Information permeability for stereo matching

Cevahir Cigla\textsuperscript{a,*}, A. Aydin Alatan\textsuperscript{b}

\textsuperscript{a} ASELSAN Inc., Turkey
\textsuperscript{b} Electrical and Electronics Department, Middle East Technical University, Turkey

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 12 November 2012
Received in revised form 4 March 2013
Accepted 8 April 2013
Available online 18 April 2013

\textbf{Keywords:}
Dense disparity map
Information permeability
Successive weighted summation
Separate filtering

\textbf{A B S T R A C T}

A novel local stereo matching algorithm is introduced to address the fundamental challenge of stereo algorithms, accuracy and computational complexity dilemma. The time consuming intensity dependent aggregation procedure of local methods is improved in terms of both speed and precision. Providing connected 2D support regions, the proposed approach exploits a new paradigm, namely separable successive weighted summation (SWS) among horizontal and vertical directions enabling constant operational complexity. The weights are determined by four-neighborhood intensity similarity of pixels and utilized to model the information transfer rate, permeability, towards the corresponding direction. The same procedure is also utilized to diffuse information through overlapped pixels during occlusion handling after detecting unreliable disparity assignments. Successive weighted summation adaptively cumulates the support data based on local characteristics, enabling disparity maps to preserve object boundaries and depth discontinuities. According to the experimental results on Middlebury stereo benchmark, the proposed method is one of the most effective local stereo algorithms providing high quality disparity models by unifying constant time filtering and weighted aggregation. Hence, the proposed algorithm provides a competitive alternative for various local methods in terms of achieving precise and consistent disparity maps from stereo video within fast execution time.

1. Introduction

In recent years, many algorithms have been developed for stereo disparity estimation and an excellent taxonomy for these algorithms is given in [1], where the classification is performed based on matching cost, disparity optimization and disparity refinement stages. Actually, diversity of the algorithms mostly relies on the second stage where optimization approach determines the main characteristics (complexity, precision) of the methods. In that manner, stereo matching algorithms can be analyzed with four fundamental optimization approaches, local, global, cooperative and semi-global.

In local methods [2–12], disparity map assignment is achieved via “winner-take-all” optimization by treating each disparity candidate independently. The matching cost function is aggregated through summation or averaging over a support region and the disparity providing minimum cost is assigned to the corresponding pixel. The second group involves global optimization algorithms [13–21], which are more complex, whereas more precise compared to the local methods. These methods are formulated in an energy-minimization framework and the objective is to optimize the global energy for the estimated disparity map. Cooperative methods [22–25] have been developed to unify advantages of local and global methods by handling occlusions, object boundaries where depth discontinuities are observed and un-textured regions.
These methods, similar to region based global algorithms, rely on the assumption that scenes are composed of non-overlapping planar patches all of which correspond to pixel groups involving color-wise similar pixels. The last class is the semi-global methods [26–29] which involve dynamic programming (DP) optimization. These methods are provided to decrease computational complexity of the global algorithms which are NP-hard in general. From that point of view, global optimization is performed for each scan-line (row) independently resulting in a polynomial time.

Rapid execution time, robust disparity estimation and low memory requirement are the fundamental constraints of stereo matching algorithms, especially in consumer electronics applications. Moreover, for the next generation 3D TVs, robotic applications and surveillance systems, stereo estimation algorithms should require less memory, provide less computational complexity and less precision loss in the extracted 3D models. In that manner, local window-based methods are the strongest candidates for faster disparity calculations [30–38]. The local methods typically do not involve iterative steps which provide simplicity and fast operations, since they do not utilize full cost volume they require low memory compared to other methods. As a result, these methods become available for real-time implementations on special platforms [37] and the efforts on efficient window based algorithms gain popularity by the requirement of disparity maps on variety of 3D systems.

In this paper, a novel paradigm, namely information permeability, is introduced engaging computationally efficient two pass integration approach by weighted and connected support regions. Such an approach enables non-iterative diffusion among wide support area providing not only high quality disparity maps, but also yielding constant operation time. The proposed method introduces further computation reduction, at the same time competing disparity map quality to the highest accuracy local methods. The method exploits only aggregation and occlusion handling with no additional post-processing which might further increase the computational complexity. The paper is organized as follows; in Section 2 related work on stereo matching is given. In Section 3, the proposed approach is introduced in detail. Section 4 is devoted to the comprehensive experiments and finally, in Section 5, conclusions and the future directions of this work are given.

2. Related work

The fundamental property of disparity maps that enable high quality processing is accuracy at depth discontinuities and object boundaries. In general, local methods do not achieve high precision at those regions due to improper selection of window sizes; however, in recent years, adaptive support weight approaches [3] have gained attention due to their high quality disparity map estimates. These methods provide weighting among the support window based on intensity similarity. Hence, for each pixel, similarly colored neighboring pixels have higher weights (higher support) during the aggregation; this approach provides uniform weight distribution for smooth regions increasing robustness against noise and sharp weight transition for object boundaries preserving edge characteristics. However, this approach requires independent calculation of support weights for each pixel within a neighborhood that increases computational complexity. Therefore, several approximations over adaptive weights (AWs) are proposed in order to increase efficiency.

To our best knowledge, the most efficient constant time complexity bilateral filter approximation is introduced in [41] where a piece-wise linear discretization of intensity levels is exploited to achieve weighted averaging within a window. Besides the methods in [43,44] utilize various fast approximations of bilateral filter to enable fast stereo matching. In [42], an alternative edge-preserving filtering technique is presented, image guided filtering, where AWs are provided by a local linear model between the guidance image and filtered data. In [5], guided image filtering is applied for stereo matching where the computational complexity of AWs is reduced drastically by use of box filters over mean and variance of the cost volume.

An extension of adaptive support weights is given in [12] by introducing geodesic distance which forces the support regions to involve connected pixels (i.e. geometry constraint), with an increase in computational complexity. The connected supports prevent the information pass between different objects with similar color distributions; hence, the depth accuracy is increased. Adaptive weight (AW) based algorithms provide crisp disparity assignments at object boundaries and smooth variation at other regions; however, the computational complexity is increased, especially as the support window size becomes larger. So far, the aforementioned techniques exploits soft weights providing intensity adaptive weight distribution; this characteristics is converted to hard weights in [4] by arbitrary support region to yield constant time complexity. The aggregation is performed by two passes in vertical and horizontal direction over integral images providing constant weights within each support region. The most obvious advantage of the approach presented in [4] is the constant time filtering independent of window size, which enables prompt operation. Hence, this method [4] seems to be the most efficient alternative for the computationally complex AW method [3], providing connected and arbitrary support regions depending on local color variation among the images. However, assigning constant weights within the support region decreases the performance of disparity estimation which may not be desired in certain cases. Hence, a local method should utilize connected, as well as weighted support regions to obtain high accuracy disparity maps, which is confirmed by [12]. However, computation complexity should be minimized by exploiting constant time operations, such as two pass integral images introduced in [4] to enable real-time implementations.

