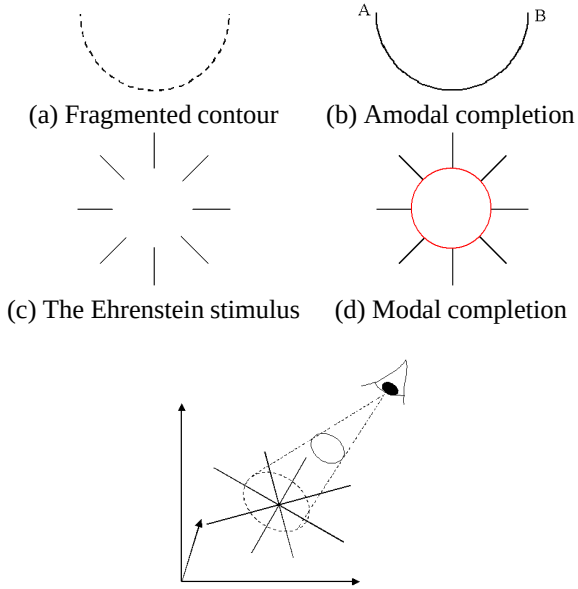
Junction Inference and Classification for Figure Completion using Tensor Voting

Philippos Mordohai and Gérard Medioni
Institute for Robotics and Intelligent Systems,
University of Southern California,
Los Angeles, CA 90089, USA
{mordohai, medioni}@iris.usc.edu

Abstract

We address the issues associated with figure completion, a perceptual grouping task. Endpoints and junctions play a critical role in contour completion by the human visual system and should be an integral part of a computational process that attempts to emulate human perception. A significant body of evidence in the psychology literature points to two types of completion, modal (or orthogonal) and amodal (or parallel). We provide a computational framework which implements both types of completion and integrates a fully automatic decision making mechanism for selecting between them. It proceeds directly from tokens or binary image input, infers descriptions in terms of overlapping layers and labels junctions as T, L and endpoints. It is based on first and second order tensor voting, which facilitate the propagation of local support among tokens. The addition of first order information to the original framework is crucial, since it makes the inference of endpoints and the labeling of junctions possible. We illustrate the approach on several classical inputs, producing interpretations consistent with those of the human visual system.



(a) Fragmented contour (b) Amodal completion

(c) The Ehrenstein stimulus (d) Modal completion

(e) Layered interpretation of the Ehrenstein stimulus

Figure 1: Amodal and modal completions

1 Introduction

Figure completion is an important component of image understanding that has received a lot of attention from both the computational vision and the neuroscience community over the last few decades. While we do not claim that our approach is biologically plausible, the human visual system serves as the paradigm, since even a small fraction of its performance is still evading all attempts to emulate it. In this paper we show that the interpretations we infer are consistent with human perception even in difficult cases. Consider the fragmented contour of Fig. 1(a) which supports amodal completion of the half circle, as in Fig. 1(b). Note that completion stops at the two endpoints, marked A and B. Now consider Fig. 1(c), which is known as the Ehrenstein stimulus. This supports modal completion (defined here as in the work of Saund [1]) and produces a strong and unique perception of an illusory circle (Fig. 1(c)). Note that the outer circle is not perceived, possibly because it would require support for concave modal completion at each endpoint. What makes this input interesting to us is that both modal and amodal completion are supported by the data. This is consistent with human perception which produces a layered interpretation of the image, with a white disk occluding a set of black lines on a white background (Fig. 1(e)). In case the segments are not aligned, modal completion is still perceived, but amodal is not.

An issue that has received little attention in the computer vision literature is the selection between modal and amodal completion, especially when they compete. Most researchers assume that the type of completion that has to be performed is known in advance. For instance, modal completion starts only when the endpoints and the orientation

of completion are provided as inputs. We aim at inferring the endpoints and junctions from unoriented data, and then, automatically make decisions whether further completion is supported by the data, and which type of completion should occur. The tensor voting framework is well-suited for this task since it infers integrated descriptions of curves, junctions, regions and terminations. The latter is possible after the incorporation of first order information into the tensor voting framework [2] first presented in [3]and [4].

The paper is organized as follows. Related work is presented in the next section, an overview of the approach is presented in Section 3, the tensor voting framework in Section 4 and figure completion in Section 5. Experimental results are shown in Section 6, a summary and perspectives are given in Section 7.

