Graph Cuts using a Riemannian Metric induced by Tensor Voting

Hyung Il Koo and Nam Ik Cho
Department of EECS, INMC, Seoul National University, Seoul, 151-742, Korea
hikoo@ispl.snu.ac.kr nicho@snu.ac.kr

Abstract

In this paper, we present a new algorithm that combines the advantages of tensor voting into graph cuts. Tensor voting has been a popular tool for a number of early vision problems since it can use principles of perceptual grouping, which are not well considered in graph cuts. We attempt to encode the power of tensor voting into an energy minimization framework. For this, we assume that the tensor map obtained by tensor voting induces a Riemannian metric in image domain, and the metric is constructed according to the conventional ways of tensor interpretation. Finally, by embedding the induced Riemannian metric into the graph via edge weights, the graph cuts algorithm can have priors considering principles of perceptual grouping. The proposed method can be used in the labeling of occluded regions, object segmentation using only edge information, and boundary regularization.

1. Introduction

Graph cuts algorithm has become one of important tools for solving computer vision and image processing problems because it is numerically stable and it can efficiently find a global optimum [3, 21, 25]. In some researches on graph cuts, it has been shown that cut metrics on graphs can arbitrarily approximate the metrics in a continuous space such as a Riemannian metric [2]. The results not only provide a theoretical relation between two image segmentation methods (geodesic active contours and graph cuts) but also brought performance improvements of graph cuts by alleviating a metrication problem. Moreover the approach that combines the flux into graph cuts enabled segmentation of thin objects by largely suppressing shrinking bias and showed that graph cuts can approximate more wide class of metrics [14]. However, unfortunately, graph cuts algorithm has very limited powers in extrapolation and boundary regularization, which require the consideration of higher-order correlations or principles of perceptual grouping. A simple disocclusion problem in Fig. 1 (a) is such an example. Since the pairwise terms in conventional graph cuts algorithm are not good in handling higher-order correlations, they produce the results like Fig. 1 (b). Although it is desirable to encode higher-order correlations into the graph cuts framework [23], designing an energy function considering such information may require higher order cliques, which consequently brings difficulties in the design of potential and inference process [29]. There are similar problems with other kinds of image segmentation algorithms such as the random walker algorithm [9] or spectral clustering algorithm [24], because they do not have the ability of extrapolation which also require the consideration of higher order correlations.

Interestingly, such information is relatively easy to handle in Tensor Voting Framework (TVF), because TVF is designed to use principles of perceptual grouping. In this framework, input data are encoded as tensors, and the information is propagated by considering proximity and continuity. TVF has been used in a number of applications such as surfaces inference from stereo pair, image inpaint-
ing, image correction and epipolar geometry estimation [12, 13, 16, 26]. The tensor voting owes its success to the ability in noise suppression and dense extrapolation using principles of perceptual grouping. Moreover, TVF is a non-iterative algorithm and the efficient implementation using parallel structure of Graphics Processing Units (GPUs) is possible [20]. However, it is not clear how to extract a connected object from a dense tensor map as can be seen in Fig. 1(e). The algorithm usually results in a number of line segments rather than a single connected curve. Fig. 1(e) shows that some boundaries are not connected even after many erroneous line segments are detected. This problem can not be solved unless we consider the global structure of objects (connectivity or closure), which is not an easy problem in TVF itself. However, this problem can be resolved if we consider the problem as a labeling problem that extracts an object from background, which is what the graph cuts can do.

In this paper, we combine tensor voting with graph cuts so that perceptual grouping priors can be integrated into the energy minimization framework. For the explanation of the proposed algorithm, we first briefly review the results of [2] in the next subsection.

### 1.1. Cut metrics approximating Riemannian metrics

Results in [2] and [14] showed that the cut metric on graphs approximates any Riemannian metric by assigning proper edge \((n\text{-link})\) weights. To be precise, it is shown in [2] that given a local metric \(D(s)\), the weight of edge that links \(s\) and the \(k\)-th node should be

\[ w_k(s) = \frac{\delta^2 \cdot |e_k|^2 \cdot \Delta \phi_k \cdot \det D(s)}{2 \cdot (e_k^T \cdot D(s) \cdot e_k)^2} \tag{1} \]

where \(\delta\) is a grid size, \(e_k\) is a vector to the \(k\)-th node, \(\Delta \phi_k\) is an angle from the horizon to the direction of \(e_k\) as shown in Fig. 2, and \(\Delta \phi_k\) is defined \(\Delta \phi_1 = \phi_2 - \phi_1, \Delta \phi_2 = \phi_3 - \phi_2, \ldots\). Due to the discrete nature of graph cuts, it is required that edge weights are homogeneous in local regions, which is satisfied in most of applications including the proposed metric. Note that all experimental results in this paper are obtained using a 16-neighborhood system as illustrated in Fig. 2.

