



A novel non-parametric transform stereo matching method based on mutual relationship

Xiaobo Lai¹ · Xiaomei Xu¹ · Lili Lv¹ · Zihe Huang¹ · Jinyan Zhang¹ · Peng Huang¹

Received: 29 October 2018 / Accepted: 21 December 2018 / Published online: 2 January 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

To cope with the problem of the vast majority local stereo matching approaches that rely highly on the statistical characteristics of the image intensity, a novel non-parametric transform stereo matching method based on mutual relationship is proposed. The traditional non-parametric transform is investigated, and its limitations are analyzed. In order to take the pixels' special location information into consideration during finding stereo correspondences, the original gray values of the neighborhood pixels whose relative position is one unit greater than that of the center pixel are replaced by the gray values interpolation of the four pixels surrounding it. Then the new non-parametric transform stereo matching is performed. The proposed approach is tested with both the standard image datasets and the images captured from realistic scenery. Experimental results are compared to those of intensity-based algorithms; the percentage of bad matching pixels is almost equivalent to the other examined algorithms, and the proposed algorithm exhibits robust behavior in realistic conditions.

Keywords Stereo matching · Non-parametric transform · Mutual relationship · Bilinear interpolation · Disparity map

Mathematics Subject Classification 05C70 Factorization, matching, covering and packing

1 Introduction

With the advancement of information technology, computer vision techniques have made possible the ability to automate the image analysis process [1–4]. Stereo vision is a reliable manner to obtain images and depth data for scenery at the same time

✉ Xiaobo Lai
dmia_lab@zemu.edu.cn

¹ College of Medical Technology, Zhejiang Chinese Medical University, Hangzhou 310053, China

[5], whose accuracy of the results depends on the choice of the stereo camera system and the stereo matching algorithm utilized [6]. This problem of extracting useful information from images and videos finds application in a variety of fields such as robotics, remote sensing, virtual reality, industrial automation, etc. Stereo matching is the key-problem and the hotspot in the field of stereo vision, whose results should be a smooth and dense disparity map. Currently, researchers have proposed a number of stereo matching algorithms [7], including global [8–10] and local methods [11]. In recent years, there is a compromise between the above two, which is the semi-global matching (SGM) [12]. SGM combines the efficiency of local methods with the accuracy of global methods by approximating a 2D MRF optimization problem with several 1D scanline optimizations, which can be solved efficiently via dynamic programming. Unfortunately, their performance badly decreased when they are applied in actual scene, showing as mismatch in low-texture region, invalidation in radiometric varying area, ‘stair-case’ in non-parallel plane and ‘foreground-dilation’ when occlusion occurs. The reason for all of these effects which seriously affects the performance of the algorithm is that these areas are ‘ill-conditioned’ areas which do not satisfy with the constraints in traditional algorithms. Therefore, it would be a great challenge to accurately recover the map of the scenery.

Currently, a large number of stereo matching algorithms use some kind of intensity-based metrics as the basis of their dissimilarity measure function [7]. The correctness of their results is based on the assumption that the same feature in the two stereo images, should have ideally the same intensity. However, this assumption is often not valid in realistic conditions. Even in the case that the gains of the two cameras are perfectly tuned, the fact that the two cameras shoot from a different pose might result in different intensities for the same point because of shading reasons. Furthermore, the realistic conditions such as the case for robotic applications, the illumination is not ideal [13], which leads to large variations of the intensity values for the same features among the two images of a stereo pair. Therefore, the non-ideal illumination causes failures during an intensity-based correspondence procedure. Zabih and Woodfill [14] introduced the non-parametric transform into stereo matching society and get a correct dense disparity image, because they are less susceptible to outliers, which happens near depth discontinuities. Thus, this transform improves the location of the object border in the depth image. The transform meets the requirements of much hardware for high real time performance, and has been wildly used in projects. However, it has limitations that the amount of information the non-parametric transform associated with a pixel is not very large, and the local measures rely highly on the intensity of the central pixel.

In this paper, to overcome these limitations, a novel illumination-invariant dissimilarity measure, the mutual relationship based non-parametric transform dissimilarity measure is proposed. The mean gray value of all the pixels in the transform window is taken as the gray value of the central pixel. Original gray values of the neighborhood pixels whose relative position is one unit greater than that of the matching pixel are replaced by the gray values through bilinear interpolation. After that, to verify the effectiveness of the proposed algorithm, the non-parametric transform and stereo matching are implemented with various image pairs.

2 Traditional non-parametric transform stereo matching

Traditional non-parametric transform is based on the gray values' order of the pixels in the transform window. The transformed matrix is not represent the gray values of pixels but the relationship between the neighborhood pixels' gray values and the center pixel' gray value. Non-parametric transform mainly includes the rank transform and census transform. We describe the census transform in detail, as it is used in the paper.

2.1 Census transform

Census transform takes the relative pixels' gray values as the similarity measure of stereo matching, the whole transformation don't need complicated operation such as multiplication and evolution, but just simple operation such as data comparison, accumulating and XOR, etc. Census non-parametric transform stereo matching just need three steps to be implemented.

