A Similarity Measure under Log-Euclidean Metric for Stereo Matching

Conference Paper · December 2008

DOI: 10.1109/ICPR.2008.4761347 · Source: DBLP

Quanquan Gu
University of Virginia

A Similarity Measure under Log-Euclidean Metric for Stereo Matching

Quanquan Gu and Jie Zhou*
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 Log-Euclidean metric for stereo matching. 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 Log-Euclidean metric is adopted to calculate the distance between the generalized structure tensors. The proposed similarity measure can provide an effective and efficient way to fuse different features and is independent of illumination change and window scaling. Experiments on standard data set prove that the proposed similarity measure outperforms traditional measures such as SSD, SAD and normalized-cross-correlation (NCC).

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. For a comprehensive discussion on stereo matching methods, we refer readers to [7]. In this paper, we focus on similarity measure, which is the foundation of stereo matching. The similarity measures can be categorized into pixel-based and window-based ones. In practice, we usually choose window-based similarity measures for local optimization algorithms while pixel-based ones for global optimization algorithms. The most popular window-based similarity measures in stereo matching include sum-of-absolute-difference (SAD) [5][4], sum-of-square-differences (SSD) [8] and normalized-cross-correlation (NCC) [2]. The most widely used pixel-based similarity measures include AD [7] and SD [9]. The (S)AD and (S)SD assume brightness constancy for corresponding pixels while the NNC can compensate differences in gain and bias. All of these similarity measures are mostly adopted based only on image intensity. When using more features, e.g. image intensity and derivatives, the common strategy is just computing SAD, SSD and NNC on each feature respectively and summing them up by weight [5].

In this paper, we propose a novel similarity measure for stereo matching. First, the generalized structure tensor [6] is adopted to describe a point in the image, which fuses different features, e.g. image intensity and derivatives. After that, the similarity is measured 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. The Log-Euclidean metric [1] is used to calculate the distance between the generalized structure tensors. Furthermore, a fast algorithm is presented to calculate the generalized structure tensor to save computational cost. As far 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, the proposed similarity measure for stereo matching is introduced in detail, followed by a discussion of their good properties. The experiment results are demonstrated in Section 3. In section 4, we draw a conclusion.

2 The Proposed Similarity Measure for Stereo Matching

The 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 its good properties for stereo matching are summarized.

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

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

2.1 Structure tensor descriptor

We adopt structure tensor [3] [14] of a region as point descriptor. Structure tensor was usually used for low-level feature analysis and it gained great success in corner detection [3], optical flow estimation [14] and so on. Given a pixel $\mathbf{I}(x, y)$, structure tensor is based on the window R of size $N \times N$ centering at the pixel. The structure tensor is represented as:

$$\mathbf{T} = \begin{pmatrix} \mathbf{G} * \mathbf{I}_x^2 & \mathbf{G} * \mathbf{I}_x \mathbf{I}_y \\ \mathbf{G} * \mathbf{I}_x \mathbf{I}_y & \mathbf{G} * \mathbf{I}_y^2 \end{pmatrix}, \tag{1}$$

where $\mathbf{I}_x$ and $\mathbf{I}_y$ denote the partial derivatives along x and y, respectively. $\mathbf{G}$ is the Gaussian smooth filter:

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

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 [6] which fuses both image intensity and derivatives as:

$$\begin{aligned} \mathbf{T} &= \mathbf{G} * \mathbf{f}\mathbf{f}^T \\ &= \begin{pmatrix} \mathbf{G} * \mathbf{I}^2 & \mathbf{G} * \mathbf{I}\mathbf{I}_x & \mathbf{G} * \mathbf{I}\mathbf{I}_y \\ \mathbf{G} * \mathbf{I}_x \mathbf{I} & \mathbf{G} * \mathbf{I}_x^2 & \mathbf{G} * \mathbf{I}_x \mathbf{I}_y \\ \mathbf{G} * \mathbf{I}_y \mathbf{I} & \mathbf{G} * \mathbf{I}_y \mathbf{I}_x & \mathbf{G} * \mathbf{I}_y^2 \end{pmatrix}, \end{aligned} \tag{3}$$

where $\mathbf{f} = (\mathbf{I}, \mathbf{I}_x, \mathbf{I}_y)$, $\mathbf{I}$ is intensity, $\mathbf{I}_x$ and $\mathbf{I}_y$ are partial derivatives with respect to x and y.

