IMPROVED FINE STRUCTURE MODELING VIA GUIDED STOCHASTIC CLIQUE FORMATION IN FULLY CONNECTED CONDITIONAL RANDOM FIELDS

M. J. Shafiee, A. G. Chung, A. Wong, and P. Fieguth

Vision & Image Processing Lab, Systems Design Engineering Dept., University of Waterloo
{mjshafie, agchung, a28wong, pfieguth}@uwaterloo.ca

ABSTRACT

Markov random fields (MRFs) and conditional random fields (CRFs) are influential tools in image modeling, particularly for applications such as image segmentation. Local MRFs and CRFs utilize local nodal interactions when modeling, leading to excessive smoothness on boundaries (i.e., the short-boundary bias problem). Recently, the concept of fully connected conditional random fields with stochastic cliques (SFCRF) was proposed to enable long-range nodal interactions while addressing the computational complexity associated with fully connected random fields. While SFCRF was shown to provide significant improvements in segmentation accuracy, there were still limitations with the preservation of fine structure boundaries. To address these limitations, we propose a new approach to stochastic clique formation for fully connected random fields (G-SFCRF) that is guided by the structural characteristics of different nodes within the random field. In particular, fine structures surrounding a node are modeled statistically by probability distributions, and stochastic cliques are formed by considering the statistical similarities between nodes within the random fields. Experimental results show that G-SFCRF outperforms existing fully connected CRF frameworks, SFCRF, and the principled deep random field framework for image segmentation.

Index Terms— Stochastic Cliques, SFCRF, Guided Stochastic Cliques, CRF, Segmentation, Fully connected

1. INTRODUCTION

Interactive image segmentation [1, 2] has garnered a lot of recent interest. The goal is to label each pixel in the image as either foreground or background (i.e., foreground/background segmentation) based on user-marked areas inside the object of interest and background that often act as initial seed points to guide the segmentation process. One strategy that has proven to be effectiveness for tackling this segmentation problem is the use of random fields [3, 4, 5]. In such work, the image segmentation problem is formulated as a Maximum A Posteri-

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Ontario Ministry of Research and Innovation, and the Canada Research Chairs Program.
problem was proposed by Jegelka and Bilmes [14] and Kohli
et al. [15] (known as principled deep framework), which re-
solved the short-boundary bias problem (i.e., the excessive
smoothness over boundaries) by changing the cost of the
edges that constitute a cut in the segmentation. The proposed
models penalized the number of types of label discontinui-
ties instead of penalizing the number of label discontinuities
(which is used in regular CRFs). However, these methods
exhibit limitations when segmenting fine or complex bound-
aries.

Recently, Shafiee et al. [16] proposed a new efficient in-
ference framework based on the concept of fully connected
random fields with stochastic cliques (SFCRF). Clique for-
takes place stochastically based on a probability dis-
tribution as defined by the spatial distance and pixel-wise simi-
larity between any given two nodes in the random field. As
SFCRF only considers pixel-wise similarity and spatial dis-
tance, clique formations do not account for fine structures
(usually distributed throughout an image) in the model. As
such, an extended framework that accounts for elongated, fine
structures can potentially yield great benefits when dealing
with fine structure boundaries.

Extending upon [16], we propose a new framework for
stochastic clique formation in fully connected random fields
(called G-SFCRF) to better deal with fine structure bound-
aries. The probability distribution corresponding to the for-
formation of stochastic cliques is guided by learned fine struc-
ture models at each node of the random field. Specifically, the
probability of stochastic clique formation between two nodes
is guided by the statistical similarity between two nodes,
which is determined by the KL-divergence of the correspond-
ing fine structure models.