2 Related Work

Classical approaches on the inference of salient contours include the work of Grossberg, Mingolla and Todorovic [5, 6] who developed the *Boundary Contour System* and the *Feature Contour System* that can group fragmented and even illusory edges to form closed boundaries and regions by feature cooperation and competition in a neural network that also includes feedback mechanisms. In a seminal paper, whose performance in a variety of very hard examples has not yet been surpassed, Shashua and Ullman [7] define a locally connected network that assigns a saliency value to every image location according to the length and smoothness of curvature of curves going through it. In [8], Parent and Zucker infer the trace points of curves based on spatial integration of local information and curvature estimation. The contributions of this method include robustness to noise and applicability to real images. In the work of Heitger and von der Heydt [9] elementary curves are grouped into contours via convolution with a set of orientation selective kernels, whose responses decay with distance and difference in orientation. Hérault and Horaud [10] propose deterministic and stochastic annealing methods for the minimization of a global cost function that enforces the grouping of edgels into locally circular shapes and eliminates noise. Williams and Jacobs [11] introduce the *stochastic completion fields* for contour grouping. Their theory is probabilistic and models the contour as the motion of a particle performing a random walk. Particles decay after every step, thus minimizing the likelihood of completions that are not supported by the data or between distant points. In [12] a method for the simultaneous inference of curves and junctions based on the accumulation of local support was presented. The voting scheme is very similar to the one presented here, but data representation was less effective.

An aspect of contour completion very relevant to the results presented in this paper is the perception of illusory contours. Heitger and von der Heydt [9] propose mechanisms for both parallel and orthogonal grouping based on key points such as junctions and line ends. Gove *et al.* [13] present a paper within the BCS/FCS framework that specifically addresses the perception of illusory contours, as well as the perceived lightness of illusory surfaces. Williams and Thornber [14] describe the inference of closed illusory contours from position and orientation constraints that can be derived from line terminations. The potential for random abrupt changes of orientation has been added to the completion field framework to facilitate the inference of illusory contours with corners.

Our work differs in that we can proceed from unoriented, unlabeled data and simultaneously infer curves, junctions, regions and boundaries. Moreover, we propose an **automatic** mechanism for making decisions between modal and amodal completion without having to know the type of completion in advance. The proposed framework allows us to produce full interpretations starting from low level data.

3 Overview of the Approach

The input to our algorithm is a set of tokens in a two-dimensional space. The tokens indicate the presence of a primitive that potentially belongs to a larger configuration. They can be generated in a variety of ways, for instance with an edge detector, or by just assuming that black corresponds to foreground and white to background in a binary image. The output is a layered description of the image in terms of salient curves, junctions and regions. Our first goal is to detect salient groupings based on the support tokens receive from their neighbors. The amount of support from one token to another is in the form of a first and a second order vote whose properties (strength and orientation) depend on proximity, colinearity and cocurvilinearity. The power of tensor voting lies in the representation of each token that can simultaneously encode its behavior as a curvel, a junction or a region inlier. Therefore, tokens do not have to be classified prematurely, and decisions are made when enough information is available. The first contribution of this paper is the use of first order information, termed *"polarity"*, to infer endpoints and region boundaries, as well as label junctions.

The novelty of our current work is the mechanism for modal and amodal completion using the endpoints and junctions inferred at the previous stage. X and Y-junctions do not support completion, except for the case that an endpoint of a contour can also be connected to the junction. Endpoints, T-junctions and L-junctions offer two possibilities for completion. Either completion along the endpoint's tangent or the T-junction's stem if a corresponding keypoint with compatible orientation exists within a certain vicinity (amodal completion), or completion along the orthogonal

orientation of an endpoint, the bar of a T-junction or an edge of an L-junction (modal completion). To facilitate modal and amodal completion, two new voting fields are defined based on the original voting field. The decision whether completion is supported by the data is made by examining the support for both options at every keypoint. If at least one more keypoint supports modal or a modal completion, the current keypoint is labeled as one that supports completion, and the appropriate voting field is used. If both modal and amodal completion are supported, both are further pursued and a layered description is inferred. The process and the fields are described in more detail in Section 5.

4 First Order Augmentation to the Tensor Voting Framework

We describe here the data representation and the voting mechanism, and show how first order information has been integrated with the original second order formulation.

Data representation The representation of each token consists of a second order, symmetric, non-negative definite tensor, which is equivalent to a 2×2 matrix, and a *polarity vector*, which is described later. Another equivalent representation for the second order tensor is an ellipse, whose axes are the eigenvectors of the tensor, and whose aspect ratio is the ratio of the eigenvalues. The major axis is the preferred **normal** orientation of a potential curve going through the location. The *shape* of the ellipse indicates the certainty of the preferred orientation. That is, an elongated ellipse represents a token with high certainty of orientation. On the other hand, an ellipse with two almost equal eigenvalues represents a token with very little preference for any orientation (Fig. 2). The tensor's *size* encodes the saliency of the information encoded. Larger tensors convey more salient information than smaller ones.

Figure 2: Geometric illustration of saliency

This representation allows us to initialize with oriented or un-oriented tokens, or a combination of both. Oriented tokens (curvels) are initialized as *stick tensors* and unoriented ones (points) as *ball tensors*. A stick tensor has one non-zero eigenvalue. Its major axis (the eigenvector corresponding to λ_1) is the normal of the curvel. A ball tensor has two equal eigenvalues, and thus, is perfectly isotropic with respect to orientation. In the examples in the remainder of the paper, all tokens are initially encoded as unoriented ball tensors.

