Stereo Similarity Metric Fusion Using Stereo Confidence

Gorkem Saygili Laurens van der Maaten Emile A. Hendriks
Computer Vision Lab, Delft University of Technology, The Netherlands
Email: {g.saygili, l.j.p.vandermaaten, e.a.hendriks}@tudelft.nl

***Abstract*—Stereo confidence measures are one of the most popular research topics in stereo vision. These measures give an indication about the certainty of the matching. The main aim of using confidence measures is to filter the erroneous disparity estimations at the end of the matching process. However, they can also be incorporated at the initial step of the matching process to obtain accurate estimations before the cost aggregation. In this paper, we propose to utilize stereo confidence measures for fusing different similarity measures in order to obtain robust estimations for aggregation. Since stereo similarity measures perform differently in varying conditions, the confidence-guided fusion of them makes stereo matching more robust against errors. We evaluate the performance of our algorithm in comparison to different similarity measures on the Middleburry benchmark stereo test set. The results show significant improvements on the accuracy of initial disparity estimations with our fusion strategy compared to different similarity measures.**

I. Introduction

Stereo matching has been an extensively researched topic of computer vision especially with the introduction of the benchmark Middleburry stereo dataset [1] in the last decade. The stereo algorithms can be grouped into two classes, namely, global and local stereo matching. Global stereo matching algorithms provide accurate estimations with their global smoothness assumptions but they are computationally expensive [2], [3], [4]. In contrast, local stereo algorithms are not as computationally intensive as global algorithms but the quality of the disparity estimations can be lower because of their local inference [5], [6]. For all algorithms in both groups, cost initialization is a significant step.

In general, stereo algorithms are composed of four steps: cost initialization, cost aggregation, disparity selection and refinement. The cost initialization is performed using stereo similarity measures at pixel level. There exists a variety of different similarity measures that can be used in stereo and provide different performances at different conditions [7]. In cost aggregation step, the initial matching costs, called disparity space image (DSI), are aggregated over a neighbourhood of pixels to increase the accuracy of the measurements. At the third step, the disparity of the pixels are generally selected by using the winner-take-all algorithm. At the final step, different filters such as median and mean are used to filter out the wrong matches.

In order to improve stereo matching, recent research focuses on improving the accuracy and speed of the cost aggregation step [6], [8], [9]. However, improving the initial DSI quality can also increase the accuracy of the final disparity estimations. Klaus *et al.* [10] fused gradient and intensity information linearly in order to obtain a more robust initial DSI for further processing. Mei *et al.* [11] fused DSI of census and color similarities in order to increase the accuracy of the final estimation. While increasing the matching accuracy significantly, the aforementioned algorithms do not provide a parameter-free method to fuse any number of similarity measures. To fuse any number of similarity measures without any parameters that depends on the measures, we need to calculate the performance of different similarity metrics at different locations at the image. After the performance of each similarity measure is known, any number of similarity measures can be fused adaptively with respect to their confidence.

Stereo confidences are incorporated to measure the confidence of the matching scores. As the confidence gets higher at a pixel, the probability of that pixel to have a correct disparity increases. An extensive evaluation of stereo confidences can be found in [12]. The stereo confidences are generally used for filtering wrong disparity estimations in stereo matching algorithms [13], [14], [15], [16]. We incorporate stereo confidences in the adaptive fusion of multiple similarity measures. Since stereo confidence measures indicate the certainty of the matching, the weights of each similarity measure in adaptive fusion can be chosen proportional to their confidence measure.

In this work, we propose a novel fusion strategy based on stereo confidence and consensus of matched pixels to fuse any number of stereo similarity measures to improve the accuracy of matching. Each pixel is first matched to a pixel in the target image using each one of different similarity measures. The confidence of the matching is calculated using Left Right Difference (LRD) metric which is one of the top performer stereo confidence metrics in [12]. We build a consensus for each pixel based on the confidences and winner disparities for each similarity measure of a local neighbourhood. The similarities that provide the winner of the consensus are fused by confidence-based weighting. While providing significant increase on the accuracy, our algorithm can fuse any number of similarity measures, requires no additional parameters and can provide different combinations of similarity measures for different regions of the image. To our knowledge, we are the first to propose such an algorithm in the stereo literature.

