BELIEF PROPAGATION ON RIEMANNIAN MANIFOLD FOR STEREO MATCHING

Quanquan Gu, Jie Zhou, Senior Member, IEEE

Department of Automation, Tsinghua University, Beijing 100084, China
gqq03@mails.tsinghua.edu.cn

ABSTRACT

Stereo matching has been one of the most active research areas in computer vision for decades. Many methods, ranging from similarity measures to local or global matching cost optimization algorithms, have been proposed. As we known, stereo matching can be formulated under the framework of Markov random field (MRF), and the global optimization in stereo matching can be approximated by inference procedure. There are many exact or approximate inference algorithms, among which belief propagation is one of the most effective. In this paper, by combining Riemannian metric based similarity measure with the belief propagation algorithm, we propose a global optimization method for stereo matching, namely belief propagation on Riemannian manifold (BPRM). Experiments on benchmark dataset demonstrate the encouraging performance of our method.

***Index Terms*—** Stereo matching, Riemannian manifold, Belief propagation, Similarity measure

1. INTRODUCTION

Stereo matching has been one of the most active areas in computer vision for decades. The task of stereo matching is to find the point correspondence between two images taken from different views of the same scene. When the camera geometry is known, we usually rectify the images so that correspondence points are in the same scanline in both images and the correspondence problem is reduced to one dimensional search. Most stereo matching methods usually consist of four steps: (1) image preprocessing; (2) similarity measure selection; (3) local or global matching cost optimization; and (4) disparity postprocessing. In recent years, a large number of methods ranging from similarity measures to local or global optimization algorithms have been proposed. For a comprehensive discussion on stereo matching method, we refer readers to [1].

In our previous work [2], we have proposed a novel similarity measure under Riemannian metric for stereo matching. In this paper, we combine our similarity measure with belief propagation [3][4][5] and propose a global optimization method for stereo matching, namely belief propagation algorithm on Riemannian manifold (BPRM). The Riemannian metric based similarity measure provides an effective way to confuse pixel features, e.g. intensity and derivatives, which contain a lot of discriminative information for stereo matching. It also has many good properties such as scale invariant and illumination invariant. Belief propagation on Riemannian manifold guarantees the energy will decrease at each iteration in the optimization procedure. The main advantage of our method owes to the similarity measure. The experimental results on benchmark dataset indicate that our method has an encouraging performance.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce the novel similarity measure under Riemannian metric for stereo matching. In Section 3, we review the belief propagation algorithm under the framework of Markov random field (MRF) [3][4], and then propose our global optimization method, namely BPRM. The experimental results are demonstrated in Section 4. Finally, we draw a conclusion in Section 5.

2. SIMILARITY MEASURE UNDER RIEMANNIAN METRIC

The most popular window-based similarity measures in stereo matching include sum-of-absolute-difference (SAD), sum-of-square-differences (SSD) and normalized cross correlation (NNC). Nevertheless, all of these similarity measures describe a point by the raw region within a window.

We adopt structure tensor [6] for alternative. Since for stereo matching application, image intensity feature is indispensable. So we define a generalized structure tensor which fuses both image intensity and derivatives as follows:

$$\begin{aligned} T_n &= G * ff^T \\ &= \begin{pmatrix} G * I^2 & G * II_x & G * II_y \\ G * I_xI & G * I_x^2 & G * I_xI_y \\ G * I_yI & G * I_yI_x & G * I_y^2 \end{pmatrix}, \end{aligned} \tag{1}$$

where $f = (I, I_x, I_y)$, I is intensity, I_x and I_y are partial derivatives with respect to x and y. G is the Gaussian smooth filter as:

$$G = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{\sigma^2}\right), \tag{2}$$

This work was supported by Natural Science Foundation of China under grant 60673106 and 60573062.

where σ is the standard deviation. The generalized structure tensor represents the local orientation by its eigenvectors and eigenvalues.

The distance between point descriptors is usually used for the measurement of similarity. However, the structure tensor lie in a Riemannian manifold. In order to clarify the distance between structure tensors, we will first introduce the Riemannian geometry [7] in brief.

