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

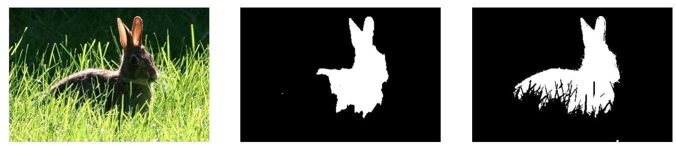
ABSTRACT

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 Posteriori (MAP) problem [6], incorporating defined interactions between neighboring pixels as the prior information via random fields as well as incorporating the likelihood of each pixel to all predefined labels. Image segmentation problems are commonly modeled using a pairwise Markov random field (MRF) or a conditional random field (CRF) utilizing the local neighborhood interactions (i.e., four or eight closest nodes) [3].



(a) Image (b) PD (c) G-SFCRF

Fig. 1. An example of an object with fine and complex boundaries. (b) shows the result of principle deep random field (PD) and (c) demonstrates the result of proposed framework (G-SFCRF) in foreground/background segmentation.

A challenge with using local MRFs/CRFs for image segmentation is that they are mainly restricted to short-range connections, resulting in excessive smoothness in the image segmentations [4, 7] (i.e., the short-boundary bias problem). Although expanding the clique structure to higher-order cliques [5, 8] have been introduced to address this problem, these methods fail to preserve fine structure boundaries. Figure 1 shows an example of an image containing an object with very fine and complex boundaries, where local MRFs and CRFs have high difficulty handling.

An approach that can tackle the short-boundary bias problem was proposed by Krähenbühl and Koltun [9], who introduced a new framework using fully connected random fields to take advantage of all connectivities in the CRF model. This new framework addressed the computational complexity of the fully connected random fields (FCRFs) by using specific potential functions and incorporating a new data structure (i.e, Permutohedral lattice) [10]. Further extensions [11, 12, 13] to the concept was proposed to relax assumptions and limitations associated with [9]. However, the efficient inference frameworks based on FCRF have specifically made use of a Gaussian feature function, which can be limiting for segmenting fine and complex boundaries as one of the advantages of CRFs is the ability to use arbitrary feature functions.

An alternative approach to tackle the short-boundary bias

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 resolved 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 discontinuities instead of penalizing the number of label discontinuities (which is used in regular CRFs). However, these methods exhibit limitations when segmenting fine or complex boundaries.

Recently, Shafiee *et al.* [16] proposed a new efficient inference framework based on the concept of fully connected random fields with stochastic cliques (SFCRF). Clique formation takes place stochastically based on a probability distribution as defined by the spatial distance and pixel-wise similarity between any given two nodes in the random field. As SFCRF only considers pixel-wise similarity and spatial distance, 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 boundaries. The probability distribution corresponding to the formation of stochastic cliques is guided by learned fine structure 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 corresponding 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):

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

$$\psi(Y,X) = \sum_{i=1}^{n} \psi_u(y_i, X) + \sum_{\varphi \in C} \psi_p(y_\varphi, X) \tag{2}$$

where Y is the state set (i.e., segmentation result) given the measurement set X (i.e., input image), $Z(X)$ is the normalization 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 between 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 factors: 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 performance 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 formation 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 represent 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 structural characteristics around the nodes have a high probability of forming cliques.

Guided Stochastic Clique Formation. The fine structures surrounding each node are modeled by a probability distribution based on the measurements in a neighborhood centered on the node.

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

$$KL(S_i||S_j) = S_i log\Big(\frac{S_i}{S_j}\Big) \tag{3}$$

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 Gaussian distribution to simplify the model. As such, the closed

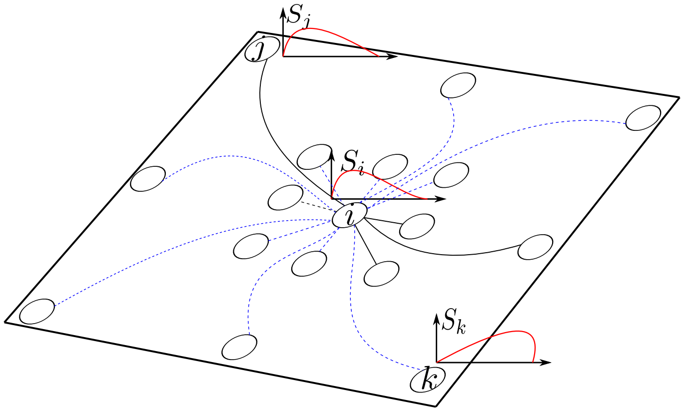


Fig. 2. An illustration of the proposed framework. The clique connectivities for node i are stochastically formed by comparing a probability distribution modeled on the surrounding fine structures (i.e., S_i) to probability distributions modeled on the fine structures surrounding other nodes, such as S_j for node j. Since the fine structures $\{S_i, S_j\}$ corresponding to two nodes i and j are similar, the probability of connectedness is higher than that for two nodes i and k, where S_i and S_k are completely different.