The similarity measures that are used in our experiments and the details about LRD is described in Section 2. In section 3, we describe our fusion algorithm. We evaluate the performance of our algorithm in Section 4. In Section 5, we draw our conclusions.

II. STEREO CONFIDENCE AND SIMILARITY MEASURES

Our algorithm starts with calculating the similarity measures. As the next step, the confidence of each similarity measure is calculated using LRD. At the last step, the similarity measures are fused using a novel confidence and consensus based similarity fusion algorithm.

In this section, we describe the similarity and confidence measures that we incorporated in our experiments. In the next section, we will introduce our fusion algorithm.

A. Similarity Measures

The DSI consist of cost measures $C_(x, y, d)$ that measures the penalty for assigning disparity d to the image location (x, y). This penalty can be calculated using different metrics. Below, we shortly describe our implementation of seven different cost measures that are often used in stereo [17], [7]: (1) absolute intensity difference (AD), (2) Rank, (3) Census, (4) Normalized Cross Correlation (NCC), (5) Zero-mean Normalized Cross Correlation (ZNCC), (6) Sobel, (7) Laplacian of Gaussian (LoG).

AD measures the absolute intensity difference between the reference (left) image $I(x, y)$ and the target (right) image $I'(x-d, y)$:

$$C_{AD}(x, y, d) = |I(x, y) - I'(x-d, y)|. \tag{1}$$

Rank [18] transform is a non-parametric image transform that models the structure of the neighbourhood of pixels by exploiting the intensity variation. Eq. 2 represents the Rank transform $RT(x, y)$ of a pixel (x, y) inside a local neighbourhood $N(x, y)$ of size 7×7 and pixel-based DSI, $C_{RT}(x, y, d)$ as:

$$\begin{aligned} &RT(x, y) = |\forall (x', y') \in N(x, y)| I(x', y') < I(x, y)| \\ &C_{RT}(x, y, d) = |RT(x, y) - RT'(x-d, y)|. \end{aligned} \tag{2}$$

Census [18] transform models the structure of 7×7 neighbourhood of pixels that is denoted by k, as represented in Eq. 3. Census is one of the most robust similarity measure against radiometric differences between stereo pairs [7] and it is calculated as:

$$\begin{aligned} &CT(x, y)[k] = \begin{cases} 1, & \text{iff } I(x_k, y_k) > I(x, y) \\ 0, & \text{otherwise,} \end{cases} \\ &\mu_c(x, y, d)[k] = \begin{cases} 1, & \text{iff } CT(x, y)[k] = CT'(x-d, y)[k] \\ 0, & \text{otherwise,} \end{cases} \\ &C_{CT}(x, y, d) = \sum_{\forall k} \mu_c(x, y, d)[k]. \end{aligned} \tag{3}$$

The pixel-based DSI of Census transformed images calculated by using Hamming distance [18].

NCC is an intensity and patch based matching method that is especially robust against Gaussian noise between the matched patches. For simplified notation, let I_p and I_{p-d} denote the pixels at (x, y) and $x-d, y$ respectively. Eq. 4 presents the pixel-wise DSI calculation using NCC:

$$C_{NCC}(p, d) = \frac{\sum_{p' \in N_p} I_{p'} I'_{p'-d}}{\sqrt{\sum_{p' \in N_p} I^2_{p'} \sum_{p' \in N_p} I'^2_{p'-d}}}. \tag{4}$$

ZNCC is similar to NCC whereas it provides more robustness against gain and offset variation between matched image patches [17]:

$$C_{ZNCC}(p, d) = \frac{\sum_{p' \in N_p} (I_{p'} - \bar{I}_p)(I'_{p'-d} - \bar{I}'_p)}{\sqrt{\sum_{p' \in N_p} (I_{p'} - \bar{I}_p)^2 \sum_{p' \in N_p} (I'_{p'-d} - \bar{I}'_p)^2}}.$$

For calculating NCC and ZNCC, we choose a patch size, $|N_p|$ of 5×5.

