Stereo Matching Using Population-Based MCMC

Joonyoung Park[1], Wonsik Kim[2], and Kyoung Mu Lee[2]

[1] DM research Lab., LG Electronics Inc., 16 Woomyeon-Dong, Seocho-Gu, 137-724, Seoul, Korea
kzoome@lge.com
[2] School of EECS, ASRI, Seoul National University, 151-742, Seoul, Korea
{ultra16, kyoungmu}@snu.ac.kr

Abstract. In this paper, we propose a new stereo matching method using the population-based Markov Chain Monte Carlo (Pop-MCMC). Pop-MCMC belongs to the sampling-based methods. Since previous MCMC methods produce only one sample at a time, only local moves are available. However, since Pop-MCMC uses multiple chains and produces multiple samples at a time, it enables global moves by exchanging information between samples, and in turn leads to faster mixing rate. In the view of optimization, it means that we can reach a state with the lower energy. The experimental results on real stereo images demonstrate that the performance of proposed algorithm is superior to those of previous algorithms.

1 Introduction

Stereo matching is one of the classical problems in computer vision [1]. The goal of stereo matching is to determine disparities, which are distances between two corresponding pixel. If we get an accurate disparity map, we can recover 3-D scene information. However, it remains challenging problem because of occluded regions, noise of camera sensor, textureless regions, etc. Stereo matching algorithms can be classified into two approaches. One is the local approach, and the other is the global approach. In the local approach, disparities are determined by comparing the intensity values in the local windows, such as SAD (Sum of Absolute Differences), SSD (Sum of Squared Differences), and Birchfield-Tomasi measure [2]. Although local approaches are fast, they have difficulties in obtaining an accurate disparity map. In the global approaches, one assumes that the disparity map is smooth in most regions. Usually, an energy function that is composed of local and global constraint is defined and solved by various energy minimization techniques. Typical global approaches include graph cut [3,4,5], belief propagation [6], and dynamic programming [7,8].

Monte Carlo method is one of the global approaches. It uses statistical sampling to obtain the solutions of some mathematical problems. Although this method was originally developed to generate samples from a given target distribution or to integrate functions in high dimensional space, it has also been applied to other types of problems such as optimization and learning problems.

Y. Yagi et al. (Eds.): ACCV 2007, Part II, LNCS 4844, pp. 560–569, 2007.

However, there are some difficulties in applying the Monte Carlo methods to vision problems as an optimizer. In general, we need to solve the vision problems in very high-dimensional solution space. Even if it is assumed to be 100 pixels in the width and the height respectively, the dimension of the image space becomes as high as 10^4. Monte Carlo methods would take infinitely long time since the acceptance rate would be almost zero in such a high-dimensional case. Moreover, we need to design a proper proposal distribution close to a target distribution. To resolve these problems, Markov Chain Monte Carlo (MCMC) methods have been tried. In MCMC, a new sample is drawn from the previous sample and the local transition probability, based on the markov chain. Contrary to simple Monte Carlo methods, the acceptance rates of MCMC methods are high enough, and the proposal distributions are designable even in the high-dimensional problem. Therefore, MCMC methods are more proper to be applied to vision problems than the Monte Carlo methods. However, difficulties still remain. Since most MCMC methods allows only local move in a large solution space, it still takes very long time to reach the global optimum.

To overcome the limitations of MCMC methods as an optimizer, Swendsen-Wang Cuts (SWC) was proposed [9,10]. In SWC, it is shown that bigger local move is possible than previous methods while maintaining the detailed balance. SWC uses Simulated Annealing (SA) [11] to find optima. Although SWC allows bigger local move, very slow annealing process is needed to approach the global optima with probability 1. This is a drawback for real applications. Therefore, usually a faster annealing process is applied for practical use in vision problems. However, the fast annealing does not guarantee the global optima but the sample is often trapped at local optima.

In this paper, we propose a new MCMC method called Population-Based MCMC (Pop-MCMC) [12,13] for stereo matching problem, trying to resolve the above problems of SWC. Our goal is to find more accurate global optima than SWC. In Pop-MCMC, two or more samples are drawn at the same time. And the information exchange is occurred between the samples. That makes it possible to perform the global move of samples. It means that the mixing rate of drawn samples becomes faster. And in the view of optimization, it means that it takes the shorter time for the samples to approach the global optima than previous methods.

