An Improved Stereo Matching Algorithm with Ground Plane and Temporal Smoothness Constraints

Cevahir Çığla[1,2] and A. Aydın Alatan [2]

[1] ASELSAN Inc.
Ankara, Turkey
[2] Middle East Technical University
Ankara, Turkey
{cevahir,alatan}@eee.metu.edu.tr

Abstract. In this study, novel techniques are presented addressing the challenges of stereo matching algorithms for surveillance and vehicle control. For this purpose, one of the most efficient local stereo matching techniques, namely permeability filter, is modified in terms of road plane geometry and temporal consistency in order to take the major challenges of such a scenario into account. Relaxing smoothness assumption of the permeability filter along vertical axis enables extraction of road geometry with high accuracy, even for the cases where ground plane does not contain sufficient textural information. On the other hand, temporal smoothness is enforced by transferring reliable depth assignments against illumination changes, reflections and instant occlusions. According to the extensive experiments on a recent challenging stereo video dataset, the proposed modifications provide reliable disparity maps under severe challenges and low texture distribution, improving scene analyses for surveillance related applications. Although improvements are illustrated for a specific local stereo matching algorithm, the presented specifications and modifications can be applied for the other similar stereo algorithms as well.

1 Introduction

The advances in robotics and automation introduce new application areas for stereo matching that provide 3D data from two cameras. In this manner, the passive technology behind such a scheme enables stereo matching [1] to be the most common way of depth extraction, especially for outdoor scenes through simple data acquisition capability. Moreover, current benchmarks [2] enabled stereo matching algorithms to improve rapidly by providing objective comparison with the available ground truth depth maps. On the other hand, these datasets have limited realism and may not model the characteristics of the real world data that involve various imperfections. Hence, new challenges are introduced in [3] for stereo matching to provide robust solutions against non-Lambertian surfaces, different lightening conditions, and complex scenes. These challenges require modifications over the well-known stereo matching algorithms.

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 134–147, 2012.

Among many alternatives [1], local stereo matching techniques [4]-[10] have been popular recently due to low memory requirement, non-iterative computation and moderate quality. As pioneered by the well-known bilateral filter (BF) [4], edge-aware filters are excessively utilized to aggregate cost functions of disparity candidates to enforce smoothness among color-wise similar regions for disparity maps. However, it is expensive to provide content adaptability, since filter coefficients alternate according to image statistics, which requires special attention for each pixel individually.

Several approximations over bilateral filter are proposed in order to increase their efficiency. In [5], constant time computation of BF has been demonstrated yielding window size independency. One of the most efficient implementation is introduced in [6], in which piecewise linear approximation is provided through discritization of image intensities into a number of intensity levels. Content adaptive filtering is conducted for these levels which are defined as Principle Bilateral Filtered Image Component (PBFIC). Filtered values of the remaining levels are calculated by the linear interpolation of the two closest quantization levels to the pixel intensity.

In [7], an alternative edge-preserving filtering technique is presented, denoted as *image guided filtering*, in which adaptive weights are provided by a local linear model between the guidance image and the filtered data. In another approach [8], guided image filtering is applied for stereo matching where the computational complexity of adaptive weights is reduced drastically by use of box filters over mean and variance of the cost volume. As a different technique, orthogonal cross aggregation is exploited by considering connected pixel groups having similar intensity characteristics [9]. The aggregation is performed by two passes in vertical and horizontal direction over integral images, which approximates BF by constant weights along arbitrary regions. This approach is recently extended to soft weights by [10] through recursive horizontal and then vertical weighted summation. In [10], for each pixel aggregation is effectively provided among connected support regions with soft weights. The recursive structure in [10] enables complete content adaptability yielding competitive performance based on Middlebury stereo benchmark.

In this study, permeability filter (PF) introduced in [10] is modified for stereo matching among challenging videos that involve various imperfections. Giving a brief summary for permeability filter and its implementation for stereo, details of the modifications for real world data are given in the following section. In Section 4, experimental results are presented for challenging stereo videos, and Section 5 is devoted to the concluding remarks.

2 Permeability Filter

In [10], an efficient approximate edge-aware filter is introduced engaging computationally efficient two pass integration approach by weighted and connected support regions. The main motivation behind PF is to calculate the output as a result of infinite impulse response (IIR) type filter in a recursive manner. For this purpose, successive weighted summations are conducted among horizontal and vertical axes that

yield adaptive 2D aggregation. The weights (μ) correspond to pixel similarities, i.e. data transfer rates among four neighboring pixels ($N^4(x)$) given as

$$\mu_{x \to z} = e^{-|I(x)-I(z)|/\sigma} \qquad z \in N^4(x) \tag{1}$$

where $I(x)$ indicates the intensity value of the pixel with index of x and σ corresponds to the smoothing factor.