Step 1 Select some pixel in the matching window as the center point, and pick a census rectangular transform window surrounding it. If the gray value of the neighborhood pixels in the transform window is less than the central pixel's gray value, set the value in corresponding position in the bit string as 1. Otherwise, set it as 0. The census measure $IC(i, j)$ can be expressed as a series of bits:

$$IC(i, j) = b_{n^2-1} \cdots b_k \cdots b_3 b_2 b_1 b_0 \tag{1}$$

where $k \in [0, \dots, n^2 - 1] / \left\lfloor \frac{n^2}{2} \right\rfloor$; $\lfloor k/n \rfloor$ means a integer division of k by n , $k \bmod n$ is a modulo n division.

The b_k parameter can be expressed as follows:

$$b_k = \begin{cases} 0, & I\left(i - \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{k}{n} \right\rfloor, j - \left\lfloor \frac{n}{2} \right\rfloor + k \bmod n\right) \geq I(i, j) \\ 1, & I\left(i - \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{k}{n} \right\rfloor, j - \left\lfloor \frac{n}{2} \right\rfloor + k \bmod n\right) < I(i, j) \end{cases} \tag{2}$$

where $I(i, j)$ denotes the intensity value for an image at a point at (i, j) in image coordinates. Bit-pixel numbering in a neighborhood depicted in Fig. 1a.

Step 2 Two census transformed vectors are compared using a similarity metric based on the Hamming distance, which is the number of bits that differ in the two-bit strings. The Hamming distance is summed over the triangular window, i.e.,

$$\sum_{(i,j) \in W} \text{Hamming}(I'_1(i, j), I'_2(i + d, j)) \tag{3}$$

where $I'_1(i, j)$ and $I'_2(i + d, j)$ represent the census transform of $I_1(i, j)$ and $I_2(i + d, j)$, respectively. d is the disparity searching range, $d_{\min} \leq d \leq d_{\max}$ (d_{\min} means the minimum disparity, d_{\max} means the maximum disparity). i and j denote the matching point coordinates in image coordinate.

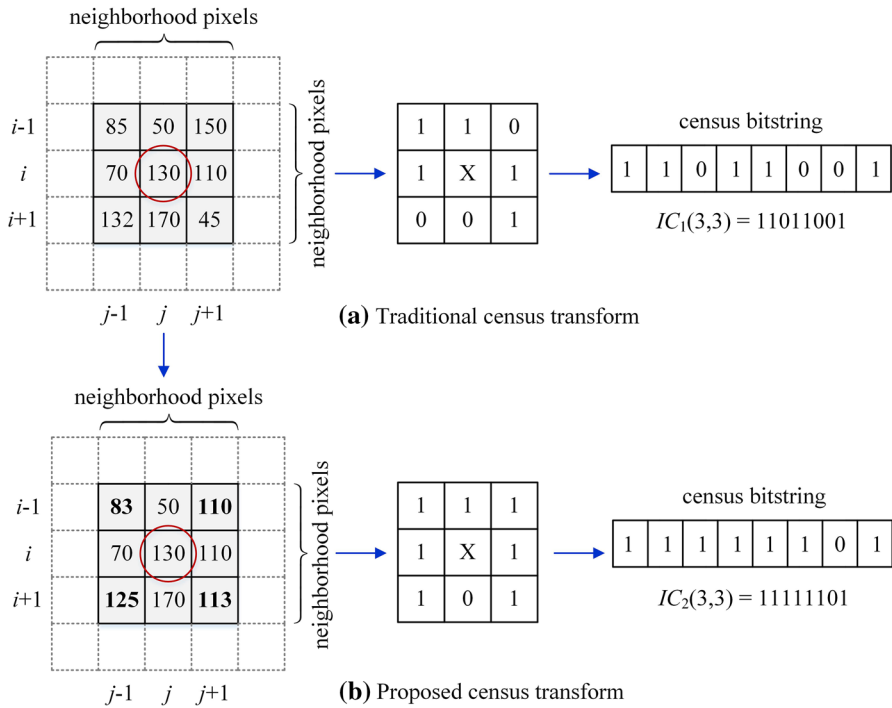


Fig. 1 Census transform

Step 3 In disparity searching range, compute the sum of Hamming distance of each pair of matching windows, locate the minimum Hamming distance, and the corresponding window's index is the disparity we try to locate.

2.2 Implementation of traditional non-parametric transform stereo matching

Although both the rank transform and the census transform are resistive to the radiometric gain and bias imbalances among images to be matched [14]. Nevertheless, the rank transformation loses information of spacial distribution of pixels since all the pixels surrounding a center one are encoded into a single value. Contrary to the rank transform, the census transform preserves the spacial distribution of pixels by encoding them in a bit string. Census transform is much more robust for blocking matching. Therefore, we choose the census transform in our approach.

For comparison of two bit-vectors of equal size, the Hamming distance can be used:

$$D_H(\mathbf{a}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^N \mathbf{a}_i \otimes \mathbf{b}_i \tag{4}$$

where \mathbf{a} and \mathbf{b} are the compared binary vectors of the same length N , \otimes denote concatenation.

Based on formula (4), it is obvious that D_H only accounts for bits that do not match.

The other measure, known also as the Tanimoto distance proposed by Sloan et al. [15], is defined as follows:

$$D_T(\mathbf{a}, \mathbf{b}) = \begin{cases} 1, & \mathbf{a} = \mathbf{b} = 0 \\ 1 - \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a} + \mathbf{b}^T \mathbf{b} - \mathbf{a}^T \mathbf{b}}, & \mathbf{a} \neq \mathbf{b} \neq 0 \end{cases} \tag{5}$$

where \mathbf{a} and \mathbf{b} are the binary vectors of the same length N .

