Stereo Matching Algorithm Based on a Generalized Bilateral Filter Model

Li Li
Shandong Provincial Key Laboratory of Digital Media Technology, Shandong Economic University, Jinan, China
School of Computer Science and Technology, Shandong University, Jinan, China
Email: lily_jn@sina.com

CaiMing Zhang and Hua Yan
Shandong Provincial Key Laboratory of Digital Media Technology, Shandong Economic University, Jinan, China
Email: {czhang, yhzhjg} @sdu.edu.cn

***Abstract*—Stereo matching is a kernel problem in stereo vision systems. Stereo algorithms can be roughly classified into local and global approaches. Local algorithms use Winner-Take-All strategy, simply taking disparity level that minimizes the aggregation costs. In this paper we present a local stereo matching algorithm with an adaptive cost aggregation strategy based on a generalized bilateral filter model. The range weight computation in the original bilateral filter is extended by the inner and outer weighted average processes. A pixel is assigned a high range weight to the central pixel not only if the patches of the two pixels are similar but also if the neighbouring patches around the two pixels are similar. The final range weight could more accurately reflect the similarity of considering two pixels. Different cost aggregation methods can be easily derived from the model by modifying parameters. In experimental section we compare four cost aggregation methods based on the generalized model and give conclusions which demonstrate the effectiveness of our proposed strategy.**

***Index Terms*—stereo matching; adaptive cost aggregation; generalized bilateral filter model; range weight; inner and outer weighted average**

I. Introduction and Related Works

Stereo matching algorithms aim at finding disparity maps based on two or more images captured from the same screen. Disparity maps reveal the differences in location of corresponding pixels along the epipolar lines. A complete survey on stereo matching algorithms can be found in paper [1] and according to authors of the paper stereo matching algorithms can be classified into local and global approaches. Most of stereo matching algorithms consist of four steps: initial matching costs computation, initial cost aggregation, disparity maps computation or optimization and disparity maps refinement. For local methods the cost aggregation step is mandatory to increase the signal to noise ratio (SNR) and is often adopted by global ones. Cost aggregation step is to aggregate raw matching costs within a support window. An ideal support window should be adjusted according to image content to include only the pixels with the same disparities. Many cost aggregation methods have been presented while this behavior is far from ideal. This paper proposes an adaptive cost aggregation strategy based on a generalized bilateral filter model.

Bilateral filter is a non-iterative feature-preserving image smoothing technique and widely used in image denosing, computer vision and compute graphics areas. As we all know the linear filter does not only smooth the noise, but also important features. However bilateral filter assigns a geometric (spatial filter) and a color proximity constraint (range filter) independently, it can smooth the image noise while preserving edge features. Given an image I , the value of pixel p in the filtered image $\tilde{I}$ is a weighted average value described as follows

$$\tilde{I}(p) = \frac{\sum_{q\in S(p)} W_C(I(q)-I(p))W_S(q-p)I(q)}{\sum_{q\in S(p)} W_C(I(q)-I(p))W_S(q-p)} \quad (1)$$

where $S(p)$ is a support region centered in pixel p, W_S and W_C are weighting functions related to spatial distance D_S and range distance D_C between p and q respectively. The higher weight should be assigned to the pixels with the both smaller spatial and range distances to the central pixel. The support weights of each pixel vary according to image content, so the bilateral filter belongs to nonlinear filter and does not average across the edges.

Adaptive Weight method (AW) [2] firstly used the bilateral filter to aggregate raw matching costs and achieved excellent results in terms of accuracy. Different from the Adaptive Window methods varying the size, the shape and the position of the windows, the AW method adopted a fixed window and assigned weights for each pixel according to image contents. In Adaptive Weight method, the weight of a pixel within the support window is obtained by applying two independent bilateral filters in the neighborhood of potential correspondence. Given a pixel p_r in the reference image I_r and the potential corresponding pixel p_l in the matching image I_l with

doi:10.4304/jsw.6.10.1906-1913

disparity d , the aggregated cost $\tilde{C}(p_r, p_l, d)$ is computed as follows

$$\tilde{C}(p_r, p_l, d) = \frac{\sum_{\substack{q_r \in S(p_r) \\ q_l \in S(p_l)}} W_C(I_r(p_r) - I_r(q_r)) W_S(p_r - q_r) W_C(I_l(p_l) - I_l(q_l)) W_S(p_l - q_l) C(q_r, q_l, d)}{\sum_{\substack{q_r \in S(p_r) \\ q_l \in S(p_l)}} W_C(I_r(p_r) - I_r(q_r)) W_S(p_r - q_r) W_C(I_l(p_l) - I_l(q_l)) W_S(p_l - q_l)} \, . \quad (2)$$

where the initial matching cost $C(q_r, q_l, d)$ is a single pixel TAD (Truncated Absolute Differences) score between corresponding pixels q_r and q_l assuming the disparity value is d , the weighting functions W_S and W_C both are Gaussian, the spatial and range distances are Euclidean distances. The AW method provides excellent results in a WTA (Winner-Take-All) framework which are comparable to some global methods without any complex reasoning.