Sobel can suppress the noise in the intensity images. Let $I_s(x, y)$ denote sobel filter of size 3×3 response of image I at pixel (x, y):

$$C_{SB}(x, y, d) = |I_s(x, y) - I'_s(x-d, y)|$$

LoG can suppress the noise and provide robustness against offset in intensities. Similar to [7], we incorporated LoG kernel with size 5×5 and standard deviation of 1:

$$\begin{aligned} &L(x, y) = \frac{-1}{\pi \sigma^4} \left(1 - \frac{x^2 + y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \\ &I_{LoG}(x, y) = I(x, y) \otimes L(x, y) \\ &C_{LoG}(x, y, d) = |I_{LoG}(x, y) - I'_{LoG}(x-d, y)|. \end{aligned}$$

AD, Rank, Census, Sobel and LoG are calculated pixel-wise, however NCC and ZNCC are calculated over a neighbourhood with intrinsic aggregation as depicted in Eq. 4 . In order to compensate this difference, the pixelwise costs are aggregated over 3×3 windows. All of the costs are normalized to have values in $[0, 1]$ before confidence estimation and fusion in order to prevent range difference in between.

B. Confidence Measure

In order to fuse multiple similarity measures, the performance of them should be measured. The confidence measures give an indication of how sure the similarity measure is on a particular location and can be used in fusion processeseqnarray.

18 confidence measures have been tested in [12]. LRD is among the top performing confidence measures because of its cross control over the left and right matching scores. Let d_1, c_1, and c_2 be the winner disparity, minimum, and second minimum costs respectively, LRD confidence, $S(x, y)$, can be calculated as:

$$S(x,y) = \frac{c_2 - c_1}{|c_1 - \min_{d'}(c'(x-d_1,y,d'))| + \epsilon}, \tag{5}$$

where $c'(x-d_1,y,d')$ denotes target to reference cost and d' represents the disparities from target to reference. As the difference between c_1 and $\min_{d'}(c'(x-d_1,y,d'))$ decrease, the confidence increases. We would expect that they would be equal if there is no error and no occlusion between the two images. Therefore, we add ϵ to the denominator to ensure we do not obtain a zero denominator.

III. Fusion of Similarity Measures

We propose a consensus-based disparity voting algorithm to fuse similarity measures, i, using LRD. Let (x_n, y_n) be the pixels around a neighbourhood $N(x,y)$ with size h_w of a pixel (x,y) and let d_n be the winner disparity of (x_n, y_n), the consensus, $U(x,y,d)$, of the pixel can be found as:

$$\mu(d_n,d) = \begin{cases} 1, & d_n = d \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$
$$U(x,y,d) = \sum_{\forall i} \sum_{(x_n,y_n)\in N(x,y)} S_i(x_n,y_n)\mu(d_n,d).$$

$U(x,y,d)$ denotes the consensus of candidate disparities for the pixel (x,y) based on a weighted sum of the confidence of every similarity measures, $S_i(x_n,y_n)$. The disparity that has the highest consensus, d^*, is chosen in order to fuse the similarity measures:

$$d^* = \operatorname*{argmax}_d (U(x,y,d)). \tag{7}$$

For each similarity measure i, the DSI of a pixel with the highest confidence having a disparity of d^* in the neighbourhood of (x,y) is chosen as the new DSI of (x,y) which is denoted as $C_i(x,y,d)$. At the next step, the DSI of all of the similarity measures $C_i(x,y,d)$ are aggregated using weighted average with respect to their confidence:

$$\alpha_i = \frac{S_i(x,y)}{\sum_{\forall i} S_i(x,y)}$$
$$C(x,y,d) = \sum_{\forall i} \alpha_i C_i(x,y,d), \tag{8}$$

where, $C(x,y,d)$ is the fused DSI measure for the pixel at (x,y).

Described fusion strategy can fuse any similarity measure and does not require any additional parameter except the consensus window size, h_w.

The proposed fusion strategy is light in terms of computational complexity. Let H, W, D and I denote the height, width, range of disparities, and number of incorporated similarity measures respectively. Measuring the confidence, LRD, requires $H \times W \times D$ computations. At the next step, calculating the consensus around the neighbourhood of size h_w requires $H \times W \times h_w \times I$ computations. Since h_w is chosen as 3×3 in all of our experiments, consensus can be measured significantly faster than the aggregation step. Finally, aggregating the similarity measures requires $H \times W \times I$ computations which is negligible since the total number of similarities, I, is small.

