Constrained Smoothness Cost in Markov Random Field Based Stereo Matching

Ba Thai[1,*], Mukhalad Al-nasrawi[2], Guang Deng[3], Robert Ross[4] and Phat Huynh[5]
[1,2,3,4,5]Department of Engineering, La Trobe University, Bundoora, Victoria 3086, Australia
[2]Foundation of Technical Education, AL-Musaib Technical College, Babil 51006, Iraq
[*]t.thai@latrobe.edu.au

***Abstract*—Stereo matching is the process of finding disparity map between a pair of stereo images. Markov Random Field (MRF) model has been used in stereo matching algorithms in which the goal is to minimize total data, smoothness and occlusion cost. Recent MRF based stereo matching algorithms assign the same smoothness cost to adjacent pixels which is not effective at objects boundaries. In this paper, we propose a method to constrain the smoothness cost such that the cost applied to disparity discontinuity within an object is higher than that of the object's boundary. Data cost function is also normalized before computing its probability distribution function (PDF) to enforce an uniform PDF for the entire image. Experimental results have demonstrated that the proposed method is competitive to the state-of-the-art stereo matching algorithms.**

I. Introduction

Stereo vision algorithms are usually based on the assumption of intensity consistent from different viewpoints and within short time interval. The goal of stereo vision is to estimate a dense disparity map from two images taken at different viewpoints. The disparity map is then used as the input for a wide range of applications, such as 3D scene reconstruction [1], collision avoidance and navigation in robotics field [2]. Disparity map estimation remains a challenging task, mainly due to variations from stereo vision system set up and unavoidable noise from the environment. Scharstein and Szeliski have provided a comprehensive overview on stereo correspondence [3].

Stereo vision algorithms can be classified into local and global approaches. The local algorithms are usually based on the block-matching or window-based techniuqes. They form a small block centering at each pixel location $p = (x, y)$ in the left image and find the corresponding disparity d such that the cost function regarding to a block centering at $q = (x - d, y)$ in the right image is minimized. The cost functions including sum of absolute difference (SAD), sum of squared different (SSD), normalize cross correlation (NCC) and sum of hamming distance (SHD) are discussed in [4], [5], [6], [7].

Global algorithms aim to optimize a global objective function. All disparities are estimated simultaneously by minimizing costs of all pixels. Let E_d, E_s and E_o respectively denote the cost of pixel's observed intensity (data), disparity discontinuity (smoothness), and occlusion map. Further let $\mathcal{I}$ is the set of pixels in left image, L_p and L_q are the disparities labelling from 0 to maximum allowable value of the disparity (d_{max}) at pixel location p and q respectively, $\mathcal{E}$ is the set of all adjacent pixels, and $\mathcal{O}$ is the set of occluded pixels. The global approaches aim to minimize the total cost which is defined as

$$E = \sum_{p \in \mathcal{I}} E_d(L_p) + \sum_{p \in \mathcal{I} \cap \mathcal{O}} E_o + \sum_{p,q \in \mathcal{E}} E_s(L_p, L_q) \quad (1)$$

The data term can be chosen from any of matching cost functions as used in local approaches, and the occlusion term is a constant added to each occluded pixel. The smoothness term plays a crucial role in determining the performance of MRF based stereo vision algorithms. By incorporating mutual information of all pixel, recent global stereo matching methods using MRF model have outperformed local approaches in terms of increasing number of good matches and eliminating collusion [8], [9], [10].

In this work, we aim to improve the MRF based global stereo matching by adaptively constraining the smoothness cost such that a pixel residing in an object's boundary has higher cost than that of a pixel in a homogeneous region. This improves the accuracy of the disparity map because two pixels belonging to two different objects should have different disparity levels.

The key contributions and organization of this work are summarized as follows. In section 2, we briefly review related work and set up a MRF model for our method. In section 3, we present the proposed method. The technique of constraining the smoothness cost is discussed in this section. In section 4, we present experimental results in comparison to the ground truth and state-of-the-art methods. In the final section, we provide concluding remarks.

II. Related Work

The MRF are undirected graphical models which do not require information of edge orientations that make them natural for images in which pixels are positioned symmetrically, and the Markov blanket of each node (pixel) is its nearest neighbors [11]. Due to these properties, the MRF has been intensively studied in the context of stereo matching.