A composition of formulae (4) and (5) is defined as follows:

$$D_{HT}(\mathbf{a}, \mathbf{b}) = D_H(\mathbf{a}, \mathbf{b}) \cdot D_T(\mathbf{a}, \mathbf{b}) \tag{6}$$

D_{HT} tends to equalize the opposing size effects of the Hamming and Tanimoto coefficients.

3 Proposed method

To some extent, traditional non-parametric transform can improve the performance of stereo matching near depth discontinuities. However, whether the rank transform or census transform, the transformation process is just the gray values' comparison between the neighborhood pixels and the central pixel. In this case, if the center pixel's gray value changes a lot affected by noises, the stability of stereo matching would be weakened. Therefore, to enhance the robustness of stereo matching, in this paper we proposed a non-parametric transform stereo matching algorithm based on mutual relationship. The overall block diagram of the stereo matching algorithm proposed in the paper is shown as Fig. 2.

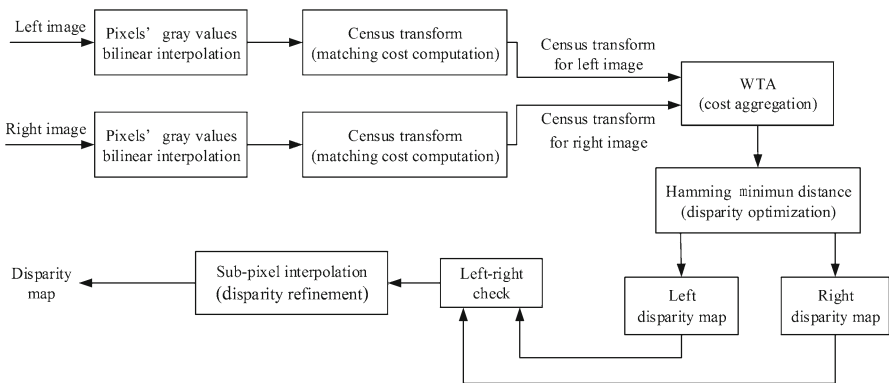


Fig. 2 Block diagram of the proposed stereo matching algorithm

3.1 Mutual relationship based non-parametric transform

Non-parametric transform based on mutual relationship can be achieved as the following steps:

- Step 1 Gray values of all pixels in the transform window are averaged, then the mean value replaced the center pixel' gray value
- Step 2 In order to take mutual relationship into consideration while finding stereo correspondences, original gray values of the neighborhood pixels whose relative position is one unit greater than that of the central pixel is replaced by the gray value through 2D bilinear interpolation. And non-parametric transform stereo matching is carried out with the gray values in the transform window

Denote the gray value of the pixel $P(x, y)$ as $I(x, y)$, a set of pixels in the rectangular transform window surrounding the central pixel $P(x, y)$ as $N(P)$, the number of the elements in $N(P)$ whose pixel gray value is less than $I(x, y)$ as $R(P)$. $\bar{I}(x, y)$ represents the average gray value of pixels in the transform window, as shown in formula (7).

$$\bar{I}(x, y) = \frac{1}{n \times n} \sum_{(i,j) \in N(P)} I(x+i, y+j) \tag{7}$$

Replace the gray value of the central pixel in the transform window with $\bar{I}(x, y)$. Then, for neighborhood pixels that are 1 relative position unit to the central pixel, their gray values remain invariant, for the neighborhood pixels of the rest areas; their gray values are obtained by bilinear interpolation. As shown in Fig. 3, the spatial mutual relationship between the neighborhood pixel p_1 and the central pixel p is represented by the gray value of pixel m_1 , which is at the same circle with pixel p_2 and pixel p_8 . Take the real time of stereo matching into consideration, we simplify the gray value of pixel m_1 to pixel m'_1 , then the gray value of pixel m'_1 can be computed by 2D bilinear interpolation with pixel p_1, p_2, p_8 and pixel p according to formula (8). Similarly, the spatial mutual relationship of pixel p_3, p_5, p_7 and pixel p also can be obtained by 2D gray value bilinear interpolation of the four pixels surrounding it, respectively.

$$I(i+u, j+v) = (1-u)(1-v)I(i, j) + (1-u)vI(i, j+1) + u(1-v)I(i+1, j) + uvI(i+1, j+1) \tag{8}$$

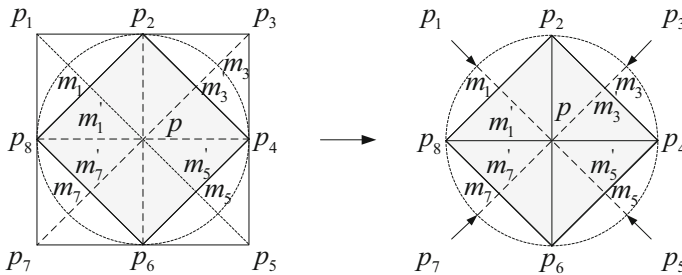


Fig. 3 Bilinear interpolation of gray value

where $I(i, j)$, $I(i + 1, j)$, $I(i, j + 1)$, and $I(i + 1, j + 1)$ is the gray value of pixels at (i, j) , $(i + 1, j)$, $(i, j + 1)$, and $(i + 1, j + 1)$, respectively.

