

Stereo correspondence based on curvelet decomposition, support weights, and disparity calibration

Dibyendu Mukherjee

Guanghai Wang

Q. M. Jonathan Wu

University of Windsor

Department of Electrical and Computer
Engineering

401 Sunset Avenue

Windsor, Ontario, N9B 3P4 Canada

E-mail: mukherjd@uwindsor.ca

Abstract. A novel multiresolution analysis-based stereo matching method using curvelets, support weights, and disparity calibration is proposed. By introducing curvelet decomposition, we obtain the curvelet coefficients in different scales and orientations, and the image points can be better described and represented by these coefficients. By using support weights, the fattening effect suffered by previous methods is reduced. As a result, false matches are reduced greatly and overall accuracy is increased. Disparity calibration smoothes the disparity map and removes the remaining outliers to further improve accuracy. The proposed method is verified and compared extensively with state of the art methods, and good results and improvements are achieved. © 2010 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3377989]

Subject terms: stereo correspondence; multiresolution analysis; curvelet decomposition; support weights; disparity calibration.

Paper 090861R received Nov. 5, 2009; revised manuscript received Jan. 7, 2010; accepted for publication Feb. 16, 2010; published online Apr. 13, 2010.

1 Introduction

Stereo reconstruction is one of the most popular techniques for estimation of depth from digital images. Stereo correspondence is a process of matching the feature points in stereo image pairs. The apparent shift of position of an image point from one image to the other is called disparity. If the neighborhood points have similar local structures of gray or color values, illumination value, etc., like in repetitive textures or textureless images, then the process becomes very difficult.

There are presently two types of approaches for stereo matching—area based and feature based.

Feature-based methods work by extracting local features from the images and matching the features across the stereo pairs. The features, such as edges, corners, and lines, are matched using local feature descriptors. Local features remain more or less unaltered across the image pairs. So, feature-based methods are robust, accurate, and fast, although they usually generate sparse disparity maps that are not appropriate for some operations like dense reconstruction. There are broadly three types of features: interest points,^{1,2} edges,³ and regions.⁴ But there are also global feature-based methods.⁵ These methods, also termed structural matching, match larger features that consist of local smaller features.

Area-based approaches rely on some statistical correlation of color or intensity values. They can be classified into two types: local- and global-based methods. The local-based methods estimate the disparity of a pixel by correlating a support window around the pixel with a similar support window in the other image. Area-based methods assume all pixels in the support window to have similar

depth. This assumption is violated in depth discontinuities and results in a fattening effect near these regions. The methods select the best correlation match for each pixel. This sometimes results in wrong disparity estimation for image points having ambiguity of depth. As a consequence, textureless regions, repetitive textured regions, and regions with depth discontinuity fail to match correctly.

Some algorithms proposed for densely textured regions are variable windows⁶ and shiftable windows.⁷ Adaptive support window methods^{8,9} try to find an optimum size support window for each pixel. An initial disparity-based adaptive support window selection method was proposed by Kanade and Okutomi.⁹ Veksler¹⁰ proposed a method to compare different sized windows based on window cost. Multiple window methods try to select an optimal support window size from a predefined range of windows.^{11–13}

In contrast to local window-based methods, global area-based methods try to estimate an optimum disparity plane that minimizes a global cost function. Some global methods proved to be very effective, like cooperative optimization,¹⁴ belief propagation,^{15,16} graph cuts,¹⁷ dynamic programming,^{18,19} etc.

Multiresolution approaches gained confidence in signal processing during the last few decades. Wavelets have already been used in stereo correspondence^{20,21} for multiresolution analysis of images. In wavelet analysis, the image is decomposed into blurred, horizontal-edge, vertical-edge, and diagonal-edge images, to make feature extraction simpler. Unfortunately, the wavelet coefficients have strong dependencies across scales and subbands. Also, the wavelets are shift-variant and the coefficients become very affected by shifts. These factors limit their application in stereo correspondences.

Contourlets and curvelets are found to be more effective as local feature descriptors. They can clearly represent in-

intrinsic geometrical structures like curvatures from images. Contourlet-based correspondence has already been implemented.²² Curvelets are relatively new and have gained recognition from researchers for image denoising, face recognition, etc. Their application in correspondence search was novel until now. They can be used for local area-based matching along with any correlation function. Curvelets decompose the image into a number of scales and orientations. A correspondence match can be found in each of these scales and orientations for each pixel of the stereo images. The best match is one of these matches, and it can be estimated by simple comparison of the correlation values or left-right disparity consistency check. However, use of plain sum of absolute differences (SAD), sum of square differences (SSD), or normalized-cross-correlation (NCC) as local correlation functions yield the fattening effect as discussed before.

In this work, we propose a novel method for stereo matching based on curvelets with modified adaptive support weights (MASW) and disparity calibration. Support weights have already prove to provide very good results.²³ In the proposed method, color information is limited to original scale, because the curvelet coefficients are treated as normal grayscale values. Thus the support weights are modified to fit grayscale images. The combination of support weights with curvelets is termed curv-MASW, and the results are verified by extensive tests and evaluations. The method is further improved by using disparity calibration (DC).²⁴ The combination is termed as curv-MASW-DC and has been tested on the Middlebury database. The results are considerably better. We are also going to explain the specific applications of curv-MASW and curv-MASW-DC.

The remaining part of the work is organized as follows. We first present a brief description of curvelets in Sec. 2, followed by a description of the MASW in Sec. 3. The method for disparity calibration is presented in Sec. 4. Implementation details are described in Sec. 5. Experimental results and comparison with current state of the art methods in the Middlebury database are presented in Sec. 6. Finally, the work is concluded in Sec. 7.

2 Curvelet Transform

Wavelets as a multiresolution analysis method has had much success. Wavelets are capable of interpreting a signal as a sum of contributions in different scales and locations. They can detect point singularities in single dimensional signals due to their effective localization and multiscaling. But for higher dimensions, they yield poor results due to the lack of orientation selectivity.

Curvelets were developed by Candes and Donoho in 1999, mainly for image analysis. They have strong directional characteristics, and due to their variable width and length with a parabolic scaling of $\text{length}^2 \sim \text{width}$, the coefficient are highly anisotropic at fine scales.