2.2 Distance between structure tensors

The distance between point descriptors is usually used for the measurement of similarity. For instance, SSD can be seen as Frobenius norm while SAD as l_1 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 fact, it lies in a Riemannian manifold. In order to calculate the distance between the generalized structure tensors, we adopt the Log-Euclidean metric [1].

The basic idea of Log-Euclidean metric [1] is to construct an equivalent relationship between Riemannian manifold and the vector space of the symmetric matrix. A symmetric positive definite matrix $\mathbf{X}$ has a unique symmetric matrix logarithm $\mathbf{L} = \log(\mathbf{X})$. Conversely, each symmetric matrix $\mathbf{L}$ is associated with a symmetric positive definite matrix $\mathbf{X}$ by the matrix exponential, $\mathbf{X} = \exp(\mathbf{L})$. Since this is one-to-one mapping, one can transfer the standard algebraic operation, addition and scalar multiplication, to the Riemannian manifold. We define the *logarithmic multiplication*, $\odot$, and the *logarithmic scalar muliplication*, $\circ$, as:

$$\mathbf{X} \odot \mathbf{Y} = \exp(\log(\mathbf{X}) + \log(\mathbf{Y})), \tag{4}$$

$$\lambda \circ \mathbf{X} = \exp(\lambda \log(\mathbf{X})) = \mathbf{X}^\lambda. \tag{5}$$

With $\odot$ and $\circ$, the Riemannian manifold has constructed a vector space structure.

When we consider only the $\odot$ on the Riemannian manifold, we have a Lie group structure, i.e. a space which is a Riemannian manifold as well as a group. Among Riemannian metrics in Lie groups, the most convenient one in practice, when it exists, is bi-invariant metric, which is invariant by multiplication and inversion. For the symmetric positive definite matrix, a bi-invariant metric, namely Log-Euclidean metric, exists and is particularly simple:

$$d(\mathbf{X}, \mathbf{Y}) = ||\log(\mathbf{X}) - \log(\mathbf{Y})||_F, \tag{6}$$

where $||\cdot||_F$ is Frobenius norm. Rewriting Eq.(6) by the equality $||\mathbf{A}||_F^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}^T) = \mathrm{tr}(\mathbf{A}^T\mathbf{A})$, we obtain:

$$d(\mathbf{X}, \mathbf{Y}) = \sqrt{\mathrm{tr}((\log(\mathbf{X}) - \log(\mathbf{Y}))^2)}. \tag{7}$$

It should be noted that the Log-Euclidean metric are invariant under orthogonal transformation and scaling.

2.3 Fast calculation of the generalized structure tensor

In this part, we revisit the calculation of the generalized structure tensor. Considering the general case that $\mathbf{f} = (\mathbf{f}_1, \ldots, \mathbf{f}_d)^T$, the computational complexity of the calculation of the generalized structure tensor is $O(HWd^2N^2)$, where H and W are the height and width of the image respectively, and N is the size of the window R. It is very time consuming. So a fast calculation algorithm is in need.

Inspired by the integral images [13][12], we derive a fast calculation algorithm for the generalized structure tensor.

The element of the generalized structure tensor defined in Eq.(3) is

$$\mathbf{T}(i, j) = \sum_{k=1}^{N^2} \mathbf{G}(i, j)\mathbf{f}_k(i)\mathbf{f}_k(j). \tag{8}$$

We construct a $H \times W \times d \times d$ tensor of the second order integral images as:

$$\mathbf{E}(x\prime, y\prime, i, j) = \sum_{x<x\prime, y<y\prime} \mathbf{G}(i, j)\mathbf{F}(x, y, i)\mathbf{F}(x, y, j), \tag{9}$$

where $\mathbf{F}(\mathbf{x}, \mathbf{y})$ is the feature vector $\mathbf{f}$ at coordinate (x, y) and $i, j = 1, \ldots, d$.