Step 3 Replace the original gray value with a new one after bilinear interpolation, put formula (8) into formula (1), the mutual relationship based non-parametric transform stereo matching achieved. Corresponding bit-pixel numbering in a neighborhood depicts in Fig. 1b, we can see the difference between the traditional method and the proposed method

3.2 Implementation of the proposed stereo matching

As shown in Fig. 2, prior to stereo matching, both the left image and the right image have been rectified, details of the principle of the image rectification can refer to Hartley [16], here will not go to any further research. A stereo pair is acquired at the same time possibly with the same acquisition. The first module is to carry out the pixels' gray values bilinear interpolation with formula (8) before census transform. The matching cost computation is represented by the Hamming distance of the two transform windows after census transform. And the cost aggregation adopts the Winner-Take-All (WTA) method which is based on local areas to sum the Hamming distance of matching windows. Once the aggregated matching cost has been computed, we need to determine which disparities best represents the scene surface. The purpose of the disparity optimization is to pick the lowest aggregated matching cost as the selected disparity at each pixel. The left-right consistency check is performed to get rid of the half-occluded pixels in the initial disparity map. After that, the sub-pixel refinement of disparity aims to improve the quality of the disparity map. It optimizes the initial disparity map by sub-pixel interpolation and obtains an accurately matched disparity map that takes the left image as the reference image.

3.3 Evaluation of the proposed stereo matching

After we have gotten the accurately matched disparity map, we need to evaluate the performance of it. Two general approaches to this are to compute error statistics with respect to some ground truth data [17] and to evaluate the synthetic images obtained by warping the reference or unseen images by the computed disparity map [18]. We compute the following two quality measures based on known ground truth data:

1. Root-Mean-Squared (RMS) error (measured in disparity units) between the computed depth map $d_C(x, y)$ and the ground truth map $d_T(x, y)$, i.e.,

$$R = \left(\frac{1}{N} \sum_{(x,y)} \left| d_C(x, y) - d_T(x, y) \right|^2 \right)^{\frac{1}{2}} \quad (9)$$

where N is the total number of pixels. (x, y) is the coordinate of matching pixel.

2. Percentage of Bad Matching Pixels (PBM)

$$B = \left(\frac{1}{N} \sum_{(x,y)} |d_C(x, y) - d_T(x, y)| > \delta_d \right) \quad (10)$$

where δ_d is a disparity error tolerance. In our algorithm of experiments, we use $\delta_d = 1$.

In addition to compute these statistics over the whole image, we also focus on three different kinds of regions: textureless regions, occluded regions, depth discontinuity regions. These regions were selected to support the analysis of matching results in typical problem areas. The statistics described above are computed for each of the three regions and their complements, e.g.,

$$B_W = \frac{1}{N_W} \sum_{(x,y) \in W} (|d_C(x, y) - d_t(x, y)| < \delta_d) \quad (11)$$

4 Experiment and analysis

4.1 Test configuration

To qualify the effectiveness of our stereo algorithm, we use data sets of Tsukuba, Venus, Teddy and Cones from the Middlebury dataset [7], whose size is 384×288 , 434×383 , 450×375 and 450×375 , respectively. The four stereo pairs have been rectified, and the maximum values of disparity searching range are 16, 20, 60 and 60, respectively. RMS in formula (9) and PBM in formula (10) are applied to evaluate the performance of the algorithms. In our experiments, we use default experimental parameters of the test platform provide. Experimental results are obtained on HP PC with T2300 1.66 GHz, 512RAM. The matching area is chosen to be square of size 9×9 pixels and the size of transform window is 5×5 pixels during all experiments, so results can be compared with.

4.2 Analysis of experimental results

We use formulae (4), (5), and (6) to compute the Hamming distance of the two bit-vectors of equal size, respectively. Both the traditional census non-parametric transform stereo matching experiments and the information-based census non-parametric transform stereo matching experiments are conducted, the results fed back by Middlebury are illustrated in Table 1. Table 1 presents the percentage of pixels whose absolute disparity error is greater than 1 in the non-occluded regions (“ R_O ” column), in all images’ regions (“ R_A ” column), and in the regions near discontinuities (“ R_D ” column). These results show the performance of the proposed mutual relationship-based stereo matching and the traditional non-parametric census transform stereo matching when applied to standard image sets, captured under ideal lighting conditions. Ours 1, ours 2, and ours 3 are the results that the mutual relationship based census non-parametric transform stereo matching employs formulae (4), (6), and (5)

Table 1 Percentage of pixels whose absolute disparity error is greater than 1 using the proposed algorithm