As described by Candes,²⁵ curvelets can be thought of as obtained by applying parabolic dilations, rotations, and translations to a specifically shaped function ψ ; they are indexed using scale a ($0 < a < 1$), location b , and orientation θ as

$$\psi_{a,b,\theta}(x) = a^{-3/4} \psi[\mathbf{D}_a R_\theta(x - b)], \quad (1)$$

with

$$\mathbf{D}_a = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix},$$

where \mathbf{D}_a is a parabolic scaling matrix, and R_θ is a rotation by θ radians. A close example of such a function would be multiscale ridgelets. The ridgelet theory was proposed by Candes. The involvement of ridgelets and formation of curvelets are well discussed in his paper.²⁶

Second-generation curvelet transforms are simpler and faster than their previous versions. There are two types of implementation—via unequally spaced fast Fourier transform (USFFT) or via wrapping.²⁷ Curvelets via wrapping has been used here because it is faster and can be applied to any size of images effectively. For a 2-D function $f[t_1, t_2]$ with $0 < t_1, t_2 < n$ (a specific length), the main steps for wrapping are as follows.

1. Take the 2-D fast Fourier transform (FFT) of $f[t_1, t_2] \rightarrow \hat{f}[n_1, n_2]$.
2. Separate $\hat{f}[n_1, n_2]$ into dyadic subbands using a scale window W_j , with j representing the j 'th scale.
3. Separate each subband into angular wedges using angular windows $W_{j,l}$, with l representing the l 'th wedge.
4. Form the product $\tilde{U}_{j,l}[n_1, n_2] \hat{f}[n_1, n_2]$, where, $\tilde{U}_{j,l}[n_1, n_2]$ is a discrete localizing window like in Eq. (1).
5. Wrap the product inside a rectangle W of size $L_{1,j} \times L_{2,j}$ (in east-west) or $L_{2,j} \times L_{1,j}$ (in north-south) around the origin to obtain $\tilde{f}_{j,l}[n_1, n_2] = W(\tilde{U}_{j,l} \hat{f})[n_1, n_2]$, where, $L_{1,j} \sim 2^j$ and $L_{2,j} \sim 2^{j/2}$.
6. Take the 2-D inverse FFT of each $\tilde{f}_{j,l}$ to obtain curvelet coefficients at scale j and orientation l .

The curvelets measure the information of a signal at specified scale and locations but only along specified orientations. Thus, they effectively represent objects with “curve-punctuated smoothness”—smoothness except discontinuity along a general curve with bounded curvature.^{25,26} Real images have lots of curves, i.e., edges along different orientations. By using curvelets, these features can be represented with much greater accuracy than wavelets.

For the current work, the images have been decomposed into curvelet coefficients of scale=3 and orientation=8. The current choice is selected as a tradeoff between speed and accuracy, and it is clearly explained in subsequent discussions. But this decomposition is not unique, and better results are possible with a different number of scales and orientations.

Some of the coefficients as grayscale images are shown in Fig. 1(a). The main advantage of using a multiresolution method for matching lies both in initial disparity estimation and in robust disparity estimation in multiple resolutions.

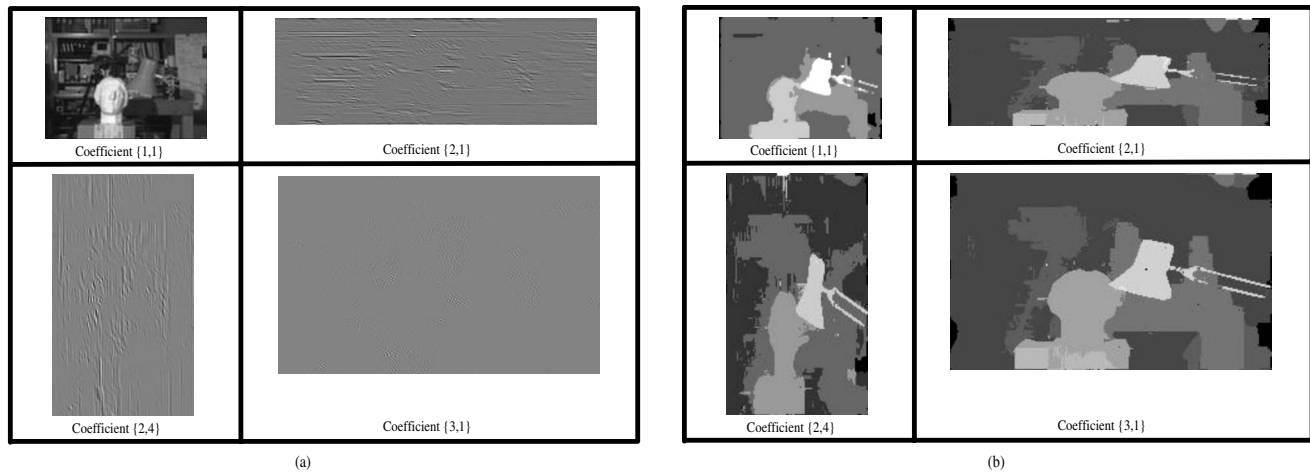


Fig. 1 (a) Four curvelet coefficients are shown for the left image of the Tsukuba pair as gray-level images. Coefficient $\{x,y\}$ denotes curvelet coefficient at scale x and orientation y . (b) Their corresponding disparity maps as processed using the proposed method.

The smallest curvelet coefficients, including the approximated image, provide smaller areas for searching. This results in higher search speed for an approximation of the correct location of the matched pixels. The actual location may have a small deviation, but with correct selection of the search window, the approximated disparity map is close to the ground truth. With the initial estimation of disparity, the search area in higher scales is reduced by a large amount. This reduction can increase computation speed and decrease uncertainty at the same time. The increase in speed is evident from the method as the search area reduces. The decrease in uncertainty is due to the decrease of choices. If the choice of a matching pixel is known to be part of a complete row of the image with an unknown exact location, then there may be a number of regions in the same row with similar textures or comparable architecture, and the matching pixel may lie in any of these regions. This increases the uncertainty in finding correct matches. The process is shown in Fig. 2. The search area chosen in the first resolution is broadened as the resolution increases.

