

Markov Chain Monte Carlo Combined with Deterministic Methods for Markov Random Field Optimization

Wonsik Kim and Kyoung Mu Lee

Department of EECS, ASRI, Seoul National University, 151-742, Seoul, Korea

{ultra16, kyoungmu}@snu.ac.kr

Abstract

Many vision problems have been formulated as energy minimization problems and there have been significant advances in energy minimization algorithms. The most widely-used energy minimization algorithms include Graph Cuts, Belief Propagation and Tree-Reweighted Message Passing. Although they have obtained good results, they are still unsatisfactory when it comes to more difficult MRF problems such as non-submodular energy functions, highly connected MRFs, and high-order clique potentials. There have also been other approaches, known as stochastic sampling-based algorithms, which include Simulated Annealing, Markov Chain Monte Carlo and Population-based Markov Chain Monte Carlo. They are applicable to any general energy models but they are usually slower than deterministic methods. In this paper, we propose new algorithms which elegantly combine stochastic and deterministic methods. Sampling-based methods are boosted by deterministic methods so that they can rapidly move to lower energy states and easily jump over energy barriers. In different point of view, the sampling-based method prevents deterministic methods from getting stuck at local minima. Consequently, a combination of both approaches substantially increases the quality of the solutions. We present a thorough analysis of the proposed methods in synthetic MRF problems by controlling the hardness of the problems. We also demonstrate experimental results for the photomontage problem which is the most difficult one among the standard MRF benchmark problems.

1. Introduction

Markov Random Field (MRF) models are of fundamental importance in computer vision. Many vision problems have been successfully formulated in MRF optimization. They include stereo matching, segmentation, denoising and inpainting, to mention just a few. Recently, Szeliski *et al.* [18] presented a comprehensive review of the standard

MRF-based vision problems and the comparative results of existing optimization methods.

The general formulation of the MRF models is as follows. Given a graph $G = \langle V, E \rangle$, where V is the set of nodes and E is the set of edges, the energy function is given by

$$E(\mathbf{x}) = \sum_{s \in V} \theta_s(x_s) + \sum_{\langle s, t \rangle \in N} \theta_{st}(x_s, x_t), \quad (1)$$

where N is a neighborhood system and the first term $\theta_s(x_s)$ is called the unary term or data term and is defined in various ways depending on the applications. For example, in stereo problems it can be intensity difference, sum of squared difference or Birchfield-Tomasi measure of corresponding pixels. In denoising problems, it can be the intensity difference between the true and the noisy pixels. In the segmentation problem, it can be the color difference between a single pixel and the histogram of the segment it belongs to. The second term $\theta_{st}(x_s, x_t)$ is called the pairwise term or smoothness term. This term usually encodes the prior knowledge into the energy function. In most applications, smoothness regularization constraints are commonly used, which compel the solution to be piecewise smooth. Widely-used smoothness models include the Potts model, the truncated linear model and the truncated quadratic model. In general cases, the optimization of the above energy function is known to be NP-hard problem [5].

There have been a lot of researches on minimizing the aforementioned energy function. For the standard grid graphs, we can see that many algorithms have a good performance [18]. The existing methods for solving this minimization can be divided into two approaches: deterministic and sampling-based methods. Some of the well-known deterministic methods are move-making algorithms. Move-making algorithms iteratively make local moves to explore the solution space. They include Iterated Conditional Modes (ICM), the Gradient Descent Method and Graph Cuts [5, 6, 11]. Graph Cuts, including α -Expansion, $\alpha\beta$ -Swap and Quadratic Pseudo-Boolean Optimization (QPBO), are the state of the art algorithms. They

iteratively optimize the binary sub-problems of the original problem. They are fast, accurate and even find global optima for some classes of functions. Another important class of deterministic methods is the message passing approach. It includes Belief Propagation (BP) [19] and Tree Reweighted Message Passing (TRW) [10, 20]. BP was originally developed for graphs without cycles. Although there is no guarantee of convergence in the case of the graph with cycles, it has been successfully applied to vision problems. One of the important properties of TRW is that it gives a lower bound of the energy function, which can be used to check optimality of the solutions.

Stochastic approaches such as sampling-based methods have also been applied to the MAP-MRF based vision problems. Sampling-based methods were originally developed to generate samples from a given target distribution or to integrate functions in high dimensional spaces. Swendsen-Wang Cuts is proposed to solve image segmentation and stereo problems [2, 3]. Recently, Kim *et al.* [9] proposed a more advanced MCMC method called Population-based MCMC (Pop-MCMC) to optimize a plane-based stereo energy model. And Jung *et al.* [7, 8] proposed window annealing method to increase mixing ratio of the MCMC method.

