

RESEARCH ARTICLE

Stereo Matching by Filtering-Based Disparity Propagation

Xingzheng Wang^{1,2}, Yushi Tian^{1,2}, Haoqian Wang^{1,2*}, Yongbing Zhang^{1,2}

1 Shenzhen Key Laboratory of Broadband Network & Multimedia Graduate School at Shenzhen, Tsinghua University, Shenzhen, China, **2** Shenzhen Institute of Future Media Technology, Shenzhen, China

* wanghaoqian@tsinghua.edu.cn



OPEN ACCESS

Citation: Wang X, Tian Y, Wang H, Zhang Y (2016) Stereo Matching by Filtering-Based Disparity Propagation. PLoS ONE 11(9): e0162939. doi:10.1371/journal.pone.0162939

Editor: Quan Zou, Tianjin University, CHINA

Received: December 24, 2015

Accepted: August 31, 2016

Published: September 14, 2016

Copyright: © 2016 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work is partially supported by the National Natural Science Foundation of China (61471213, 61571259, 61531014), Shenzhen Fundamental Research fund (JCYJ20140509172959961, JCYJ20160331185006518), Natural Science Foundation of Guangdong Province (2015A030310173) and National High-tech R&D Program of China (863 Program, 2015AA015901).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Stereo matching is essential and fundamental in computer vision tasks. In this paper, a novel stereo matching algorithm based on disparity propagation using edge-aware filtering is proposed. By extracting disparity subsets for reliable points and customizing the cost volume, the initial disparity map is refined through filtering-based disparity propagation. Then, an edge-aware filter with low computational complexity is adopted to formulate the cost column, which makes the proposed method independent on the local window size. Experimental results demonstrate the effectiveness of the proposed scheme. Bad pixels in our output disparity map are considerably decreased. The proposed method greatly outperforms the adaptive support-weight approach and other conditional window-based local stereo matching algorithms.

Introduction

Stereo matching solves the correspondence problem between stereo image pairs, which for a long time has been one of the most fundamental and challenging computer vision tasks. As designed by Scharstein and Szeliski [1], and which is widely acknowledged by later researches, a four-step framework of Stereo matching (as Fig 1 shows): matching cost computation, cost aggregation, disparity computation and disparity refinement may generate the dense and two-frame stereo problem. Most existing algorithms, including local, global as well as semi-global ones, perform all or some of these four steps.

Local stereo matching algorithms usually perform cost computation and aggregation by a simple winner-take-all (WTA) strategy. In contrast, global approaches transform the problem to an energy-minimization model, which formulate a global optimization function composed of a data term and a smoothness term, and perform global disparity optimization [2–9] by dynamic programming (DP), graph cuts (GC) [10,11] or belief propagation (BP) [12–15]. Although high accuracy can be achieved for disparity estimation through global optimization, the large computation and time cost also limits its implementation in real-time applications. Moreover, segment-based approaches utilize plane fitting from the initial disparities on each segment, based on the assumption that disparities vary smoothly and continuously within each homogeneous color segment. The result can be further improved by a global optimization

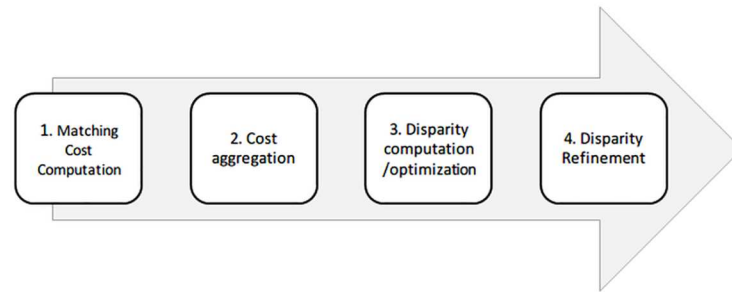


Fig 1. Four-step framework of Stereo matching.

doi:10.1371/journal.pone.0162939.g001

model to find optimal parameters of disparity plane, as the labeling space is relatively small. Recently the concept of semi-global stereo matching is also proposed based on a recognition stage that the support pixels for cost aggregation should be selected from the whole image and not restricted in a local matching window. Representatively, Hirschmuller introduced the mutual information to compute pixel wise matching cost and aggregate the costs along multiple paths that end in the current pixel [16]. Yang built a graph using all image pixels as nodes and compute the matching cost adaptively based on pixel similarity on a minimum spanning tree (MST) [17]. Matching cost aggregation plays a critical role to reduce the mismatching rate when using the per-pixel matching function. For most local stereo matching algorithms this is achieved by summing up or averaging the matching costs in a surrounding window centered by the current pixel. Using adaptive support-weight [18,19] for neighbor pixels will take edges or textures into account and bring better performance by adapting big data processing technologies [20–24].