On the other hand, the correspondences can be improved by using curvelets because of their orientation flexibility. Figure 1 shows four different curvelet coefficients as im-

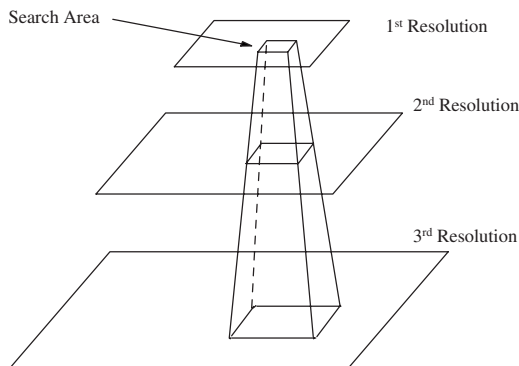


Fig. 2 Smaller search area in lower resolution is extended in larger resolutions.

ages and their corresponding disparity maps as obtained using the proposed method. A close observation of the disparity maps reveals that the disparity is more prominent along the horizontal edges for coefficient $\{2,1\}$, and along the vertical edges for coefficient $\{2,4\}$. If the better matches are taken from the two disparity maps, the resulting map will have prominent disparity difference along both horizontal and vertical directions. It is true for other orientations also. This experimentally proved fact is used in the proposed method to seek optimal disparity. Also, for scale 3 there are eight orientations that yield sufficient orientation flexibility. From the figures, it is clear that the disparity shown for coefficient 3,1 resembles the original disparity to a good extent. Using higher numbers of scales and orientations does not produce significant improvement on the disparity, so we have limited our method to third scale only.

The advantages of using curvelets make it a better choice in comparison to other multiresolution methods like the pyramid method and wavelet decomposition. The advantages also come with a burden of traversing a number of subbands for a reasonably good disparity map. In the case of wavelets, the number of subbands is limited to four for a single level and the height/width ratio remains equal. The same is not the case for curvelets. But, this problem is well compensated with the improvement of quality.

3 Modified Adaptive Support Weight

While computing the correspondences between two pixels in left and right images, the support from neighboring pixels is used for similarity measurements. The Gestalt principles of similarity and proximity are used for computation of support weights. The adaptive support weight measurements have been proposed in Ref. 23. The support weight is written as

$$w(p,q) = k \cdot f_s(\Delta c_{pq}) f_p(\Delta g_{pq}), \quad (2)$$

where Δc_{pq} and Δg_{pq} represent the color difference and spatial distance between pixels p and q . k is a proportionality constant. $f_s(\cdot)$ and $f_p(\cdot)$ represent the strength of

grouping by similarity and proximity, respectively. The color space is chosen to be CIELab because of the ease of 3-D representation of color. In the color space, Δc_{pq} is expressed as

$$\Delta c_{pq} = [(L_p - L_q)^2 + (a_p - a_q)^2 + (b_p - b_q)^2]^{1/2}, \quad (3)$$

where (L_p, a_p, b_p) are the three color components of pixel p . The strength of grouping by similarity is expressed as

$$f_s(\Delta c_{pq}) = \exp\left(-\frac{\Delta c_{pq}}{\gamma_c}\right), \quad (4)$$

where γ_c is a constant. Similarly, the strength of grouping by proximity has the expression

$$f_p(\Delta g_{pq}) = \exp\left(-\frac{\Delta g_{pq}}{\gamma_p}\right), \quad (5)$$

where Δg_{pq} can be computed using a simple Euclidean distance measure between $p(x_p, y_p)$ and $q(x_q, y_q)$ as follows:

$$\Delta g_{pq} = [(x_p - x_q)^2 + (y_p - y_q)^2]^{1/2}. \quad (6)$$

By combining Eqs. (4) and (5), Eq. (2) becomes

$$w(p, q) = k \exp\left[-\left(\frac{\Delta c_{pq}}{\gamma_c} + \frac{\Delta g_{pq}}{\gamma_p}\right)\right]. \quad (7)$$

In our implementation, the curvelets basically provide a group of 2-D matrices that are treated as gray-level images. Thus, we cannot use the color information to compute Δc_{pq} as in Eq. (3). As a replacement, we modify the support weight by including the gray-level difference value

$$\Delta c_{pq} = |m_p - m_q|, \quad (8)$$

where m_p represents the gray-level value of pixel p . Using only gray-level information, the accuracy decreases, but it is balanced by the accuracy improvement due to the search operation in multiple resolution and using curvelet coefficients. For original images, the color values can be used, but to make the process similar in each resolution, we use the norm of the CIELab color data,

$$m_p = (L_p^2 + a_p^2 + b_p^2)^{1/2}. \quad (9)$$

This gives a reasonably good approximation of the color data and is equivalent to the gray-level values.

The correspondence matching part is similar to any local method using cost function minimization. The support weights are computed in support windows around the pixel under consideration and its corresponding pixel in the other image. Then, the dissimilarity of the two pixels is computed by aggregating a raw matching cost with the support weights. Dissimilarity between a pixel p in the left image and a pixel \bar{p} in the right image $\mathbf{D}(p, \bar{p})$ is expressed as

$$\mathbf{D}(p, \bar{p}) = \frac{\sum_{q \in N_p, \bar{q} \in N_{\bar{p}}} w(p, q) w(\bar{p}, \bar{q}) e(q, \bar{q})}{\sum_{q \in N_p, \bar{q} \in N_{\bar{p}}} w(p, q) w(\bar{p}, \bar{q})}, \quad (10)$$

where N_p and $N_{\bar{p}}$ represent support windows of p and \bar{p} , respectively, where the support weights are computed, and

$e(p, q)$ represents the raw matching cost between pixel p and q . For the current experiment, this raw matching cost has been taken as the sum of absolute difference (SAD) of gray values. For Eq. (10), $w(p, q)$ represents the support weight between pixel p and its neighbor q inside the support window of N_p , as described in Eq. (7). The final disparity of the pixel p corresponds to the pixel \bar{p} for which $\mathbf{D}(p, \bar{p})$ is minimized,

$$d_{\text{final}} = \arg \min_{\bar{p}} \mathbf{D}(p, \bar{p}). \quad (11)$$

4 Disparity Calibration

Disparity calibration is based on the assumption that points with similar color or short spatial distances have similar disparity. For any point in the image, a calibration window is chosen around it. The size of the window is optimized according to the criteria that the pixels in the window have similar color or short distance. Then, the occurrence number or frequency of every possible disparity is checked in the window. The disparity with the highest occurrence number is assigned to the point. The process of disparity calibration can be summarized as follows.

1. Find the maximum and minimum disparities ($d_{\text{max}}, d_{\text{min}}$).
2. For each disparity d in $[d_{\text{max}}, d_{\text{min}}]$, define zero matrix \mathbf{C}_d with the same size as the image. There will be a series of matrices.
3. In the original disparity map \mathbf{D} , find the pixels with disparity d , assign 1 to the corresponding elements in \mathbf{C}_d .
4. Result disparity $\mathbf{D}(i, j)$ for pixel (i, j) is expressed as

$$\mathbf{D}(i, j) = \arg \max_d \left\{ \sum_{x=-1/2(W_x-1)}^{x=1/2(W_x-1)} \sum_{y=-1/2(W_y-1)}^{y=1/2(W_y-1)} \times w[(i, j)(i+x, j+y)] \times C_d(i+x, j+y) \right\}, \quad (12)$$

where W_x and W_y represent the width and height of the calibration window, respectively, and $w(x, y) = \exp[-(\Delta c_{pq}/\gamma_i + \Delta g_{pq}/\gamma_p)]$ represent the MASW of pixel (x, y) with $\gamma_i < \gamma_c$ to make the color constraint stricter.