However the bilateral filter in itself has some limits such as poor smoothing in high gradient regions, smoothing and blunting cliffs, and time demanding and so on. Many modified methods against AW method have been proposed in recent years. There are generally three directions to study new algorithms. First, improve the disparity accuracy. For example, the Segment-based Support approach (SS) applies segment information [3]. The two bilateral filters assign weights 1 to those pixels belonging to the same segment of the central pixel and the range filter weights of (2) to those pixels outside the segment. By using segment information and removing the spatial filter weights, the SS method can further improve the accuracy of disparity maps. Second, improve the algorithm speed. For example, a simplified asymmetrical strategy against the AW method was proposed in the paper [4]. The bilateral filter is enforced on the reference image only and weights are computed by means of a two pass approach. These simplifications yield real-time implementation and worse but reasonable results compared with AW method. Third, trade off accuracy and speed. These methods often yield real-time or near real-time implementation and mainly include decreasing the algorithm complexity without degrading accuracy, code optimization for a given algorithm and designing algorithm on a more powerful hardware platform. For example, a Fast Bilateral Stereo method (FBS) combines the traditional local approaches with a symmetric adaptive weights strategy based on two independent bilateral filters applied on a regular block basis [5]. At each block a single weight is assigned to pixels within block. The spatial filter weight is that of the central pixel within block and the range filter weight is the average value of pixels within block. Disparity maps yielded by FBS method are, in general, less noisy compared to the AW method and one can trade accuracy for speed and vice versa by modifying the block size. Recently a stereo system based on hardware-aware algorithm is proposed in the paper [6]. The paper presents two stereo algorithms: exponential step size adaptive weight (ESAW) and exponential step size message propagation (ESMP). The ESAW can reduce computation complexity without sacrificing disparity accuracy by exponential step size aggregation. In this paper, our focus is along the first path.

Many other stereo matching methods related with bilateral filter have also been presented. Authors of paper [7] proposed a framework to derive a function that locally captures the plausibility of the underlying geometric and photometric constraints independently enforced by supports of neighboring points. Danny Barash [8] pointed out that the nature of bilateral filter resembles that of anisotropic diffusion. Bilateral filter and anisotropic diffusion can be related to adaptive smoothing and can be viewed as the generalization of the latter. So many stereo matching methods based on the PDE model have been presented [9-11]. These PDE-based methods are usually used in a variational framework and belong to global matching methods whose implementations are similar to adaptive weight methods.

This paper proposes a local stereo matching algorithm with an adaptive cost aggregation strategy based on a generalized bilateral filter model. In original bilateral filter the range weight is computed based on photometric differences between two pixels no considering the neighbouring pixels' similarities around the two pixels. Based on the above observation we firstly present a generalized bilateral filter model and enforce it in cost aggregation step. To improve robustness to noise and outliers, the modified truncated L1 norm is used as initial cost function. By adjusting parameters different cost aggregation methods can be derived from the proposed model. After cost aggregation the final disparity map can be obtained by a WTA strategy without any post processing steps. In our paper, there are three main distinguishing features as follows

First, we present a generalized bilateral filter model in which the range weight is extended by inner and outer weighted average processes to decrease the noise and outliers. In this model, the range weight can be more accurate to reflect the similarity of two considering pixels, so cost aggregation based on the model can be executed to achieve more accurate disparity maps especially fitting for real images having different source and light conditions confirmed by our experiments.

Second, the patch shape and size of inner and outer weighted average processes can be adjusted independently and different cost aggregation strategies can be easily derived from the model by modifying a few parameters fitting for different applications.

Third, the idea of the inner and outer weighted average processes can be extended and realized by different filter methods which can yield the other cost aggregation algorithms to trade off the accuracy and the speed of the system.

The rest of the paper is organized as follows: we give a detailed explanation for each part in Sections II and III; show some experimental results to compare different cost aggregation methods in Section IV and conclusions and future work are given in Section V.