In this paper a filtering-based stereo matching algorithm is proposed. Based on the perception that aggregating matching costs in a rectangular window is equivalent to filter the cost volume, we conducted the adaptive support-weight cost aggregation through filtering. Particularly we used the edge-aware guided filter which has linear time complexity with respect to the image size, it would reduce the computational complexity and greatly save the running time. Then the reliable and unreliable points were obtained through crosschecking of two rough-estimated disparity maps. Lastly, the high confidence disparity estimates are propagated from reliable points to unreliable ones by filtering a customized cost volume. A feedback-based optimization can also be achieved by integrating the initial disparity map into the guide image and re-execute the filtering progress. The main novelty of the proposed method is the utilize of reliable points, by propagating of these point, high performance of stereo matching would be achieved.

The rest of this paper is organized as follows. Section 2 explains the idea that matching cost aggregation is equivalent to the cost volume filtering. In Section 3 the concept of reliable point and disparity subset is first introduced, then we show how to customize a new cost volume to propagate the useful information from reliable points to unreliable ones by applying edge-aware filtering. Experimental results are demonstrated in Section 4, along with some analysis and evaluation. Finally, we conclude our paper and discuss the future work in Section 5.

Cost Aggregation by Cost Volume Filtering

The robust dissimilarity measure taking both SAD and gradient into account [13] is used as the pixel-wise matching cost function. Then a cost aggregation procedure which sums up costs in a window is usually implemented for local stereo matching. In this way, cost aggregation is

equivalent to applying filtering on the initial cost volume. Furthermore, the adaptive support-weight can be achieved by an edge-aware filter such as bilateral filter. For non-linear filtering each output pixel is calculated in $O(r^2)$ if the kernel size is $r * r$, thus the computational complexity shoots up as the kernel size increases. The so-called $O(1)$ or constant time bilateral filter [25,26], meaning the computational complexity is invariant to the kernel size, is designed for fast implementation.

In this paper we adopt the guided filter proposed by *He et al.* [27] to conduct the cost volume filtering. The guided filter is based on a local linear model, assuming that in a local window the filtering output q can be expressed as a linear transform of the guided image I :

$$q_i = a_k^T I_i + b_k, \forall i \in \omega_k \tag{1}$$

Where (a_k, b_k) are linear coefficients which are constant in ω_k . Considering the constraints from the filtering input p the linear coefficients can be derived as:

$$a_k = (\Sigma_k + \epsilon U)^{-1} \left(\frac{1}{|\omega_k|} \sum_{i \in \omega_k} I_i p_i - \mu_k \bar{p}_k \right) \tag{2}$$

$$b_k = \bar{p}_k - a_k^T \mu_k$$

Here, Σ_k is the $3 * 3$ covariance matrix of I in ω_k , and U is a $3 * 3$ identity matrix, μ_k is the mean of image I . Then compute the output q using:

$$q_i = \bar{a}_i^T I_i + \bar{b}_i \tag{3}$$

The guided filter is also edge-aware like bilateral filter, yet has better performance near edge locations. Another advantage of guided filtering is that the time complexity is only $O(1)$ for each pixel and $O(N)$ for an image of N pixels. In contrast, traditional edge-aware filters such as bilateral filter has $O(r^2)$ time complexity for each pixel when the local filter window has size of $r * r$, thus for the whole image it's $O(Nr^2)$ time. As the size of filter kernel increases, the time cost for such filters increases rapidly as well. Although the guided filter follows the local linear assumption, its computation complexity is unrelated to the local window size. This property makes it more practical for cost volume filtering.

When performing guided filtering on the cost volume C , the input image to be filtered is a certain slice at disparity candidate d in the cost volume, and the input color image of reference view is used as the guide image. An initial disparity map can be generated using the winner-take-all (WTA) strategy based on the filtered cost volume.

