Image denoising with parallel Markov random fields

Matthew Larkins[1], Talita Perciano[2]
University of California, San Diego[1]
Lawrence Berkeley National Laboratory[2]

ABSTRACT

Markov random fields (MRFs) are a versatile tool for image processing. They allow robust modeling of image processing problems, but their optimization is computationally expensive. The computational problem can be overcome by parallelizing the optimization process, but this is difficult to do because every pixel or region in the graph is dependent on others. Researchers have proposed and used the Linear and Parallel (LAP) algorithm for breaking up the graph that maintains the necessary relationships between random variables and does not limit models to being lower order with submodular energy functions. The authors propose a parallel MRF model that uses LAP for breaking up the graph and applies a variation of the Iterated Conditional Modes (ICM) model for image denoising on the subgraphs. First, the authors briefly describe the LAP algorithm. Then, the ICM method is described with a slight modification to work with the subgraphs produced by LAP. To evaluate the model's performance, the authors present a runtime scalability analysis of the model and compare its accuracy for removing Poisson and salt and pepper noise against three leading classical denoising methods: Total Variation (TV), bilateral filtering, and a wavelet domain denoising method. Finally, the authors present conclusions and directions for future research.

I. INTRODUCTION

Because of the physical limitations of current image capturing devices, noise in unprocessed images is inevitable. For example, random electron spin or EM field fluctuations can mix with the desired signal in sensitive input sensors, resulting in some amount of erroneous values in the final image. This noise can make image analysis difficult. For example, classifying a very noisy image with a Convolutional Neural Network (CNN) results in a much less confident and often incorrect prediction where a denoised version could have been successfully classified. Image denoising is therefore an important preprocessing step for many image processing tasks, such as classification, localization, and analyzing images with the visual cortex.

Markov random fields are a versatile tool for image processing. They allow for intuitive modeling of an image processing problem and a standard approach to its solution involving defining a set of random variables over the image and changing them in the way that minimizes an energy function of those variables. One drawback of Markov random fields is their computational complexity. Optimizing an energy function and estimating parameters for an MRF model can be prohibitively complex for larger images.

One solution to the computational problem is to parallelize the MRF optimization and parameter estimation processes. Parallelization itself is a challenging problem because breaking up a graph in an MRF model requires maintaining dependencies between the nodes in the graph. Mizrahi et al developed the Linear and Parallel (LAP) algorithm for breaking up graphs that preserves the required dependencies[1]. Furthermore, it allows for more robust modeling that other high-performance methods did not allow. For example, LAP partitioning can be used on MRF models with nonsubmodular energy functions, whereas move-making methods such as graph cut require the model to have a submodular energy function.

Perciano et. al implemented MPI-PMRF, a parallel segmentation model using LAP to partition the MRF before performing parameter estimation and optimization in parallel[2]. They showed the runtime performance gains expected with parallelization with 95% accuracy in segmentation results on synthetic data.

The authors build on the LAP and MPI-PMRF work with PMRF-D, a parallel MRF model for image denoising. The model uses the LAP algorithm to break up the graph and implements the classical iterative conditional modes (ICM)[3] model to perform denoising on the subgraphs in parallel. The authors show similar runtime performance gains with parallelization and find that the denoising accuracy compares competitively with other classical methods including total variation (TV) and a wavelet-based denoising method.

II. LINEAR AND PARALLEL

The framework represents an image as a 2-D graph with a smoothness-based prior distribution as does the classical ICM algorithm. It then implements LAP and distributes those graph partitions over multiple nodes for parallel optimization. This algorithm breaks up the graph by selecting a maximal clique (e.g. the green nodes in Fig. 1a) and grouping it with every node in the graph that has a neighbor in the maximal clique (e.g. the yellow nodes in

Fig. 1a). This auxiliary MRF has enough context for the optimization process to be done for the random variables represented by the base nodes. This process is continued until every node in the graph is represented as a base node in an auxiliary MRF.

