Image and Vision Computing 35 (2015) 31–49

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Enhanced disparity estimation in stereo images ☆

Georgios A. Kordelas [a,b,*], Dimitrios S. Alexiadis [a], Petros Daras [a], Ebroul Izquierdo [b]

[a] Information Technologies Institute, Centre for Research and Technology-Hellas, 6th km CharilaouThermi, GR-57001 Thessaloniki, Greece
[b] Electronic Engineering and Computer Science Department, Queen Mary University of London, Mile End Road, E1 4NS London, UK

ARTICLE INFO

Article history:
Received 27 April 2014
Received in revised form 26 November 2014
Accepted 12 December 2014
Available online 6 January 2015

Keywords:
Stereo vision
Stereo matching
Disparity estimation
Scanline optimization
Outliers handling

ABSTRACT

This paper presents a novel stereo disparity estimation method, which combines three different cost metrics, defined using RGB information, the CENSUS transform, as well as Scale-Invariant Feature Transform coefficients. The selected cost metrics are aggregated based on an adaptive weight approach, in order to calculate their corresponding cost volumes. The resulting cost volumes are then merged into a combined one, following a novel two-phase strategy, which is further refined by exploiting scanline optimization. A mean-shift segmentation-driven approach is exploited to deal with outliers in the disparity maps. Additionally, low-textured areas are handled using disparity histogram analysis, which allows for reliable disparity plane fitting on these areas. Finally, an efficient two-step approach is introduced to refine disparity discontinuities. Experiments performed on the four images of the Middlebury benchmark demonstrate the accuracy of this methodology, which currently ranks first among published methods. Moreover, this algorithm is tested on 27 additional Middlebury stereo pairs for evaluating thoroughly its performance. The extended comparison verifies the efficiency of this work.

1. Introduction

Stereo reconstruction is one of the most active research fields in computer vision [1]. Though mature, the task of estimating dense disparity maps from stereo image pairs is still challenging, since there is still space for improving accuracy and providing new ways of handling uniform areas, depth discontinuities and occlusions. Several approaches have been proposed so far, targeting the improvement of the reconstruction accuracy and/or minimization of the computational cost. Section 1.1 reports on the approaches in this field. Paper's contribution is described in Section 1.2. While, Section 1.3 compares the proposed work to state-of-the-art methods.

1.1. Review of previous work

The work in [1] presents a complete taxonomy of approaches used for stereo disparity estimation. The categorization of the approaches is based on the following four generic steps, into which most of the stereo algorithms can be decomposed: 1. matching cost computation; 2. cost (support) aggregation; 3. disparity computation/optimization; and 4. disparity refinement. Several metrics have been proposed in the literature for the computation of matching costs between pixels. Prevalent pixel-based cost measures include the absolute difference of image intensity values, gradient-based measures and non-parametric transforms such as spearman, CENSUS and rank [2]. The CENSUS transform has been successfully used for disparity estimation and several modifications of it have been presented [3–5]. Many approaches combine various cost measures in order to boost accuracy. The work in [6] is based on the self-adapting dissimilarity measure, which combines the sum of absolute intensity differences and a gradient based measure. The works in [7–9] exploit a combination of absolute intensity differences, as well as the hamming distance of CENSUS transform coefficients. The matching cost values over all pixels and all disparities form the initial disparity space image (DSI) or the initial cost volume.

In order to reduce matching ambiguity, the pixel-based matching costs are aggregated spatially over support regions in the DSI. The performance evaluations on different cost aggregation approaches [10,11] show that adaptive-weight [12] and segment-support [13] outperform the rest of cost aggregation approaches. More recent cost aggregation methods include successive weighted summation [8] and guided image filter [14,15].

The disparity optimization step includes local, global, cooperative and semi-global methods. Local methods [8,12–14,16,17] put emphasis on matching cost computation and cost aggregation. The final disparity map is computed by applying a simple local winner-take-all (WTA) approach independently for each pixel. Global optimization methods aim at assigning a disparity label to each pixel, so that a global cost function is minimized over the whole image area. Efficient techniques include Graph Cuts [18], Belief Propagation [6] and cooperative optimization [19]. In an additional category of approaches, the energy function is minimized on a subset of points of the stereo pair (semi-global methods), for instance along 1D paths. Such approaches, which decrease the computational complexity compared to global optimization

☆ This paper has been recommended for acceptance by Philippos Mordohai.

* Corresponding author at: Information Technologies Institute, Centre for Research and Technology-Hellas, 6th km CharilaouThermi, GR-57001 Thessaloniki, Greece.
E-mail addresses: kordelas@iti.gr (G.A. Kordelas), dalexiad@iti.gr (D.S. Alexiadis), daras@iti.gr (P. Daras), ebroul.izquierdo@eecs.qmul.ac.uk (E. Izquierdo).

http://dx.doi.org/10.1016/j.imavis.2014.12.001
0262-8856/

algorithms, involve Dynamic Programming [20] or Scanline Optimization [21] techniques.

The disparity results have to be refined, since they are "polluted" with outliers in occluded areas, depth discontinuities and uniform areas that lack texture. Several stereo algorithms, such as those in [21,22], use segmented regions for reliable outlier handling. The work in [9] uses iterative region voting and proper interpolation to fill outliers.

1.2. Contributions of this paper

In this paper, a methodology for accurate dense disparity estimation is proposed. Most significant contributions of this work include the following:

- The algorithm acquires a combined cost volume by exploiting three types of cost metrics. The first cost metric combines RGB-CENSUS information, the second one uses only CENSUS information and the third one SIFT (Scale Invariant Feature Transform) information. The cost metrics are aggregated using adaptive weights and their cost volumes are acquired. A reliable two-phase strategy is then followed to merge the individual cost volumes into a combined one.
  This approach, to the extent of our knowledge, is the first one that combines efficiently RGB, CENSUS and SIFT information.
- This method exploits mean-shift image segmentation in several stages of this approach. In our approach plane fitting is applied just to segments that correspond to large uniform areas and not to all segments. This fact reduces the dependency of our method from the result of the disparity plane fitting, which may be of reduced accuracy for small segment areas, due to the decreased number of contained disparities. Also, a metric is used to verify if planar fitting is successful, since not all large uniform areas can be considered as planar.
  Segmentation is also useful in the disparity optimization step. In more detail, the mean-shift segmentation maps of the stereo pair are used to introduce a new criterion for the definition of the smoothness penalty terms that are used in the semi-global scanline optimization method of [21] (previously exploited, among other works, in [9,23–25]). The modified scanline method is employed for the optimization of the combined cost volume.
  Moreover, segmentation is exploited for the occlusion handling task, where an efficient strategy that incorporates mean-shift segmentation-based occlusion handling to successfully cope with occluded areas is presented.
- Handling of large uniform areas is based on disparity histogram analysis, which removes outlier disparities from large uniform regions, before applying disparity plane fitting in each region using the remaining reliable disparities.

Except for the major contributions, some secondary contributions are the following:

- A weighted variant of the original CENSUS transform, which improves the disparity accuracy, is proposed.
- Disparity refinement at disparity discontinuities is performed by applying a two-step disparity edges refinement approach. The first step handles disparity errors at depth discontinuities in a coarser level and the second one in a finer level.

This approach, by encompassing the aforementioned contributions, manages to rank 1st among already published methods in the Middlebury Stereo Evaluation benchmark [26] and gives superior results on an additional dataset of 27 stereo pairs.

1.3. Proposed methodology and state-of-the-art methods

This method is the first one that combines RGB, CENSUS and SIFT information by utilizing an efficient strategy. There are several works that use RGB and/or CENSUS information, such as [3,4,7–9,25], but they do not exploit the SIFT information, which could probably improve their performance. However, the approaches that use SIFT descriptors, or similar ones (such as SURF-Speeded Up Robust Features [27]), for the case of short-baseline stereo disparity estimation, are limited. For instance, the work in [28] combines mutual information, SIFT descriptor, and segment based plane-fitting to robustly find correspondences for stereo image pairs which undergo radiometric variations. The paper in [29] uses SURF key points for the initial disparity estimation, which is further improved by using graph cuts for disparity plane assignment.

Many methods, such as [6,19,29–31], exploit image segmentation algorithms in order to separate images into segments and then solve the disparity estimation problem by assigning, in various ways, a disparity plane for each estimated segment of the scene. In contrast to this class of approaches, the proposed method applies plane fitting only to large segments that correspond to low-textured areas. Additionally, in order to prevent application of plane fitting to low-textured areas that are not (near) planar, a metric is used to verify if plane fitting is successful.

The disparity histogram analysis, described in this paper, could be used as preprocessing step in algorithms that perform plane fitting using methods that are sensitive to outliers, such as the least square error (LSE) based plane fitting algorithm, which is used in [30,31]. Even plane fitting algorithms that are insensitive to outliers, such as RANSAC (Random Sample Consensus) [32], could be fostered by our outlier filtering technique, since their computational cost would be reduced in case the data to be fitted contains less outliers. Disparity estimation methods that exploit RANSAC plane fitting include [22,29,31].