Disparity Propagation of Reliable Points

The left-right consistency check is widely used to verify the accuracy of the disparity estimation for each pixel. In this work pixels that pass the cross-check are marked as reliable points, and accordingly the others are unreliable ones. It is reasonable that the information of reliable points in the cost volume has higher confidence and should do favor to other pixels.

A. Building Disparity Subsets

Denote the full disparity range as D_{full} , a small subset $D_{sub}(p)$ is built for each reliable pixel p containing a couple of disparity candidates corresponding to the $|D_{sub}(p)|$ minimal matching costs in the cost volume. The size of the subset should be small as $|D_{sub}(p)| \ll |D|$. The extracted disparity values have the highest confidence among all candidates for the current pixel. Disparities out of this subset will be punished when filtering the cost volume. [Fig 2 \(a\)](#)

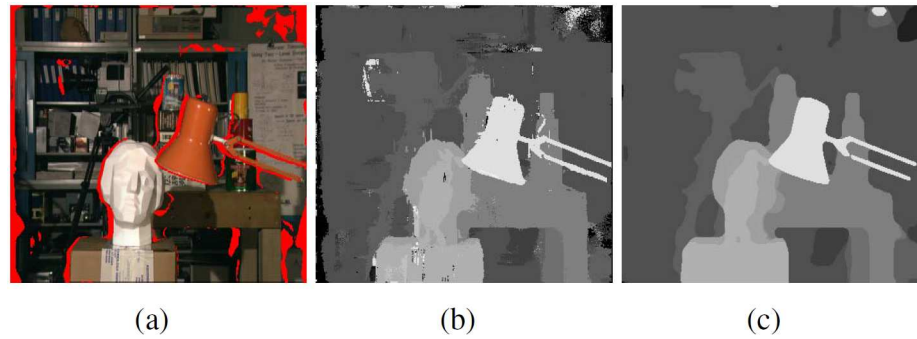


Fig 2. The “Tsukuba” test image. (a) Unreliable regions marked in red. (b) Initial disparity map. (c) Refined disparity map after filtering-based disparity propagation.

doi:10.1371/journal.pone.0162939.g002

shows the result of the cross-check based on the initial disparity map in (b). The unreliable pixels failing the cross-check are marked in red. Computing disparities in these red regions has to consult to the useful information from nearby reliable points.

B. Disparity Propagation Base on Customized Cost Volume

We customize a new cost volume based on the initial disparity map and the disparity subsets of reliable points:

$$C'_d(p) = \begin{cases} |d - D(p)| + \delta(p, d), & p \in P_{reliable} \\ 0, & else \end{cases} \quad (4)$$

Where $P_{reliable}$ is the set of reliable pixels. D is the dense disparity map, $\delta(p, d)$ is the penalty factor. The costs for unreliable points are set to 0 hence to eliminate the negative impact of wrong estimates, while for reliable points, disparities closer to the initial estimate $D(p)$ still have smaller values in the new cost volume. $\delta(p, d)$ is the penalty term to avoid being too far away from the disparity subset in the new estimation.

$$\delta(p, d) = 2 - \exp(-\lambda_1 \sum_{\substack{d_s \in D_{sub}(p) \\ |d - d_s| \leq 1}} |d - d_s|) - \exp(-\lambda_2 \sum_{\substack{d_s \in D_{sub}(p) \\ |d - d_s| > 1}} |d - d_s|) \quad (5)$$

Where $\lambda_1 < \lambda_2$ indicating that disparities far from the subset $D_{sub}(p)$ will lead to larger penalty. However, the penalty is limited as we use the exponential function.

Once again we perform the WTA optimization at each pixel and get a refined disparity map. A constant-time median filter is then applied as post-processing to fill holes and remove peaks.

C. Integrating Disparity into Guided Filtering

The RGB image of the reference view is used as the guide image when filtering the cost volume. Notice that before the new volume is generated an initial disparity map is available. The estimated disparity can feed back to the filtering procedure as an extra channel of the guide image, and bear a hand in filtering the customized cost volume. We integrate the disparity channel into the RGB reference image and form a new RGB-D guide image. For guided filtering this means simply replace the three-dimensional vector I_i in Eqs (2) and (4) with a four-

dimensional one. Then disparity propagation using the new guide image and the customized cost volume will lead to refinement and optimization of the disparity map.