This paper describes how Pop-MCMC is designed for stereo matching, and how the performance is comparing with the other methods like SWC or Belief Propagation. In section 2, we present how Pop-MCMC is applied to stereo matching. In section 3, we show the experimental results in the real problem. In the final section, we conclude this paper with discussions.

2 Proposed Algorithm

Segment-Based Stereo Energy Model. In order to improve the accuracy of the disparity map, various energy models have been newly proposed for stereo problem. Among them, we choose the segment-based energy model which is

one of the popular models [15,16,17,18]. In a segment-based energy model, the reference image is over-segmented. Mean-shift algorithm is often used for the segmentation[14]. And then, we assume that each segment can be approximated by a part of a plane in the real world. Each segment is defined as a node $v \in V$, and neighboring nodes s and t are connected with edges $\langle s,t\rangle \in E$. Then we construct a graph $\mathbf{G} = (V, E)$. And the energy function is defined by

$$E(\mathbf{X}) = \sum_{v \in V} C_{\mathrm{SEG}}(f_v) + \sum_{\langle s,t\rangle \in N} \beta_{s,t} \mathbf{1}(f_s \neq f_t), \tag{1}$$

where X represents the current state of every segment, f_v is an estimated plane for each segment, $C_{\mathrm{SEG}}(f_v)$ is a matching cost, and $\beta_{s,t}$ is a penalty for different neighboring nodes of s and t, which are defined by

$$C_{\mathrm{SEG}}(f_v) = \sum_{(x,y) \in V} C(x, y, f_v(x, y)), \tag{2}$$

$$\beta_{s,t} = \gamma \cdot (\text{mean color similarity}) \cdot (\text{border length}), \tag{3}$$

where function $C(x, y, f_v(x, y))$ is Birchfield-Tomasi cost. By varying γ, we can control the relative effect of matching cost and smoothness cost.

We firstly need to make a list of the planes for assigning them to each segment. For each segment, we calculate a new plane and add it to the list. The process to find a new plane is following.

We represent a plane with following equation.

$$d = c_1 x + c_2 y + c_3, \tag{4}$$

where x and y is the location of the pixel, d is the value of the disparity. From every pixel in a segment and initially assigned disparity values, we can construct the following algebraic equation.

$$\mathbf{A}\,[c_1, c_2, c_3]^{\mathrm{T}} = \mathbf{B}, \tag{5}$$

where ith row of the matrix $\mathbf{A}$ is $[x_i, y_i, 1]$ and ith row of the matrix $\mathbf{B}$ is d_i. The values of c_1, c_2, c_3 can be obtained by the least squares method. Once we find the values of the parameters, we can distinguish outlier pixels based on the values of the parameters. Then, the least squares method is repeated to exclude the outliers and improve the accuracy of c_1, c_2, c_3.

After obtaining the list of planes, we group the segments and calculate planes again in order to improve the accuracy of planes. To this end, each segment is firstly assigned to a plane in the list that has lowest C_{SEG} value. Then we group the segments which is assigned to the same plane. And for each group, the above plane fitting is repeated again. At last, we have the final list of the planes to use.

