IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 6, 2016, accepted August 11, 2016, date of publication August 17, 2016, date of current version September 16, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2601069

Local Stereo Matching Based on Support Weight With Motion Flow for Dynamic Scene

JIACHEN YANG[1], (Member, IEEE), HUANLING WANG[1], ZHIYONG DING[1], ZHIHAN LV[2], WEI WEI[3], AND HOUBING SONG[4], (Senior Member, IEEE)

[1]School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China
[2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen, 518055 China
[3]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048 China
[4]Department of Electrical and Computer Engineering, West Virginia University, Montgomery, WV 25136, USA

Corresponding author: H. Song (houbing.song@mail.wvu.edu)

This work was supported in part by the Natural Science Foundation of China under Grant 61471260 and Grant 61271324 and in part by the Natural Science Foundation of Tianjin under Grant 16JCYBJC16000.

ABSTRACT Stereo matching is one of the most important and challenging subjects in the field of stereo vision. The disparity obtained in stereo matching can represent depth information in 3-D world to a great extent and shows great importance in stereo field. In general, stereo-matching methods primarily emphasize static image. However, the information provided by dynamic scene can be used fully and effectively to improve the results of stereo matching for dynamic scene, such as video sequences. In this paper, we propose a dynamic scene-based local stereo-matching algorithm which integrates a cost filter with motion flow of dynamic video sequences. In contrast to the existing local approaches, our algorithm puts forward a new computing model which fully considers motion information in dynamic video sequences and adds motion flow to calculate suitable support weight for accurately estimating disparity. Our algorithm can perform as an edge-preserving smoothing operator and shows improved behavior near the moving edges. The experimental results show that the proposed method achieves a better depth map and outperforms other local stereo-matching methods in disparity evaluation.

INDEX TERMS Stereo matching, disparity, support weight, motion flow, dynamic scene.

I. INTRODUCTION

The chief objective of computer vision is to endow computers with human-like depth vision capabilities, therefore stereo matching is one of the most active research topics in this field. In fact, numerous stereo-matching algorithms for estimating disparity can be classified into two general methods: global and local [1]. The global methods compute all disparities of an image simultaneously by optimizing a global-energy function [2]–[4], which produce accurate disparity maps. But global methods are usually computationally expensive and sometimes require many parameters that are difficult to determine. Unlike most global stereo matching methods, local methods utilize the color or intensity values within a finite support window to determine the disparity for each pixel. Therefore, local methods compute disparities within an image with a simple structure and are generally efficient and easy to implement.

How to select an appropriate matching window for each pixel has thus been a main goal of local methods. To reduce image ambiguity and improve accuracy, local methods commonly aggregate support from neighboring pixels in a given size-constrained window. This is the implied assumption that all pixels in the window are from the same depth, i.e., they have the same disparity. Several adaptive-window algorithms have been proposed to solve the problem of optimizing the size and shape of the window, and some results have been achieved [5]–[8].

However, finding the optimal support window with an arbitrary shape and size is extremely difficult and generally known as an ''NP-hard problem''. The smoothness assumption that pixels in the window have the same disparity is broken at depth discontinuities in which the window contains pixels of both background and foreground disparities. This leads to the well-known foreground fattening effect. Thus, one of the most successful local solutions based on weight using a fixed-size square window has been proposed, which defines each pixel in the window with different support weights. Yoon and Kweon proposed an adaptive support weight (ASW) approach [9] that adjusts the support weights of the window pixels by using the photometric and geometric distance with respect to the center pixel. Following this pioneering work, numerous improvements have been made in subsequent algorithms regarding the basics of ASW. Geodesic support [10] defines the weights within one window by computing the geodesic distance to the center pixel. Segment support [11] improves the reliability of adaptive support aggregation by adding additional segmentation processes. Cost filter [12] obtains consistent edge-preserving results by using a guided filter. There are also a number of other local methods that have been proposed [13]–[16].

A cost filter is a quality edge-preserving method which has been recognized as one of the best local methods for the Middlebury dataset [17]. However, it still contains errors regarding disparity estimation in textureless (flat) areas, which have different characteristics compared to edges. Sometimes it is difficult to obtain extremely precise disparity maps only considering the limited information from a given image pair. However, when working with video, it is limited to apply the existing image-based methods to obtain disparity directly. In contrast to image-based methods, we must utilize additional information from video frames to improve the disparity map.

Actually, stereo video disparity estimation is at an early developmental stage, whereas stereo image disparity estimation is at a mature one. Few approaches that adopt flow vectors or spatial-temporal characteristics have been proposed [18]–[21]. There is still enormous room for development in the area of video-based stereo matching.

In this study, we propose a video-based local stereo-matching algorithm that integrates a cost filter with motion flow of video sequences. In video processing, motion is a critical feature, and moving objects in general have a high degree of saliency and can be clearly distinguished from the background. However, most disparity methods have difficulty in dealing with fast-moving edges in video scenes. To solve this problem, we integrate motion flow into a local stereo matching algorithm to calculate the appropriate support weight. This method can reduce errors in depth discontinuities and object edge areas. The experimental results demonstrate that the proposed method achieves better depth maps and outperforms other local stereo matching methods with respect to video disparity evaluation.

