EFFICIENT BAYESIAN INFERENCE USING FULLY CONNECTED CONDITIONAL RANDOM FIELDS WITH STOCHASTIC CLIQUES

M. J. Shafiee, A. Wong, P. Siva, P. Fieguth

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

ABSTRACT

Conditional random fields (CRFs) are one of the most powerful frameworks in image modeling. However practical CRFs typically have edges only between nearby nodes; using more interactions and expressive relations among nodes make these methods impractical for large-scale applications, due to the high computational complexity.

Recent work has shown that fully connected CRFs can be tractable by defining specific potential functions. In this paper, we present a novel framework to tackle the computational complexity of a fully connected graph without requiring specific potential functions. Instead, inspired by random graph theory and sampling methods, we propose a new clique structure called stochastic cliques. The stochastically fully connected CRF (SFCRF) is a marriage between random graphs and random fields, benefiting from the advantages of fully connected graphs while maintaining computational tractability.

The effectiveness of SFCRF was examined by binary image labeling of highly noisy images. The results show that the proposed framework outperforms an adjacency CRF and a CRF with a large neighborhood size.

***Index Terms*—** Conditional Random Fields, Random Graph, Stochastic Clique, Stochastically Fully Connected Random Fields

1. INTRODUCTION

Structural learning is a well-known approach in computer vision. Simple methods like naïve Bayes, to complex methods like Markov random fields (MRFs), have made strong contributions to computer vision challenges such as image denoising, segmentation, image labeling and super resolution [1, 2, 3]. These algorithms can be illustrated from two different points of view:

1. Based on their probabilistic formulations and associated parameters, and
2. Based on the representation of their graphical models [4].

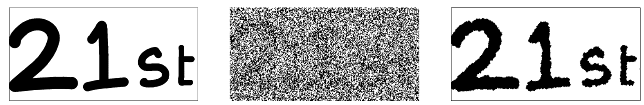
The graphical models are of interest here; these are divided into directed and undirected graphs, however it is the undirected graphical models which are more appropriate in image modeling.

MRFs are the most common undirected graphical model used in computer vision. The MRFs assert an assumption of conditional independence of states on measurements. In contrast, the conditional random field proposed by Lafferty et al. [5] relaxes this assumption by explicitly modeling the conditional dependence of the states on the measurements. The impact of CRFs has been investigated in [6, 7, 8].

Basic CRF models utilize unary and pairwise potentials on local neighborhoods. These adjacency CRFs lead to smoothed image boundaries because they are not able to capture the long range relations among nodes, thus a number of methods have been proposed to capture such non-local relations.

The basic CRF framework was expanded to incorporate non-local relations through hierarchical connectivity [9]. Similarly, fully connected CRFs frameworks were proposed for semantic image labeling[10]. However, the complexity of inference in those models limits the usage of them to hundreds nodes or fewer.

Krähenbühl and Koltun [2] proposed a tractable inference procedure based on specific potential functions. They modeled the multi-class image segmentation based on a fully connected CRF, and the edge potentials were obtained using Gaussian kernels. Based on these new feature functions, they



(a) True image (b) Observation (c) Label map

Fig. 1. SFCRF binary label classification result. The goal is to label the observed, noisy image into foreground and background. Here the given image is distorted with 90% salt & pepper noise, however the SFCRF still produces an accurate label map.

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Ontario Ministry of Economic Development and Innovation and Canada Research Chairs Program.

formulated the inference as a filtering problem.

Following this method, Zhang and Chen [11] relaxed the Gaussian assumption to any distribution by using a stationary constraint. They showed that the spatial potentials over two pixels depend only on their relative positions. Therefore, they could encode more statistics by different distribution. Champbell et. al. [12] generalized the pairwise potentials to a non-linear dissimilarity measure. They presented the pairwise terms as density estimates of the conditional probability, and the probabilities were expressed by a dissimilarity measure. Ristovski et al. [13] proposed new continuous CRF on a fully connected graph. Their framework is similar to [2] but they targeted the regression problems with continuous outputs.