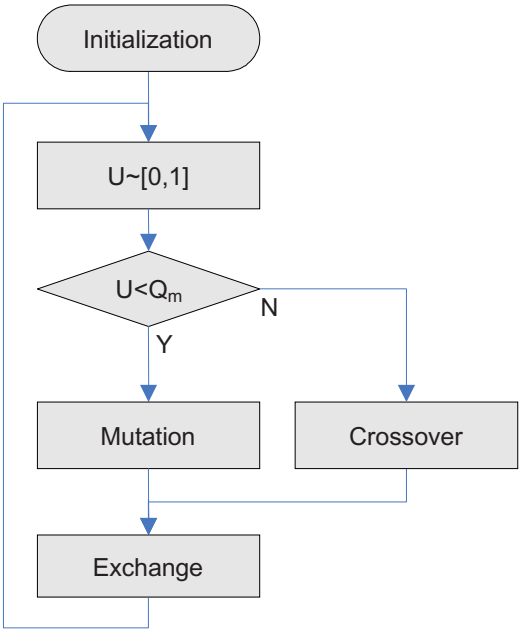

Fig. 1. The overall flow chart of Pop-MCMC applied to stereo matching

Design of Pop-MCMC. Given the probability distribution $\pi(\mathbf{X}) \propto \exp\{-E(\mathbf{X})\}$, our aim is to find the state $\mathbf{X}$ where the probability is maximized. In Pop-MCMC, we draw multiple samples from multiple chains at the same time with respect to the following distributions.

$$\pi_i(\mathbf{X}) = \pi(\mathbf{X})^{\frac{1}{T_i}} \propto \exp\left\{-\frac{E(\mathbf{X}_i)}{T_i}\right\}, \tag{6}$$

where T_i is the temperature of ith chain. Each sample from each chain is called a chromosome. Chromosomes interact with each other and this helps performing global moves.

The overall flow of Pop-MCMC is illustrated in figure 1. The three moves in Pop MCMC, which are mutation move, crossover move, and exchange move, are repeatedly performed and the samples are generated at each iteration. In this process, we firstly choose a random number U between 0 and 1, and compare U with the value of Q_m. Depending on value of U, we choose one move among mutation and crossover. By varying the parameter value Q_m, we can control the rates between the global move (crossover) and the local move (mutation) easily. A proper value of Q_m can be adjusted according to the given problem, the model, or the number of chains. For example, if large number of chains are used, more global move will be needed than local move since the global move enables the samples exchange their information each other. Next, we will describe the detailed design of each move for stereo problem.

1. Mutation move

In the case of mutation move, we apply the original MCMC move. we borrow the MCMC kernel from SWC [9,10] and a few modifications are made to the kernel. Basically, the MCMC kernel is used for a randomly chosen chain. In SWC, the graph nodes are probabilistically clustered as the first step. For each edge $e = \langle s, t \rangle \in E$, if labels of node s and node t are different, we delete that edge. Otherwise, we determine whether delete it or not according to the edge probability q_e. Then we randomly choose a cluster and propose a new label according to the predefined proposal distribution. A label here means estimated plane. Instead of accepting every proposal, we determine the acceptance according to the acceptance ratio. And, the acceptance ratio is determined by Metropolis-Hastings rule [9]. Edge probability and proposal distribution of new labels are designed by following equations, respectively.

$$q_e = 1 - \exp\left(-\frac{\text{(mean color similarity)}}{\frac{C_{\text{SEG}}(f_{v_1})}{N(v_1)} + \frac{C_{\text{SEG}}(f_{v_2})}{N(v_2)} + 2} \right), \tag{7}$$

$$q(l' \mid V_0, A) = \exp\left[-\left\{ \frac{\sum_{v \in V_0} C_{\text{SEG}}(f_v = l')}{\sum_{v \in V_0} N(v)} + 1 - \prod_{\langle v_1, v_2 \rangle \in N, v_1 \in V_0, v_2 \notin V_0} \mathbf{1}(l' = f_{v_2}) \right\} \right], \tag{8}$$

where v_1 and v_2 represent neighboring nodes, $N(v)$ is the number of the pixels in the node v, l' is the newly proposed label, and A is the current sample. In equation (7), the more similar the intensities of the connected nodes and the lower the matching costs are, the higher the probability that the edge remains is. And the size of the segments is not considered because of normalizing terms. In equation (8), when the nodes in the cluster V_0 have low matching costs and there exist the neighboring nodes which have the same label, the value of $q(l' \mid V_0, A)$ becomes high.

2. Exchange move

Exchange move is originated from parallel tempering [19,20]. In this move, we choose two chains and determine whether we have to exchange the chromosomes of two chains or not by the Metropolis-Hastings rule. Note that for the exchange move, there is no need of special design for stereo matching. So, when the ith and jth chains are selected, we simply calculate the Metropolis-Hastings rate by

$$\gamma_e = \exp\left[\left\{ E(\mathbf{X}_i) - E(\mathbf{X}_j) \right\} \left(\frac{1}{T_i} - \frac{1}{T_j} \right) \right], \tag{9}$$

where X_i and T_i are the current state and temperature of ith chain.