The story is different, however, when it comes to more difficult MRF problems. More difficult MRF models are inevitable to incorporate realistic image priors into the models. (*e.g.*, occlusion terms in stereo and texture information in denoising and segmentation.) Examples of more difficult MRF models include non-submodular functions, globally conditioned MRF models, highly connected MRF models and higher-order clique potentials. To those difficult examples, most existing methods are not applicable, and even with some applicable algorithms the results are far from the global optimum.

Note that α -Expansion and $\alpha\beta$ -Swap cannot be applied to non-submodular energy functions. To solve those functions, we should truncate the non-submodular terms. Consequently, it seriously degrades the quality of solutions in difficult MRF problems. In contrast, QPBO can solve those functions without truncation. As the difficulty of problems increases, however, it produces more unlabeled pixels, which yields unsatisfactory results. The number of unlabeled pixels depends on the strength of unary and pairwise terms, the number of non-submodular terms and the connectivity of the graph structure [16]. To resolve this problem, probing is proposed by Rother *et al.* [16] but it still leaves unlabeled pixels. In addition, every Graph Cut based method can handle only pairwise graphs. Although Boros and Hammer [4] showed how to reduce higher-order clique potentials into equivalent pairwise potentials, the number of terms in the energy function grows exponentially with the clique size. Message passing algorithms are also degraded as the difficulties of the problems increase. The gap

between the solutions and the lower bounds of TRW-S can be efficient measures of qualities of the solutions. As Komodakis and Paragios [12] mentioned, the solutions and the lower bounds do not converge in difficult MRF problems. Recently, efficient BP methods for higher-order clique potential are proposed by Nwogu and Corso [15] and Lan *et al.* [13].

Sampling-based methods also have weaknesses. Although applicable to any class of MRF problems, they are usually slower than deterministic methods even in the simple MRF problems [9], and do not lower the energy state substantially [7, 8]. If difficulties of the problems increase, we do not think they can solve the problems in a practical timescale.

Therefore, we definitely need a more efficient optimization technique to cope with difficult MRF vision problems. Our main idea is to combine sampling-based and deterministic methods so that we can take advantages of the strength of both approaches. Our new algorithm is mainly inspired by the work of Strens *et al.* [17]. They used direct search optimization (downhill simplex method and differential evolutions) in the framework of Pop-MCMC to increase sampling performance. Our paper is organized as follows. We present the advantages of combining stochastic and deterministic algorithms. In section 3, Pop-MCMC is briefly introduced. Then, we propose the new combined method in Section 4. Section 5 gives the experimental results. The final section concludes the paper with discussion.

2. Toward difficult MRF problems: combination of stochastic and deterministic methods

To deal with difficult MRF problems, we propose to combine deterministic methods with the sampling-based stochastic method. The sampling-based method is boosted by combination with deterministic methods. The sample can rapidly move into lower energy state owing to the deterministic methods. Moreover, it can effectively jump from one basin to another over the energy barrier. Consequently, this property increases the mixing rate and yields faster convergence and better solutions. On the other hand, the sampling-based method helps deterministic methods not to be stuck in local minima. In addition, in terms of the landscape of the solution space, the combination reduces the number of local minima. The reason is as follows. The number and the location of local minima vary with respect to the applied methods. This is because each method has different neighborhood structures in the solution space owing to its own operators or search techniques. This elimination of local minima is illustrated in Fig 1. If we effectively combine two methods in Fig 1 (a) and (b), all the minima but the overlapped ones will be eliminated. Although this

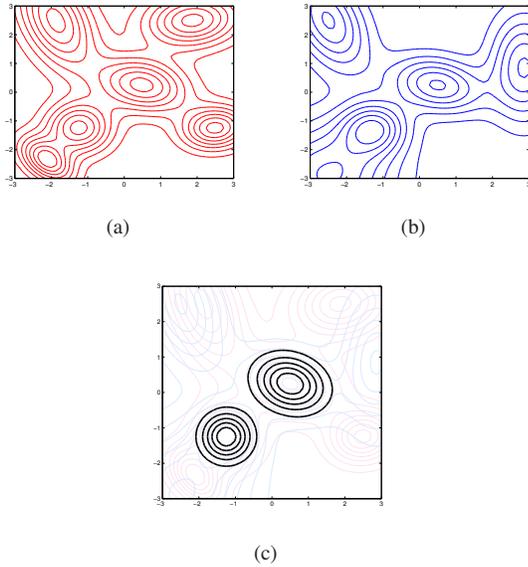


Figure 1. The energy landscape is altered by the combination of multiple methods: (a) The structure of local minima viewed from an algorithm, (b) another algorithm, and (c) combined algorithm. In (c), many local minima are eliminated and only two overlapping local minima have survived. They are drawn with black thick contours.

scenario of elimination is not exactly the same as in the proposed approach, it may be helpful in getting intuition on the advantages of the combination of different methods.

The whole procedure of the proposed algorithm is based on Pop-MCMC. As in Pop-MCMC, the proposed algorithm runs several chains in parallel and makes samples move by the proposals under the Metropolis-Hastings rule. In this work, we design new effective proposals which exploit deterministic methods.

