Statistics-Based Matching Cost Function for Intensity Transformation

Phuc Nguyen Hong, Chang Wook Ahn*  

* Corresponding author. Tel.: +82312994677; email: cwan@skku.edu  
Manuscript submitted October 30, 2015; accepted April 18, 2016.  
doi: 10.17706/ijcce.2016.5.6.455-464

Abstract: Stereo matching is a challenging task due to stereo images being affected by many factors such as radiometric distortion, sun and rain flare, flying snow, occlusion and object boundaries. However, most of the existing stereo matching methods assume that corresponding pixels in left and right images have the same intensity; accordingly, the stereo matching methods use simple matching cost functions. As a result, they degrade their performance significantly when operating with real-world stereo images whose intensities of corresponding pixels can be arbitrarily transformed. State-of-the-art matching cost functions under conditions of distorted intensities between stereo images, such as the census transform, adaptive normalized cross-correlation, and the support local binary pattern perform with limited accuracy. In this paper, we propose a novel matching cost function based on a robust order statistic coefficient and segmentation that can operate accurately under various conditions of transformed intensities between stereo images. Based on the robust order statistic coefficient, the proposed matching cost function can tolerate local, monotonically nonlinear changes in intensities between the left and right images. By using the information of segmentation, the matching cost function operates accurately at object boundaries. The qualitative and quantitative experimental results obtained using stereo images in different datasets under various conditions show that our proposed matching cost function outperforms state-of-the-art matching cost functions in indoor and outdoor stereo images with various radiometric distortions.

Key words: Matching cost, order statistics, stereo matching.

1. Introduction

Stereo matching aims to find the depth for each pixel in the left image (e.g., using the left image as a reference image) of a stereo pair of images. It is still studied intensively because of its vital applications such as interactive robot navigation, self-driving cars, view interpolation, and three-dimensional reconstruction [1], [2]. According to Scharstein et al. [3], most of the stereo matching methods perform four steps: the matching cost computation, cost aggregation, disparity computation/optimization, and disparity refinement processes. Each step is important for achieving a high-quality disparity map, where the matching cost computation is a basic and important step that significantly affects the accuracy of the disparity map.

According to Hirschmuller and Scharstein [4], stereo images, even when captured under controlled environments, are inherently affected by radiometric distortion. However, the majority of existing stereo matching functions exploit the color consistent assumption and use simple matching cost functions [5]-[7].
so their performances are degraded severely in real stereo images whose intensities are arbitrarily changed [8].

Local stereo matching methods typically consist of matching cost computation, cost aggregation, disparity computation, and disparity refinement steps, such as [9]-[13]. Clearly, matching cost computation is compulsory for all stereo matching methods, so the performance of a stereo matching method heavily relies on the accuracy and robustness of the matching cost function used. A matching cost function computes a matching cost value for each pixel, \( p \), in the left image with a disparity hypothesis, \( d \), to create a matching cost image space, \( C \), in which \( C(d(p)) \) is a matching cost value for \( p \) and \( d \). From \( C \), we can obtain a disparity value for \( p \) as follows:

\[
D(p) = \arg \min_d \left( C_d(p) \right)
\]

(1)

where \( D \) is a disparity map.

The contributions of our work are as follows. First, we show that although the order statistics coefficient (OSC) is a good measure for computing the correlation between two sets of data points, it is not applicable to stereo matching as a matching cost function. This is due to the OSC becoming undefined when all data points have equal values. We propose an ROSC, which can overcome the weakness of the OSC while keeping the important properties of OSC. Second, we propose a matching cost function using segmentation and an ROSC, abbreviated SROSC. As a result, the proposed matching cost function addresses the fattening effect and radiometric distortion problems in stereo matching. Finally, we conduct experiments to evaluate the proposed matching cost function and compare it with state-of-the-art matching cost functions using stereo images captured with radiometric distortion from different datasets. Experimental results show that the proposed matching outperforms the state-of-the-art matching cost functions in different datasets.

1.1. OSC

The OSC, which was proposed in [14], can measure the monotonically non-linear association of two sets of data points using order statistics and the rearrangement inequality [15]. The OSC works on the principle that two sets of data points are highly correlated if large (small) values of one set are associated with small (large) values of another set.

