DOUBLY-MRF STEREO MATCHING

Li Cheng and Terry Caelli

University of Alberta
Department of Computing Science
Edmonton, A.B. Canada
T6G 2E8

ABSTRACT

We examine a new double-layered Markov Random Field probabilistic framework for stereo matching and use Belief Propagation for approximate inference. Our initial experimental results are promising and future developments are discussed.

1. INTRODUCTION

To address the problem of stereo matching [1], and inspired by the region based methods [2], we propose a probabilistic method that assumes two layers of MRFs (Markov Random Fields): the pixel layer and the region layer. In the pixel MRF layer, the disparity matching score of a given pixel is defined via the MAP (Maximum of Posteriori) probability of pixel disparity. In the region MRF layer, the hypothetical region's disparity plane parameter vector (a, b, c) is determined by the MAP probability of the parameter vector.
Further, instead of the non-causality of the traditional MRF, we augment each MRF layer with some causal connections to model stereo naturally. Exact solution for such complicated loopy Markov network models is far from reachable. For both MRF layers, a BP (Belief Propagation) algorithm [3] is used for approximate inference via MAP. The BP algorithm has been theoretically [3] proved to converge to the true means for such pairwise networks as MRF, with successful results (e.g. [4]).

2. THE PROPOSED ALGORITHM

More formally, we model the stereo problem via a double-layer MRF, under the following constraints.

- *Segmentation constraint*: Given over-segmentation [5] of the reference view image (the left image), pixel disparities are continuous (smooth) within each region and discontinues only exist between regions.

This project was funded by the Natural Sciences and Engineering Research Council of Canada.

- *Planar fitting constraint*: The disparity map for each region can be described via planar fitting ($\{a, b, c\}$ tuple) - by fitting $d = ax + by + c$ to a given set of pixel coordinates $\{x, y\}$.
- *Planar hypothesis space constraint*: The space of planar parameters for a given region is limited to within the clique of this region plus the region's current planar parameter.

First, some notations. Let:

- $z \in \{1, ..., Z\}$ is a region z chosen from a total Z;
- $i_z \equiv i$ denotes pixel i which belongs to region z;
- θ_z is a $\{a, b, c\}$ tuple that describes the planar parameters for a given region z. At the pixel level, we use $\theta_{zi} \equiv \theta_i$ for the planar parameter θ_z that apply to pixel $i_z \equiv i$;
- D_z define the disparity map for region z; while d_i is the disparity value for pixel i;
- Y_z correspond to the observed matching difference using the given disparity map for region z, while y_i is the matching difference value for pixel i;
- ∂z denotes the clique for region z, including all the neighbouring regions of z; ∂i means the clique for pixel i, including all the neighbouring pixels of i; $\partial z \backslash x$ further restricts the set by excluding region x from the set; $\partial i \backslash j$ has the similar interpretation.

2.1. The MRF for the region layer

Based on the previous assumptions, we can now define the region layer as a MRF augmented with causal connections (see Fig. 1). Following this causal model we can easily derive the following equations:

$$p(D_z, Y_z|\theta_z) = p(Y_z|D_z)p(D_z|\theta_z) \quad (1)$$

0-7803-7663-3/03/$17.00 ©2003 IEEE

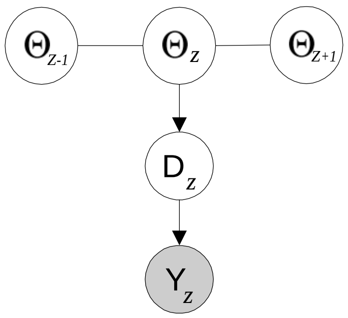

Fig. 1. A 1-D illustration of the region augmented MRF model. The grey node denotes an observable variable while the white ones the unobservable variables.

so the posterior probability for D_z, given θ_z and Y_z is:

$$p(D_z|Y_z, \theta_z) = cp(D_z, Y_z|\theta_z) \quad (2)$$

where c is a constant number and can be omitted.
Similarly, we have:

$$p(\theta_z, D_z|\theta_{\partial z}, Y_z) = p(D_z|\theta_z, Y_z)\ p(\theta_z|\theta_{\partial z}) \quad (3)$$

and the posterior probability for θ_z, given the planar parameters of its neighbouring regions, D_z and Y_z, is:

$$p(\theta_z|\theta_{\partial z}, D_z, Y_z) = cp(\theta_z, D_z|\theta_{\partial z}, Y_z).$$