Stereo matching algorithm	Tsukuba			Venus			Teddy			Cones		
	$R_{\bar{O}}$	R_A	R_D	$R_{\bar{O}}$	R_A	R_D	$R_{\bar{O}}$	R_A	R_D	$R_{\bar{O}}$	R_A	R_D
Ours 1	5.59	7.54	18.8	7.50	8.99	35.0	17.4	25.7	36.9	10.2	19.9	22.6
Ours 2	5.92	7.86	19.1	8.80	10.3	36.0	18.7	26.1	37.7	11.1	20.8	23.2
Ours 3	6.81	8.79	19.2	13.3	14.7	36.6	19.7	27.9	37.9	12.8	22.3	24.0
Tradition 1	6.51	8.43	19.7	10.5	12.0	36.7	20.1	28.3	38.2	10.5	21.5	23.9
Tradition 2	7.30	9.29	20.7	13.5	14.9	37.4	20.6	28.9	38.6	12.9	22.6	24.1
Tradition 3	8.97	11.0	21.7	17.9	19.3	39.9	21.3	29.3	41.3	13.5	23.0	25.3
CoopRegion	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01
AdaptingBP	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32
LCD + AdaptWq	5.98	7.84	22.2	14.5	15.4	35.9	20.8	27.3	38.3	8.90	17.2	20.0
Inflection	7.95	9.54	28.9	4.41	5.53	31.7	17.7	25.1	44.4	14.3	21.3	38.0

to compute the Hamming distance of the two bit-vectors, respectively. Tradition 1, tradition 2, and tradition 3 are the results that the traditional non-parametric census transform stereo matching employs formulae (4), (6), and (5) to compute the Hamming distance of the two bit-vectors, respectively. As we can see from Table 1, regardless of the mutual relationship based census non-parametric transform stereo matching or the traditional non-parametric transform stereo matching, the one uses formula (4) to compute the similarity measure achieves better results than that uses formulae (5) and (6). Therefore, we only choose the proposed algorithm to implement the experiments' comparison.

Figure 4 depicts the performance of our algorithm in detail. The first column and second column are standard image sets, from top to bottom: the Tsukuba, Venus,

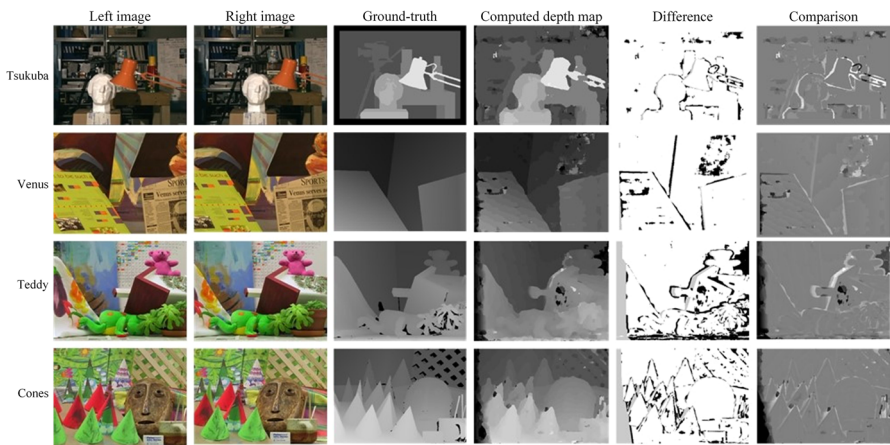


Fig. 4 Results of the proposed algorithm for Middlebury data sets

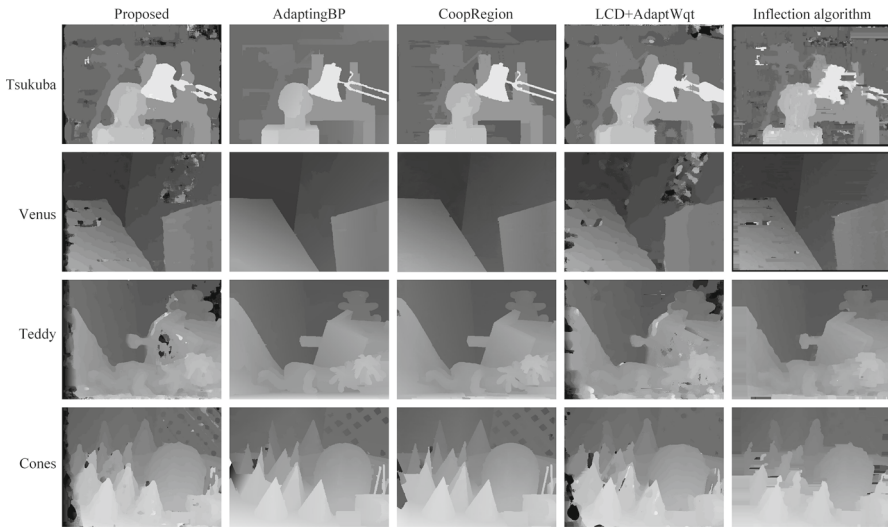


Fig. 5 Results comparison of stereo matching between our algorithm and several algorithms in Middlebury

Teddy, and Cones. The third column are ground-truth disparity maps provided by the Middlebury dataset. The 4th column are the computed depth maps with the proposed algorithm performed. Four images in the 5th column show the differences between the computed depth maps and the ground truth maps. The black points represent the bad matching pixels in the non-occluded areas of the images; the gray points represent the bad matching pixels in the occluded areas, and the white points in large portions represent the accurate matching pixels in whole images. The 6th column shows the results comparison between the accurate matching pixels (gray areas) and the bad matching pixels (the white and black areas) acquired by the proposed algorithm.

