

Performance Evaluation of Multiresolution Methods in Disparity Estimation

Dibyendu Mukherjee, Gaurav Bhatnagar, and Q.M. Jonathan Wu

Department of Electrical and Computer Engineering
University of Windsor, Windsor, ON, Canada N9B 3P4
mukherjd@uwindsor.ca, goravb@uwindsor.ca, jwu@uwindsor.ca

Abstract. Disparity estimation from stereo imagery has gained substantial interest of research community from its commencement with the recent trend being the use of multiresolution methods. Existing multiresolution based methods are relatively independent and do not, in general, relate to a continuous progress in the research. As a result, the relative advantages and disadvantages of a particular multiresolution method in disparity estimation are hard to understand. Present work is an effort to put different multiresolution methods together to highlight their expediency and suitability along with the comparison to get a better understanding. Three different frameworks are used having different strengths and limitations followed by the comparison in the terms of time complexity, quality of matching and effect of different levels of decomposition. Qualitative and quantitative results have been provided for four types of standard multiresolution methods.

1 Introduction

Disparity estimation between stereo images is a fundamental task in computer vision to determine the dense depth map. Recently, searching disparity in multiple resolutions has been an interesting topic of research. But, an extensive study would reveal that less research has been carried out to compare different multiresolution methods and how detail sub-bands contribute to enhance the performance. The multiresolution based methods bridge a gap between the global and local methods in terms of speed and accuracy. They are faster than global methods due to the reduction in searching area whereas more accurate when compared to the local methods due to higher number of matching features and refinement of disparity in different resolutions. A lot of research have been reported on wavelet based stereo estimation [3,7,11]. Recently, Bhatti and Nahavandi [1] have proposed an effective way of stereo matching based on multi-wavelets that actually involves different detail subbands. Further, Mukherjee *et al.* [8] have proposed the use of curvelets associated with support weights to enhance the results. Ding *et al.* [4] have proposed the use of shift invariant contourlet transform in stereo. However, these methods are apparently less popular due to the lack of proper exploration of the performances and comparison among them.

The goal of present work is to provide a comparative study of different multiresolution methods for stereo disparity estimation with relative performances. The primary reason for such a study is to highlight the progress in a relatively less popular domain

and motivate the researchers to make more advancements. Also, it allows to analyze each algorithm and improve their performances. Finally, it highlights the limitations of a method with respect to the others and facilitate the choice of an appropriate multiresolution method for specific needs.

The paper briefly discusses the use of Wavelet Transform, Curvelet Transform, Multiresolution Singular Value Decomposition and Contourlet Transform in disparity estimation and their comparison on well-known Middlebury stereo datasets. For comparison, three frameworks have been used viz. fixed window based SAD, segmented window based SAD and approximate band based disparity estimation. To the best of our knowledge, this work is the first attempt to explore and evaluate the performance of different multiresolution methods in disparity estimation by above mentioned frameworks.

The paper is organized as follows. Section 2 provides an overview of different multiresolution methods. Section 3 describes the frameworks used for evaluation and comparison of their relative performances in disparity estimation. Section 4 gives the experimental results and comparison tables. Finally, the concluding remarks are depicted in Section 5.

2 Multiresolution Methods

This section provides a brief overview of the different multiresolution methods used in the paper.

2.1 Wavelet Transform

Wavelets are time limited functions that are localized in frequency [7]. For their transient nature, they are fundamentally different from the functions used for Fourier Transform. Wavelet transform of a signal constitutes of an approximate and a detail part. Approximate part is of low frequency and detail part consists of the high frequency components. Wavelet transform of an image decomposes it into an approximate sub-band and a detail part consisting of horizontal, vertical and diagonal subbands. For the implementation, Daubechies wavelets have been used.

2.2 Curvelet Transform

Curvelets are new multiscale representations that are better suited to objects with smooth curvatures [2]. They were designed to represent edges and singularities along curves. Roughly, to represent an edge to square error $1/N$ requires $1/N$ wavelets but only $1/\sqrt{N}$ curvelets. Unlike wavelets, curvelet transform (CT) can measure the information of a signal at specified scales, positions along specified orientations. The orientation flexibility makes it a powerful feature descriptor for disparity estimation. For the implementation, *Curvelets via Wrapping* [2] has been used with the codes provided by the Curvelab.