The remainder of this study is organized as follows. Section 2 presents a related cost filter local stereo matching method. Improvements to the cost filter method for videos are described in Section 3. We show experimental results and provide analysis in Section 4. Section 5 concludes the study.

II. RELATED WORK

As mentioned previously, a cost filter is recognized as one of the best local methods and obtains better results especially on edge-preserving compared to other methods. This is mainly due to the fact that a cost filter can distinguish the same-side edge part effectively by means of the algorithm, and can calculate the appropriate weights for pixels in the support window. Cost filters use the weights of the guided filter [22], which we briefly review now [12]. To illustrate, we just take a grayscale guidance image I as example, i and j are pixel indexes. The weight $W_{i,j}$ is defined as:

$$W_{i,j} = \frac{1}{|\omega|^2} \sum_{k:(i,j)\in\omega_k} \left(1 + \frac{(I_i-\mu_k)(I_j-\mu_k)}{\sigma_k^2+\epsilon}\right) \tag{1}$$

where ϵ is a smoothness parameter, and μ_k and σ_k are the mean and the variance of I in a squared window ω_k with dimensions $(2r+1)\times(2r+1)$, centered at pixel k. We denote the number of pixels in this window with $|\omega|$.

Next, we will explain why the filter weights can preserve edges of I. The numerator $(I_i-\mu_k)(I_j-\mu_k)$ will be greater than zero and have a positive effect if I_j is located on the same side of the edge as I_i , but will be less than zero and have a negative effect in the opposite case. Thus the term $1+\frac{(I_i-\mu_k)(I_j-\mu_k)}{\sigma_k^2+\epsilon}$ is large for pixel pairs on the same side of the edge, and small otherwise. Hence, pixels are not averaged if they are separated by an image edge.

The strength of the averaging is controlled by the parameter ϵ in eq.(1). If $\sigma_k^2 \ll \epsilon$ (then μ_k is similar to I_i and I_j), the numerator is much smaller than the denominator in eq.(1). Hence, the kernel converges to an (unweighed) low-pass filter: $W_{i,j}=\frac{1}{|\omega|^2}\sum_{k:(i,j)\in\omega_k}1$.

The filter weights are similarly defined for color images:

$$W_{i,j} = \frac{1}{|\omega|^2} \sum_{k:(i,j)\in\omega_k} (1+(I_i-\mu_k)^T(\Sigma_k+\epsilon U)^{-1}(I_j-\mu_k)) \tag{2}$$

where I_i, I_j and μ_k are 3×1 (color) vectors, and the covariance matrix Σ_k and identity matrix U are of size 3×3. The weights are high in regions that are self-similar to the central pixel, and low otherwise.

Overall, the effectiveness of the cost filter is outstanding. However, there still exist problems when we observe the results carefully. The algorithm provides weights for the pixels on the different sides of the edge, but the values of these weights should be small and nearly to zero. This is shown and compared to the proposed method in Section 4. The numerator $(I_i-\mu_k)(I_j-\mu_k)$ will be greater than zero when the pixel on the other side of the edge has similar color distribution to that of the central pixel. In this situation, the pixel can easily be considered to be on the same side and given the wrong weights. Despite the few errors in weight, they may still influence the process of stereo matching, which will eventually lead to failure in disparity estimation.

To solve this problem and improve the algorithm to be suitable for stereo matching in videos, we intend to use the motion information from the video. In videos, motion is a critical feature, which covers spatial and temporal characteristics simultaneously and is also the biggest difference with the static image. Therefore, motion has a better chance to effectively remove error weights and eventually obtain an accurate stereo-matching result of video sequences. When the object moves, it will progress to different positions in the video sequences, and the edge of foreground and background objects can be distinguished according to the change of positions. We then combine the edge with the cost filter to remove errors of weight on the different side. Accordingly, we can provide appropriate weights for every pixel and obtain an accurate stereo-matching result.

III. STEREO MATCHING BASED ON VIDEOS

In this section, we will introduce our improved algorithm based on a cost filter for video stereo matching.

A. COST COMPUTATION

For each pixel i in the left image I_{left} and each allowed disparity d, the cost volume represents the dissimilarity between pixel i and the pixel at coordinates $i - d$ in the right image I_{right}. In particular, we use a truncated absolute difference between the color and gradient (TAD C+G) at the matching points. This model has been proven to be robust to illumination changes.

The color difference $M(i, d)$ for the matching pixel i at disparity d is defined as

$$M(i, d) = ||I_{left}(i) - I_{right}(i - d)|| \tag{3}$$

where $I(i)$ denotes the value of the color distribution in RGB space at pixel i.