Many methods, such as [6,9,19], are evaluated using just the four well-known stereo pairs of the Middlebury Stereo Online Evaluation Benchmark and they manage to rank among the top methods. However, there are additional Middlebury stereo pairs that can be used to present a more thorough and complete evaluation. The presented approach, except for the well-known stereo pairs, uses 27 more stereo pairs for assessing the overall performance of this approach.

The rest of this paper is organized as follows. In Section 2, the proposed method is presented in detail. Section 3 provides information on the parameters used, as well as the experimental results, while conclusions are drawn in Section 4.

2. Proposed method

The proposed algorithm is divided into four steps, as visualized in the flowchart of Fig. 1. The values of the parameters defined throughout this section are analyzed in Section 3.

2.1. Preprocessing steps

2.1.1. Rectified stereo pairs

The input stereo image pair is rectified, so that the epipolar lines become horizontal [33]. Therefore, the search of point-correspondences between the two images can be performed along the same horizontal epipolar line. Except for limiting searching area, rectified input makes the application of optimization algorithms simpler, such as the scanline optimization used in this work that uses specific path directions. Additionally, since the

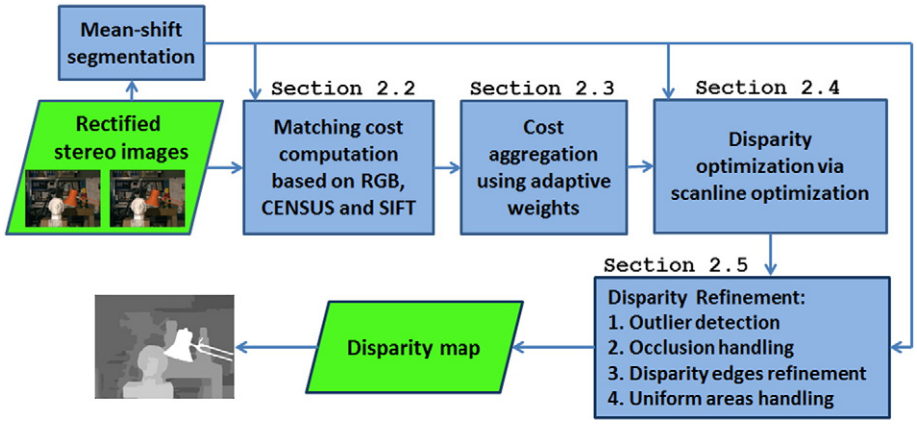


Fig. 1. Flowchart of the proposed method.

resulting rectified images have similar scale and the epipolar lines have the same orientation, it is feasible to define and compare adaptive support areas of the same size and orientation between two rectified images. The rectification process is beyond the scope of this paper. Any efficient existing algorithm, such as the one in [33], can be used for this task.

2.1.2. Mean-shift segmentation

Low-level image segmentation, which groups pixels into homogeneous regions based on color or texture, is exploited by numerous disparity estimation methods ([6,13,19,34,35]) in various ways, since it assists to acquire disparity results of high accuracy.

In this work, the stereo images are initially segmented into non-overlapping regions using a state of the art mean-shift segmentation software (EDISON software [36]), which relies on color and edge information. Detailed information about the mean-shift segmentation and the EDISON software can be found in [37–39]. The parameters used for the mean-shift segmentation are the segmentation spatial radius h_s, which is set to $h_s = 3$ and the segmentation feature space radius h_r, which is set to $h_r = 3$. The selection of these strict values ensures that the segmentation map will be of high reliability, meaning that most likely a segment will not overlap a depth discontinuity, and this fact is verified also in [13,35]. The mean-shift segmentation map for the "Tsukuba" left image (see Fig. 2a) is visualized in Fig. 2b. The pixels that belong to the same mean-shift segment have an individual label and their mean color value is computed. Let the labels image be denoted as Lab($\mathbf{x}$). The segmentation maps of the left and the right image are computed once and then used in the following algorithmic steps.

2.2. Matching cost computation

This stage considers cost metrics for estimating the similarity between two pixels. In this work, absolute differences (AD) of RGB values, AD of weighted CENSUS coefficients and AD of SIFT coefficients are exploited to define the used cost metrics. This choice is made for the following reasons: i) Exploitation of RGB information gives better results in areas where depth discontinuities exist; ii) CENSUS is able to cope with radiometric changes and noise [9], while iii) the exploitation of SIFT improves the results in textured unconcluded areas, as verified in Section 2.3.2.

2.2.1. Weighted CENSUS transform

In order to define the original CENSUS transform [3], a function ξ, which represents the relationship between the intensity of a pixel $\mathbf{x} = (x, y)^T$ and a neighbor pixel $\mathbf{x}_n$, is used:

$$\xi(\mathbf{x}, \mathbf{x}_n) = \begin{cases} 1, \text{if } I(\mathbf{x}_n) < I(\mathbf{x}) \\ 0, \quad \text{otherwise}, \end{cases} \tag{1}$$

where $I(\mathbf{x})$ represents the image intensity of pixel $\mathbf{x}$.

Fig. 2. (a) Left "Tsukuba" image and (b) its mean-shift segmentation map.

The CENSUS transform for pixel $\mathbf{x}$ is computed by comparing its intensity with the intensity of other pixels $\mathbf{x}_n$ that lie within a square window $\mathcal{N}(\mathbf{x})$ around $\mathbf{x}$. The results of these comparisons are then concatenated into a single CENSUS binary vector. Thus, the CENSUS transform of a pixel $\mathbf{x}$ is defined as:

$$\mathrm{CENSUS}(\mathbf{x}) = \underset{\mathbf{x}_n \in \mathcal{N}(\mathbf{x})}{\otimes} \xi(\mathbf{x}, \mathbf{x}_n), \tag{2}$$

where $\otimes$ represents the concatenation operation.

This paper proposes a modification of the original CENSUS transform defined as a weighted CENSUS transform. In the proposed transform (see Fig. 3) the bit string that is generated from the original CENSUS transform for a central pixel $\mathbf{x}$, is multiplied by a weight vector, whose elements correspond to the weights between $\mathbf{x}$ and each pixel $\mathbf{x}_n \in \mathcal{N}(\mathbf{x})$. The weight between the central pixel $\mathbf{x}$ (red circle in Fig. 3) and a pixel $\mathbf{x}_n$ (green circle in Fig. 3) is defined as:

$$\mu(\mathbf{x}; \mathbf{x}_n) = 1 - \beta \cdot \Delta_e(\mathbf{x}; \mathbf{x}_n), \tag{3}$$

where $\Delta_e(\mathbf{x}; \mathbf{x}_n)$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_n$ and β is a constant parameter. The window size of the weighed CENSUS transform is set experimentally to 5×5. The weight vector gives greater weights for pixels closer to the central pixel, since they are considered as more reliable than those which lie further. Let us denote as $\mathbf{CEN}(\mathbf{x}; c)$ the weighted CENSUS transform at pixel $\mathbf{x}$ for the color band $c \in \{R, G, B\}$.

2.2.2. SIFT-based cost

The SIFT coefficients are extracted densely from an image using the SIFT implementation that was used in the work of [40], which deals with visual concept classification. In detail, the parameters used for the SIFT coefficients extraction were selected as: Size of subregions $N_P = 1$, Scale of Gaussian Derivatives $\sigma_{DOG} = 1$, and number of subregions $N_S = 2$. These parameters define a SIFT descriptor composed of $N_S \times N_S$ subregions with subregions' size equal to $N_P \times N_P$ pixels. The horizontal and vertical responses for SIFT are calculated using a Gaussian derivative filter, while the diagonal responses are calculated using a fast anisotropic Gaussian derivative filter [41], both using a scale of σ_{DOG}. When, a larger support area was used for the extraction of the SIFT descriptor vector (by increasing N_P and/or N_S), the foreground fattening effect [1] was becoming more intense in the estimated disparity map. Let us denote $\mathbf{SIFT}(\mathbf{x}; c)$ as the SIFT descriptor at pixel $\mathbf{x}$ for the color band $c \in \{R, G, B\}$.