IV. Experiments

To evaluate the performance of our algorithm, we performed tests on the benchmark Middlebury dataset [1]. We evaluate the performance by finding the percentage of disparity estimations that has a difference more than one from the ground truth disparity. We checked the appearance of errors in different locations of the image such as non-occlusion (nonocc), all pixels (all) and locations close to disparity discontinuities (disc). Our algorithm has only one parameter which is the consensus window size, h_w. In all of our experiments, we choose h_w is equal to 3×3, so that we incorporate 8 neighbouring pixels when building the consensus.

We first explore the improvement that is obtained in the initial DSI with the proposed fusion strategy compared to other similarity measures and possible fusion strategies. We created four different fusion strategies and tested our algorithm with them using seven similarity measures as presented in Table I.

Most utilizes the DSI of the most confident similarity measure for each location.

Average is the fusion strategy that takes the average of all of the similarity measure's costs.

Conf denotes the adaptive aggregation of DSI using the weights that are obtained from the LRD confidence of the similarity measures at each pixel.

Voting uses our consensus strategy but without confidences. The confidence in the disparity voting is always equal to 1.

Voting+Conf is based on our consensus strategy and adaptive aggregation of the similarities using LRD confidence measure.

The best performer in all of the dataset images and in all regions is Voting+Conf which is also our proposed strategy. Voting scenario with our consensus algorithm also performs significantly better than the remaining algorithms and single similarity measures. Adaptive strategy performs on par with Average which highlights the importance of proposed consensus algorithm. Directly choosing the highest confidence is the worst performer of the fusion strategies. Census is the top performer of the individual similarity measures. In contrast, it is computationally expensive compared to other similarities. NCC and ZNCC are the next top performers after Census. The worst performer is the LoG. Computationally the cheapest similarity is the AD which performs better than Sobel, LoG and Rank in some images.

The most commonly used similarity metric is the AD and most of the algorithms [11], [10] aim to fuse AD with different similarities in an effective way. In our next experiment, we fused AD with each of the other similarity measures and compare our results with other algorithms. The results are presented in Table II. AD with Census is the top performer

TABLE I. PERCENTAGE OF ERRONEOUS DISPARITY VALUES OF PROPOSED ALGORITHM WITH SEVEN DIFFERENT SIMILARITY MEASURES (SINGLE SIM.) AND MULTIPLE SIMILARITIES (MULT. SIM.) WITH VARIOUS FUSION STRATEGIES.

		Tsukuba			Venus			Teddy			Cones		
		nonocc	*all*	*disc*	*nonocc*	*all*	*disc*	*nonocc*	*all*	*disc*	*nonocc*	*all*	*disc*
Single Sim.	**AD**	21.5	23.2	21.6	27.2	28.4	30.7	35.7	42.2	41.3	37.4	42.2	40.5
	Rank	29.8	31.2	34.0	23.8	30.8	33.6	36.1	42.5	43.9	23.0	31.5	36.0
	Census	17.1	18.8	**22.6**	**12.6**	**14.0**	**25.0**	**15.0**	**23.6**	**28.7**	**7.1**	**17.2**	**17.6**
	NCC	**16.5**	**18.2**	26.6	13.4	14.8	30.4	16.7	25.2	33.7	9.7	19.9	25.9
	ZNCC	18.6	20.3	27.2	13.8	15.3	32.3	18.6	27.0	36.4	10.5	20.6	26.3
	Sobel	23.4	25.0	32.2	27.7	28.9	36.5	40.7	46.7	48.3	28.6	36.4	41.9
	LoG	38.4	39.7	43.8	41.5	42.4	49.9	54.0	38.6	60.7	40.7	47.2	53.5
Mult. Sim.	**Most**	15.1	16.9	21.0	12.8	14.3	25.5	16.4	25.0	28.9	8.4	18.7	19.7
	Average	15.5	17.3	22.8	12.6	14.1	27.8	15.3	24.1	28.9	7.7	18.1	19.4
	Conf	14.7	16.6	21.1	12.6	14	25.8	15.7	24.4	28.6	7.8	18.3	19.1
	Voting	12.3	14.2	19.0	9.3	10.8	23.5	12.4	21.3	25.9	5.7	16.1	15.4
	Voting+Conf	**11.7**	**13.6**	**17.9**	**8.3**	**9.7**	**22.8**	**12.3**	**21.2**	**25.3**	**5.3**	**15.8**	**14.1**

TABLE II. PERCENTAGE OF ERRONEOUS DISPARITY VALUES OF AD SIMILARITY MEASURE AND ITS FUSION WITH OTHER SIMILARITY MEASURES.

