Anisotropic local high-confidence voting for accurate stereo correspondence

Jiangbo Lu [*,†], Gauthier Lafruit [†], and Francky Catthoor [*,†]

[*] Department of Electrical Engineering, University of Leuven, Belgium

[†] Multimedia Group, IMEC, Kapeldreef 75, B-3001, Leuven, Belgium

ABSTRACT

We present a local area-based, discontinuity-preserving stereo matching algorithm that achieves high quality results near depth discontinuities as well as in homogeneous regions. To address the well-known challenge of defining appropriate support windows for local stereo methods, we use the anisotropic Local Polynomial Approximation (LPA) - Intersection of Confidence Intervals (ICI) technique. It can adaptively select a near-optimal anisotropic local neighborhood for each pixel in the image. Leveraging this robust pixel-wise shape-adaptive support window, the proposed stereo method performs a novel matching cost aggregation step and an effective disparity refinement scheme entirely within a local high-confidence voting framework. Evaluation using the benchmark Middlebury stereo database shows that our method outperforms other local stereo methods, and it is even better than some algorithms using advanced but computationally complicated global optimization techniques.

Keywords: Stereo matching algorithm, view interpolation, shape-adaptive, anisotropic local approximation

1. INTRODUCTION

Depth from stereo is an important computer vision topic that has attracted intensive research interests for decades. A substantial amount of work has been done on stereo correspondence, which is systematically surveyed and evaluated by Scharstein and Szeliski.[1] In general, casting a stereo problem as a global optimization problem usually leads to high quality disparity estimation results. However, most of these global techniques are too computationally expensive and sometimes require many parameters that are hard to determine. On the contrary, local area-based stereo methods, without global reasoning, are complexity-wise more feasible for a number of practical real-time applications, e.g., online view synthesis and robot navigation.

In fact, our research is clearly motivated by two basic observations. First, several local stereo methods[2,3] are proposed recently, even outperforming many global optimization based algorithms in terms of disparity estimation accuracy. Secondly, local stereo methods actually prevail in real-time vision systems, achieving extremely high disparity estimation throughputs, e.g., recent algorithms on graphics processing units (GPUs).[4–6] In this paper, we hence center the study on local area-based stereo correspondence algorithms. Specifically, we are primarily concerned with the disparity estimation quality, and this paper presents an advanced local stereo algorithm, yielding novel insights to the algorithmic question, "how far can we go with the local stereo matching method?"

The fundamental challenge to all stereo algorithms is to properly tackle image ambiguity problem in the correspondence search, resulting from image noise, occlusion, lack of texture, and repetitive patterns. Typically, local area-based approaches choose to aggregate the matching cost over a given support window to increase the robustness to noise and texture variation. An implicit assumption for these approaches is that all pixels in a support window are from similar depth in a scene, and therefore, that they have similar disparities. Although this assumption functions considerably well for homogeneous regions, it results in the "foreground-fattening" phenomenon for regions near depth discontinuities or occluding boundaries, where a naive support window scheme mistakenly aggregates the contribution of pixels from different depths. Therefore, to obtain accurate results at depth discontinuities as well as on homogeneous regions, an appropriate support window for each

Further author information: (Send correspondence to Jiangbo Lu.)
E-mail: {Jiangbo.Lu, Gauthier.Lafruit, and Francky.Catthoor}@imec.be, Telephone: +32 (0) 16 28 87 49

Image Processing: Algorithms and Systems VI, edited by Jaakko T. Astola,
Karen O. Egiazarian, Edward R. Dougherty, Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 6812, 68120J, © 2008 SPIE-IS&T · 0277-786X/08/$18

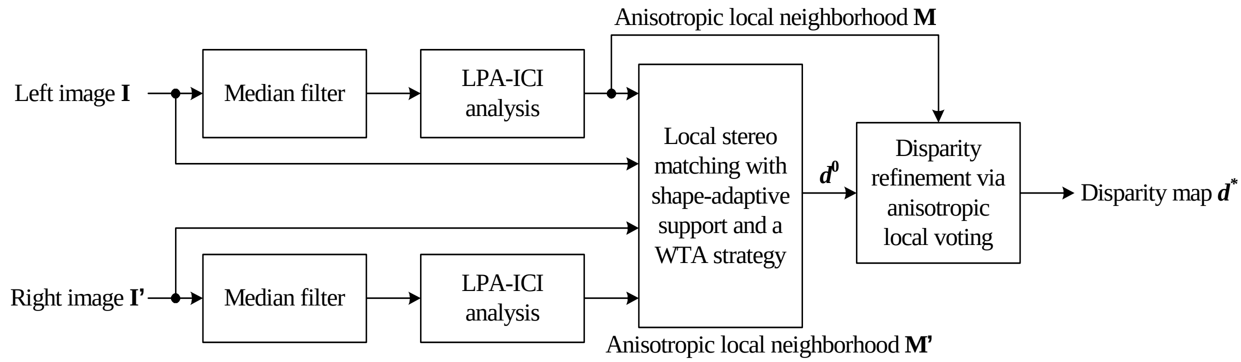


Figure 1. The flowchart of the proposed stereo matching algorithm.

pixel should be decided adaptively. Towards this end, many local stereo methods have been proposed, and they roughly fall into three categories according to their techniques: 1) adaptive-window methods,[2,7,8] trying to find an optimal support window for each pixel, 2) multiple-window methods,[9,10] selecting an optimal support window among predefined multiple windows, and 3) methods using adaptive support-weights,[3,11] assigned to the pixels in a local support window with a fixed shape and size.

Even though the aforementioned methods improve the correspondence estimation quality, there are some inherent limitations associated with them. Firstly, the shape of a local window is not general,[2,7–10] therefore their rectangular or constrained-shaped windows are inappropriate for pixels near arbitrarily shaped depth discontinuities. In addition, some of these techniques[7,11] are highly dependent on the initial disparity estimation, which may be erroneous.

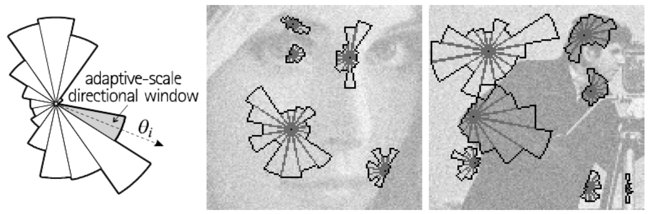
In this paper, we propose a novel stereo correspondence algorithm, which produces high-quality disparity estimation results at depth discontinuities as well as in homogeneous regions. As a state-of-the-art local stereo method, the proposed algorithm is built upon on an elegant local approximation technique,[12,13] combining two independent nonparametric estimation ideas: the *local polynomial approximation* (*LPA*) and the *intersection of confidence intervals* (*ICI*) *rule*. Asymptotically, these LPA-ICI estimators allow to get a near-optimal quality of the signal recovery. Applying this LPA-ICI technique to a stereo image pair, we can derive an anisotropic local neighborhood with high confidence for each pixel. This resulting pixelwise varying-scale multidirectional neighborhood plays a central role in our stereo matching process. Based on the generated robust local support masks, our stereo method performs the entire disparity estimation and refinement within a local high-confidence voting framework. To the best of our knowledge, this is the first attempt at integrating the spatially adaptive signal processing theory of this kind with the local stereo matching research problem.