During the voting process, tokens cast tensor votes to their neighbors within a radius defined by the *scale of the voting field*, σ, the only free parameter of the framework. The votes are also second order, symmetric, non-negative definite tensors which are accumulated at the receiver by tensor (matrix) addition. The size of the accumulated tensor indicates the saliency of the receiver and its shape indicates the most likely structure type, curve, region or junction, it belongs to. This scheme allows all tokens to interact with their neighbors regardless of their type. Analysis of the resulting tensor field at the end of the voting process produces the desired description.

This formulation of the tensor voting framework, as strictly second order, is very effective for many perceptual organization problems (see [2]), but it is incapable to detect structure terminations such as the endpoints of curves. To address this shortcoming, first order information needs to be integrated into the framework [4]. Polarity vectors are now associated with each token and encode the support the token receives for being a termination of a perceptual structure. The term polarity refers to the magnitude of the polarity vector and is large when the majority of the token's neighbors lie on the same side. The direction of the polarity vector indicates the direction of the inliers of the perceptual structure whose potential boundary is the token under consideration.

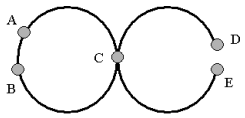
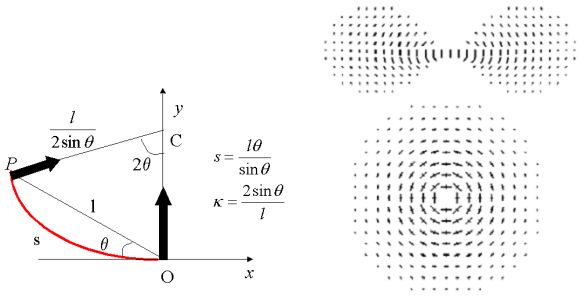


Figure 3: A contour with a junction and two endpoints

Figure 3 illustrates the necessity of the first order augmentation to the framework. The contour consists of a number of "smooth inliers", such as A and B, which can be inferred from their neighbors under the constraint of good continuation. Consider point C, which is an X-junction where two smooth segments of the contour intersect. It can be represented as having both curve normals simultaneously. This is a second order discontinuity that is described by the presence of a salient ball tensor. On the other hand, the endpoints D and E of the contour are also smooth continuations of their neighbors and are represented as salient stick tensors. Therefore, they cannot be qualitatively discriminated from points A and B using second order properties only. The property that makes these pairs of points very different is that D and E are terminations of the curve while A and B are not. This can be captured by the polarity vector, whose magnitude is very close to zero for A and B, which are surrounded by other curvels, and locally maximum for D and E, which receive stronger support from one side of the contour only. Polarity allows us to infer richer descrip-



(a) Stick vote (b) Voting fields

Figure 4: Second order vote cast by a stick tensor located at the origin and the stick and ball voting fields

tions that make the distinction between interior points and endpoints, that was missing before. The same principle is applied for regions and is also useful for junction labeling, since L, T and W-junctions have high polarity, whereas X and Y-junctions do not.

Tensor Voting Once the inputs, oriented or unoriented, have been encoded with tensors, the information they contain is propagated to their neighbors via first and second order votes. The design of the voting field is explicitly described in [2].

Second order voting We first consider the case of a second order vote cast by a voter with a pure stick tensor. In the absence of other information, the arc of the *osculating circle* (the circle that shares the same normal as a curve at the given point) at O that goes through P is the most likely smooth path, since it maintains constant curvature (Fig. 4(a)). The second order vote is also a stick tensor and has a normal lying on the radius of the osculating circle at P. The magnitude of the vote is a function of the distance between the tokens and the curvature of the path. The influence from a token to another attenuates with distance, to minimize interference from unrelated tokens, and curvature, to favor straight over curved continuation when both are possible. Moreover, no votes are cast if the receiver is at an angle larger than 45^o with respect to the orientation of the voter. The *saliency decay function* is:

$$DF(s,\kappa,\sigma) = e^{-(\frac{s^2+c\kappa^2}{\sigma^2})} \quad (1)$$

where s is the arc length OP, κ is the curvature, c is a constant that controls the degree of decay with curvature, and σ is the *scale of voting*, which determines the effective neighborhood size.

The 2-D second order stick voting field for a unit stick voter located at the origin and aligned with the y-axis is defined in the following equation. The field is illustrated in Fig. 4(b) where the line segments show the normal orientation of the vote cast at every location, while their length is proportional to the magnitude of the vote.

$$\mathbf{S}_{SO}(d,\theta) = DF(s,\kappa,\sigma) \begin{bmatrix} -sin(2\theta) \\ cos(2\theta) \end{bmatrix} [-sin(2\theta) \quad cos(2\theta)] \quad (2)$$

The ball voting field $\mathbf{B}(P)$ can be derived from the stick voting field $\mathbf{S}(P)$ by integrating the votes cast by a stick voter that spans the unit circle. It is illustrated in Fig. 4(b). It can be shown that any tensor can be decomposed into a stick component of size $\lambda_1 - \lambda_2$ parallel to the tensor's maximum eigenvector, and a ball component of size λ_2. Therefore, if the voters are arbitrary tensors, they are first decomposed into two components that vote separately. For stick tensors of arbitrary size the magnitude of the vote is given by (1) multiplied by the the size of the stick $\lambda_1 - \lambda_2$. Similarly, the magnitude of votes cast by arbitrary ball tensors has to be multiplied by λ_2. Since the size of the neighborhood is fixed, the complexity of tensor voting is linear with respect to the number of tokens.