3. Population-based MCMC

Markov Chain Monte Carlo (MCMC) methods have been used to sample target distributions. Along with Simulated Annealing, MCMC has also been used to obtain optimum samples of target functions. In MCMC, a new sample is drawn from the previous sample with a local transition probability, based on the Markov chain. Since most MCMC methods allow only local moves, in a large solution space it takes a very long time to reach the global optimum. To overcome the limitations of MCMC, recently Population-based MCMC has been applied to the vision problem [9]. It obtained lower energy state much faster than other existing sampling methods.

Pop-MCMC or evolutionary Monte Carlo is a stochastic simulation method that combines a population of Metropolis-Hastings samplers and Evolutionary Algorithms to improve the performance of MCMC samplers.

Pop-MCMC generates multiple chains in parallel with different temperatures, and multiple samples are drawn at the same time. Samples can exchange information with each other. This enables global moves of samples which consequently makes the mixing rate of drawn samples faster. In terms of optimization, the fast mixing rate means fast convergence to the global optimum.

Let us define the target distribution of i th chain as follows.

$$\pi_i(\mathbf{x}) = \pi(\mathbf{x})^{\frac{1}{T_i}}, \quad (2)$$

where $\pi(\mathbf{x})$ is an original target distribution, and T_i is the temperature of the i th chain. In the chain with high temperature, the target distribution is nearly flat, where the heights of barriers between local optima are very low. Therefore, the samples in such chains can freely wander in contrast to the samples in a chain with low temperature. By exchanging these higher-temperature configurations with the configuration of a low temperature of our interest, we can allow the low temperature simulation to sample configurations much more efficiently than with local Metropolis updates only. This leads to a faster mixing rate between samples, and helps the escape from local minima. Given an original target distribution $\pi(\mathbf{x})$, a new expanded target distribution is defined as follows.

$$\pi^*(\mathbf{x}_{1:N}) = \prod_{i=1}^N \pi_i(\mathbf{x}_i), \quad (3)$$

where N is the number of chains to use. We assume that $\pi_k \equiv \pi$ for at least one chain $k \in \{1, \dots, N\}$. $\mathbf{x}_{1:N} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ is a population composed of samples of N chains. Each component \mathbf{x}_i in the vector $\mathbf{x}_{1:N}$ is labeled as a *chromosome*. The goal of Pop-MCMC is to generate samples $\mathbf{x}_{1:N}$ which follow the new target distribution π^* . A collection of chromosomes from the k th chain, which has the target distribution $\pi_k = \pi$, is what we want to obtain in the end.

In general, Pop-MCMC employs three types of moves [9]: mutation, exchange and crossover. Mutation move is same as the conventional Metropolis-Hastings move for a single chain MCMC. Exchange and crossover moves help samples exchange information with each other. This makes it possible to efficiently explore the solution space.

4. Combined Method

As we mentioned above, by combining stochastic and deterministic methods, we can take advantages of both approaches. Careless combination of the two methods becomes, however, problematic. If we simply apply deterministic methods as the kernel of the sampling method,

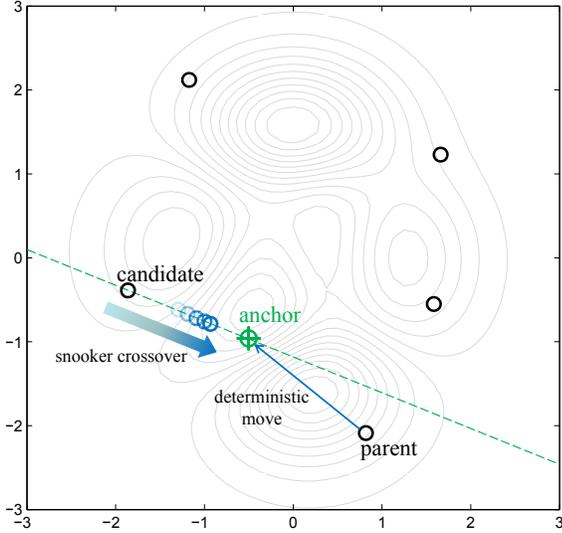


Figure 2. Dynamic anchor-based proposal

it might violate the reversibility condition of the Markov Chain Monte Carlo. Consequently, it is impossible to satisfy detailed balance. That is, we are not able to sample from the target probability function in that way. In the following subsections, we explain three novel algorithms in which the Pop-MCMC method and the deterministic methods are elegantly combined through a new proposals by the snooker crossover move [14]. Note that in our current framework, we do not use the mutation and exchange moves since we have empirically found that our new crossover proposal is so powerful and other moves are redundant.