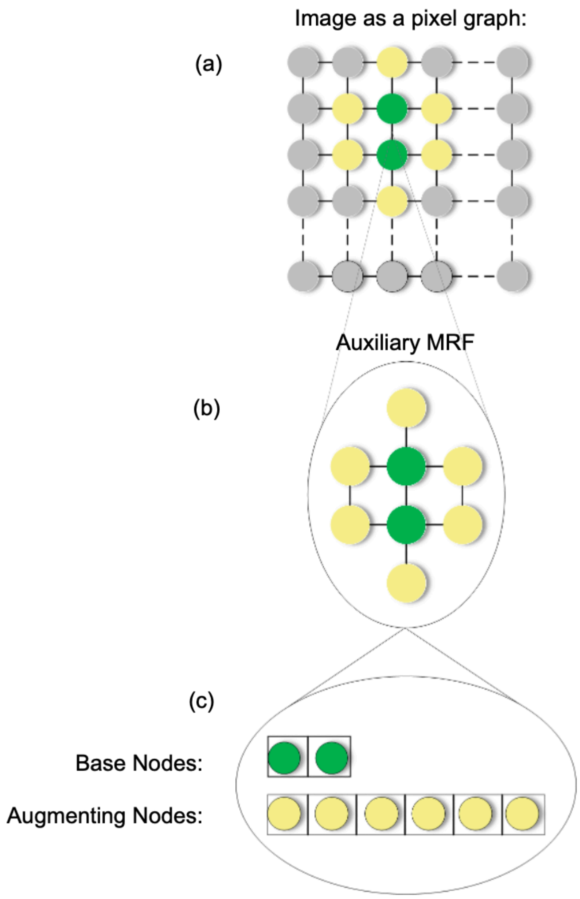

Figure 1: LAP Visualization

III. ITERATED CONDITIONAL MODES

To denoise the auxiliary MRFs, a variation of ICM optimization is applied on each subgraph, which may be distributed out to multiple compute nodes in a system. See Algorithm 1. The ICM algorithm assumes a smoothness-based prior and penalizes differences between spatially nearby pixels in the image. This penalty is made by the pairwise term, which is the sum weighted by λ in the energy function. The prior also assumes that there is some probability that a given pixel value is correct. The energy function ensures this by penalizing proposed changes to a pixel with the data term limited by ψ_{max}. The algorithm iteratively proposes new values for each pixel and chooses the proposed value that results in the minimum energy. In the classical ICM method, the pairwise term consists of adjacent pixels in a 4-connected or 8-connected graph. In PMRF-D, however, pixels may not necessarily be adjacent. Take, for example, the top-most base node (green) and the bottom-most augmenting node (yellow) in Figure 1a. These two nodes are separated by the other base node in the auxiliary MRF, and yet the augmenting node will be considered in the pairwise term for the top base node. In practice, this slight variation in the traditional ICM method produces results comparable to the original method, as the pixels are near enough to be expected to be similar in value in a smooth image. Notice that only the base nodes are assigned to. The augmenting nodes provide the required context for the base nodes, but the augmenting nodes themselves are separated from their adjacencies (i.e. their dependencies)[1].

Algorithm 1: Iterative Conditional Modes

$let\ \boldsymbol{A} =$ the set of Augmenting Nodes
$let\ \boldsymbol{B} =$ the set of Base Nodes
$let\ \boldsymbol{X} = \{x \in \mathbb{Z} \mid 0 \le x \le 255\}$
$for\ b_i\ in\ \boldsymbol{B}$:
    $for\ x_i\ in\ \boldsymbol{X}$:
        $b_i = argmin(E(\boldsymbol{A}, b_i, x_i))$
    $end\ for$
$end\ for$
$where$

$$E(\boldsymbol{A}, b_i, x_i) = \min((b_i - x_i)^2, \psi_{max}) + \lambda \textstyle\sum_j (a_j - x_i)^2$$

$where\ \psi_{max}$ the max penalty for the data term
$where\ \lambda$ the pairwise term's weighting term

IV. PERFORMANCE ANALYSIS

PMRF-D's performance was analyzed for runtime scalability and accuracy. To measure the framework's scalability, the research team ran it on an 8-node cluster with number of processes varying sequentially from 1 to 8 and then by increments of 10 from 10 to 30. Each trial was run 3 times and the runtime for each number of processed was recorded as the average of these three times. The input image was the 788 x 610 grayscale image seen in Figure 3b.