The paper is organized as follows: we start with the stereo problem statement and overview of the proposed algorithm in Section 2, where the notation is also defined and to be used throughout the rest of the paper. Section 3 then presents our stereo matching method in details. Experimental results are shown in Section 4, along with the discussions. Finally, we conclude our work and outline the future research directions in Section 5.

2. PROBLEM STATEMENT AND ALGORITHM OVERVIEW

Given a pair of stereo images, $\mathbf{I}$ and $\mathbf{I}'$, the goal of a *dense* stereo algorithm is to establish pixel correspondences between the images, generating a disparity vector estimate at each pixel. For the scope of this paper, we assume the input stereo images are *rectified*, i.e., the epipolar lines are aligned with corresponding scanlines. In this case, the disparity vector degenerates to a scalar d, defined as the horizontal shift between the correspondences X and X' from the left ($\mathbf{I}$) and right image ($\mathbf{I}'$), respectively. Let the coordinate vectors $X = (x, y)$ and $X' = (x', y')$, then the coordinates are correlated as follows:

$$x' = x - d, \quad y' = y \; . \tag{1}$$



(a) (b)

Figure 2. Anisotropic LPA-ICI. (a) Sectorial structure of the anisotropic neighborhood achieved by combining a number of adaptive-scale directional windows. (b) Some of these windows selected by the ICI for the noisy *Lena* and *Cameraman* images. Image courtesy of the Institute of Signal Processing, Tampere University of Technology.[13]

According to the taxonomy,[1] stereo algorithms generally perform four steps: matching cost computation, cost aggregation, disparity computation and optimization, and disparity refinement. Similar to most local stereo approaches, the proposed algorithm places a key emphasis on the cost aggregation step to reduce the ambiguity in matching. Besides the significant contribution in this step, we also propose a novel disparity refinement technique that noticeably enhances the accuracy of the final disparity map.

The entire flowchart of the proposed stereo algorithm is shown Fig. 1. Our algorithm first applies a 3×3 median filter to two stereo images, suppressing the impact of image noise as well as very subtle non-Lambertian surfaces. Next, the LPA-ICI technique is adopted to address the most challenging problem for local stereo matching paradigm, i.e., optimally deciding a shape-adaptive local support neighborhood for each pixel. Thanks to this robust nonlinear estimator, we can produce a pixelwise anisotropic local voting mask with high confidence, namely, $\mathbf{M}_X$ for the pixel X in $\mathbf{I}$, and $\mathbf{M'}_{X'}$ for the pixel X' in $\mathbf{I'}$. Integrating this mask information within a local voting framework, the proposed method performs matching cost aggregation in a pointwise adaptive manner. Then, a WTA (Winner-Takes-All) strategy, usually coupled with local stereo methods, chooses at each pixel the disparity d_X^0 associated with the minimum cost value. Finally, we make use of the reliable mask $\mathbf{M}_X$ to further refine the disparity estimate d_X^0, resulting in our final disparity value d_X^* for the pixel X.

3. DENSE STEREO VIA ANISOTROPIC LOCAL HIGH-CONFIDENCE VOTING

As illustrated in Fig. 1, the proposed dense stereo matching method is composed of three major components: 1) LPA-ICI analysis, 2) anisotropic aggregation of matching cost, and 3) disparity refinement using a local voting scheme. Central to our disparity estimation and refinement—2) and 3)—is an anisotropic local high-confidence voting mechanism, as to be discussed in Section 3.2 and 3.3. Nevertheless, it is the LPA-ICI technique[12] that assures the success of the proposed voting framework, supplying a near-optimal pixelwise shape-adaptive voting mask. We therefore start with a brief introduction of the LPA-ICI technique, before delving into the ideas of our stereo method.

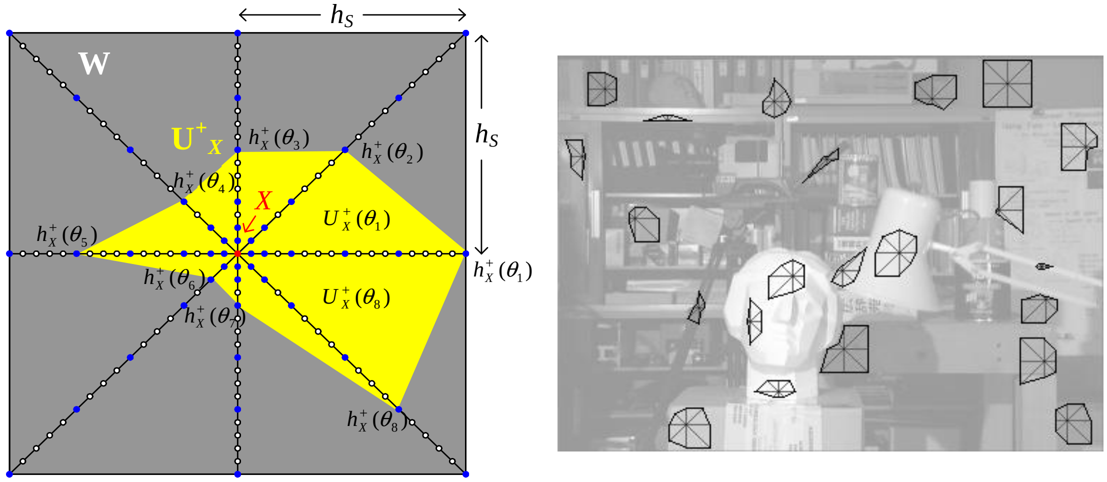
3.1 LPA-ICI technique

Anisotropic LPA-ICI[12] is an efficient technique recently developed for robust analysis and restoration of noisy data, adapting to the unknown smoothness of the signal. When applying it to 2D image processing, multidirectional sectorial-neighborhood estimates are first calculated for every point, and then the ICI rule is exploited for the adaptive selection of the size of each sector. As a result, the estimator is anisotropic and the shape of its support adapts to the structures present in the image. Fig. 2 shows a sectorial structure of the anisotropic neighborhood, and some examples of these anisotropic neighborhoods for the *Lena* and *Cameraman* images.[13]