A manifold M is a topological space which is locally homeomorphism to a Euclidean space. The derivatives at point X lie in a vector space T_X, called tangent space.

A Riemannian manifold is a differential manifold in which each tangent space has a Riemannian metric $< y, y >$. The inner product induces a norm $||y||$.

The minimum length curve connecting two points on the manifold is called the geodesic. The distance between $X, Y \in M$ is the length of the geodesic. let $y \in T_X$, there exist an exponential map,$\exp_X : T_X \mapsto M$. In general, the exponential map is one to one in a neighborhood of X and maps the y to the point reached by the geodesic. The inverse map, called logarithm map, $\log_X : M \mapsto T_X$, maps the Y to a tangent vector with smallest norm. So we can take this smallest norm to measure the distance between X and Y:

$$d^2(X, Y) = d^2(X, \exp_X(y)) = ||y||_X^2 =< y, y >_X . \quad (3)$$

The structure tensor, which is a symmetric positive definite matrix, forms a Riemannian manifold. According to [8][9], we define a Riemannian metric like that:

$$< y, z >_X = \mathrm{tr}(X^{-1/2} y X^{-1} z X^{-1/2}). \quad (4)$$

The exponential map associated to the above Riemannian metric is

$$\exp_X(y) = X^{1/2} \exp(X^{-1/2} y X^{-1/2}) X^{1/2}. \quad (5)$$

By Eq.(5) we can obtain the logarithm map

$$y = \log_X(Y) = X^{1/2} \log(X^{-1/2} Y X^{-1/2}) X^{1/2}. \quad (6)$$

Submit Eq.(6) to Eq.(3)

$$\begin{aligned} d^2(X, Y) &= ||y||_X^2 =< y, y >_X \\ &= < \log_X(Y), \log_X(Y) >_X \\ &= \mathrm{tr}(\log^2(X^{-1/2} Y X^{-1/2})). \quad (7) \end{aligned}$$

It is just the distance between structure tensors. Furthermore, Eq.(7) is equivalent to

$$d(X, Y) = \sqrt{\sum_{k=1}^{d} \log^2 \lambda_k(X, Y)}, \quad (8)$$

where $\lambda_k(X, Y)$ are the generalized eigenvalues of X and Y.

3. BELIEF PROPAGATION ON RIEMANNIAN MANIFOLD

In this part, we first describe the stereo matching task under the framework of Markov random filed (MRF) [3][4], which belongs to undirected probabilistic graph. Let $\mathcal{P}$ be the set of pixels in an image and $\mathcal{L}$ be a finite set of labels. The labels correspond to the disparities that we want to estimate at each pixel. Suppose the label of each pixel p in the image is a random variable, denoted by l_p. Each variable has a state space of dimension k, etc, number of disparity level. On MRF, we define two types of potential functions, denoted by $\Phi(\cdot)$ and $\Psi(\cdot, \cdot)$. While $\Phi(l_p)$ represents the suitability of assigned label, $\Psi(l_p, l_q)$ reflects the compatibility of assigned labels between neighboring pixels.

With the potential functions defined, the joint probability of the MRF can be written as:

$$p(\mathcal{P}, \mathcal{L}) = \prod_{\mathcal{P}} \Phi(l_p) \prod_{(p,q) \in \mathcal{N}} \Psi(l_p, l_q), \quad (9)$$

where $\mathcal{N}$ are the edges in the four-connected image grid for computational simplicity.

By taking negative logarithm probabilities on Eq.(9), we obtain:

$$E(\mathcal{P}, \mathcal{L}) = \sum_{\mathcal{P}} - \log \Phi(l_p) + \sum_{(p,q) \in \mathcal{N}} - \log \Psi(l_p, l_q). \quad (10)$$

We rewrite Eq.(10) as follows:

$$E(f) = \sum_{p \in \mathcal{P}} D_p(l_p) + \sum_{(p,q) \in \mathcal{N}} V(l_p, l_q), \quad (11)$$

$D_p(l_p)$ is the cost of assigning label l_p to pixel p, and is referred to data cost. $V(l_p, l_q)$ measures the cost of assigning labels l_p and l_q to two neighboring pixels, and is referred to smoothness cost. Note that the global optimization based stereo matching corresponds to the maximum a posteriori (MAP) estimation problem for the MRF.