Successive weighted summation (SWS) rules for horizontal filter are given as,

$$\begin{aligned} C^{LtoR}(x) &= C(x) + \mu^R(x-1)\, C^{LtoR}(x-1) \\ C^{RtoL}(x) &= C(x) + \mu^L(x+1)\, C^{RtoL}(x+1) \\ C^H(x) &= C^{LtoR}(x) + C^{RtoL}(x) \end{aligned} \tag{2}$$

where $C(x)$ corresponds to the data to be filtered (vertical index, y, is dropped for simplicity), C^{LtoR} and C^{RtoL} are the left-to-right and right-to-left aggregations; when the scans are completed, the input is updated by the unification of these values that provides horizontal accumulation, $C^H(x)$. The same update and progress rule is utilized along vertical axis to extend supporting regions to 2D as follows:

$$\begin{aligned} C^{TtoB}(y) &= C^H(y) + \mu^B(y-1)\, C^{TtoB}(y-1) \\ C^{BtoT}(y) &= C^H(y) + \mu^U(y+1)\, C^{BtoT}(y+1) \\ C^F(y) &= C^{TtoB}(y) + C^{BtoT}(y) \end{aligned} \tag{3}$$

Top-to-bottom (C^{TtoB}) and bottom-to-top (C^{BtoT}) aggregations are conducted on the horizontally filtered data and the final aggregated cost, C^F, is obtained by summing both as given in (3). This approach has some similarities with separable linear-time-invariant (LTI) filters based on the orthogonal decomposition. The characteristics of the permeability filter depend on the transfer rate distribution of pixels, in which data penetration is prevented along low weights (edges) and permitted along smooth color variation that provide high transition rates.

The effective filter coefficients for two pixels are illustrated in Figure 1 after horizontal and vertical SWS. It is important to note that the support regions are not restricted by any pre-defined window sizes which provides complete content adaptability. In Figure 1.a, the effective horizontal weight distribution of each pixel on the same column is given, where lighter regions correspond to higher weights. It is clear that color-wise smooth regions provide high correlative supports within, and the data transfer is prevented along edge regions. In Figure 1.b, vertical effective weights of the corresponding pixels are illustrated which are the result of two pass vertical transfer. The final 2D effective support weights in Figure 1.c are constructed by further weighting horizontal support region with vertical weights.

The effect of PF can be simulated by direct calculation of weighted summation through the supporting weights in Figure 1.c which correspond to contribution of neighboring pixels. However, compared to direct implementation, a permeability filter requires only six additions and four multiplications per pixel.

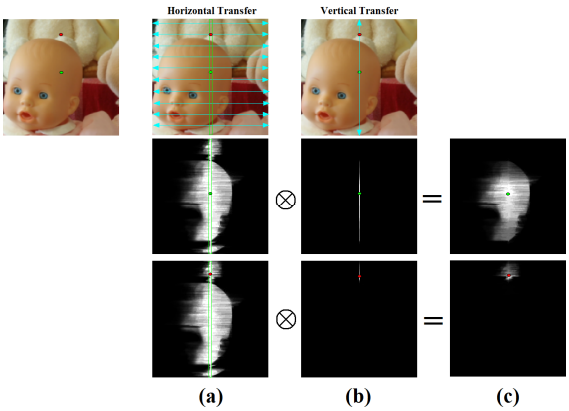

Fig. 1. (a) Horizontal effective weights of the pixels on the same column, (b) vertical effective weights for squared pixel, (c) 2D effective weights after horizontal and vertical SWS