#### 1.2. Our approach

In this paper, we propose a framework that integrates TVF into the energy minimization framework. It is achieved by encoding perceptual information obtained from TVF into the edge \((n\text{-link})\) weights in the graph. The proposed method is quite different from the conventional methods where only pixel affinities were encoded into the edge weights [1, 9, 24, 22] because our method can encode higher-order correlation such as constant curvature into edge weights.

The tensor map obtained by TVF is the result of addition of many votes, and the map contains several kinds of perceptual information. Especially in the case of 2D, a stick tensor gives the information on curves (curve saliency and curve normals). By designing a Riemannian metric so that a desirable curve (with respect to the tensor map) has smaller geodesic length and vice versa, the tensor map obtained using TVF can be encoded into the Riemannian metric. Moreover, by using the method that approximates Riemannian metric to a cut metric [2], this process reduces to assigning edge \((n\text{-link})\) weights so that they have the information of dense tensor map.

Experimental results show that the proposed algorithm provides improved results compared to the conventional methods by using better image priors. Also we can solve a labeling problem even for the occluded regions (disocclusion) as can be seen in Fig. 1 (a) and (c). Objects can also be extracted from imperfect edge information (contrast information) as can be seen in Fig. 1 (d) and (f).

#### 2. A Riemannian metric induced by TVF

In this section, we define a Riemannian metric on the domain where tensors are densely estimated. The definition of the metric is based on the tensor decomposition and its interpretation. For the explanation of the metric, we begin with the review of TVF and we define a Riemannian metric induced by TVF followed by comparisons with the conventional ones.

##### 2.1. Tensor Voting

In tensor voting (2D-case), each input data is encoded as a ball tensor or a stick tensor, and the token casts a vote to its neighborhood using predefined voting field. The magnitude of the vote is decayed by saliency decay function:

\[ DF(r, \kappa, \sigma) = e^{-\frac{r^2 + \kappa^2}{2\sigma^2}} \tag{2} \]
where \( r \) is the arc length between two points, \( \kappa \) is the curvature, \( c \) is a constant controlling decay with \( \kappa \), and \( \sigma \) is the scale of voting \([8, 19]\). This \textit{saliency decay function} controls voting strength in such a way that a small portion of strength are given to distant points and the points that are connected to the curves with high curvature.

At the receiving site, the collected votes are combined through tensor addition, which produces the dense tensor map \( T : P \to \mathbb{R}^{d \times 2} \) where \( P \) is a set of sites (the domain of an image). Then the local features at the site \( s \) can be extracted by decomposing the tensor \( T(s) \):

\[
T(s) = \lambda_1 \hat{e}_1 \hat{e}_1^T + \lambda_2 \hat{e}_2 \hat{e}_2^T \quad (3)
\]

where \( \lambda_i \) are eigenvalues in decreasing order and \( \hat{e}_i \) are corresponding eigenvectors (for simplicity, we will sometimes omit the site \( s \) for the eigenvalues and eigenvectors). In this representation, \( \hat{e}_1 \hat{e}_1^T \) is a 2D stick tensor with \( \hat{e}_1 \) indicating the normal direction of curve and \( \hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T \) is a 2D ball tensor. And \( \lambda_1 - \lambda_2 \) is called curve saliency and \( \lambda_2 \) is called junction saliency. If \( \lambda_1 \gg \lambda_2 \), it is very likely that the token belongs to a curve whose normal direction is \( \hat{e}_1 \). In this case, \( (\lambda_1 - \lambda_2, \hat{e}_1) \) provides a local information on the curve, \textit{i.e.}, the existence of a curve and its orientation. On the other hand, \( \lambda_1 \simeq \lambda_2 \) means that the token has no preference to the orientation, which occurs when there coexist multiple orientations in that token (junction) or the token lies inside a region \([8, 19]\).