We compare the computed disparity maps' performance between our proposed algorithm and several typical algorithms in Middlebury, as shown in Fig. 5. From top to bottom: the Tsukuba, Venus, Teddy, and Cones. From left to right: our algorithm, AdaptingBP algorithm proposed by Claus et al. [9], CoopRegion algorithm proposed by Wang and Zheng [19], LCD + AdaptWqt algorithm proposed by Nalpantidis and Gasteratos [20], and inflection algorithm proposed by Olague et al. [21], respectively. Combined with Fig. 5 and Table 1, in terms of correct matching rates, it can be seen that our algorithm is not good as the CoopRegion algorithm (rank 1st in Middlebury) and the AdaptingBP algorithm (rank 2nd in Middlebury), but is better than the local area-based inflection algorithm and the LCD + AdaptWqt algorithm. In terms of matching efficiency or algorithm computation complexity, suppose the matching area is chosen to be square of size $W \times W$ pixels and the size of transform window is $C \times C$ pixels, the image size is $N \times N$ pixels, the disparity searching range is d . The computation complexity of our algorithm is mainly concentrated in the matching cost computation and the cost aggregation, so the directly computed complexity is $O(WCN^2d)$. Similarly, the computation complexity of the LCD + AdaptWqt algorithm and the inflection algorithm are $\propto O(W^2N^2d)$, the computation complexity of the CoopRe-

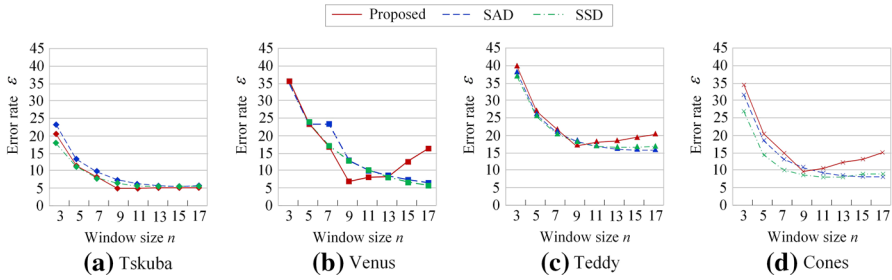


Fig. 6 Error rate of non-occluded regions with different window size

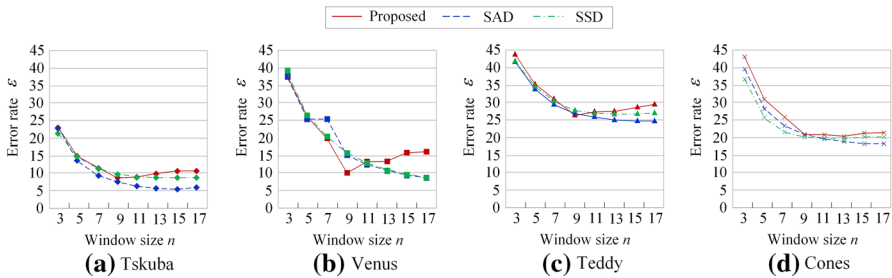


Fig. 7 Error rate of textureless regions with different window size

gion algorithm and the AdaptingBP algorithm also are $\propto O(W^2N^2d)$. However, it should be pointed out that the CoopRegion algorithm and the AdaptingBP algorithm need repeat iterations, which reduce the computation efficiency. In conclusion, the proposed algorithm is superior to those tested stereo matching algorithms in matching efficiency.

In addition, the WTA cost aggregation approach we implemented has parameters controlling the size of the support region to be used. Figures 6 and 7 plot the error rate of the disparity maps generated by different approaches with different settings, respectively, in which (a) is the curves of Tskuba obtained with three stereo matching algorithms, (b) is the curves of Venus obtained with three stereo matching algorithms, (c) is the curves of Teddy obtained with three stereo matching algorithms, (d) is the curves of Cones obtained with three stereo matching algorithms. Conclusions can be drawn as follows:

1. In general, the error rate is high when small window small support regions are used and starts to decrease as the support regions get larger. This is reasonable as small support regions cannot effectively suppress noises. However, increasing the support window size over a threshold will also increase the error rate. This is mainly because larger support windows introduce more mismatches at areas near depth discontinuity boundaries.
2. It is generally difficult to pick a good parameter for different similarity measure functions and different input dataset. For example, the best window size is 9×9 pixels for the Teddy dataset with ours and 13×13 pixels for the Cones one.

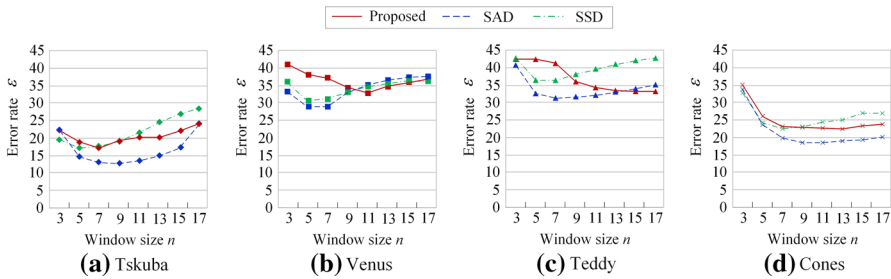


Fig. 8 Error rate of discontinuity region with different window size

However, the best window size is 13×13 pixels for the Teddy dataset with the SSD algorithm and 11×11 pixels for the Cones one.