2.1 Stereo Matching via PF

Application of an edge-aware filter to local stereo matching is provided by conducting aggregation on disparity dependent cost functions. For this purpose, pixel-wise cost values of each disparity candidate are calculated as follows:

$$\begin{aligned} C_d^{SAD}(x) &= \left| I_{left}(x) - I_{Right}(x+d) \right| \\ C_d^{CENSUS}(x) &= Ham(CT_{left}(x), CT_{Right}(x+d)) \\ C_d(x) &= \alpha . C_d^{SAD}(x) + (1-\alpha) C_d^{CENSUS}(x) \end{aligned} \tag{4}$$

where $C_d^{SAD}(x)$ corresponds to the *SAD* cost value of the pixel (x) in the left image for disparity d, I_{left} and I_{right} are left and right images. For the *census* measure $C_d^{CENSUS}(x,y)$, Hamming distance (*Ham*(.)) between the bit streams of the correspondences in the *Census* Transformed images (*CT*) is calculated. This step involves noisy measures that yield insufficient correlation between candidate pixel matches in left and right images. Hence, these cost values are filtered, especially through edge-aware filters, to provide crisp and reliable correlation measures. Then, for each pixel the disparity candidate with minimum filtered cost is determined which finalizes the initial disparity estimation. These operations are conducted for left and right pairs independently resulting in two disparity maps. The consistency between stereo pairs is enforced by cross check and occlusion handling steps.

As analyzed in [11], each step has an influence on the quality of the estimated disparity maps. Among all, the most crucial step is the filtering of cost values that determines the computational complexity and accuracy of the algorithms. In [10], it has been shown through experiments on well known Middlebury stereo pairs that permeability filter is one of the most efficient techniques providing comparable even better performance against various edge-aware filters. Besides, an extensive discussion is given in [12] for the utilization of various cost functions. According to [12], unification of sum of absolute difference (SAD) and Census transform (CT) yield robust measures providing high quality estimates compared to the remaining cost

functions. Tough, in this study, permeability filter and SAD+CT measure are considered as the fundamental tools for cost calculation and aggregation steps.

It is agreed that the real world data have several imperfections that affect the performance of stereo matching algorithms. In this manner, direct application of state-of-the-art techniques for challenging data might not provide reliable estimates, since they are developed on well-defined static stereo pairs with reduced imperfections, such as Middlebury stereo benchmark [2]. Typical results for direct application of SAD+CT based cost calculation and PF [10] are illustrated in Figure 2 and Figure 3 addressing lack of texture, high occlusion and temporal inconsistencies. It is obvious that the performance of stereo matching is not sufficient to determine scene geometry. Though, several modifications are required over traditional local approaches to obtain robust stereo matching for real world data having imbalances between left-right pairs.

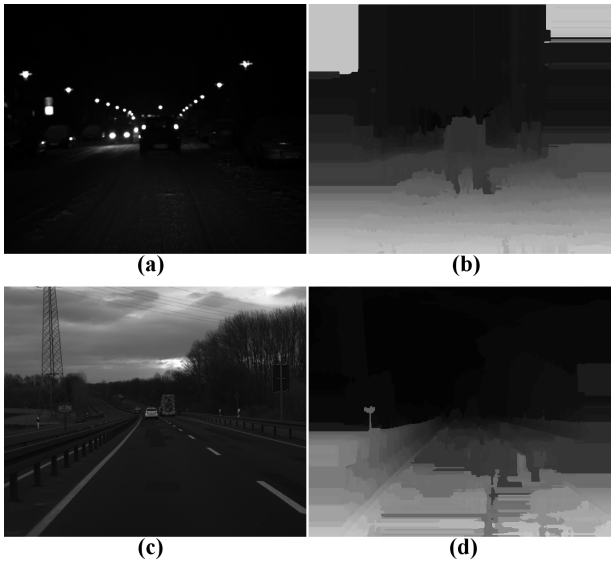

Fig. 2. (a), (c) two frames with lack of texture, (b), (d) estimated disparity maps involve resolution and quantization loss

Fig. 3. First row: Three consecutive frames involving large occluded regions and texture inconsistencies, second row: erroneous disparity maps due to imperfections

3 Proposed Stereo Matching Algorithm

In this study, a stereo matching algorithm dedicated to surveillance and automatic vehicle control applications is introduced. The real world data for these types of applications involve various inconsistencies between stereo pairs and along consecutive frames corresponding to spatial and temporal imbalances. In order to develop reliable stereo matching techniques for these types of data, specific scene structure and requirements should be analyzed that determine fundamental road maps for algorithm development. Thus, an analysis is given in the following section addressing the specific problems and characteristics of the input stereo data. Then, several modifications are proposed to boost up stereo matching performance.

3.1 Constraints on Surveillance and Vehicle Control Application

For the problem of surveillance and vehicle control, stereo camera is ported in front of a car and the road is observed by this camera setup. Two fundamental aims in such a scenario are estimation of 3D structure of the road terrain and distances of the vehicles in the field of view at a time instant. These aims yield two implications; high disparity differences are observed among pixels on the same column with different vertical coordinates due to the orientation of road ground plane (as illustrated in Figure 4) and scene change is not severe among consecutive frames due to physical limitations on the vehicle speed.