II. Initial Matching Costs

In Adaptive Weight method, TAD score between two potential corresponding pixels is computed as initial matching cost. The TAD matching cost is given by

$$u(p_r, p_l)) = | I_r(p_r) - I_l(p_l) |$$
$$C(p_r, p_l, d) = \min(u(p_r, p_l), T) \quad (3)$$

where $p_r(x, y)$ and $p_l(x-d, y)$ are potential corresponding pixels in the reference image I_r and matching image I_l respectively with disparity d, T is the predetermined threshold parameter. For color image absolute difference u is the summation of that of three channels in RGB color space (different from the original AW method with CIELab color space) used in our compared experiments. In order to better handle the outliers and noise, the truncated L1 norm is employed as matching cost function in this paper

$$C(p_r, p_l, d) = -\log[\delta_M + (1-\delta_M)\exp(-\frac{u}{\sigma_M})] \quad (4)$$

where δ_M and σ_M are predetermined parameters. In Section IV comparison experiments are held on the performances of two cost functions (3) and (4).

III. Proposed Adaptive Cost Aggregation

A. A generalized bilateral filter model

In original bilateral filter, the range weight W_C is the function of single differences between pairs of connected pixels in (1). The single difference has a limited ability to express the similarity of related two pixels and the information of their neighborhoods should be included in rang weight computation. So we modify the bilateral filter formulation and range weight W_C is computed based on patch similarity around two connected pixels. In this section we explain the generalized bilateral filter model and apply it in next cost aggregation step.

Firstly the patch range distance D_C is defined as below

$$D_C(p_i, p_j) = (\sum_m G_\sigma(m) u(p_{i+m}, p_{j+m})) / \sum_m G_\sigma(m) \quad (5)$$

where the denominator is the normalization coefficient by the sum of all applied weights, u is the TAD score of two corresponding pixels and given by (3), p_{i+m} and p_{j+m} are the neighborhood of p_i and p_j respectively, $|m| < N$ and N is the size of patch around p_i and p_j, the size and shape of the patch of p_i and p_j are all the same, $G_\sigma(m) = \exp(-|m| / 2\delta_\sigma)$ is a Gaussian function with the radius δ_σ which represents the patch size. So the distance is not the single difference between two related pixels and is a weighted average value of pixels in the patches around the two pixels to decrease the noise influence. Then the range weight is computed by

$$W_C^1(p_i, p_j) = \exp(-D_C(p_i, p_j) / 2\sigma_C) \quad (6)$$

where W_C^1 is a Gaussian function and converts the range distance D_C to the similarity weight. The value W_C^1 represents the similarity of two patches around p_i and p_j respectively. We can further sum up the patch similarity weights of their corresponding neighbors to remove the noise and outliers. Then the range weight W_C^1 is modified as follows

$$W_C^3(p_i, p_j) = (\sum_k G_\rho(k) W_C^1(p_{i+k}, p_{j+k})) / \sum_k G_\rho(k) \quad (7)$$

where $G_\rho = \exp(-|k| / 2\delta_\rho)$ is a similar function as G_σ in (5). When we substitute the equations (5) and (6) into (7), (7) can be rewritten as below

$$W_C^3(p_i, p_j) = (\sum_k G_\rho(k) \exp[-((\sum_m G_\sigma(m) u(p_{i+k+m}, p_{j+k+m})) / \sum_m G_\sigma(m)) / 2\sigma_C]) / \sum_k G_\rho(k) \quad (8)$$

From the above equation we observe that there are two weighted average processes. For pixels p_i and p_j, we compute the patch distances of all patches at positions p_{i+k} and p_{j+k} taken with the offset k around p_i and p_j respectively. Each patch distance is computed as the weighted average value of initial matching costs for each pair pixels and controlled by G_σ. This is called the inner weighted average process. Then the patch distance is transformed into patch similarity weight by function (6). We then compute the average of these patch similarity weights as the finial range weight between p_i and p_j. This is called the outer weighted average process controlled by G_ρ. Thus the pixel p_j will contribute to the pixel p_i with a high weight not only if the patches around p_i and p_j are similar, but also if the neighboring patches p_{i+k} and p_{j+k} resemble each other.

Then the bilateral filter equation (1) can be modified into

$$\tilde{I}(p) = \frac{\sum_{q \in S(p)} W_C^3(I(q) - I(p)) W_S(q-p) I(q)}{\sum_{q \in S(p)} W_C^3(I(q) - I(p)) W_S(q-p)} \quad (9)$$

In the above equation, the original range weight W_C is replaced with the new range weight W_C^3. We call this equation a generalized bilateral filter model. Different cost aggregation strategies can be derived from the model by modifying two parameters δ_σ and δ_ρ respectively.