The disparity range ($d_{\text{max}}, d_{\text{min}}$) is already estimated from the initial and final disparity calibration. Thus, the implementation of disparity calibration is straightforward.

In our implementation, the output of curv-MASW is improved using disparity calibration. The results of the process are shown in Sec. 6.2.

5 Implementation

The implementation steps in this section describe the disparity estimation using curv-MASW-DC. The steps are as follows.

1. Perform curvelet transform of the grayscale stereo pairs.
2. Compute the correspondence match in the lowest approximate scale of curvelets by Eq. (10).
3. Shift to the next scale, first orientation.
4. Compute $\text{scale_factor} = (\text{size of current scale image}) / (\text{size of previous scale image})$.
5. For each pixel, divide the coordinate of the pixel by scale_factor to get the coordinates in the previous scale. Then, compute the $\text{initial_curr_disparity} = \text{scale_factor} * \text{disparity in the previous scale}$.
6. Take a range of search around the $\text{initial_curr_disparity}$.
7. Find the best match using MASW from Eq. (10).
8. If this is the last orientation, go to step 10.
9. Shift to the next orientation and go to step 5.
10. For each pixel, there is a match in each orientation at the current scale. Best match can be found by a normal correlation value check or a left-to-right and right-to-left disparity consistency check.
11. Smooth the $\text{initial_curr_disparity}$ using disparity calibration.
12. If this is the last curvelet scale, go to step 13, else go to step 3.
13. For original image pair, use the norm of CIE Lab color images using Eqs. (9) and (10) and compute the final disparity map using MASW.

The implementation does not depend on any initial disparity estimation provided externally. The initial estimation is already performed in the lowermost scale, i.e., the approximation. The approximated disparity map is then improved in each scale using different orientations of curvelets. Therefore, the proposed method performs a complete disparity computation rather than pre- or postprocessing.

The curv-MASW method excludes the disparity calibration, i.e., it does not contain step 11. We discuss the effectiveness of the curv-MASW and curv-MASW-DC methods, their relative differences, and their applications in Sec. 6.

The initial disparity map is not totally correct due to the approximated scale but can define the range of search, and also has the limit of maximum search. With the curvelet coefficients obtained at $\text{scale}=3$, $\text{orientation}=8$, the approximated image for a Tsukuba pair has a dimension of 97×29 . In this dimension, using the window size 21×21 , the maximum disparity obtained is 5. If this is multiplied by a scale factor equal to the ratio of the original image size and approximated image size, the disparity in the original image size has a maximum range of 15. This reduces the search area at the border regions, even if the approximate disparity map does not have any disparity value in these areas, as shown in Figs. 3(a) and 3(d).

The support weights reduce the fattening effect by a considerable amount. We compare the results of the initial disparity maps with support weights [as shown in Figs. 3(a) and 3(c)] and with NCC [shown in Figs. 3(b) and 3(d)]. The Cones and Aloe disparity maps for the approximated image created with support weights are clearer and have sharp disparity changes along the edges. The disparity maps with NCC show poor results along the edges and

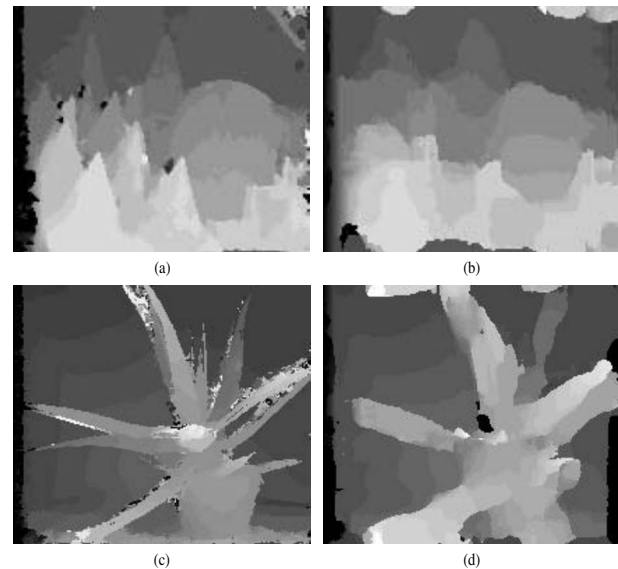


Fig. 3 Comparison of the initial disparity map generated with support weight and NCC. (a) Cones with support weights. (b) Cones with normalized cross correlation. Similarly, (c) Aloe with support weights and (d) Aloe with normalized cross correlation.

occluding areas. The reason for the reduction of the fattening effect is well explained in the paper by Yoon and Kweon.²³

The discontinuities in the images at the occlusion boundaries are not properly recovered in the disparity map using local methods due to inaccurate support window size selection. Using support weights, this limitation is largely removed. The support weights are based on the color similarity and distance information of the neighboring pixels, and they can well predict the disparity in the discontinuities, even in a fixed window size. It is further improved with the use of unique curvelet coefficients, which add the edge and orientation information to the pixels. Finally, disparity calibration smoothes the disparity map by assigning disparity to those pixels that still remain to be matched. Thus, by combining the methods, the disparity maps have been highly improved and are comparable to the results of global methods.

6 Experimental Results

We tested the proposed method extensively with 2001, 2003, 2005, and 2006 datasets from Middlebury and compared the results with state of the art methods in the literature. Even if our method is kept limited to using only curvelets and support weights, the results are of good quality. This is the case when step 11 is removed from the implementation steps in Sec. 5. Some of the results are shown in Sec. 6.1 and extensively compared with the original support weight results. With the inclusion of disparity calibration, the results are improved and are presented in Sec. 6.2.