### 2.2. Riemannian Metric induced by TVF

In order to connect the tensor map with graph cuts, we assume that \( T(s) \) induces a local Riemannian metric at \( s \). Conventional tensor interpretation tells that

- a unit vector is the most desirable curve segment when the vector is normal to \( \hat{e}_1(s) \) (\textit{i.e.}, parallel to \( \hat{e}_2(s) \))

- a unit vector is the most undesirable curve segment when the vector is parallel to \( \hat{e}_1(s) \)

at the site \( s \). Hence we design a local metric \( D(s) \) so that the minimum length occurs at the former case:

\[
\hat{e}_2(s) = \arg \min_{\|v\|=1} \sqrt{v^T D(s) v} \quad (5)
\]

and the maximum at the latter case:

\[
\hat{e}_1(s) = \arg \max_{\|v\|=1} \sqrt{v^T D(s) v}. \quad (6)
\]

This means that the two eigenvectors of a local metric are \( \hat{e}_1(s) \) and \( \hat{e}_2(s) \), and \( D(s) \) is given by

\[
D(s) = \phi(\lambda(s)) \cdot \hat{e}_1(s) \cdot \hat{e}_1^T(s) + \psi(\lambda(s)) \cdot \hat{e}_2(s) \cdot \hat{e}_2^T(s) \quad (7)
\]

where \( \lambda(s) \) is the normalized value of curve saliency \( (\lambda_1(s) - \lambda_2(s)) \) so that its variance is unity. This metric results in the local distance map in the form of ellipse. Since the larger curve saliency \( \lambda(s) \) means the larger preference on the orientation, the ellipse becomes more elongated for larger \( \lambda(s) \), and the ellipse should be a circle when \( \lambda(s) = 0 \) (no orientation preference).

Among many possible choices satisfying aforementioned conditions \( (\phi(x), \psi(x)) \) is an increasing function with \( \phi(0) = 1 \), we set \( \phi(x) = e^{kx} \) and \( \psi(x) = e^{-kx} \) so that the area of local distance map is fixed. The local metric is illustrated in Fig. 3. As can be seen in the figure, the curve segment normal to \( \hat{e}_1 \) introduces smaller cost. This definition naturally results in the geodesic length of a curve \( C \) on \( T \):

\[
|C|_T = \int_C \sqrt{\tau_s^T D(s) \tau_s} ds \quad (8)
\]

\[
= \int_C \sqrt{e^{k\lambda(s)}|\tau_s^T \hat{e}_1|^2 + e^{-k\lambda(s)}|\tau_s^T \hat{e}_2|^2} ds \quad (9)
\]

\[
= \int_C \sqrt{2 \sinh(k\lambda(s))|\tau_s^T \hat{e}_1|^2 + e^{-k\lambda(s)}} ds \quad (10)
\]

and it satisfies the following intuitive conditions:

- If a segment of \( C \) is not normal to \( \hat{e}_1 \), the length of the segment increases as the angle deviation \( \tau_s^T \hat{e}_1 \) and the curve saliency \( \lambda(s) \) are increasing.

- If a segment of \( C \) passes a region where \( \lambda(s) \) is small, the length of the segment increases as the curve saliency decreases (even if the segment is normal to \( \hat{e}_1(s) \)).

- If \( \lambda(s) \) is small, the length should be a Euclidean length.

### 2.3. Summary

Now we explain the way of using the \( D(s) \) in Eq. 7 in the energy minimization framework. In graph cuts, the energy
function is usually given by
\[ E(\{L_p\}_{p \in \mathcal{P}}) = \sum_{p \in \mathcal{P}} V_p(L_p) + \mu \sum_{(p,q) \in \mathcal{E}} w_{p,q} \delta(L_p, L_q) \] (11)
where \( L_p \in \{0, 1\} \) is a label at the site \( p \), \( V_p(\cdot) \) is unary term which encodes the likelihood of foreground/background, \( \mathcal{E} \) is a set of pairwise neighborhood (See Fig. 2) and \( \delta(\cdot, \cdot) \) is
\[ \delta(L_p, L_q) = \begin{cases} 1 & L_p \neq L_q \\ 0 & L_p = L_q \end{cases} \] (12)
Then, by assigning \( w_{p,q} \) using Eq. 1 and Eq. 7, we can make the second term in Eq. 11 be the geodesic length of the object boundaries [2]. Because the minimization of Eq. 11 using graph cuts results in tradeoff between the first term and the second term, where the former is related with the pixel-wise observation and the latter with higher-order correlations or perceptual grouping, the proposed framework can impose better smoothness constraints. The summary of this process can be found in Table 1.

The drawback of the proposed framework is that we have to determine the scale of voting (\( \sigma \) in Eq. 2). Although it is known that results are not sensitive to reasonable selection of scale [8], \( \sigma \) is still the critical parameter for the success of our framework.