It is important to note that edge adaptive local methods [4,41], (and [42] with constant complexity) exploit box filters via integral images which require scanning of the whole image. Two-pass weighted integration approach introduced in this paper resembles integral image phenomenon in terms of horizontal and vertical scanning; however,
box filters are not exploited, in which supporting windows are pre-defined; instead, aggregation is achieved adaptively during passes in four directions.

The horizontal scanning idea of the proposed approach has some similarities with DP [26–27] due to scan line directions, while there are also fundamental differences. DP algorithms operate on full matching costs along a scan line, requiring cost values for each disparity candidate at a certain time. In semi-global matching [28], DP is extended to multiple directions in order to provide reliable support. For this purpose, DP is conducted in 2D, including neighboring pixels among various scanning directions. On the other hand, proposed method operates on image space, such that each disparity candidate is processed independently as a consequence of local winner-take-all optimization which does not require full cost volume. Moreover, additional vertical scan in the proposed approach extends aggregation in 2D which requires less operation compared to multiple scans in [28].

In [39], arbitrary support region [4] and scan line optimization ideas are unified within an additional refinement stage to obtain high quality disparity maps with additional computational complexity. The diffusion based approaches [32,33], utilize 4-neighborhood color similarities to update cost values related to the idea introduced in this study; however, they require high number of iterations to enlarge support regions. In this manner, the proposed algorithm eliminates iteration and support area dependency, while aggregates cost values effectively along large regions by four passes that save computation and memory.

3. Proposed stereo matching algorithm

Similar to most local stereo matching methods, the proposed approach involves the following steps: cost calculation, cost aggregation, minimization and occlusion handling [10]. Once the cost values are calculated based on color similarities between stereo pairs, aggregation of these values is provided by averaging or summation over various support regions. These two steps are conducted sequentially and independently for each disparity candidate. Initial disparity map is obtained by assigning disparity values having minimum supported cost values. In the final step, the occluded and unreliable regions are detected among initial disparity maps and handled by post processing. Cost aggregation and occlusion handling steps are the most discriminative steps that determine the performance of the algorithms in terms of computational complexity and accuracy.

The key idea behind most of these innovative steps in this manuscript is to obtain weighted averaging over support regions within constant time. Although, applications of the integral image paradigm to vision problems have led size-independent complexity for any matching or searching step, obtaining a weighted sum for the same regions requires additional iterations further decreasing efficiency of integral images. In that manner, a novel concept, namely permeability filtering (PF), is introduced in this paper involving intensity-dependent two-pass integration over some cost data. The same filtering structure is also exploited after minor modifications for occlusion handling to finalize the disparity maps.

3.1. Cost calculation

There are various metrics to calculate visual similarity of pixels between stereo images. In [45], an excellent analysis of the common endeavored cost functions is given in terms of computational complexity and matching reliability. According to the analysis Census Transform [46,47] is one of the best performing cost functions providing robust pixel matching. Census Transform, \( \overline{CT}(x, y) \) of a pixel \((x, y)\), is the bit stream obtained through comparison of intensity levels between neighboring pixels according to

\[
\overline{CT}(x, y) = \begin{cases} 
1 & \text{if } I(x, y) > I(x_n, y_n) \\
0 & \text{else} 
\end{cases}, \quad (x_n, y_n) \in N(x, y) \tag{1}
\]

where \( n \) corresponds to the clock-wise (or counter clock-wise) order of neighboring pixels \((x_n, y_n)\) in the bit stream. Hence, within a specified window (of size \(k \times k\)), the pixel is represented by a binary codeword of length \(k^2-1\), forming the transformed image. During matching stage, Hamming distances [46] between the corresponding bit streams are calculated as a cost function. The superior performance of SAD-Census Transform comes with an increase in computational complexity especially due to Hamming distance calculation. On the other hand, sum of absolute difference (SAD) is one of the most endeavored metrics due to its computational simplicity. Hence, in this study these two metrics are exploited to provide a trade-off between accuracy and complexity, as recently proposed in [39]. Assuming that the stereo pair is horizontally aligned, the cost values corresponding to a candidate disparity \(d\) are determined by shifting pixels in one image onto the other image along horizontal direction, as

\[
C_d(x, y) = \min( \sum_{i=1}^{3} |I_{Left}(x, y, i) - I_{Right}(x + d, y + i)|, T)
\]

\[
C_d^{CENSUS}(x, y) = \text{Ham}(\overline{CT}_{Left}(x, y), \overline{CT}_{Right}(x + d, y))
\]

\[
C_d(x, y) = \alpha C_d^{SAD}(x, y) + (1-\alpha)C_d^{CENSUS}(x, y), \tag{2}
\]

where \( C_d^{SAD}(x, y) \) corresponds to the SAD cost value of the pixel \((x, y)\) in the left image for disparity \(d\), \(I_{Left}\) and \(I_{Right}\) are three-channel (RGB) left and right images. During SAD cost calculation, occluded or overlapped regions cannot be handled, since 3D structure is not known; hence, truncation of the cost values is performed to keep maximum SAD below a level \((T)\) for enabling disparity values to be assignable depending on local neighboring support. For the census measure \( C_d^{CENSUS}(x, y) \), the Hamming distance \((\text{Ham}[]))\) between the bit streams of the correspondences in the census Transforme images (CT) is calculated. In this study, for the census transform a \((5 \times 5)\) window is exploited. Once the SAD and CENSUS cost measures are obtained, the cost function \( C_d(x, y) \) is set as the linear combination of both according to a weighting factor, \(\alpha\). The selected values of the parameters are further discussed in the experimental results section.
3.2. Cost aggregation

Cost aggregation is the most important step effecting the general performance of the proposed and all other local stereo matching algorithms. In general, for each candidate disparity value, aggregation is performed along spatial directions independently, and then, the aggregated cost values are compared in order to determine the best disparity candidate with the minimum cost. The proposed aggregation approach involves three main steps; firstly, for each pixel depending on color similarity between neighboring pixels, permeability weights, in four directions are calculated in the image space. These weights correspond to the ratio of intensity similarity information that is to be propagated through the corresponding direction and the propagated information should be reset in case of a high contrast edge. These permeability weights are utilized to perform aggregation in horizontal and then vertical directions that are conducted on cost space for each disparity candidate sequentially. The order of horizontal and vertical aggregation can be interchanged optionally; in this study, the presented ordering is utilized, although a similar performance is also observed for the other option. The three step aggregation approach is an extension to the two pass integration approach introduced in [4], so that weighted integration is also achieved.