The first implication contradicts with the smoothness assumption of the stereo matching algorithms, enforcing the same disparity levels among color-wise similar neighbor pixels. The disparity variation might be lost during local stereo matching due to uniform texture variation on the road. The similarities between pixels yield same disparity levels introducing compaction as illustrated in Figure 2.b-d. This case is also valid under extreme lightening conditions such as night views and flare situations. Therefore, a relaxation is required to allow disparity changes.

The second implication enforces temporal smoothness between consecutive frames by cumulating temporal data, which increases robustness and accuracy of the estimation. Though, each frame should be related to the previous frames instead of independent operations.

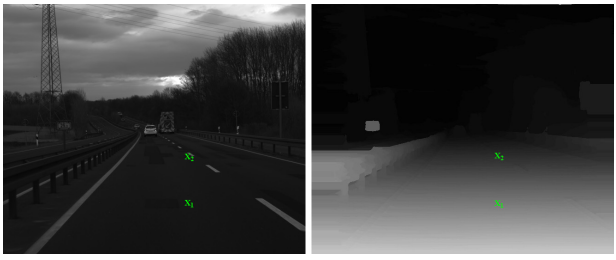

Fig. 4. Marked pixels with almost same intensity values on the same column have 21 disparity level differences in the disparity map

3.2 Proposed Improvements

In order to address aforementioned spatial and temporal specifications, two improvements are introduced for permeability based stereo matching.

3.2.1 Spatial Improvements

In order to relax smoothness of pixels only in vertical direction, a *vertical damping factor* is included during SWS along vertical axis. This goal is achieved by updating the formula in (3) as follows:

$$\begin{aligned} C^{TtoB}(y) &= C^{H}(y) + \lambda.\mu^{B}(y-1).C^{TtoB}(y-1) \\ C^{BtoT}(y) &= C^{H}(y) + \lambda.\mu^{U}(y+1).C^{BtoT}(y+1) \\ C^{F}(y) &= C^{TtoB}(y) + C^{BtoT}(y) \end{aligned} \tag{5}$$

where constant λ scaling is applied to the update term involving previously aggregated values in vertical direction. The effect of this parameter depends on the vertical distance between two pixels which is observed as a power term. Actually, this term can be considered as a spatial range function in vertical axis which damps the accumulation of cost values independent of color-wise similarity. In this way, vertical aggregation is softened enabling disparity variation between vertical neighboring pixels with same texture characteristics. The relaxation of disparity values eliminates compaction of disparity levels providing smoothly varying road terrain. It is important to note that scaling factor affects the degree of relaxation which should be limited to preserve smoothness of the disparity map.

In Figure 5, the effect of scaling factor on the distribution of effective aggregation weights is illustrated for three pixels on different types of scenes from the challenge database [3]. The support regions for the corresponding pixels extend to distant pixels in vertical direction due to lack of texture, in which pixels are forced to be located on the same disparity levels. On the other hand, utilization of a damping factor limits the support regions along vertical axis, yielding relaxation. It is clear that as λ decreases, support regions tend to shrink and after a certain level vertical aggregation is not observed. The effect of vertical damping on disparity estimation is illustrated in Figure 6 in which disparity resolution is preserved with the introduction of vertical scale. As expected, noise artifacts are observable when the scale factor is decreased below a certain level. According to parametric analyses, setting λ to the value 0.9 is optimal to provide a balance between disparity resolution and smoothness.

3.2.2 Temporal Improvements

Treating each frame in a stereo video independently is a sub-optimal solution to provide disparity. In such a case, flickers and inconsistencies within disparity maps of consecutive frames could be observed which decreases accuracy of estimation. Therefore, relations between following frames should be exploited to provide consistent estimates.