4.1. Combination of MCMC and move-making deterministic algorithms

In this subsection, we introduce our first algorithm called MCMC-D (Markov Chain Monte Carlo with Dynamic anchor-based proposal) which effectively combines the sampling-based method and deterministic methods. Our MCMC-D algorithm has a similar structure to Pop-MCMC. It runs multiple chains at different temperatures T_i . Given the energy function $E(\mathbf{x})$, we first change the domain of \mathbf{x} from the discrete domain to the continuous domain according to:

$$E_{real}(\mathbf{x}) = E(\text{round}(\mathbf{x})). \quad (4)$$

From now on, \mathbf{x} is a continuous random variable. And then, we define probability distribution of chain i as

$$\pi_i(\mathbf{x}_i) = \frac{1}{Z} \exp\left\{-\frac{E_{real}(\mathbf{x}_i)}{T_i}\right\}, \quad (5)$$

where Z is a normalizing constant. Note that π_i is defined on real number space. The appropriate sequence of the tem-

peratures depends on the given energy function. It is empirically determined. In this method, we propose a new scheme to make a proposal which exploits move-making deterministic algorithms. Using deterministic algorithms, we dynamically generate anchors, so that it is called *dynamic anchor*-based proposal. It satisfies reversibility as well as detailed balance. This proposal is iteratively applied to update populations until convergence. Now, we explain how dynamic anchor-based proposal works.

The dynamic anchor-based proposal is well illustrated in Fig 2. The black circles represent the current population. Given the current population $\mathbf{x}_{1:N} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, two samples are randomly chosen to be a candidate \mathbf{x}_c and a parent \mathbf{x}_p , respectively. Then, we apply a deterministic move to the parent. That is, with the parent as the initial point, we perform single or multiple iterations of any move-making deterministic algorithm (*e.g.* α -Expansion, $\alpha\beta$ -Swap, and QPBO). The result of the deterministic move from the parent is considered as an anchor point. We call it a dynamic anchor point because it is generated and destroyed dynamically while the algorithm is running. With this dynamic anchor point and the candidate \mathbf{x}_c , we perform snooker crossover. A newly-generated sample \mathbf{y}_c lies on the line going through the anchor and the candidate according to:

$$\mathbf{y}_c = \mathbf{x}_c \kappa \exp(s) + M(\mathbf{x}_p)(1 - \kappa \exp(s)), \quad (6)$$

where s and κ are control parameters of snooker crossover and $M(\cdot)$ is the move by the move-making deterministic algorithm. s is the random variable taken from the predefined set S with probability distribution $r(s)$. The set S can be designed as any set closed under the operator \bar{s} , which is defined by $1 - s$. The parameter s controls the distance between the newly-generated sample and the anchor. Small s results in the new sample being close to the anchor and large s results in the new sample being far away from the anchor. κ can be fixed either by $+1$ and -1 or randomly chosen among $+1$ and -1 with equal probability. κ will decide if the new sample, started from the candidate, passes over the anchor or not. When κ is -1 , the newly-generated sample lies on the ray from the anchor in the opposite direction to the candidate. That is, the new sample passes over the anchor point. When κ is $+1$, the newly-generated sample lies on the ray from the anchor through the candidate. When s is 0 and κ is $+1$, the candidate \mathbf{x}_c and the new sample \mathbf{y}_c are exactly the same.

The candidate \mathbf{x}_c is substituted with the new sample \mathbf{y}_c according to the Metropolis-Hastings rule with the acceptance probability:

$$\alpha = \min(1, \gamma), \quad (7)$$

where

Algorithm 1 MCMC-D algorithm

- 1: (Initialize)
 - 2: Initialize the population $X_{1:N}$
 - 3: Set the temperatures $T_1 < T_2 < \dots < T_N$
 - 4: **repeat**
 - 5: $c \sim \{1, 2, \dots, N\}$
 - 6: $p \sim \{1, 2, \dots, N\} - \{c\}$
 - 7: $A \leftarrow M(\mathbf{x}_p)$
 - 8: (Snooker crossover)
 - 9: $\kappa \sim \{+1, -1\}$
 - 10: $s \sim S$
 - 11: $\mathbf{y}_c \leftarrow \mathbf{x}_c \cdot \kappa \exp(s) + A \cdot (1 - \kappa \exp(s))$
 - 12: Determine whether accept the new population or not by the Metropolis-Hastings rule.
 - 13: **until** The algorithm converges.
-

$$\begin{aligned}
 \gamma &= \frac{\pi_c(\mathbf{y}_c)\pi_p(\mathbf{y}_p)}{\pi_c(\mathbf{x}_c)\pi_p(\mathbf{x}_p)} \cdot \frac{q(\mathbf{x}_c, \mathbf{x}_p|\mathbf{y}_c, \mathbf{y}_p)}{q(\mathbf{y}_c, \mathbf{y}_p|\mathbf{x}_c, \mathbf{x}_p)} \\
 &= \frac{\pi_c(\mathbf{y}_c)}{\pi_c(\mathbf{x}_c)} \cdot \frac{q(\mathbf{x}_c|\mathbf{y}_c, \mathbf{x}_p)}{q(\mathbf{y}_c|\mathbf{x}_c, \mathbf{x}_p)} \\
 &= \frac{\pi_c(\mathbf{y}_c)}{\pi_c(\mathbf{x}_c)} \cdot \frac{r(1-s)}{r(s)} \\
 &= \exp\left[\frac{E(\mathbf{x}_c) - E(\mathbf{y}_c)}{T_c}\right] \cdot \frac{r(1-s)}{r(s)}.
 \end{aligned} \tag{8}$$

Note that the reverse move is attained by selecting $1 - s$ from S .