Experimental Results

Our method is evaluated on the standard Middlebury benchmark. All the experiments run on a PC platform equipped with Intel Core i5 CPU and 4GB memory. The size of the disparity subset $D_{sub} = 2$, and parameters λ_1 and λ_2 are set to 0.04 and 1.2, respectively. Fig 3 shows the left view of the four test image pairs “Tsukuba”, “Venus”, “Teddy”, “Cones” and their corresponding ground truth disparity. The third column shows the disparity maps generated using the proposed algorithm. Bad pixels with absolute disparity error larger than 1.0 are marked out in the last column. It can be observed that the proposed algorithm recovers satisfactory disparity maps even for complicated scenes. Most of the bad pixels lie near edges where occlusion often occurs and is challenging for all stereo matching algorithms.

Table 1 gives the quantitative evaluation indicators of our results with error threshold = 1, along with results of some other representative algorithms such as adaptive support-weight approach and global optimization algorithms using Graph Cuts or Belief Propagation. Our method outperforms the others with respect to the average percent of bad pixels, especially for the “Cones” image pair on which our method is in the top 3 considering all 3 indicators including bad pixels in regions near discontinuities, non-occluded regions and all regions. For the other test image pairs our results still occupy the advanced level among all algorithms listed here. The proposed guided filtering-based disparity propagation outperforms the adaptive-weight approach for all test images except for “Tsukuba”. Moreover, our algorithm maintain the edge-aware property just as adaptive-weight yet has obvious advantage on computation complexity, as guided filtering is $O(1)$ time for each pixel and regardless of the filter kernel size. This makes it easier to use bigger local window size when handling images of large size without concern of sharp rise of the time cost.

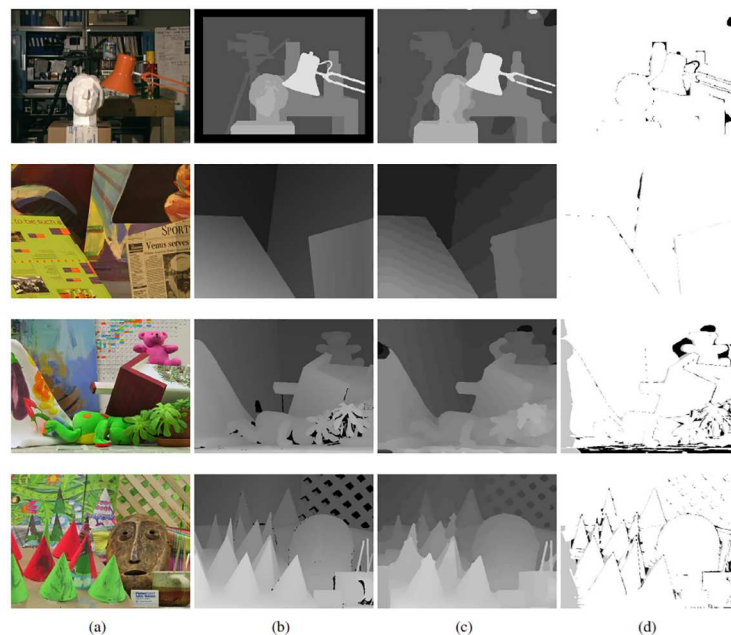


Fig 3. Results of the proposed algorithm. (a) Left view of the input image pair. (b) Ground truth disparity map. (c) Resulting disparity map using our method. (d) Bad pixels with error larger than 1.0.

doi:10.1371/journal.pone.0162939.g003

Table 1. Middlebury error rates of different algorithms (Error Threshold = 1).

Algorithm	Tsukuba			Venus			Teddy			Cones			Bad pixels(%)
	nonoc	all	disc	nonoc	all	disc	nonoc	all	disc	nonoc	all	disc	
Ours	1.73	2.14	7.54	0.33	0.75	4.5	7.39	13.1	17.9	2.57	8.52	7.56	6.17
CostAggr+occ	1.38	1.96	7.14	0.44	1.13	4.87	6.8	11.9	17.3	3.6	8.57	9.36	6.2
MVSegBP [7]	1.06	2.78	5.57	0.2	0.61	2.02	6.53	11.3	14.8	5.29	11.3	14.5	6.34
RandomVote	4.85	5.54	17.7	0.13	0.45	1.86	5.4	9.54	14.8	2.62	7.93	7.54	6.53
GradAdaptWgt [11]	2.26	2.63	8.99	0.99	1.39	4.92	8	13.1	18.6	2.61	7.67	7.43	6.55
AdaptWeight [10]	1.38	1.85	6.9	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26	6.67
EnhancedBP [6]	0.94	1.74	5.05	0.35	0.86	4.34	8.11	13.3	18.5	5.09	11.1	11	6.69
SemiGlob [8]	3.26	3.96	12.8	1	1.57	11.3	6.02	12.2	16.3	3.06	9.75	8.9	7.5
RealtimeBP	1.49	3.4	7.87	0.77	1.9	9	8.72	13.2	17.2	4.61	11.6	12.4	7.69
GC+occ [3]	1.19	2.01	6.24	1.64	2.19	6.75	11.2	17.4	19.8	5.36	12.4	13	8.26