In our experiment, we choose the truncated linear model as smoothness cost

$$V(l_p, l_q) = \min(c|l_p - l_q|, V_{max}), \quad (12)$$

where c is the slope of linear model and V_{max} is the upper bound of the increase.

The optimal labeling with minimum energy is approximated by belief propagation [3][4][5]. The max-product BP algorithm works by passing messages around the MRF. The method is iterative, with messages from all nodes being passed in parallel. Each message is a vector of dimension k. Let $m_{p \to q}^t$ be the message that node p sends to a neighboring node q at iteration t. When using negative log probabilities all entries in $m_{p \to q}^0$ are initialized to zero, and at each iteration new messages are computed in the following way

$$m^t_{p\to q}(l_q) = \min_{l_p}(V(l_p, l_q) + D(l_p) + \sum_{s\in\mathcal{N}(p)\setminus q} m^{t-1}_{s\to p}(l_p)), \tag{13}$$

where $\mathcal{N}(p)\setminus q$ denotes the neighbors of p other than q. After T iterations a belief vector is computed for each node,

$$b_q(l_q) = D_q(l_q) + \sum_{p\in\mathcal{N}(q)} m^T_{p\to q}(l_p). \tag{14}$$

The approximate optimal labeling l^*_q is obtained individually by

$$l^*_q = \arg\min b_q(l_q). \tag{15}$$

When it comes to our similarity measure, each pixel, i.e. node in MRF, is represented by a generalized structure tensor. So the MRF is established on a Riemannian manifold. Note that only the item $D(l_p)$ in Eq.(13) depends on the Riemannian metric directly. Thus in order to combine our similarity measure with belief propagation, we need to replace the calculation of data cost $D_p(l_p)$ using either region-wise SSD, SAD or NNC by

$$D_p(l_p) = \sqrt{\sum_{k=1}^{d} \log^2 \lambda_k(T_p, T_{p+l_p})}, \tag{16}$$

where T_p is the structure tensor descriptor of pixel p in reference image and T_{p+l_p} is the structure tensor descriptor of pixel p' in target image.

We summarize the belief propagation on Riemannian manifold method (BPRM) as follows:

Input: I_r and I_t: reference image and target image
d_{min},d_{max}: disparity range
T_{size}: structure tensor size
σ: Gaussian standard deviation
Output: Disparity image
foreach *pixel* p_1 *in* I_r **do**
    foreach $l \in [d_{min}, d_{max}]$ **do**
        There is a pixel p_2 in I_t
        Compute generalized structure tensors T_1 and T_2 with respect to p_1 and p_2 by Eq.(1)
        Calculate $D(l)$ by Eq.(16)
    end
end
Run belief propagation by Eq.(13)
foreach *pixel* p_1 *in* I_r **do**
    Calculate the disparity l by Eq.(15)
end

Algorithm 1: Belief propagation on Riemannian manifold (BPRM)

4. EXPERIMENTAL RESULTS

The proposed method was evaluated using the Middlebury evaluation website[1] provided by [1]. There are four benchmark image pairs, namely ”Tsukuba”, ”Venus”, ”Teddy” and ”Cones”. For more details about the dataset and evaluation method, please refer to [1] or the website.