To present the anisotropic LPA-ICI technique[12] in more detail, we consider noisy image observations $\mathbf{Z}$ of the form

$$\mathbf{Z}(X) = \mathbf{Y}(X) + \eta(X) , \tag{2}$$

where $\mathbf{Y}$ is the original image, $\eta(X) \sim \mathcal{N}(0, \sigma^2)$ is independent Gaussian white noise, X is a spatial variable. The anisotropic LPA-ICI proceeds as follows. At first, for every specified direction θ_k, $k = 1, ..., K$, a varying-scale family of directional-LPA convolution kernels $\{g_{h,\theta_k}\}_{h \in H}$ is used to obtain a corresponding set of directional



(a) (b)

Figure 3. (a) Fast implementation of the anisotropic LPA-ICI technique. (b) Sample anisotropic local neighborhoods for the *Tsukuba* image.

varying-scale estimates $\{\hat{\mathbf{Y}}_{h,\theta_k}\}_{h\in H}$, $\hat{\mathbf{Y}}_{h,\theta_k} = \mathbf{Z} * g_{h,\theta_k}$, $h \in H$, where H is the set of scales (i.e., the sector sizes exemplified in Fig. 2(a)). These estimates are then compared according to the ICI rule,[12,13] and it searches for a largest local vicinity of the point of estimation, where the local polynomial approximation assumptions fit well to the data. As a result, an adaptive scale $h^+_X(\theta_k) \in H$ is defined for every direction θ_k for the pixel location X. The final anisotropic LPA-ICI estimate $\hat{\mathbf{Y}}(X)$ is yielded by "fusing" together the corresponding adaptive-scale estimates $\hat{\mathbf{Y}}_{h^+_X(\theta_k),\theta_k}(X)$.

However, in this paper we are not interested in this anisotropic estimate $\hat{\mathbf{Y}}(X)$. Instead, for the purpose of the local stereo matching, we are especially concerned with the pointwise anisotropic neighborhood achieved by the LPA-ICI technique. Moreover, to approach an optimal trade-off between the complexity and the estimation performance, we do not need a large variety of different shapes as broad as in the examples of Fig. 2. In practice, a much simplified neighborhood structure[13] is used in our implementation. More specifically, narrow one-dimensional "linewise" directional LPA kernels $\{g_{h,\theta_k}\}_{h\in H}$, with $H = \{1, 2, 3, 5, 9, 13, 18\}$, are used for $K = 8$ directions. As shown in Fig. 3(a), the candidate scales $h_X(\theta_k)$ along each direction θ_k are marked as blue dots. After the ICI-based selection of the adaptive-scales $\{h^+_X(\theta_k)\}_{k=1}^8$, the anisotropic neighborhood $\mathbf{U}^+_X$ is constructed as a polygonal hull of the adaptive-scale directional supports decided by the neighboring $h^+_X(\theta_k)$, i.e., $\mathbf{U}^+_X = \bigcup_{k=1}^8 \mathbf{U}^+_X(\theta_k)$. Fig. 3 illustrates the fast implementation of an anisotropic neighborhood $\mathbf{U}^+_X$, and some examples for the *Tsukuba* image.

Provided with such a robust pixelwise anisotropic neighborhood $\mathbf{U}^+_X$, we produce a binary voting mask $\mathbf{M}_X$ for each pixel location X in the left image $\mathbf{I}$. As a 2D matrix, $\mathbf{M}_X$ is defined over a square window $\mathbf{W}$ of size $(2 \cdot h_S - 1) \times (2 \cdot h_S - 1)$ ($h_S = 18$ in our case), as follows,

$$\mathbf{M}_X(r) = \begin{cases} 1, & \forall r \in \mathbf{U}^+_X \\ 0, & \forall r \in \mathbf{W} \setminus \mathbf{U}^+_X \ , \end{cases} \tag{3}$$

where r is a relative displacement vector with respect to the center pixel position X. Likewise, an anisotropic local voting mask $\mathbf{M'}_{X'}$ can be derived for every pixel X' of the right image $\mathbf{I'}$.

Up to this point, the technique is presented implicitly for grayscale images. When dealing with the color images, we adopt the same color-space transformation scheme[13] before the LPA-ICI analysis. To ensure a high-confidence of the estimated anisotropic neighborhood, we apply the LPA-ICI technique to luminance and chrominance channels separately. In the end, a min-filter is used to choose the smallest scales from these three

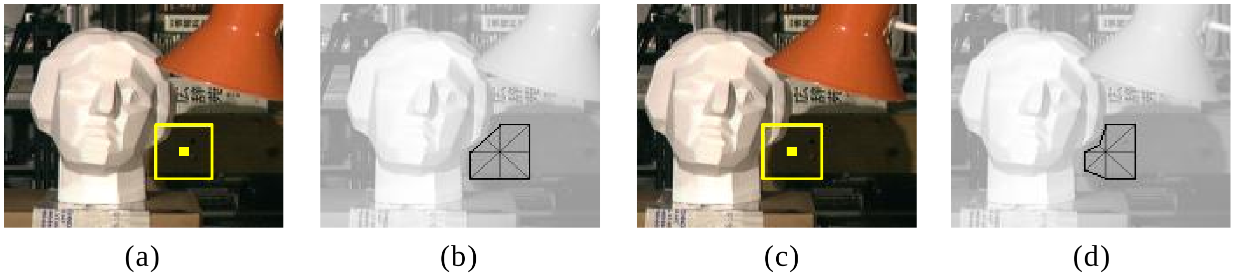


Figure 4. Correspondence matching with shape-adaptive local windows. (a) Close-up on the reference image. (b) Close-up on the reference anisotropic neighborhood. (c) Close-up on the target image. (d) Close-up on the target anisotropic neighborhood.

channels, deciding the final pointwise adaptive-scales. This is a method more conservative than the *structural constraint in luminance-chrominance space*.[13]

3.2 Anisotropic aggregation of correspondence matching cost

The key idea of our matching cost aggregation is to adaptively aggregate raw matching costs of inlier pixels while rejecting the outliers, and this is enabled by generating a combined local voting mask. Taking both high-confidence anisotropic masks $\mathbf{M}_X$ and $\mathbf{M'}_{X'}$ as the inputs, the combined local voting mask for the matching pair (X, X') is given by $\mathbf{M}_X \cdot \mathbf{M}'_{X'}$. In fact, this joint voting process is essential to select reliable pixels from similar depths for both reference and target windows, while excluding outliers that would otherwise contaminate the aggregated cost. An implicit assumption we make here is that pixels in the same anisotropic local neighborhood belong to the same surface, and hence they usually have similar disparity values. This is also an assumption widely adopted in recent segmentation-based stereo algorithms,[14] and it especially holds true in our case, where local neighborhoods have a fairly limited size.