Consider \( X = \{x_1, x_2, \ldots, x_m\} \) and \( Y = \{y_1, y_2, \ldots, y_m\} \), two time series of length \( m \). Forming the two time series pairwise results in a set, \( Z = \{z_1, z_2, \ldots, z_m\} \), where \( z_i = (x_i, y_i) \). Rearranging \( Z \) according to the values of \( x_i \), we get a new ordered set \( Z_1 = \{(x^1_i, y^1_i), \ldots, (x^2_i, y^2_i)\} \), where \( x^1_i \leq \ldots \leq x^2_i \) are called the order statistics of \( X \), and \( y^1_i, \ldots, y^2_i \) are the associated concomitants [16], [17]. On the other hand, rearranging \( Z \) according to the values of \( y_i \), we obtain a set, \( Z_2 = \{(x_i^1, y_i^1), (x_i^2, y_i^2)\} \), where \( y_i^1 \leq \ldots \leq y_i^2 \). The OSC is defined as follows:

\[
\text{OSC}(L, R) = \frac{\sum_{i=1}^{m} (l_i - l_{m+i+1}) r_i^0}{\sum_{i=1}^{m} (l_i - l_{m+i+1}) r_i^1}
\]

(2)

The value of the OSC is within \([-1, +1]\). A larger OSC value indicates that \( X \) and \( Y \) have a higher correlation. Conversely, when \( X \) and \( Y \) have a low degree of correlation, the smaller is the OSC value.
1.2. Motivation to Improve OSC

The OSC is highly sensitive to changes in association, tolerates monotone nonlinear transformations, and is robust to noise [14]. This makes the OSC suitable to use as a metric for measuring the image similarity. Given two image regions of the same size, set Z can be constructed from the two image regions and the OSC value can be computed using Z to measure the relationship between the image regions. An OSC value represents the degree of correlation between the two regions.

We denote \( d = [d0]^T \) for the disparity hypothesis, and \( p = [xy]^T \) for a pixel in the left image; the corresponding pixel in the right image is \( p = p - d \). Let \( L_{(p)} \) and \( R_{(p')} \) be two pixel windows of the same size, \( m \). For the sake of simplicity, we denote \( L \) and \( R \) by \( L_{(p)} \) and \( R_{(p')} \), respectively. For window-based matching cost functions in stereo matching, for each pixel \( p \) in the left image, a window of pixels, \( L_{(p)} \), centered at \( p \) is exploited to compute the matching value. Accordingly, the number, \( d \), of candidate windows, \( R_{(p')} \), in the right image is constructed and then each of the candidate windows is measured for the degree of image similarity. The OSC can be used to measure the degree of similarity between pairs of pixel windows (one window from the left image and one window from the set of candidate windows from the right image). A larger OSC value of a window pair indicates more correlation (image similarity) between them. An OSC-based matching cost function can tolerate local, monotonically nonlinear intensity transformation between pair of pixel windows. This is a very important property for a matching cost function because stereo images in real-world situations are affected by many factors and their intensities can be arbitrarily transformed such as globally/locally linear/non-linear intensity changes.

However, a problem preventing the OSC from being applicable to stereo matching is that the OSC is undefined for a pixel window whose pixels have equal intensities. Pixels from severely textureless image regions or from images captured under illumination and exposure configurations that are too large/small are likely have the same intensity. Furthermore, road-driving images often contain homogenous intensity regions such as walls, the sky, and sun flare-affected regions. To illustrate this, we use some images from the Middlebury and Kitti datasets, and compute their textureless map in which a pixel is considered textureless if the variance of its corresponding pixel window is zero (all pixels over its window have the same intensity).