2.3 Multiresolution Singular Value Decomposition

Multiresolution SVD (MRSVD) is a new addition to the family of multiresolution methods. In 2001, Kakarala and Ogunbona [6] generalized the SVD to MRSVD. The idea is

to replace the low and high pass filters in decomposition by SVD. The image is subdivided in blocks of size $p \times q$ and a feature vector matrix R_l from the blocks is generated. Eigen vector matrix U_l is obtained by Singular value decomposition of $\bar{R}_l \bar{R}_l^t$ where, \bar{R}_l is the zero-mean representation of R_l . Finally, the rows of the transform domain representation $U_l^t \bar{R}_l$ yield the different subbands. The number of subbands depend on the size of block size $p \times q$. For our implementation, block size is kept 2×2 to get subbands equivalent to WT.

2.4 Contourlet Transform

A key feature in multidimensional signal being the directional features, Contourlet Transform (CONT) decomposes an image into several directional subband by combining the Laplacian pyramid with a directional filter at each scale [5]. Due to the cascade structuring, the multi-scale and directional stages are independent of each other and each scale can be separately decomposed into any number of directions provided the number of directions is a power of 2. Contourlets have a similarity to curvelets in terms of their directionality and parabolic scaling nature. For implementation, the code provided in Contourlet Toolbox by Minh. Do. is used with '9-7' filter for pyramidal decomposition stage and 'pkva' filter for direction decomposition stage.

3 Comparison Framework

In order to evaluate performances, three different frameworks are used for a number of reasons. A comparison among methods does not only rely on the performance of the methods but also on the comparison framework used. Using three different frameworks, a better knowledge on the performance of the methods can be obtained. Also, the results in Section 4 show that by using better cost functions, the results can be greatly improved. In Section 3.1, the support window is kept as fixed size and include all subbands of multiresolution method. Next, segmented window based cost function is used for more accurate disparity estimation in Section 3.2. Finally, in Section 3.3, the classical approximate band based disparity estimation is used to highlight the effects of simple multiple resolution searching. In all the cases, SAD is used as out cost function. Further, in order to keep an equivalent level of WT and MRSVD decomposition, single level is used in sub-sections 3.1 and 3.2 whereas 2-level is used in sub-section 3.3. As per as CT and CONT are concern, one cannot use level of decomposition because these are decomposed into scale and orientation. Hence, scale 2 and orientation 8 are used for sub-sections 3.1 and 3.2 whereas scale 3 and orientation 8 is used for sub-section 3.3.

3.1 Fixed Window SAD

Fixed Window SAD (fixWin) uses SAD as cost function and refines disparity in multiple resolution and subbands over a fixed size window in the following algorithm -

1. Perform multiresolution analysis of the gray scale stereo pairs.
2. Compute the correspondence match in the approximate image using SAD.
3. Shift to the subband.

4. Compute $scale_factor = (size\ of\ current\ subband\ image)/(size\ of\ previous\ subband\ image)$.
5. For each pixel, divide the coordinate of the pixel by $scale_factor$ to get the coordinates in previous scale. Then, compute the $initial_curr_disparity = scale_factor * disparity\ in\ previous\ scale$.
6. Take a range of search around the $initial_curr_disparity$.
7. Find the best match using SAD.
8. If this is the last subband, go to step 10.
9. Shift to the next subband. Go to step 5.
10. Take a range of search in final image based on the refined disparity map recovered after step 9.
11. Find the final disparity map using SAD.

For WT and MRSVD, only four subbands are obtained while CT and CONT yield a number of subbands in different scales and orientations. In the algorithm, it is assumed that if current subband is at scale s and orientation o , then next subband will be at scale s and orientation $o + 1$, if orientation $o + 1$ exists, else it will be the first orientation in next scale. The algorithm implicitly mentions two parameters - the search range for refinement and the fixed support window size. To keep consistency in quality of estimated disparity, both are kept constant for each comparison.

3.2 Segmented Window SAD

The major disadvantage of fixed window is fattening effect in depth discontinuities. As all points in a support window are considered to be of similar depth, the regions with discontinuity in depth are poorly matched. If the pixels in the window are kept to same region, the results can be improved. Based on This idea, this section describes a segmented support window (segWin) that works particularly better in these regions and can largely remove fattening effect. For this method, color images are used to produce multiresolution output of each color band separately. Each color band yields a set of multiresolution subbands. In other words, each subband has three gray value components each corresponding to a color. These “tricolor” subbands are used to segment the image into different regions by simple thresholding. The pseudo code below shows the procedure for the same.