Define a $d \times d$ matrix as

$$\mathbf{E}_{x,y} = \begin{pmatrix} \mathbf{E}(x,y,1,1) & \ldots & \mathbf{E}(x,y,1,d) \\ & \vdots & \\ \mathbf{E}(x,y,d,1) & \ldots & \mathbf{E}(x,y,d,d) \end{pmatrix} \quad (10)$$

The computational complexity of constructing the integral images $\mathbf{E}_{x,y}, 1 < x < W, 1 < y < H$, is $O(HWd^2)$.

Let $R(x\prime, y\prime; x\prime\prime, y\prime\prime)$ be the window, where $(x\prime, y\prime)$ is the upper left coordinate and $(x\prime\prime, y\prime\prime)$ is the lower right coordinate, then the generalized structure tensor of the window bounded by $(1, 1)$ and $(x\prime, y\prime)$ is

$$\mathbf{T}_{R(1,1;x\prime,y\prime)} = \mathbf{E}_{x\prime,y\prime}, \quad (11)$$

and the generalized structure tensor of window $R(x\prime, y\prime; x\prime\prime, y\prime\prime)$ can be computed as:

$$\mathbf{T}_{R(x\prime,y\prime;x\prime\prime,y\prime\prime)} = \mathbf{E}_{x\prime\prime,y\prime\prime} + \mathbf{E}_{x\prime,y\prime} - \mathbf{E}_{x\prime\prime,y\prime} - \mathbf{E}_{x\prime,y\prime\prime} \quad (12)$$

After constructing the integral image, the generalized structure tensor can be calculated in a computational complexity of $O(d^2)$. So the total computational complexity is $O(HWd^2) + HWO(d^2) = O(HWd^2)$. Comparing it with the original computational complexity $O(HWd^2N^2)$, we can find that the computational cost is reduced a lot. It should also be noted that the computational cost is similar with those traditional similarity measures such as SAD, SSD, and NNC.

2.4 Properties

The proposed similarity measure owns good properties for stereo matching as follows.

First of all, the proposed similarity measure provides an effective way to fuse different features. For example, the feature vector $\mathbf{f}$ can also be defined as [12]:

$$\mathbf{f} = \begin{bmatrix} \mathbf{I} & |\mathbf{I}_x| & |\mathbf{I}_y| & \sqrt{\mathbf{I}_x^2 + \mathbf{I}_y^2} & |\mathbf{I}_{xx}| & |\mathbf{I}_{yy}| \end{bmatrix}, \quad (13)$$

where $\mathbf{I}_{xx}$ and $\mathbf{I}_{yy}$ are second-order partial derivatives of intensity with respect to x and y, and $\sqrt{\mathbf{I}_x^2 + \mathbf{I}_y^2}$ are the magnitude of the gradient. In our experiments, the feature vector is selected as $f = (\mathbf{I}, \mathbf{I}_x, \mathbf{I}_y)$ 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, but on 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.

3 Experimental Results

We perform the experiments on standard Middlebury dataset[1] provided by [7]. There are four benchmark image pairs, namely ”Tsukuba”, ”Venus”, ”Teddy” and ”Cones”.

The performance is evaluated using the quality measures provided in Middlebury dataset, including the percentage of ”bad” pixels for different regions: non-occluded region (nonocc), whole image (all) and pixels near discontinuities (disc).

To illustrate the performance of the proposed similarity measure, i.e. LE, we compare it with AD, SD, SAD, SSD, and NNC. We combine the similarity measures with local optimization algorithm, winner take all (WTA) [7], and the global optimization algorithm, belief propagation (BP) [11][10], respectively, to obtain 9 algorithms, SAD + WTA, SSD + WTA, NNC + WTA, LE + WTA, SAD + BP, SSD + BP, SD + BP, AD+ BP and LE + BP. The window size N for LE, SAD, SSD and NNC is set by the grid $\{5, 7, 9, 11\}$, and the parameter σ for LE is set by the grid $\{0.5, 1, 1.5\}$. The best results are selected.

Table.1 lists the qualitative results of the similarity measures combining with WTA. The first column contains the names of the algorithms. Table.2 presents the results of the similarity measures combining with BP. It is obvious that either combining with WTA or BP, the proposed similarity measure outperforms the other similarity measures such as SAD, SSD, NNC, AD and SD. It is worth noticing that the LE+ WTA is even better than SAD+BP and SSD+BP in some items of the evaluation measures.

In Figure.1, some results of LE+BP algorithm are provided.