6.1 Results from Curvelets with Modified Adaptive Support Weights

The experiments were carried out on the Middlebury dataset^{28,29} and the results are compared on the Middlebury

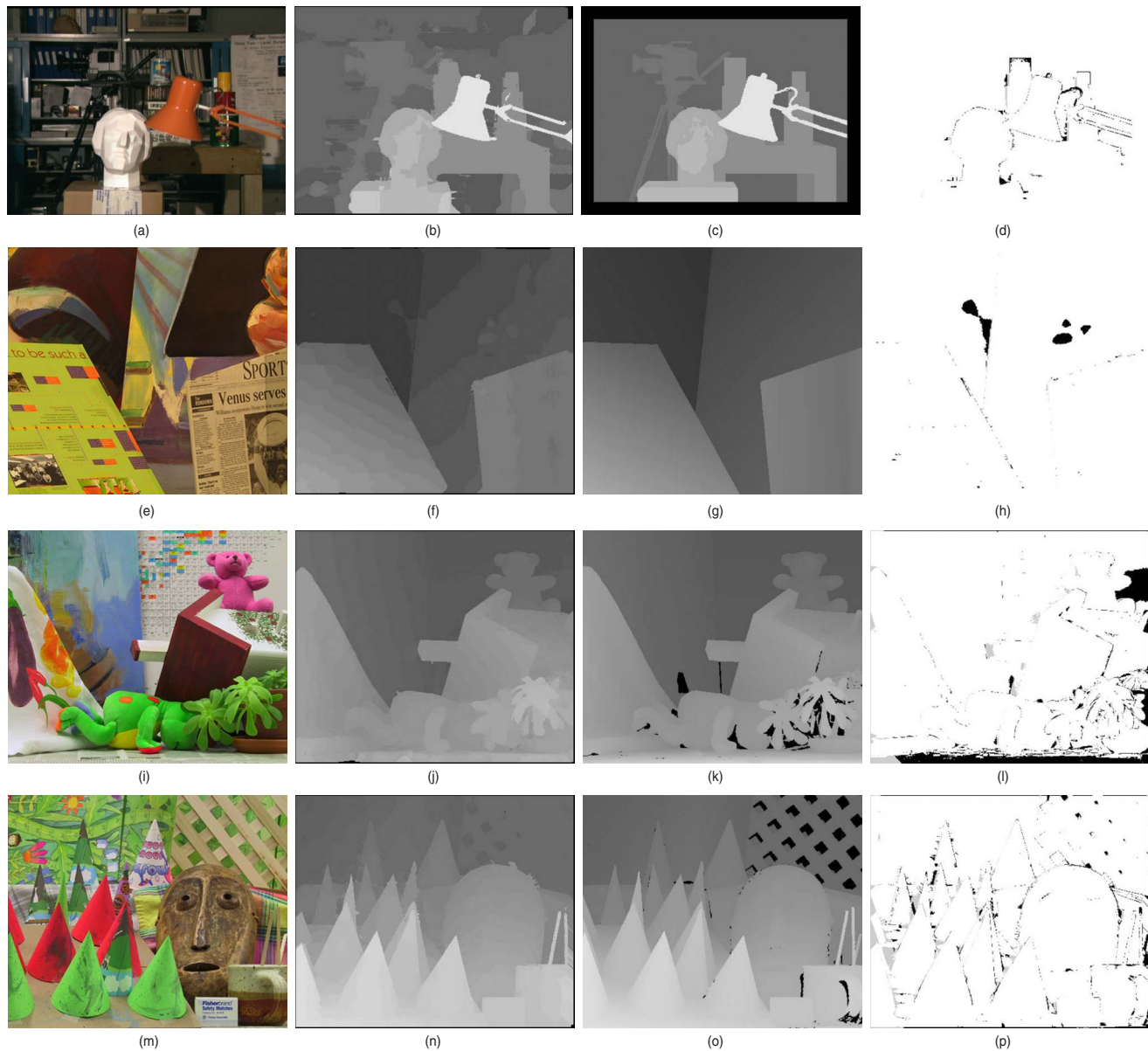


Fig. 4 Dense disparity maps for the Middlebury images using curv-MASW and their corresponding groundtruths. (a) Tsukuba left image, (b) corresponding curv-MASW, (c) groundtruth, (d) bad pixels, (e) Venus left image, (f) corresponding curv-MASW, (g) groundtruth, (h) bad pixels, (i) Teddy left image, (j) corresponding curv-MASW, (k) groundtruth, (l) bad pixels, (m) Cones left image, (n) corresponding curv-MASW, (o) groundtruth, and (p) bad pixels.

website with other methods. The main comparison has been done with the original support weight approach,²³ termed AdaptWeight for simplicity.

The parameters in the tests are specified as follows: size of support window = 33×33 , $\gamma_p = 36$, $\gamma_c = 7$, and $k = 1.5$. For the smallest approximation, window size has been taken as 21×21 . The search area for higher scales has ± 10 pixel deviation from the original disparity position of the initial disparity map and reduces to ± 5 pixels for the original size image pairs.

The experimental results are shown in Fig. 4. The numbers in Table 1³⁰⁻³² represent the percentage of bad pixels compared to the groundtruth: for all pixels “all,” pixels in the nonoccluded regions “nonocc,” and pixels near depth

discontinuities “disc.” Our method outperforms most of the local methods for stereo matching, even without including disparity calibration. It is also better than some of the global-based methods like RegionTreeDP, RealTimeBP, etc. Of course there are still some local and global methods better than curv-MASW. But the results, including disparity calibration from Sec. 6.2, show that the method is comparable to the best global and local methods. In this section, our main concern is to show the improvement using curvelets. Thus, more results have been provided with other datasets from the Middlebury database^{33,34} in Fig. 5, along with the results using the original support weights. These results have been evaluated with the formula for percentage of bad pixels defined as,

Table 1 Comparison with state of the art methods. The proposed method is denoted as curve-MASW.

Methods	Tsukuba			Venus			Teddy			Conces		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
CoopRegion ¹⁴	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01
DoubleBP ¹⁶	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.30	9.63	2.90	8.78	7.79
GeoSup ³⁰	1.45	1.83	7.71	0.14	0.26	1.90	6.88	13.2	16.1	2.94	8.89	8.32
AdaptDispCalib ²⁴	1.19	1.42	6.15	0.23	0.34	2.50	7.80	13.6	17.3	3.62	9.33	9.72
Curv-MASW	1.40	1.84	7.42	1.00	1.11	4.42	7.85	8.84	16.8	3.82	6.22	8.24
CostAggr+occ ³¹	1.38	1.96	7.14	0.44	1.13	4.87	6.80	11.9	17.3	3.60	8.57	9.36
RegionTreeDP ¹⁸	1.39	1.64	6.85	0.22	0.57	1.93	7.42	11.9	16.8	6.31	11.9	11.8
AdaptWeight ²³	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26
SegTreeDP ³²	2.21	2.76	10.3	0.46	0.60	2.44	9.58	15.2	18.4	3.23	7.86	8.83
RealtimeBP ¹⁵	1.49	3.40	7.87	0.77	1.90	9.00	8.72	13.2	17.2	4.61	11.6	12.4
GC+occ ¹⁷	1.19	2.01	6.24	1.64	2.19	6.75	11.2	17.4	19.8	5.36	12.4	13.0