2. METHODOLOGY

The foreground/background image segmentation problem is
formulated as a MAP problem where the underlying graph
is a fully connected conditional random field with stochastic
cliques (SFCRF):

\[
P(Y|X) = \frac{1}{Z(X)} \exp(-\psi(Y, X))
\]  

(1)

\[
\psi(Y, X) = \sum_{i=1}^{n} \psi_u(y_i, X) + \sum_{\varphi \in C} \psi_p(y_{\varphi}, X)
\]

(2)

where \( Y \) is the state set (i.e., segmentation result) given the
measurement set \( X \) (i.e., input image), \( Z(X) \) is the normal-
ization constant, and \( \psi(\cdot) \) is the combination of the unary
potential function \( \psi_u(\cdot) \) and the pairwise potential function
\( \psi_p(\cdot) \). The unary potential encodes the likelihood model of
each random variable \( y_i \) and its corresponding measurement,
while the pairwise potential represents the relationship be-
tween random variables in a clique structure \( \varphi \in C \), where
\( C \) is the set of all stochastic clique structures. The underlying
graph in the random field is fully connected while the pairwise
clique structures incorporated in the inference procedure are
specified stochastically. Any pair of nodes can form a clique
based on a probability distribution.

In [16], the probability distribution incorporates two fac-
tors: i) spatial distance, and ii) color similarity. Therefore,
closer nodes have a higher probability of forming clique
connectivities than far nodes, and similar nodes construct
clique connectivities with higher probability than dissimilar
nodes. While such an approach provides good modeling per-
fomance for general coarser structure boundaries, it does not
provide an accurate clique formation for complex and fine
structure boundaries for a number of reasons. First, using
only pixel-wise similarity to guide stochastic clique forma-
tion does not take into account the fine structures surrounding
each node and the fine structures were not considered when
determining the probability of stochastic clique formation.
As a result, the constructed underlying graph did not rep-
resent the relationships between nodes with highly similar
fine structures, and leads to smoothing over fine structures
in the segmentation results. Second, the use of a spatial
distance between two nodes, while appropriate for larger,
coarser structures, can be limiting for modeling complex,
fine, elongated structures (as seen in Figure 1).

To overcome these issues, we propose a new stochastic
clique formation framework where fine structures around
each node are modeled via probability distributions, and the
statistical similarity between these fine structure models are
used to guide the formation of stochastic cliques. The new
framework can form a graph where nodes with similar struc-
tural characteristics around the nodes have a high probability
of forming cliques.

Guided Stochastic Clique Formation. The fine struc-
tures surrounding each node are modeled by a probability dis-
tribution based on the measurements in a neighborhood cen-
tered on the node.

To form clique connectivities, every possible pair of nodes
is compared based on their corresponding fine structure mod-
els. Figure 2 shows the underlying graph created by this
framework. As fine structures surrounding nodes are mod-
eled by distribution functions, the KL-divergence is used to
assess the statistical similarity between each pair of nodes via
their associated distributions:

\[
KL(S_i||S_j) = S_i \log \left( \frac{S_i}{S_j} \right)
\]

(3)

where \( KL(\cdot) \) is the KL-divergence between two distributions
\( S_i \) and \( S_j \), and \( S_i \) and \( S_j \) are the fine structure models
of nodes \( i \) and \( j \) in the random field, respectively. A high KL-
divergence corresponds to low statistical similarity. Since
a relatively small neighborhood of measurements is used to
model the surrounding fine structures for a given node, the
underlying probability distribution is assumed to be a Gauss-
ian distribution to simplify the model. As such, the closed
The performance of the proposed guided clique formation approach within the SFCRF framework was compared with different state-of-the-art CRF inference frameworks for the task of interactive image segmentation. The CSSD dataset [17] was utilized to compare all methods, and consists of 200 natural images with manually annotated ground truth segmentation. Furthermore, a new dataset composed of 20 natural images with very fine structures was created to evaluate the ability of the tested algorithms in preserving complex and fine structure boundaries during segmentation. The images were selected from the MSRA dataset [18]; therefore, we named the subset of images as MSRA-FS.

**Competing Algorithms.** The proposed guided SFCRF (G-SFCRF) framework was compared against different CRF frameworks:

- **Fully-connected CRF (FCRF)** [9] A fully-connected CRF with an efficient inference by use of Gaussian potential functions and permutohedral lattice [19].
- **SFCRF** [16] A fully-connected CRF with stochastic cliques based on a distance-intensity stochastic clique formation strategy.

The same unary potentials were used in all methods and were computed via five-component Gaussian mixture models trained on the pixel color intensities within each seed region (i.e., foreground, background).