B. Different cost aggregation strategies

Based on the generalized bilateral filter model (9), the final cost aggregation equation is expressed by

$$\tilde{C}(p_r, p_l, d) = \frac{\sum_{\substack{q_r \in S(p_r) \\ q_l \in S(p_l)}} W_C^3(p_r, q_r) W_S(p_r, q_r) C(q_r, q_l, d)}{\sum_{\substack{q_r \in S(p_r) \\ q_l \in S(p_l)}} W_C^3(p_r, q_r) W_S(p_r, q_r)} \quad (10)$$

where p_r and p_l are potential corresponding pixels in the reference image I_r and the matching image I_l respectively with disparity d, $S(p_r)$ and $S(p_l)$ are the support windows around p_r and p_l respectively, $W_S = \exp(-|p_r - q_r|/2\sigma_S)$ is computed by Gaussian function based on the spatial distance same with (2), W_C^3 is computed by (8). In the above equation, we simply execute the filter on the reference image only and the method can be extended to symmetric filter easily. We call the above equation a cost aggregation strategy based on a generalized bilateral filter (symbolized as CA_GBF).

From the generalized bilateral filter model (9), other cost aggregation methods can be obtained too. First, let $\delta_\rho \to 0$, leading to the following outer weighting function

$$G_\rho = \begin{cases} 1 & if \;\; k = 0 \\ 0 & if \;\; k \neq 0 \end{cases} . \qquad (11)$$

The equation (10) then simplifies to

$$\tilde{C}(p_r, p_l, d) = \frac{1}{M_{pr,pl}} \sum_{\substack{q_r \in S(p_r) \\ q_l \in S(p_l)}} W_C^1(p_r, q_r) W_S(p_r, q_r) C(q_r, q_l, d) . \qquad (12)$$

where the range weight W_C^3 is replaced with W_C^1 expressed by (6), the normalization denominator is symbolized as $M_{pr,pl}$. The above equation only considers the inner weighted average process and only the patch similarity of two neighboring pixels p_r and q_r is computed here. The equation (12) is called a cost aggregation strategy based on the inner weighted filter (symbolized as CA_IWF).

Second, let $\delta_\sigma \to 0$, the range weight W_C^3 in (10) is replaced by

$$W_C^2(p_i, p_j) = (\sum_k G_\rho(k) W_C(p_{i+k}, p_{j+k})) / \sum_k G_\rho(k) . \qquad (13)$$

where W_C is the weighting function based on the single pixel difference similar with the function in (2). The range weight is computed only considering the outer weighted average process and the patch similarity simplifies to the similarity of two considering pixels.

The equation (10) then simplifies to

$$\tilde{C}(p_r, p_l, d) = \frac{1}{M_{pr,pl}} \sum_{\substack{q_r \in S(p_r) \\ q_l \in S(p_l)}} W_C^2(p_r, q_r) W_S(p_r, q_r) C(q_r, q_l, d) . \qquad (14)$$

We call the above equation a cost aggregation strategy based on the outer weighted filter (symbolized as CA_OWF). As a third example, let both $\delta_\rho \to 0$ and $\delta_\sigma \to 0$, then the generalized model simplifies to the original bilateral filter and leads to

$$\tilde{C}(p_r, p_l, d) = \frac{1}{M_{pr,pl}} \sum_{\substack{q_r \in S(p_r) \\ q_l \in S(p_l)}} W_C(p_r, q_r) W_S(p_r, q_r) C(q_r, q_l, d) . \qquad (15)$$

where $W_C(p_r, q_r) = \exp(-u(p_r, q_r)/2\sigma_C)$ and here the range weight is computed based on the single pixel difference no considering our proposed two weighted average processes. Similarly we call (15) a cost aggregation strategy based on the asymmetric bilateral filter (symbolized as CA_ABF).

C. Disparity map computation

After the cost aggregation step, the final disparity map is obtained in a WTA framework without any post processing

$$D(p_r) = \arg\min_{d \in D}(\tilde{C}(p_r, p_l, d)) . \qquad (16)$$

where D represents the set of all allowed disparities.

IV. Experimental Results

This section we aim at assessing the performance of our proposed cost aggregation strategy based on a generalized bilateral filter model. We use the Middlebury stereo benchmark provided by authors of the paper [1] to evaluate performances of different cost functions (described by (3) and (4)) and different cost aggregation methods derived from our model (that is CA_GBF, CA_IWF, CA_OWF and CA_ABF) and compare the performances of our approach with the other state-of-art cost aggregation strategies.