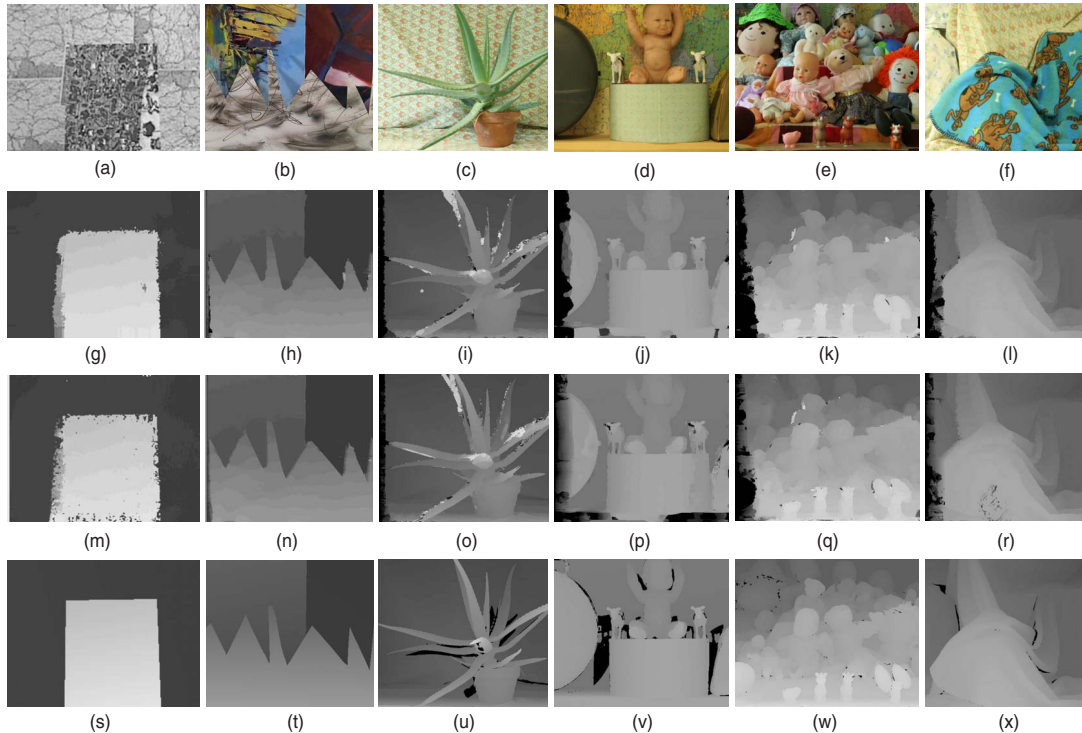


Fig. 5 Comparison of curv-MASW with AdaptWeight for Middlebury datasets. Each column consists of a different set of left images of a stereo pair, disparity map produced by curv-MASW, disparity map produced by AdaptWeight, and the corresponding groundtruth. (a) Map, (b) Sawtooth, (c) Aloe, (d) Baby, (e) Dolls, (f) Cloth, (g) through (l) represent corresponding curv-MASW results, (m) through (r) represent corresponding AdaptWeight results, and (s) through (x) represent corresponding groundtruths.

Table 2 Comparisons of curv-MASW with the original adaptive support weight method.

Methods	Image					
	Map	Sawtooth	Aloe	Baby	Dolls	Cloth
Curv-MASW	0.8212	0.5382	0.1666	0.2692	0.3595	0.2103
AdaptWeight	0.8554	0.5472	0.2114	0.3960	0.4494	0.2926