- Figure 8 plots the error rates in the areas near depth discontinuities. As shown in the figure, the number of mismatches starts to increase after the support window size passes 9×9 . This suggests that, in areas that contain geometry details, changing the location of the window only is not good enough, especially for fixed-sized square-window approach.

To evaluate the performance of the proposed algorithm in real-life conditions, we captured a stereo pair of 320×240 with standard binocular stereo vision device, then stereo matching is achieved with the proposed algorithm. Like most vision algorithms, the results of stereo processing can contain errors. They are two major sources of these errors: lack of sufficient image texture for a good match, and ambiguity in matching when the correlation window straddles a depth boundary in the image. Areas with insufficient textures are rejected as low confidence, so the confidence filter is used to eliminate stereo matches that have a low probability of success because of lack of image texture. Each camera has a slightly different view of the scene, and at the boundaries of an object, there will be an area that can be viewed by one camera but not the other. Such occluded areas cause problems for stereo matching, and application of a uniqueness filter can eliminate these errors. Figure 9 shows a disparity image produced by the proposed algorithm. Higher disparities (closer objects) are indicated by white color. In the figure, the closest disparities are around 60, while the furthest are about 3. Figure 9a is one of a stereo pair after filtered. Figure 9b is the disparity from the proposed algorithm, without any post-processing. In Fig. 9c, textureless regions are rejected by texture filter and they appear black in the picture. Application of a uniqueness filter eliminates the errors in portions of the image with disparity discontinuities, as in Fig. 9d. As shown in Fig. 9, both the texture filter and the uniqueness filter are proved to be effective at eliminating bad matches.

The previous presented results are obtained for image pairs consisting of images with the same illumination each ideally. However, real-life conditions are not ideal and may lead to large variations of the intensity values for the same features among the two images of a stereo pair. To evaluate the stereo matching robustness of the proposed algorithm in real-life conditions, the right image of the real scene is manually processed with specialized software, and its luminosity is altered. The amount of alterations

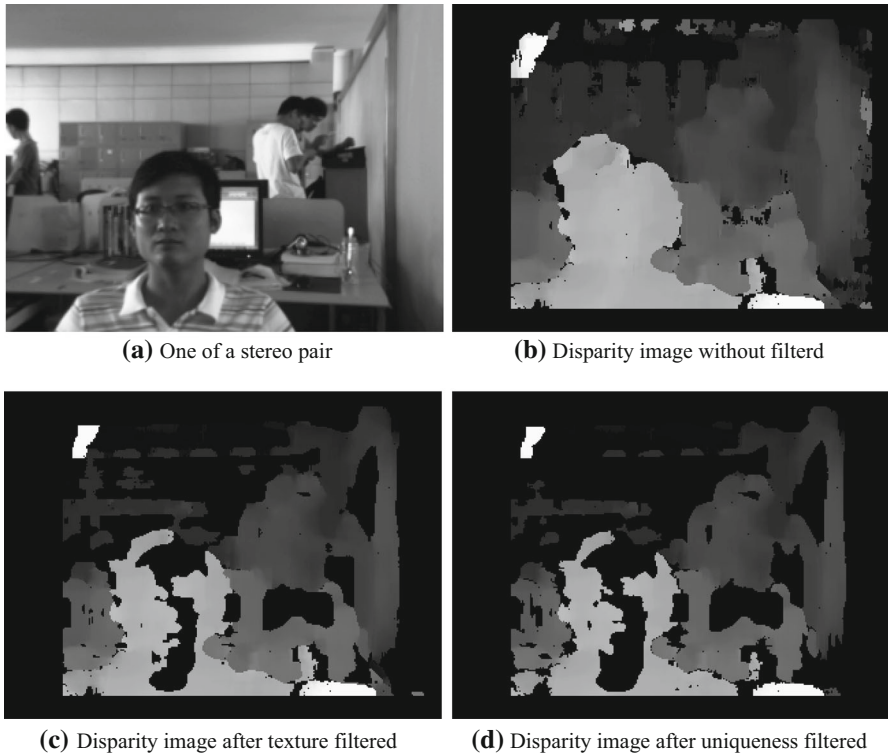


Fig. 9 Post-filters applied to the disparity map

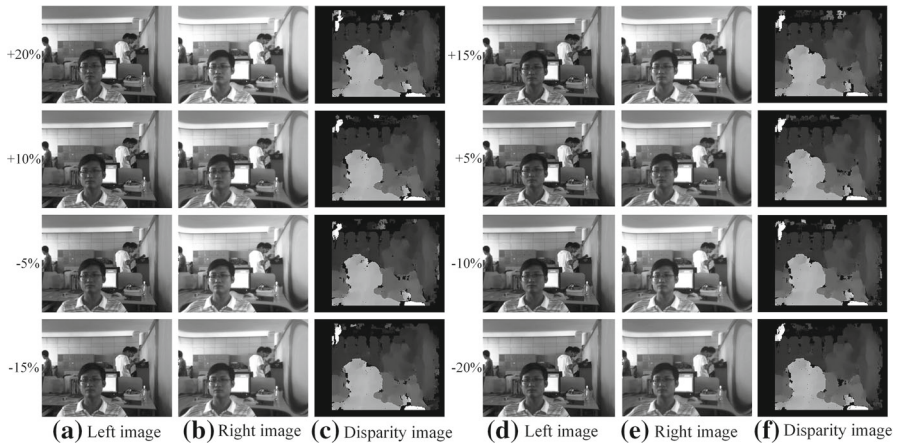


Fig. 10 Stereo matching of the proposed algorithm with right image luminosity altered

ranged from -20% to $+20\%$ with 5% increments. The input images as well as the results of the proposed method are shown in Fig. 10. The (c) column and the (f) column of Fig. 10 show the disparity images computed by the proposed algorithm. It can be

seen that the quality of the proposed algorithm's results almost remain practically the same for every tested lighting condition.