First order voting Now, we turn our attention to the generation of first order votes. Note that the first order votes are cast according to the second order information of the voter. This occurs because polarity vectors have to be initialized to zero since no assumption about structure terminations can be made. The first order vote cast by a unitary stick tensor at the origin is *tangent* to the osculating circle and is directed towards the voter. Its magnitude is equal to that of the second order vote, according to (1). The first order voting field for a unit stick voter aligned with the y-axis is:

$$\mathbf{S}_{FO}(d,\theta) = DF(s,\kappa,\sigma) \begin{bmatrix} -cos(2\theta) \\ -sin(2\theta) \end{bmatrix} \quad (3)$$

The *first order ball voting field* can be derived from the first order stick voting field in a way similar to the second order case.

Vote collection and interpretation Votes are accumulated at each location by tensor addition in the case of the second order votes, which are in general arbitrary second order tensors, and by vector addition in the case of the first order votes, which are vectors.

Analysis of the second order votes can be performed once the eigensystem of the accumulated second order 2×2 tensor has been computed. Then the tensor can be decomposed into the stick and ball components:

$$T = (\lambda_1 - \lambda_2)\hat{e}_1\hat{e}_1^T + \lambda_2(\hat{e}_1\hat{e}_1^T + \hat{e}_2\hat{e}_2^T) \quad (4)$$

where $\hat{e}_1\hat{e}_1^T$ is a stick tensor, and $\hat{e}_1\hat{e}_1^T + \hat{e}_2\hat{e}_2^T$ is a ball tensor. If $\lambda_1 \gg \lambda_2$, this indicates certainty of one normal orientation, therefore high curve saliency. If $\lambda_1 \approx \lambda_2 > 0$, the dominant component is the ball and there is no preference of orientation. This indicates either a token that belongs to a region, which is surrounded by neighbors from the same region from all directions, or a junction where two or more curves intersect and multiple curve orientations are present simultaneously. The terms ball and junction saliency are used interchangeably and correspond to λ_2.

The result of first order voting is the polarity vector. Since the first order votes are also weighted by the saliency of the voters and attenuate with distance and curvature, their vector sum points to the direction from which the most salient contributions were received. Low polarity indicates a token that is in the interior of a curve or region, therefore surrounded by neighbors whose votes cancel each other out. On the other hand, high polarity indicates a token that is on or close to a boundary, thus receiving votes from only one side with respect to the boundary, at least locally. For the example of Fig. 3, A and B have high curve saliency and low polarity, C has high junction saliency and low polarity, and D and E have high curve saliency and high polarity.

The inference of perceptual structures is performed taking into account both first and second order information. Curves are extracted after non-maximal suppression of curve saliency. The two neighbors along the inferred normal at each site are examined and the non-maximal ones are suppressed. The non suppressed tokens serve as candidates during curve extraction, which starts from the most salient tokens and proceeds as long as curve candidates with compatible orientations can be grouped in the curve, or until an endpoint is reached. Endpoints are detected as local maxima of polarity. Non-maximal suppression is performed along the polarity vector this time, which is parallel to the contour's tangent. Regions are inferred due to their high λ_2, while region boundaries also have maximal polarity along the direction of the polarity vector. Finally, junctions are peaks of λ_2. Suppression, in this case, is performed with respect to λ_2 inside a small window. Furthermore, junctions can be given a preliminary label based on their polarity, since L, T and W-junctions have high polarity, whereas X and Y-junctions do not.

5 Figure completion

Now that endpoints and junctions have been inferred and labeled automatically, we need to consider how they interact to infer figure completion. There are two types of completion that have to be considered: modal and amodal. In amodal completion endpoints extend along their tangent and T-junctions along their stem. In modal completion connections occur orthogonally to the tangent of endpoints, along the bar of T-junctions, or along one of the edges of an L-junction. Amodal and modal completion are termed parallel and orthogonal grouping by Heitger and von der Heydt [9]. The proposed framework makes automatic decisions on which type of completion occurs. Either type has to be supported by at least one other key point within a certain range. For instance, amodal completion from an endpoint has to be supported by another endpoint with similar orientation and opposite polarity. Then, tensor voting is performed to infer a contour connecting the two endpoints, as with the endpoints of the curve segments of Fig. 1(a). Amodal completion between pairs of endpoints produces the contour of Fig. 1(b).