Instead of a single candidate, we can also pick multiple candidates at each iteration. The opposite extreme of using a single candidate is to take all the samples as candidates except the parent. Our MCMC-D that employs a move-making deterministic algorithm is summarized in Algorithm 1.

4.2. Combination of MCMC and non-move-making deterministic algorithms

MCMC-D algorithm exploits move-making deterministic algorithms such as α -Expansion, $\alpha\beta$ -Swap, and QPBO. This method cannot, however, exploit message passing methods since message passing methods do not make moves. In this subsection, we propose a new algorithm called MCMC-S (Markov Chain Monte Carlo with Static anchor-based proposal) which exploits non-move-making algorithms such as message passing algorithms. The idea is similar to MCMC-D. The only difference is the anchor points. In MCMC-S, we initially run message passing methods and save solutions before the algorithm starts. Those solutions are used as the anchor points while the algorithm is running. We call these anchor points the *static*

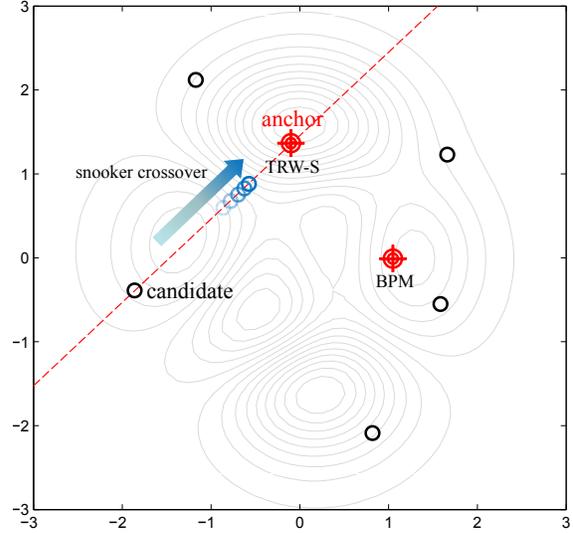


Figure 3. Static anchor-based proposal

Algorithm 2 MCMC-S algorithm

- 1: (Initialize)
 - 2: Initialize the population $X_{1:N}$
 - 3: Set the temperatures $T_1 < T_2 < \dots < T_N$
 - 4: Run message passing methods to get solutions F_1, F_2, \dots, F_K
 - 5: **repeat**
 - 6: $c \sim \{1, 2, \dots, N\}$
 - 7: $k \sim \{1, 2, \dots, K\}$
 - 8: $A \leftarrow F_k$
 - 9: (Snooker crossover)
 - 10: $\kappa \sim \{+1, -1\}$
 - 11: $s \sim S$
 - 12: $\mathbf{y}_c \leftarrow \mathbf{x}_c \cdot \kappa \exp(s) + A \cdot (1 - \kappa \exp(s))$
 - 13: Determine whether accept the new population or not by the Metropolis-Hastings rule.
 - 14: **until** The algorithm converges.
-

anchor points since they are fixed until the algorithm terminates. We call the new proposal the static anchor-based proposal. The static anchor-based proposal is illustrated in Fig 3, and MCMC-S algorithm is summarized in Algorithm 2.

4.3. Combination of MCMC and general deterministic algorithms

In this subsection, we propose our third algorithm called MCMC-GD (Markov Chain Monte Carlo combined with General Deterministic methods) which unifies MCMC-D and MCMC-S. In MCMC-GD algorithm, we use both of dynamic and static anchor-based proposals. At each iteration we choose a random number U uniformly from interval $[0, 1)$, and compare U with predefined dynamic anchor-