### 2.4. Comparisons to other metrics

The idea of using an anisotropic Riemannian metric has already been used in the literature including the framework of level set, geodesic active contour, and graph cuts. Because this approach can consider the edge strength and the edge orientation simultaneously, it provides better results compared to the case of isotropic metric [1, 7]. However, in most cases, the metric is induced by local observations and they do not have a power of perceptual grouping. For example, the local metric in [2] is defined as
\[ D(s) = (1 - g(|\nabla I|)) \cdot u \cdot u^T + g(|\nabla I|) \cdot I_{2 \times 2} \] (13)
where \( |\nabla I| \) is an image gradient, \( u = \frac{\nabla I}{|\nabla I|} \) and \( g(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \). This metric can be considered as a special case of the proposed one where the curve saliency is defined as edge strength, the normal vector is defined as an edge direction, and the tensor voting is skipped. Another approach that is worth mentioning is flux-based one. Flux-based methods help to segment narrow elongated objects such as vessels by suppressing shrinking bias [27]. In [14], the authors showed that flux-based method can be integrated into graph cuts: the flux makes regional t-links and it helps to avoid shrinking. Although the flux-based method is also an attractive approach, it requires the shape prior which is not always possible. Without shape priors, it has limited applications in such cases that normal vectors are directed from the object to the background (for example, the intensity gradient is oriented consistently inwards or outwards along boundaries like vessels) [27] or resolving \( \pm 180^\circ \) degree ambiguity of the normal vector is required [17].

### 3. Applications

In this section, we present three exemplary applications among many ones that the proposed method can be applied. The first application is the object segmentation using edge information. The second is to improve the quality of image segmentation by using boundary regularization (especially texture segmentation), and the final application is disocclusion.

#### 3.1. Object segmentation using edges (contrast)

Edge information is one of the most important clues for object recognition. Human can segment objects using imperfect edges, but the automatic segmentation using only edges is not a simple task because there are a number of false edges (not boundary edges). They come from the noise, background clutter and high contrast region within the object. Moreover, some boundaries are missing due to the low contrast. Fig 1(d) shows such an example. Even after the tensor voting and curve inference, it is not clear whether it is boundary or not.

In order to consider the connectivity as well as dense tensor map, we pose the problem as a labeling problem and propose a new criterion for object segmentation. The pro-
Figure 5. Object segmentation using edge segments.

The ratio of two functionals are used in several image segmentation applications [11, 14]. Note that \(1/2\) for a compact set [10].

In estimating a dense tensor map \(T\), we use a default voting field \([8]\), and the input token is obtained by an edge detection algorithm. Since TVF can use several types of token, the method works even for edge map without orientation as shown in Fig. 4. However, estimating the edge orientation is a better choice for performance and efficiency. Our edge detection method is similar to a conventional edge detection algorithm, except that we use a 7 \(\times\) 7 mask and the orientation is also estimated (we use a stick tensor in voting process). We minimize the Eq. 14 by Dinkelbach’s method \([15, 6]\), which consists of a finite number of minimization of

\[
E^\mu(S) = |\partial S|^T - \mu |S|
\]

for given \(\mu\). Since \(|\partial S|^T\) can be approximated to a discrete one by assigning weights of \(n\)-links using Eq. 1 and Eq. 7, and \(-\mu |S|\) can be encoded as weights of \(t\)-links, the functional can be discretized to the form of Eq. 11. And the functional can be minimized via graph cuts [4].

An experimental result on synthetic data can be found in Fig. 1(f), which resembles a way of human’s perception. Experimental results on real world images are also shown in Fig. 5. As can be seen in Fig. 5 (b), there are many false edge segments. TVF suppresses some of them and the estimated normal field guides the segmentation occurs along the true object boundary. Because the method is based on only contrast, it also works on contrast reversing edges as can be seen Fig. 5 (e) and (f) (Similar examples can be found in [11, 23]). The experimental result on an example which consists of contrast-reversing edges with adjacent texture background is also shown in Fig. 5 (g) and (h).

3.2. Boundary Regularization

Since the proposed method endows better boundary smoothness conditions, it can be incorporated in the framework of interactive image segmentation [1]. Image segmentation problem is also modeled as a minimization problem:

\[
E(\{L_p\}_{p \in P}) = \sum_{p \in P} V_p(L_p) + \mu \sum_{(p,q) \in E} w_{p,q} \delta(L_p, L_q).
\]

where \(\mu\) is a constant controlling the effect of the smoothness term. In conventional approaches, the contrast dependent weights \(\{w_{p,q}\}\) are used to obtain the clean boundaries in color segmentation [1]. However, in texture segmentation, the contrast dependent term gives little information about the object boundaries and therefore \(\{w_{p,q}\}\) are designed to minimize the Euclidean length of the boundaries (i.e., \(D(s) = I_{2 \times 2}\) is used in Eq. 1).