For cost function comparison, the parameters are set as threshold 40 for the TAD cost function (3), $\delta_M = 10^{-7}$ and $\sigma_M = 2$ for the truncated L1 norm cost function (4). The two cost functions both are used in the CA_ABF method expressed by (15). Parameters of cost aggregation are $\sigma_C = 15$, $\sigma_S = 10.5$ and support window size 21×21. The corresponding disparity maps of the Middlebury test images (Tsukuba, Venus, Teddy and Cones) are plotted in Fig. 1. By the Middlebury stereo benchmark we compute the percentage of bad pixels (i.e. pixel whose absolute disparity error is greater than 1) for pixels in non-occluded areas, pixels in occluded areas, pixels near depth discontinuities described by authors [1]. For pixels in occluded areas, the results of four images all are better using the L1 norm function than the TAD function, which manifests the L1 norm function more robust to outliers.

In cost aggregation comparison experiments, we adopt a constant parameter setting across four test images: aggregation support window size $= 21 \times 21$, $\delta_\sigma = 1.5$, $\delta_\rho = 1.5$, $\sigma_C = 15$, $\sigma_S = 10.5$. It is worth noting that the patch shape for the inner and outer weighted average processes in our method is a rectangle window for simplicity and the patch size is 3×3 to reduce the computational time. The L1 norm is taken as the cost function with the same parameters as the first experiment. These parameters have been found empirically. Different parameters δ_σ and δ_ρ are shown in Table I for different cost aggregation methods. Quantitative comparative results for different cost aggregation methods are given in Table II. The focus here is the evaluation of the raw cost aggregation methods which do not deal explicitly with occlusions and we only report the percentage of bad pixels for pixels in non-occluded areas (Vis.) and near depth discontinuities (Dis.). From the table, we can find that for pixels in non-occluded areas the results of the

CA_OWF method are almost the best of the four methods. Compared with CA_ABF, the CA_OWF method

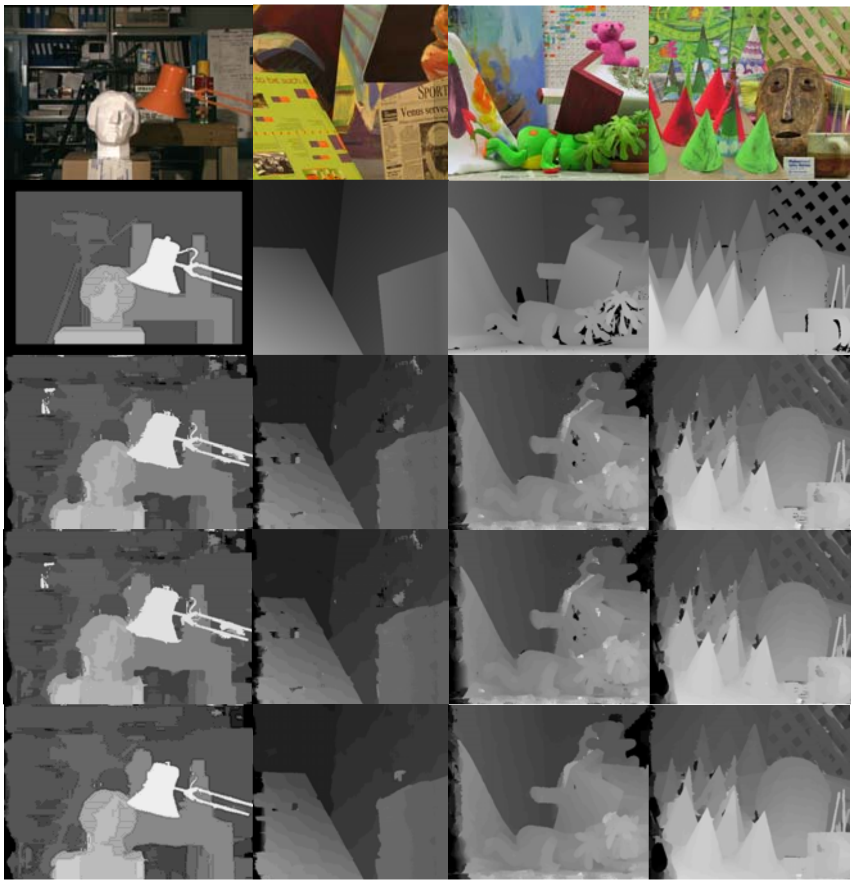

Figure 1. Disparity maps for the "Tsukuba", "Venus", "Teddy" and "Cones" images. The top is the reference images. The second row is the truth disparity maps. The third row is the results using L1 norm function. The fourth row is the results using the TAD cost function. The last row is our optimal results using the CA_OWF method.

TABLE I.
PARAMETERS OF OUR COST AGGREGATION METHODS. THE BLANK FIELDS MEAN NO PARAMETERS SET BY CORRESPONDING METHODS..