The absolute difference $G(i, d)$ of the gradients is defined as

$$G(i, d) = ||\nabla_x(I_{left}(i)) - \nabla_x(I_{right}(i - d))|| \tag{4}$$

where $\nabla_x(I(i))$ denotes the gradient in the x direction computed at pixel i .

The final cost function $C(i, d)$ is defined as

$$C(i, d) = \alpha \cdot min(T_c, M(i, d)) + (1 - \alpha) \cdot min(T_g, G(i, d)) \tag{5}$$

where α balances the color and gradient terms, and T_c and T_g are truncation values that contribute to reduce the influence of occluded pixels on the matching result.

B. COST AGGREGATION AND DISPARITY SELECTION STRATEGY

Aggregated cost volume represents the matching cost between pixel i and the pixel at coordinates $i - d$ of the right image. The aggregated cost volume is a weighted average of all pixels in the same window, which is defined as

$$C'_{i,d} = \sum_j W_{i,j}(I)\, C(j, d) \tag{6}$$

where $C'_{i,d}$ is the aggregated cost volume, and i and j are pixels in the support window. The weight $W_{i,j}$ represents the influence of pixel in the support window.

Once the aggregated cost volume of pixels is determined, the final disparity is obtained by adopting an accepted rule of winner-take-all (WTA). In other words, we regard the disparity d_i that corresponds to the minimum value of the aggregated cost volume as the final disparity of pixel i. This can be expressed mathematically as

$$d_i = arg \min_d C'_{i,d} \tag{7}$$

C. PROPOSED VIDEO-BASED WEIGHT

In order to solve the problem of weight calculation in a cost filter as mentioned in Section 2, additional measures are required, such as using motion flow in the video to improve the weights.

Motion has been used rarely for support weight calculation within a localized window, although it is a crucial factor in video processing. In the ASW method, proximity and similarity are treated as measures of independent standards. We thus model motion flow in the same manner. In addition, the local methods require pixel-based computation, hence we use the classic optical flow method with the weighted non-local term [23], which is a state-of-the-art optical-flow method.

Once the method has been determined, we must consider how the motion flow affects the result. A larger moving speed will produce a larger motion flow, and the edge of foreground and background of objects can be distinguished easily, which will contribute to provide a more appropriate weights for each pixel and finally produce a more accurate stereo-matching result. On the contrary, when the moving speed is relatively small, the effect of motion flow will be small, which will result in less accurate stereo-matching results. In fact, the motion difference between the two pixels is calculated by measuring optical flow. There exist two methods to calculate motion difference: absolute flow endpoint difference (ED) and angular difference (AD) [24]. In our model, ED is employed because AD in the region that has large motion is down-weighted and more likely to produce an error. As such, ED becomes the preferred measure of flow accuracy [24]. We regard $f_i = (u_i, v_i)$ and $f_j = (u_j, v_j)$ as the flow vectors of pixel i and j, respectively. The truncated motion difference is defined as

$$\Delta f_{i,j} = min(||f_i - f_j||, T_\tau) \tag{8}$$

where T_τ is a truncation value. In this manner, the influence of abnormal optical flow similar to T_c and T_g can be reduced.

At the beginning of our method, we integrate the motion difference simply by referring to the formula used in the ASW method. However, the effect is unsatisfactory and many errors occur on the edge part of the object. We then focus on the essential fact that the optical flow is an estimated value and cannot be exactly correct. Thus, we should consider additional factors to reduce the estimated errors and obtain the appropriate formula of weight rather than just add motion difference to the formula. After extensive exploration, a model that integrates motion with color similarity is put forward, which is denoted as

$$W'_{i,j} = W_{i,j} \cdot \exp(-\Delta f_{i,j} \cdot \Delta c_{i,j}/\gamma) \tag{9}$$

where $W_{i,j}$ is the weight defined in cost filter, $\Delta c_{i,j}$ is the color dissimilarity, and γ is an empirical parameter. $\Delta c_{i,j}$ is defined as

$$\Delta c_{i,j} = min(||I_i - I_j||, T_\eta) \tag{10}$$

where I_i and I_j are pixel distribution in RGB color space, and T_η is a truncation value.

The proposed formula is based on the understanding that there exists a correlation between color similarity and motion. In addition, the two pixels that have the same color distribution as in the flat areas of an object surface seem to have a similar motion trend. Moreover, because color is an observed quantity, the exact color value can be obtained directly. The addition of color can optimize the model that only considers an estimated motion. Finally, the proposed model improves performance at the edges compared to the cost filter. In general, we define a new weight calculation model that considers the correlated relation between similarity and motion.