	Tsukuba			Venus			Teddy			Cones		
	nonocc	*all*	*disc*	*nonocc*	*all*	*disc*	*nonocc*	*all*	*disc*	*nonocc*	*all*	*disc*
AD	21.5	23.2	21.6	27.2	28.4	30.7	35.7	42.2	41.3	37.4	42.2	40.5
AD+Rank	17.8	19.5	21.1	18.9	20.2	24.0	21.2	29.3	29.3	9.6	19.7	18.7
AD+Census	**11.4**	**13.2**	**15.8**	**7.78**	**9.28**	**20.5**	**11.6**	**20.6**	**23.5**	**4.9**	**15.4**	**12.7**
AD+NCC	13.8	15.5	16.7	18.0	19.2	26.6	22.3	30.0	32.7	20.0	28.7	26.2
AD+ZNCC	12.7	14.5	17.9	8.6	10.1	25.3	13.0	21.9	27.6	6.3	16.7	16.0
AD+Sobel	12.4	14.2	18.2	16.3	17.7	24.5	22.2	30.2	30.9	10.1	20.1	19.8
AD+LoG	14.9	16.7	18.6	19.5	20.8	28.7	25.3	32.9	33.7	17.1	26.3	25.6

among the others whereas fusion with LoG and Sobel also provides significantly higher accuracy results compared to their individual performances. The performance of each fusion is higher than the performance of individual similarities except the fusion with NCC. The accuracy of LRD on NCC is not as high as its accuracy with other similarities. Therefore, the accuracy of the fusion with NCC hampered significantly.

The results that are presented demonstrate the performance of fusion over the initial DSI. In order to explore the performance of our algorithm with aggregation, we perform two experiments with AD-Census and AD-Sobel features. AD and Census features are fused using exponential functions by Mei *et al.* [11] (AD-Census) in order to obtain higher accuracy using the function in Eq. 9.

$$C_{AC}(x,y,d) = 2 - e^{-\frac{C_{AD}}{\sigma^2_{AD}}} - e^{-\frac{C_{Census}}{\sigma^2_{Census}}} \tag{9}$$

Additionally, Klaus *et al.* [10] fused Gradient and AD features linearly (Weighted Average) as given in Eq. 10.

$$C_{WA}(x,y,d) = (1-\alpha)C_{AD}(x,y,d) + \alpha C_{SB} \tag{10}$$

σ_{AD} and σ_{Census} are given as 10 and 30 respectively. However, optimum w for $C_{WA}(x,y,d)$ is not explicitly given by Klaus *et al.* [10]. We tried different w values and experimentally found that w equal to 0.9 gives the best results. The results of our experiments for different-sized aggregation windows are presented in Fig. 1. The results are averaged over all of the four dataset images and errors are evaluated over all pixels inside the images. In both of our experiments, AD is the least accurate similarity metric in the aggregation. In the fusion of AD with Census experiment in Fig. 1.a, our algorithm performs the best especially for smaller aggregation windows. AD-Census performs on par with the Census similarity for smaller-sized windows. Yet, Census outperforms AD-Census as the aggregation window size increases. In the fusion of AD with Sobel similarity experiment, our proposed method is the best performer for all sizes of aggregation windows. Sobel similarity and Weighted Average [10] performs similarly in terms of accuracy. Yet, for smaller sized windows, Weighted Average is slightly better than Sobel and significantly better than AD similarities. In all of the experiments, the accuracy is saturated to a constant value when the aggregation window size approach to its largest size (13×13). The reason for this is as the aggregation size increase, the disparity at the discontinuities are smoothed as well as the disparities at homogeneous regions as depicted in Fig. 1.c-d and this causes errors. In general, aggregation window size is chosen between 3×3 and 9×9 to avoid errors at discontinuities. Our algorithm can achieve the highest possible accuracy even at the smaller window sizes.

In order to explore the effect of consensus region size, h_w, we perform an experiment with different h_w and show the results in Fig. 2. As h_w increases, the resulting accuracy increases. Since the computational complexity also increases with h_w, larger values requires more computations. In all of our experiments, we choose h_w as 3×3.