2.2.3. Cost metrics

In this section, the way that RGB, weighed CENSUS and SIFT are used to define the similarity between pixels, is described. Given a pixel $\mathbf{x}$ on the left image (reference image) $I_l(\mathbf{x})$, the corresponding pixel on the right image (target image) I_r for a candidate disparity d will be $I_r(\mathbf{x^d})$, where $\mathbf{x^d} = \mathbf{x} - \mathbf{d}$ and $\mathbf{d} = (d, 0)^{\mathbf{T}}$, since the input stereo images are rectified and consequently the disparity has only a horizontal component. The individual pixel similarity measures $C_{RGB}(\mathbf{x}; \mathbf{d})$, $C_{CENSUS}(\mathbf{x}; \mathbf{d})$ and $C_{SIFT}(\mathbf{x}; \mathbf{d})$, between $\mathbf{x}$ and $\mathbf{x^d}$, are given from:

$$C_{RGB}(\mathbf{x}; \mathbf{d}) = \sum_{c \in \{R,G,B\}} \left| I_l(\mathbf{x}; c) - I_r\left(\mathbf{x^d}; c\right) \right|, \tag{4}$$

$$C_{CENSUS}(\mathbf{x}; \mathbf{d}) = \sum_{c \in \{R,G,B\}} \left\| \mathbf{CEN}_l(\mathbf{x}; c) - \mathbf{CEN}_r\left(\mathbf{x^d}; c\right) \right\|_1, \tag{5}$$

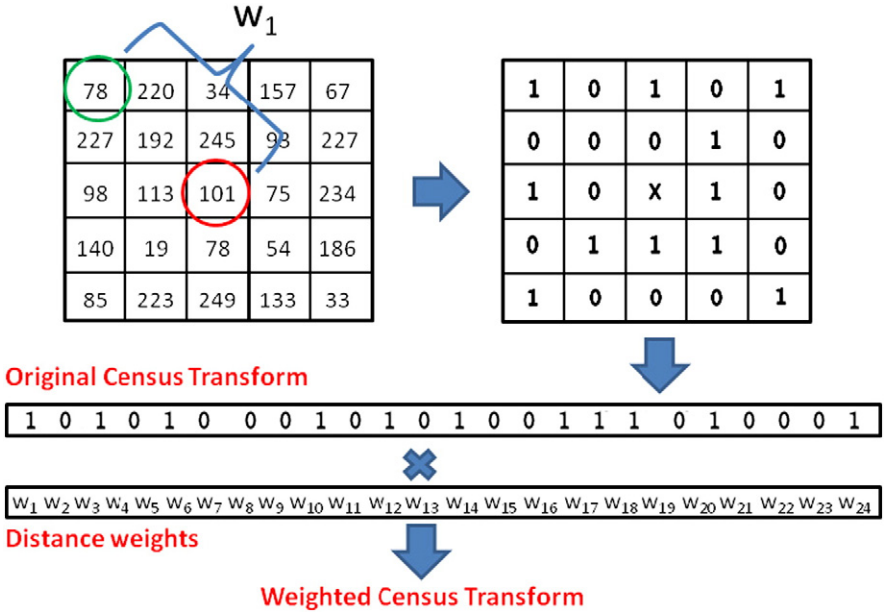


Fig. 3. Weighted CENSUS transform.

$$C_{\mathrm{SIFT}}(\mathbf{x};\mathbf{d}) = \sum_{c\in\{R,G,B\}} \left\| \mathbf{SIFT}_l(\mathbf{x};c) - \mathbf{SIFT}_r\left(\mathbf{x}^{\mathbf{d}};c\right) \right\|_1. \tag{6}$$

Using the aforementioned measures three different matching costs are defined. A RGB-CENSUS combination cost $C_{R-C}(\mathbf{x};\mathbf{d})$ (following the paradigm of algorithms [7–9]), a pure weighted CENSUS-based cost $C_{CEN}(\mathbf{x};\mathbf{d})$ and a SIFT-based cost $C_S(\mathbf{x};\mathbf{d})$, which are given from:

$$C_{R-C}(\mathbf{x};\mathbf{d}) = \rho(C_{RGB}(\mathbf{x};\mathbf{d}),\lambda_{RGB}) + \rho(C_{CENSUS}(\mathbf{x};\mathbf{d}),\lambda_{CEN}), \tag{7}$$

$$C_{CEN}(\mathbf{x};\mathbf{d}) = \rho(C_{CENSUS}(\mathbf{x};\mathbf{d}),\lambda_{CEN}), \tag{8}$$

$$C_S(\mathbf{x};\mathbf{d}) = \rho(C_{SIFT}(\mathbf{x};\mathbf{d}),\lambda_{SIFT}), \tag{9}$$

where $\rho(C_y,\lambda_y) = 1 - e^{(-Cy/\lambda y)}$.

The exponential function $\rho(C_y,\lambda_y)$ has the advantage of mapping the values of a measure in the range of [0, 1]. This allows different types of measures with different ranges to be scaled into the same range and then to be combined. Additionally, this function allows trimming of outlier values of C_y, depending on the value of λ_y.

2.3. Cost aggregation

2.3.1. Adaptive cost aggregation

In order to reduce matching ambiguity, the pixel-based matching costs $C_{R-C}(\mathbf{x};\mathbf{d})$, $C_{CEN}(\mathbf{x};\mathbf{d})$ and $C_S(\mathbf{x};\mathbf{d})$ are aggregated spatially over support regions around each pixel. According to the evaluation studies of [10,11], the adaptive weight approach [12] produces reasonably accurate disparity maps. Thus, this aggregation approach with slight modifications is used in this work.

More specifically, adaptive support-weight based aggregation applies weights to each of the pixels surrounding the pixel of interest. The adaptive-support weights notion is based on the Gestalt principles of similarity and proximity [12]. The similarity principle assumes that the more similar color a surrounding pixel has to the central pixel of interest, the more likely it is to belong to the same surface, while the proximity principle assumes that the closer a surrounding pixel is to the central pixel of interest, the more likely it is to belong to the same surface.

In order to describe the adaptive-support weight notion with mathematic expressions, a pixel of interest $\mathbf{x}$ and a neighbor pixel $\mathbf{x}_n$ are considered. The adaptive weight between $\mathbf{x}$ and $\mathbf{x}_n$, is given by:

$$w(\mathbf{x};\mathbf{x}_n) = e^{\left(\frac{-\Delta I(\mathbf{x},\mathbf{x}_n)}{\gamma_c}\right)} \cdot e^{\left(\frac{-\Delta e(\mathbf{x},\mathbf{x}_n)}{\gamma_e}\right)}, \tag{10}$$

where γ_e and γ_c are constant parameters, $\Delta e(\mathbf{x};\mathbf{x}_n)$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_n$ and $\Delta I(\mathbf{x};\mathbf{x}_n)$ is given by:

$$\Delta I(\mathbf{x};\mathbf{x}_n) = \sqrt{\sum_{c\in\{R,G,B\}} |I(\mathbf{x};c) - I(\mathbf{x}_n;c)|^2}. \tag{11}$$

Similar to [16], the adaptive weights are computed on the input stereo images after applying a median filter that uses a 2×2 neighborhood in order to alleviate the impact of image noise and subtle non-Lambertian effects.

The adaptive weight approach used in this work has two slight modifications compared to the original work of [12]. Experimental results in [13] proved that the use of the RGB color space for computing color similarity decreases the possibility that pixels belonging to different depths are being aggregated in the same support region. For this reason the RGB color similarity is used, contrary to the CIELAB similarity used in [12]. Additionally, instead of using all pixels in the square support region, only pixels within radius R_S from the central pixel are used. In this way, the support region becomes symmetric around the central pixel $\mathbf{x}$ of interest.

A weight support mask is generated for a pixel $\mathbf{x}$ on the left stereo image, denoted as $w_l(\mathbf{x};\mathbf{x}_n)$. Similarly, a weight support mask is generated for the right stereo image around the corresponding pixel $\mathbf{x}^{\mathbf{d}}$ and is denoted as $w_r(\mathbf{x}^{\mathbf{d}};\mathbf{x}_n^{\mathbf{d}})$. Both $w_l(\mathbf{x};\mathbf{x}_n)$ and $w_r(\mathbf{x}^{\mathbf{d}};\mathbf{x}_n^{\mathbf{d}})$ are taken into consideration to define the aggregated cost $V(\mathbf{x};\mathbf{d})$ between $\mathbf{x}$ and $\mathbf{x}^{\mathbf{d}}$ as:

$$V(\mathbf{x};\mathbf{d}) = \frac{\sum_{\mathbf{x}_n\in S_L,\mathbf{x}_n^{\mathbf{d}}\in S_R} w_l(\mathbf{x};\mathbf{x}_n)\cdot w_r\left(\mathbf{x}^{\mathbf{d}};\mathbf{x}_n^{\mathbf{d}}\right)\cdot C(\mathbf{x}_n;\mathbf{d})}{\sum_{x_n\in S_L,x_n^d\in S_R} w_l(\mathbf{x};\mathbf{x}_n)\cdot w_r\left(\mathbf{x}^{\mathbf{d}};\mathbf{x}_n^{\mathbf{d}}\right)}, \tag{12}$$

where S_L defines the support region around pixel $\mathbf{x}$ on the left image and S_R is the support region around pixel $\mathbf{x}^{\mathbf{d}}$, on the right image, as it is visualized in Fig. 4. If cost $C(\mathbf{x};\mathbf{d})$ is replaced by $C_{R-C}(\mathbf{x};\mathbf{d})$, $C_{CEN}(\mathbf{x};\mathbf{d})$ or $C_S(\mathbf{x};\mathbf{d})$, the aggregated cost volumes $V_{R-C}(\mathbf{x};\mathbf{d})$, $V_{CEN}(\mathbf{x};\mathbf{d})$ and $V_{SIFT}(\mathbf{x};\mathbf{d})$ can be estimated, respectively. The schematic representation of a cost volume is depicted in Fig. 5a.

