclusion

A novel approach of transforming symmetric gradient information into asymmetric DDFs is proposed, along with a method by which to diffuse them. The accuracy is shown to be an improvement over current convolution-based approaches and the asymmetric representation allows for the distinction between non-π periodic structures. Future work will investigate a multi-scale as well as a 3D implementation.

References

1. Nicolescu, M., Medioni, G.: Motion segmentation with accurate boundaries - a tensor voting approach. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. Volume 1. (2003) 382–389
2. Tschumperlé, D., Deriche, R.: Diffusion pde's on vector-valued images. IEEE Signal Processing Magazine (2002) 16–25
3. Bolles, R., Baker, H., Marimont, D.: Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision **1** (1987) 7–55

4. Jiang, X.: On orientation and anisotropy estimation for online fingerprint authentication. IEEE Trans. On Signal Processing **53** (2005) 4038–4049
5. Rao, A., Jain, R.: Computerized flow field analysis: Oriented texture fields. IEEE Trans. On Pattern Analysis and Machine Intelligence **14** (1992) 693–709
6. Jähne, B.: Spatio-Temporal Image Processing: Theory and Scientific Applications. Volume 751. Springer-Verlag, Berlin (1993)
7. Adelson, E., Bergen, J.: Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America **2** (1985) 284–299
8. Freeman, W., Tanaka, K., Ohta, J., Kyuma, K.: Computer vision for computer games. In: 2nd International Conference on Automatic Face and Gesture Recognition. (1996) 100–105
9. Iverson, L., Zucker, S.: Logical/linear operators for image curves. IEEE Trans. on Pattern Analysis and Machine Intelligence **17** (1995) 982–996
10. Perona, P.: Steerable-scalable kernels for edge detection and junction analysis. In: Proc. European Conference on Computer Vision. (1992) 3–18
11. Simoncelli, E., Farid, H.: Steerable wedge filters for local orientation analysis. IEEE Trans. On Image Processing **5** (1996) 1377–1382
12. Malik, J., Belongie, S., Shi, J., Leung, T.: Textons, contours and regions: Cue integration in image segmentation. In: Proc. IEEE International Conference on Computer Vision. Volume 2. (1999) 918–925
13. Heeger, D.: Optical flow using spatiotemporal filters. International Journal of Computer Vision **1** (1988) 279–302
14. Yu, W., Daniilidis, K., Sommer, G.: Rotated wedge averaging method for junction characterization. In: Proc. IEEE Computer Vision and Pattern Recognition. (1998) 390–395
15. Michelet, F., Germain, C., Baylou, P., Costa, J.D.: Local multiple orientation estimation: Isotropic and recursive oriented network. In: Proc. IEEE International Conference on Pattern Recognition. Volume 1. (2004) 712–715
16. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Machine Intell. **33** (1990) 629–639
17. Perona, P.: Orientation diffusions. IEEE Trans. On Image Processing **7** (1998) 457–467
18. Tang, B., Sapiro, G., Caselles, V.: Direction diffusion. In: International Conference on Computer Vision. (1999) 1245–1252
19. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110
20. Tong, W., Tang, C., Mordohai, P., Medioni, G.: First order augmentation to tensor voting for boundary inference and multiscale analysis in 3d. IEEE Trans. On Pattern Analysis and Machine Intelligence **26** (2004) 594–611
21. Parent, P., Zucker, S.: Trace inference, curvature consistency, and curve detection. IEEE Trans. On Pattern Analysis and Machine Intelligence **11** (1989) 823–839
22. Ben-Shahar, O., Zucker, S.: Hue fields and color curvatures: A perceptual organization approach to color image denoising. In: IEEE Computer Vision and Pattern Recognition. Volume 2. (2003) 713–720
23. Savadjiev, P., Campbell, J., Pike, G., Siddiqi, K.: 3d curve inference for diffusion mri regularization. In: International Conference on Medical Image Computing and Computer Assisted Intervention. (2005) 123–130
24. anonymous for peer review: An asymmetrical diffusion framework for junction analysis. In: British Machine Vision Conference (to appear). (2006)