The perceptual results of our fusion strategy and two of the single similarity measures are presented in Fig. 2 (AD and LoG) and 4 (AD and Census) respectively. The locations where significant improvements are achieved are surrounded with red. Even without any aggregation, our algorithm provides high accuracy at both uniform color regions and disparity discontinuity locations. Fig. 3 shows the result for the fusion of AD and LoG features. The LoG similarity is computationally cheaper than Census similarity. Yet, both quantitative and perceptual results show that the difference between their fusion with AD similarity is not as high as their individual performances. Therefore, our algorithm can provide high accuracy with higher frame rates compared to using single and computationally expensive measures.

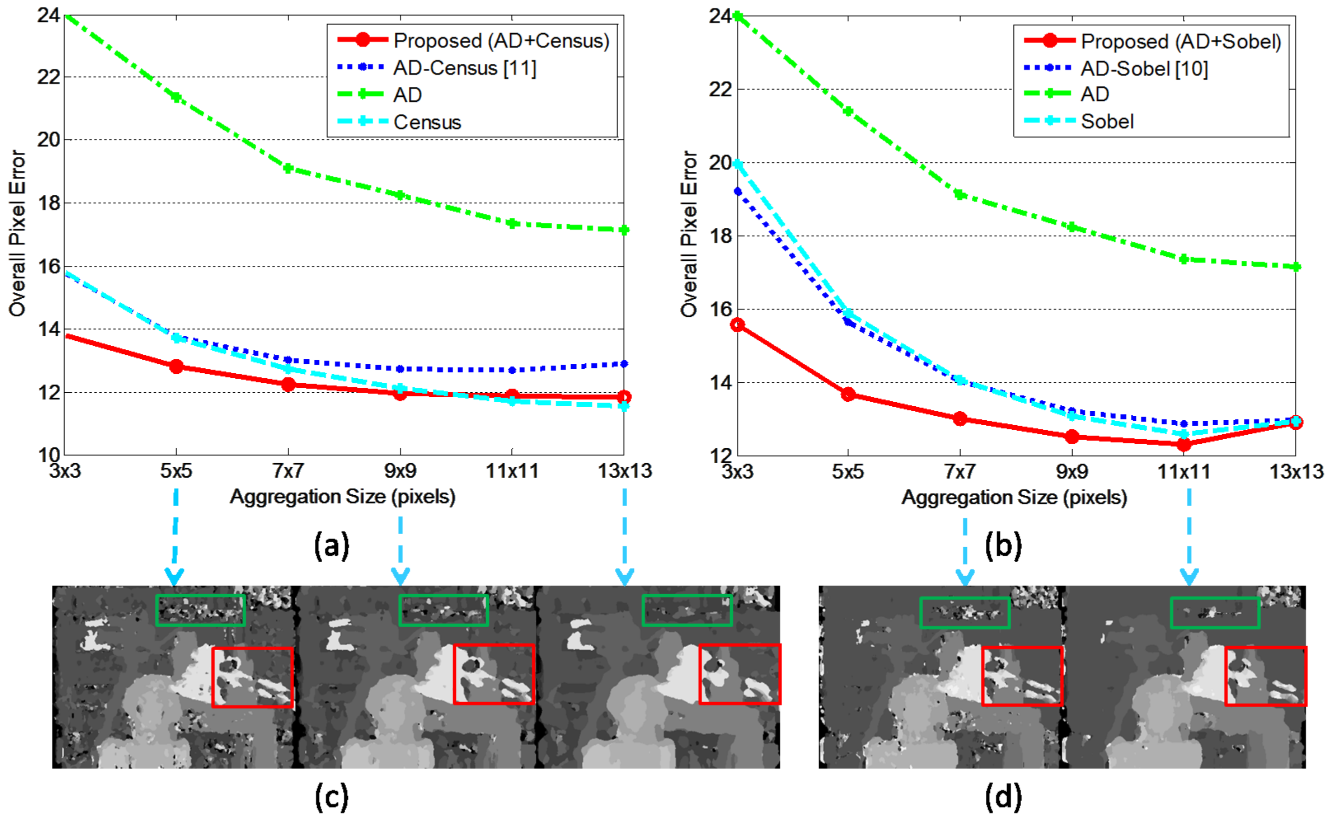

Fig. 1. The proposed algorithm performance with respect to (a) AD and Census similarity measures and AD-Census [11], (b) AD and Sobel similarity measures and Weighted Average [10]. (c-d) The results with Census (c) and Sobel (d) for aggregation size indicated with arrows. Red and Green indicate the error and improvement as the aggregation window increase respectively.