Algorithm 3 MCMC-GD algorithm

```
1: (Initialize)
2: Initialize the population  $X_{1:N}$ 
3: Set the temperatures  $T_1 < T_2 < \dots < T_N$ 
4: Run message passing methods to get solutions
    $F_1, F_2, \dots, F_K$ 
5: repeat
6:    $c \sim \{1, 2, \dots, N\}$ 
7:   if  $U \sim [0, 1] < Q_D$  then
8:      $p \sim \{1, 2, \dots, N\} - \{c\}$ 
9:      $A \leftarrow M(\mathbf{x}_p)$ 
10:  else
11:     $k \sim \{1, 2, \dots, K\}$ 
12:     $A \leftarrow F_k$ 
13:  end if
14:  (Snooker crossover)
15:   $\kappa \sim \{+1, -1\}$ 
16:   $s \sim S$ 
17:   $\mathbf{y}_c \leftarrow \mathbf{x}_c \cdot \kappa \exp(s) + A \cdot (1 - \kappa \exp(s))$ 
18:  Determine whether accept the new population or not
    by the Metropolis-Hastings rule.
19: until The algorithm converges.
```

based proposal rate Q_D which controls the relative weight of the dynamic and static anchor-based proposals. According to the value of U , we choose either the dynamic or static anchor-based proposals as the next proposal. MCMC-GD is described in Algorithm 3.

5. Experimental results

5.1. Analysis of synthetic MRF problems

To analyze the performance of the proposed algorithm thoroughly, we apply our algorithm on the synthesized MRF problems. Unlike real problems, we can easily control the hardness of the problems. We control the ratio of the non-submodular terms and the coupling strength of the graphs. For comparison, we also apply other methods including QPBO, TRW-S and BPM to the same problems.

For our experiments, we constructed 30 by 30 grid graphs with four-neighborhood structures. In graph construction, we followed the synthetic MRF construction in Komodakis's work [12]. We set the unary term of each node with a randomly generated number from Gaussian distribution $\mathcal{N}(0, 1)$. The pairwise terms were set as:

$$\theta_{st}(x_s, x_t) = \begin{cases} 0 & \text{if } x_s = x_t, \\ \lambda_{st} & \text{if } x_s \neq x_t, \end{cases} \quad (9)$$

where λ_{st} was drawn from $|\mathcal{N}(0, \sigma^2)|$ for submodular terms and from $-|\mathcal{N}(0, \sigma^2)|$ for non-submodular terms. We varied the percentage of non-submodular terms with parameter

ρ . We also varied the parameter σ to control the coupling strength.

We applied different algorithms to randomly generated MRF problems while changing parameters. ρ was set to 1%, 25%, and 50% and σ was set to 0.1, 2, 4, 6, and 8. For each parameter setting, experiments were repeated 20 times with different random number seeds and the average of the final energies was obtained. We also performed the proposed three algorithms, QPBO, TRW-S, and BP. For our algorithms, we used 100 chains and the temperature of i th chain was set to i . At each iteration, we selected single candidate. For the dynamic anchor-based proposal, we used a single iteration of QPBO algorithm. In the single iterations of QPBO, the proposal label was drawn from a uniform distribution and unlabeled pixels were fixed to current labels. For the static proposal, static anchors were obtained using TRW-S and BP. For the snooker crossover, κ was randomly chosen among +1 and -1 with equal probability. $\exp(s)$ was drawn from $\{0.1, 0.5, 2, 10\}$ with probability 0.5, 0.2, 0.2, and 0.1, respectively. Q_D was set to 0.9. In QPBO algorithm, unlabeled pixels were set to current labels.

The results of synthetic MRF problems are summarized in Fig 4. The x -axis are coupling strength and the y -axis are relative energy given by $100 \times (\text{energy of solution}) / (\text{minimum energy})$. We obtained better results by combination. MCMC-D which is combined with QPBO is better than QPBO alone in most cases. And, MCMC-S which is combined with TRW-S and BP is also better than the original TRW-S and BP. The enhancement in MCMC-S is, however, not as large as MCMC-D. Thus, under the condition that the performances of the deterministic methods are similar, MCMC-D is more effective than MCMC-S. Note that MCMC-GD which employs both the dynamic and static anchor-based proposals always obtain the lowest energy among all other methods. The running time of our algorithms are set to 8 seconds. QPBO is fastest among all the methods. It takes less than 0.1 second to converge. For TRW-S and BP, the maximum number of iterations is set to 2000. TRW-S takes 3~5 seconds to terminate and BP takes 7~9 seconds to terminate. All the experiments are performed on the Intel Quad Core 2.4GHz PC platform.

5.2. Results on a real problem: Photomontage

We also applied our algorithm to a practical vision problem known as photomontage [1, 18]. Among all the benchmark MRF problems in Szeliski *et al.*'s comparative study [18], the photomontage is considered as the most difficult problem due to the intrinsic property of the energy formulation. It is because the energy of the photomontage problem is dominated by the smoothness cost. As shown in the previous subsection, large coupling strength makes the problem more difficult. In addition, the function itself