To measure the framework's accuracy, the research group used three metrics for quantifying denoising performance: structural similarity (SSIM)[5], peak signal-to-noise ratio (PSNR), and mean squared error (MSE). In all of these metrics, the denoised output was compared to the original. The research group tested the framework on two types of noise, salt and pepper and Poisson. Salt and pepper noise was added to a microscopic pollen sample image, and Poisson noise was added to a ceramic sample image taken from a micro-CT imaging device. To see how the framework compares to other commonly used methods, the research group selected three leading denoising methods: total variation (TV)[6], BayesShrink[7] (a wavelet domain method), and bilateral filtering[8].

V. RESULTS

A. Scalability

The scalability analysis (Figure 2) shows runtime performance gains as the number of processes is increased. Doubling the number of processes from 1 to 2 and then from 2 to 4 corresponds to halving the runtime. The runtime tends to decrease less quickly as the number of threads is increased beyond 4. This is expected to be because of the limited number of compute nodes available to the testing system. Since the system has 8 nodes, doubling the number of processes from 4 to 8 puts a greater load on the system than doubling it from 1 to 2 and from 2 to 4 did. It is expected that the observed halving of runtime when doubling the number of threads will be seen beyond 8 nodes given a system with more processors available. The analysis continued beyond 8 nodes with only minor changes in performance, and this is believed to be because the system nearly reaches its capacity once it is using all 8 of its nodes.

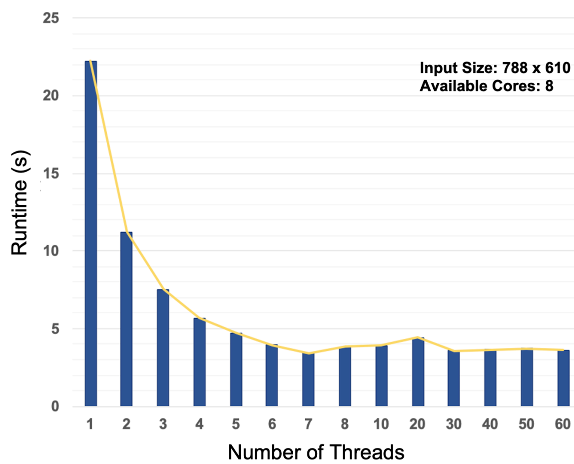

Figure 2: Scalability Analysis

B. Accuracy

The research team first compared the accuracy of PMRF-D on an 84 x 84 RGB pollen grain image with added salt and pepper noise against total variation, wavelet, and bilateral filtering denoising methods (Fig. 3a). In this experiment, the TV implementation performed best, with PMRF-D performing second best, followed by the BayesShrink wavelet method and bilateral filtering. This ranking was followed by all three metrics used (SSIM, PSNR, and MSE).

The research team then compared PMRF-D's accuracy on a 789 x 611 8-bit grayscale ceramic sample image with Poisson noise added against the same three methods (Fig 3b). In this experiment, BayesShrink performed best, followed by PMRF-D, TV, and bilateral filtering.

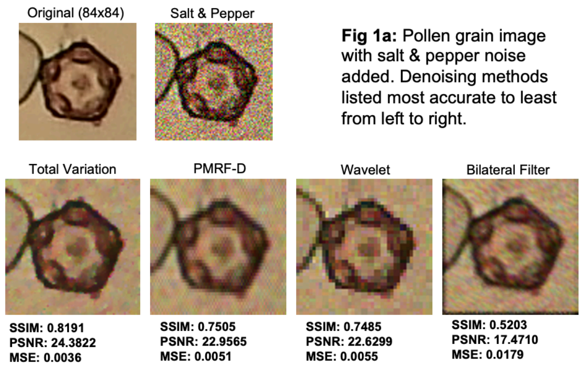

Fig 1a: Pollen grain image with salt & pepper noise added. Denoising methods listed most accurate to least from left to right.