D. OCCLUSION FILLING AND POST-PROCESSING

After the disparity map is acquired, a problem with occlusion occurs in which the disparity is discontinuous in some parts of the image. To solve this problem, a left-right consistency check is performed to detect unreliable pixels. Unreliable pixels are those that have different disparities on the left and right images. To obtain a dense disparity map, we adopt a post-processing strategy of [25]. The occluded pixels are assigned the lowest disparity value of the spatially closest non-occluded pixels that lie on the same scanline (pixel row). This strategy generates streak-like artifacts in the disparity map, and we post-process the filled-in pixels. We perform edge-preserving smoothing on the filled-in regions by using a weighted bilateral median filter.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the proposed method, we use stereo videos for processing and acquire a disparity image. The experimental parameters of our proposed method are set to constant values, and empirically defined as:

$$\{r, \epsilon, \alpha, T_c, T_g, T_\tau, T_\eta, \gamma\} = \{9, 0.01^2, 0.1, 0.028, 0.008, 0.8, 0.03, 0.2\}$$

A. SUPPORT WEIGHT IMPROVEMENT COMPARED WITH COST FILTER

For the purpose of comparing the weight calculated by cost filter [12] and our proposed method objectively, we employ one video in [28] that contains the same frame as the one in the cost filter to do experiment. For comparison, we just use the frame in the study of cost filter and the same five points to show the weights. The image and corresponding optical flow image are shown in Fig. 1. The weights of points are shown in Fig. 2. As Fig. 2(a1)-(e1) reveal (see highlighted regions marked with red ellipses), several pixels on the different sides of the center pixel are assigned wrong weights as calculated by the cost filter. Although the overall effect is not unsatisfactory, acquiring more accurate results is possible. By contrast, our proposed method can remove the error weights effectively, as Fig. 2(a2)-(e2) show, which ensure that subsequent processing is accurate. It is thanks to the motion flow that can distinguish the object and background. This shows the potential of motion flow in improving the performance of stereo matching for dynamic scene. What's more, the estimated optical flow is slightly inaccurate, which causes weight problems. To solve this problem, we make use of the observed color variable to optimize the model. In particular case that no motion occurs in the video, the optical flow difference of all pixels is zero, and our method is the same as that of the cost filter. Accordingly, our proposed model integrating the cost filter with motion and color is suitable for weight calculation.

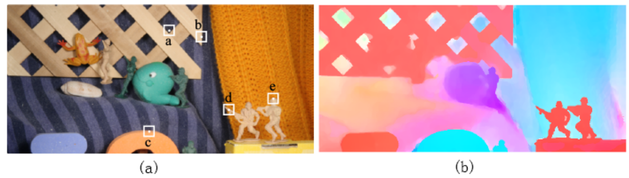


FIGURE 1. (a) original image. (b) corresponding optical flow image.

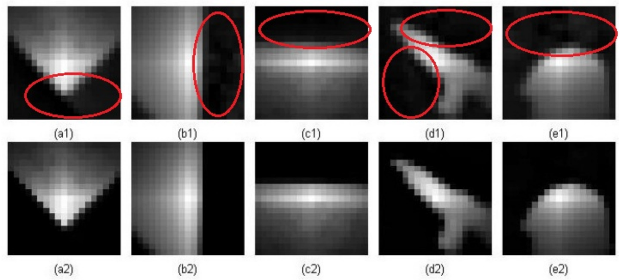


FIGURE 2. The weights calculated at different positions. (a1)-(e1) use cost filter, (a2)-(e2) add motion and color based on cost filter. Note that cost filter have more errors (see highlighted regions) but our method performs well.

B. DISPARITY ESTIMATION RESULTS ON STEREO VIDEOS

It is shown in [22] that ASW and the cost filter are the best among the stereo matching methods based on adaptive weight, because ASW performs better on the average rank, while the cost filter produces a lower average error. Therefore, we compare our proposed method with ASW and the cost filter.

In this section, we first provide the disparity image results before occlusion filling and post-processing. We mainly focus on moving parts, because our method is the same as the cost filter in static parts when no motion occurs. Fig. 3 demonstrates the original image of a moving hand and the corresponding optical flow image. Fig. 4 illustrates the left and right disparity maps. As mentioned previously, the optical flow is an estimated value and visible errors occur, as shown in Fig. 3(e)-(h). Therefore, calculating weights fully with optical flow is inappropriate. Integrating the optical flow

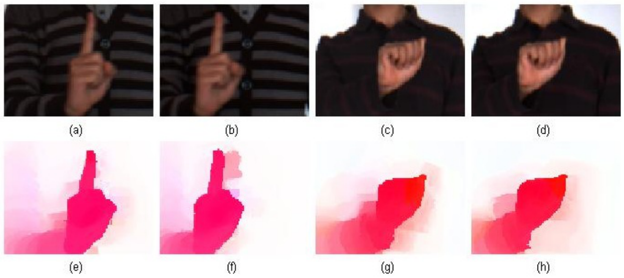


FIGURE 3. (a)-(b) original left and right image of "finger". (c)-(d) original left and right image of "fist". (e)-(h) corresponding optical flow image with upper original image.