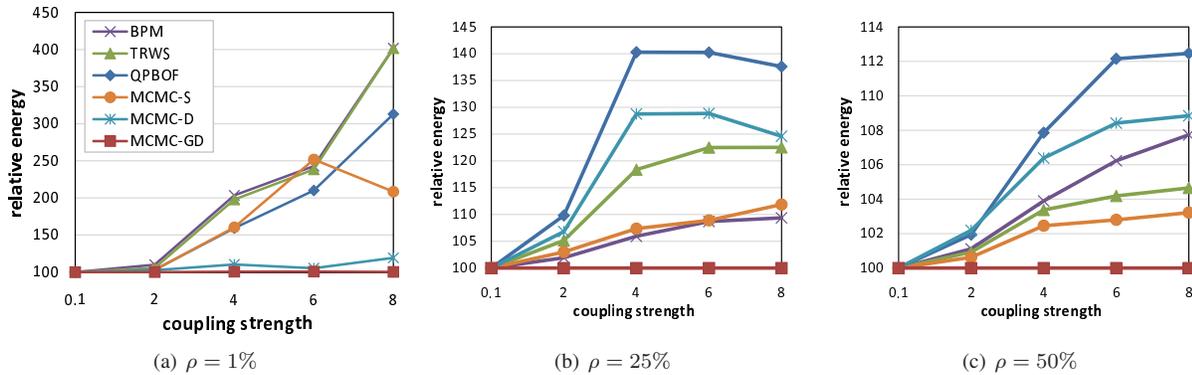


Figure 4. Experimental results on the synthetic MRF problems

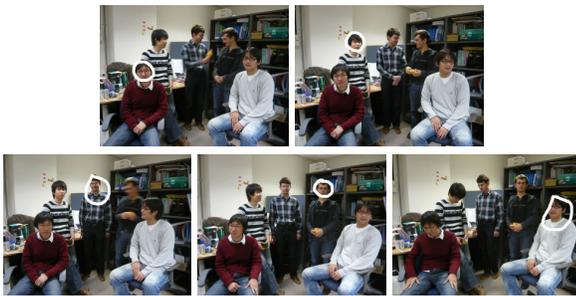


Figure 5. Input images of photomontage. The white circles in the input images represent the user strokes.

is non-submodular which consequently leads the truncation for α -Expansion method. We also empirically found that fewer user strokes and larger clutter in the image made the problem even harder. In this experiment, the energy model was set to be the same as that in Szeliski *et al.*'s paper [18] (second benchmark in photomontage). We used five input images so that the number of labels was also five. We applied our MCMC-GD algorithm as well as other methods.

Now, the settings for MCMC-GD algorithm are as follows. First of all, we used 100 chains and the temperature of i th chain was set to $i \times 100$. The optimal choice of temperatures depends on the scale of minimum energy of the target problem. At each iteration, we selected every sample except the parent as candidates. This is better than single candidate because single move of the move-making deterministic algorithm takes long time in this problem. For the dynamic anchor-based proposal, we used five iterations of α -Expansion algorithm. Note that the number of iterations was set to be the same as the number of labels. For the static proposal, static anchors were obtained using TRW-S. The parameter settings for the snooker crossover were the same as used for the synthetic MRF problems in the previous subsection.

The input images are shown in Fig 5. User strokes are

represented by the white circles. Quantitative results are provided in Fig 6. Upper row shows the resulting photomontage image of the each algorithm, and the lower row exhibits the corresponding color-coded image according to the labeling. Fig 7. presents the comparative energy plots of all the test algorithms against running time in seconds. Note that MCMC-GD algorithm still reached the lowest energy state among all other methods. The preprocessing time for obtaining the static anchor was not counted on the graph.

6. Conclusion

Although there have been great advances in solving simple MAP-MRF based vision problems, solution of more complex MRF models are still remain as challenging ones. Examples of the complex MRF models include non-submodular energy functions, highly connected MRFs and high-order clique potentials. Most existing optimization methods have inherent limitations in solving those difficult problems. In this paper, we propose new efficient algorithms called MCMC-D, MCMC-S, and MCMC-GD that can cope with those difficult MRF problems. They are sampling-based method (Pop-MCMC) combined with deterministic methods. By combination, the deterministic methods help the sampling-based method to rapidly move into lower energy state. Moreover, the deterministic methods make the sampling-based method jump easily from one basin to another over the energy barrier. Consequently, mixing rate is increased and we achieve faster convergence and better solutions. On the other hand, the sampling-based method helps deterministic methods not to be stuck in local minima. We experimentally showed that the proper combination of the two different approaches could substantially improve the overall performance. Our new energy minimization framework will be useful in solving many challenging vision problems. And consequently, this will encourage the design of better yet more complex energy models for practical vision applications.