Since it is required that each method runs with constant parameters on all 4 image pairs, except for the disparity ranges, we set $\sigma = 1.5$ in Eq.(2), $c = 1$ and $V_{max} = 20$ in Eq.(12), and the size of window to calculate the generalized structure tensor is 5×5 for all of the 4 image pairs. Qualitative results are shown in Figure1. Quantitative results are given in Table1, which was generated by the Middlebury evaluation website, listing the percentage of ”bad” pixels for different regions: non-occluded region (nonocc), whole image (all) and pixels near discontinuities (disc). There are totally 34 algorithms provided by the Middlebury evaluation website with which we compare our method. For the space limit, we only list the top 12 algorithms in Table.1. The first column contains the names of the algorithms. The items of the second column are the respective average ranks of the algorithms. The subscript numbers indicate the rank of each method in each column. We can find that the performance of our method is encouraging, which owes to the similarity measure. Although it is not the best one, it is among the top 10. Especially, our method acquired very good results on ”Teddy” and ”Cones” since these two images are rich of texture and our Riemannian metric based similarity measure is very suitable for this kind of images. It should be noted that our method does not adopt any sophisticated strategies such as segment-based preprocessing, occlusion handling, adaptive weighted measure, which are adopted extensively in most of the other 34 methods, especially in the top 6 algorithms. We believe that the performance of our method will improve a lot if we adopt those strategies, however, it is beyond the attention of our work in this paper.

5. CONCLUSIONS

In this paper, we combine our similarity measure with the belief propagation algorithm, and propose a global optimization method for stereo matching, namely BPRM. The experimental results demonstrate the satisfying performance of our method on Middlebury stereo evaluation test bed.

6. REFERENCES

[1] Daniel Scharstein and Richard Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[1] http://vision.middlebury.edu/stereo/

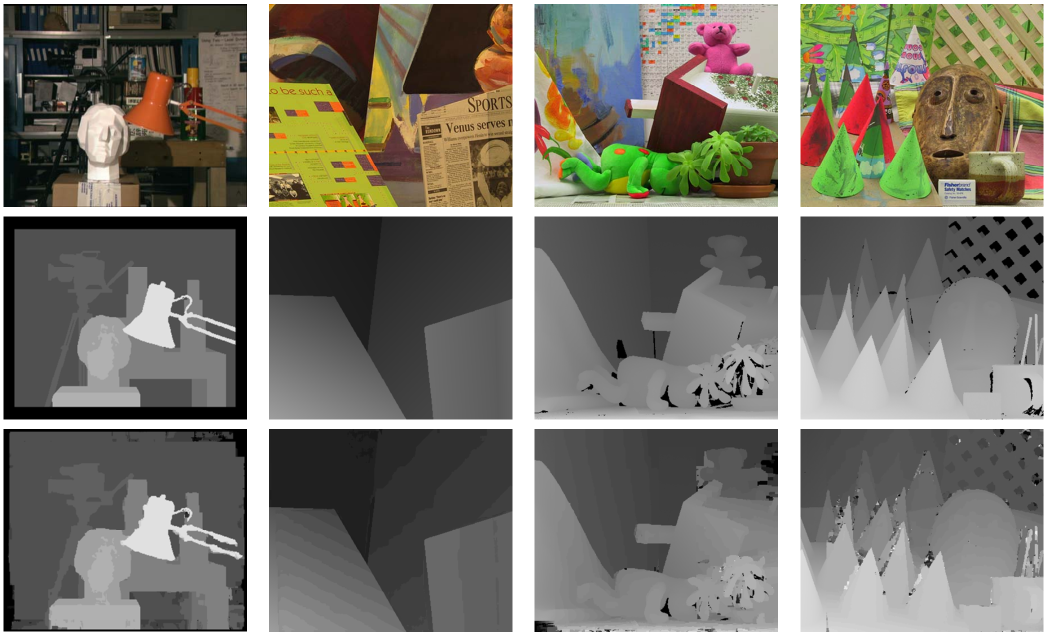

Fig. 1. Results using the Middlebury datasets: Tsukuba, Venus, Teddy and Cones. The top row is original images, the mid row is disparity ground truth, and the bottom row is disparity images acquired by our method.

Table 1. Middlebury stereo evaluation on different algorithms, ordered according to their overall performance. The subscript numbers indicate the rank of each method in each column. Our method ranks seven.