3.2.1. Permeability weight calculation

“Permeability” is a common term in biomedical engineering, used to derive mathematical model of the behaviors of cell membranes by defining ratio of molecules that can pass through. The same idea can be used to model the transfer ratio of information through an RGB pixel in an image depending on the application, such as edge-preserving color filtering. In this work, following the common assumption in stereo matching, which states that similarly colored pixels belong to similar depth surfaces, permeability is being utilized to support cost values of similarly colored regions. For that purpose, a permeability metric is constructed to set high transfer ratio for color-wise smooth regions and low transition ratio for discontinuous color regions.

Among many alternatives, in this study, the exponential function depending on color differences is utilized to model data permeability, \( \mu_{x \to y} \), from pixel \( x \) to \( y \) as

\[
\mu_{x \to y} = F_R(I_x, I_y) \quad y \in N^4_x
\]

\[
F_R(I_x, I_y) = \min(e^{-\Delta R/2\sigma}, e^{-\Delta G/2\sigma}, e^{-\Delta B/2\sigma})
\]

(3)

where \( \Delta R \), \( \Delta G \) and \( \Delta B \) indicate the absolute difference between the \( R \), \( G \) and \( B \) values of the two spatially neighboring pixels, \( \sigma \) corresponds to the smoothing factor, \( N^4_x \) is the 4-neighborhood of pixel \( x \). Permeability, \( \mu \), is assigned by taking minimum data transfer-rate among three channels, forcing smooth regions to have similar color values for each channel. In order to enable data transfer in a wider range of directions, permeability weights are calculated in four directions, as illustrated in Fig. 1, for each pixel. The absolute color differences are calculated between the center pixel and the first neighboring pixel in the corresponding direction. At that point, median filter with a window size \((3 \times 3)\) can be applied to the input images before permeability calculation to reduce possible noise and improve reliability. As a result, each pixel characteristic is modeled by different \( \mu \) weights, yielding values in the range of \([0, 1]\). The range extends from information-wise impermeable to transparent characteristics, as the value approaches to “1”.

3.2.2. Horizontal support

Horizontal support is achieved by successive weighted summation (SWS) of the cost values, at each time for one disparity candidate, along right and left scan order. Horizontal processing is applied for each row independently; hence, inter scan line relations are not considered at that step. For the right scan order, the cost values are updated starting from the first pixel to the last, and vice-versa, for the left scan order, as in Fig. 2. Both operations are performed by using the initial cost values and two aggregated cost values are obtained after SWS in two directions. The final horizontally supported cost value is obtained by the summation of right and left aggregated cost (\( C^{R}_{d} \) and \( C^{L}_{d} \)) values. In this operation, the most crucial step is SWS, which enables constant execution complexity. In the conventional integral image paradigm, such a weighted summation cannot be achieved through single pass.

In SWS, cost value of a pixel is updated through weighted addition of the previous pixels updated cost value which is scaled by related permeability coefficient according to

\[
C^{R}_{d}^{\text{SWS}}(x) = C_{d}(x) + \mu_{x-1}^{R} C^{R}_{d}(x-1)
\]

(4)

where \( C_{d}(x) \) is the input cost value corresponding to disparity \( d \), for the pixel with horizontal index \( x \) (for simplicity row index is dropped), \( C^{R}_{d}(x) \) is the updated
value in raster scan (left-to-right), $y^R_{x-1}$ is the right permeability coefficient of the previous pixel indicating the transfer ratio. The data from the previous pixel is transferred to the next pixel by permeability weighted cost values. Due to this successive approach, the effect of distant pixels can be transferred by successive multiplications of the $\mu^R$ values. This characteristic can be observed by analyzing the update rule given in (4) and turning it into closed form as

$$
C^{\text{LR}}_{d}(y) = C_d(x) + \mu^R_{x-1} C^{\text{LR}}_{d}(x-1)
$$

$$
= C_d(x) + \mu^R_{x-1} \left[ C_d(x-1) + \mu^R_{x-2} C^{\text{LR}}_{d}(x-2) \right]
$$

$$
= \ldots
$$

$$
= C_d(x) + \sum_{i=1}^{x-1} \left\{ C_d(x-i) \prod_{j=1}^{i} \mu^R_{x-j} \right\}
W^{\text{LR}}_{x-i}
$$

$$
= C_d(x) + \sum_{i=1}^{x-1} C_d(x-i) W^{\text{LR}}_{x-i}
$$

(5)

where the updated cost values are written in an explicit form until the leftmost pixel is reached in this direction. The effective weight $\left(W^{\text{LR}}_{x-y}\right)$ between a pixel with index $y$ and another pixel with index $x (x > y)$ can be obtained by successively multiplying all the right permeability weights between indices $y$ and $x$. The support of the cost values may extend to large distances as long as high permeability weights are observed consecutively. Once an impermeable pixel (i.e., $\mu = 0$) is observed on the left of a pixel, then the information behind the impermeable region cannot be propagated to the right due to multiplicative effects. Actually, this property provides support regions to be connected, i.e., a pixel is supported by previous pixels on the left as long as there is smooth color transition among the path. Some of the approaches from the literature do not enforce such a constraint [4,38]. The set of equations explains the recursive nature of permeability calculation.

For the inverse raster scan, SWS starts from the last pixel on the right and the update rule is reversed as

$$
C^{\text{R}}_{d}(x) = C_d(x) + \mu^R_{x+1} C^{\text{R}}_{d}(x + 1)
$$

$$
C^{\text{R}}_{d}(x) = C_d(x) + \sum_{i=1}^{M} \left\{ C_d(x + i) \prod_{j=1}^{i} \mu^R_{x+i} \right\}
W^{\text{R}}_{x+i}
$$

(6)

where $M$ defines the width of the row, $C^{\text{R}}_{d}(x)$ corresponds to the right-to-left aggregated cost value of the pixel with column index $x$. The final horizontal support ($C^{\text{HOR}}_{d}$) is obtained by the summation of right and left SWS values according to,

$$
C^{\text{HOR}}_{d}(x) = C^{\text{LR}}_{d}(x) + C^{\text{R}}_{d}(x)
$$

(7)

Given the closed form effective representation of right and left scan SWS in (5) and (6), the computational complexity is constant; in other words, for each pixel, horizontal aggregated costs are extracted by a total of two multiplications and two additions for left and right scan orders and one final addition; resulting in two multiplications and three additions per pixel. It is also important to note that, these operations are performed for each disparity candidate to aggregate cost values. As mentioned previously, the effective weights of neighboring pixels can also be determined by successive multiplication of permeability weights in order to make the analysis of the proposed SWS. In Fig. 3, effective horizontal weight distributions for a point on 1D signal is presented. The weight distribution on the right corresponds to the weights of neighboring points that are obtained by summing left and right effective weights according to open forms given in (5) and (6) after two pass SWS. Thus, a brute force aggregation for the point can be calculated by averaging neighbor points with the corresponding effective weights, which is much more complex than the proposed technique. It is clear that, as long as there is a sharp edge, information transfer is prevented, corresponding to an edge-aware characteristic.