1.3. ROSC

The idea of an ROSC is that we monotonically increase the value of each element in ordered sets \( Z_1 \) and \( Z_2 \) by its rank in \( Z_1 \) and \( Z_2 \). Similar to the OSC, an ROSC forms the two time series pairwise first, resulting in a set, \( Z_1 = \{z_1, z_2, \ldots, z_m\} \), and rearranges \( Z \) according to the values of and \( y_i \) to result in new ordered sets, \( Z_1 = \{(x_1, y_{i1}), \ldots, (x_m, y_{im})\} \) and \( Z_2 = \{(x_1, y_{i1}), \ldots, (x_m, y_{im})\} \), respectively. After this, increasing the value of each element in \( Z_1 \) obtains a new set, \( Z_1' = \{(x_1', y_{i1'}), \ldots, (x_m', y_{im'})\} \), where \( x_i' = \lambda(x_i) \) and \( y_i' = \lambda(y_i) \). Similarly, multiplying each element in \( Z_2 \) by its rank in ordered set \( Z_2 \) obtains a new set, \( Z_2' = \{(x_1', y_{i1'}), \ldots, (x_m', y_{im'})\} \), where \( y_i' = \lambda(y_i) \) is the rank of element \( x_i \) in the sorted set \( Z_1 \), and \( \lambda(x_i') \) and \( \lambda(y_i') \) are the ranks of elements \( y_i' \) and \( y_i' \) in the sorted set, \( Z_2 \). An ROSC is defined as follows:
Here, by exploiting the sorted sets, $Z_1$ and $Z_2$, the ranks of $x_i$ and $y_i$, can be computed as $x_i^1 = i$ and $y_i^1$. An ROSC has several properties as follows.

Property 1: An ROSC is within the interval [-1, 1].
Proof. According to (18), rearrangement follows.

$$\sum_{i=1}^{m} x_i^1 y_i^1 \leq \sum_{i=1}^{m} x_i^1 y_i^2 \leq \sum_{i=1}^{m} x_i^1 y_i^1$$

(4)

and

$$\sum_{i=1}^{m} x_{m-i+1}^1 y_i^1 \leq \sum_{i=1}^{m} x_{m-i+1}^1 y_i^2 \leq \sum_{i=1}^{m} x_{m-i+1}^1 y_i^1$$

(5)

Subtracting (4) by (5) and dividing the difference by $\sum_{i=1}^{m} (x_i^1 - x_{m-i+1}^1)y_i^1$. We have $-1 \leq \text{ROSC}(x, y) \leq 1$.

which completes the proof

Property 2:

$$\text{ROSC}(X, Y) = -1(+1)$$

when $X$ and $Y$ have a monotonic decreasing (increasing) relationship.

Proof. Assuming $y_i = \varphi(x_i)$ if $y_i = \varphi(.)$ is a decreasing function we have $y_i^2 = y_{m+i+1}^1$ for all $i$.

Substituting this into Eq.(3), We have $\text{ROSC}(X, Y) = -1$. Similarly, if $\varphi(.)$ is an increasing function, it results in $y_i^2 = y_i^1$ for all $i$ and eventually $\text{ROSC}(X, Y) = 1$.

1.4. SROSC
Performing robustly with stereo images captured under different conditions is crucial to a matching cost function due to real-world stereo images being affected by many factors. Like Census, an ROSC, when applied to stereo matching as a matching cost function, can tolerate local, monotonically nonlinear intensity transform between stereo images; therefore, an ROSC-based matching cost function has good potential to operate well in stereo images taken in various situations. From the inherent benefits of an ROSC, we deeply investigate an ROSC-based matching cost function in the context of stereo matching and propose a matching cost function that combines an ROSC and segmentation. In the proposed matching cost function, segmentation is exploited to use pixels that locates in the same segmented regions. To do this, we use a common assumption in stereo matching that states that pixels having similar pixel intensities are very likely to be in the same image structure, and have similar disparities. This assumption is exploited in [5] to reduce the fattening effect. Fig. 1 shows the flowchart of the proposed matching cost function. In our work, we use the mean-sift algorithm [16] to segment the left image of a stereo pair.

In window-based matching cost functions, a window size with radius \( r \) is required to determine a set of pixel neighbors for each pixel, \( p \). The coordinates of the given window, \( w \), are computed as follows.

\[
W = \left\{ (x_w, y_w) \mid -r \leq x_w \leq r \wedge -r \leq y_w \leq r \right\}
\]

Let \( I \) be the left and \( I^i \) the right images. In the proposed cost function, the coordinate of \( L_p \) for pixel \( p \) is determined as follows.

\[
L_p = \left\{ q \mid q \in I, x_q = x_w + x_p, y_q = y_w + y_p \right\}
\]

Similarly, the coordinate of \( R_p \) for pixel \( p' \) in the right image, \( I^i \) is determined as follows

\[
R_p = \left\{ q' \mid q' \in I^i, x_{q'} = x_w + x_p - d, y_{q'} = y_w + y_p \right\}
\]

In our work, we use the information of segmented image regions to determine if two pixels are in the same segmented region. Let \( N_p \) and \( N_{p'} \) be a set of pixel neighbors for \( p \) and \( p' \), respectively, and \( p \) be the index of a segmented image region in the segmented left image contains \( p \).