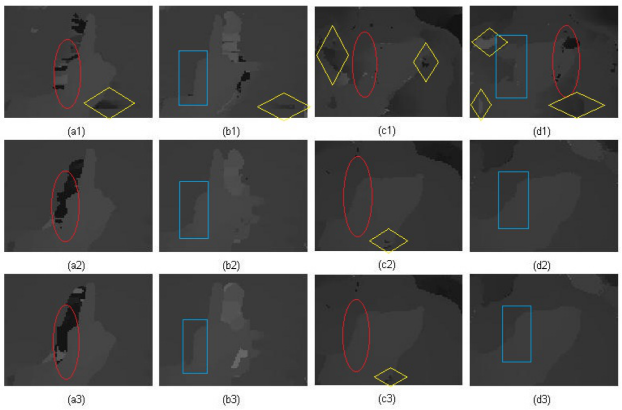


FIGURE 4. Left and right disparity map for "finger" and "fist". (a1)-(d1) acquired by ASW. (a2)-(d2) acquired by cost-filter. (a3)-(d3) acquired by our proposed method.

and color similarity into the weight represents an improved strategy.

Next, we analyze the results of Fig. 4. In the wrong disparity parts marked with yellow diamonds, ASW produces multiple wrong disparity parts in the finger and fist images, whereas the cost filter and our proposed method show no problems in the finger image and fewer parts of wrong disparity in the fist image. In the occlusion parts marked with red ellipses, ASW and the cost filter produce rough edges in the finger image, whereas our proposed method produces a smooth and clear boundary that is consistent with the original image. The three methods perform similarly with respect to the fist image. In non-occlusion parts marked with blue rectangles in the finger image, the cost filter produces a rough edge and flat corner between the palm and arm that should be nearly at a right angle, whereas this problem does not occur when ASW and our method are employed. Regarding the blue rectangles in the fist image, ASW produces a projecting part with obvious wrong disparity, and the cost filter produces a bulge sandwiched between the hand and arm. By contrast, our proposed has no problem in these parts.

Overall, our proposed method represents a major improvement over the other two methods. This is primarily due to the improvement of our weight model, and also illustrates the feasibility of our model. In fact, distinguishing objects at different depths and acquiring accurate disparity with single image pair, especially on the edges, is difficult. Therefore, errors occur easily during weight calculation. However, our method make full use of optical flow according to the motion information given by the object in the video to solve this problem. Our method can distinguish edges of an object and allocate appropriate weight to pixels on both sides of object's edge, removing the negative influence of errors in which pixels near edge have undeserved weights. Therefore, our method can achieve better results than the other methods, and eventually can obtain a more accurate disparity image after following the same occlusion filling and post-processing procedures.

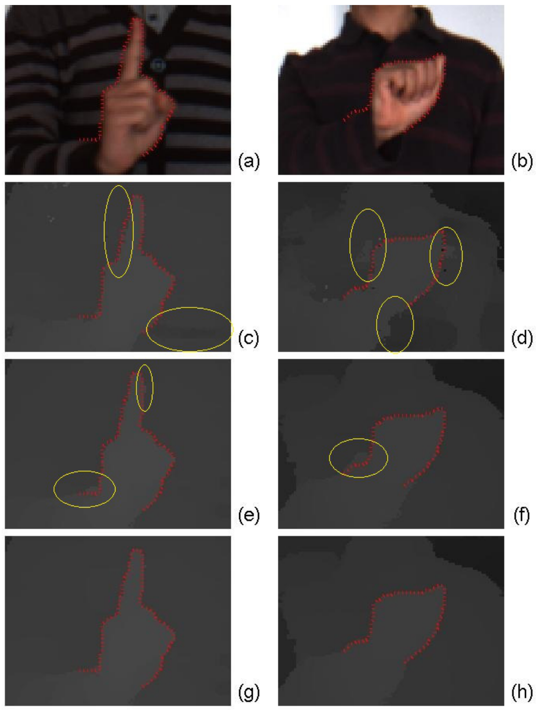


FIGURE 5. Final disparity maps for "finger" and "fist" images. (a) and (b) original "finger" and "fist" image. (c) and (d) acquired by ASW. (e) and (f) acquired by cost filter. (g) and (h) acquired by our proposed method.

In order to fully understand the validity and integrality of our method, we show the disparity image after all processing has completed. Based on the left and right disparity images acquired in the previous process, occlusion parts are filled and non-occlusion parts are processed. The final disparity maps are shown in Fig. 5. To compare the results more analytically, we draw the outline of a moving hand and mark the wrong disparity parts with yellow ellipses on disparity maps. ASW show several errors in the background and a deviation relative to the outline in Fig. 5(c). In addition, it produces many obvious errors that cannot distinguish depth at the edge of the hand in Fig. 5(d). Cost filter errors appear in the corner between the palm and arm, and cost filter shows some differences with the outline observed in Fig. 5(e) and (f). By contrast, our proposed method performs expertly, especially with respect to the parts of the hand and background.