3.2.3. Vertical support

In horizontal aggregation, information transfer is performed for each row independently, though vertical relation is not considered. However, to obtain robust cost measures vertical pass is also required extending the support region from 1D to 2D. In that manner, vertical aggregation is performed on the horizontally supported data via SWS, as introduced in the previous section. SWS is applied in upwards and downwards directions independently; then, the partial supports are unified to obtain the final vertical support. The update rule during vertical SWS is similar to its horizontal counterpart by the use of vertical permeability weights.

![Fig. 3. Effective weight distribution of a point on 1D signal after the proposed SWS approach.](image-url)
3.2.4. Effective support

After the vertical pass, 2D support regions are obtained, since the vertical aggregation is performed over the horizontal supported data. In Fig. 4, the final effective support region of a pixel (squared) after horizontal and vertical SWS is illustrated which is valid for any disparity candidate. In Fig. 4b, horizontal weights of the pixels on the same column with corresponding pixel are illustrated, where red color indicates higher weights. It is clear that color-wise smooth regions provide strong supports and data transfer is prohibited along edge regions. The aggregated cost values for these pixels are calculated through weighted averaging according to the weight distribution along the row. SWS provide efficient calculation of these values through left and right scan by only 2 multiplications and 3 additions. In Fig. 4c, the vertical effective weights of the corresponding pixel are illustrated. Finally, the overall 2D support weights of the corresponding pixel are illustrated in Fig. 4d which $C_{\text{rad}}(x)$ represents the effective weights obtained after SWS.

The proposed four direction SWS approach, consecutive horizontal and vertical scans, enables 2D connected and weighted support regions by constant operation, 6 additions and 4 multiplications per pixel for each disparity candidate. Considering direct implementation of the resultant support region, depending on the region size with $(W \times W)$ the computation complexity per pixel for each disparity is $2W^2$ additions and $W^2$ multiplications, which is much larger. On the other hand, the guided image filtering method provides constant operational complexity for the effective edge-preserving filter which involves 18 box filters (3 additions and 2 subtractions per pixel for each), 21 multiplications, 21 additions and 1 $(3 \times 3)$ matrix inversion per pixel per disparity. The proposed method remains quite efficient due to separable horizontal and vertical filtering with the complexity of 6 additions and 4 multiplications; compared to [42] and the other approximations of AWs approach [27,41–44]. The operational complexity of the arbitrary support region approach introduced in [4], with $(12W)$ subtractions and 4 additions per pixel for each disparity, has less disparity dependent complexity (4 additions) than the proposed SWS. However, the first term with $12W$ subtractions, which is independent of number of disparity candidates, yields an additional complexity effective for low number of disparities. Further analyses are provided in the experimental results section.

3.2.5. Comparison of support Regions against other relevant methods

In Fig. 5, the effective support region (valid for any disparity candidate) provided by the proposed approach is compared with well-known AWs approach [3], its extension with geodesic constraint [12] and arbitrary regions [4] in terms of support region reliability. The image guided filtering [42] which is an efficient alternative of [3], provides similar effective weight characteristics as in AWs approach, analyzed in detail by [42] and [5]: hence, the weight distribution of this method is not included. It is clear that AWs does not provide connected regions; on the other hand, geodesic support [12] yields connected and crisp regions with an increase in computational complexity. It is important to note that, for the proposed approach, performing vertical aggregation over horizontally aggregated data extends the support window to 2D among connected support regions as an approximation of geodesic support. However, due to orthogonal scanning, some precision loss is expected, especially along thin and tilted objects with non-vertical and non-horizontal structures. As observed in Fig. 5, for the 2nd image crop, the support area provided by the proposed method is limited compared to AW [3] and geodesic filter [12], which may affect the matching performance. On the other hand, depending on local characteristics, such as smoothness, larger support regions can be provided through the proposed method, since there is no constraint on support area limited by any pre-defined window size unlike all other edge-aware

![Fig. 4](image_url) (a) Colored image crop, (b) horizontal effective weights of the pixels on the same column, (c) vertical effective weights for squared pixel, (d) 2D effective weights after horizontal and vertical SWS.
filters. Hence, permeability filtering extends support areas in certain cases, while losing precision along thin and tilted objects with much lower computation complexity compared to that of [12].

The approach introduced in [4] does not involve weighting; instead exploits simple operations, addition and subtraction. On the other hand, the arm length calculation in [4] requires additional operations along horizontal and vertical direction, increasing the computational complexity to a certain extend. In Fig. 5, smooth variation of the slanted surfaces is modeled by support regions with constant (hard) weights for [4]. This property typically results in loss of disparity resolution, in such a way that similarly colored pixels located at different disparity values force their neighboring pixels to reach the same disparity. Therefore, small disparity variations may be settled to a mean disparity of the support region which decreases the accuracy of estimation. In the proposed approach, however, effective soft weights prevent shrinking of depth variation by decreasing the effect of distant pixels. Based on Fig. 5, it is clear that the boundaries of the support regions are similar for both of the methods; the filtering method in [4] is extended to AWs by the proposed information permeability approach with further efficiency in computation.

3.3. Minimization

As a common procedure to most of the stereo algorithms, minimization procedure is performed by winner-take-all approach. For this purpose, the aggregation of cost values is calculated for each disparity candidate independently and sequentially, and then the candidates with minimum costs are assigned to the corresponding pixels.

3.4. Occlusion handling

The cost calculation, aggregation and minimization steps are performed for both of the images, and at the
end of these steps two initial disparity maps are obtained for the stereo pair. As illustrated in Fig. 6b, initial disparity maps involve errors at the occluded regions, in which the true correspondences cannot be obtained due to invisibility. Those regions should be handled in such a way that reliable information is diffused to assign geometrically consistent disparity values. Hence, as a first step, the reliable and occluded region detection is performed by a cross-check between two disparity maps. The results of the occluded region detection are given in Fig. 6c for the Teddy stereo pair with red colored unreliable and occluded regions.

Although, there could be different ways to compensate for the occlusions, handling of those regions is performed by the extension of the proposed two-pass filtering approach. Permeability-based filtering is applied over the disparity map, taking reliability of the regions into account and the filtered disparities are assigned to the occluded regions. Such an approach provides information diffusion over the occluded pixels based on color-wise similar and reliable pixels. In general, occluded regions are located at the local background (further to camera) of the corresponding regions, since foreground region (closer to camera) is always visible in both of the images. Hence, occlusion handling algorithms [27] diffuse information from background to the occluded regions in various ways. In the proposed approach, diffusion of background information is further supported by color-wise similarity between pixels.