2.3.2. Combination of aggregated cost volumes

In the beginning of this section, we explain the reasons for using the specified cost metrics $C_{R-C}(\mathbf{x};\mathbf{d})$, $C_{CEN}(\mathbf{x};\mathbf{d})$ and $C_S(\mathbf{x};\mathbf{d})$. In this paragraph, we describe the details of combining their corresponding cost volumes $V_{R-C}(\mathbf{x};\mathbf{d})$, $V_{CEN}(\mathbf{x};\mathbf{d})$ and $V_{SIFT}(\mathbf{x};\mathbf{d})$ to produce a combined cost volume.

The proposed approach uses a combination of RGB and CENSUS information via Eq. (7) in order to compute V_{R-C}. However, after extensive experiments, it was deduced that the cost volume V_{CEN} computed using only weighted CENSUS information, could be efficiently exploited to

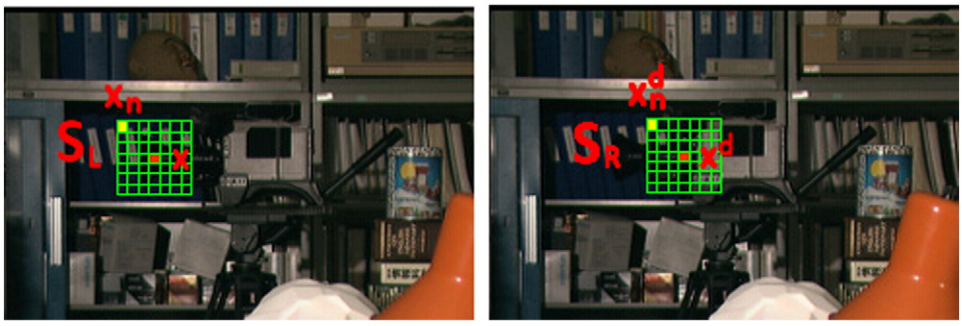


Fig. 4. Adaptive weights support region on reference and target "Tsukuba" images.

refine V_{R-C}. Additionally, it was noticed that the winner-take-all-estimated disparity map from V_{SIFT} is reliable for unoccluded textured areas. This fact is exploited here to further refine V_{R-C}. The reason for not combining directly the SIFT information with the RGB and CENSUS information (for instance using an equation similar to Eq. (7), with an additional term for SIFT information), is that the ability of the SIFT-based metric to provide accurate disparity estimates at unoccluded textured areas degrades significantly when SIFT is combined directly with other cost metrics, as experimentally verified. In the following, an efficient two-phase strategy for combining V_{R-C}, V_{CEN} and V_{SIFT} is described. This strategy is built upon the aforementioned conclusions regarding V_{CEN} and V_{SIFT}.

2.3.2.1. First combination phase. During the first phase, $V_{CEN}(\mathbf{x};\mathbf{d})$ is used to refine $V_{R-C}(\mathbf{x};\mathbf{d})$. The Peak Ratio confidence measure, one of the best confidence measures according to [42], is used for this purpose.

2.3.2.1.1. Peak ratio confidence measure. Let us consider that we have the curve of cost variation along disparity $\mathbf{d}$ for a pixel $\mathbf{x}$ from cost volume $V(\mathbf{x};\mathbf{d})$. This term is depicted visually with green color in the visual representation of a cost volume $V(\mathbf{x};\mathbf{d})$ in Fig. 5a and an example of cost variation curve is shown in Fig. 5c. Let us define as $G(\mathbf{x}) = \min_{\mathbf{d}}\{V(\mathbf{x};\mathbf{d})\}$ the global minimum of $V(\mathbf{x};\mathbf{d})$ and as $L(\mathbf{x})$ the second local minimum of $V(\mathbf{x};\mathbf{d})$. Then, the peak ratio confidence measure is defined as:

$$R(\mathbf{x}) = \frac{L(\mathbf{x})}{G(\mathbf{x})}. \tag{13}$$

Finally, let

$$\alpha(\mathbf{x}) = arg \min_{\mathbf{d}}\{V(\mathbf{x};\mathbf{d})\}, \tag{14}$$

be the optimum disparity value that gives the global minimum of $V(\mathbf{x};\mathbf{d})$. The higher $R(\mathbf{x})$, the more confident the global minimum of $V(\mathbf{x};\mathbf{d})$ is.

Based on this confidence measure, the optimum disparity for a pixel $\mathbf{x}$, as estimated from $V_{CEN}(\mathbf{x};\mathbf{d})$, will be propagated to $V_{R-C}(\mathbf{x};\mathbf{d})$. The curves of cost variation along disparity for $V_{R-C}(\mathbf{x};\mathbf{d})$ and $V_{CEN}(\mathbf{x};\mathbf{d})$ are depicted in Fig. 5b and Fig. 5c, respectively.

In more detail, for a pixel $\mathbf{x}$, the confidence $R_{R-C}(\mathbf{x}) = \frac{G_{R-C}(\mathbf{x})}{L_{R-C}(\mathbf{x})}$ based on $V_{R-C}(\mathbf{x};\mathbf{d})$ (Fig. 5b), is estimated. Similarly, the confidence $R_{CEN}(\mathbf{x}) = \frac{G_{CEN}(\mathbf{x})}{L_{CEN}(\mathbf{x})}$ based on $V_{CEN}(\mathbf{x};\mathbf{d})$ (Fig. 5c), is estimated.

In case that $R_{CEN}(\mathbf{x}) > R_{R-C}(\mathbf{x})$, at the disparity position $\alpha_{CEN}(\mathbf{x})$ (position of the global minimum of $V_{CEN}(\mathbf{x};\mathbf{d})$), the corresponding value of $V_{R-C}(\mathbf{x};\alpha_{CEN}(\mathbf{x}))$ is modified according to:

$$V_{R-C}(\mathbf{x};\alpha_{CEN}(\mathbf{x})) \leftarrow \min_{\mathbf{d}}\{V_{R-C}(\mathbf{x};\mathbf{d})\} - \delta \tag{15}$$

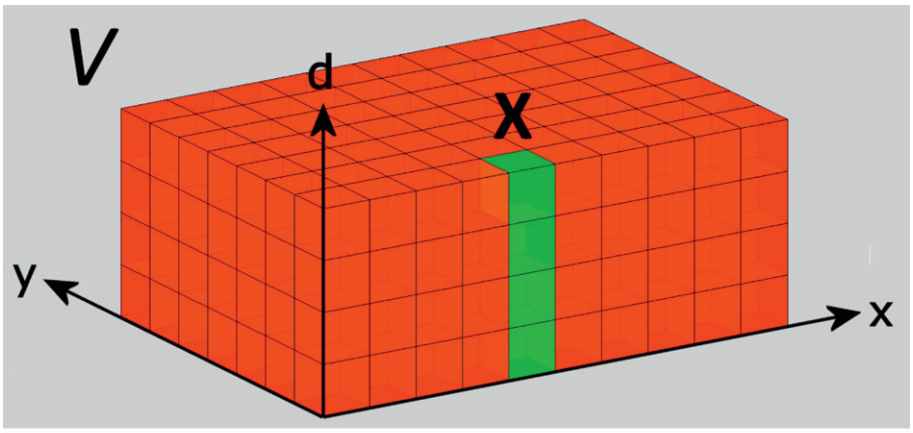
with $\delta \to 0^+$, so that the global minimum of $V_{R-C}(\mathbf{x};\mathbf{d})$ to coincide with the one of $V_{CEN}(\mathbf{x};\mathbf{d})$. The part of the curve that changes after this step is depicted with green color in Fig. 5b. The case $R_{CEN}(\mathbf{x}) > R_{R-C}(\mathbf{x})$ means that the global minimum of $V_{CEN}(\mathbf{x};\mathbf{d})$ is more confident than that of $V_{R-C}(\mathbf{x};\mathbf{d})$. In this case, information about the disparity that gives this global minimum is propagated to $V_{R-C}(\mathbf{x};\mathbf{d})$. After executing the first phase, $V'_{R-C}(\mathbf{x};\mathbf{d})$ is acquired. The winner-take-all (WTA) of $V'_{R-C}(\mathbf{x};\mathbf{d})$ gives the disparity map of Fig. 6a.

2.3.2.2. Second combination phase. In a second phase, $V_{SIFT}(\mathbf{x};\mathbf{d})$ is used to refine $V'_{R-C}(\mathbf{x};\mathbf{d})$. The WTA of $V_{SIFT}(\mathbf{x};\mathbf{d})$ gives the SIFT-based disparity map $d_{SIFT}(\mathbf{x})$ (see Fig. 6b), which provides reliable disparities in textured unoccluded areas where depth does not change. This is evident in Fig. 6b for the disparity of the left "Tsukuba" image.