Methods / Parameters	CA_ABF	CA_IWF	CA_OWF	CA_GBF
δ_σ		1.5		1.5
δ_ρ			1.5	1.5

decreases the Vis. and Dis. errors more for Teddy image and Venus image than errors for Tsukuba image and Venus image, mainly because of the weighted average process increasing the Dis. errors for latter two images. The CA_GBF method creates the best results of the whole Vis.+Dis. error among four cost aggregation methods. This manifests that the two weighted average processes can effectively eliminate noise in non-occluded areas while well preserving depth discontinuities. However the results of the CA_IWF method could even become worse than that of CA_ABF. Because the inner and outer weighted average processes adopt the Gaussian functions which smooth the data and blur edges simultaneously. This experiment also shows that the effect of the inner weighted average is different from that of the outer weighted average. The decreasing of the error mainly depends on the outer weighted average process for the patch similarity weights. So the following experiments adopt the CA_OWF method to compare the performances of our strategy with others. On the other hand, from the experimental results, we can see that images have different change rules when they use different cost aggregation methods because of their different features.

The weighted average process for range weight computation in our method can decrease noise influence fitting for real images which usually have different light source or distortions. To manifest this, we have the same comparison experiments on other two datasets which are available at Middlebury test bed. Each dataset has 9 different images that exhibit 3 different exposure and 3 different lighting variations. Fig. 2 shows the both exposure and lighting variations of the left image of the Art dataset. We compare the CA_OWF method with the method based on the original bilateral filter (that is the CA_ABF method). The same parameters as the above experiment are adopted expect of the $\sigma_C = 10$ for the Art dataset in this experiment. The quantitative comparison results are given in Table III. From the table, we can see that almost all the results in the non-occluded and near discontinuity areas of CA_OWF method are better than that of CA_ABF method. The outer weighted average process can effectively smooth non-occluded areas while

Figure 2. The left image of the Art dataset with three different exposures and under three different light conditions.

preserving well depth discontinuities. The disparity maps of the Art dataset are plotted in Fig. 3.

To evaluate our proposed cost aggregation method, we compare the results of our approach with the state-of-art cost aggregation strategies. In our method, we adopt the CA_OWF method with optimal parameters minimizing the Vis.+Dis. error on the whole dataset: window size $=31\times 31$, $\delta_\rho=1.5$, $\sigma_C=10$, $\sigma_S=15.5$. We have reported in Table IV the results obtained by our method and the other five top performing cost aggregations strategies [2,3,5,12,13]. It is worth noting that these results reported by authors [4] are obtained by using the original cost function proposed by authors of each paper and the results for AW and SS available on the Middlebury evaluation site including the post processing steps are not used. From the Table IV, we can see that our proposed method has accuracy comparable to the best performing cost aggregation strategies. The AW method outperforming the accuracy of our method is mainly due to the symmetric strategy used by it while asymmetric strategy by our method. However our method runs faster than the AW method and the AW run is 3226 seconds according to authors [4] while our method takes 320 seconds without any accelerating techniques on Teddy. Furthermore our method can decrease the noise influence due to our weighted average process. The results of our proposed method are plotted in Fig. 1.

TABLE II.
QUANTITATIVE COMPARISON RESULTS OF FOUR COST AGGREGATION METHODS FOR FOUR TEST IMAGES.

Methods	Tsukuba		Venus		Teddy		Cones	
	Vis.	*Dis.*	*Vis.*	*Dis.*	*Vis.*	*Dis.*	*Vis.*	*Dis.*
CA_ABF	3.77	11.73	5.21	15.61	13.69	25.43	10.32	20.60
CA_OWF	3..51	12.15	4.84	15.85	13.09	24.91	9.03	18.83
CA_IWF	3.93	11.51	5..52	14.66	13.36	24.44	10.50	20.58
CA_GBF	3.72	11.68	4.97	14.11	13.07	24.17	9..90	19.61

TABLE III.
QUANTITATIVE COMPARATIVE RESULTS OF THE CA_OWF METHOD WITH THE CA_ABF METHOD FOR REAL TEST IMAGES.