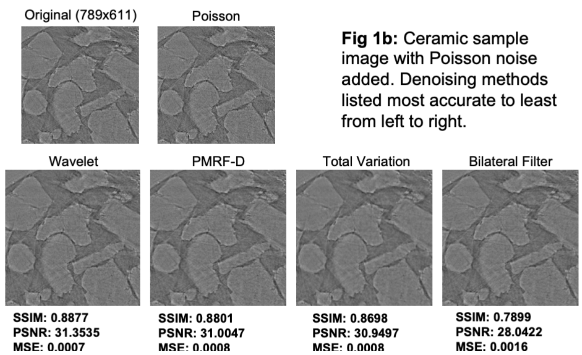

Fig 1b: Ceramic sample image with Poisson noise added. Denoising methods listed most accurate to least from left to right.

VI. DISCUSSION

Denoising images can be prohibitive in cases where the images are large. The runtime analysis experiment presented here suggests that given enough computational power, large images can be denoised in a relatively short amount of time. This can be useful, for example, when training convolutional neural networks to denoise images given a large set of noisy data with no ground truth images to pair with the noisy ones. Other use cases for PMRF-D include using it as a preprocessing step before running classification or object tracking on images.

In terms of accuracy, PMRF-D outperformed TV in one experiment, BayesShrink in another, and bilateral filtering in both. In the case where wavelet denoising outperformed PMRF-D, the difference in performance was only slight. This suggests that PMRF-D may be preferable to TV and wavelet methods in some cases, especially for large images because of the runtime performance scaling that PMRF-D offers.

VII. CONCLUSIONS AND FUTURE WORK

The research team found that PMRF-D delivers runtime scalability that allows it to denoise large images relatively quickly while outputting results with accuracy comparable to leading methods. Other directions for research include implementing the model with an energy function that preserves edges, as ICM blurs edges indiscriminately. A more powerful scalability analysis is also necessary, since the one presented here was limited by the number of compute nodes available.

VIII. ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships (SULI) program.

IX. REFERENCES

[1] Y. D. Mizrahi *et. al.*, “Linear and Parallel Learning of Markov Random Fields,” *IEEE Trans Pattern Anal Mach Intell*, vol. 31, no. 10, pp. 1871-1887, Oct 2010.

[2] T Perciano *et. al.*, “Reduced-Complexity Image Segmentation Under Parlallel Markov Random Field Formulation Using Graph Partitioning”, *2016 IEEE International Conference on Image Processing*, pp.1259-1263, September 2016.

[3] Besag, Julian. "On the Statistical Analysis of Dirty Pictures." *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 48, no. 3, 1986, pp. 259–302. *JSTOR*, www.jstor.org/stable/2345426. Accessed 31 July 2020.

[4] Blake, Andrew and Kohli, Pushmeet, "Introduction to Markov Random Fields," *Markov Random Fields for Image Processing,* pp. 1-6, MIT Press, 2011

[5] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13, 600-612. https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf, DOI:10.1109/TIP.2003.819861

[6] Chambolle, A., An algorithm for total variation minimization and applications, Journal of Mathematical Imaging and Vision, Springer, 2004, 20, 89-97.

[7] Chang, S. Grace, Bin Yu, and Martin Vetterli. "Adaptive wavelet thresholding for image denoising and compression." Image Processing, IEEE Transactions on 9.9 (2000): 1532-1546. DOI: 10.1109/83.862633

[8] Tomasi, C. and Manduchi, R., "Bilateral Filtering for Gray and Color Images." IEEE International Conference on Computer Vision (1998) 839-846. DOI:10.1109/ICCV.1998.710815