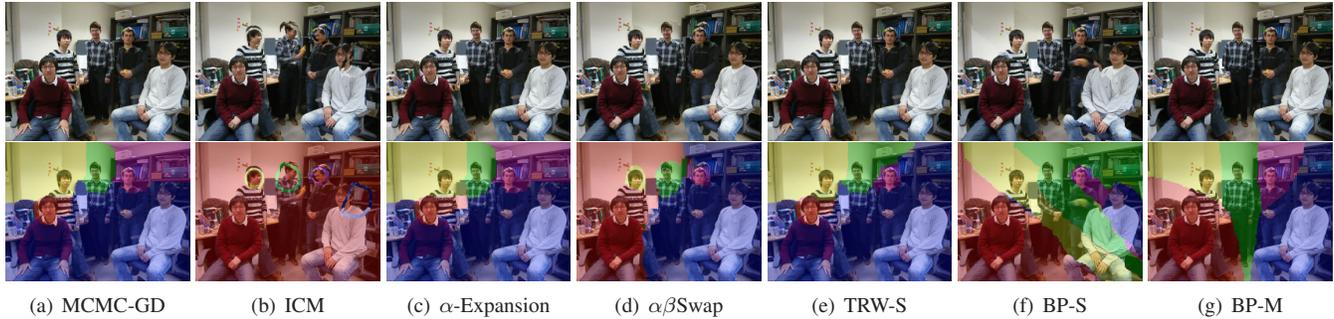


Figure 6. Photomontage results. Upper row shows the resulting photomontage images of the each algorithm. Lower row represents the corresponding color-coded image according to the labeling

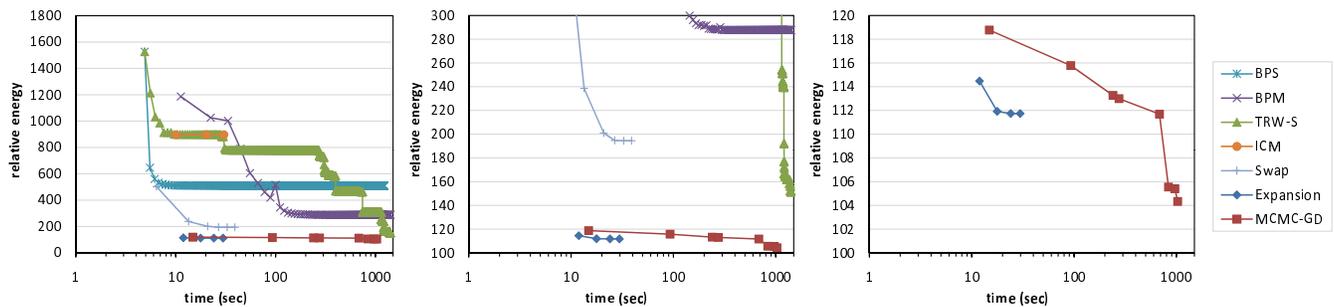


Figure 7. Experimental results on the photomontage problem

Acknowledgments

This research was supported in part by the Defense Acquisition Program Administration and Agency for Defense Development, through IIRC [UD070007AD], and in part by the IT R&D program of MKE/IITA [2008-F-030-01], Korea.

References

- [1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, and B. Curless. Interactive digital photomontage. In *ACM SIGGRAPH*, 2002.
- [2] A. Barbu and S. C. Zhu. Multigrid and multi-level Swendsen-Wang cuts for hierarchic graph partition. In *CVPR*, pages 731–738, 2004.
- [3] A. Barbu and S. C. Zhu. Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities. *PAMI*, 27(8):1239–1253, 2005.
- [4] E. Boros and P. L. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.
- [6] Y. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 26(9):1124–1137, 2004.
- [7] H. Y. Jung, K. M. Lee, and S. U. Lee. Toward global minimum through combined local minima. In *ECCV*, 2008.
- [8] H. Y. Jung, K. M. Lee, and S. U. Lee. Window annealing over square lattice Markov random field. In *ECCV*, 2008.
- [9] W. Kim, J. Park, and K. M. Lee. Stereo matching using population-based MCMC. *IJCV*, 30(6):1068–1080, 2008.
- [10] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.
- [11] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.
- [12] N. Komodakis and N. Paragios. Beyond loose LP-relaxations: Optimizing MRFs by repairing cycles. In *ECCV*, 2008.
- [13] X. Y. Lan, S. Roth, D. P. Huttenlocher, and M. J. Black. Efficient belief propagation with learned higher-order Markov random fields. In *ECCV*, 2006.
- [14] F. Liang and W. H. Wong. Real-parameter evolutionary monte carlo with applications to bayesian mixture models. *Journal of the American Statistical Association*, 96:653–666, 2001.
- [15] I. Nwogu and J. J. Corso. (BP)2: Beyond pairwise belief propagation labeling by approximating Kikuchi free energies. In *CVPR*, 2008.
- [16] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, 2007.
- [17] M. J. A. Strens, M. Bernhardt, and N. Everett. Markov chain Monte Carlo sampling using direct search optimization. In *ICML*, 2002.
- [18] 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. *PAMI*, 30(6):1068–1080, 2008.
- [19] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In *ICCV*, pages 900–907, 2003.
- [20] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. MAP estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches. *IEEE Trans. Information Theory*, 51(11):3697–3717, 2005.