		CA_ABF method		CA_OWF method	
		Vis.	*Dis.*	*Vis.*	*Dis.*
Art Dataset	Art1-1	17.00	24.39	16.65	24.39
	Art1-2	18.25	20.52	17.30	19.80
	Art1-3	18.31	22.53	17.35	22.03
	Art2-1	16.22	21.92	15.82	22.04
	Art2-2	13.41	18.25	12.59	17.60
	Art2-3	17.18	22.09	16.40	21.72
	Art3-1	16.24	17.00	15.62	16.50
	Art3-2	17.28	16.75	16.36	15.85
	Art3-3	28.60	24.70	27.28	24.21
Book Dataset	Book1-1	16.26	27.49	15.15	26.37
	Book1-2	17.49	29.92	15.76	27.99
	Book1-3	25.02	39.43	23.62	37.95
	Book2-1	18.79	31.89	17.86	31.76
	Book2-2	22.55	33.86	22.64	32.22
	Book2-3	19.74	31.81	17.46	30.45
	Book3-1	14.60	27.16	13.94	27.31
	Book3-2	19.41	30.17	17.52	28.83
	Book3-3	18.26	34.01	17.31	33.26

V. Conclusions

We have proposed a cost aggregation strategy based on a generalized bilateral filter model. The model extends the range weight computation of the bilateral filter in two steps: the inner weighted average process of the pixels' range distances in the patch and the outer weighted average process of the patch similarity weights. The range weight will be high not only if two pixels have similar patches but also if their neighboring pixels have similar patches. So the range weights in our model can accurately reflect two pixels' similarity. And the weighted average process can effectively remove noise and outliers in real images. By modifying the inner and outer parameters, different cost aggregation strategies are easily derived from the model. On the other hand the weighted average process can be realized by different filter methods and yield other cost aggregation strategies fitting for different applications. Within the evaluation framework for cost aggregation strategies, experimental results confirm the effectiveness of our proposed method.

In the future we plan to adopt efficient accelerating calculations to decrease the computational time of our method. We are also interested in analyzing resemblance between bilateral filter and PDE model to further improve the accuracy of bilateral filter method.

Acknowledgment

The authors would like to thank financial supports from National Natural Science Foundation of China under Grant Nos. 60970048, Natural Science Foundation of Shandong Province Grant Nos. 2009ZRB019SF.

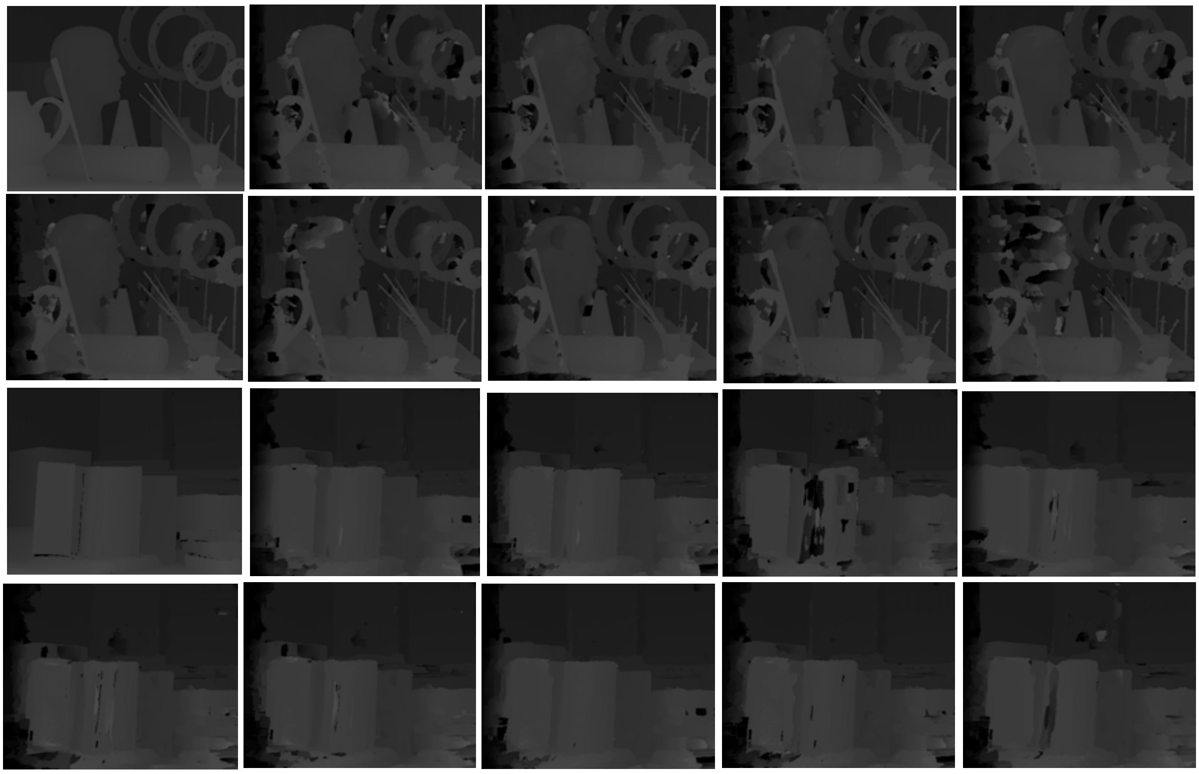

Figure 3. The disparity maps of the CA_OWF method for the Art and Book datasets. The first image of the two datasets is truth disparity maps respectively for the Art and Book datasets.