1. $SAD = 0$;
2. For i IN $-W$ TO W
3. For j IN $-W$ TO W
4. IF $dist(sBL(x + i, y + j, :), sbL(x, y, :)) < thresh$
5. $SAD = SAD + (|sBL(x + i, y + j, :) - sBR(x + i + d, y + j, :)|)$
6. END IF
7. END FOR
8. END FOR

Here, $sBL(\cdot)$ and $sBR(\cdot)$ denote the “tricolor” subbands with $sBL(x, y, :)$ denoting three components of pixel (x, y) , $dist$ is a function that returns the euclidean distance

between the two component vectors of two pixels, d denotes the disparity, $(2 * W + 1)$ is the support window size and $thresh$ is the threshold for segmentation. Only the pixels that meet the criteria in IF condition are chosen for SAD computation.

The extra parameter involved in this method is the threshold value. This value needs to be kept at minimum range to remove the interregional interference.

3.3 Approximate Band Based Matching

The study is concluded with classical approximate band searching that ignores detail subbands and searches in approximate bands progressively till it reaches the original image resolution. The search region is projected into higher resolutions and searching is limited only to this region. This work is an effort to highlight the effect of low filtered approximate band for stereo matching. The method (segWinApprox) uses segWin and the comparison has been done in single and double level of decomposition.

4 Experimental Results

The well-known Middlebury datasets Tsukuba [9] and Cones [10] are used for generating comparative results. The window size has been kept 9×9 in original scale, 7×7 for level 1 of decomposition for fixWin and 33×33 in original scale and all subbands for segWin and segWinApprox. In Table 1, the comparison of different methods is provided for fixWin. The results are also compared with original SAD (origSAD) result in terms of percentage of bad pixels. The percentage is calculated based on the following formula -

$$\frac{1}{N} \sum_{x,y} (|d_c(x,y) - d_t(x,y)|) > \delta_{thresh} \tag{1}$$

where, N is the total number of pixels in the image, d_c and d_t represent the computed and ground truth disparity maps respectively, and δ_{thresh} is the threshold for bad pixels (usually equal to 1.0).

In Table 2, the execution times of different methods are compared for fixWin in terms of seconds. For time complexity, it is important to mention that all methods have been run in a Pentium 4, 2 GHz machine with MATLAB. It is evident from the results that multiresolution methods provide comparative results in faster computations. The qualitative results can be seen from Figure 1.

Table 3 shows the comparison among different multiresolution methods based for segWin. The results, as can be seen from Figure 2, are much better than fixWin and are quite immune to the fattening effect. As segwin is much slower compared to fixed

Table 1. Comparisons of methods based on fixed window SAD

Methods	WT	CT	MRSVD	CONT	OrigSAD
Images	% of bad pixels				
Tsukuba	10.44	9.38	10.53	10.57	11.82
Cones	21.79	20.81	20.41	21.69	21.55

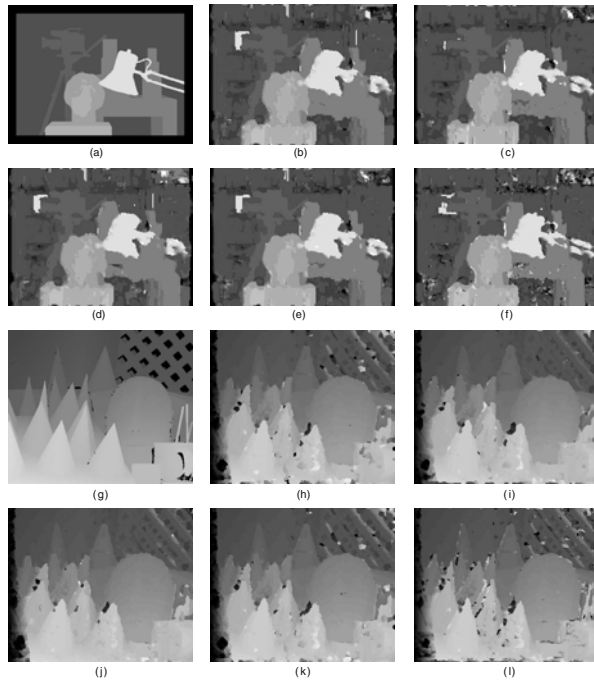


Fig. 1. Disparity maps generated using Fixed Window SAD. (a),(g) groundtruth, (b),(h) WT, (c),(i) CT, (d),(j) MRSVD, (e),(k) CONT and (f),(l) OrigSAD.