(a) Parallel field (b) Orthogonal field

Figure 5: Voting fields used for amodal and modal completion. P is the polarity vector at the voting endpoint or junction. Voting is allowed only towards the opposite half-plane

New voting fields, based on the original second order stick voting field, have to be defined to facilitate the propagation of information from the voting endpoints and junctions. The orientation and saliency of the votes are the same as in (1) and (2), but the fields are one-sided. This is because amodal completion can only occur along one direction: away from the curve that was inferred at the first stage or in the direction that would make the T-junction an X-junction. Modal completion is possible only towards the direction that results in a convex illusory contour. This is due to a strong preference in the human visual system for convex modal completions [12]. Even though the two fields are orthogonal, the polarity vector, in both cases, indicates the direction opposite to completion. Figure 5 shows the fields used for these cases. Before voting, the following cases have to be considered for each keypoint:

- There is no possible modal or amodal continuation due to the absence of other keypoints that support either option. In this case endpoints are just terminations, like A and B in Fig. 1(b).
- Amodal completion is supported by another keypoint of the same type with similar orientation and opposite polarity, while modal is not (Fig. 1(a)). Endpoints and stems of T-junctions cast votes with the parallel field along their tangent.
- Modal completion is possible, but amodal is not (Fig.

6). Support in this case has to come from keypoints with similar curve orientation and polarity. Completion occurs orthogonally to the tangent of endpoints or along the bars of T-junctions and edges of L-junctions, using the orthogonal field.

- Both types of completion are possible (Fig. 1(c)). In this case, the modal completion is perceived as occluding the amodal completion. In the case of the Ehrenstein stimulus, a disk is perceived to occlude the crossing lines (Fig. 1(e)).

Once the above cases have been examined, the keypoints can be labeled with respect to whether they support completion or not. The appropriate field is used according to the type of completion. Analysis of the votes is performed as in the previous section and curves, junctions and endpoints are inferred. now, junctions can be fully labeled.

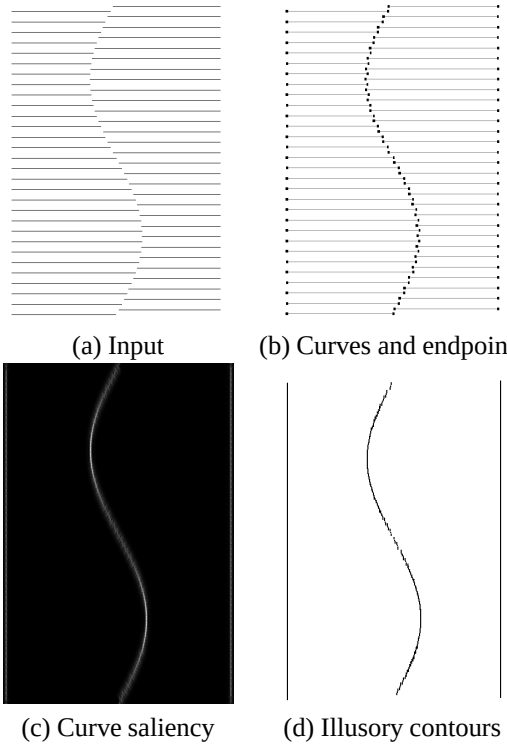
6 Experimental Results

We now present experimental results on a variety of examples. A small scale is used for the original data and a large scale (typically 20 to 30 times larger) is used to infer completions at the second stage, where the distance between tokens is considerably larger.

Modal completion The input consists of a set of unoriented tokens that form line segments and can be seen in Fig. 6(a). The curves and endpoints detected after a first pass of tensor voting can be seen in Fig. 6(b). The endpoints are marked in black and their inferred normals are orthogonal to the segments. Three illusory contours can be inferred after voting using the orthogonal field of Fig. 5(b). Curve saliency and the illusory contours can be seen in Fig. 6(c) and (d). Higher intensity corresponds to higher saliency, as in all saliencies shown in this paper. Even though the convexity of the sinusoidal contour changes, it is still inferred since locally some endpoints propagate votes that support the contour. Note that no other contours can be inferred from the curve saliency map even though the orientation of the endpoints is not normal to the sinusoidal contour at all locations except the peaks.

Koffka crosses An interesting perceptual phenomenon is the Koffka crosses (see also [14]). The perceived illusory contour changes from a circle to a square as the arms of the cross become wider. Two examples of Koffka crosses can be seen in Fig. 7(a) and (b). The leftmost one is narrow and produces an illusory disk, while the rightmost is wide and produces an illusory square.