Fig. 4 visualizes the basic idea of the proposed stereo matching method, which is implemented with shape-adaptive local support windows for both stereo images. More rigorously, the aggregated matching cost between the pixel X in $\mathbf{I}$ and X' in $\mathbf{I}'$, $E(X, X')$, can be expressed as

$$E(X, X') = \frac{\sum_{r\in\mathbf{W}} \mathbf{M}_X(r) \cdot \mathbf{M}'_{X'}(r) \cdot e_{X,X'}(r)}{\sum_{r\in\mathbf{W}} \mathbf{M}_X(r) \cdot \mathbf{M}'_{X'}(r)} \ . \tag{4}$$

The pixel-based raw matching cost $e_{X,X'}(r)$ in (4) is computed using all color channels,

$$e_{X,X'}(r) = \min\left(\sum_{c\,\in\{\mathrm{R,G,B}\}} \left|\mathbf{I}_c(X+r) - \mathbf{I}'_c(X'+r)\right|, \tau\right) \ , \tag{5}$$

where τ is the threshold used for the truncated absolute difference measure.

After the matching cost aggregation, the disparity of each pixel is simply selected by the WTA (Winner-Takes-All) method without any global reasoning as

$$d_X^0 = \arg\min_{d\in D} E(X, X') \equiv \arg\min_{d\in D} E\left(X, X-(d,0)\right) \ , \tag{6}$$

where $D = \{d_{min}, \cdots, d_{max}\}$ is the set of all possible disparities.

Actually, we also notice that our joint voting mask (based on the LPA-ICI technique) is in spirit close to bilateral filtering used in a recent stereo method.[3] Despite that the bilateral filter is a good edge-preserving smoothing filter, its color dissimilarity weighting term is solely dependent on the center pixel's intensity. This makes its adaptive smoothing capability drop significantly, when the filter kernel is centered on an outlier pixel.[15] In contrast, the LPA-ICI technique roots in the multiple statistical hypotheses testing, and its optimal scale adaptation is decided by the best trade-off between the bias and variance of the estimate. From a theoretical

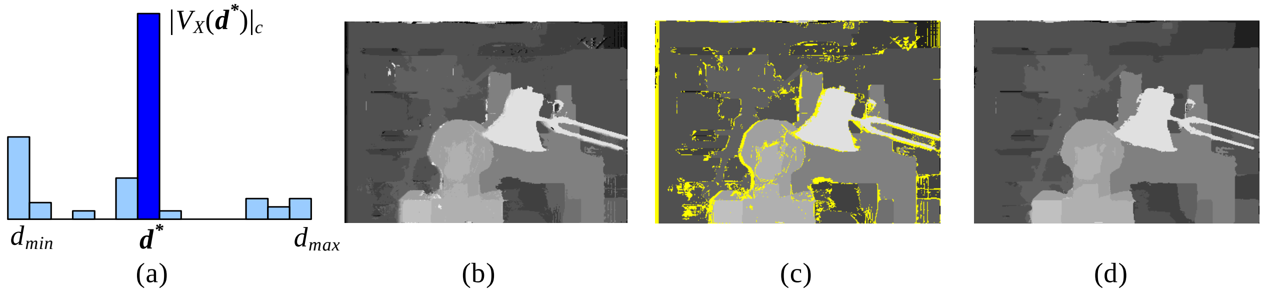


Figure 5. Disparity refinement using an anisotropic local voting scheme. (a) The histogram of the estimated disparities. (b) Disparity map d_X^0 before the refinement, with error rates@{all, untex., disc.} = {3.04, 3.39, 10.24}. (c) Disparity map change mask. (d) Disparity map d_X^* after the refinement, with error rates@{all, untex., disc.} = {2.29, 1.98, 9.39}.

perspective, the LPA-ICI technique is hence more robust. Furthermore, the anisotropic local neighborhood produced by the LPA-ICI analysis is a connected region anchored on the center pixel, while the bilateral filter per its soft-weighting design cannot enforce this connectivity constraint. This may make the bilateral filter-based stereo method[3] problematic, when dealing with the image regions with repetitive textures but varying depths.

3.3 Disparity refinement

Although the disparity estimate d_X^0 obtained after the WTA is of a good accuracy, we can further refine it with a local high-confidence voting scheme. Unlike most local stereo methods that hardly perform any effective post-processing to improve the results of the WTA, our disparity refinement step can noticeably enhance the quality of the disparity map, for both depth discontinuities and homogeneous regions. Moreover, the proposed disparity refinement scheme invokes little additional computational complexity, because it utilizes the pixelwise anisotropic local neighborhood, the result already available from previous steps.

The principle of our disparity refinement can be explained statistically by the *plurality voting* concept. Basically, if the disparity estimation for each pixel inside the same anisotropic local neighborhood is considered as performing an identical experiment repeatedly, the outcomes of different repetitions are independent, because the estimation and the WTA are performed independently for each pixel. However, since we assume with high confidence that pixels in the same anisotropic local neighborhood have similar disparity values, all the observed independent outcomes are very likely to build a distribution with a unique peak. This peak corresponds to the common mean (or the "optimal" disparity on a discrete grid) that we use to regulate and refine the disparity estimates.

To implement the proposed voting scheme, we first create a histogram of the estimated disparities d_{X+r}^0 from the WTA, with $X+r$ in the vicinity of the anchor pixel X. More clearly, for each disparity d from the given disparity search set D, a corresponding set of *effective votes*, $V_X(d)$, from the anisotropic local neighborhood is constructed as

$$V_X(d) = \left\{ r \mid r \in \mathbf{U}_X^+ \text{ and } d_{X+r}^0 = d \right\} \ . \tag{7}$$

Then, based on this histogram characterizing the distribution of effective votes (see Fig. 5(a)), the final disparity at X is set to the disparity d_X^*, i.e., the histogram bin with the maximum vote number

$$d_X^* = \arg\max_{d \in D} |V_X(d)|_c \ , \tag{8}$$

where $|V_X(d)|_c$ denotes the cardinality of the set $V_X(d)$. Finally, we use a simple disparity extrapolation step to refine the disparities of the trivial image border region.

From Fig. 5(b-d), we notice that the typical phenomenon of the foreground overextension over occluded pixels is mitigated after our refinement, and the disparity accuracy for the low-textured image parts is improved as well. The quantitative Middlebury stereo evaluation service[16] also confirms that the proposed disparity refinement reduces the error rates for all sorts of *Tsukuba* image regions (i.e., "all", "untextured", and "discontinuity").