3. Crossover move

In crossover move, two chains are firstly selected as in exchange move. And then, instead of exchanging the whole chromosomes of two chains, we exchange only selected parts of chromosomes. After that, it is decided with the acceptance ratio whether the new sample is accepted or not. The typical types of the crossover moves are 1-point crossover and 2-point crossover moves. However, these are the methods based on the fact that the chromosomes are 1-dimensional vectors, so it is improper to apply them to the stereo matching problem in which the chromosomes are 2-dimensional images. However, 1-point and 2-point crossover moves have an advantage of low computational complexity because the proposal distributions are canceled out each other. Therefore, we designed the crossover move to maintain this advantage and to be proper to chromosomes whose dimensions are two. Detailed algorithms are as follows.

We firstly choose two chains randomly, and we construct the cluster V_0 in a similar way as in SWC. However, there are two differences in constructing V_0 compared with SWC. First, q_e is constant, not adaptively determined with the matching costs or the intensities of the input image since there is no need for the nodes of the cluster V_0 to have same labels here. It is computationally efficient to use q_e as a constant value because the proposal distribution part in the Metropolis-Hastings rate is canceled. Second, when we calculate the probability q_e, we do not consider whether the labels of the nodes are the same or not. So resulting cluster can have nodes with different labels. Note that, in the mutation move, we should remove the edges connecting the nodes that have different labels, in order to compute the Metropolis-Hastings rate efficiently. But in the case of crossover move, the proposal distribution part is completely eliminated by the Metropolis-Hastings rate and eventually it enables high efficiency in computation. And this free construction of V_0 also helps faster convergence.

The process after constructing the cluster V_0 is similar to the 1-point crossover move. From the chromosomes of two selected chains, new chromosomes are proposed by exchanging the labels of the nodes which belong to the cluster V_0. Then the acceptance ratio $\alpha = \min(1, \gamma_c)$ of the newly proposed chromosomes are calculated, and the next sample is determined. By substituting equation (6) to Metropolis-Hastings rule, we can obtain γ_c as follow.

$$\gamma_c = \exp\left[\frac{E(\mathbf{X}_i) - E(\mathbf{Y}_i)}{T_i} + \frac{E(\mathbf{X}_j) - E(\mathbf{Y}_j)}{T_j}\right]. \tag{10}$$

3 Experimental Results

In this section, we evaluate the performance of the proposed algorithm by comparing with other methods such as SWC and BP (Belief Propagation) [6]. For the experiment, the test images from Middlebury website are used [21].

We employed the segment-based energy model in 1 for this test. The resulting disparity maps are shown in figure 2. And table 1 shows bad pixel rates of test