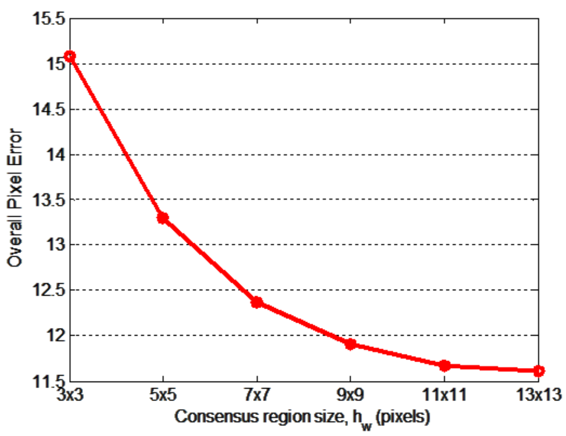

Fig. 2. The performance of our algorithm with respect to different consensus region sizes, h_w

V. Conclusion

In this paper, we presented a novel consensus-based similarity fusion algorithm using stereo confidences. To our knowledge, it is the only algorithm in stereo literature that does not require additional parameters and can be applied to fuse any number of similarity measures. The experiments show significant accuracy increase compared to individual similarity performances and other possible fusion strategies. The proposed algorithm can be further generalized for utilizing in multi-view disparity estimation algorithms.

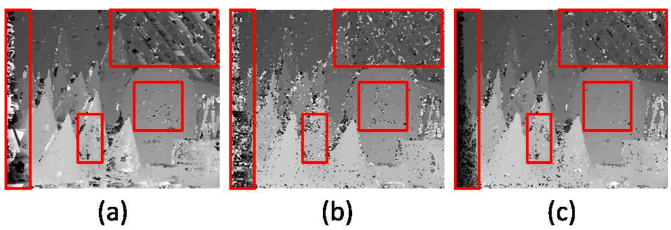

Fig. 3. AD and LoG fusion result on Cones dataset: (a) Color image, (b) ground truth disparity, (c) AD similarity, (d) LoG similarity, (e) proposed algorithm results (AD+LoG) respectively. Some of the significant differences are marked in red.

References

[1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[2] G. Saygili, L. van der Maaten, and E. A. Hendriks, “Improving segment based stereo matching using surf key points,” in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2973–2976.