The aforementioned methods try to reduce the $O(N^3)$ computational complexity of exact inference on a fully connected graph by approximation algorithms and filter-based frameworks. They addressed the problem by defining specific potential functions to manage the inference as a filtering approach. However using specific feature functions limits the effectiveness of CRF in modeling, as the one strengths of the CRF is the selection of arbitrary feature functions.

In this work, we propose a novel CRF framework on a fully connected graph. Inspired by Monte Carlo sampling, we propose a new stochastic clique structure, which allows the computational complexity of the fully connected graph to be reduced without limiting the CRF with specific feature functions. Because of this flexibility, the new framework preserves the merits of the standard CRF, as any arbitrary function can be selected as the potential function. The proposed stochastically fully connected random fields (SFCRF) is a mixture between random graph theory [14] and random fields theory, sampling the fully connected random field while allowing for computational tractability. An illustration of the proposed SFCRF for binary labeling is shown in Figure 1.

2. METHODOLOGY

Stochastically fully connected conditional random fields (SFCRFs) are fully connected random fields in which cliques are defined stochastically. The term *fully connected* refers to the fact that each node in the graph is connected to all other nodes of the graph, however the cliques for each node are determined based on distribution probabilities, so the number of pairwise cliques in the graph may not be the same as the number of neighborhood pairs.

The goal is to model $P(Y|X)$, the conditional probability of the state set Y given the measurement X. The conditional random field (CRF) approach to expressing $P(Y|X)$ is to write it as

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

where $Z(X)$ is the partition function and $\psi(.)$ is some combination of unary and pairwise potential functions:

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

Here $y_i \in Y$ is a single state in the set $Y = \{y_i\}_{i=1}^n$, $y_\varphi \in Y$ is the subset of states (clique), and $X = \{x_j\}_{j=1}^n$ is the set of observations.

Each node i has a set of neighbors

$$N(i) = \Big\{ j | j = 1 : n, j \neq i \Big\} \tag{3}$$

where $|N(i)| = n - 1$. The specified clique structure C is, in this paper, taken to be the pairwise clique

$$C = \Big\{ C_p(i) \Big\}_{i=1}^{n} \tag{4}$$

$$C_p(i) = \Big\{ (i,j) | j \in N(i), 1_{\{i,j\}}^S = 1 \Big\} \tag{5}$$

although the formulation can be generalized for other clique structures. $C_p(i)$ for node i is determined based upon a stochastic indicator neighbor function, $1_{\{i,j\}}^S = 1$, to distinguish whether two nodes can construct a clique or not. This function itself is a combination of probability distributions. For image modeling, this function must consider the spatial relation among the nodes and must involve data driven into the model; therefore, the proposed indicator function is the combination of spatially driven and data driven probabilities:

$$1_{\{i,j\}}^S = \begin{cases} 1 & P_{i,j}^s \cdot Q_{i,j}^d \geq \gamma \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$1_{\{i,j\}}^S = 1$ has two responsibilities:

1. Incorporate the spatial information ($P_{i,j}^s$)
2. Involve the data relation among the states ($Q_{i,j}^d$)

The threshold γ determines the sparsity of the graph.

The potential function $\psi_u(.)$ and $\psi_p(.)$ from (2) are the combinations of unary and pairwise feature functions and their corresponding weights λ, respectively:

$$\psi_u(y_i, X) = \sum_{j=1}^{K} \lambda_j^u f_j(y_i, X) \tag{7}$$

$$\psi_p(y_\varphi, X) = \sum_{\{y_i, y_j\} \in y_\varphi, k=1}^{K'} \lambda_k^p f_k(y_i, y_j, X) \tag{8}$$

where the λ constants control the importance of each feature function in the energy formulation calculated in the training stage. Since the SFCRF is a generalization of the CRF, the various λ parameters can be tied [15] in some situations.