For this purpose, a background weight map, \( \text{BackW}(x, y) \), is generated, in which occluded regions are assigned to the value zero, and reliable regions are assigned to a range of values between \( [\varphi, 1] \) depending on the disparity values, as

\[
\text{BackW}(x, y) = \begin{cases} \varphi & \text{occluded} \\ f(d) & \text{else} \end{cases}
\]  

(8)

where \( d \) is the disparity value of the pixel \((x, y)\), \( \varphi \) is a constant (set to 0.1 throughout experiments) and \( f \) is a linear mapping function that favors pixels at local background, as given in Fig. 7(a). Once the background weights are obtained for left and right images, disparity maps are weighted by these maps as

\[
\begin{align*}
D^\text{weight}_\text{Left}(x, y) &= D_\text{Left}(x, y) \cdot \text{BackW}_\text{Left}(x, y) \\
D^\text{weight}_\text{Right}(x, y) &= D_\text{Right}(x, y) \cdot \text{BackW}_\text{Right}(x, y)
\end{align*}
\]  

(9)

The next step of occlusion handling is determination of appropriate disparity values for the occluded regions that is achieved by permeability based filtering of the weighted disparity and the background maps. At that step, normalization is required to assign disparity values in the range of minimum and maximum disparities. The permeable filtering provides weighted summation over disparity values, possibly resulting in larger values than the disparity values.

In order to provide range-filtered disparity maps, \( D^\text{Norm} \), the filtered data should be normalized by the total effective weights calculated for each pixel, as

\[
D^\text{Norm}(x, y) = \frac{F_{\text{Per}}\left(D^\text{weight}(x, y)\right)}{F_{\text{Per}}(\text{BackW}(x, y))}
\]  

(10)

where \( F_{\text{Per}} \) is the proposed edge-aware filter, \( D^\text{weight} \) is the weighted disparity maps. The filtered and normalized disparity values, \( D^\text{Norm} \), in (10) are assigned to the unreliable or occluded pixels detected in cross-check. As the final step of occlusion handling, a median filter with window size of \((3 \times 3)\) is applied to remove possible noisy assignments. In Fig. 8, the resultant disparity maps are illustrated after occlusion handling with and without background favoring. It is obvious that occlusion handling is a critical step to increase the reliability, as some corrected regions (shown by green circles) are observed.

**Fig. 6.** (a) Left-right color views of Teddy stereo pair, (b) estimated disparity maps without occlusion handling, bright regions correspond to small disparities (i.e. closer to camera), (c) detection of occluded and unreliable regions.
In Fig. 8(b), some leakage (circled in red) from the foreground object is observed, when background favoring is not performed, due to color-wise similarity between foreground and background pixels; whereas, the proposed approach (with background favoring) handles these cases as illustrated in Fig. 8(c). The method assigns geometrically consistent disparity values to the large occluded regions, the leftmost circle, as well. Hence, unification of background favoring and color adaptive filters for the occlusion handling enable crisp and edge-preserved disparity maps which is important for high accuracy.

4. Experimental results and discussions

The test bench in [1] provides an environment to evaluate and compare stereo disparity map estimation algorithms based on four different stereo images with their ground truth disparity maps. Although the test bench lacks a comparison of the algorithms in terms of computational complexity and rendering quality, it still gives a clear idea about the visual quality of the estimated disparity maps on static images. In order to evaluate the robustness of a stereo matching algorithm, further experiments are also required, such as rendering quality. For this purpose, the experiments are performed in three categories; in the first category, performance tests are conducted on stereo images via the test bench provided by [1] based on ground truth disparity map comparison. Moreover, computational complexities are also compared for the proposed algorithm against some well-known local methods [4,5]. In the other category, the rendering quality is evaluated through the use of estimated disparity maps. In the final section, the reasoning behind the parameter selection is presented with in-depth analyses of the algorithmic blocks.

4.1. Comparison against ground truth depth data

The performance evaluation against ground truth disparity data is performed using the Middlebury stereo test bench where color images with their ground truth disparity maps are illustrated in the first two rows of Fig. 9. In the third row, disparity maps estimated by the proposed algorithm are given. During the estimation, truncation parameter, $T$, and weight factor, $\alpha$, in (2) is set to 15 and 0.2, respectively, while smoothing factor, $s$, in (3) is selected as 8. According to the error maps between the estimated and ground truth disparities given in the last row, the proposed method preserves disparity transitions at object boundaries and provides crisp maps.

4.1.1. Accuracy

In Table 1, quantitative results taken from the online Middlebury stereo database are illustrated for local methods. The quality of disparity maps in Table 1, are calculated after comparison with the ground truth over visible pixels based on a disparity difference threshold of 1 pixel. According to these results, the proposed method ranks 30 (for SAD+Census) and 60 (for SAD) among 140 submissions involving complex global optimization methods (such as Belief Propagation), while it remains in the 2nd and 7th ranks for two different cost functions within pure local-based methods. Moreover, the proposed method provides the highest quality disparity maps for Tsukuba, Teddy and Cones stereo images compared to the other local methods, which is another promising result. The
utilization of Census Transform enhances the accuracy of the estimated disparity maps considerably over sole SAD utilization.

On the other hand, the proposed approach outperforms algorithms based on Dynamic Programming optimization [26–31], which share some similarities in terms of scan line processing, as well as the optimum diffusion technique [33] introduced recently; while two extension of [4] that exploit cross support regions, [39,40,48] have better accuracy (Rank 2, Rank 6 and Rank 23). Mei et al. [39] unifies multi-directional scan line optimization and arbitrary shaped support regions with additional post processing, which is much more complex than the proposed approach. In [40], a simpler scan line optimization along ground control points and a disparity refinement step are proposed, which are limited tools and optimized for the given data set. In this manner, they involve some non-local steps over [4], which could also boost-up the performance of the proposed approach. Besides, multi-directional extension of the cross filter [4] is exploited in [48] that has the rank 23; this extension increases the accuracy as well as computational complexity. Recently, histogram based local aggregation, yielding accurate disparity maps, is proposed in [49], where the memory requirement is increased due to histogram integrals. It is important to note that, the recently developed algorithms exploit additional processes or complex cost functions over the well known edge-aware filters to provide high accuracy. Hence, such techniques are valid for the proposed approach as a future work to further increase the

Fig. 9. First row: images (Tsukuba, Venus, Teddy, Cones) from Middlebury stereo database 0, second row: ground truth disparity maps, third row: disparity maps by the proposed algorithm using SAD, fourth row: disparity maps by the proposed algorithm using SAD+CENSUS, fifth row: disparity errors for (SAD+CENSUS) larger than 1 disparity level where gray corresponds to errors at occluded regions and black is for non-occluded regions.
overall accuracy. The Guided Filter [5] outperforms the proposed approach in Table 1, even though its average error is higher. The improvement of the permeability filter over the arbitrary support region approach [4] is almost 25% that illustrates the advantage of soft weights compared to the hard weights.