The black pixels of these images are encoded as unoriented ball tensors and the endpoints of the segments are



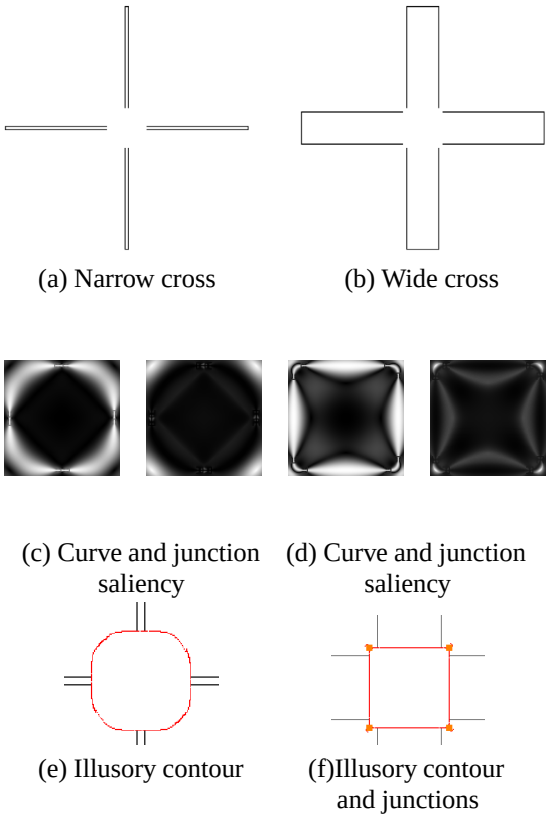
(a) Input (b) Curves and endpoints

(c) Curve saliency (d) Illusory contours

Figure 6: Modal completion. Starting from unoriented inputs, endpoints are detected and used as inputs for modal completion

extracted as before. Modal completion is possible, since the endpoints can be hypothesized as T-junctions with zero contrast bars, and voting is performed using the orthogonal field. Curve and junction saliencies are shown in Fig. 7(c) and (d). Note that the saliencies in each map are normalized independently so that white corresponds to the maximum and black to the minimum. The maximum junction saliency is 90.4% of the maximum curve saliency for the wide cross and only 9.8% of the maximum curve saliency for the narrow cross, where no junctions are inferred. Figures 7(e) and (f) show the inferred modal completion. Intermediate widths of the arms produce intermediate shapes of the illusory contour, such as rounded squares, which are consistent with human perception.

The case of amodal completion from the detected endpoints to compatible endpoints must also be considered. The parallel voting field of Fig. 5(a) should be used in this case. Figures 8(a) and (b) show the curve and junction saliencies for the wide Koffka cross of Fig. 7(b). Four contours that connect corresponding endpoints and four X-junctions are inferred (Fig. 8(c)). This interpretation is also consistent with human perception, which is a layered interpretation of the scene that consists of a white square, oc-



(a) Narrow cross (b) Wide cross

(c) Curve and junction saliency (d) Curve and junction saliency

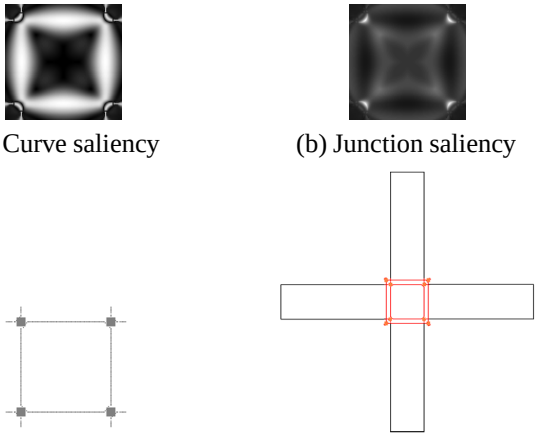
(e) Illusory contour (f)Illusory contour and junctions

Figure 7: Koffka crosses. Inputs, saliency maps and inferred contours and junctions (marked as squares)

cluding a cross, on top of a white background.

We also performed an experiment with one arm of the wide cross missing. The input can be seen in Fig. 9(a). Voting with the orthogonal field produces the saliency and polarity maps seen in Fig. 9(b-d). The polarity map has four local maxima: at the two L-junctions and at two of the six voting endpoints. The inferred description consists of three straight contours, two L-junctions and two endpoints at the wrongly hypothesized T-junctions, A and B, which based on polarity can now be correctly labeled as L-junctions. Additional modal completion is now possible starting from the L-junctions that results in the contour AB on Fig. 9(f).

Amodal completion through a region (Poggendorff illusion) In the final example of this paper we attempt to explain the Poggendorff illusion that occurs when a straight line does not appear perfectly straight when it passes through a region. The input can be seen in Fig. 10(a) and consists of unoriented points that form a line and a region. The region boundaries are inferred based on their high λ_2 and high polarity and are used as inputs to a second pass of tensor voting along with the curvels. After the second pass, four L-junctions and two T-junctions are inferred. There is no support for completion based on the L-junctions so they do not contribute any further. The T-junctions should be treated carefully in this case. The orientation of the stems is locally unreliable, since as one approaches the junction, curve saliency decreases almost to zero while junction saliency increases. Even if the orientation θ of the line farther away from the stem is used, some uncertainty remains. On the other hand, the bar and the stem of a T-junction are expected to be orthogonal [12]. Combining these two sources of evidence, we can set the orientation of the stem equal to $\alpha 90^o + (1-\alpha)\theta$. Voting from the two T-junctions using the parallel field, for a range of values of α, produces a curve saliency map similar to the one of Fig. 10(b) where completion is not straight and an inflection point always exists in the contour (Fig. 10(c)). This is one possible explanation for the illusion. Also, note that the illusion does not occur if the line is orthogonal to the region boundaries. The saliency map in the absence of the region is the one shown in Fig. 10(e) where straight continuation occurs, as expected.