TABLE IV.
Quantitative comparison results of our method with the five top performing cost aggregation methods for four test images.

Methods	Tsukuba		Venus		Teddy		Cones	
	Vis.	*Dis.*	*Vis.*	*Dis.*	*Vis.*	*Dis.*	*Vis.*	*Dis.*
SS[3]	2.19	7.22	1.38	6.27	10.5	21.2	5.83	11.8
AW[2]	3.33	8.87	2.02	9.32	10.52	20.84	3.72	9.37
FBS[5]	2.95	8.69	1.29	7.62	10.71	20.82	5.23	11.34
SB[12]	2.25	8.87	1.37	9.4	12.7	24.8	11.1	20.1
VW[13]	3.12	12.4	2.42	13.3	17.7	25.5	21.2	27.3
Our Method	2.34	10.0	3.4	13.68	12.44	25.0	7.61	16.86

REFERENCES

[1] D. Scharstein and R. Szeliski. “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms”, *International Journal of Computer Vision*, 2002, pp. 7-42.
[2] K. J. Yoon and I. S. Kweon, “Adaptive support-weight approach for correspondence search”, *PAMI*, 2005, 28(4), pp. 650-656.
[3] F. Tombari, S. Mattoccia, and L. Di Stefano “Segmentation-based adaptive support for accurate stereo correspondence”, *In Proc. Of PSIVT 2007*, 2007.
[4] M. Gong, R.G. Yang, W. Liang, and M.W. Gong. “A performance study on different cost aggregation approaches used in real-time stereo matching”, *Int. Journal Computer Vision*, 75(2), pp. 283–296, 2007.
[5] S. Mattoccia, S. Giardino, and A. Gambini. “Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering”, *In Proc. Of ACCV 2009*, 2009.
[6] Y. Wei, T. Chen, F. Franchetti, and J. C. Hoe. “High performance stereo vision designed for massively data parallel platforms”, *IEEE Transactions on Circuits and Systems For Video Technology*, 2010.
[7] S. Mattoccia. “A locally global approach to stereo correspondence”, *In Proc. Of ICCV Workshop 2009*, pp. 1763-1770, 2009.
[8] Barash D. “Bilateral Filtering and Anisotropic Diffusion: Towards a Unified Viewpoint”, *Hewlell-Packard Laboratories Technical Report*, HPL-2000-18(R.1), 2000.
[9] D. Scharstein and R. Szeliski. “Stereo matching with nonlinear diffusion”, *Int. J. of Computer Vision*, 28(2), pp. 155–174, 1998.
[10] R.B. Ari and N.A. Sochen. “Variational stereo vision with sharp discontinuities and occlusion handling”, *In Proc. of ICCV 2007, Rio de Janeiro, Brazil*, IEEE Computer Society Press, pp. 1–7, 2007.
[11] H. Zimmer, A. Bruhn, L. Valgaerts, M. Breuß, J. Weickert, B. Rosenhahn, and H.P. Seidel. “PDE-based anisotropic disparity-driven stereo vision”, *Vision, Modeling, and Visualization 2008: Proceedings*, October 8-10, 2008.
[12] M. Gerrits and P. Bekaert. “Local stereo matching with segmentation-based outlier rejection”, *In Proc. of CRV 2006*, pages 66–66, 2006.
[13] O. Veksler. “Fast variable window for stereo correspondence using integral images”, *In Proc. of CVPR 2003*, pp. 556–561, 2003.

Li Li works as an instructor in the School of Computer Science and Technology at the Shandong Economic University. She received the B.E. degree in Motor Engineering from Shandong Technology University in 1998 and M.E. degree from Shandong University in 2000, and now is currently a Ph.D. student in the School of Computer Science and Technology at the Shandong University. Her research interests lie in computer vision, especially in 3D reconstruction and stereo matching.

Hua Yan works as an associate professor in the School of Computer Science and Technology of Shandong Economic University. She received the B.S. degree in Physics in 1997, M.E. degree in Communication and Information System in 2004 and D Sc Tech degree in Communication and Information System in 2007 from Shandong University. Dr. Yan’s research interests include image and video processing, multimedia data retrieval and super resolutions.