Algorithm	Avg. Rank	Tsukuba			Venus			Teddy			Cones		
		nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
AdaptingBP	2.5	1.11_5	1.37_2	5.79_6	0.10_1	0.21_2	1.44_1	4.22_3	7.06_2	11.8_3	2.48_1	7.92_3	7.32_1
DoubleBP	3.8	0.88_1	1.29_1	4.76_1	0.14_4	0.60_{10}	2.00_6	3.55_2	8.71_5	9.70_1	2.90_3	9.24_{10}	7.80_2
SubPixDoubleBP	4.8	1.24_9	1.76_{10}	5.98_8	0.12_2	0.46_4	1.74_4	3.45_1	8.38_4	10.0_2	2.93_4	8.73_7	7.91_3
AdaptOvrSegBP	8.8	1.69_{19}	2.04_{17}	5.64_5	0.14_3	0.20_1	1.47_2	7.04_{13}	11.1_7	16.4_{11}	3.60_{10}	8.96_9	8.84_9
PlaneFitBP	9.1	0.97_4	1.83_{11}	5.26_4	0.17_6	0.51_5	1.71_3	6.65_8	12.1_{11}	14.7_7	4.17_{17}	10.7_{17}	10.6_{16}
SymBP+occ	9.4	0.97_3	1.75_9	5.09_3	0.16_5	0.33_3	2.19_7	6.47_7	10.7_6	17.0_{14}	4.79_{21}	10.7_{18}	10.9_{17}
Our Method	**9.8**	$\mathbf{1.17_6}$	$\mathbf{2.74_{18}}$	$\mathbf{5.92_7}$	$\mathbf{0.56_{15}}$	$\mathbf{0.74_{14}}$	$\mathbf{6.60_{17}}$	$\mathbf{6.77_{10}}$	$\mathbf{8.32_3}$	$\mathbf{14.2_6}$	$\mathbf{3.32_9}$	$\mathbf{6.68_1}$	$\mathbf{9.66_{12}}$
Segm+visib	10.4	1.30_{13}	1.57_3	6.92_{16}	0.79_{18}	1.06_{16}	6.76_{19}	5.00_4	6.54_1	12.3_4	3.72_{11}	8.62_6	10.2_{14}
C-SemiGlob	10.6	2.61_{26}	3.29_{21}	9.89_{23}	0.25_9	0.57_7	3.24_{12}	5.14_5	11.8_8	13.0_5	2.77_2	8.35_5	8.20_4
SO+borders	10.8	1.29_{12}	1.71_6	6.83_{13}	0.25_{10}	0.53_6	2.26_8	7.02_{12}	12.2_{12}	16.3_9	3.90_{13}	9.85_{13}	10.2_{15}
DistinctSM	12.1	1.21_8	1.75_8	6.39_10	0.35_{11}	0.69_{13}	2.63_{11}	7.45_{17}	13.0_{15}	18.1_{17}	3.91_{14}	9.91_{15}	8.32_6
OverSegmBP	12.3	1.69_{20}	1.97_{14}	8.47_{20}	0.50_{14}	0.68_{12}	4.69_{15}	6.74_9	11.9_{10}	15.8_8	3.19_7	8.81_8	8.89_{10}

[2] Quanquan Gu and Jie Zhou, "A novel similarity measure under riemannian metric for stereo matching," in *ICASSP*, 2008, pp. 1073–1076.

[3] D. Scharstein O. Veksler V. Kolmogorov A. Agarwala M. Tappen R. Szeliski, R. Zabih and C. Rother, "A comparative study of energy minimization methods for markov random fields," in *ECCV(2)*, 2006, pp. 16–29.

[4] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters," in *ICCV*, 2003, pp. 900–907.

[5] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient belief propagation for early vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41–54, 2006.

[6] Hai-Yun Wang and Kai-Kuang Ma, "Accurate optical flow estimation using adaptive scale-space and 3D structure tensor," in *ICIP (2)*, 2002, pp. 301–304.

[7] W.M. Boothby, *Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, 1975.

[8] Xavier Pennec, Pierre Fillard, and Nicholas Ayache, "A riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.

[9] O. Tuzel, F. M. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds," in *CVPR*, 2007, pp. 1–8.