For simplicity, we assume that the hypothesis space for planar parameters of region z is limited to the planar parameters ($\theta_{\partial z \bigcup z}$) of its cliques where:

$$p(\theta_z|\theta_{\partial z \cup z}) = \prod_{x \in \partial z \cup z} p(\theta_z|\theta_x)$$

and $p(\theta_z|\theta_x)$ is further defined as:

$$p(\theta_z|\theta_{x, x \in \partial z \cup z}) = c \exp\{-l(d_1 + d_2)^2 / \ \sigma^2_{rgn_{nbr}}\}.$$

where rgn_{nbr} refer to the standard deviation for neighboring regions. Also, for z's neighbouring region x, l represents the border length between x and z, d_1 and d_2 are the distances between the segment center and the depth planes, for region z and x, respectively. The term $d_1 + d_2$ describes the depth discontinuity between the two regions (see [6] for details).

As a consequence the posterior likelihood for θ_z can be simplified as:

$$p(\theta_z|\theta_{\partial z}, D_z, Y_z) \propto p(D_z|\theta_z, Y_z) p(\theta_z|\theta_{\partial z \cup z}).$$

Using the Belief Propagation(BP) algorithm, the MAP principle (max-product) for the region layer becomes (refer to Eq. 1 – 5 in [4] for detailed explanation):

Planar id	region 1 (i.e. x_1)	region 2 (x_2)
1	θ_1	θ_2
2	θ_2	θ_1
3	θ_3	θ_3
4	θ_4	θ_5
5	θ_5	θ_6
6		θ_7

Table 1. Shows the hypotheses for regions labeled 1 and 2 in Fig. 2, defined by the parameters (θ_i) of the possible clique neighbours.

$$\theta_z^* = \arg\max_{\theta_z}\{\max_{\theta_x, x \neq z} p(\theta, D|Y)\}. \quad (4)$$

For this MRF layer the joint probability over the region disparity D and the region planar parameters θ can be written as [4]:

$$p(\theta, D|Y) = \prod_{x \in \partial z \cup z} p(\theta_z|\theta_x) \prod_y p(D_y|\theta_y, Y_y)$$

By substitution into Eq. 4, we obtain the MAP estimation at node z:

$$\theta_z^* = \arg\max_{\theta_z}\{p(D_z|\theta_z, Y_z) \prod_{x \in \partial z} m_z^x(\theta_z)\} \quad (5)$$

where x runs over all region nodes neighbouring of node z. $m_z^x(\theta_z)$ is the message flow from node x to node z, defined by:

$$m_z^x(\theta_z) = \max_{\theta_x}\{p(D_x|\theta_x, Y_x) \prod_{y \in \partial x \setminus z} m_x^y(\theta_x)\}. \quad (6)$$

This message passing scheme is simply illustrated in Fig. 2.

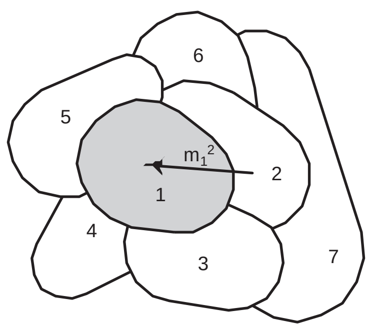

Fig. 2. An example of message passing in the region MRF layer. The message from site 2 to 1 is: $m_1^2 = \max_{x_2} p(\theta_{x_1}|\theta_{x_2}) p(y_2|\theta_{x_2}) m_2^6 m_2^7 m_2^5 m_2^3$. Also refer to Table. 1 for the related hypothesis space of site 1 and 2.

2.2. The MRF for the pixel layer

Figure 3 shows the proposed transition from the region MRF layer to the pixel MRF layer. Similar to the previous section, we derive the following equations according to Fig. 3.

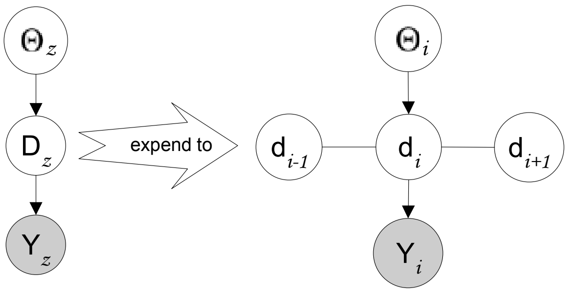

Fig. 3. 1-D illustration of the pixel augmented MRF model, and the transition from the region MRF layer to the pixel MRF layer.