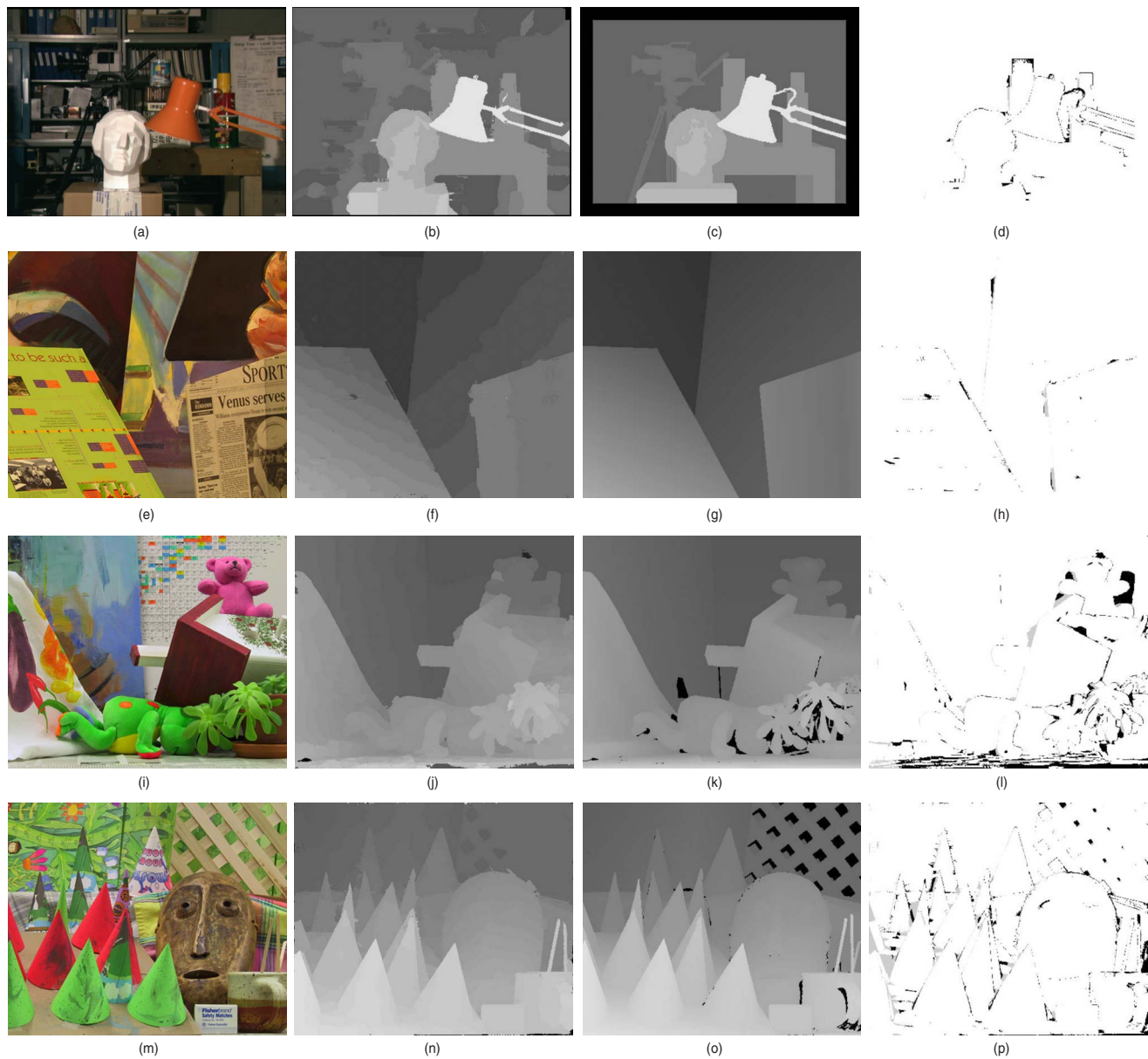


Fig. 6 Dense disparity maps for the Middlebury images using curv-MASW-DC, the corresponding groundtruths, and the bad pixels from curv-MASW-DC. (a) Tsukuba left image, (b) corresponding curv-MASW-DC, (c) groundtruth, (d) bad pixels, (e) Venus left image, (f) corresponding curv-MASW-DC, (g) groundtruth, (h) bad pixels, (i) Teddy left image, (j) corresponding curv-MASW-DC, (k) groundtruth, (l) bad pixels, (m) Cones left image, (n) corresponding curv-MASW-DC, (o) groundtruth, and (p) bad pixels.

Table 3 Comparison with state of the art methods. The proposed method is denoted as curv-MASW-DC.

Methods	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
CoopRegion ¹⁴	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01
DoubleBP ¹⁶	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.30	9.63	2.90	8.78	7.79
Curv-MASW-DC	1.36	1.80	7.18	0.46	0.85	4.17	5.65	6.71	14.8	3.23	5.51	8.45
GeoSup ³⁰	1.45	1.83	7.71	0.14	0.26	1.90	6.88	13.2	16.1	2.94	8.89	8.32
SymBP+occ ³⁵	0.97	1.75	5.09	0.16	0.33	2.19	6.47	10.7	17.0	4.79	10.7	10.9
AdaptDispCalib ²⁴	1.19	1.42	6.15	0.23	0.34	2.50	7.80	13.6	17.3	3.62	9.33	9.72
Curv-MASW	1.40	1.84	7.42	1.00	1.11	4.42	7.85	8.84	16.8	3.82	6.22	8.24
CostAggr+occ ³¹	1.38	1.96	7.14	0.44	1.13	4.87	6.80	11.9	17.3	3.60	8.57	9.36
RegionTreeDP ¹⁸	1.39	1.64	6.85	0.22	0.57	1.93	7.42	11.9	16.8	6.31	11.9	11.8
AdaptWeight ²³	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26

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

where N is the total number of pixels in the image, d_c and d_t represent the computed and groundtruth disparity maps, respectively, and δ_{thresh} is the threshold for bad pixels (usually equal to 1.0). The results are shown columnwise, with each column representing the left image of a stereo pair, its disparity map generated with curv-MASW, AdaptWeight, and the corresponding groundtruth. The estimated percentage of bad pixels are tabulated in Table 2. We can see that the proposed method has a lower percentage of bad pixels.

From the comparison result in Table 1, we can see that curv-MASW has better performance than the original support weights, RegionTreeDP, RealtimeBP, etc. methods. In comparison to support weights, the Tsukuba and Venus pairs do not show considerable improvements, while the Cones and Teddy pairs are significantly improved. The reason for this lies in the initial disparity map creation. For Tsukuba and Venus, the maximum disparity is much smaller compared to Cones and Teddy. For smaller disparities, the initial disparity map creation process does not produce effectively better results than the other methods. Nevertheless, with increasing amount of disparity, the method performs better than others due to the uniqueness of the curvelet coefficients in lower resolutions. This difference is prominent in Cones and Teddy. The Cones pair has been listed in the top two in the Middlebury evaluation for all matching regions. The time complexity does not change too much with the variation of image size, because the main searching operation is performed in lowest scales, and higher scale operations have fixed search areas.

The time complexity can be reduced greatly if some of the higher order scales and orientations are omitted, e.g., for third scales, some orientations have repetitive textures and can be omitted with a small decrease in accuracy. The

result can be further improved by using a higher number of scales and angles or using better outlier removal methods in combination.

The simple combination of curvelets and support weights is fast and can produce disparity maps suitable for applications that do not need highly accurate disparity maps. An example would be a preprocessing step for a global-area-based method.

6.2 Combination with Disparity Calibration

Curv-MASW in combination with disparity calibration has been carried out on a current Middlebury dataset, and results have been compared with other methods. The results are shown in Fig. 6, and comparisons with some of the best global and local methods are tabulated in Table 3,³⁵ where curv-MASW-DC stands for the proposed method, and AdaptDispCalib denotes the original disparity calibration. Currently, curv-MASW-DC performs best among the local area-based methods.

As seen from the results, curv-MASW-DC significantly improved the Cones and Teddy disparity maps. The improvement is due to the early disparity estimation at the lower resolutions, the uniqueness of the curvelet coefficients, and smoothing of the disparity map using calibration. For AdaptDispCalib, the disparity range has to be provided for a smaller search range and reduction of disparity ambiguity. For curv-MASW-DC, this range is readily provided by lower resolution data. The results are better compared to existing local methods, and thus it is one of the best in area-based methods.

The results also confirm the previous conclusion that the proposed algorithm performs better with an increased amount of disparity. The improvements on Cones and Teddy images made them rank first in the Middlebury

evaluation for all region matching. Curv-MASW-DC itself produces highly accurate disparity maps and involves a relatively lower number of parameters dependent on image size and type, in comparison to other methods. Thus, the proposed method can be thought of as a complete stereo-matching process, unlike curv-MASW.

7 Conclusion

We propose a new multiresolution approach for a stereo correspondence search. The approach adopts curvelets as the multiresolution method, support weights as the cost minimization method, and disparity calibration as a smoothing function and outlier minimizer. The method does not depend on initial disparity estimation and does not suffer from the foreground fattening effect. The outputs are comparable to state of the art algorithms. The algorithm can be used as a preprocessing step for some global energy minimization methods, or as a complete matching process, as already explained.