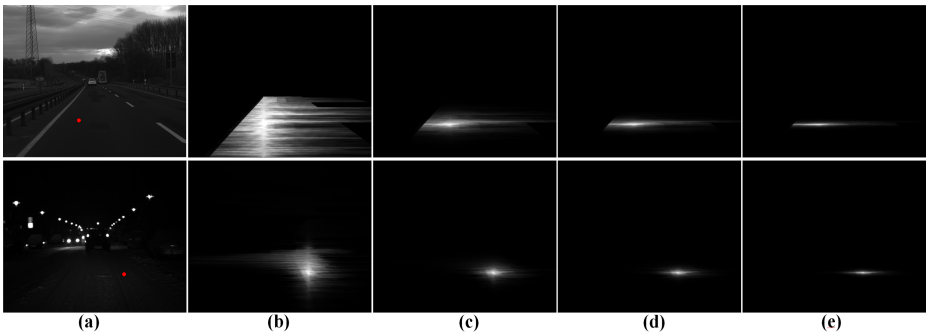

Fig. 5. (a) Intensity views; effective weight distribution for marked pixels via (b) no vertical scaling (λ=1.0) , (c) λ=0.95, (d) λ=0.9 and (e) λ=0.8

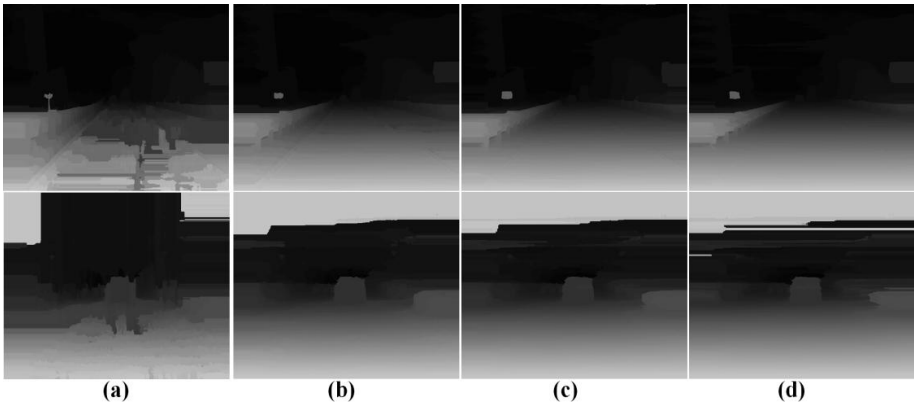

Fig. 6. Estimated disparity maps via (a) no vertical scaling, (b) λ=0.95, (c) λ=0.9 and (d) λ=0.8

Based on the analysis given in Section 3, it can be argued that instant scene changes are not expected in stereo video captured by a system installed within a car. Hence, temporal smoothness among consecutive frames is a valid assumption. On the other hand, flares or reflections, as well as windshield wiper on a vehicle, introduce observable changes in the captured video as illustrated in Figure 3. During frame-by-frame processing, these types of cases cannot be handled that degrades estimation accuracy, whereas exploiting temporal data should improve the performance.

For this purpose, two modifications are proposed in this study that aims to transfer reliable data along time axis. Stemming from the assumption of temporal smoothness, a conventional scene change analysis technique is applied by comparing histograms of two consecutive frames. Under normal conditions in which reflections, flares and other sources of disturbances do not introduce inconsistency; hence, histogram change is expected to be limited. On the other hand, there is significant difference in the histograms for unexpected scene changes. Typical frames with high histogram changes are illustrated in Figure 7 for three different stereo video.

As the first step, the percentage of histogram change is calculated between previous and current frames as follows:

$$\Delta^t = \frac{1}{N} \sum_{i=1:255} \left| Hist^t(i) - Hist^{t-1}(i) \right| \tag{6}$$

where N is the total number of pixels, $Hist^t$ is the histogram of the frame at time instant t. The rate of change (Δ) is utilized as a weighting function to model in temporal transfer of data relating permeability weights. Hence, permeability weights in the current frame are weighted by the permeability values in the previous frame as follows:

$$\overline{\mu}_t(x) = (1 - \Delta^t).\overline{\mu}_t(x) + \Delta^t.\overline{\mu}_{t-1}(x) \tag{7}$$

where $\overline{\mu}_t$ is the permeability vector involving weights in four fundamental directions for time instant t. The update formula in (7) enforces utilization of permeability weights in the current frame, as long as significant scene change is not observed. On the other hand, when there is significant scene change, which is not an expected case, permeability weights of the previous frame are utilized. Once the weights are calculated, histogram of the current frame is updated by the change factor for the analysis of the next frame as follows:

$$Hist^t(i) = (1 - \Delta^t).Hist^t(i) + \Delta^t.Hist^{t-1}(i) \tag{8}$$

Fig. 7. Inconsistent frames yielding significant changes in histogram for three different scenes given in the first column

The histogram update in (8) provides robustness against multiple inconsistent consecutive frames. Therefore, data from the last reliable frame is transferred to the frames involving severe flares, reflections and sudden large occlusions as soon as a consistent frame is encountered with similar histogram characteristics.