doi:10.1371/journal.pone.0162939.t001

The running time of proposed method is as Table 2 shows, All the experiments run on a Laptop T450s equipped with Intel Core i5-5200U CPU and 4GB memory.

We also test various combinations of parameters including the size of disparity subsets as well as λ_1, λ_2 in the penalty term. Table 3 gives the average percent of bad pixels on the above 4 test image pairs when using various combinations of parameters, which is denoted in the form of $(|D_{sub}|, \lambda_1, \lambda_2)$. Our experience suggests that the size of disparity subsets $|D_{sub}|$ in the range [2,3], and [0.02,1.0], [0.8,2.0] for λ_1, λ_2 respectively can achieve approximately optimal results. The selection of parameters is based on the experimental results on the test images, as listed in Table 3. Empirically the size of disparity subsets is set to be small, for the initial disparity value of a reliable points is regarded to be of relatively high confidence, thus large penalty should be given to avoid large deviation from the subsets.

Conclusions

This paper proposes a stereo matching algorithm based on disparity propagation using edge-aware filtering. By extracting disparity subsets for reliable pixels and define a new cost volume

Table 2. Running time of proposed method.

Image	Running time (s)
cones	11.591
teddy	12.044
venus	2.618
tsukuba	4.570

doi:10.1371/journal.pone.0162939.t002

Table 3. Error rates for various parameters $(|D_{sub}|, \lambda_1, \lambda_2)$.

Parameters	Bad pixels(%)
(1,0.04,1.2)	6.25
(2,0.04,1.2)	6.17
(2,0.04,0.8)	6.18
(2,0.10,1.2)	6.18
(3,0.04,1.2)	6.31
(4,0.04,1.2)	6.35

doi:10.1371/journal.pone.0162939.t003

accordingly, mismatches in the initial disparity map are corrected through disparity propagation from nearby reliable points. The guided filtering is integrated to conduct the propagation in $O(1)$ time, which shows great advantage compared to traditional window-based cost aggregation methods. Future work will focus on how to customize a more reasonable cost volume which is essential to the disparity propagation. Occlusion handling and post processing of the disparity map also remain to improve.

Acknowledgments

This work is partially supported by the NSFC fund (61471213, 61571259, 61531014), Shenzhen Fundamental Research fund (JCYJ20140509172959961, JCYJ20160331185006518), Natural Science Foundation of Guangdong Province(2015A030310173) and National High-tech R&D Program of China (863 Program,2015AA015901)

Author Contributions

Conceptualization: XZW YST.

Formal analysis: XZW YST.

Investigation: XZW YST.

Methodology: XZW YST.

Project administration: HQW YBZ.

Resources: HQW YBZ.

Software: YST.

Supervision: HQW YBZ.

Validation: XZW YST.

Writing – original draft: XZW YST.

Writing – review & editing: XZW YST.