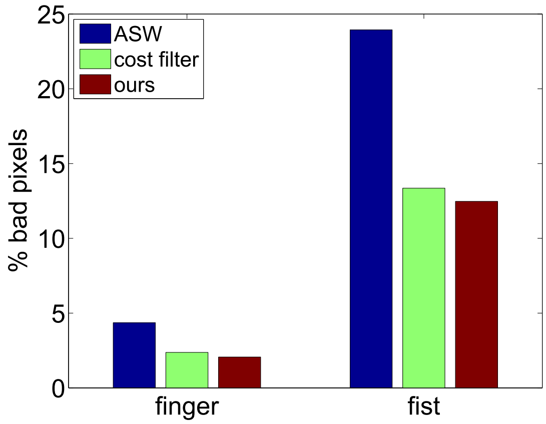


FIGURE 6. The performance comparison for "finger" and "fist" disparity maps.

Besides, our method produces a more explicit boundary between the moving hand and background (consistent with the outline) in Fig. 5(g) and (h). For more intuitive comparison, we show a histogram of rough percent of bad pixels in Fig. 6. Our proposed method has minimal error percent in images, and outperforms ASW and the cost filter methods, which show the effectiveness of our approach.

C. QUANTITATIVE EVALUATION

The quantitative evaluation of disparity maps from stereo videos is hindered by the general lack of ground truth disparity maps. In order to ensure objective quantitative evaluation, we adopt a synthetic dataset including five stereo sequences with known ground truth disparity maps, provided by [29] (see in Fig. 7).

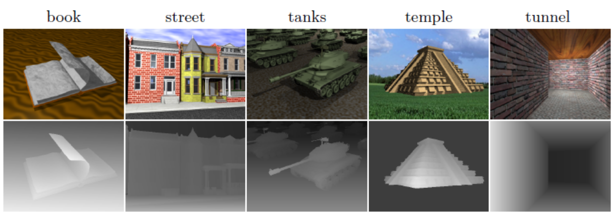


FIGURE 7. Selected frames and disparity maps from synthetic stereo video sequences.

To understand why our method performs better than cost filter, we exhibit the improvement of adding motion information relative to the cost filter in Fig. 8. We select frames at a fixed interval to show the results and offer overall results in the ensuing paragraphs. Note that the book sequence has almost only half of the frames of the other four sequences, thus the interval of book is smaller. Our method has a large improvement in book and temple sequences, because these two sequences have clear and distinguishable movement that lead to more accurate weight. The improvement in tanks and tunnel sequences are relatively small, because the frames have overall movement that lead to small difference in motion and some troubles in distinguishing objects, but our method still achieve better results. We believe that the poor performance compared with cost filter on street sequences is the result of the existence of a lot of texture, inconspicuous edge information and tiny range of depth differences in the scene. These factors lead to unstable and obscure optical flow, and finally a bad performance of disparity estimation is obtained. We will adopt a more ideal method to overcome this problem in the future.

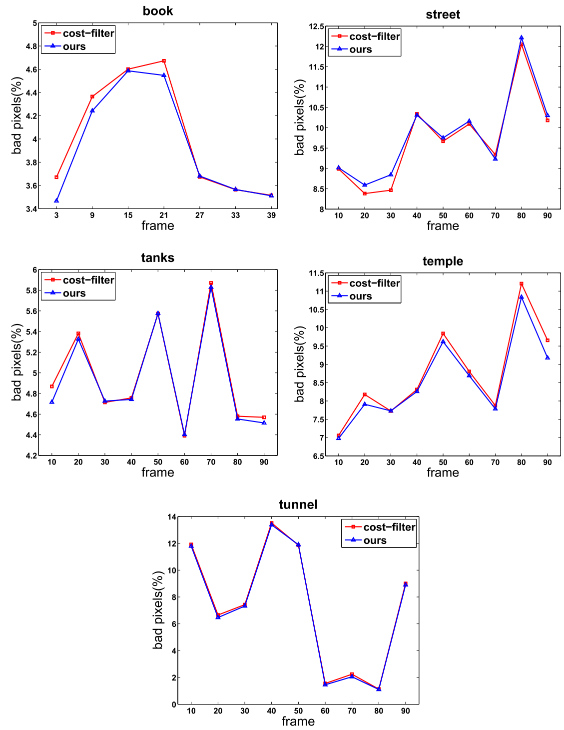


FIGURE 8. Bad pixels comparison of selected frames.

To ensure the integrity of experiment, we add the video based method denoted as "DCB grid" and "temporal DCB grid" in [29] for better comparison. The performance of all methods are shown in Table 1. We process all sequences and use the mean error (percentage of bad pixels, threshold of 1) as a standard to evaluate the performance. Our approach generates results that are visually comparable or better than the compared methods.

TABLE 1. Quantitative evaluation of performance.