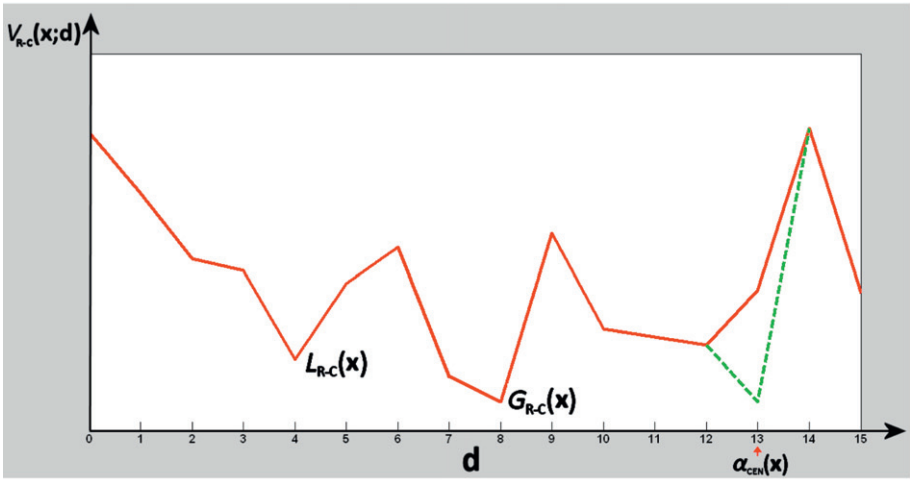
2.3.2.2.1. Detection of reliable disparities. In order to find the regions in $d_{SIFT}(\mathbf{x})$ that are reliable, the mean-shift color segmentation map (see Section 2.1.2) is used. If $n(S)$ denotes the number of pixels in a color segment S and $n_f(S)$ is the number of pixels that have the most frequent disparity in this segment according to d_{SIFT}, then $P_{\mathbf{x}}(S) = \frac{n_f(S)}{n(S)}$ is defined. If $P_{\mathbf{x}}(S) \geq 90\,\%$, then it is assumed that the disparities inside this segment are reliable (since the vast majority of pixels have the same disparity value).

According to the above, reliable disparities in $d_{SIFT}(\mathbf{x})$ are propagated to $V'_{R-C}(\mathbf{x};\mathbf{d})$ in the following way: For every pixel $\mathbf{x} \in S$, the disparity estimate $d_{SIFT}(\mathbf{x})$ is propagated to $V'_{R-C}(\mathbf{x};\mathbf{d})$ according to:

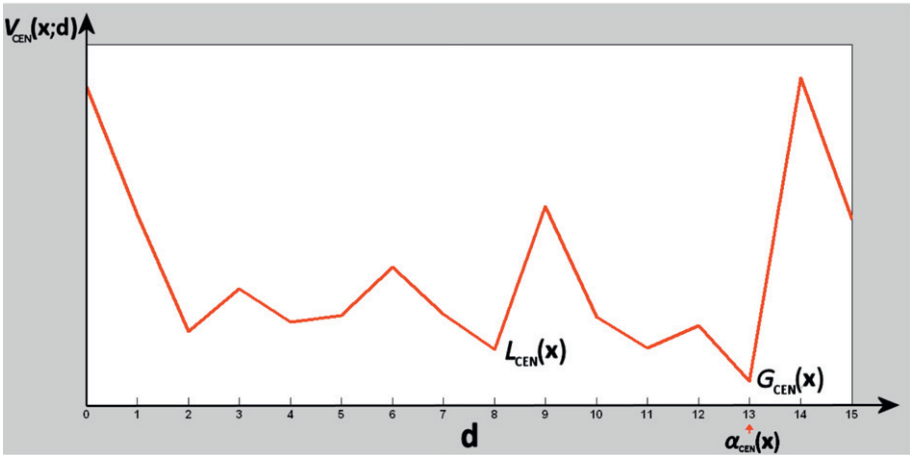
$$V'_{R-C}(\mathbf{x};d_{SIFT}(\mathbf{x})) \leftarrow \min_{\mathbf{d}}\{V'_{R-C}(\mathbf{x};\mathbf{d})\} - \delta \tag{16}$$



(a)



(b)



(c)

Fig. 5. (a) Cost volume visualization and cost variation along disparity for a pixel x of (b) V_{R-C} and (c) V_{CEN}.

with $\delta \rightarrow 0^{+}$. After executing this second phase, $V''_{R-C}(\mathbf{x};\mathbf{d})$ is acquired. Let the WTA-estimated disparity map from $V''_{R-C}(\mathbf{x};\mathbf{d})$ be $d_{LR}(\mathbf{x})$. After applying a 3×3 median filter on $d_{LR}(\mathbf{x})$, in order to remove spurious disparities, the disparity map of Fig. 7a is generated.

Comparing Figs. 6a and 7a, it is evident that $d_{SIFT}(\mathbf{x})$ can be exploited to efficiently enhance the results in unoccluded textured regions.

Except for the visual demonstration of using the two-phase combination strategy to improve the generated disparity map, an additional numeric evaluation is presented in Section 3.2.2.

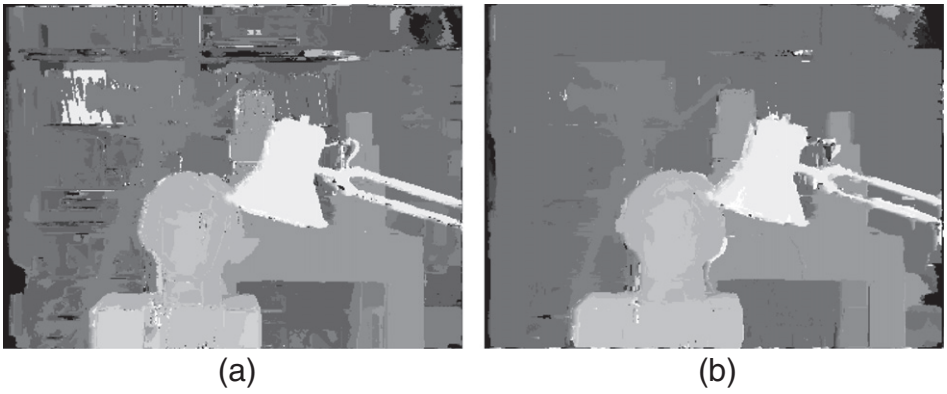


Fig. 6. Disparity maps after applying WTA to (a) V'_{R-C} and (b) V_{SIFT}.

2.4. Disparity optimization

2.4.1. Outliers detection

The disparity map $d_{LR}(\mathbf{x})$ is acquired considering as reference image the left image of the stereo pair. If the right image is considered as reference, then the disparity map $d_{RL}(\mathbf{x})$ of Fig. 7b is acquired. The disparity maps $d_{LR}(\mathbf{x})$ and $d_{RL}(\mathbf{x})$ are taken into consideration to detect problematic areas, especially outliers in occluded regions and depth discontinuities. A prevalent strategy for detecting outliers is the Left–Right consistency check [24]. In this strategy, the outliers are disparity values that are not consistent between the two maps and therefore, they do not satisfy the relation:

$$|d_{LR}(\mathbf{x}) - d_{RL}(\mathbf{x} - d_{LR}(\mathbf{x}))| \leq T_{LR}. \tag{17}$$

At this point of our algorithm, the threshold for outliers detection is set equal to $T_{LR} = 1$. With this value, pixels with difference equal to 1 in the Left–Right consistency check are not considered as outliers. This is plausible, since disparity in stereo images usually varies smoothly along the epipolar lines, in regions without depth discontinuities. Fig. 8a shows the outliers map $O_1^{T_{LR}=1}(\mathbf{x})$ for $T_{LR} = 1$. The blue regions are the outlier regions.

Let X_{OUT} be the set of outlier pixels. Before applying scanline optimization, the cost values in $V''_{R-C}(\mathbf{x}; \mathbf{d})$ that correspond to outlier pixels $\mathbf{x} \in X_{OUT}$ are set to zero. This way, the costs of these pixels do not affect the optimization result.

2.4.2. Semi-global, scanline optimization

The optimization of the combined cost volume $V''_{R-C}(\mathbf{x}; \mathbf{d})$ is based on the semi-global, scanline optimization method of [21], which aggregates matching costs in 1D equally from multiple path directions.

This work considers four path directions $\mathbf{r}$, namely left-to-right, right-to-left, up-to-down and down-to-up, which are denoted as $\mathbf{r}_{lr} = [+1, 0]^{\mathbf{T}}$, $\mathbf{r}_{rl} = [-1, 0]^{\mathbf{T}}$, $\mathbf{r}_{ud} = [0, +1]^{\mathbf{T}}$ and $\mathbf{r}_{du} = [0, -1]^{\mathbf{T}}$, respectively (see Fig. 8b).