The two sets, \( N_p \) and \( N_{p'} \), are determined as follows:

\[
N_p = \left\{ q \mid q \in L_p, \delta\left(p, q\right) = 1 \right\}
\]

And

\[
N_{p'} = \left\{ q' \mid q' \in R_p, q \in N_p, q' - q = d \right\}
\]

where \( \delta\left(p, q\right) \) is the indicator function which determines if the two pixels, \( p \) and \( q \), are in a segmented image region. The indicator function is computed as follows:
\[
\delta(p, q) = \begin{cases} 
1 & \text{if } q = p \\
0 & \text{otherwise}
\end{cases}
\] (11)

Once \( L_p \) and \( R_p \) are determined, a matching cost value at pixel \( p \) in the left image and a disparity hypothesis, \( d \) is computed as follows:

\[
C_d(p) = \frac{1 - \text{ROSC}(L_p, R_p)}{2}
\] (12)

An SROSC matching cost value is in interval \([0, 1]\). A smaller value of an SROSC indicates a higher degree of similarity between \( N_p \) and \( N_p' \).

2. Experimental Results

We conducted experiments to investigate the proposed matching cost function using the Middlebury dataset and compared it with state-of-the-art matching cost functions including the ANCC, Census, ADCensus, and SLBP. We used the available source code of the ANCC in [17]. We set the window size to be the same (11×11) for the test matching cost functions. For the other parameters of the ANCC and ADCensus, we used the same values as the original papers, [5] and [18]. For SROSC, we used the mean-sift algorithm [19] to segment the left images of stereo pairs. The mean-sift segmentation parameters for the color bandwidth, spatial bandwidth, and minimum number size are 5, 4, and 20, respectively. All of the matching cost functions in this paper are evaluated using the average percentage of erroneous pixels in all zones, except occlusions, from disparity maps obtained using Eq. (1) and are computed at an error threshold of 1 pixel.

Fig. 2. The result of the test matching cost function using a winner-takes-all strategy with the Cloth2 stereo image pair with varying exposures. (a) Left image is for I2E1 (b) Right image is for I2E2. (c) Disparity map of Census (error of 10.4 %). (d) Disparity map of ANCC (error of 7.8 %). (e) Disparity map of ADCensus (error of 17.0 %). (f) Disparity map of SLBP (error of 10.3 %). (g) Disparity map of SROSC (error of 7.2 %). (h) Ground truth.
2.1. Middlebury Dataset

To measure the strengths of the test matching cost functions in stereo images with radiometrical distortion, we used the Middlebury dataset [20], which contains stereo pairs captured under different lighting and exposures. The illuminations are indexed as 1, 2 and 3, and the exposures are indexed as 0, 1 and 2 with the available disparity ground truth. For simplicity, we used [I2E1/I2E2] as the abbreviation for a pair of stereo images for which the left image is with illumination 2 and exposure 1 (I2E1) and the right image is with illumination 2 and exposure 2 (I2E2). We fixed the left image to I2E1, while the right image changed with the eight remaining different conditions: I1E0, I1E1, I1E2, I2E0, I2E2, I3E0, I3E1, and I3E2. Stereo images with different exposures are global intensity transformations, whereas those with different illuminations are local changes of intensities.

Fig. 2 shows the results of the test matching cost functions under different exposure settings. Fig. 2(a) shows left images with I2E1, and Fig. 2(b) shows right images with I2E2. The disparity maps of Census, the ANCC, AD Census, SLBP, and SROSC are shown in Fig. 2(c)-Fig. 2(g), respectively. Fig. 2(h) shows the ground truth of the left image.