5 Conclusions

To overcome the limitations of traditional non-parametric transform stereo matching, a mutual relationship based non-parametric transform dissimilarity measure is proposed in the paper. The proposed algorithm is tested on various image pairs; experimental results are compared to those of area-based stereo algorithm. The error rates of the proposed algorithm are nearly the same as the tested algorithm. Moreover, in order to evaluate the performance of the proposed algorithm is able to confront lightness non-uniformities, stereo matching is implemented with the right image luminosity altered from -20% to 20% . Additionally, post filters are applied to eliminate mismatches caused by textureless regions and discontinuities. The proposed algorithm is able to deal with the problem caused by the luminosity non-uniformities in realistic conditions and at the same time preserves the details of the scene. It exhibits good, robust and reliable behavior.

Acknowledgements This work is funded in part by National Natural Science Foundation of China (Grant No. 61602419), and also supported by Natural Science Foundation of Zhejiang Province of China (Grant Nos. LY16F10008, LQ16F020003).

Compliance with ethical standards

Conflict of interest We declare that all authors have no conflicts of interest in the authorship or publication of this contribution.

References

1. Zhang J, Liu P, Zhang F, Song Q (2018) Cloudnet: ground-based cloud classification with deep convolutional neural network. *Geophys Res Lett* 45(16):8665–8672
2. Xu C, Xu L, Gao Z, Zhao S, Zhang H, Zhang Y, Du X, Zhao S, Ghista D, Liu H (2018) Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. *Med Image Anal* 50:82–94
3. Gao Z, Xiong H, Liu X, Zhang H, Ghista D, Wu W, Li S (2017) Robust estimation of carotid artery wall motion using the elasticity-based state-space approach. *Med Image Anal* 37:1–21
4. Gao Z, Li Y, Sun Y, Yang J, Xiong H, Zhang H, Liu X, Wu W, Liang D, Li S (2017) Motion tracking of the carotid artery wall from ultrasound image sequences: a nonlinear state-space approach. *IEEE Trans Med Imaging* 37(1):273–283
5. Pollefeys M, Nister D, Frahm J, Akbarzadeh A, Mordohai P, Clipp B, Engels C, Gallup D, Kim S, Merrel P, Salmi C, Sinha S, Talton B, Wang L, Yang Q, Stewenius H, Yang R, Welch G, Towles H (2008) Detailed real-time urban 3D reconstruction from video. *Int J Comput Vis* 78(2–3):143–167
6. Hu X, Mordohai P (2012) A quantitative evaluation of confidence measures for stereo vision. *IEEE Trans Pattern Anal Mach Intell* 34(11):2121–2133
7. Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comput Vis* 47(1/2/3):7–42
8. Zitnick C, Kang S (2007) Stereo for image-based rendering using image over-segmentation. *Int J Comput Vis* 75(1):49–65

9. Claus A, Sormann A, Kamer K (2006) Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: IEEE international conference on pattern recognition. IEEE, pp 15–18
10. Yang Q, Wang L, Ahuja N (2010) A constant-space belief propagation algorithm for stereo matching. In: IEEE International Conference on Computer Vision and Pattern Recognition, IEEE, pp 1458–1465
11. Birchfield S, Tomasi C (1998) A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Trans Pattern Anal Mach Intell* 20(4):401–406
12. Hirschmuller H (2008) Stereo processing by semiglobal matching and mutual relationship. *IEEE Trans Pattern Anal Mach Intell* 30(2):328–341
13. Clancar G, Kristan M, Karba R (2004) Wide-angle camera distortions and nonuniform illumination in mobile robot tracking. *Robot Auton Syst* 46(2):125–133
14. Zabih R, Woodfill J (1994) Non-parametric local transforms for computing visual correspondence. In: European conference on computer vision. Springer, pp 151–158
15. Sloan KR, Tanimoto SL (1979) Progressive refinement of raster images. *IEEE Trans Comput C-28*(11):871–874
16. Hartley RI (1999) Theory and practice of projective rectification. *Int J Comput Vis* 35(2):115–127
17. Barron JL, Fleet DJ, Beauchemin SS (1994) Performance of optical flow techniques. *Int J Comput Vis* 12(1):43–77
18. Szeliski R (1999) Prediction error as a quality metric for motion and stereo. In: IEEE international conference on computer vision. IEEE, pp 781–788
19. Wang Z, Zheng Z (2008) A region based stereo matching algorithm using cooperative optimization. In: IEEE conference on computer vision and pattern recognition. IEEE, pp 1–8
20. Nalpantidis L, Gasteratos A (2010) Stereo vision for robotic applications in the presence of non-ideal lighting conditions. *Image Vis Comput* 28(6):940–951
21. Olague G, Fernandez F, Pérez C, Lutton E (2006) The infection algorithm: an artificial epidemic approach for dense stereo correspondence. *Artif Life* 12(4):593–615