Algorithm	*book*	*street*	*tanks*	*temple*	*tunnel*
Ours	**4.02**	9.85	**4.93**	**8.55**	**7.25**
Cost filter	4.08	**9.69**	4.97	8.74	7.27
DCB grid	9.46	12.44	6.15	9.12	9.68
Temporal DCB grid	11.62	10.45	11.25	8.63	24.20

The best results are produced by our method which outperforms the other techniques on all datasets and gets higher quality (except street relative to cost filter). Because the stereo sequences have relatively clear and discernible motion and edge information, the obtained optical flow in our method has a good effect and improves performance compared with cost filter, the DCB grid and temporal DCB grid methods.

Due to 3D contents can bring users stereo perception and immersive viewing experiences, 3D video has become a popular research field in recent years. Video transmission has been recently deployed in Vehicular networks [30]–[32], in the future work, we will also focus on the study of real-time stereo video transmission system in Vehicular networks.

V. CONCLUSION

In this study, we present a video-based stereo-matching method that integrates motion flow and color similarity with an image-based cost-filter local method. Our proposed method is proven to be more suitable for video disparity estimation than other methods. In local stereo matching, support weight is a crucial factor that influences the accuracy of a disparity map. To obtain a more precise support weight, a correlated support model is introduced. We consider that object motion flow takes advantage of the benefits of motion and thus combine color similarity to refine the model for video disparity estimation. The experimental results show that our proposed method achieves an improved depth map, especially at the edge of moving objects, and outperforms the other local stereo-matching methods in terms of video disparity evaluation.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.

[2] M. G. Mozerov and J. van de Weijer, "Accurate stereo matching by two-step energy minimization," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1153–1163, Mar. 2015.

[3] S. Lee, J. H. Lee, J. Lim, and I. H. Suh, "Robust stereo matching using adaptive random walk with restart algorithm," *Image Vis. Comput.*, vol. 37, pp. 1–11, May 2015.

[4] F. Cheng, H. Zhang, D. Yuan, and M. Sun, "Stereo matching by using the global edge constraint," *Neurocomputing*, vol. 131, pp. 217–226, May 2014.

[5] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1073–1079, Jul. 2009.

[6] J. Lu, K. Shi, D. Min, L. Lin, and M. N. Do, "Cross-based local multipoint filtering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 430–437.

[7] U. Raghavendra, K. Makkithaya, and A. K. Karunakar, "Anchor-diagonal-based shape adaptive local support region for efficient stereo matching," *Signal, Image Video Process.*, vol. 9, no. 4, pp. 893–901, 2013.

[8] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 467–474.

[9] K.-J. Yoon and I.-S. Kweon, "Locally adaptive support-weight approach for visual correspondence search," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2005, pp. 924–931.

[10] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann, "Local stereo matching using geodesic support weights," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2093–2096.

[11] F. Tombari, S. Mattoccia, and L. Di Stefano, "Segmentation-based adaptive support for accurate stereo correspondence," in *Proc. 2nd Pacific-Rim Symp. Image Video Technol.*, Heidelberg, Germany, Dec. 2007, pp. 427–438.

[12] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3017–3024.

[13] J. Lu, H. Yang, D. Min, and M. N. Do, "Patch match filter: Efficient edge-aware filtering meets randomized search for fast correspondence field estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1854–1861.

[14] Y. Yang, M. Gao, J. Zhang, Z. Zha, and Z. Wang, "Depth map super-resolution using stereo-vision-assisted model," *Neurocomputing*, vol. 149, pp. 1396–1406, Feb. 2015.

[15] D. Min, J. Lu, and M. N. Do, "Joint histogram-based cost aggregation for stereo matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2539–2545, Oct. 2013.

[16] D. Chen, M. Ardabilian, and L. Chen, "A fast trilateral filter-based adaptive support weight method for stereo matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 730–743, May 2015.

[17] D. Scharstein and R. Szeliski. (2002). *Middlebury Stereo Evaluation-Version 2*. [Online]. Available: http://vision. middlebury. edu/stereo/eval

[18] M. Bleyer and M. Gelautz, "Temporally consistent disparity maps from uncalibrated stereo videos," in *Proc. 6th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2009, pp. 383–387.

[19] R. Khoshabeh, S. H. Chan, and T. Q. Nguyen, "Spatio-temporal consistency in video disparity estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP).*, May 2011, pp. 885–888.

[20] Q. Liu, Y. Yang, Y. Gao, R. Ji, and L. Yu, "A Bayesian framework for dense depth estimation based on spatial–temporal correlation," *Neurocomputing*, vol. 104, pp. 1–9, Mar. 2013.

[21] J. Jiang, J. Cheng, B. Chen, and X. Wu, "Disparity prediction between adjacent frames for dynamic scenes," *Neurocomputing*, vol. 142, pp. 335–342, Oct. 2014.

[22] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Computer Vision*. Heidelberg, Germany: Springer, Sep. 2010, pp. 1–14.

[23] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2432–2439.

[24] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.

[25] A. Hosni, M. Bleyer, and M. Gelautz, "Secrets of adaptive support weight techniques for local stereo matching," *Compt. Vis. Image Understand.*, vol. 117, no. 6, pp. 620–632, Jun. 2013.