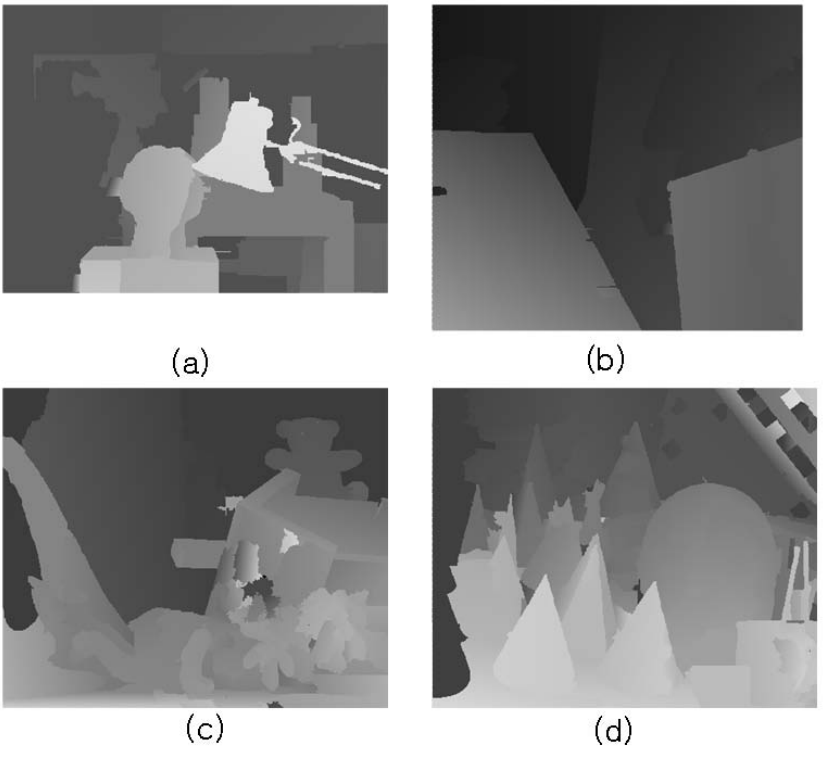

Fig. 2. Results of the proposed algorithm: the disparity maps of (a) Tsukuba (b) Venus (c) Teddy (d) Cones

images. Note that there are some limitations in the segment-based energy model. When real world objects are piecewise planar, the results might quite good. However, for the cases of Teddy and Cones that include objects with curved surfaces, the performances seem bad. And also, for a fronto-parallel plane, a non-segment based energy model can be superior to the segment-based energy model because of smaller number of the labels. In addition, occlusion or visibility are not considered here, so it may results in bad pixels in the disparity map.

We compared the performance of Pop-MCMC with those of SWC and BP in the view of energy minimization. The same energy model was applied to each method. In case of SWC, we followed the Barbu's work [9,10]. In Figure 3, the

Table 1. The bad pixel rate for each test image

Test images	Tsukuba	Venus	Teddy	Cones
Bad Pixels(%)	1.38	1.21	14.7	13.1

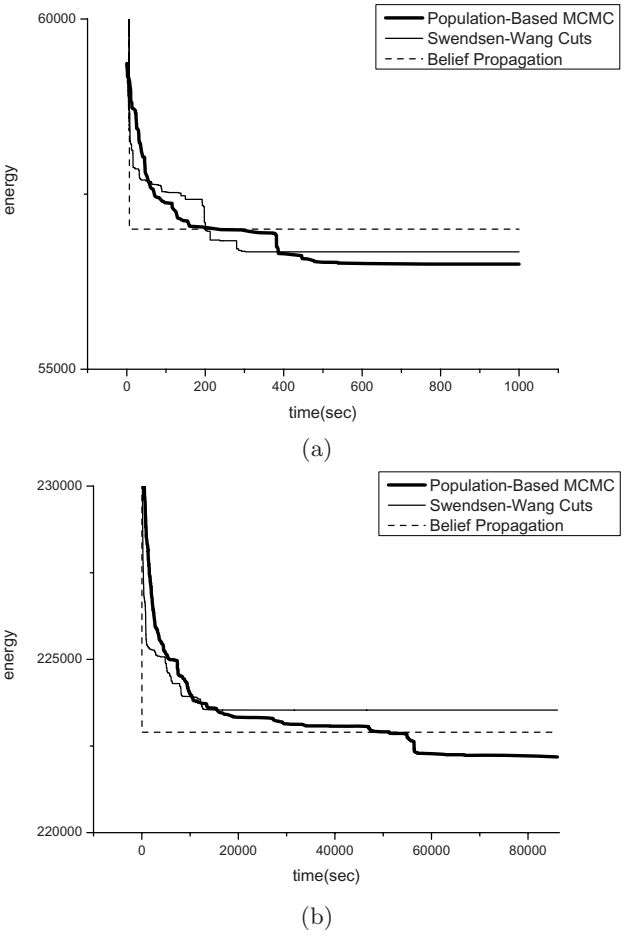

Fig. 3. Convergence comparison of energy minimization methods for (a) Tsukuba (b) Cones

convergence graphs of each method are presented. In the early part of each graph, SWC is faster than Pop-MCMC. That is caused by the characteristics of the energy function. In the energy function used in stereo matching, the local minima are concentrated near the global minimum. In the view of large scale, it can be considered as the energy function which has only one minimum (more exactly, one bunch of minima). So, SWC gets close to the vicinity of global minimum faster than Pop-MCMC. However, once the samples approach near the global minimum, there exist many local minima, which are no longer concentrated at one part. At this stage, SWC is easy to be trapped at the local minimum, while Pop-MCMC is likely to approach the global minimum nearer. As illustrated in Figure 3, we can get to the state of the lower energy with Pop-MCMC method than the other methods.