Table 2. Comparison execution time of different methods based on fixed window SAD

Images \ Methods	WT	CT	MRSVD	CONT	OrigSAD
Tsukuba	13.9313	19.3346	14.1931	17.3685	25.0329
Cones	28.2972	44.7665	28.5641	28.7104	130.3596

Table 3. Comparison of methods based on segmented window SAD

Images \ Methods	WT	CT	MRSVD	CONT
	% of bad pixels			
Tsukuba	6.23	5.72	6.35	6.63
Cones	19.31	17.69	17.64	19.68

Table 4. Comparisons of methods based on approximate band matching in multiple resolution

Images \ Methods	WT		CT		MRSVD		CONT	
	% of bad pixels							
	L1	L2	L1	L2	L1	L2	L1	L2
Tsukuba	8.73	9	8.42	8.33	9.2	9.4	8.78	8.85
Cones	23.59	27.27	22.72	25.4	21.24	22.61	23.51	27.54

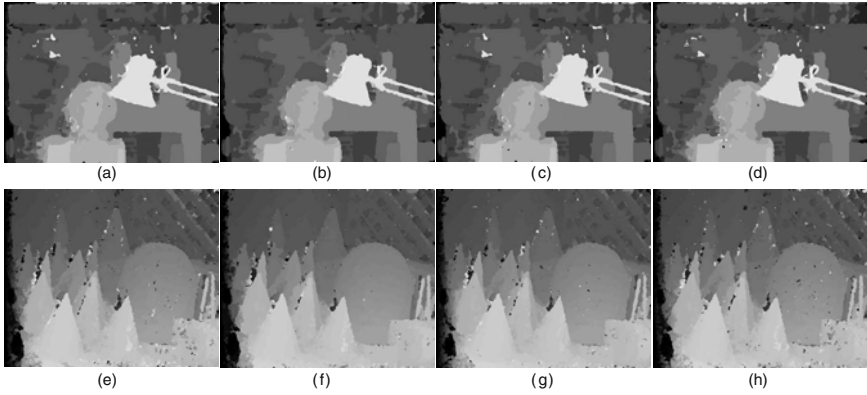


Fig. 2. Disparity maps generated using Segmented Window SAD. (a),(e) WT, (b),(f) CT, (c),(g) MRSVD and (d),(h) CONT.