form of (3) is

$$KL(S_i||S_j) = \frac{(\mu_i - \mu_j)^2}{2\sigma_j^2} + \frac{1}{2} \cdot \Big(\frac{\sigma_i^2}{\sigma_j^2} - 1 - log(\frac{\sigma_i^2}{\sigma_j^2})\Big) \tag{4}$$

where $S_i = \mathcal{N}(\mu_i, \sigma_i^2)$ and $S_j = \mathcal{N}(\mu_j, \sigma_j^2)$ are the Gaussian distributions of the fine structure models of nodes i and j, respectively. The proposed structure similarity measure $KL(\cdot)$ is combined with the pixel intensity similarity measure (used by Shafiee *et al.* [16]) to formulate the stochastic clique probability distribution $P_C(i, j)$, which can be expressed by:

$$P_C(i,j) = \frac{exp(-KL(S_i||S_j)) \cdot exp(E(i,j))}{\gamma} \tag{5}$$

where $E(\cdot)$ encodes the pixel intensity similarity-measure, and γ is the sparsity factor that specifies the expectation of the number of clique connectivities. It can be seen that $P_C(i, j)$ stochastically favors nodes with similar fine structures and pixel intensities. As such, the set of all stochastic clique structures C can be defined by

$$C = \Big\{(i,j)|P_C(i,j) \geq \hat{u}\Big\} \tag{6}$$

where $\hat{u}$ is a randomly generated number from $U(0, 1)$, a uniform distribution over the unit interval.

The proposed approach provides a SFCRF where nodes surrounded by similar fine structures are connected in the underlying graph. Thus, each node is usually affected by other nodes with similar structure and pixel intensity properties, and reduces the amount of smoothing produced by the penalty function. As a result, the proposed method better preserves the fine structure boundaries in the image segmentation.

3. RESULTS & DISCUSSION

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[1] 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].

- **Principled deep random field (PD)** [15] A new penalty function implemented based on a hierarchical framework.

- **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 σ 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:

$$F1 = 2 \cdot \frac{Pr.Re}{Pr+Re},\ Pr = \frac{TP}{TP+FP}\ ,\ Re = \frac{TP}{TP+FN}$$

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

[1]The dataset is publicly available at http://www.eng.uwaterloo.ca/ mjshafie/MSRA-FS.html