Let $L_{\mathbf{r}}$ be a path that is traversed in the direction $\mathbf{r} \in \{\mathbf{r}_{lr}, \mathbf{r}_{rl}, \mathbf{r}_{ud}, \mathbf{r}_{du}\}$. The path cost $L_{\mathbf{r}}(\mathbf{x}; \mathbf{d})$ of pixel $\mathbf{x}$ at disparity $\mathbf{d}$ is computed recursively from:

$$L_{\mathbf{r}}(\mathbf{x}; \mathbf{d}) = V''_{R-C}(\mathbf{x}; \mathbf{d}) + min\{L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}; \mathbf{d}), L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}; \mathbf{d} \pm 1) + \pi_1, \min_i L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}; i) + \pi_2\} - \min_i L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}; i) \tag{18}$$

where $i \in [\text{disparity range}]$ and $\mathbf{x} - \mathbf{r}$ denotes the previous pixel along the path direction. π_1 and π_2 are two smoothness penalty terms (with $\pi_1 \leq \pi_2$) for penalizing disparity changes of neighboring pixels. The works in [9,24] make the assumption that often a depth discontinuity (i.e. disparity change) coincides with an intensity edge.

Fig. 7. (a) $d_{LR}(\mathbf{x})$ and (b) $d_{RL}(\mathbf{x})$ disparity maps.

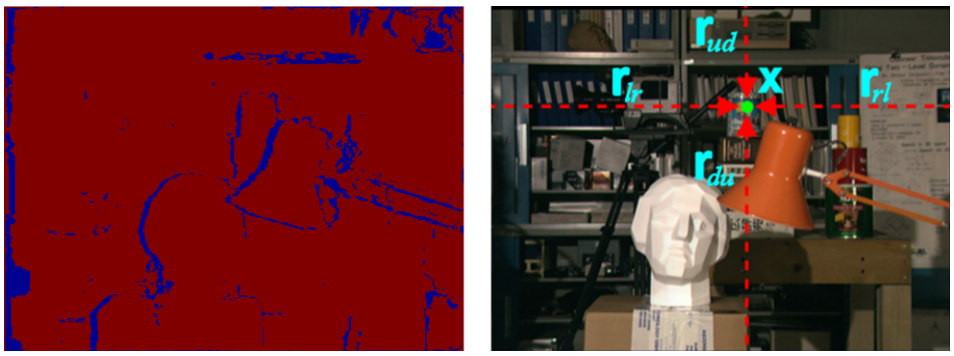


Fig. 8. (a) Outliers map $O_1^{T_{LR}=1}(\mathbf{x})$ for threshold $T_{LR} = 1$ and (b) path directions used for cost volume optimization.

In our approach, two criteria are used to check depth discontinuity and consequently compute the smoothing terms π_1 and π_2. The first criterion, as in [24], checks the intensity difference between the current pixel and the previous one along the considered path direction. The intensity difference, on the two images, is defined as:

$$\nabla(\mathbf{x}) = \max_{c\in\{R,G,B\}} |I_l(\mathbf{x};c) - I_l(\mathbf{x}-\mathbf{r};c)| \tag{19}$$

and

$$\nabla\left(\mathbf{x}^{\mathbf{d}}\right) = \max_{c\in\{R,G,B\}} \left|I_r\left(\mathbf{x}^{\mathbf{d}};c\right) - I_r\left(\mathbf{x}^{\mathbf{d}}-\mathbf{r};c\right)\right|. \tag{20}$$

The second criterion, proposed in our approach, checks whether the current pixel and the previous one along the considered path direction belong to the same mean-shift segment.

$$(\pi_1,\pi_2) = \begin{cases} (\Pi_1,\Pi_2), & \text{if } \left(\nabla(\mathbf{x}) \le \tau_{so} \;\&\; \nabla\left(\mathbf{x}^{\mathbf{d}}\right) \le \tau_{so}\right) \\ \left(\frac{\Pi_1}{1.5},\frac{\Pi_2}{1.5}\right), & \text{if } \left(\mathrm{Lab}_l(\mathbf{x}) == \mathrm{Lab}_l(\mathbf{x}-\mathbf{r}) \;\&\; \mathrm{Lab}_r\left(\mathbf{x}^{\mathbf{d}}\right) == \mathrm{Lab}_r\left(\mathbf{x}^{\mathbf{d}}-\mathbf{r}\right)\right) \\ \left(\frac{\Pi_1}{4},\frac{\Pi_2}{4}\right), & \text{if } \left(\nabla(\mathbf{x}) \le \tau_{so} \;\&\; \nabla\left(\mathbf{x}^{\mathbf{d}}\right) > \tau_{so}\right) \text{ or} \left(\mathrm{Lab}_l(\mathbf{x}) == \mathrm{Lab}_l(\mathbf{x}-\mathbf{r}) \;\&\; \mathrm{Lab}_r\left(\mathbf{x}^{\mathbf{d}}\right) \ne \mathrm{Lab}_r\left(\mathbf{x}^{\mathbf{d}}-\mathbf{r}\right)\right) \\ \left(\frac{\Pi_1}{4},\frac{\Pi_2}{4}\right), & \text{if } \left(\nabla(\mathbf{x}) > \tau_{so} \;\&\; \nabla\left(\mathbf{x}^{\mathbf{d}}\right) \le \tau_{so}\right) \text{or} \left(\mathrm{Lab}_l(\mathbf{x}) \ne \mathrm{Lab}_l(\mathbf{x}-\mathbf{r}) \;\&\; \mathrm{Lab}_r\left(\mathbf{x}^{\mathbf{d}}\right) == \mathrm{Lab}_r\left(\mathbf{x}^{\mathbf{d}}-\mathbf{r}\right)\right) \\ \left(\frac{\Pi_1}{10},\frac{\Pi_2}{10}\right), & \text{otherwise.} \end{cases} \tag{21}$$

Based on these two criteria, π_1 and π_2 are defined according to Eq. (21), where $\Pi_1 = 0.2, \Pi_2 = 0.6$ are constant parameters, $\tau_{so} = 10$ is a threshold for color difference and Lab_l and Lab_r are the labels images after applying mean-shift segmentation (see Section 2.1.2) to the left and right images, respectively. The denominator for the second condition of Eq. (21), has been slightly increased to 1.5 so as to compensate for segmentation errors.

To sum up, while the approaches in [9,24], employ only intensity based criteria to define π_1 and π_2, the proposed method introduces an additional criterion based on mean-shift segmentation that assists in enhancing the disparity estimation results, as it is experimentally verified in Section 3.2.4.

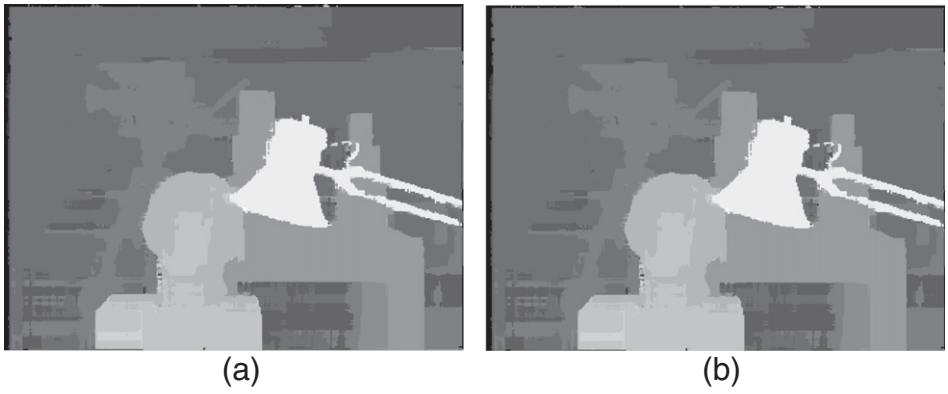


Fig. 9. (a) $d'_{LR}(\mathbf{x})$ and (b) $d'_{RL}(\mathbf{x})$ disparity maps after optimization.

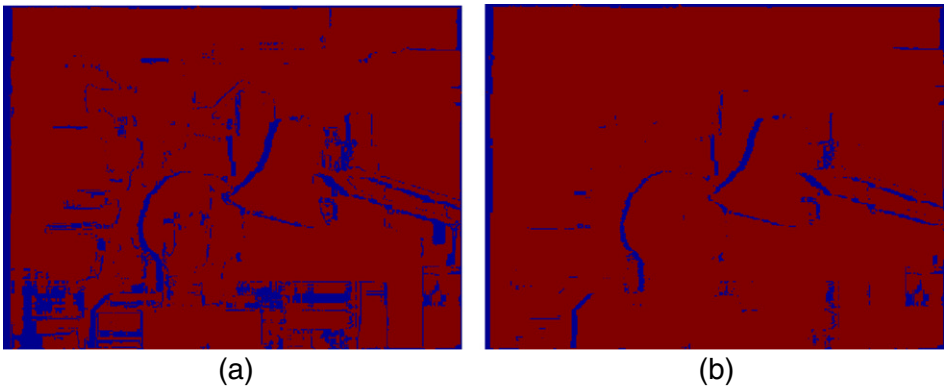


Fig. 10. Outliers map (a) $O_2^{T_{LR}=0}(\mathbf{x})$ for threshold $T_{LR} = 0$ and (b) $O_2^{T_{LR}=1}(\mathbf{x})$ for threshold $T_{LR} = 1$.