$$p(d_i|y_i, \theta_i, d_{\partial i}) = c * p(y_i|d_i)\, p(d_i|d_{\partial i}, \theta_i)$$

and the MAP principle for the pixel layer is:

$$d_i^* = \arg\max_{d_i}\{p(y_i|d_i) \prod_{j \in \partial i} m_i^j(d_i)\} \tag{7}$$

where j runs over all pixels neighbouring pixel i, $m_i^j(d_i)$ is the message flow from pixel j to pixel i defined by:

$$m_i^j(d_i) = c \max_{d_j}\{p(d_i|d_j)p(y_j|d_j) \prod_{k \in \partial j \setminus i} m_j^k(d_j)\}. \tag{8}$$

The resultant belief function (b) becomes

$$b_i(d_i) = p(y_i|d_i) \prod_{j \in \partial i} m_i^j(d_i). \tag{9}$$

2.3. Combining the pixel and region MRF layers

We can establish the following equations by assuming *iid* distribution for pixels within the current region z obtaining (see Fig. 3):

- The likelihood of the observed intensity differences given the disparities:

$$p(Y_z|D_z) = \prod_{i \in R_z} p(y_i|d_i) \tag{10}$$

- The likelihood of the disparities given the underlying planar parameters for regions:

$$p(D_z|\theta_z) = \prod_{i \in R_z, \partial i \in R_z} p(d_i|d_{\partial i}, \theta_i) \tag{11}$$

Here, the pixel likelihood (the data term) is defined by:

$$p(y_i|d_i) = \exp\{-\frac{(y_i^{ref} - y_{i|d_i}^{mat})^2}{2\sigma_Y^2}\}$$

where $i \in D_z$; ref defines a pixel from the reference view and mat a pixel from the second view; $y_{i|d_i}$ indicates the matching pixel in the second view located using i and d_i from the reference view - under the epipolar constraint; σ_Y is the standard deviation for observation likelihood.

Also the pixel prior term (the smoothness term) is:

$$p(d_i|d_{\partial i}, \theta_i) = \exp\{-\frac{\sum_{i \in D_z, j \in \partial i}(d_j - d_i)^2}{2\sigma_D^2}\}$$

where $i \in D_z$; σ_D is the standard deviation for disparity.

To combine such evidence with the region layer evidence, we now combine Eq. 1, 2, 10, 11 and Eq. 9 together, as:

$$p(D_z|\theta_z, Y_z) = \prod_{i \in R_z} p(y_i|d_i) \prod_{\partial i \in R_z} p(d_i|d_{\partial i}, \theta_i) \tag{12}$$

$$= \prod_{i \in R_z} b_i(d_i). \tag{13}$$

According to Eq. 5, 6, the MAP principle for the region θ estimation could be rewritten as

$$\theta_z^* = \arg\max_{\theta_z}\{\prod_{i \in R_z} b_i(d_i) \prod_x m_z^x(\theta_z)\}. \tag{14}$$

Meanwhile the message is:

$$m_z^x(\theta_z) = \max_{\theta_x}\{\prod_{i \in R_z} b_i(d_i) p(\theta_z|\theta_{x, x \in \partial z \cup z}) \prod_{y \in \partial x \setminus z} m_x^y(\theta_x)\}. \tag{15}$$

Due to the fact that the numerical result of $\prod_{i \in R_z} b_i(d_i)$ is intractable it is very difficult to compute Eq. 13, 14, 15, directly. So we approximate it as:

$$p(D_z|\theta_z, Y_z) = \exp\{\frac{\sum_{i \in R_z} \log b_i(d_i)}{n}\} \tag{16}$$

where n counts the number of pixels in the current region. This way, the rules in Eq. 14 and 15 become:

$$\theta_z^* = \arg\max_{\theta_z} b_z(\theta_z) \tag{17}$$

and

$$m_z^x(\theta_z) = \max_{\theta_x}\{b_{\theta_z} p(\theta_z|\theta_{x, x \in \partial z \cup z}) \prod_{y \in \partial x \setminus z} m_x^y(\theta_x)\}. \tag{18}$$

3. FLOWCHART OF THE ALGORITHM

The resultant flowchart of the algorithm is as follows:

1. Perform over-segmentation of the reference view; Here we use the Mean-Shift procedure [5].

2. At the pixel layer update the pixel disparity estimation using the BP algorithm under the MAP principle (Eq. 7, 8) for each region - loop until the messages converge.

3. According to the current pixel level disparity estimation, run the plane fitting (we use the Normal Equation for plane fitting [2]) to update the planar parameters for the region layer - repeat until converge.

4. At the region layer update the regional disparity estimation using the BP algorithm under the MAP principle (Eq. 14, 15).

