A NOVEL SIMILARITY MEASURE UNDER RIEMANNIAN METRIC 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 areas in computer vision for decades. Many methods, ranging from similarity measures to local or global matching cost optimization algorithms, have been proposed. In this paper, we propose a novel similarity measure under Riemannian metric. A generalized structure tensor is applied to describe a point and the similarity is measured by the distance between the associated tensors. Since the structure tensor lies in a Riemannian manifold, the distance between structure tensors is the geodesic distance on Riemannian manifold. We will show that our similarity measure provides an efficient way to fuse different features and it is independent of illumination change and window scaling. Experiments on standard dataset prove that our similarity measure outperforms many traditional measures such as SSD, SAD and normalized cross-correlation (NCC).

***Index Terms*—** Stereo matching, similarity measure, structure tensor, Riemannian metric

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 this paper, our interest mainly focuses on similarity measure since it is the foundation of the stereo matching. The similarity measures can be classified into pixel-based method and window-based method. In practice, we usually choose window-based method which uses a window centering at the point of interest to describe it. The popular window-based similarity measures in stereo matching include sum-of-square-differences (SSD) [2], sum-of-absolute-difference (SAD) [3] and normalized cross correlation (NCC) [4]. The SAD and SSD assume brightness constancy for corresponding pixels while the NNC can compensate differences in gain and bias. All of these methods are mostly adopted based only on image intensity. When fusing more features, the common strategy is just computing SAD, SSD and NNC on each feature respectively and summing them up.

In this paper, we propose a novel similarity measure for stereo matching. First, we adopt the structure tensor [5] to describe a point in the image, which is generalized to fuse different features, e.g. image intensity and derivatives. After that, we can measure the similarity by the distance between pair-wise structure tensors. The structure tensors do not lie in a vector space, otherwise, they form a positive definite matrix space, which is a Riemannian manifold. So we can calculate the distance under Riemannian metric for the measurement of similarity in stereo matching. The similar solution has been applied to object tracking [6] and human detection [7]. As we know, there is little work related with the method mentioned above in stereo matching area.

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed similarity measure for stereo matching in detail, discuss their good properties and give an application example in stereo matching. The experiment results are demonstrated in Section 3. In section 4, we draw a conclusion and point out the future works.

2. OUR SIMILARITY MEASURE FOR STEREO MATCHING

Similarity measure for stereo matching generally consists of two aspects: (1) point descriptor and (2) similarity measurement. In the following, we will discuss our similarity measure for stereo matching from these two aspects respectively. Then we summarize its good properties for stereo matching. After that, we will combine the proposed similarity measure with WTA (winner-take-all) strategy to derive a prototype stereo matching algorithm as its application.

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

2.1. Structure Tensor Descriptor

We usually describe a point in an image by the intensity, color, derivatives and even higher order derivatives. The most popular window-based similarity measures in stereo matching include SAD, SSD and NNC. Nevertheless, all of these similarity measures describe a point by the raw region within the window.

We adopt structure tensor [5] [8] of a region for alternative. Structure tensor was usually used for low-level feature analysis and gained great success in corner detection [5], optical flow estimation [8] and so on. Given a pixel $I(x, y)$, structure tensor is based on the window W centering at the pixel. The naive structure tensor is represented as:

$$T_n = \begin{pmatrix} G * I_x^2 & G * I_x I_y \\ G * I_x I_y & G * I_y^2 \end{pmatrix}, \tag{1}$$

where Ix and Iy denote the partial derivatives in x and y, respectively. G is the Gaussian smooth filter as:

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

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

$$\begin{aligned} T_n &= G * ff^T \\ &= \begin{pmatrix} G * I^2 & G * II_x & G * II_y \\ G * I_x I & G * I_x^2 & G * I_x I_y \\ G * I_y I & G * I_y I_x & G * I_y^2 \end{pmatrix}, \end{aligned} \tag{3}$$

where $f = (I, I_x, I_y)$, I is intensity, I_x and I_y are partial derivatives with respect to x and y.