Graph $G(V, E)$ is the realization of the SFCRF where V is the set of nodes of the graph representing the states

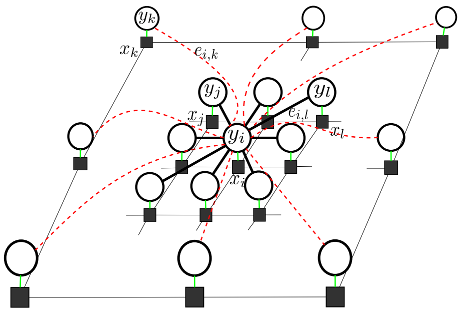


Fig. 2. A realization of a stochastically fully connected conditional random field graph. A connectivity between two nodes is determined based on a distribution; each two nodes in the graph can be connected according to a probability drawn from this distribution. There is a measurement x_i corresponding to each node y_i. The connectivity of each pair of two nodes y_i and y_k is distinguished by the edge $e_{i,k}$. Closer nodes are connected with a higher probability (black solid edges), whereas two nodes with a greater separation are less probable to be connected (red dashed edges).

$Y = \{y_i\}_{i=1}^n$, and the E is the set of edges of the graph with $|E| \leq \frac{n(n+1)}{2}$. Corresponding to each vertex in the graph $G(.)$ there is an observation $x_i \in X$. The edges in $G(.)$ are randomly sampled, thus G is a realization of a random graph [14]. Based on the Erdös-Rényi theorem [14] if the probability p' of the random graph $\hat{G}_{n,p'}$ is greater than $\frac{\log n}{n}$ the graph is connected with a high probability. As a result, the proposed graph $G(.)$ is connected, has at least $n-1$ edges even for large values of γ, and satisfies a Gibbs distribution [16].

Figure 2 illustrates an example of a SFCRF. As can be seen, each node in the graph can be connected to all other nodes, although to improve visualization the connectivities of the centered node are highlighted. The probability of connecting two nodes as a clique is different for each pair of nodes. According to $P^s_{i,j}$, the connectedness probability of two nodes and the distance between them have inverse relation. However, there is a possibility for two distantly separated nodes y_i and y_k to be connected, as illustrated in Figure 2, which is how the SFCRF takes advantageous of the fully connected CRF.

By the amalgamation of random graph theory and random fields theory, the proposed SFCRF provides the merits of fully connected random fields by sampling the configuration of a fully connected random fields, leading to a much smaller computational complexity than that of fully connected random fields.

3. RESULTS & DISCUSSION

To demonstrate the power of the SFCRF we performed experiments on binary classification datasets. The first dataset is EnglishHnd, a set of handwritten characters [17], containing 3410 images grouped into 62 equally sized classes: 10 classes for digits, 26 classes for upper case letters, and 26 classes for lower case letters. We corrupted the given images with noise, and the problem goal is to classify the pixels of the noisy images as foreground and background. Salt & pepper noise at 80% and 90%, and Gaussian noise at 220% and 300%, where the Gaussian noise percentage is characterised by

$$\% = \left(\frac{\sigma}{\text{dynamic range}} \right) \times 100 \tag{9}$$

The images have a size of 480×360; therefore, the total number of pairwise connections of the fully connected graph is approximately 2.99×10^{10}. According to the random graph theory [14] mentioned earlier, the selection probability must be greater than 3.03×10^{-5} for the graph to be connected. We conducted the experiments with $\gamma = 0.9993$, meaning that the selection probability is equal to 7×10^{-4}, leading to an expected number of pairwise cliques to be 2.09×10^7 with an average of 121 pairwise cliques per node.

To test the effectiveness of the proposed framework, we compared our proposed SFCRF against two other CRFs of different neighborhood sizes. Of the two compared approaches, the first is a regular CRF with adjacent neighbors (CRF-N3) where each node is connected to its closest eight neighbours (those within a 3×3 block); the second model, CRF-N11, has a larger neighborhood, where each node is connected to its closest 120 neighbors (those within an 11×11 block). The exact number of pairwise cliques of CRF-N3 is 1.38×10^6 (8 pairwise cliques per node) and of CRF-N11 is 2.07×10^7(120 pairwise cliques per node).

All experiments were conducted on an Intel Core i7-4770 @ 3.4 GHz. All three methods (CRF-N3, CRF-N11 and SFCRF) are implemented in Matlab, whereas the potential calculation was computed in C++ code integrated with Matlab through the Mex interface. The average computational time for each iteration of the inference step is $0.04s$ for CRF-N3 and $3.85s$ for CRF-N11, a significant difference caused by the change in degree of connectivity between the two models. In contrast, the average runtime per iteration for the SFCRF configuration is $2.7s$. Thus the inference time is decreased, while the flexibility in edge connectivity, in principle allowing arbitrarily distant connections, is increased based on this new clique structure.