A typical 2-lattice MRF used in global stereo matching is shown in Fig.1. This is a piece-wise network with two random

978-1-5090-2896-2/16/$31.00 ©2016 IEEE

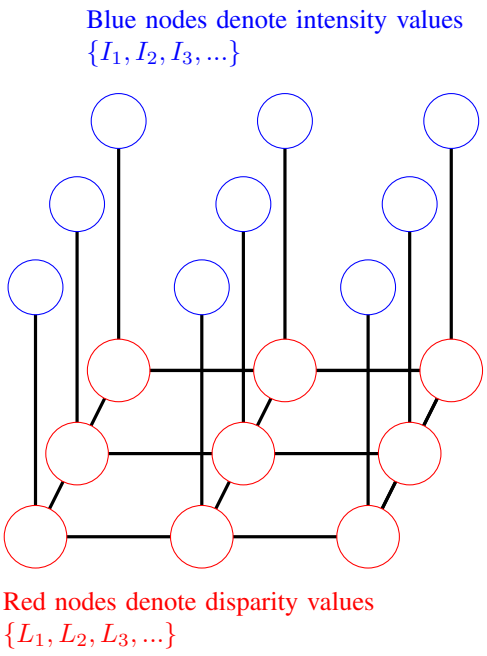

Fig. 1: A MRF model for a 3x3 image patch

fields including the intensity values denoted by $\{I_1, I_2, I_3, ...\}$ as the measured layer and the disparity label denoted by $\{L_1, L_2, L_3, ...\}$ as the hidden layer. Let S denote the smoothness variable. The occlusion term is omitted for the simplicity. By applying Bayes rule, the joint probability of the MRF can be defined as

$$P(S, L|I) = \frac{P(I|S, L)P(S, L)}{P(I)} \tag{2}$$

As illustrated in Fig.1, only the hidden lattice has smoothness term, whereas each pixel intensity is measured independently of its neighborhood. Hence, the likelihood term $P(I|S, L)$ can be simplified into $P(I|L)$. Also $P(I)$ is constant. Consequently, the joint posterior of the MRF becomes

$$P(S, L|I) \propto P(I|L)P(S, L) \tag{3}$$

The disparity levels assigned to all pixels are determined by minimizing the total cost in Equation (1) or by maximizing the posterior in Equation (3). Each pixel at location p in the left image can be assigned to different disparity levels $d = 0, 1, ..., d_{max}$, at each assignment there is a cost associated with the matching pixel at location $q = p+d$ in the right image which is determined by the data cost function. Also there is a cost applied to adjacent pixel I_q when it is assigned to each disparity level that is different to that of pixel at p. This is due to the assumption that adjacent pixels should have similar disparity level. This cost is determined by the smoothness cost function. Let $\phi(I_p, L_p)$ denote the data cost function of pixel p with disparity L_p given intensity observation I_p and further let and $\psi(L_p, L_q)$ denote the smoothness cost function of pixel q given disparity level of pixel p. As discussed in [12], the joint posterior can be defined as

$$P(I_1, I_2, ..., I_N, L_1, L_2, ..., L_N) \propto \prod_{p \in N} \exp(-\phi(I_p, L_p) \prod_{p,q \in N} \exp(-\psi(L_p, L_q)) \tag{4}$$

where N is the number of pixels.

III. Proposed method

The likelihood and prior are often defined with an assumption that observation noises follow an independent identical distribution [12]. This assumption implies that any adjacent pixels should have similar disparity level. However this may not hold true if two adjacent pixels reside in two different objects with different depths. In order to tackle this, we introduce a constrained model which is defined as

$$P(I_1, I_2, ..., I_N, L_1, L_2, ..., L_N) \propto \prod_{p \in N} \lambda_D \exp(-\lambda_D \phi(I_p, L_p) \prod_{p,q \in N} \lambda_S \exp(-\lambda_S \psi(L_p, L_q)) \tag{5}$$

where λ_D and λ_S are constrained parameter for the data and smoothness cost functions respectively. The likelihood can be assumed to be independent of disparity continuity as the intensity is pixel-based [12], hence λ_D can be defined as a constant, normally $\lambda_D = 1$. On the other hand, λ_S should be different between textureless area and object's boundaries. We propose a method of determining λ_S based on the standard deviation of data which is detailed in later discussion. Because the cost functions return high values in object boundaries and low values in textureless areas, we propose to normalize cost functions before computing probability to achieve uniform distribution for all pixels. Let ϕ_n and ψ_n be the normalized data and smoothness cost function relatively. They are defined as

$$\begin{aligned} \phi_n(I_p, L_p) &= \frac{\phi(I_p, L_p)}{\sum_{L=0}^{d_{max}} \phi(I_p, L_p)} \\ \psi_n(L_p, L_q) &= \frac{\psi(L_p, L_q)}{\sum_{L=0}^{d_{max}} \psi(L_p, L_q)} \end{aligned} \tag{6}$$

In the following, we first derive the formulation for λ_S, we then discuss the method to approximate the maximum a posterior (MAP) defined in Equation (5).