The FCRF was evaluated with the standard deviations of the Gaussian pairwise potential function set as (3, 3) and its weight was 5; the standard deviations of the bilateral potential function were (20, 20, 0.08) with the weight of 10, where the first two values show the spatial standard deviation and the last one is the color standard deviation (i.e., the image dynamic range is [0, 1]). The contrast-dependent Potts pairwise potential used in PD method was selected similarly to [15] where \( \sigma \) was the mean of the color gradients of the image and \( \lambda = 1.5 \) and \( \theta = 0.0005 \). Both SFCRF frameworks (SFCRF and the proposed G-SFCRF) were evaluated with the same rate of clique formation, with the expected number of cliques for each node set to 80.

To assess the methods quantitatively, F1-scores were applied to both the segmentation region and boundary:

\[
F1 = 2 \cdot \frac{Pr \cdot Re}{Pr + Re} , \quad Pr = \frac{TP}{TP + FP} , \quad Re = \frac{TP}{TP + FN}
\]

where \( TP, FN, \) and \( FP \) are the number of true positives, false negatives, and false positives, respectively.

---

1The dataset is publicly available at http://www.eng.uwaterloo.ca/ mjshafie/MSRA-FS.html
Fig. 3. Example segmentation results produced on the MSRA-FS and CSSD datasets. As can be seen, the proposed G-SFCRF method provides strong preservation capabilities for very fine structure boundaries.

Region F1-Score: The conventional region F1-score was evaluated based on the region-of-interest specified by the ground truth images, and compared the methods generally.

Boundary F1-score: Boundary accuracy was used to compare algorithms with respect to preserving fine structure boundaries. Motivated by [20], the extracted boundary of the ground truth was assumed as the positive class with all other pixels specified as the negative class. The distance tolerance of 1 pixel was used in the calculation of this evaluation measure. Thus, the boundary detected by the algorithm was considered to be a true positive if it was within 1 pixel of a ground-truth boundary.

The two F1-scores were applied based on [21], where the F1-scores were computed for the biggest object in the image.

Table 1 shows the quantitative comparisons of the tested methods in terms of the region F1-score and the boundary F1-score. The F1-scores were reported for two datasets: CSSD and MSRA-FS. It can be observed that the proposed guided stochastic clique formation (G-SFCRF) led to a better underlying graph, resulting in a more accurate segmentation than the previous SFCRF framework with respect to both the region-based and boundary-based F1-scores. The proposed framework also outperformed the PD method, which is regarded as the state-of-the-art in random field-based image segmentation approaches, and performed comparably to the FCRF approach in terms of the F1-scores. However, a visual assessment shows that the proposed method preserves boundaries and fine structures better than the FCRF approach.

Table 1. Quantitative comparison of the proposed G-SFCRF method and other frameworks via F1-scores. G-SFCRF shows good performance in the segmentation of the image into foreground and background, as reported by the region F1-score, while preserving the fine structured boundaries, as shown by the boundary F1-score.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region F1-score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSSD</td>
<td>0.8582</td>
<td>0.8314</td>
<td>0.8643</td>
<td>0.8664</td>
</tr>
<tr>
<td>MSRA-FS</td>
<td>0.9162</td>
<td>0.8493</td>
<td>0.9102</td>
<td>0.9171</td>
</tr>
<tr>
<td><strong>Boundary F1-Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSSD</td>
<td>0.4658</td>
<td>0.4708</td>
<td>0.4717</td>
<td>0.4991</td>
</tr>
<tr>
<td>MSRA-FS</td>
<td>0.5806</td>
<td>0.4453</td>
<td>0.5166</td>
<td>0.5851</td>
</tr>
</tbody>
</table>

The proposed method is capable of preserving very fine structure boundaries. This is especially evident in the segmentation results for the sheep, eagle, and red door image examples.

4. CONCLUSION & FUTURE WORK

In this work, we proposed a new framework to guide the stochastic clique formation for fully connected conditional random fields for the purpose of preserving fine structure boundaries in image segmentation. Reported results show that the new framework can preserve very fine structure boundaries and outperform other state-of-the-art CRF inference frameworks for the tested datasets. Future work includes the investigation of other distributions to model the fine structure surrounding nodes in the random field more accurately for different types of images.
5. REFERENCES