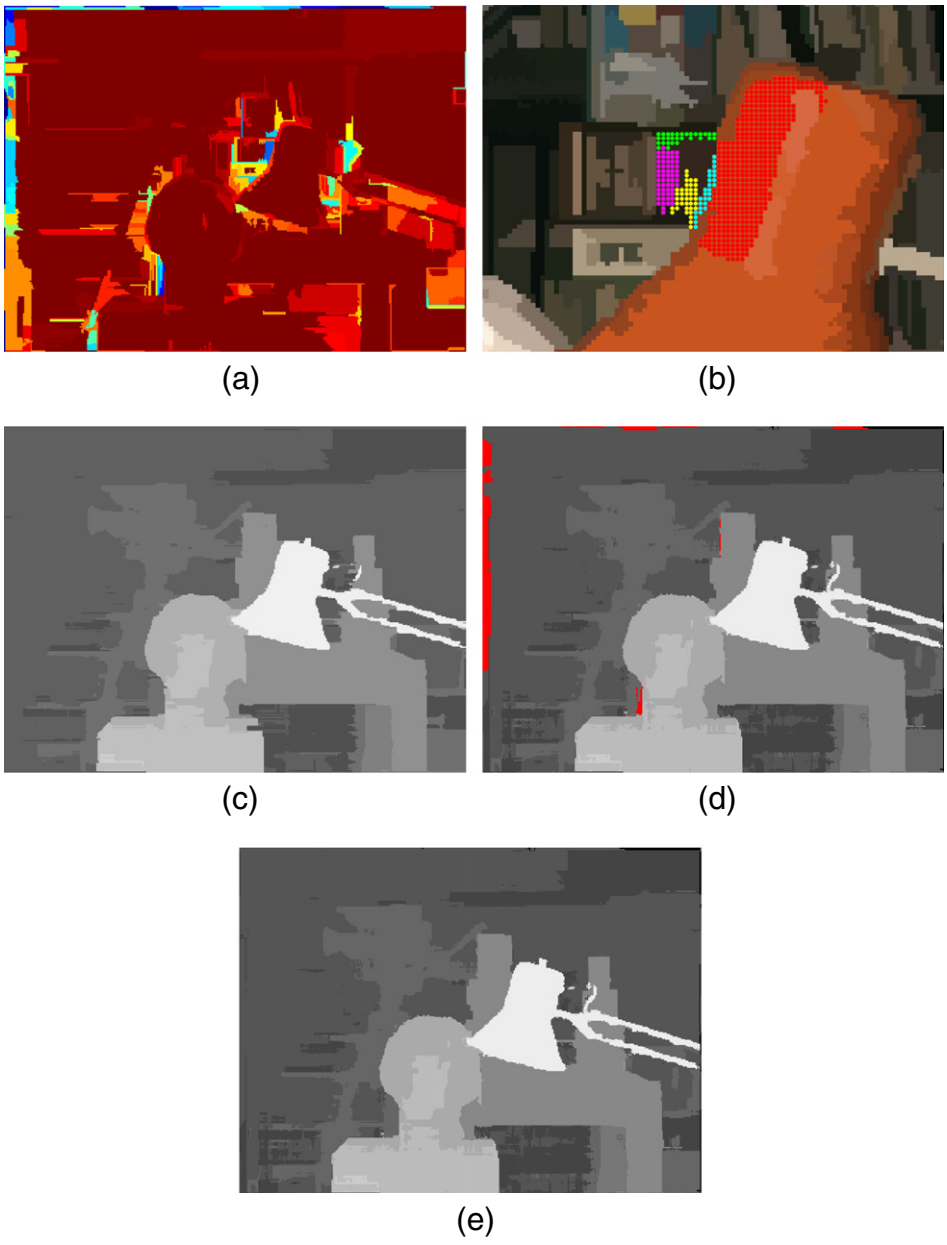


Fig. 11. (a) Reliability map, (b) an unreliable segment and its neighboring segments, (c) disparity map after applying basic occlusion handling, (d) disparity map mean-shift based segmentation occlusion handling and (e) disparity map after combined occlusion handling.

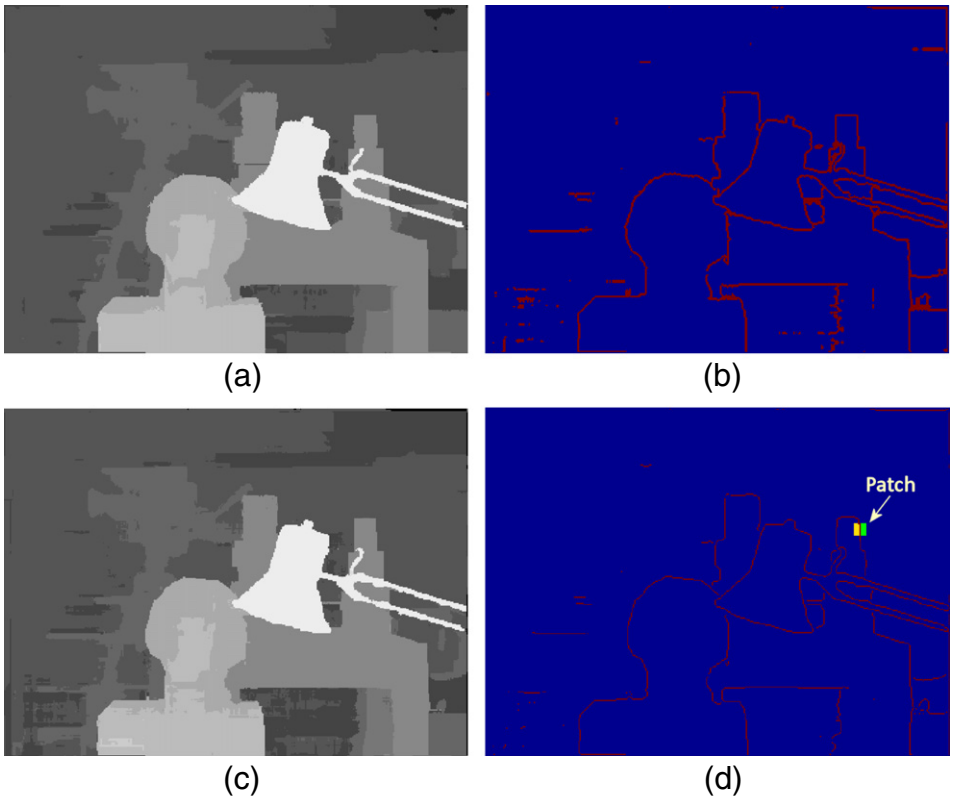


Fig. 12. (a) Disparity map, (b) its disparity edges and (c) disparity map after coarse discontinuity refinement (d) canny disparity edge detection.

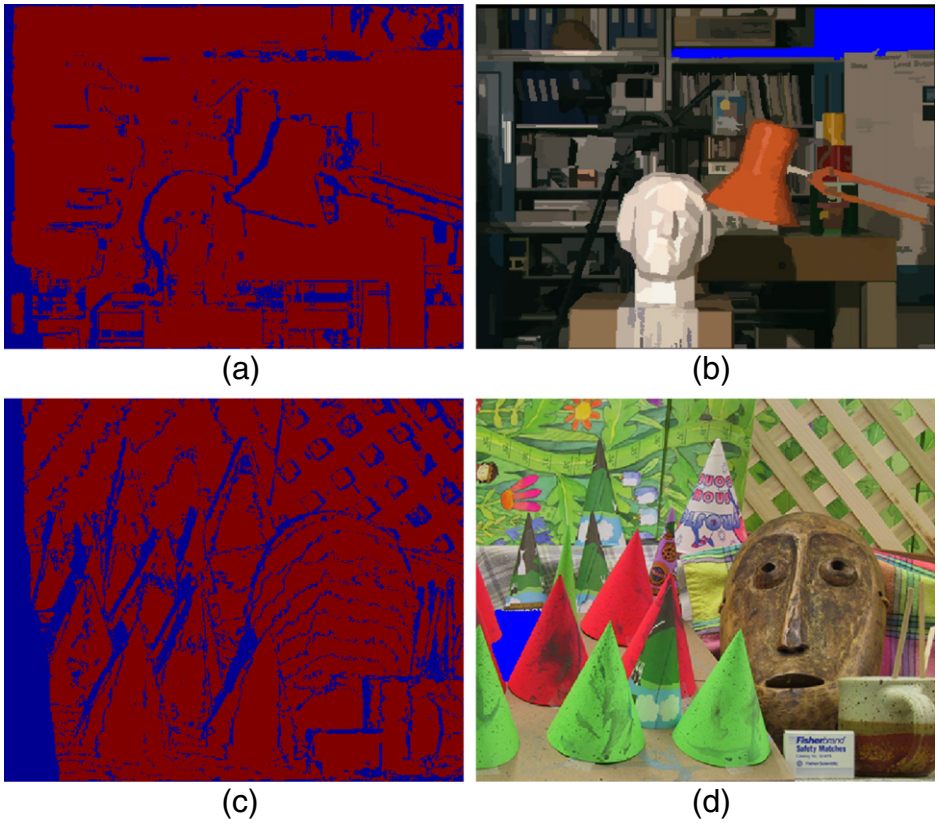


Fig. 13. Inlier pixels (red regions) in $O_U(\mathbf{x})$ for (a) the left "Tsukuba" image and (c) the left "Cones" image. A segment on (b) the left "Tsukuba" image and (d) the left "Cones" image (shown with blue).