4. EXPERIMENTS

We have experimented with the proposed algorithm with many standard stereo pairs. In Fig. 4, two set of results are shown. The top two rows are for a'Map' image and the remaining two rows are 'Lamp and Head' images. For the Map images, in scan line order, there are: the reference (left) view image, the disparity map obtained after Belief Propagation at the pixel MRF layer, the disparity map after the consecutive BP running on the region MRF layer, and finally, the obtained confidence map. The last two rows follow the same set of configuration. These results are comparable with recent methods [6, 1]. Note that besides the plane-fitted disparity map, this algorithm also outputs the confidence map which measures, for each pixel, the confidence of choosing the current disparity value. The confidence for pixel i is computed as:

$$confidence_i = \frac{b_i + \lambda}{(1 - b_i) + \lambda}$$

where λ is a constant, and b_i is the approximated marginal probability for pixel i obtained after the Belief Propagation (white $= 1$, black $= 0$).

5. CONTRIBUTION AND FUTURE WORK

We propose a new probabilistic framework for stereo matching, and use Belief Propagation for approximate inference. Good results are obtained using the proposed algorithm. However, in some cases, the image segmentation results may not be consistent with the corresponding disparity changes. We plan to address this problem in the future work.

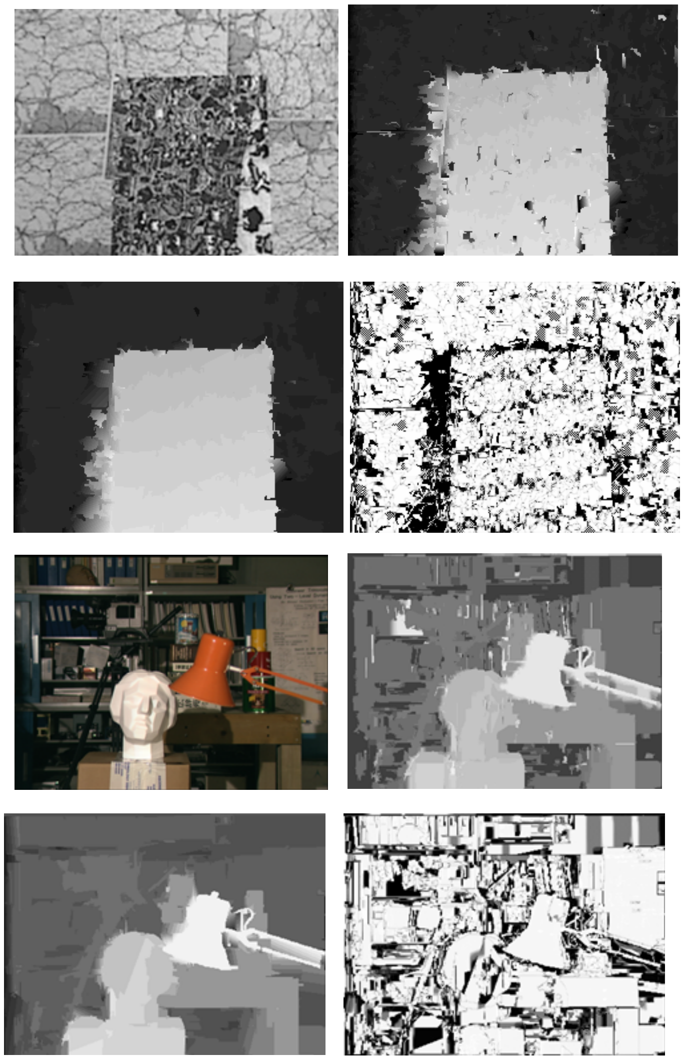

Fig. 4. Some results of our stereo algorithm (See text for details).

6. REFERENCES

[1] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *IEEE Workshop on Stereo and Multi-Baseline Vision, Kauai, HI.*, 2001.

[2] Hai Tao and Harpreet S. Sawhney, "Global matching criterion and color segmentation based stereo," in *Proc. Workshop on the Application of Computer Vision (WACV2000), pp. 246-253, December*, 2000.

[3] W. T. Freeman J. Yedidia and Y. Weiss, "Understanding belief propagation and its generalizations," in *IJCAI 2001*, 2001.

[4] E. C. Pasztor W. T. Freeman and O. T. Carmichael, "Learning lowlevel vision," *Intl. J. Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

[5] P. Meer D. Comanicu, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[6] Harpreet S. Sawhney Hai Tao and Rakesh Kumar, "A global matching framework for stereo computation," in *Proc. IEEE Intl. Conf. on Computer Vision*, 2001.