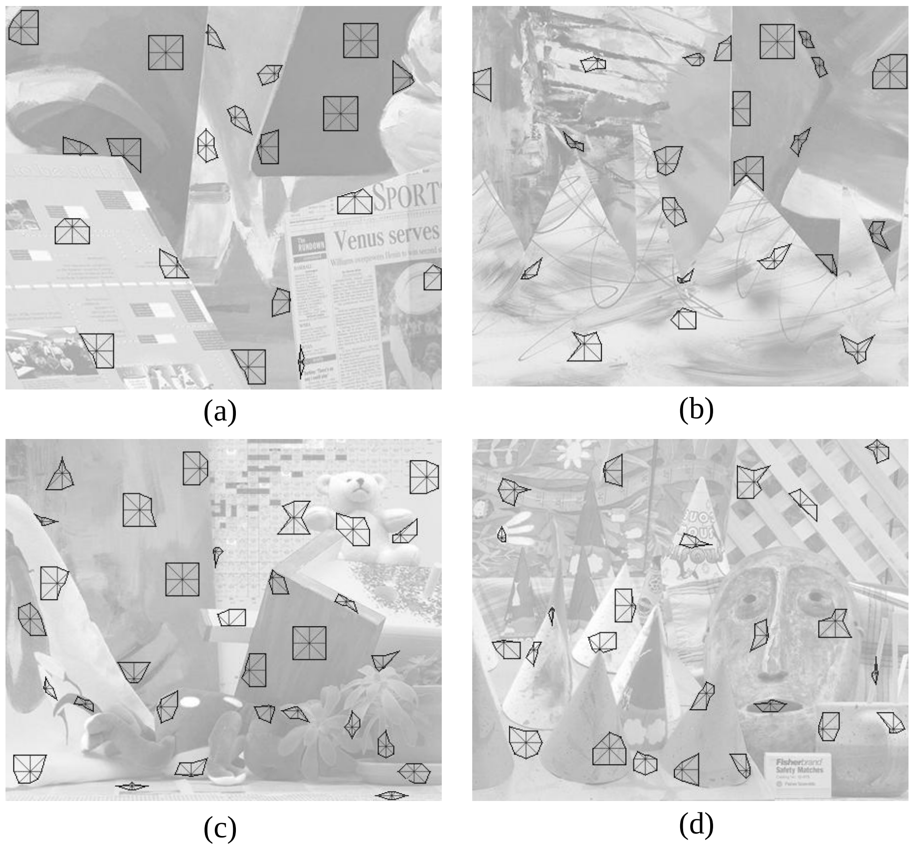


Figure 6. Sample adaptive anisotropic local neighborhoods: (a) *Venus*. (b) *Sawtooth*. (c) *Teddy*. (d) *Cones*.

4. EXPERIMENT RESULTS AND DISCUSSIONS

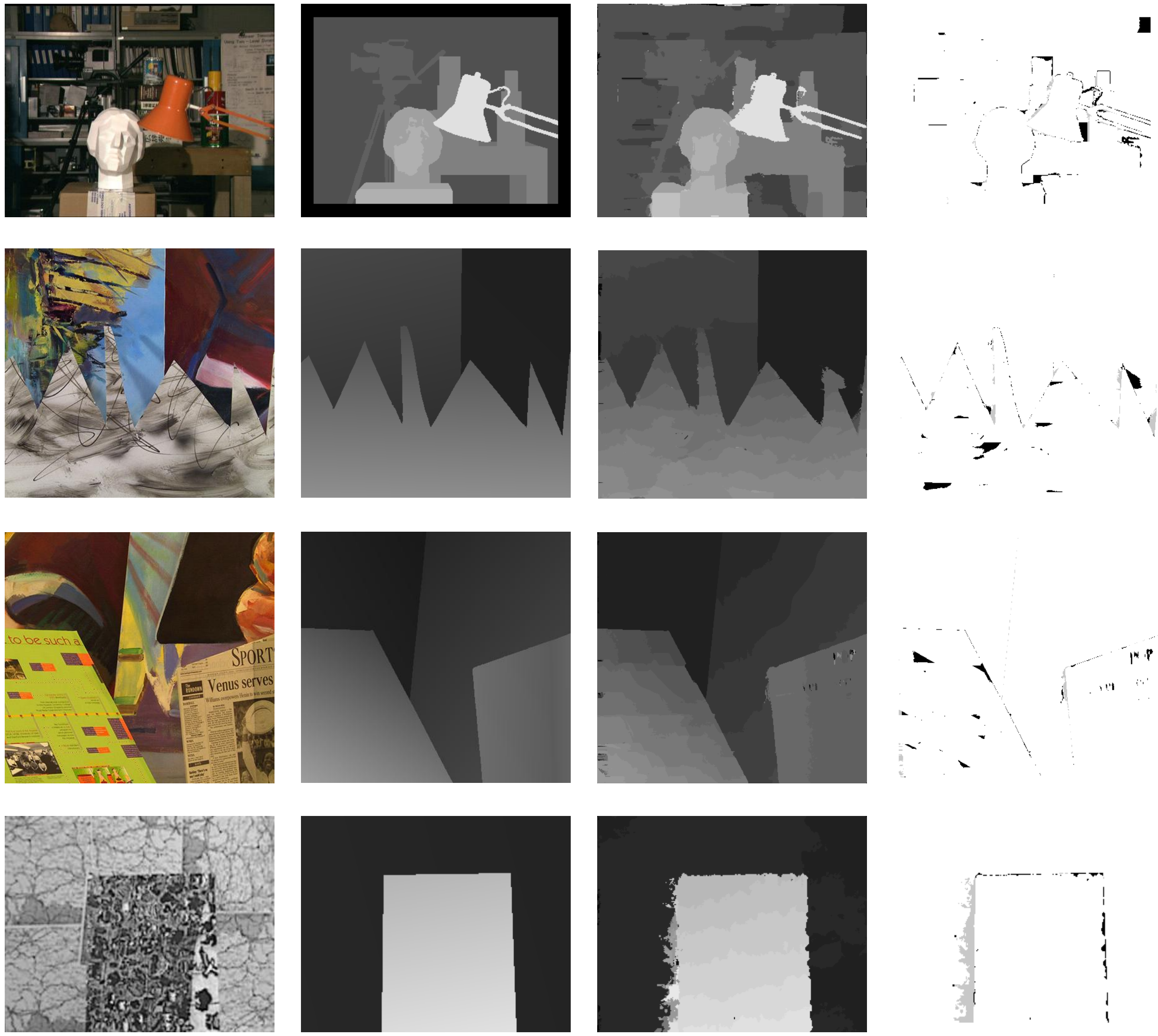
This section presents the experimental results of 1) estimation of anisotropic neighborhoods, 2) evaluation of the stereo correspondence accuracy, and 3) interpolated virtual views. The anisotropic LPA-ICI analysis module in our implementation is based on the public Matlab software*. We set our parameters constant across all experiments, i.e., $H = \{1, 2, 3, 5, 9, 13, 18\}$, $K = 8$, and $\tau = 80$, as the only parameters discussed in Section 3.