4 Conclusion

In this paper, we present a stereo matching algorithm using Pop-MCMC. Pop-MCMC uses multiple chains, and establish faster mixing rate by exchanging information between chromosomes. As a consequence, it is shown that the proposed Pop-MCMC method reaches the global optimum faster than other energy minimization method including SWC and BP. We have a plan to apply and analyze the performance of the proposed method to more sophisticated stereo energy models including occlusion handling and visibility terms as well as the segmentation problem.

Acknowledgments. This work was supported in part by the ITRC program by Ministry of Information and Communication and in part by Defense Acquisition Program Administration and Agency for Defense Development, through the Image Information Research Center, Korea.

References

1. Scharstein, D., Szeliski, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. Microsoft Research Technical Report MSR-TR-2001-81 (2001)
2. Birchfield, S., Tomasi, C.: A Pixel Dissimilarity Measure That is Insensitive to Image Sampling. IEEE Trans. Pattern Analysis and Machine Intelligence (1998)
3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Analysis and Machine Intelligence 23(11), 1222–1239 (2001)
4. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: Proc. Int'l. Conf. Computer Vision, pp. 508–515 (2001)
5. Kolmogorov, V., Zabih, R.: Multi-camera scene reconstruction via graph cuts. In: Proc. European Conf. Computer Vision, pp. 82–96 (2002)
6. Sun, J., Zheng, N.N., Shum, H.Y.: Stereo matching using belief propagation. IEEE Trans. Pattern Analysis and Machine Intelligence 25(7), 787–800 (2003)
7. Ohta, Y., Kanade, T.: Stereo by intra- and inter-scanline search. TPAMI 2, 449–470 (1985)
8. Veksler, O.: Stereo Correspondence by Dynamic Programming on a Tree. Computer Vision and Pattern Recognition (2005)
9. Barbu, A., Zhu, S.C.: Graph Partition By Swendsen-Wang cuts: International Conf. Computer Vision, pp. 320–327 (2003)
10. Barbu, A., Zhu, S.C.: Multigrid and multi-level Swendsen-Wang cuts for hierarchic graph partition. Computer Vision and Pattern Recognition, 731–738 (2004)
11. Kirkpatrick, S., Gelatt, C., Vecchi, M.: Optimization by simulated annealing. Science 220(4598), 671–680 (1983)
12. Liang, F., Wong, W.H.: Evolutionary Monte Carlo: Applications to Model Sampling and Change Point Problem. Statistica Sinica 10, 317–342 (2000)
13. Jasra, A., Stephens, D.A.: On Population-Based Simulation for Static Inference (2005)
14. Comanicu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Machine Intell. 24, 603–619 (2002)

15. Tao, H., Sawhney, H.S., Kumar, R.: A Global Matching Framework for Stereo Computation. In: Proc. Int'l. Conf. Computer Vision, vol. 1, pp. 532–539 (2001)
16. Bleyer, M., Gelautz, M.: Graph-based surface reconstruction from stereo pairs using image segmentation. In: Proc. SPIE, Videometrics VIII, vol. 5665, pp. 288–299 (2005)
17. Hong, L., Chen, G.: Segment-Based Stereo Matching Using Graph Cuts. Computer Vision and Pattern Recognition, I, 74–81 (2004)
18. Klaus, A., Sormann, M., Karner, K.: Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. In: International Conf. Pattern Recognition, pp. 15–18 (2006)
19. Geyer, C.J.: Markov chain Monte Carlo maximum likelihood, Computing Science and Statistics, 156–163 (1991)
20. Hukushima, K., Nemoto, K.: Exchange Monte Carlo method and application to spin glass simulations. J. Phys. Soc. Jpn. 65, 1604–1608
21. `http://cat.middlebury.edu/stereo`