2.2. Distance between Structure Tensors

It is usually use the distance between point descriptors for the measurement of similarity. For instance, SSD can been seen as Frobenius norm while SAD as $l1$ norm, and the NCC is the angle between two vectors. However, the structure tensor does not lie in a vector space since the structure tensor space is not closed after multiplying a negative scalar. In order to clarify the distance between structure tensors, we will first introduce the Riemannian geometry [9] 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 for measuring the distance between X and Y:

$$d(X, Y)^2 = d(X, \exp_X(y))^2 = ||y||_X^2 =< y, y >_X . \tag{4}$$

The structure tensor, which is symmetric positive definite matrix, forms a Riemannian manifold. According to [10], we define a Riemannian metric like that:

$$< y, z >_X = tr(X^{-1/2} y X^{-1} z X^{-1/2}). \tag{5}$$

The exponential map associated to the above Riemannian metric is

$$\exp_X(y) = X^{1/2} \exp(X^{-1/2} y X^{-1/2}) X^{1/2}. \tag{6}$$

By Eq.(6) we can obtain the logarithm map

$$y = \log_X(Y) = X^{1/2} \log(X^{-1/2} Y X^{-1/2}) X^{1/2}. \tag{7}$$

Submit Eq.(7) to Eq.(4)

$$\begin{aligned} d^2(X, Y) &= ||y||_X^2 =< y, y >_X \\ &=< \log_X(Y), \log_X(Y) >_X \\ &= tr(\log^2(X^{-1/2} Y X^{-1/2})). \end{aligned} \tag{8}$$

It is just the distance between structure tensors. Furthermore, Eq.(8) is equivalent to

$$d(X, Y) = \sqrt{\sum_{k=1}^{d} \log^2 \lambda_k(X, Y)}, \tag{9}$$

where $\lambda_k(X, Y)$ are the generalized eigenvalues of X and Y and this problem can been solved by SVD decomposition.

2.3. Properties

The proposed similarity measure has a number of good properties for stereo matching.

First of all, the proposed similarity measure provides an effective way to fuse different features. As we have seen above, the structure tensor descriptor can be generated by feature vector f. For example, the feature vector f can also be defined as

$$f = \left[I \quad |I_x| \quad |I_y| \quad \sqrt{I_x^2 + I_y^2} \quad |I_{xx}| \quad |I_{yy}| \right], \tag{10}$$

where I_{xx} and I_{yy} are second order partial derivatives of intensity with respect to x and y, and $\sqrt{I_x^2 + I_y^2}$ are the magnitude of the gradient. In our experiments, we selected the feature vector as $f = (I, I_x, I_y)$ to compromise between accuracy and computation efficiency.

The second advantage of our similarity measure is scale invariant since the order of structure tensor descriptor does not depend on the window size, otherwise, it is determined by the dimension of the feature vector. This property enables comparing two windows without being restricted to the same window size. It can help us to easily design an asymmetric window size matching algorithm for stereo matching.

Thirdly, our similarity is invariant to varying illumination since the structure tensor descriptor contains the partial derivatives which can compensate largely the illumination change. When the two images taken under different light conditions, or the cameras which taking the images have chromatic aberration, it contributes a lot to overcome these difficulties.

2.4. Algorithm

Here we will show that how to combine our similarity measure with cost optimization algorithm. The proposed similarity can be adopted in both local and global cost optimization algorithms such as graph cut [11]. Taking into account of simplicity, We choose WTA local optimization algorithm for example. The algorithm is summarized in Table.1. It also provides a framework to compare our similarity measure with SAD, SSD and NCC measures in an unbiased way.

Table 1. The flow chart of the stereo matching algorithm which combines the proposed similarity measure with WTA.