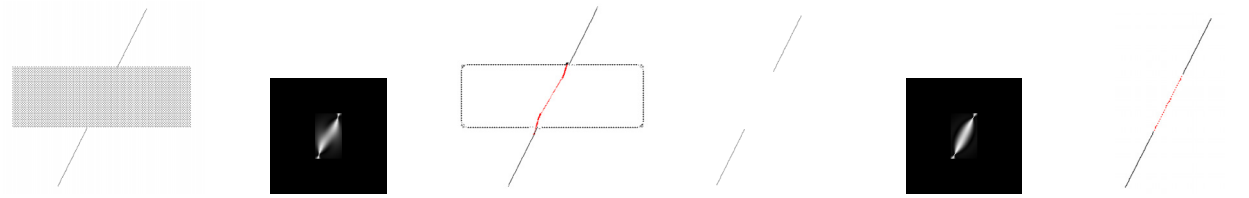
(a) Curve saliency (b) Junction saliency

(c) Illusory contours and junctions (d) All completions

Figure 8: Amodal contour completion and illustration of all possible completions, including the occluded ones

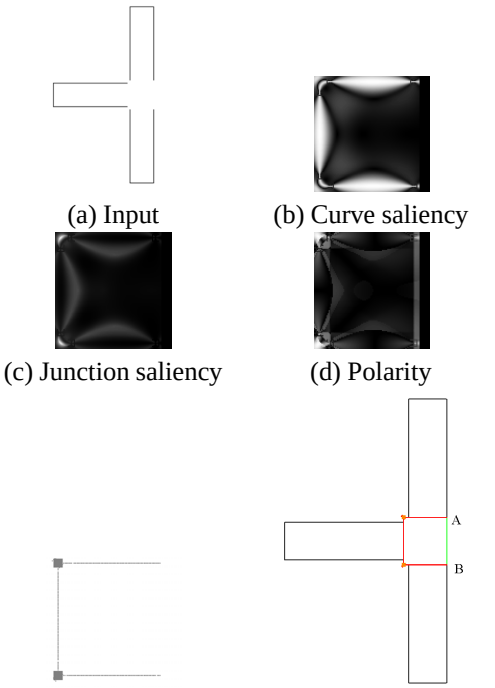
7 Concluding Remarks

In this paper we have presented a first order augmentation to the tensor voting framework that allows us to infer richer descriptions than with the strictly second order formulation of the framework. It enables us to detect the endpoints of curves and the boundaries of regions, as well as label junctions. Moreover, we have addressed figure completion of



(a) Input (b) Curve saliency (c) Completion (d) Line input (e) Curve saliency (f) Completion

Figure 10: Possible explanation of the Poggendorff illusion



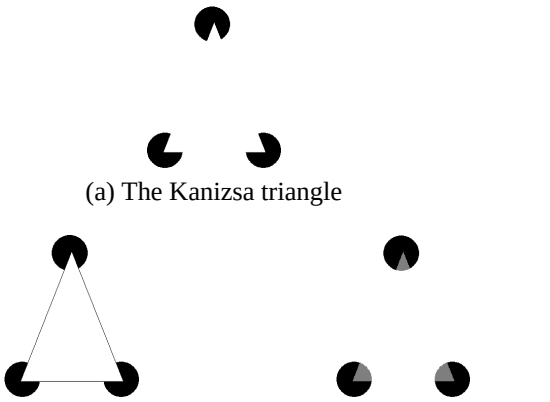
(a) Input (b) Curve saliency (c) Junction saliency (d) Polarity (e) Illusory contours (f) Visible completions

Figure 9: Modal completion for the three-armed cross

both modal and amodal type in an automatic and integrated way. Our framework does not require a priori knowledge of the type of completion that has to be performed, but can infer it from the data. We demonstrated how it can be applied to the interpretation of line drawings that exhibit complex perceptual phenomena.

Consistently with the body of work on tensor voting, we have avoided premature decisions based on local operators. For instance, we do not classify a location as an L-junction just because a corner detector produces a strong response. Instead, we only assign a preliminary label based on the results of first and second order voting, which for an L-junction have to produce high ball saliency and high polarity. The final labeling occurs only after the completion possibilities have been examined.