The reason for this occurring in Census is that it encodes local structures and produces bit strings by comparing the intensity of an anchor pixel with that of its neighbors. Census encodes a pixel pair in a binary value that represents a relative order; this causes the Census encoding to be ambiguous in textureless image regions in which relative order of pixel pairs more likely change under radiometrical distortions. In cases in which anchor pixels are intensity-transformed, the encoding output of Census is significantly affected. The AD Census disparity maps are visually more erroneous than the Census ones due to stereo images with radiometric distortion; the AD had a negative effect in AD Census.

![Graphs showing quantitative comparisons of the test matching cost functions using a winner-takes-all strategy with Middlebury stereo image pairs with varying illumination and exposure configurations.](image-url)
The ANCC relies on the color of an anchor pixel to compute a weight for each pixel in a window. When the colors of anchor pixels are radiometrically distorted or different largely from the colors of their neighbors, this leads to weights being largely changed. In addition, the ANCC’s assumption of a Lambertian reflectance object likely does not hold under radiometric distortions between stereo images. The disparity maps produced by SLBP reduce errors more compared to Census due to its use of more constraints on the relative orders among pixels. In contrast, as a result of SROSC using more pixels and being sensitive to the degree of a relative order, the SROSC disparity maps look more accurate than the SLBP ones. Overall, the disparity maps produced by SROSC are more accurate than those of ADCensus, Census, the ANCC, and SLBP.

Fig. 3 shows the quantitative comparison of the test matching cost functions using the different Middlebury sub-datasets. In this figure, I1E0 denotes the right image with illumination 1 and exposure 0. Census and the ANCC are comparable in different sub-datasets. For each sub-dataset, the ANCC and ADCensus displayed more fluctuation with different illumination and exposure settings than Census, SLBP, and SROSC due to the ANCC exploiting pixel colors to compute weights and ADCensus incorporating the AD that uses the color-consistent assumption. Census, SLBP, and SROSC were more stable as a result of them being based on the relative orders of pixels for computing their matching values; therefore, they can operate accurately and the order of pixels is preserved. SROSC had the most robust and accurate performance in all of the test sub-datasets. In a sub-dataset, the test matching cost functions performed differently for different illumination and exposure configurations, and SROSC was generally superior to the other test matching cost functions.

2.2. Time Complexity

To measure the computation time of the proposed data cost, we used the Aloe images with a resolution of 427×370 and a disparity range of 70. We only measure the computation time required to compute its matching cost image space. Our test platform is a PC equipped with Intel core i7, a 4.00 GHz CPU, and 8.00 GB of memory. Fig. 4 shows the computation times required to compute the disparity image space of the proposed matching cost method with different window sizes. In Fig. 4, win5 indicates that SROSC uses a window size of 5×5. SROSC, with an 11×11 window size, took about 122 seconds to compute the matching cost image space, whereas SROSC, with a 5×5 window size, only needed about 21 seconds.

3. Conclusion

In this paper, we introduced a new matching cost function that operates robustly and accurately with stereo images that are in different indoor and outdoor scenes and situations. The proposed matching cost function exploited segmentation to improve the performance of pixels near object boundaries. We evaluated the proposed data cost and the state-of-the-art data costs using various stereo matching datasets,
and the experimental results show that the proposed matching cost method was superior to the state-of-the-art matching cost methods across the datasets.

A robust order statistics coefficient can tolerate a monotone transformation of data points. In this paper, a robust order statistics coefficient is applied in stereo matching as a matching cost function. However, a robust order statistics coefficient can be used for image similarity; therefore, applying it in other computer vision applications such as template matching, texture classification, and optical flow is our future work.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2015R1D1A1A02062017).

References


**Phuc Nguyen Hong** received the BS in science service from Auckland University of Technology (AUT), New Zealand in 2013. She is currently working toward the combine master and PhD degrees at the Computer Engineering Department, Sungkyunkwan University, Suwon, Republic of Korea.

Her research interests include genetic algorithms and computer vision.

**Change Wook Ahn** received the PhD degree at the Department of Information and Communications, Kwang-Ju Institute of Science and Technology, Kwang-ju, Republic of Korea.

From 2007 to 2008, he was a research professor in Gwangju Institute of Science & Technology (GIST). He is currently an assistant professor in Sungkyunkwan University (SKKU), Republic of Korea.

His research interests include genetic algorithms, multiobjective optimization, neural networks, and the applications of evolutionary techniques to wireless networks.