Algorithm
Input: I_r and I_t: reference image and target image
d_{min},d_{max}: disparity range
W_{size}: search window size
σ: gaussian standard deviation
Initialize: convolute I_r and I_t with gaussian kernel to generate I'_r and I'_t
For each pixel p_1 in I'_r
For $d \in [d_{min}, d_{max}]$, there is a p_2 in I'_t
-Compute generalized structure tensors T_1 and T_2 with respect to p_1 and p_2
-Calculate $dis(T_1, T_2) = \sqrt{\sum_{k=1}^{d} \log^2 \lambda_k(T_1, T_2)}$
End
$disparity = \arg\min_{d \in [d_{min}, d_{max}]} dis$
End
Output: Disparity image

3. EXPERIMENTAL RESULTS

We performed the experiments on widely-used Middlebury data sets. The data sets consist of dozens of data, each of which has 9 images and ground-truth disparities.

To evaluate our results, we introduce a quality metric named percentage of bad matching pixels similar with that in [1], but without considering different kinds of regions:

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x, y) - d_T(x, y)| > \delta_d), \qquad (11)$$

where δ_d is a disparity error tolerance. In our experiments, we use $\delta_d = 1$, the σ in Eq.(2) is set as 1.5 and the search window size W_{size} is 9.

To compare the proposed similarity measure with other measures fairly, the same features are used for SAD, SSD, NNC and our similarity measure. However, in our experiments, we found that the performances of SAD, SSD and NNC which fused intensity and partial derivatives are not always as good as those based only on intensity. Since there is not an effective way to fuse features for these methods rather than summing up the cost associated with each feature. This counts for the result. So we list the results of SAD, SSD and NNC based only on intensity and on intensity and derivatives in Table.2 to compare with our similarity measure. Figure.1 demonstrates the results on "Sawtooth" and "Teddy" data.

From Table.2, we can see that the percentage of bad matching of our similarity measure is lower than other measures on this data set. In Figure.1, the red marks in disparity image indicate the pixels which are bad matching. It is obvious that the count of bad matching pixels by our similarity measure is much smaller, especially in textureless region and disparity discontinuous region.

Table 2. Bad matching percentage of the four Similarity measures on five datas. SAD1, SSD1, NNC1 denote the measures based only on intensity; SAD2, SSD2, NNC2 denote the measures based on intensity and derivatives.

Data	Sawtooth	Venus	tsukuba	Cones	Teddy
SAD1	0.0066	0.0124	0.0207	0.0284	0.0421
SAD2	0.0071	0.0056	0.0277	0.0269	0.0165
SSD1	0.0076	0.0103	0.0205	0.0231	0.0316
SSD2	0.0096	0.0060	0.0210	0.0258	0.0180
NNC1	0.0085	0.0047	0.0211	0.0303	0.0124
NNC2	0.0014	0.0045	0.0191	0.0296	0.0108
Ours	0.0010	0.0031	0.0182	0.0227	0.0100

4. CONCLUSIONS

In this paper, we have proposed a novel similarity measure under Riemannian metric for stereo matching. Our similarity measure has many excellent properties. The experiment results demonstrate the advantages of our similarity measure over traditional measures. Future work will devote to combining our similarity measure with global optimization algorithm such as graph cut [11] to explore a high performance stereo matching approach.