[26] D. Chen, M. Ardabilian, and L. Chen, "A novel trilateral filter based adaptive support weight method for stereo matching," in *Proc. BMVC*, Oct. 2013, pp. 1–11.

[27] Z. Lee, J. Juang, and T. Q. Nguyen, "Local disparity estimation with three-moded cross census and advanced support weight," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1855–1864, Dec. 2013.

[28] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.

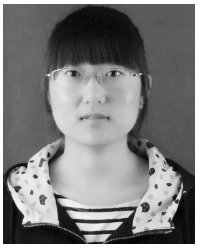
[29] C. Richardt, D. Orr, I. Davies, A. Criminisi, and N. A. Dodgson, "Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid," in *Computer Vision*. Heidelberg, Germany: Springer, Sep. 2010, pp. 510–523.

[30] M. A. Yaqub, S. H. Ahmed, S. H. Bouk, and D. Kim, "FBR: Fleet based video retrieval in 3G and 4G enabled vehicular ad hoc networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

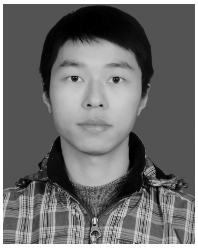
[31] M. Shojafar, N. Cordeschi, and E. Baccarelli, "Energy-efficient adaptive resource management for real-time vehicular cloud services," *IEEE Trans. Cloud Comput.*, vol. PP, no. 99, pp. 1–14, 2016.

[32] E. Baccarelli, N. Cordeschi, A. Mei, M. Panella, M. Shojafar, and J. Stefa, "Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: Review, challenges, and a case study," *IEEE Netw.*, vol. 30, no. 2, pp. 54–61, Mar./Apr. 2016.

JIACHEN YANG received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, Tianjin, China, in 2005 and 2009, respectively. He is an Associate Professor with Tianjin University. In 2014, he was a Visiting Scholar with the Department of Computer Science, School of Science, Loughborough University, U.K. His research interests include computational electromagnetic, wireless power transfer, the designs and applications of active and passive planar antennas, pattern recognition, and digital image processing.



HUANLING WANG received the B.S. degree in engineering from the Hebei University of Technology in 2015. She is currently pursuing the M.S. degree with the School of Electronic Information Engineering, Tianjin University, Tianjin, China. Her research interests include batch data processing, machine learning, digital image processing, stereo image feature detection and description, and video processing.



ZHIYONG DING received the B.S. and M.S. degrees from the School of Electronic Information Engineering, Tianjin University, Tianjin, China, in 2013 and 2016, respectively. His research interests include stereo vision (such as capture, transmission, feature detection, and description), hand gesture recognition, and coding.

ZHIHAN LV received the Ph.D. degree in computer applied technology from the Ocean University of China in 2012. Before that, he has 16 months full-time research experience with the Centre National de la Recherche Scientifique, Paris, from 2010 to 2011. After then, he has fulfilled two-year post-doctoral research experience with Umea University and an invited teaching experience with the KTH Royal Institute of Technology, Sweden. Since 2012, he has been an Assistant Professor with the Chinese Academy of Science. His research interests include wireless power transfer, the applications of active and passive planar antennas, 3-D visualization, and human–computer interaction.

WEI WEI received the M.S. and Ph.D. degrees from Xian Jiaotong University in 2005 and 2011, respectively. He is currently a Lecturer of Computer Science with the Xi'an University of Technology. His research interests include computational electromagnetics, wireless sensor networks application, mobile computing, distributed computing, and pervasive computing.

HOUBING SONG (M'12–SM'14) received the M.S. degree in civil engineering from the University of Texas, El Paso, TX, in 2006, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in 2012. In 2012, he joined the Department of Electrical and Computer Engineering, West Virginia University, Montgomery, WV, where he is currently an Assistant Professor and the Founding Director of both the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us), and the West Virginia Center of Excellence for Cyber-Physical Systems sponsored by the West Virginia Higher Education Policy Commission. In 2007, he was an Engineering Research Associate with the Texas A & M Transportation Institute. He is the Editor of four books, including *Smart Cities: Foundations, Principles and Applications* (Hoboken, NJ, USA: Wiley, 2017), *Security and Privacy in Cyber-Physical Systems: Foundations, Principles and Applications* (Chichester, U.K.: Wiley, 2017), *Cyber-Physical Systems: Foundations, Principles and Applications* (Waltham, MA, USA: Elsevier, 2016), and *Industrial Internet of Things: Cybermanufacturing Systems* (Cham, Switzerland: Springer, 2016). He is the author of more than 100 articles. His research interests include cyber-physical systems, Internet of Things, cloud computing, big data analytics, connected vehicle, wireless communications and networking, and optical communications and networking.

Dr. Song is a member of ACM. He was the very first recipient of the Golden Bear Scholar Award, the highest faculty research award at the West Virginia University Institute of Technology in 2016.