Estimation of anisotropic neighborhoods. Fig. 6 shows the results of the LPA-ICI analysis on different input images. The resulting anisotropic local neighborhood appropriately adapts its shape and scale for various types of image locations. Because of this robust discontinuity-preserving local neighborhood for each pixel, our stereo method is placed on solid ground, when performing the proposed correspondence search within a local high-confidence voting framework.

Evaluation using Middlebury stereo data sets. To evaluate the performance of the proposed local stereo method, we base our experiments on the rectified stereo images from the benchmark Middlebury stereo database.[16] It contains diversified stereo datasets with ground truth, featuring different parallax and occlusions. Fig. 7 and Fig. 8 show the visual results of our method for the original Middlebury stereo data sets and the new data sets, respectively. The proposed method yields accurate results at the depth discontinuities as well as in the homogeneous regions for the testbed images.

Using the online Middlebury stereo evaluation service,[16] we summarize the quantitative results of our method and those of other competitive local stereo methods[2,3,8,10,17,18] in Table 1, roughly in descending order of overall performance. The numbers in Table 1 represent error rates, i.e., the percentage of "bad" pixels whose

*Pointwise Shape-Adaptive DCT Demobox, http://www.cs.tut.fi/∼foi/SA-DCT/.



(a) (b) (c) (d)

Figure 7. Dense disparity maps for *Tsukuba*, *Sawtooth*, *Venus*, and *Map* images. (a) Left image. (b) Ground truth. (c) Our result. (d) Bad pixels (error > 1).

absolute disparity error is greater than 1. We report the error rates separately for all pixels ("all"), pixels in untextured regions ("untex.", except for the *Map* image), and pixels near depth discontinuities ("disc."). As shown in Table 1, the proposed method is highly ranked among the state-of-the-art local area-based methods. Although the adaptive-weights method[3] is ranked before ours, it may have employed extra optimization techniques to generate the high-quality disparity maps (unexposed in the authors' public code available at http://vision.middlebury.edu/stereo/code/). For this reason, it is not absolutely fair to position this algorithm as a local stereo method, when assessing the merits of different local area-based techniques. Even so, our method has a quantitative performance very close to this adaptive-weights method,[3] with an even better performance for the *Venus* and *Map* images. Compared to the other methods, the proposed method performs particularly better for depth discontinuities, because it can preserve arbitrarily shaped depth discontinuities well, while the methods using rectangular or constrained-shaped windows cannot.

However, the proposed method has a worse performance than other local methods on the *Map* images. This

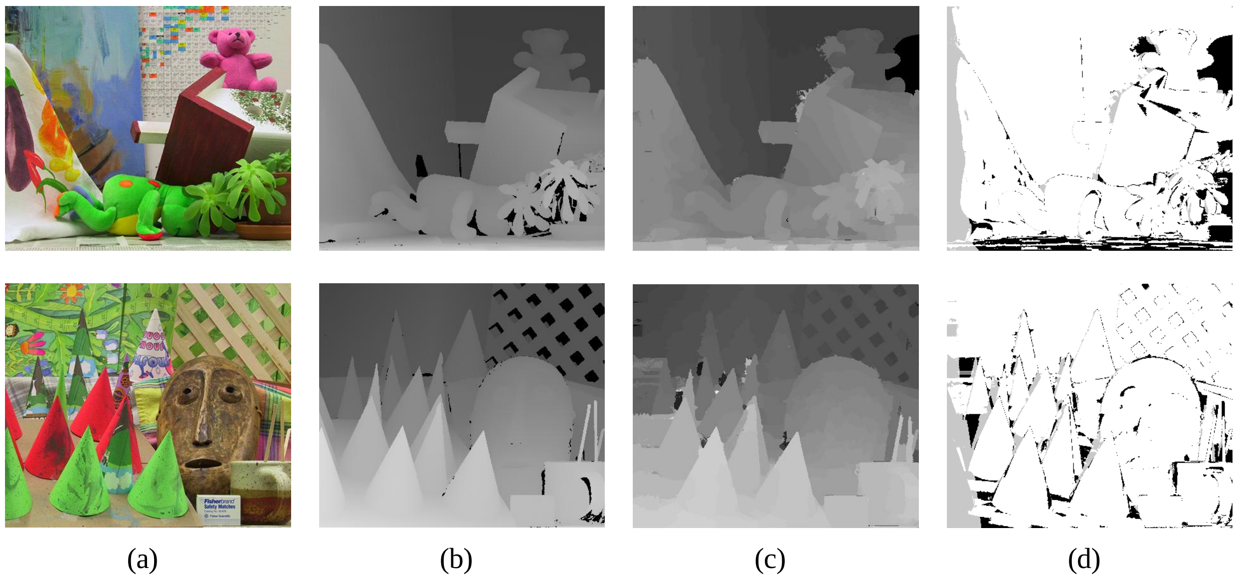


Figure 8. Dense disparity maps for *Teddy* and *Cones* images. (a) Left image. (b) Ground truth. (c) Our result. (d) Bad pixels (error > 1).

Table 1. Quantitative evaluation results for the original benchmark Middlebury stereo database.