4 Conclusions

In this paper, we propose a novel similarity measure under Log-Euclidean metric for stereo matching. A generalized structure tensor is applied to describe a point and the similarity is measured by the distance between the associated tensors under Log-Euclidean met-

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

Table 1. Qualitative results of the LE+WTA, SSD+WTA, SAD+WTA and NNC+WTA algorithms.

Algorithm	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
LE+WTA	7.22	8.94	22.5	6.56	8.06	37.0	18.3	25.2	35.9	13.4	21.8	27.0
SSD+WTA	9.95	11.9	31.8	6.80	8.14	41.8	20.2	26.9	40.0	13.2	21.3	32.9
SAD+WTA	8.94	10.9	26.0	7.46	9.04	42.0	20.6	27.7	36.7	15.0	23.4	29.9
NNC+WTA	14.3	16.3	41.1	7.59	8.53	40.6	46.8	50.5	68.3	55.8	57.4	68.0

Table 2. Qualitative results of the LE+BP, AD+BP, SD+BP, SAD+BP and SSD+BP algorithms.

Algorithm	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
LE+BP	1.16	2.77	5.51	1.96	3.59	24.7	9.88	19.0	18.9	6.50	15.4	14.2
AD+BP	2.67	4.59	12.0	3.91	5.56	39.4	12.3	20.2	21.9	15.2	23.6	24.0
SD+BP	5.28	7.35	11.8	3.44	5.05	26.6	17.3	24.5	23.4	54.0	57.7	55.9
SAD+BP	6.39	8.40	25.0	3.71	5.35	41.9	19.3	26.5	36.1	14.4	22.9	29.5
SSD+BP	8.57	10.5	31.7	4.25	5.57	41.4	19.8	26.3	39.8	13.1	20.6	32.9

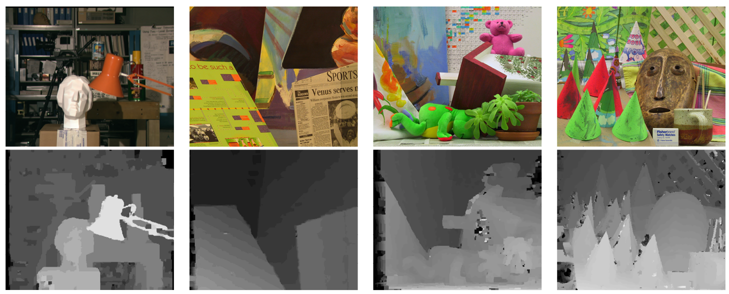

Figure 1. Results on the Middlebury dataset. The top row is original images, and the bottom row is disparity images acquired by LE + BP.

ric. We also present a fast calculation algorithm of the generalized structure tensor. Experimental results prove that the proposed similarity measure outperforms many traditional measures such as SSD, SAD and NCC.

References

[1] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the log-euclidean framework. In *MICCAI*, pages 115–122, 2005.

[2] M. J. Hannah. *Computer matching of areas in stereo images.* PhD thesis, Dept. of Computer Science, Stanford University, 1974.

[3] C. Harris and M. J. Stephens. A combined corner and edge detector. In *Alvey Conference*, pages 147–152, 1988.

[4] L. Hong and G. Chen. Segment-based stereo matching using graph cuts. In *CVPR (1)*, pages 74–81, 2004.

[5] A. Klaus, M. Sormann, and K. F. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *ICPR (3)*, pages 15–18, 2006.

[6] J. Malcolm, Y. Rathi, and A. Tannenbaum. A graph cut approach to image segmentation in tensor space. In *CVPR*, 2007.

[7] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002.

[8] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *CVPR*, pages 310–315, 1991.

[9] J. Sun, Y. Li, and S. B. Kang. Symmetric stereo matching for occlusion handling. In *CVPR (2)*, pages 399–406, 2005.

[10] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, M. T. V. Kolmogorov, A. Agarwala, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *ECCV*, pages volume 2, pages 16–29, 2006.

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

[12] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *ECCV (2)*, pages 589–600, 2006.

[13] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR (1)*, pages 511–518, 2001.

[14] H.-Y. Wang and K.-K. Ma. Accurate optical flow estimation using adaptive scale-space and 3D structure tensor. In *ICIP (2)*, pages 301–304, 2002.