The other temporal modification is provided by enforcing smoothness of disparity values along pixels with low intensity change between consecutive frames. For this purpose, temporal permeability weights are calculated for each pixel as,

$$\mu^t(x) = \exp(-\left|I^t(x) - I^{t-1}(x)\right| / \sigma), \tag{9}$$

where I^t is the intensity image for time instant t, σ is a scaling factor (set as 16). Temporal permeability relates the change of the corresponding pixel in time, which is utilized to enforce disparity values of the previous frame to the estimation of current disparity value. This is provided by including a smoothness term in the cot function as

$$C^{new}{}_d(x) = C_d(x) + \mu^t(x).\left\|d - D^{t-1}(x)\right\|, \tag{10}$$

where d is the candidate disparity value and $D^{t-1}(x)$ is the estimated previous disparity value of pixel x.

4 Experimental Results

The improvement of stereo matching performance is validated by utilization of a recent challenging stereo video dataset provided by [3]. Algorithm parameters are kept constant among all video sequences as well as same intermediate steps are utilized for the original and modified versions of PF based stereo matching. The evaluation of estimation performance is provided by visual comparison including spatial and temporal consistency due to lack of ground truth disparity maps. For this purpose, estimated disparity maps are illustrated in Figure 8 to Figure 11 for 4 different challenging videos. For the sake of completeness, disparity maps for five instants with 5-10 frames differences are illustrated.

In Figure 8, results of the *Crossing Cars* [3] sequence are shown, that involve consistent frames with no specific visual artifact. In the second row and third row disparity estimates with and without the proposed modifications in Sections 3.2.1. and 3.2.2 are illustrated. The effect of modifications is observable for temporal consistency of the disparity assignment of the regions below the moving car.

On the other hand, in Figure 9, for *Night and Snow* [3] sequence involving lack of texture and severe lightening changes, spatial and temporal modifications boost-up stereo matching performance. Proposed vertical damping (λ=0.9) yield spatially smooth ground plane, which is especially observable in 4th and 5th frames. Besides, estimated disparity maps are consistent among time, which is robust against changes due to headlights of cars and effects of windshield wiper.

The results of *Rain Blur* [3] sequence are illustrated in Figure 10. Lack of texture, reflections from wet road and motion of windshield wiper are the main causes of challenges in *Rain Blur* sequence. According to visual interpretation, it is obvious that modifications improve quality and consistency of disparity maps with increased disparity resolution and robustness against temporal fluctuations. However, reflections are still problematic due to insufficiency of cost function that is valid for each frame.

The effect of sun flares is observed in *Sun Flare* [3] sequence, introducing severe degradation in disparity estimation, when spatial and temporal modifications are not

exploited. On the other hand, temporal modifications especially increase reliability of estimation process providing for this sequence. This improvement is clearly observable as the strength of sun flare increases.

The robustness against windshield wiper motion and rain flares is further illustrated in Figure 13 for the *Rain Flare* [3] sequence. Modifications on [10] provide obvious improvement on the estimated disparity maps.

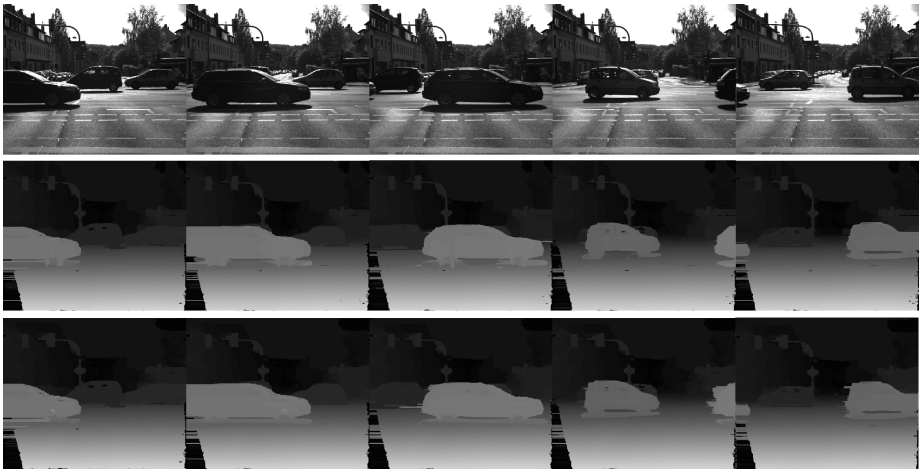

Fig. 8. First row: 5 frames from *Crossing Cars* sequence, second row: disparity maps via [10], last row: proposed stereo matching results after spatial and temporal modifications