4.1.2. Complexity

When the aggregation methods utilized by algorithms given in Table 1 are compared in terms of the computational complexity, excluding the proposed algorithm, image guided filtering [42] and the cross-based method [4], all of the methods involve high computational complexity, such that they exploit adaptive support weights on large local windows with some additional operations. The fundamental method of AWS [3] spends almost 100 s with proposed window sizes for Tsukuba image on the environment utilized in this work (Core Duo 1.80 GHz 2 G Ram CPU), which makes the other algorithms in Table 1 to remain much complex in the case where no special hardware is utilized. However, there are efficient approximations of AWS such as [41,43,44], that have comparable efficiency with guided filter [42]. An analysis is also given for the best AW approximation [41] in extended comparison section for sake of completeness. For this reason, the proposed method is compared to [4,42] which are proven to be efficient aggregation algorithms in terms of computational complexity in this section.

It is important to note that the algorithm presented in [5] originally runs on a special parallel platform; however in this study, guided filtering technique [42] exploited in [5] is implemented (in C++) as well as the method in [4], on the current computer environment to compare aggregation steps in CPU. Among the local methods, unless different cost functions and occlusion handling approaches are utilized, total run time is mostly determined by the cost aggregation step. Therefore, a fair comparison in terms of computational complexity between the proposed method. [4, 42] is obtained by the aggregation times in Table 3, where the proposed parameters in the corresponding papers are exploited.

In Table 2, the cost aggregation step for single image is compared for four Middlebury stereo pairs for the given image sizes and disparity ranges. As mentioned previously, the proposed two pass SWS approach requires 6D additions and 4D multiplications per pixel where D corresponds to number of disparity candidates, while the method introduced in [4] requires 4D additions and the image guided filtering [42] requires almost 129D additions, 21D multiplications and D inversion of 3 x 3 matrices (28 multiplications and 14 additions) for each pixel. According to this table, the proposed method consumes less time compared to [4], although higher number of operations is performed. The reason behind such a result is due to the fact that, in [4], arm length calculation requires fix number of operations independent of disparity range. The calculation of arm length requires (12W) subtractions and comparisons, where W is the arm length taken as 20. Thus the total complexity for [4] turns out to be (4D+12W) additions; especially for low number of disparity candidates, the second term has influence on the computational complexity. As the number of disparity levels increases the gap between the proposed approach and [4] is decreased. Compared to guided filter [42], the proposed aggregation technique runs 10–15 times faster, which is an expected result, when the required operations (additions and multiplications) are compared. According to the overall performance, it is clear that the proposed algorithm provides faster aggregation in CPU compared to the state of the art.

The other important constrain for stereo algorithms is the memory requirements which need to be mentioned under complexity characteristics. One of the advantages of local methods is that they do not depend on number of disparity candidates due to utilization of WTA local optimization. Hence, cost calculation and aggregation are calculated for each disparity candidate sequentially among the same memory by replacement, while only the minimum cost and its disparity value are stored per pixel. The proposed approach utilizes data of size (3 x W x H), during horizontal and vertical passes, where W and H are the image sizes. Considering the pixel-wise cost data as (W x H); for scanning, additional (W x H) is required and one for storing minimum cost values. The memory requirement for [4] during aggregation of cost values is also (3 x W x H), which corresponds to one additional (W x H) for integral image calculation and one for minimum cost storage. On the other hand, the requirement goes up to (49 x W x H) for color guided filtering in [42] due to multiple integral image calculation in RGB channels.

Table 1: Rankings of the selected local methods in middlebury database.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Rank</th>
<th>Avg. error [%]</th>
<th>Tsukuba</th>
<th>Venus</th>
<th>Teddy</th>
<th>Cones</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD Census [39]</td>
<td>2</td>
<td>3.97</td>
<td>1.07</td>
<td>0.09</td>
<td>4.10</td>
<td>2.42</td>
</tr>
<tr>
<td>RDP [40]</td>
<td>6</td>
<td>4.57</td>
<td>0.97</td>
<td>0.21</td>
<td>4.84</td>
<td>2.53</td>
</tr>
<tr>
<td>HistoAggr [49]</td>
<td>18</td>
<td>5.20</td>
<td>1.93</td>
<td>0.16</td>
<td>5.88</td>
<td>2.41</td>
</tr>
<tr>
<td>Cross LMF [48]</td>
<td>23</td>
<td>5.13</td>
<td>2.46</td>
<td>0.27</td>
<td>5.50</td>
<td>2.34</td>
</tr>
<tr>
<td>Cost Filter [5]</td>
<td>27</td>
<td>5.55</td>
<td>1.51</td>
<td>0.20</td>
<td>6.16</td>
<td>2.71</td>
</tr>
<tr>
<td>Proposed (SAD-Census)</td>
<td>30</td>
<td>5.50</td>
<td>1.06</td>
<td>0.32</td>
<td>5.60</td>
<td>2.65</td>
</tr>
<tr>
<td>GeoSup [12]</td>
<td>36</td>
<td>5.80</td>
<td>1.45</td>
<td>0.14</td>
<td>6.88</td>
<td>2.94</td>
</tr>
<tr>
<td>AdaptDisp [6]</td>
<td>44</td>
<td>6.10</td>
<td>1.19</td>
<td>0.23</td>
<td>7.80</td>
<td>3.62</td>
</tr>
<tr>
<td>GeoDiff [33]</td>
<td>46</td>
<td>5.49</td>
<td>1.88</td>
<td>0.38</td>
<td>5.99</td>
<td>2.84</td>
</tr>
<tr>
<td>Distinct SM [7]</td>
<td>59</td>
<td>6.14</td>
<td>1.21</td>
<td>0.35</td>
<td>7.45</td>
<td>3.91</td>
</tr>
<tr>
<td>Proposed (SAD)</td>
<td>60</td>
<td>6.33</td>
<td>1.06</td>
<td>1.00</td>
<td>5.86</td>
<td>4.06</td>
</tr>
<tr>
<td>SegSupport [8]</td>
<td>62</td>
<td>6.44</td>
<td>1.25</td>
<td>0.25</td>
<td>8.43</td>
<td>3.77</td>
</tr>
<tr>
<td>CostAggr-Occ [9]</td>
<td>67</td>
<td>6.20</td>
<td>1.38</td>
<td>0.44</td>
<td>6.80</td>
<td>3.60</td>
</tr>
<tr>
<td>AdaptWeight [3]</td>
<td>76</td>
<td>6.67</td>
<td>1.38</td>
<td>0.71</td>
<td>7.88</td>
<td>3.97</td>
</tr>
<tr>
<td>Fast Bilateral [43]</td>
<td>83</td>
<td>7.31</td>
<td>2.38</td>
<td>0.34</td>
<td>9.83</td>
<td>3.10</td>
</tr>
<tr>
<td>VarCross [4]</td>
<td>89</td>
<td>7.60</td>
<td>1.99</td>
<td>0.62</td>
<td>9.75</td>
<td>6.28</td>
</tr>
</tbody>
</table>

* Involve non-local operations.

Table 2: Execution time comparison for cost aggregation step.