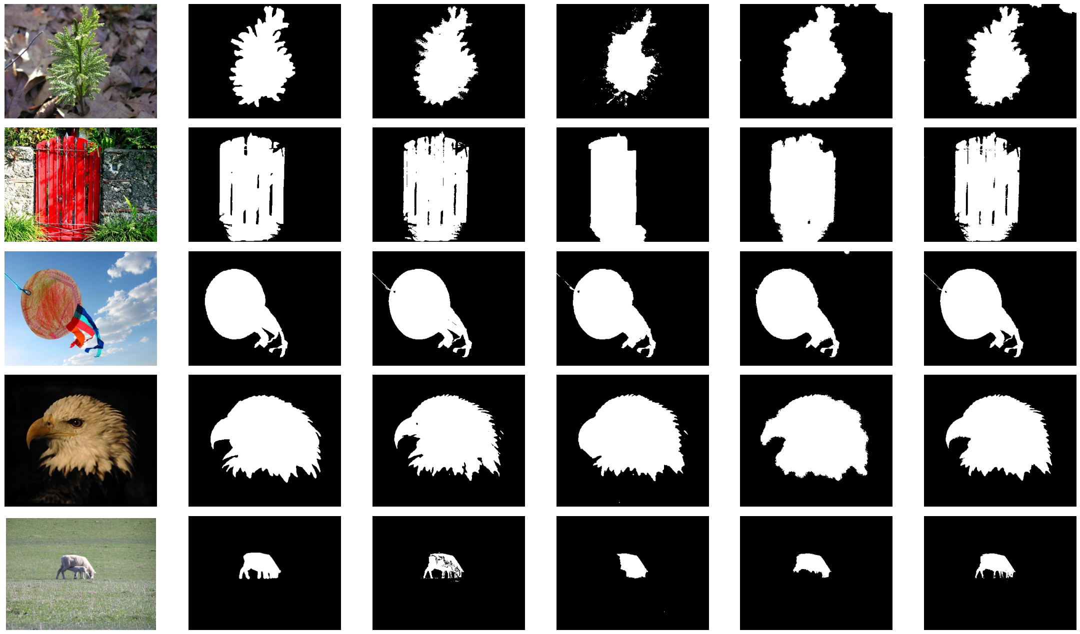


Image GT FCRF [9] PD [15] SFCRF [16] G-SFCRF

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.

Figure 3 shows examples of segmentation results produced by the tested methods for both datasets (CSSD and MSRA-FS). It can be observed that 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.

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.

	FCRF [9]	PD [15]	SFCRF [16]	G-SFCRF
Region F1-score				
CSSD	0.8582	0.8314	**0.8643**	**0.8664**
MSRA-FS	**0.9162**	0.8493	0.9102	**0.9171**
Boundary F1-Score				
CSSD	0.4658	0.4708	0.4717	**0.4991**
MSRA-FS	0.5806	0.4453	0.5166	**0.5851**

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 differen types of images.

5. REFERENCES

[1] A. Protiere and G. Sapiro, "Interactive image segmentation via adaptive weighted distances," *Transactions on Image Processing (TIP). IEEE*, 2007.

[2] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in nd images," in *International Conference on Computer Vision (ICCV)*. IEEE, 2001.

[3] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International Journal of Computer Vision (IJCV). Springer*, 2009.

[4] J. Verbeek and W. Triggs, "Scene segmentation with crfs learned from partially labeled images," in *Advances in Neural Information Processing Systems (NIPS)*, 2007.

[5] X. He, R. Zemel, and MA Carreira-Perpindn, "Multiscale conditional random fields for image labeling," in *Conference on Computer vision and pattern recognition (CVPR)*. IEEE, 2004.

[6] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI). IEEE*, 2008.

[7] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller, "Multi-class segmentation with relative location prior," *International Journal of Computer Vision (IJCV). Springer*, 2008.

[8] P. Kohli, P. Torr, et al., "Robust higher order potentials for enforcing label consistency," *International Journal of Computer Vision (IJCV). Springer*, 2009.

[9] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with gaussian edge potentials," in *Advances in Neural Information Processing Systems (NIPS)*, 2011.

[10] A. Adams, J. Baek, and M. Davis, "Fast high-dimensional filtering using the permutohedral lattice," in *Computer Graphics Forum*, 2010.

[11] K. Ristovski, V. Radosavljevic, S. Vucetic, and Z. Obradovic, "Continuous conditional random fields for efficient regression in large fully connected graphs," in *Conference on Artificial Intelligence (AAAI)*, 2013.

[12] Y. Zhang and T. Chen, "Efficient inference for fully-connected CRFs with stationarity," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012.

[13] S. Zheng, M. Cheng, J. Warrell, P. Sturgess, V. Vineet, C. Rother, and P. Torr, "Dense semantic image segmentation with objects and attributes," in *International Conference on Computer Vision (ICCV)*. IEEE, 2014.

[14] S. Jegelka and J. Bilmes, "Submodularity beyond submodular energies: coupling edges in graph cuts," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011.

[15] P. Kohli, A. Osokin, and S. Jegelka, "A principled deep random field model for image segmentation," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013.

[16] M. J. Shafiee, A. Wong, P. Siva, and P. Fieguth, "Efficient bayesian inference using fully connected conditional random fields with stochastic cliques," in *International Conference on Image Processing (ICIP)*. IEEE, 2014.

[17] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013.

[18] M. Cheng, N. Mitra, X. Huang, P. Torr, and S. Hu, "Global contrast based salient region detection," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI). IEEE*, 2014.

[19] A. Adams, J. Baek, and M. Davis, "Fast high-dimensional filtering using the permutohedral lattice," in *Computer Graphics Forum*, 2010.

[20] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI). IEEE*, 2011.

[21] S. Alpert, M. Galunm, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2007.