The results presented here are encouraging. However, there are still numerous issues that need to be addressed, even in simple line drawings. Consider for instance Fig. 11(a) that depicts the Kanizsa triangle [15]. It contains six L-junctions and each junction supports two possibilities for completion. Either completion along the straight edge which produces the triangle of Fig. 11(b), or completion along the circular edge which produces the three disks seen in 11(c). This example is an excellent demonstration of a scenario that cannot be handled by the current state of our research. Moreover, note that making the decision on one vertex of the triangle affects the other two vertices. As demonstrated by numerous visual illusions, drawings of impossible objects and in [1], for instance, locally consistent perceptions that are globally impossible are accepted by the human visual system. Therefore, the decision on one vertex does not automatically resolve the other two. What is clear, however, is that the computer vision, psychology and neuroscience literature provide abundant examples for which a more sophisticated decision making mechanism than the one presented here are needed. This is one of the axes of our future research.



(a) The Kanizsa triangle

(b) Completion of the triangle (c) Completion of the disks

Figure 11: The Kanizsa triangle and two possible interpretations

In this paper we have also scratched the surface of inferring hierarchical descriptions. Typically processing occurs in two stages: in the first stage tensor voting is performed on the original low level tokens, while in the second stage completion based on the previously inferred structures is performed. The processing stages are three in case regions are present in the dataset. Then, region boundaries are inferred in the first stage and interact with other tokens at the second stage. Completion now occurs at the third stage. More complicated scenarios may include more stages. It is reasonable to assume that scale increases from stage to stage, as the distances between the "active" points increase. A more systematic investigation of the role of scale in this context is also required. It is possible that the interpretation of certain inputs changes as the scale of voting varies. Our objective is to develop a more complete approach that integrates low level edge and corner detectors, possibly similar to [16], to process real images, as in [17][18].

Acknowledgment

This research has been supported by the National Science Foundation grant IIS 03 29247.

References

[1] E. Saund, "Perceptual organization of occluding contours of opaque surfaces," *CVIU*, vol. 76, no. 1, pp. 70–82, October 1999.

[2] G. Medioni, M.S. Lee, and C.K. Tang, *A Computational Framework for Segmentation and Grouping*, Elsevier, 2000.

[3] W.S. Tong, C.K. Tang, and G. Medioni, "First order tensor voting and application to 3-d scale analysis," in *CVPR*, 2001, pp. I:175–182.

[4] W.S. Tong, C.K. Tang, P. Mordohai, and G. Medioni, "First order augmentation to tensor voting for boundary inference and multiscale analysis in 3d," *PAMI*, vol. 26, no. 5, pp. 594–611, May 2004.

[5] S. Grossberg and E. Mingolla, "Neural dynamics of form perception: Boundary completion," *Psychological Review*, pp. 173–211, 1985.

[6] S. Grossberg and D. Todorovic, "Neural dynamics of 1-d and 2-d brightness perception: A unified model of classical and recent phenomena," *Perception and Psychophysics*, vol. 43, pp. 723–742, 1988.

[7] A. Shashua and S. Ullman, "Structural saliency: The detection of globally salient structures using a locally connected network," in *ICCV*, 1988, pp. 321–327.

[8] P. Parent and S.W. Zucker, "Trace inference, curvature consistency, and curve detection," *PAMI*, vol. 11, no. 8, pp. 823–839, August 1989.

[9] F. Heitger and R. von der Heydt, "A computational model of neural contour processing: Figure-ground segregation and illusory contours," in *ICCV*, 1993, pp. 32–40.

[10] L. Herault and R. Horaud, "Figure-ground discrimination: A combinatorial approach," *PAMI*, vol. 15, no. 9, pp. 899–914, September 1993.

[11] L.R. Williams and D.W. Jacobs, "Stochastic completion fields: A neural model of illusory contour shape and salience," *Neural Computation*, vol. 9, no. 4, pp. 837–858, 1997.

[12] G. Guy and G. Medioni, "Inferring global perceptual contours from local features," *IJCV*, vol. 20, no. 1/2, pp. 113–133, 1996.

[13] A. Gove, S. Grossberg, and E. Mingolla, "Brightness perception, illusory contours, and corticogeniculate feedback," *Visual Neuroscience*, vol. 12, pp. 1027–1052, 1995.

[14] L.R. Williams and K.K. Thornber, "Orientation, scale, and discontinuity as emergent properties of illusory contour shape," *Neural Computation*, vol. 13, no. 8, pp. 1683–1711, 2001.

[15] G. Kanizsa, *Organization in Vision*, Praeger, New York, 1979.

[16] U. Koethe, "Integrated edge and junction detection with the boundary tensor," in *ICCV*, 2003, pp. 424–431.

[17] S. Mahamud, L.R. Williams, K.K. Thornber, and K. Xu, "Segmentation of multiple salient closed contours from real images," *PAMI*, vol. 25, no. 4, pp. 433–444, April 2003.

[18] E. Saund, "Finding perceptually closed paths in sketches and drawings," *PAMI*, vol. 25, no. 4, pp. 475–491, April 2003.