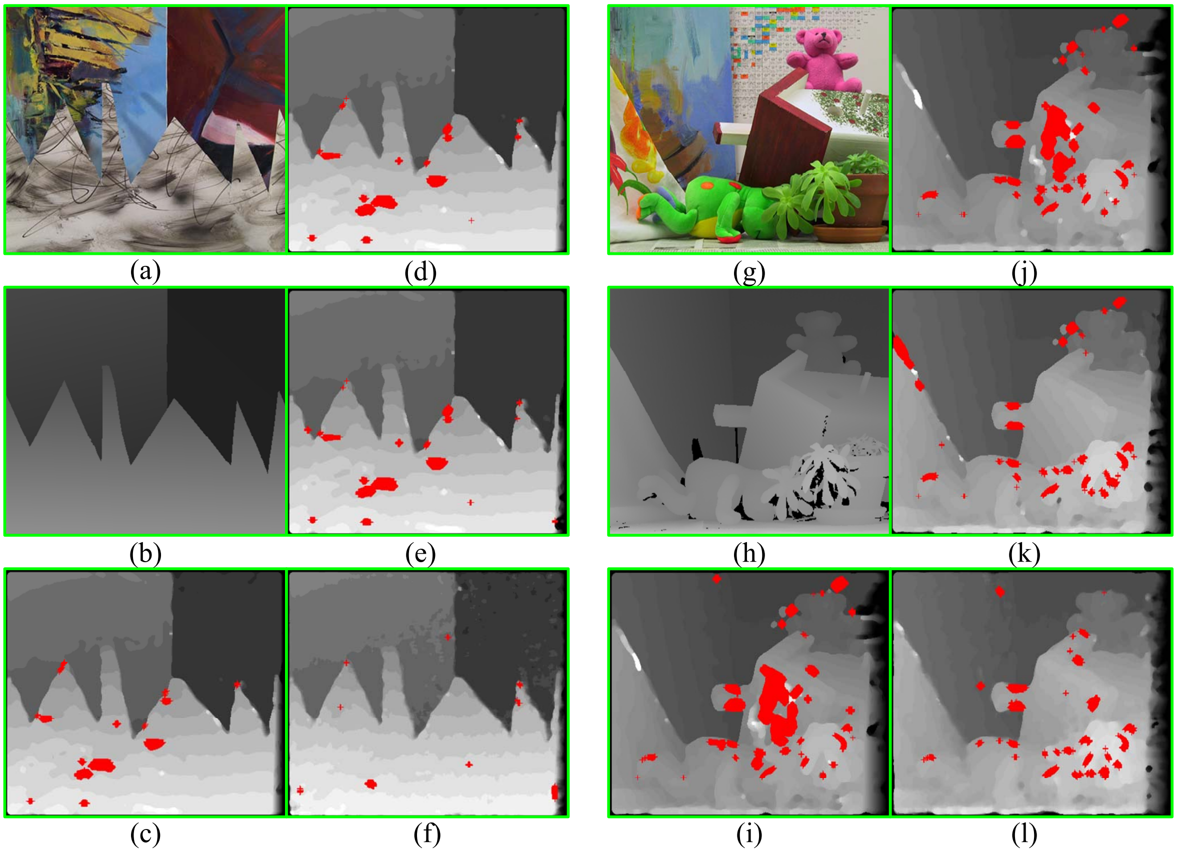

Fig. 1. Results on "Sawtooth" and "Teddy". (a)-(f) are respectively the original image, ground-truth disparity, SAD result, SSD result, NNC result and our result of "Sawtooth". (g)-(l) are respectively the original image, ground-truth disparity, SAD result, SSD result, NNC result and our result of "Teddy". The red crosses in disparity images label the pixels that are bad matching.

5. 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.

[2] E P Simoncelli, E H Adelson, and D J Heeger, "Probability distributions of optical flow," in *Proc Conf on Computer Vision and Pattern Recognition*, Mauii, Hawaii, 1991, pp. 310–315, IEEE Computer Society.

[3] T.[Takeo] Kanade, "Development of a video-rate stereo machine," in *Image Understanding Workshop*, 1994, pp. I:549–557.

[4] Marsha Jo. Hannah, *Computer matching of areas in stereo images.*, Ph.D. thesis, Dept. of Computer Science, Stanford University, 1974.

[5] C. Harris and M. J. Stephens, "A combined corner and edge detector," in *Alvey Conference*, 1988, pp. 147–152.

[6] Fatih Porikli, Oncel Tuzel, and Peter Meer, "Covariance tracking using model update based on lie algebra," in *CVPR (1)*, 2006, pp. 728–735.

[7] O. Tuzel, F. M. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds," in *IEEE Computer Vision and Pattern Recognition or CVPR*, 2007, pp. 1–8.

[8] 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.

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

[10] 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.

[11] Vladimir Kolmogorov and Ramin Zabih, "Computing visual correspondence with occlusions via graph cuts," in *ICCV*, 2001, pp. 508–515.