Table 1 shows the F1-score for the SFCRF and other CRFs subject to the stated noise. The ground truth labels are obtained by binarizing the true images manually. The results in the table show that the proposed SFCRF outperforms the regular CRFs in all cases.

Figure 3 shows some results based on data from the EnglishHnd dataset. It is shown that SFCRF can classify the images even when they are distorted by the high level of the noise (i.e. 300%).

The SFCRFs are also examined with grayscale images from the Weizmann segmentation dataset [18]. The goal is to segment the noisy images into foreground and background.

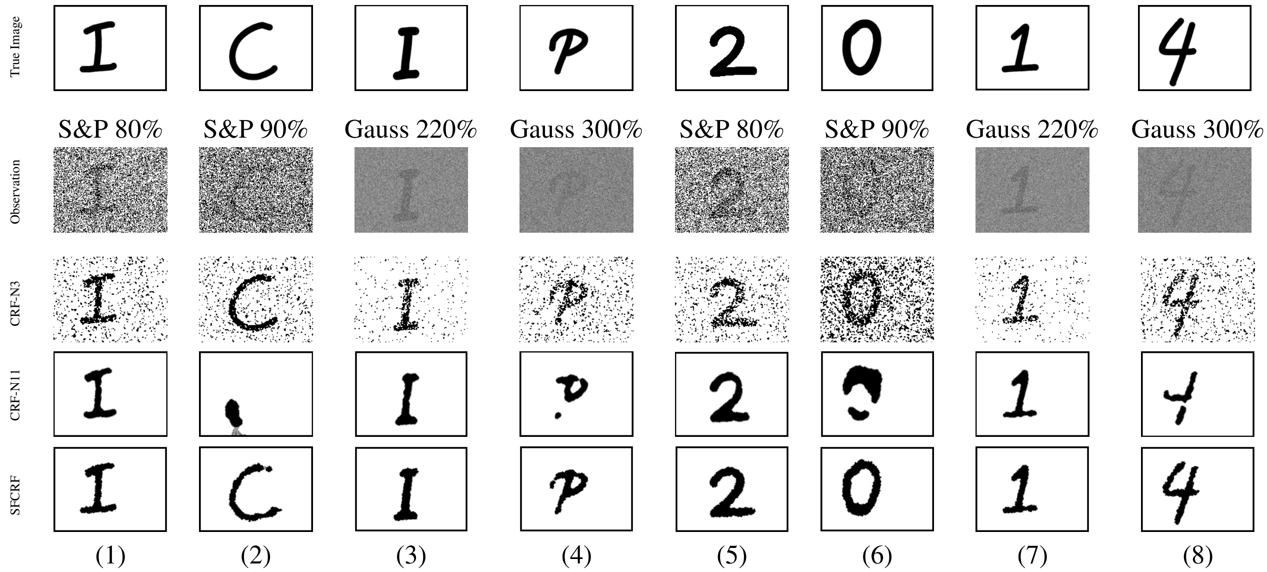


Fig. 3. Qualitative results of SFCRF; the proposed algorithm is examined based on two noise types with two strengths. The results clearly show how the SFCRF outperforms both local and non-local CRFs.

Table 1. Quantitative results based on the EnglishHnd dataset [17]. The proposed framework is examined by two noise types with two different levels. The SFCRF is compared with the regular CRF (CRF-N3) and a CRF with a neighborhood size of 11 (CRF-N11). The per-iteration run time of each method is reported; all methods were run with an equal number of iterations.

	CRF-N3	CRF-N11	SFCRF
Salt & Pepper (80%)	0.488	0.872	0.931
Salt & Pepper (90%)	0.235	0.313	0.859
Gaussian (220%)	0.566	0.818	0.895
Gaussian (300%)	0.391	0.646	0.842
Time per Iteration (s)	0.04	3.85	2.70

Figure 4 shows the segmentation results of SFCRFs for noisy images in comparison with CRF-N11. The images are distorted with a high level of noise to the point that the objects are not obvious in the images. The results demonstrate that SFCRFs can detect the object in low SNR images.