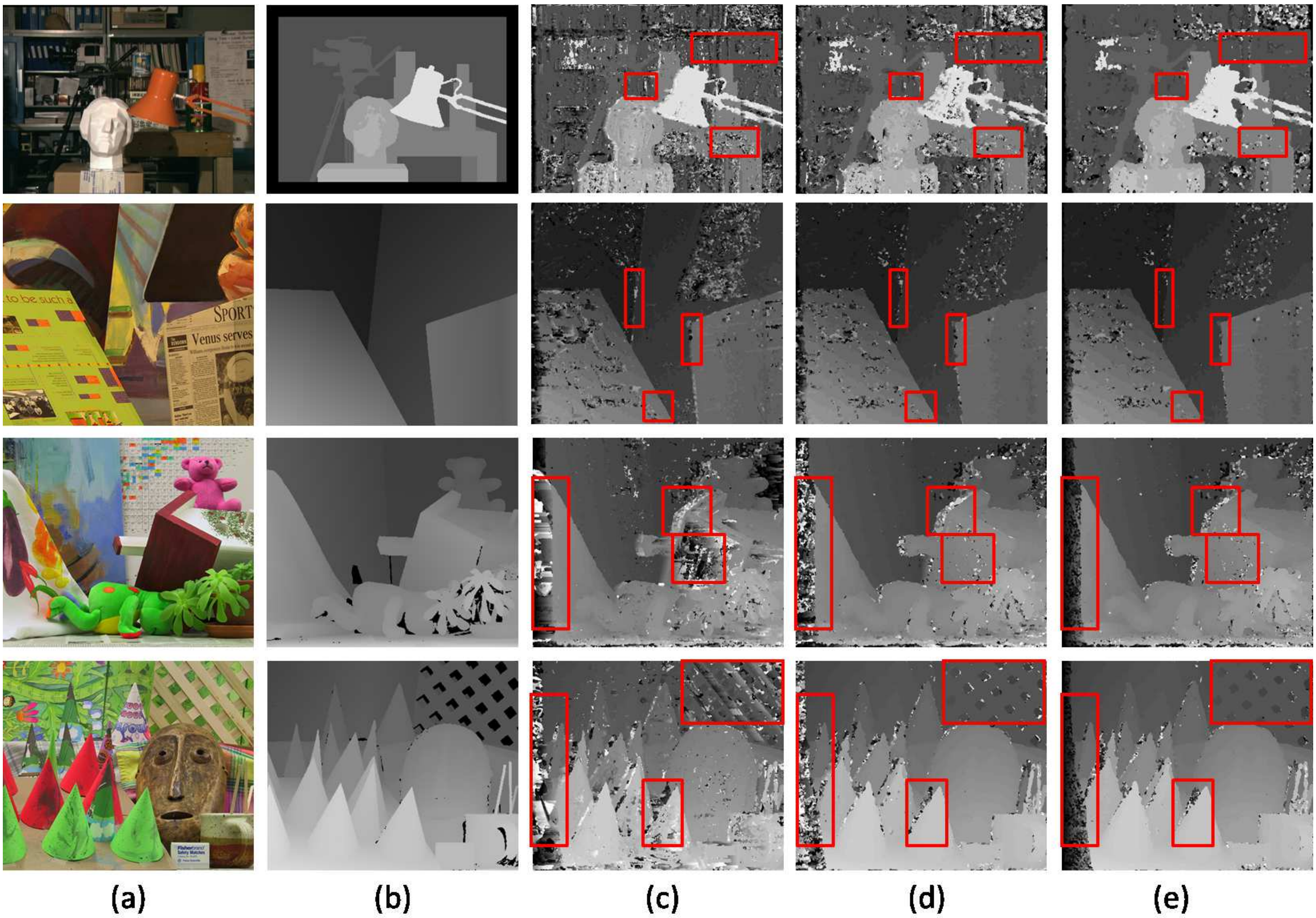

Fig. 4. (a) Color images, (b) ground truth disparities, (c) AD similarity, (d) Census similarity, (e) proposed algorithm results (AD+Census) respectively. Some of the significant differences are marked in red.

[3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International journal of computer vision*, vol. 70, no. 1, pp. 41–54, 2006.

[4] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 508–515.

[5] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 4, pp. 650–656, 2006.

[6] A. Hosni, M. Bleyer, and M. Gelautz, "Secrets of adaptive support weight techniques for local stereo matching," *Computer Vision and Image Understanding*, 2013.

[7] H. Hirschmuller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 9, pp. 1582–1599, 2009.

[8] D. Min, J. Lu, and M. N. Do, "A revisit to cost aggregation in stereo matching: How far can we reduce its computational redundancy?" in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1567–1574.

[9] Q. Yang, "A non-local cost aggregation method for stereo matching," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1402–1409.

[10] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3. IEEE, 2006, pp. 15–18.

[11] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 467–474.

[12] X. Hu and P. Mordohai, "A quantitative evaluation of confidence measures for stereo vision," 2012.

[13] P. Mordohai, "The self-aware matching measure for stereo," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1841–1848.

[14] K.-J. Yoon and I. S. Kweon, "Distinctive similarity measure for stereo matching under point ambiguity," *Computer Vision and Image Understanding*, vol. 112, no. 2, pp. 173–183, 2008.

[15] A. Milella and R. Siegwart, "Stereo-based ego-motion estimation using pixel tracking and iterative closest point," in *Computer Vision Systems, 2006 ICVS'06. IEEE International Conference on*. IEEE, 2006, pp. 21–21.

[16] P. Steingrube, S. K. Gehrig, and U. Franke, "Performance evaluation of stereo algorithms for automotive applications," in *Computer Vision Systems*. Springer, 2009, pp. 285–294.

[17] J. Banks and P. Corke, "Quantitative evaluation of matching methods and validity measures for stereo vision," *The International Journal of Robotics Research*, vol. 20, no. 7, pp. 512–532, 2001.

[18] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Computer VisionECCV'94*. Springer, 1994, pp. 151–158.