A. Smoothness constrained parameter

The smoothness term, or sometimes referred to as the disparity continuity, enforces adjacent nodes in hidden lattice of the MRF network to have the same disparity level. A cost is applied if an adjacent node takes a different disparity level. A heavier cost is enforced to a larger different level. The cost is determined by the smoothness cost function which is usually one of the followings: potts, truncated linear, and truncated quadratic model [13]. In this work we do not aim to explore the cost function but its constrained parameter. Because it does not necessary for adjacent pixels at object boundaries to have same disparity level, in this work we apply a smaller cost for

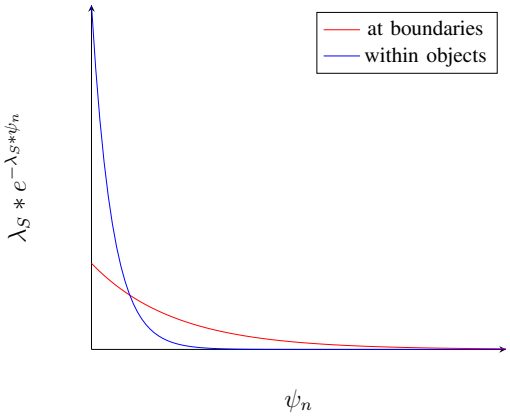

Fig. 2: Effect of λ_S to the smoothness cost

pixels at object boundaries than that of pixels at flat regions. The effect of constrained parameter on the smoothness cost function is illustrated in Fig. 2.

Since the algorithm requires smoothness cost to be computed, we propose a method to derive λ_S from ψ to reduce computational complexity. Fig. 3 illustrates an example where the blue pixel resides within an object (assume the object is near textureless) and the red pixel lies in the boundary of the object. For each pixel I_p in left image, we need to compute cost for I_{p+d} $(d \in 0, ..., d_{max})$ matching pixels in the right image. The score of data cost for the blue pixel is very small compared to that of the red pixel. We can compare the mean of data cost over all disparity levels to determine if the pixel lies on boundary within an object. However, this only holds true in ideal scenario when the objects are assumed to be featureless. In order to differentiate object boundaries and textures, we exploit the local standard deviation to estimate how spread out the data score is. In this method, we assume that the standard deviation at boundaries is greater than that in featured area and it has minimum value at featureless area. In addition, as discussed above, the bigger λ_S indicates the strong disparity correlated between adjacent pixels. This implies $\lambda_S \propto \sigma$. As the result, we can define λ_S as

$$\lambda_S = \exp(-\sigma) \tag{7}$$

where

$$\begin{aligned} \sigma &= \sqrt{\frac{\sum_{L_p=0}^{d_{max}} (\psi(L_p, L_q) - \mu)^2}{d_{max}}} \\ \mu &= \frac{\sum_{L_p=0}^{d_{max}} \psi(L_p, L_q)}{d_{max}} \end{aligned} \tag{8}$$

B. MAP approximation using belief propagation method

The disparity map is defined by estimating a set of label $L_1, L_2, ..., L_N$ in Equation 4 such that the joint probability is maximized. This method is referred to as Maximum A Posterior (MAP) estimation. In order to approximate MAP,