References

1. Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comput Vision*. 2002; 47(1–3):7–42.
2. Pan Z, Zhang Y, Kwong S. Efficient motion and disparity estimation optimization for low complexity multiview video coding. *IEEE Trans Broadcast*. 2015; 61(2):166–76.
3. Gu B, Sheng VS, Tay KY, Romano W, Li S. Incremental support vector learning for ordinal regression. *IEEE Transactions on Neural networks and learning systems*. 2015; 26(7):1403–16. doi: [10.1109/TNNLS.2014.2342533](https://doi.org/10.1109/TNNLS.2014.2342533) PMID: [25134094](https://pubmed.ncbi.nlm.nih.gov/25134094/)
4. Qian Y, Yao F, Jia S. Band selection for hyperspectral imagery using affinity propagation. *IET Comput Vision*. 2009; 3(4):213–22.
5. Su Z, Luo X, Deng Z, Liang Y, Ji Z. Edge-preserving texture suppression filter based on joint filtering schemes. *IEEE Trans Multimedia*. 2013; 15(3):535–48.
6. Liang-qun L, Wei-xin X, Zong-xiang L. Bearings-only maneuvering target tracking based on truncated quadrature Kalman filtering. *AEU-International Journal of Electronics and Communications*. 2015; 69(1):281–9.
7. Luo X, Ming Z, You Z, Li S, Xia Y, Leung H. Improving network topology-based protein interactome mapping via collaborative filtering. *Knowledge-Based Systems*. 2015; 90:23–32.
8. Huang Y, Chen X, Zhang J, Zeng D, Zhang D, Ding X. Single-trial ERPs denoising via collaborative filtering on ERPs images. *Neurocomputing*. 2015; 149:914–23.

9. Li Y, Liu W, Huang Q. Traffic anomaly detection based on image descriptor in videos. *Multimedia Tools and Applications*. 2016; 75(5):2487–505.
10. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Trans Pattern Anal Mach Intell*. 2001; 23(11):1222–39.
11. Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts. *Proceedings of International Conference on Computer Vision*. 2001;2:508–15.
12. Sun J, Zheng N-N, Shum H-Y. Stereo matching using belief propagation. *IEEE Trans Pattern Anal Mach Intell*. 2003; 25(7):787–800.
13. Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *Proceeding of 18th International Conference on Pattern Recognition (ICPR'06)*. 2006;3:15–8.
14. Larsen ES, Mordohai P, Pollefeys M, Fuchs H. Temporally consistent reconstruction from multiple video streams using enhanced belief propagation. *Proceeding of 11th International Conference on Computer Vision*. 2007:1–8.
15. Montserrat T, Civit J, Escoda OD, Landabaso J-L. Depth estimation based on multiview matching with depth/color segmentation and memory efficient belief propagation. *Proceeding of 16th IEEE International Conference on Image Processing (ICIP)*. 2009:2353–6.
16. Hirschmuller H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans Pattern Anal Mach Intell*. 2008; 30(2):328–41. PMID: [18084062](#)
17. Yang Q. A non-local cost aggregation method for stereo matching. *Proceeding of Computer Vision and Pattern Recognition (CVPR 2012)*. 2012:1402–9.
18. Yoon K-J, Kweon IS. Adaptive support-weight approach for correspondence search. *IEEE Trans Pattern Anal Mach Intell*. 2006; 28(4):650–6. PMID: [16566513](#)
19. De-Maeztu L, Villanueva A, Cabeza R. Stereo matching using gradient similarity and locally adaptive support-weight. *Pattern Recognit Lett*. 2011; 32(13):1643–51.
20. Zou Q, Hu Q, Guo M, Wang G. HAlign: Fast multiple similar DNA/RNA sequence alignment based on the centre star strategy. *Bioinformatics*. 2015; 31(15):2475–81. doi: [10.1093/bioinformatics/btv177](#) PMID: [25812743](#)
21. Zeng X, Zhang X, Zou Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Briefings Bioinf*. 2016; 17(2):193–203.
22. Xia Z, Wang X, Sun X, Wang B. Steganalysis of least significant bit matching using multi-order differences. *Security and Communication Networks*. 2014; 7(8):1283–91.
23. Ren Y-J, Shen J, Wang J, Han J, Lee S-Y. Mutual verifiable provable data auditing in public cloud storage. *Journal of Internet Technology*. 2015; 16(2):317–23.
24. Liu Y, Zeng X, He Z, Zou Q. Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans Comput Biol Bioinf*. 2016.
25. Porikli F. Constant time $O(1)$ bilateral filtering. *Proceeding of Computer Vision and Pattern Recognition (CVPR 2008)*. 2008:1–8.
26. Yang Q, Tan K-H, Ahuja N. Real-time $O(1)$ bilateral filtering. *Proceeding of Computer Vision and Pattern Recognition (CVPR 2009)*. 2009:557–64.
27. He K, Sun J, Tang X. Guided image filtering. *IEEE Trans Pattern Anal Mach Intell*. 2013; 35(6):1397–409. doi: [10.1109/TPAMI.2012.213](#) PMID: [23599054](#)