Algorithm	*Tsukuba*			*Sawtooth*			*Venus*			*Map*	
	all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all	disc.
Adapt.weights[3]	1.29	0.61	6.72	0.97	0.34	4.82	0.99	0.89	6.66	1.13	11.55
Our method	**2.29**	**1.98**	**9.39**	**1.32**	**0.32**	**4.77**	**0.80**	**0.61**	**3.67**	**0.70**	**9.53**
Variable win.[2]	2.35	1.65	12.17	1.28	0.23	7.09	1.23	1.16	13.35	0.24	2.98
Compact win.[8]	3.36	3.54	12.91	1.61	0.45	7.87	1.67	2.18	13.24	0.33	3.94
Cooperative[17]	3.49	3.65	14.77	2.03	2.29	13.41	2.57	3.52	26.38	0.22	2.37
Bay. diff.[18]	6.49	11.62	12.29	1.45	0.72	9.29	4.00	7.21	18.39	0.20	2.49
Shiftable win.[10]	5.23	3.80	24.66	2.21	0.72	13.97	3.74	6.82	12.94	0.66	9.35

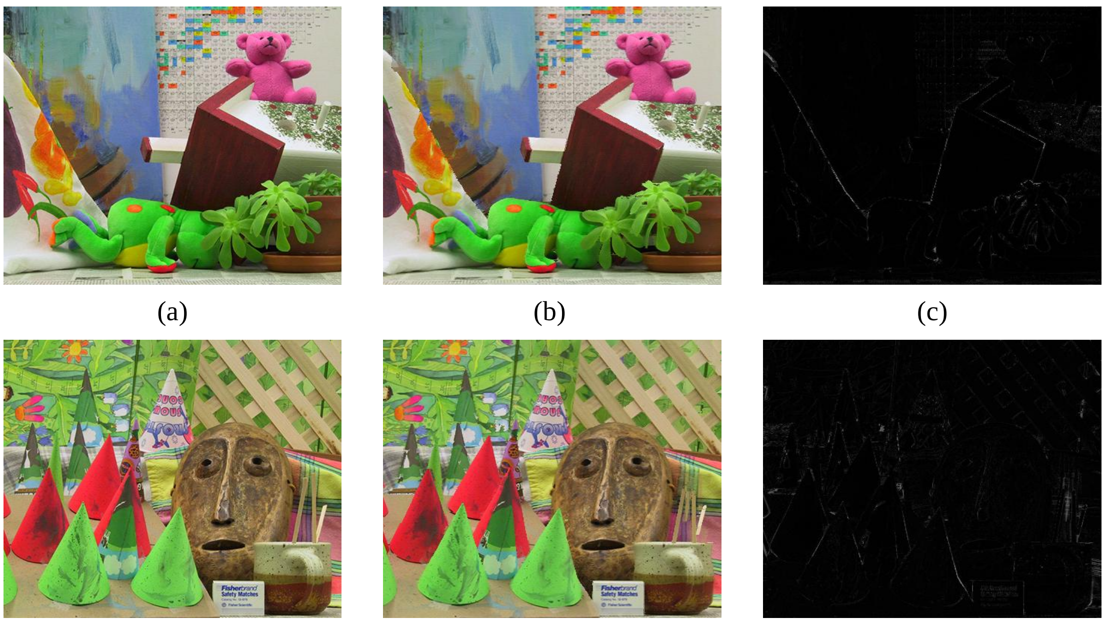
is because the images are highly textured, while our anisotropic local neighborhood tends to be very small for highly-textured regions, resulting in insufficient support for the pixels under consideration.

Table 2 presents the quantitative evaluation results for the new Middlebury stereo database, which includes more challenging stereo images. The performance of the proposed stereo method is comparable to the global methods,[14, 19] and it is even better than some advanced but complex correspondence algorithms.[20, 21]

Interpolated views. Though high-quality view interpolation is not within the scope of this paper, we demonstrate that our estimated disparity maps are capable of generating photorealistic novel view images. Using the disparity map estimated for the left viewpoint, we first generated an initial center view image based on the baseline view interpolation algorithm.[22] To appropriately render half-occluded regions and border regions, we implemented a variant of the adaptive view synthesis scheme.[23] Fig. 9 shows the center view interpolation results in comparison to the ground truth center images. Without involving any advanced but complex method (e.g., boundary matte extraction[24]) in the interpolation process, our depth-image-based rendering technique leads to impressive interpolation results, as shown in Fig. 9(b, e). It is also found that our absolute difference maps are visually close (with even less residual energy for *Teddy*) to those maps obtained from a specifically designed image-based rendering approach,[25] where a loopy belief propagation method within a Markov Random Field framework is adopted for depth estimation, plus the boundary matting for high-quality view interpolation.

Table 2. Quantitative evaluation results for the new benchmark Middlebury stereo database.

Algorithm	*Tsukuba*			*Venus*			*Teddy*			*Cones*		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
Segm+visib[14]	1.30	1.57	6.92	0.79	1.06	6.76	5.00	6.54	12.3	3.72	8.62	10.2
Adapt.weights[3]	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26
MultiCamGC[19]	1.27	1.99	6.48	2.79	3.13	3.60	12.0	17.6	22.0	4.89	11.8	12.1
Our method	**2.29**	**2.88**	**8.94**	**0.80**	**1.11**	**3.41**	**10.5**	**15.9**	**21.3**	**6.13**	**13.2**	**13.3**
TensorVoting[21]	3.79	4.79	8.86	1.23	1.88	11.5	9.76	17.0	24.0	4.38	11.4	12.2
GenModel[20]	2.57	4.74	13.0	1.72	3.08	16.9	6.86	15.0	19.2	4.64	14.9	11.4



(a) (b) (c)

(d) (e) (f)

Figure 9. View interpolation results. (a) Ground truth center *Teddy* image. (b) Our interpolated center *Teddy* image. (c) Absolute difference image between (a) and (b). (d) Ground truth center *Cones* image. (e) Our interpolated center *Cones* image. (f) Absolute difference image between (d) and (e).

Discussions. In addition to its high functional performance in correspondence search, the proposed stereo method based on the LPA-ICI technique is also computationally efficient. On the one hand, the LPA-ICI technique is fast, because it is based on convolutions against one-dimensional kernels for a very limited number of directions (e.g., $K = 8$). On the other hand, as a local stereo method, without incurring any time-consuming iterative processing or global information propagation, the proposed method has a good potential to be optimized for a fast execution speed. Meanwhile, performing the computation locally is also a nice implementation advantage, because it can be greatly accelerated on the platforms with powerful parallel processing capability, e.g., modern graphics hardware.[6]

In comparison to the bilateral filter-based stereo method,[3] our proposed method shows a significant advantage in reducing the memory consumption during the execution. Rather than storing an adaptive support-weight for every pixel in a given window by 32-bit floating-point numbers,[3] the proposed stereo method only needs 8 bytes to store the adaptive-scales $\{h_X^+(\theta_k)\}_{k=1}^8$ for an entire support window. By looking up the pre-stored set of possible anisotropic neighborhood patterns, the proposed binary voting mask $\mathbf{M}_X$ can be easily computed online from $\{h_X^+(\theta_k)\}_{k=1}^8$. We only need to perform the pointwise OR-logic for each element in the window 7 (or $K-1$)