<table>
<thead>
<tr>
<th>Image resolution</th>
<th>Tsukuba</th>
<th>Venus</th>
<th>Teddy</th>
<th>Cones</th>
</tr>
</thead>
<tbody>
<tr>
<td>384 x 288</td>
<td>343 x 383</td>
<td>450 x 375</td>
<td>450 x 375</td>
<td></td>
</tr>
<tr>
<td>Disparity candidate</td>
<td>16</td>
<td>20</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Var. cross [4] (s)</td>
<td>0.134</td>
<td>0.192</td>
<td>0.536</td>
<td>0.572</td>
</tr>
<tr>
<td>Guided filter [42] (s)</td>
<td>1.145</td>
<td>2.214</td>
<td>6.758</td>
<td>6.942</td>
</tr>
<tr>
<td>Proposed (s)</td>
<td>0.116</td>
<td>0.174</td>
<td>0.501</td>
<td>0.515</td>
</tr>
</tbody>
</table>
The accuracy comparison for the extended stereo set.

<table>
<thead>
<tr>
<th>Error %</th>
<th>$\Delta d &gt; 1$</th>
<th>$\Delta d = 1$</th>
<th>$\Delta d &gt; 2$</th>
<th>$\Delta d = 2$</th>
<th>Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visible</td>
<td>All</td>
<td>Visible</td>
<td>All</td>
<td>(s)</td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>7.98</td>
<td>14.15</td>
<td>6.46</td>
<td>10.34</td>
<td>2.034</td>
</tr>
<tr>
<td>Guided filter [42]</td>
<td>8.40</td>
<td>15.06</td>
<td>6.80</td>
<td>11.82</td>
<td>25</td>
</tr>
<tr>
<td>Adapt. sup. [3]</td>
<td>9.54</td>
<td>16.94</td>
<td>7.42</td>
<td>13.10</td>
<td>1754</td>
</tr>
<tr>
<td>Constant time AW [41]</td>
<td>11.32</td>
<td>18.40</td>
<td>7.21</td>
<td>15.04</td>
<td>19.86</td>
</tr>
</tbody>
</table>

The percentages of erroneous pixels are given in Table 3 for two different error criteria. According to these results, the proposed approach has the best accuracy which is followed by guided filter [42] for one disparity level difference; while for two level differences; geodesic filter [12] has the second best performance. The proposed approach yields ~10% improvement in precision for one disparity level error, while this improvement goes up to 15%, when precision is sacrificed with additional one level difference, among all pixels. In the last column of Table 3, execution times are also tabulated for each aggregation methodology; and it is clear that proposed approach has competitive efficiency in terms of complexity with [4], which is the fastest approach. Compared to the results provided in Table 2, the advantage of low operational requirement of [4] is obvious as the number of disparity candidates increases, such that for the given resolution (720 × 576), both approaches consume almost similar times for 90 disparity candidates. In Table 2, the proposed approach is faster that [4] due to low number of disparity candidates. Besides, proposed approach requires almost 90% low computation in CPU compared to the second best accurate aggregation method, guided filter, for the given data set.

Another important observation from Table 3 is that, the fastest approximation of AW [41] decreases estimation accuracy, while requiring 10 times more processing than the proposed approach. In conclusion, permeability filter is a good alternative for edge-aware filters with its fast execution and accuracy for stereo matching.

In Fig. 10, estimated disparity maps for five scenes with different characteristics are illustrated for visual interpretation of the results given in Table 3. According to Fig. 10, all methods perform well for textured regions such as the Dolls stereo pair in the middle; however, as the regions become un-textured, Monopoly and Midd1 pairs, the methods yield erroneous estimates except for the proposed approach. This result is due to the specific window size definition of the state-of-the-art filters which limits diffusion of information from large areas. On the other hand, the proposed technique does not exploit any window size definition and aggregate cost values depending on the texture characteristics. In these types of regions, permeability weights are close to 1, which enables high transition rates increasing support regions that provide much more reliable estimates.

The proposed aggregation utilizes orthogonal information pass for fast operation, which may results in possible precision loss, especially for non-vertical and non-horizontal, round shaped regions that is worth to mention. However, permeability filtering provides compatible accuracy against state-of-the-art, based on extensive tests performed through various stereo data involving complex structures rather than horizontal and vertical edges. In this manner, this drawback is overcome to a certain extend by the utilization of census transform among a small window (5 × 5).

4.2. Rendering quality evaluation

In the same database [2], multiple-view versions of the stereo images are also provided which enables the
measurement for rendering quality. The quality of the estimated disparity maps are measured by rendering virtual image (middle view) from the left-right pair based on disparity map and taking frame differences between the original and rendered views. For this scenario, 2\textsuperscript{nd} and 4\textsuperscript{th} images are considered as the reference views with halved baseline compared to the experimental setup in the previous section. The average peak signal-to-noise ratio (PSNR) of the virtual views for all stereo pairs is calculated as 37.82 dB, which is quite high with visually pleasing rendering results (attached as a supplementary material due to lack of space). For the sake of completeness, the results for Dolls and Reindeer pairs are illustrated; in

\textbf{Fig. 10.} Top-to-bottom: color views (Aloe, Art, Dolls, Monopoly, Midd1), disparity maps via AW, O(1) AW, guided filter, geodesic support, arbitrary shaped cross filter, proposed approach correspondingly.

\textbf{Fig. 11.} rendered middle view of Dolls and Reindeer are given with PSNR of 36.79 dB and 35.22 dB, where the object boundaries are well preserved in a complex environment according to the error map in bottom-right. In Fig. 11(b), most of the error is due the repeated patterns at the background involving multiple good matching pixels, which is actually one of the fundamental problems of stereo vision.

4.3. Analysis of the proposed algorithm

According to visual interpretations among the resulting disparity maps in \textbf{Fig. 10}, the proposed method preserves
object boundaries, in which obvious intensity variations occur, and models smooth changes at the background with high precision. Moreover, the qualitative evaluations over ground truth disparity maps and rendering capabilities prove the promising performance of the algorithm. Based on the computational complexity analysis, the proposed approach can be argued as the fastest method without any special hardware implementation among Top-10 local methods of the Middlebury test bench, as of February 2013.

In order to further analyze the properties of the proposed method, extensive tests are provided involving detailed time analysis, the effect of occlusion handling and finally, the reasoning behind parameter selection. Use of various cost functions has an effect on the computational complexity independent of the aggregation step as analyzed in [45]. In Table 4, total computational times are presented for the proposed disparity estimation methodology exploiting two cost metrics involving cost calculation, aggregation, minimization and occlusion handling steps. It is clear that aggregation and SAD cost calculations have almost the same complexity. Census Transform during cost calculation is almost 5 times more complex than SAD term and has the highest share in terms of consumed time. Hence, accuracy improvements provided by Census transform increase complexity to a certain extend. In general, aggregation is the most time consuming step for local stereo algorithms; however, such a rule is not valid for the proposed scheme due to high efficiency of aggregation.