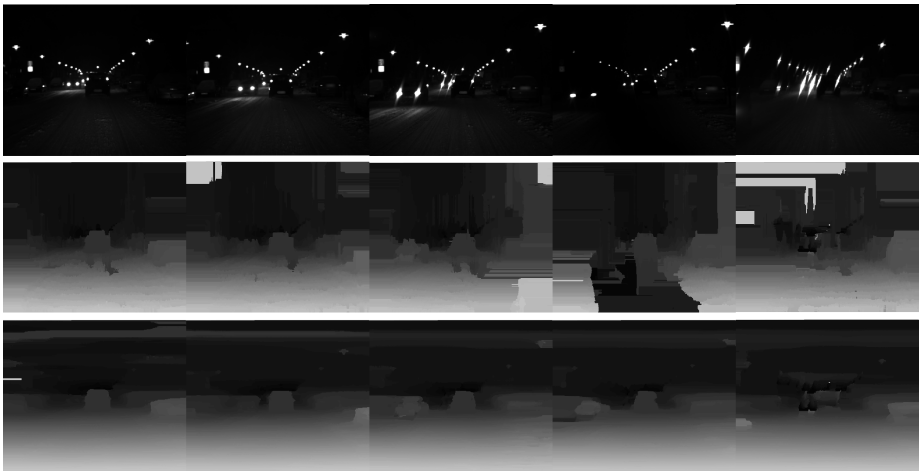

Fig. 9. First row: 5 frames from *Night and Snow* sequence, second row: disparity maps via [10], last row: proposed stereo matching results after spatial and temporal modifications

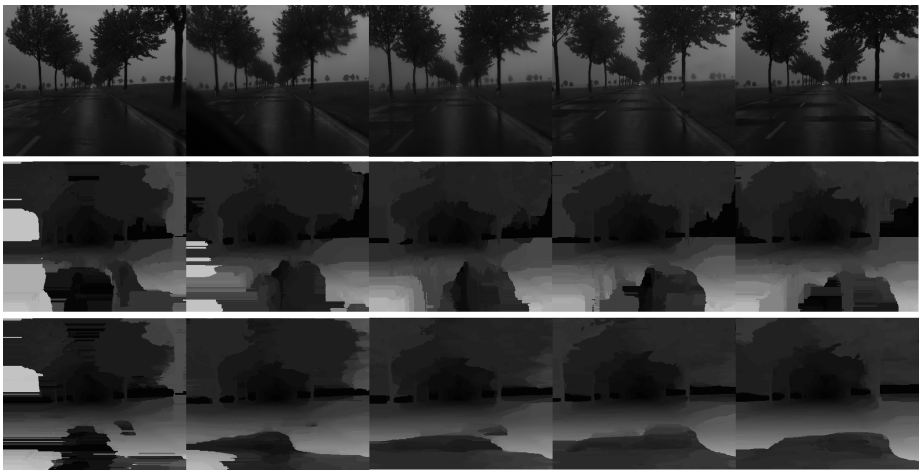

Fig. 10. First row: 5 frames from *Rain Blur* sequence, second row: disparity maps via [10] last row: proposed stereo matching results after spatial and temporal modifications