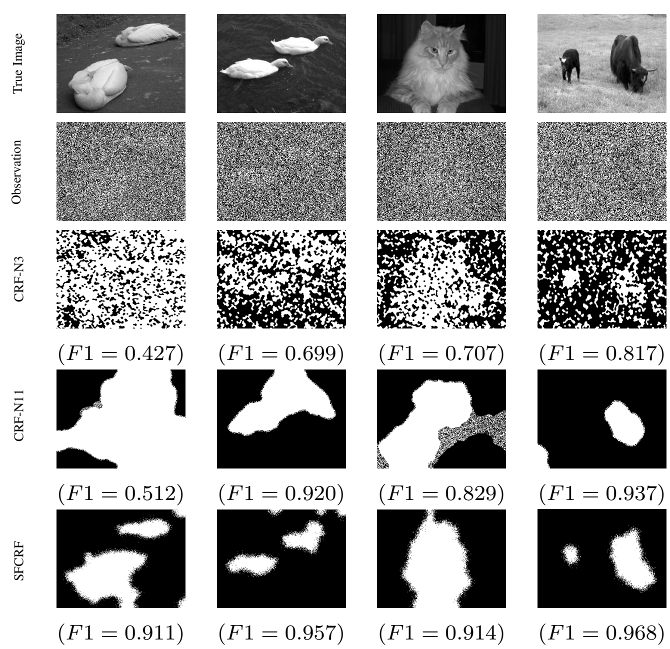


Fig. 4. The effectiveness of the SFCRF in the segmentation of noisy images from the Weizmann dataset. The first row shows the true images, the second the images after distortion, and the three last rows the segmentation results of the CRF-N3, CRF-N11 and the SFCRF respectively. The images are distorted with salt & pepper noise at 90%. The corresponding F1-score for each result is shown after the image. The results clearly illustrate the applicability of the SFCRF to natural images.

4. CONCLUSION & FUTURE WORK

In this paper we proposed a new framework of fully connected conditional random fields with a tractable computational inference stage by incorporating a novel stochastic clique structure. Through the stochastic clique, the dense graph structure is converted to a sparse one, significantly reducing the computational complexity while preserving the effectiveness of a fully connected random field. Results show that the proposed approach performs well for image labelling.

5. REFERENCES

[1] M. Malfait and D. Roose, "Wavelet-based image denoising using a markov random field a priori model," *IEEE Transactions on Image Processing*, 1997.

[2] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in Neural Information Processing Systems*, 2011.

[3] D. Kong, M. Han, W. Xu, H. Tao, and Y. Gong, "A conditional random field model for video super-resolution," in *International Conference on Pattern Recognition*. IEEE, 2006.

[4] M. Wainwright and M. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, 2008.

[5] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *International Conference on Machine Learning*, 2001.

[6] F. Sha and F. Pereira, "Shallow parsing with conditional random fields," in *North American Chapter of the Association for Computational Linguistics on Human Language Technology*. Association for Computational Linguistics, 2003.

[7] H. Zhao, C. Huang, M. Li, and B. Lu, "Effective tag set selection in chinese word segmentation via conditional random field modeling," in *Proceedings of PACLIC*, 2006.

[8] X. He, R. Zemel, and M. Carreira-Perpinan, "Multiscale conditional random fields for image labeling," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2004.

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

[10] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," in *International Conference on Computer Vision*. IEEE, 2007.

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

[12] N. Campbell, K. Subr, and J. Kautz, "Fully-connected crfs with non-parametric pairwise potentials," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2013.

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

[14] P. Erdos and A. Renyi, "On random graphs i.," *Publ. Math. Debrecen*, 1959.

[15] C. Sutton and A. McCallum, "An introduction to conditional random fields for relational learning," *Introduction to statistical relational learning*, 2007.

[16] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984.

[17] T. de Campos, B. Babu, and M. Varma, "Character recognition in natural images," in *International Conference on Computer Vision Theory and Applications*, 2009.

[18] A. Sharon, G. Meirav, B. Ronen, and B. Achi, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2007.