The effect of proposed occlusion handling step is illustrated in Fig. 12, where the disparity estimates are presented with and without occlusion handling. It is clear that occlusion handling step fixes the errors at depth discontinuities where foreground objects occludes background objects and provide crisper maps preserving discontinuities. Apart from visual interpretation, the results are also compared against the Middlebury stereo benchmark in Table 5, and the increase in the accuracy is obvious, when the left-right consistency is enforced through occlusion handling. The effect of depth favoring is also observable in Table 5, where the accuracy decreases quickly as long as depth favoring is not exploited. Hence, it is clear that depth based weighting is critical for the occlusion handling.

The proposed occlusion handling method is further compared with the approach utilized in Guided Stereo [5] which has the best performance after permeability filter. For this purpose, the initial disparity estimation is conducted by the proposed aggregation over the extended stereo database; then the occlusion handling methods are applied on the same initial estimates.

The precision of these approaches are evaluated for two cases; in the first scenario, all pixels are considered during error calculation. In the second scenario, frame boundaries are excluded, since there is no available information for the completion of these large regions (around 50–100 pixel width depending on disparity range). Therefore, the error rates are recalculated by ignoring the left-most frame boundary which is not visible in the other view to observe the occlusion performance within the image. According to the results given in Table 6 for both cases, the proposed occlusion handling approach enables more reliable (20–30%) completion of missing regions compared to the method introduced in [5]. It is expected to observe the decrease of erroneous pixel percentage when the frame boundaries are discarded. Compared to the traditional

![Fig. 11. First row: stereo left image of a scene from (a) Dolls, (b) Reindeer stereo pair and their estimated disparity maps; second row: the virtual view between left-right pair, error map larger than 10 RGB average value.](image-url)
The background copy approach, utilized in the previous section, the proposed depth favoring edge-adaptive occlusion handling method has superior performance with approximately 25% improvement in accuracy.

For the sake of completeness, effects of critical parameters to the performance of the algorithm are also examined. In that manner, smoothing parameter, \( \sigma \), and weighting factor, \( \alpha \), for SAD-Census Transform unification are tested while the truncation value, \( T \), in (2) is fixed at 15, \( \phi \) in (9) is set to 0.1, and the census transform window is chosen as \( 5 \times 5 \). Comparison is provided on the extended stereo pairs in order to provide much reliable measure. In Fig. 13, the average erroneous pixel percentage on the estimated disparity maps with respect to the ground truth disparity maps is illustrated based on the variation of smoothness factor. In addition, the effect of smoothness parameter on the disparity maps can be observed for Art pair, correspondingly. It is clear that, for high \( \sigma \) values, smoother disparity maps are obtained, while losing details by some amount. On the other hand, as \( \sigma \) parameter is decreased to 2, details are preserved with higher susceptibility to noise, resulting in salt-and-pepper type artifacts in the disparity maps, which decrease the quality. According to the bad pixel percentage plot and visual interpretation, setting \( \sigma \) in the range of [4–16] provides good matching quality, which also covers the selected \( \sigma \) throughout this study.

The effect of SAD-Census Transform weighting factor is illustrated in Fig. 14, where the smoothing parameter is set to its optimum value, 8. In all error types, the best accuracy is obtained when \( \alpha \) is set to 0.2, i.e. 80% of the weight is provided for Census measure. It is interesting to note that, as the weight of SAD measure decreases below 0.2, the accuracy of the estimate is also decreased. This result is compatible with the in depth analysis provided by [45] stating that combination of SAD and Census transform yields higher accuracy than utilization of only SAD or Census measure.

### 5. Conclusions

In this work, a novel information permeability filtering approach is presented for local stereo matching problem. The proposed filtering provides constant operational time for color adaptive weighted aggregation of cost values, unifying fast operation and high accuracy. Successive weighted summation is applied in horizontal and vertical directions, enabling the advantages of separate filtering for achieving 2D support regions. Moreover, the same idea is exploited to handle occlusions for which reliable background disparity is also propagated towards these regions. According to the experimental results, the proposed method has the 2\(^{nd}\) rank in terms of accuracy among pure local based stereo matching algorithms in Middlebury online test bench, while outperforms the whole state-of-the-art (by 5\%) for the experiments on an extended stereo dataset. Moreover, based on a computational analysis, the fastest execution time is observed among the top 10 local

### Table 5

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Rank</th>
<th>Avg. error [%]</th>
<th>Tsukuba</th>
<th>Venus</th>
<th>Teddy</th>
<th>Cones</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAD+Census</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>30</td>
<td>5.50</td>
<td>1.54</td>
<td>0.88</td>
<td>13.1</td>
<td>9.16</td>
</tr>
<tr>
<td>No depth Favor</td>
<td>68</td>
<td>6.48</td>
<td>1.81</td>
<td>1.48</td>
<td>14.9</td>
<td>10.3</td>
</tr>
<tr>
<td>No Occ.</td>
<td>81</td>
<td>7.32</td>
<td>2.59</td>
<td>1.88</td>
<td>17.00</td>
<td>10.8</td>
</tr>
<tr>
<td>SAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>60</td>
<td>6.33</td>
<td>1.53</td>
<td>1.56</td>
<td>13.2</td>
<td>10.5</td>
</tr>
<tr>
<td>No depth Favor</td>
<td>105</td>
<td>9.14</td>
<td>2.58</td>
<td>3.67</td>
<td>14.9</td>
<td>13.4</td>
</tr>
<tr>
<td>No Occ.</td>
<td>114</td>
<td>10.1</td>
<td>3.43</td>
<td>5.13</td>
<td>16.2</td>
<td>15.3</td>
</tr>
</tbody>
</table>

### Table 6

<table>
<thead>
<tr>
<th>Avg. error [%]</th>
<th>All pixels</th>
<th>Exclude frame boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>11.2</td>
<td>7.6</td>
</tr>
<tr>
<td>Guided filter [5–42]</td>
<td>12.9</td>
<td>9.6</td>
</tr>
<tr>
<td>Background copy</td>
<td>14.2</td>
<td>10.6</td>
</tr>
</tbody>
</table>

Fig. 12. First row: disparity maps without occlusion handling, second row: disparity maps after occlusion handling.
methods, in CPU. It is important to note that the proposed method utilizes only filtering of some cost functions with no additional local-global optimizations. Therefore, its performance can be further increased by additional optimization steps after the aggregation stage, which remains as a future work of this study. Besides, the parallelization availability of the proposed algorithm encourages to further speed-up disparity estimation process to real-time by specialized GPU platforms. Finally, the proposed filtering idea can be extended to various image processing applications, related with edge-aware image filtering, such as depth up-scaling, abstraction and segmentation.

Fig. 13. Top: average bad pixel percentage plot for all Middlebury stereo pairs based on smoothness factor, bottom: disparity maps for the Art sequence at specific $\sigma$ values.

Fig. 14. The effect of weighting factor between SAD and census measures on the estimation accuracy.
Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.image.2013.04.001.

References