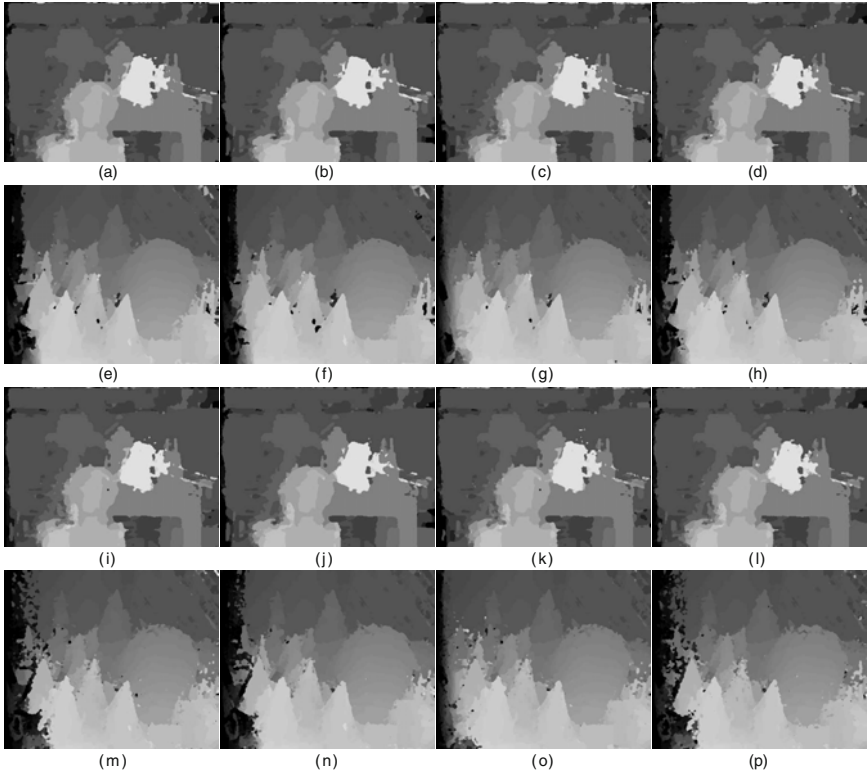


Fig. 3. Disparity maps generated using Approximate subbands and Segmented Window SAD. (a)-(h) represent level 1 and (i)-(p) represent level 2 decomposition. (a),(e),(i),(m) WT, (b),(f),(j),(n) CT, (c),(g),(k),(o) MRSVD and (d),(h),(l),(p) CONT.

window based method, time complexity analysis is not provided. Of course, the execution time depends on the platform and can be improved. Finally, Table 4 shows the quantitative comparison for segWinApprox. The results are not as good as segWin due to the unavailability of other subbands.

Comparison results reveal that CT and MRSVD generally give better results than those of CONT and WT. Moreover, the results of CT are better in lower search range as in Tsukuba pairs and MRSVD provides better results in higher search ranges as in Cones pairs. This can be attributed to the fact that same level decomposition of CT generates more subbands than those of MRSVD. Multiple subband searching, although effective for accuracy improvement, may lead to *more choice for disparity* as the search range goes higher. This apparent ambiguity results in relatively less accurate disparity estimation in higher disparity ranges for CT.

The results of segWinApprox are quite close to segWin results with improper edge restoration and reduction in minute details because approximate bands do not contain the detail parts and the edge or minute details of an image are lost. As the level of decomposition increases, the result deteriorates due to the higher resolution approximate that progressively reduce information. As the *initial disparity estimation* is done in the *smallest* approximate band at highest level of decomposition, the initial map loses detailed information. The refinements in lower level approximations are unable to provide details. Thus, higher level of decomposition normally reveals less accurate disparity maps although execution speed increases as the maximum search is done in lowest resolution.

5 Conclusions

This paper provides a study on the use of different multiresolution methods for disparity estimation with an effort to highlight the power of multiple resolution processing and the expediency and suitability of the details obtained from the subbands. This study also provides references to previous works that were relatively independent till now. Finally, the goal of the comparisons provided in the paper is to provide a better understanding of different multiresolution methods, their relative strengths and drawbacks and their effects on stereo matching.

Acknowledgments

The work is supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

1. Bhatti, A., Nahavandi, S.: Depth estimation using multiwavelet analysis based stereo vision approach. *International Journal of Wavelets Multiresolution and Information Processing* 6(3), 481–497 (2008)
2. Candes, E.J., Demanet, L., Donoho, D.L., Ying, L.: Fast discrete curvelet transforms. *Multi-scale Modelling and Simulation* 5, 861–899 (2006)

3. Caspary, G., Zeevi, Y.Y.: Wavelet-based multiresolution stereo vision. In: Proceedings of the International Conference on Pattern Recognition, vol. 3, pp. 680–683 (2002)
4. Ding, H., Fu, M., Wang, M.: Shift-invariant contourlet transform and its application to stereo matching. In: Proceedings of the International Conference on Innovative Computing, Information and Control, pp. 87–90 (2006)
5. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing* 14(12), 2091–2106 (2005)
6. Kakarala, R., Ogunbona, P.: Signal analysis using a multiresolution form of the singular value decomposition. *IEEE Transactions on Image Processing* 10(5), 724–735 (2001)
7. Mallat, S.: Wavelets for a vision. *Proceedings of IEEE* 84(4), 604–614 (1996)
8. Mukherjee, D., Wang, G., Wu, Q.: Stereo matching algorithm based on curvelet decomposition and modified support weights. Accepted at IEEE International Conference on Acoustics, Speech and Signal Processing (2010)
9. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47(1-3), 7–42 (2002)
10. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 195–202 (2003)
11. Zhang, W., Zhang, Q., Qu, L., Wei, S.: A stereo matching algorithm based on multiresolution and epipolar constraint. In: Proceedings of the International Conference on Image and Graphics, pp. 180–183 (2004)