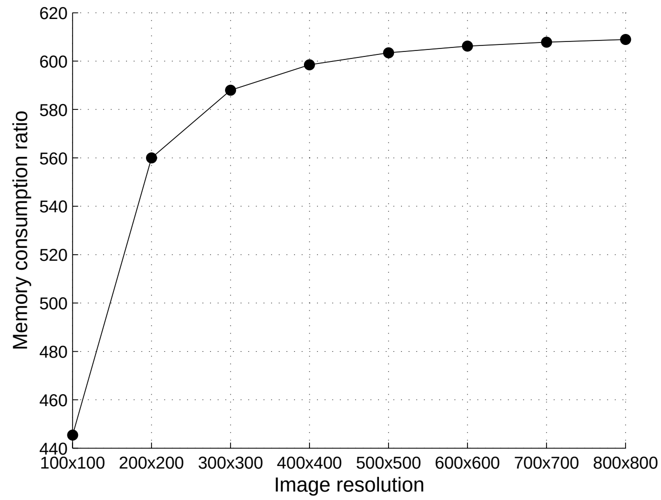


Figure 10. The ratio of the memory consumed by the bilateral filter-based stereo method[3] and that of our method.

times. To compare the memory consumption, we fix the local window size to 35×35, $K = 8$, and the dimension of the adaptive-scale set H to 7. Including the memory overhead of pre-stored tables, Fig. 10 shows that the proposed stereo method consumes significantly less memory than the bilateral filter-based stereo method[3] does.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel area-based stereo correspondence algorithm, yielding new insights to the realm of local stereo methods. Based on the anisotropic LPA-ICI technique, we cast the proposed local stereo matching and disparity refinement within a local high-confidence voting framework. By generating a pixel-wise combined local voting mask, the proposed method addresses the fundamental matching ambiguity problem in the key cost aggregation step. In addition, our disparity refinement scheme utilizing the anisotropic local neighborhood, also noticeably improves the accuracy of the estimated disparity map. Experiments on the standard stereo benchmarks show that our method produces accurate piecewise smooth disparity results, and it is highly ranked among the best performing local stereo methods. In particular, the proposed method preserves the depth discontinuities much better than other local methods, because our anisotropic neighborhoods adapt to the local image structures very well.

In our future work, we plan to research the complexity and estimation accuracy trade-off of the anisotropic LPA-ICI technique. The LPA-ICI analysis naturally accommodates this complexity-performance scalability; for instance, we can adjust the number of adaptive scales, as well as the value of each individual scale. To accurately compare the respective execution time of different local stereo methods, we also intend to implement our entire stereo algorithm and some other local methods within a unified software platform.

Based on this first trial in exploiting the recent progress in local approximation techniques for practical stereo matching problem, more research interests may be triggered to investigate the performance of other alternative structure-preserving adaptive signal filtering methods, different from the LPA-ICI presented in this paper.

REFERENCES

1. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int'l Journal of Computer Vision* **47(1)**, pp. 7–42, May 2002.
2. O. Veksler, "Fast variable window for stereo correspondence using integral images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, **vol. 1**, pp. 556–561, 2003.

3. K. J. Yoo and S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Machine Intell.* **28**(4), pp. 650–656, April 2006.
4. M. Gong and R. Yang, "Image-gradient-guided real-time stereo on graphics hardware," in *Proc. IEEE Conf. 3-D Digital Imaging and Modeling*, pp. 548–555, 2005.
5. J. Lu, S. Rogmans, G. Lafruit, and F. Catthoor, "Real-time stereo correspondence using a truncated separable Laplacian kernel approximation on graphics hardware," in *Proc. IEEE Int'l Conf. Multimedia and Expo*, pp. 1946–1949, 2007.
6. J. Lu, G. Lafruit, and F. Catthoor, "Fast variable center-biased windowing for high-speed stereo on programmable graphics hardware," in *Proc. IEEE Int'l Conf. Image Process.*, pp. 568–571, 2007.
7. T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiments," *IEEE Trans. Pattern Anal. Machine Intell.* **16**(9), pp. 920–932, Sept. 1994.
8. O. Veksler, "Stereo correspondence with compact windows via minimum ratio cycle," *IEEE Trans. Pattern Anal. Machine Intell.* **24**(12), pp. 1654–1660, Dec. 2002.
9. A. Fusiello, V. Roberto, and E. Trucco, "Efficient stereo with multiple windowing," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 858–863, 1997.
10. S. B. Kang, R. Szeliski, and C. Jinxiang, "Handling occlusions in dense multi-view stereo," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, **vol. 1**, pp. 103–110, 2001.
11. Y. Xu, D. Wang, T. Feng, and H.-Y. Shum, "Stereo computation using radial adaptive windows," in *Proc. IEEE Int'l Conf. Pattern Recognition*, **vol. 3**, pp. 595–598, 2002.
12. V. Katkovnik, K. Egiazarian, and J. Astola, *Local approximation techniques in signal and image processing*, SPIE Press, Monograph Vol. PM157, Sept. 2006.
13. A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.* **16**(5), pp. 1395–1411, May 2007.
14. M. Bleyer and M. Gelautz, "A layered stereo algorithm using image segmentation and global visibility constraints," in *Proc. IEEE Int'l Conf. Image Process.*, pp. 2997–3000, 2004.
15. E. P. Bennett and L. McMillan, "Video enhancement using per-pixel virtual exposures," in *Proc. ACM SIGGRAPH*, pp. 845–852, 2005.
16. D. Scharstein and R. Szeliski, *Middlebury stereo vision page.* http://vision.middlebury.edu/stereo/.
17. C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Trans. Pattern Anal. Machine Intell.* **22**(7), pp. 675–684, July 2000.
18. D. Scharstein and R. Szeliski, "Stereo matching with nonlinear diffusion," *Int'l Journal of Computer Vision* **28(2)**, pp. 155–174, 1998.
19. V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proc. European Conference on Computer Vision*, pp. 82–96, 2002.
20. C. Strecha, R. Fransens, and L. V. Gool, "Combined depth and outlier estimation in multi-view stereo," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, **vol. 2**, pp. 2394–2401, 2006.
21. P. Mordohai and G. Medioni, "Stereo using monocular cues within the tensor voting framework," *IEEE Trans. Pattern Anal. Machine Intell.* **28**(6), pp. 968–982, June 2006.
22. S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," in *ICCV*, 2007.
23. J. Lu, S. Rogmans, G. Lafruit, and F. Catthoor, "High-speed stream-centric dense stereo and view synthesis on graphics hardware," in *Proc. IEEE Int'l Workshop on Multimedia Signal Process.*, pp. 243–246, 2007.
24. C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video interpolation using a layered representation," in *Proc. ACM SIGGRAPH*, pp. 600–608, 2004.
25. C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *Int'l Journal of Computer Vision* **75(1)**, pp. 49–65, Oct. 2007.