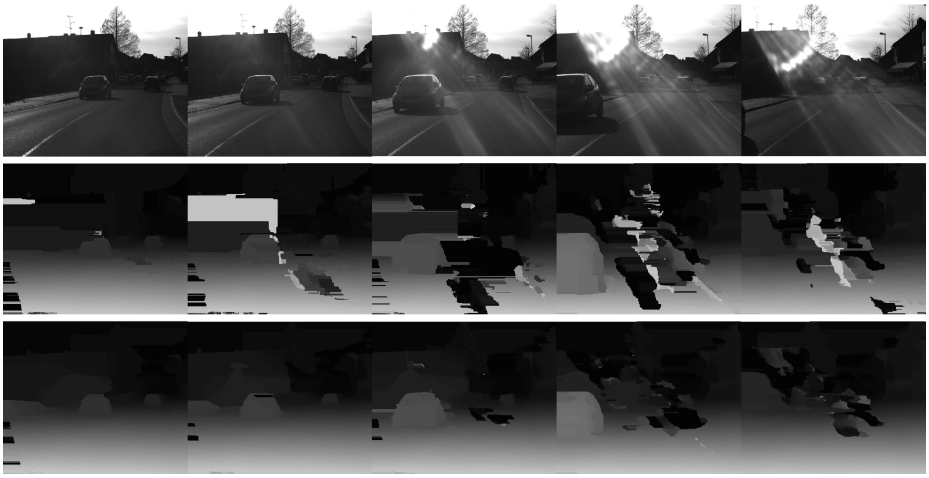

Fig. 11. First row: 5 frames from *Sun Flare* sequence, second row: disparity maps via [10], last row: proposed stereo matching results after spatial and temporal modifications

Apart from visual quality, the proposed approach enables prompt processing with low computational complexity and memory requirement. In this manner, temporal data is transferred by only keeping the previous disparity map rather than full cost volume that requires large memory. Besides, as shown in [10], permeability filter is one of the most efficient edge-aware filtering techniques in literature. In table 1, computation time of aggregation via permeability filter and the well known *guided filter* [7] are illustrated for a pair with resolution of (720x576) on a 3.06GHz Intel Core i7 CPU with 6 GB RAM. During experiments, the number of disparity candidates is set to 50. It is clear that, permeability filter enables almost 12 times faster operation compared to *guided*

filter. This is an expected result according to the required number of additions and multiplications for aggregation, such that permeability filter exploits six additions and four multiplications, while guided filter involves 107 additions and 43 multiplications on the average.

Table 1. Computation time comparison for aggregation of 50 disparity candidates on stereo pair with size of 720x576

Computation time (msec)	Permeability Filter [10]	Guided Filter [7]
720x576x50	**1034**	**12720**

5 Conclusion

In this study, a novel approach is presented for providing spatial and temporal improvements for stereo matching algorithms that improves estimation accuracy for challenging videos. Specifically dedicated to automatic vehicle control and surveillance applications, the proposed modifications exploit general scene characteristics of stereo video captured by a camera imported in front of a car. The specifications of road plane geometry and temporal consistency between consecutive frames enable further assumptions for stereo matching to increase robustness against several imperfections, such as lack of texture, sun flare, rain flare, reflections, changing lightening and occlusions due to motion of windshield wiper. For this purpose, some modifications are provided for the stereo matching algorithm introduced in [10], that is one of the most efficient local techniques. In this manner, a vertical damping scale is included relaxing smoothness of disparity maps along vertical axis which prevents disparity resolution loss due to lack of texture. Besides, temporal consistency is enforced by scene change analysis and transfer of reliable data to consecutive frames. According to the extensive experiments on a recent challenging stereo video dataset, the proposed modifications provide reliable disparity maps under severe challenges and low texture distribution improving scene analyses for surveillance related applications.

References

[1] Sharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal on Computer Vision 47, 7–42 (2002)
[2] http://vision.middlebury.edu/stereo/
[3] http://hci.iwr.uni-heidelberg.de//Static/challenge2012/
[4] Yoon, K.-J., Kweon, I.S.: Adaptive support weight approach for correspondence search. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(4), 650–656 (2006)
[5] Porikli, F.: Constant time O(1) bilateral filtering. In: CVPR (2008)
[6] Qingxiong, Y., Kar-Han, T., Ahuja, N.: Real-Time O(1) Bilateral Filtering. In: CVPR, pp. 557–564 (2009)

[7] He, K., Sun, J., Tang, X.: Guided Image Filtering. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 1–14. Springer, Heidelberg (2010)
[8] Rhemann, C., Hosni, A., Bleyer, M., Rother, C., Gelautz, M.: Fast cost-volume filtering for visual correspondence and beyond. In: CVPR 2011 (2011)
[9] Zhang, K., Lu, J., Lafruit, G.: Cross based stereo matching using orthogonal integral images. IEEE Transactions on Circuits and Systems for Video Technology 19(7) (July 2009)
[10] Cigla, C., Alatan, A.A.: Efficient Edge-Preserving Stereo Matching. In: ICCV LDRMC Workshop (2011)
[11] Tombari, F., Mattoccia, S., Di Stefano, L., Addimanda, E.: Classification and Evaluation of Cost Aggregaton methods for stereo Correspondence. In: International Conference on Computer Vision and Pattern Recognition (2008)
[12] Hirschmüller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(9), 1582–1599 (2009)