Experimental results using the Euclidean metric on texture segmentation can be found in Fig. 6. Fig. 6(a) is the input with user interaction, (b) represents \(V_p(L_p)\) which is obtained by a training machine using the features of [5]. Fig. 6 (c) and (d) show the segmentation results for two different values of \(\mu\). The boundaries are not clean even for a larger \(\mu\) where thin and elongated parts are not correctly segmented due to shrinking bias. Fig. 6 (e) is a magnified version of (a) where errors are occurred in (c) and (d) (the leftmost leg of the mask). It shows the difficulties in estimating the labels of pixels using only local observation.
In order to handle the problem, we use the data term in Fig. 6 (b) in two ways: one is to use the data term as a regional bias (as the first term in Eq. 17) and the other is to infer the metric of the domain $\mathcal{P} - \mathcal{O}$. Applying the edge detection method to Fig. 6 (b) and using TVF, we obtain tensor map $\mathbf{T}$ and it results in edge weights in Eq. 17. Fig. 6 (f) shows an experimental result of the proposed method. Although it is not clear that the resulting mask is more similar to ground truth (for example, hamming distance can be an objective measure), the result is much pleasing to human because the result is connected and has smooth boundary.

3.3. Disocclusion

Disocclusion is an essential process for the object recognition, robot vision, and image restoration [12, 18]. In many disocclusion literatures, the problem is considered as finding an optimal path connecting two end points. For example, the curve inference on the dense tensor map was performed in [12]. In their approach, they found a curve by detecting high curve saliency points and connecting them using B-spline because conventional TVF produces a number of line segments.

Unlike the conventional approaches based on the inference of curves, we pose the disocclusion problem into a labeling problem on the occluded region $\mathcal{O}$. Since labels on $\mathcal{P} - \mathcal{O}$ are already known, the energy function is given by

$$E(\{L_p\}_{p \in \mathcal{P}}) = \sum_{p \in \mathcal{P}} V_p(L_p) + \sum_{(p,q) \in \mathcal{E}} w_{p,q} \delta(L_p, L_q),$$

(18)

where $w_{p,q}$ is induced by the tensor map $\mathbf{T}$ and

$$V_p(L_p) = \begin{cases} 0 & p \in \mathcal{O} \\ \infty & p \notin \mathcal{O} \text{ and } L_p \neq \hat{L}_p \\ 0 & p \notin \mathcal{O} \text{ and } L_p = \hat{L}_p. \end{cases}$$

(19)

Here $\hat{L}_p$ is a known label, and $L_p, \hat{L}_p \in \{0, 1\}$. In this case, the tensor map $\mathbf{T}$ is obtained by applying the edge detection method on the defined region ($\mathcal{P} - \mathcal{O}$). The experimental result on a simple case can be found in Fig. 1(c). Since the first term in Eq. 18 plays a role of hard constraints, the minimization of Eq. 18 results in the minimization of the second term while satisfying the constraints. It is equivalent to find a minimum length curve (or several curves) on $\mathcal{T}$. Since a minimum length curve $\mathcal{C}$ on $\mathcal{T}$ means the most natural curve (with respect to $\mathcal{T}$), this framework provides a natural disocclusion.

This labeling-based formulation of disocclusion has several advantages over the conventional ones. First, this approach do not require specification of the starting point and the end point of each line segment, which is required in conventional approaches [12, 28]. In addition, the orders of connection are not required. And the proposed method can find a global optimum using graph cuts. Experimental results on more complex cases are shown in Fig. 7. When the second term in Eq. 18 is set to the Euclidean length (i.e., $D(s) = I_{2 \times 2}$), the disocclusion performance is degraded as can be seen in (c) and (d). However, the proposed method naturally extends the known boundaries and results in much improved results as can be seen in Fig. 7 (e) and (f).

4. Conclusions

In this paper, we have presented a new framework that combines the advantages of tensor voting and graph cuts, which are two popular tools for computer vision. The proposed method is based on the assumption that the tensor map induces a Riemannian metric. From this assumption, a local metric is designed according to the conventional ways of tensor interpretation, and the local metric is encoded in the form of edge weights in a graph. Experimental
results show that the induced metric can resolve the problems with conventional ones: it can infer a single object boundary rather than a number of line segments in edge-based segmentation, it can consider higher-order correlation of boundaries, and it also allows a labeling on the occluded region.

References

[14] V. Kolmogorov and Y. Boykov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In ICCV, 2005. 1, 2, 4, 5
[28] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In CVPR, 2008. 6