In the future, we plan to improve the algorithm using better local cost minimization methods to perform more accurately in lower disparities.

Acknowledgments

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

References

1. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(10), 1615–1630 (2005).
2. C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.* **37**(2), 151–172 (2000).
3. C. Schmid and A. Zisserman, "The geometry and matching of lines and curves over multiple views," *Int. J. Comput. Vis.* **40**(3), 199–233 (2000).
4. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.* **65**(12), 43–72 (2005).
5. Y. Wang, "Principles and applications of structural image matching," *J. Photogram. Remote Sens.* **53**(3), 154–165 (1998).
6. O. Veksler, "Fast variable window for stereo correspondence using integral images," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, vol. **1**, pp. 556–561 (2003).
7. J. Kim, K. Lee, B. Choi, and S. Lee, "A dense stereo matching using two-pass dynamic programming with generalized ground control points," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, vol. **2**, pp. 1075–1082 (2005).
8. Y. Boykov, O. Veksler, and R. Zabih, "A variable window approach to early vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(12), 1283–1294 (1998).
9. T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(9), 920–932 (1994).
10. O. Veksler, "Stereo correspondence with compact windows via minimum ratio cycle," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(12), 1654–1660 (2002).
11. A. Bobick and S. Intille, "Large occlusion stereo," *Int. J. Comput. Vis.* **33**(3), 181–200 (1999).
12. A. Fusiello, V. Roberto, and E. Trucco, "Efficient stereo with multiple windowing," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, pp. 858–863 (1997).
13. S. Kang, R. Szeliski, and C. Jinxjang, "Handling occlusions in dense multi-view stereo," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, vol. **1**, pp. 103–110 (2001).
14. Z. Wang and Z. Zheng, "A region based stereo matching algorithm using cooperative optimization," *IEEE Conf. Computer Vis. Patt. Recog.*, pp. 1–8 (2008).
15. Q. Yang, L. Wang, and R. Yang, "Real-time global stereo matching using hierarchical belief propagation," *Br. Mach. Vision Conf.*, pp. 989–998 (2006).
16. Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(3), 492–504 (2009).
17. V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions via graph cuts," *Proc. Intl. Conf. on Computer Vision*, pp. 508–515 (2001).
18. C. Lei, J. Selzer, and Y. H. Yang, "Region-tree based stereo using dynamic programming optimization," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, pp. 2378–2385 (2006).
19. L. Wang, M. Liao, M. Gong, R. Yang, and D. Nister, "High-quality real-time stereo using adaptive cost aggregation and dynamic programming," *Proc. 3rd Intl. Symp. 3D Data Process. Visual. Transmission*, pp. 798–805 (2006).
20. G. Casparly and Y. Y. Zeevi, "Wavelet-based multiresolution stereo vision," *Proc. 16th Intl. Conf. Patt. Recog.*, vol. **3**, pp. 680–683 (2002).
21. S. Mallat, "Wavelets for a vision," *Proc. IEEE* **84**(4), 604–614 (1996).
22. H. Ding, M. Fu, and M. Wang, "Shift-invariant contourlet transform and its application to stereo matching," *Proc. 1st Intl. Conf. Innovative Comput. Info. Control*, pp. 87–90 (2006).
23. K. Yoon and I. Kweon, "Locally adaptive support-weight approach for visual correspondence search," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, vol. **2**, pp. 924–931 (2005).
24. Z. Gu, X. Su, Y. Liu, and Q. Zhang, "Local stereo matching with adaptive support-weight, rank transform and disparity calibration," *Pattern Recogn. Lett.* **29**(9), 1230–1235 (2008).
25. E. J. Candes, "What is...a curvelet?," *Not. Am. Math. Soc.* **50**(11), 1402–1403 (2003).
26. E. J. Candes and D. L. Donoho, "Curvelets—a surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surface Fitting*, pp. 105–120, Vanderbilt University Press, Nashville, TN (2000).
27. E. J. Candes, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.* **5**, 861–899 (2006).
28. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.* **47**(1–3), 7–42 (2002).
29. D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, vol. **1**, pp. 195–202 (2003).
30. A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann, "Local stereo matching using geodesic support weights," *Proc. Intl. Conf. on Image Processing*, pp. 2093–2096, IEEE, Piscataway, NJ (2009).
31. D. Min and K. Sohn, "Cost aggregation and occlusion handling with wls in stereo matching," *IEEE Trans. Image Process.* **17**(8), 1431–1442 (2008).
32. Y. Deng and X. Lin, "A fast line segment based dense stereo algorithm using tree dynamic programming," *Eur. Conf. Computer Vision*, pp. 201–212 (2006).
33. D. Scharstein and C. Pal, "Learning conditional random fields for stereo," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, pp. 1–8 (2007).
34. H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, pp. 1–8 (2007).
35. J. Sun, Y. Li, S. Kang, and H. Shum, "Symmetric stereo matching for occlusion handling," *Proc. IEEE Conf. Computer Vis. Patt. Recog.*, vol. **2**, pp. 399–406 (2005).



Dibyendu Mukherjee received his bachelors degree in electronics and telecommunications engineering from Bengal Engineering and Science University, India, in 2006. From 2006 to 2008, he worked as a consultant in PricewaterhouseCoopers Private Limited, and since 2008, he has been pursuing his MAsc from University of Windsor, Canada, in the Department of Electrical and Computer Engineering. His research interests include stereo correspondence analysis, 3-D reconstruction, robot localization, and tracking and motion detection. He is also associated with IEEE as a graduate student member.



Guanghui Wang received his MS degree in control theory and application from Jilin University of Technology in 2000. He is currently a research fellow with the Department of Electrical and Computer Engineering, University of Windsor, Canada. His research interests include structure from motion, camera calibration, artificial intelligence, and robot localization and navigation. He has published more than 50 peer-reviewed papers in major journals and conferences. He has served as a reviewer and program committee member for many journals, research councils, and conferences.



Jonathan Wu received his PhD degree in electrical engineering from the University of Wales, Swansea, United Kingdom, in 1990. He is currently a professor in the Department of Electrical and Computer Engineering at the University of Windsor, Canada. He holds the Tier 1 Canada Research Chair in automotive sensors and sensing systems. He has published more than 100 peer-reviewed papers in areas of computer vision, image processing, intelligent systems, robotics, microsensors and actuators, and integrated microsystems. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.