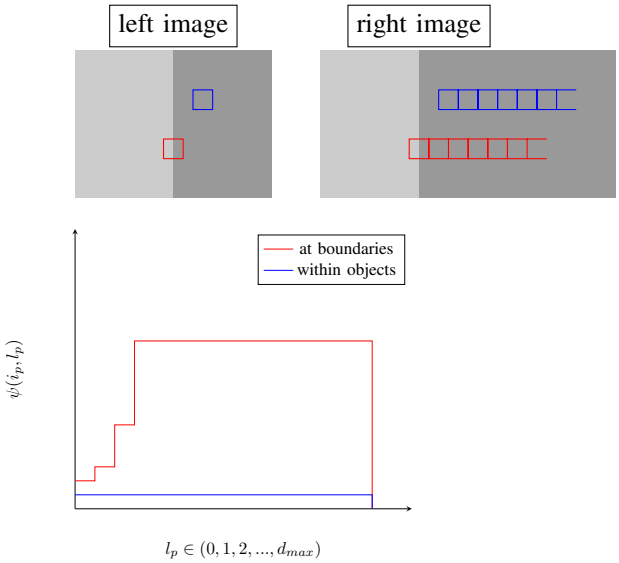

Fig. 3: Simulation of smoothness cost for pixels at boundary and within object

we use Loopy Belief Propagation algorithm pioneered by Pearl [16] in which the existence of loops in the network is considered. Each node sends its belief about the disparity level probability of a pixel in the neighbourhood to that pixel. Let $m_{p,q}(d)$ denotes the message that node at location p sends to node at location q about disparity level d. Each node has its initial belief about its disparity probability, this initial belief will be updated when it receives messages from its neighbourhood pixels. Three main algorithms for message update are "Sum-Product", "Max-Product", and "Min-Sum". We use max-product because it approximates max joint posterior of the network.

The proposed algorithm includes the following steps

1) Initialize all messages as uniform distributions
2) Normalize data energy

$$\phi_n(I_p, L_p) = \frac{\phi(I_p, L_p)}{\sum_{L_p=0}^{d_{max}} \phi(I_p, L_p)} \tag{9}$$

3) Set $\lambda_D = 1$ and compute λ_S as in Equation (7) and (8).
4) Update messages iteratively. A message sent from pixel p to pixel q given disparity level d is computed as

$$\begin{aligned} m_{p,q}(d) = \max_{0 \le L_q \le d_{max}} [&\lambda_D \exp(-\lambda_D \phi_n(I_p, L_q)) \\ &\lambda_S \exp(-\lambda_S \psi_n(L_p, L_q)) \\ &\prod_{k \in \Omega_p, k \neq p} m_{k,p}(L_p)] \end{aligned} \tag{10}$$

where Ω_p denote the neighboorhood of pixel p.
5) For a pixel at location p, the belief corresponding to disparity level $d \in \{0, 1, 2, ..., d_{max}\}$ is denoted by $b_p(d)$ and it is computed as

$$b_p(d) = \lambda_D \exp(-\lambda_D \phi_n(I_p, d)) \prod_{k \in \Omega_p, k \neq p} m_{k,p}(L_p) \tag{11}$$

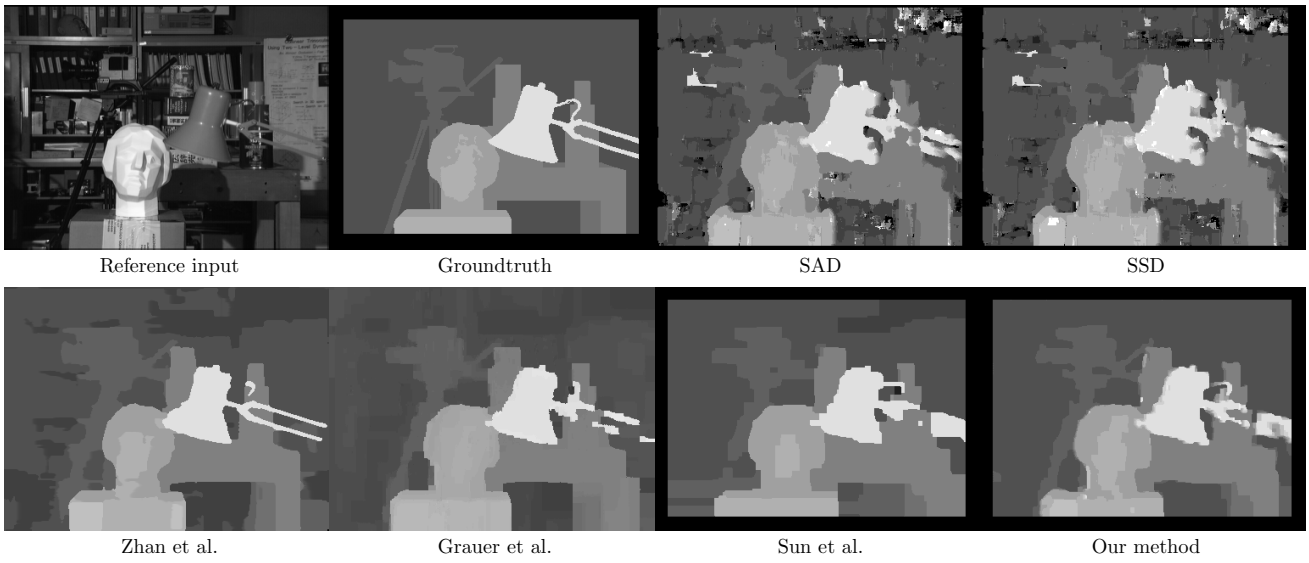

Fig. 4: The experiment results. The images of Zhan et al. [14], Grauer et al. [15] and Sun et al. [12] were taken from the Middlebury Stereo Evaluation database (available at *http://vision.middlebury.edu/stereo*)

6) For a pixel at p, find the maximum belief among $\{b_p(0), b_p(1), ..., b_p(d_{max})\}$. The disparity level corresponding to the maximum belief is the final result for that pixel.

IV. Experimental results

The proposed algorithm has been tested using the Tsukuba image pairs. Fig. 4 demonstrates the result of our method in comparison with the state-of-the-art methods which are evaluated by Middlebury Stereo Evaluation (the well-known benchmark in stereo matching algorithms). The proposed algorithm was implemented in C++ using a 3.6GHZ computer.

We first experimented our algorithm in comparison with SAD and SSD which are the typical local algorithms. As shown in Fig. 4, the proposed method outperforms the tested local methods in terms of disparity accuracy. The proposed method was also compared with Zhan et al.'s method which was ranked number one by the Middlebury Stereo Evaluation at the time we conducted our experiment. Zhan et al.'s method has high accuracy for foreground objects such that the lamp or the table. In contrary, our method produced less errors in the background. Finally, the proposed method is comparable with other MRF based methods including Grauer et al. and Sun et al. methods.

V. Conclusion

In this paper, we have proposed a method to improve accuracy of MRF based stereo matching algorithm. The idea is to control the smoothness cost function by introducing a constrained parameter such that the probability distribution function corresponding to smoothness decreases faster at object's boundary compared to that in flat regions. We derived the smoothness constrained parameter by using the standard derivation of the smoothness cost function corresponding to one pixel in the left image and other pixels in the right image which has disparity difference from 0 to d_{max}. Experimental results have shown that our method produces accurate disparity maps which are comparable to those produced by the state-of-the-art methods.

References

[1] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Proc. Intelligent Vehicles Symposium (IV)*, 2011, pp. 963–968.

[2] D. Murray and J. J. Little, "Using real-time stereo vision for mobile robot navigation," *Autonomous Robots*, vol. 8, no. 2, pp. 161–171, 2000.

[3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1-3, pp. 7–42, 2002.

[4] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. CVPR*, 2007, pp. 1–8.

[5] N. Lazaros, G. C. Sirakoulis, and A. Gasteratos, "Review of stereo vision algorithms: from software to hardware," *Int. J. Optomechatronics*, vol. 2, no. 4, pp. 435–462, 2008.

[6] R. Mayoral, G. Lera, and M. J. Pérez-Ilzarbe, "Evaluation of correspondence errors for stereo," *Image and Vision Computing*, vol. 24, no. 12, pp. 1288–1300, 2006.

[7] P. Huynh, R. Ross, J. Devlin, and B. Thai, "Constrained sliding window correspondence algorithm for fast stereo vision," in *Proc. Industrial Electronics and Applications (ICIEA)*, 2015, pp. 1943–1948.

[8] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proc. ICPR*, vol. 3, 2006, pp. 15–18.

[9] Y. Xie, N. Liu, S. Liu, and N. Barnes, "Stereo matching using subsegmentation and robust higher-order graph cut," in *Proc. Digital Image Computing Techniques and Applications (DICTA)*, 2011, pp. 518–523.

[10] J. Liu, P. Delmas, G. Gimel'farb, and J. Morris, "Stereo reconstruction using an image noise model," in *Proc. Digital Image Computing Techniques and Applications (DICTA)*, 2005, pp. 69–69.

[11] S. Marsland, *Machine learning: an algorithmic perspective*. CRC press, 2015.

[12] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 787–800, 2003.

[13] D. Scharstein and R. Szeliski, "Stereo matching with nonlinear diffusion," *Int. J. Comput. Vision*, vol. 28, no. 2, pp. 155–174, 1998.
[14] Y. Zhan, Y. Gu, K. Huang, C. Zhang, and K. Hu, "Accurate image-guided stereo matching with efficient matching cost and disparity refinement," *IEEE Transactions on Circuits and Systems for Video Technology*, no. 99, 2015.
[15] S. Grauer-Gray and C. Kambhamettu, "Hierarchical belief propagation to reduce search space using cuda for stereo and motion estimation," in *Proc. Workshop on Applications of Computer Vision (WACV)*, 2009, pp. 1–8.
[16] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.