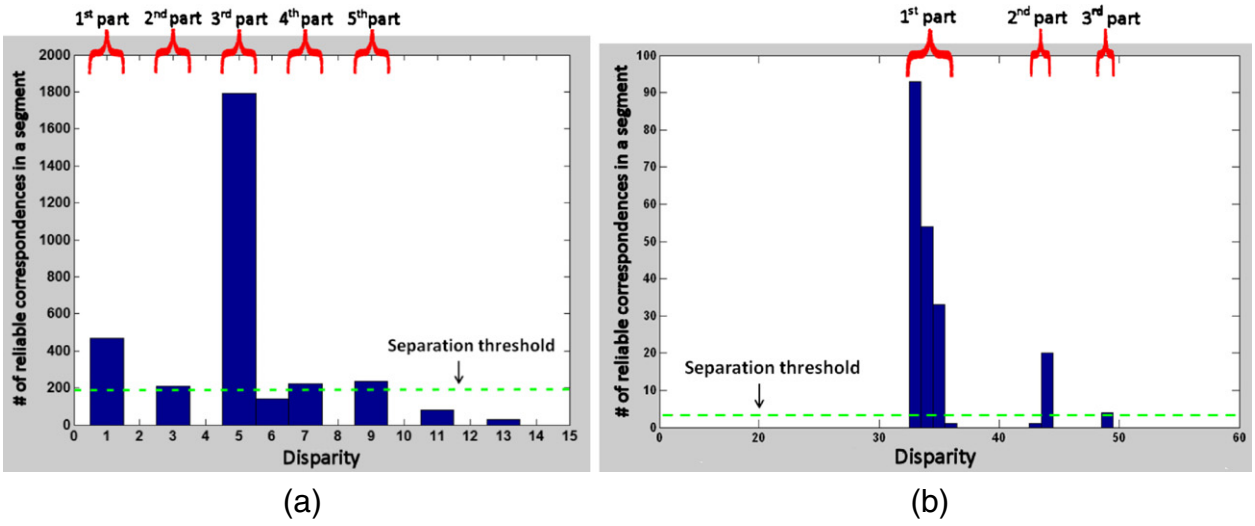


Fig. 14. Disparity histogram of the inlier pixels in a segment on (a) the left "Tsukuba" image and (b) the left "Cones" image.

After computing the four path costs, the final cost volume $V'''_{R-C}(\mathbf{x};\mathbf{d})$ is calculated from:

$$V'''_{R-C}(\mathbf{x};\mathbf{d}) = \frac{\sum_{\mathbf{r}\in\{\mathbf{r}_{lr},\mathbf{r}_{rl},\mathbf{r}_{ud},\mathbf{r}_{du}\}} L_{\mathbf{r}}(\mathbf{x};\mathbf{d})}{4}. \tag{22}$$

The WTA of $V'''_{R-C}(\mathbf{x};\mathbf{d})$ gives disparity map $d'_{LR}(\mathbf{x})$ (see Fig. 9a). If the right image is considered as reference image, then the disparity map $d'_{RL}(\mathbf{x})$ (see Fig. 9b) is acquired.

2.5. Disparity refinement

The disparity map, after cost volume optimization, contains a large number of outliers in occluded regions, uniform areas and near depth discontinuities. With the algorithmic steps, described through this section, these problematic areas can be handled efficiently in order to get a disparity map of high accuracy. The contribution of each refinement step, in the handling of problematic areas, is described later in Section 3.2.3.

2.5.1. Occlusion handling

2.5.1.1. Outliers detection. The disparity maps of the left image $d'_{LR}(\mathbf{x})$ (see Fig. 9a) and the right image $d'_{RL}(\mathbf{x})$ (see Fig. 9b) are taken into account so as to detect problematic areas. According to Left–Right consistency check:

$$\left|d'_{LR}(\mathbf{x}) - d'_{RL}\left(\mathbf{x} - d'_{LR}(\mathbf{x})\right)\right| \le T_{LR}, \tag{23}$$

for $T_{LR} = 0$ and $T_{LR} = 1$ the outliers maps $O_2^{T_{LR}=0}(\mathbf{x})$ (see Fig. 10a) and $O_2^{T_{LR}=1}(\mathbf{x})$ (see Fig. 10b) are acquired, respectively.

The detected outliers (due to occlusions) have to be filled with reliable disparities from neighboring areas. Occlusion handling is performed by combining two occlusion handling schemes that are executed independently. The first occlusion scheme is very simple and the second scheme is performed by exploiting the mean-shift color segmentation map.

2.5.1.2. Basic occlusion handling. The basic occlusion handling strategy is performed for the outlier map $O_2^{T_{LR}=0}(\mathbf{x})$ (see Fig. 10a). In more detail, an outlier pixel $\mathbf{x}$ is filled by the disparity of its closest inlier pixel. Practically, the disparity values of $\mathbf{x}$'s left nearest inlier pixel $\mathbf{x}_l$ and $\mathbf{x}$'s right nearest inlier pixel $\mathbf{x}_r$

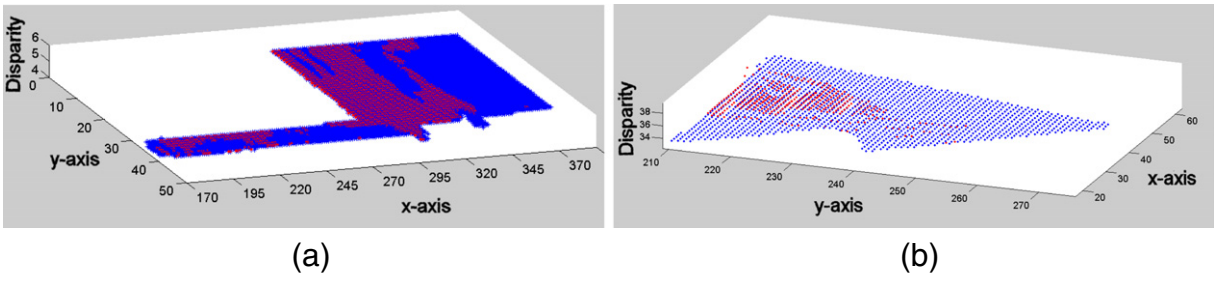


Fig. 15. Fit a plane (blue) to a segment, applying PCA on the reliable subset of pixels (red) for a segment on (a) the left "Tsukuba" image and (b) the left "Cones" image.

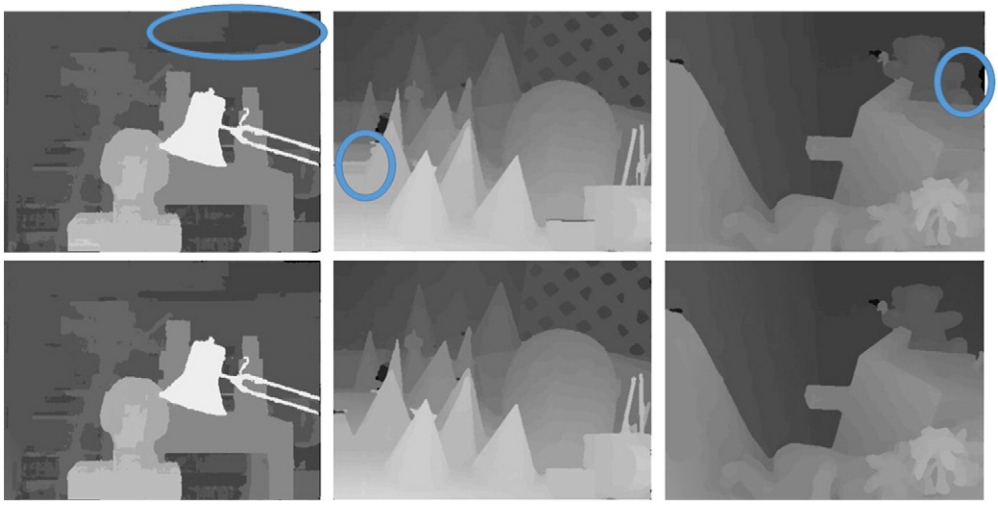


Fig. 16. Disparity maps before (1st row) and after applying uniform region handling (2nd row).

are denoted as $d'_{LR}(\mathbf{x}_l)$ and $d'_{LR}(\mathbf{x}_r)$, respectively. Then, the disparity value of $\min(d'_{LR}(\mathbf{x}_l), d'_{LR}(\mathbf{x}_r))$ is assigned to $\mathbf{x}$. The disparity map, after the basic occlusion handling, is visualized in Fig. 11c.

2.5.1.3. Mean-shift segmentation-based handling. There is high probability that the candidate "outlier" points for $T_{LR} = 1$, are not actual outliers. Instead, it is probable that there is a slight difference in the disparity estimation between the left and the right disparity maps. The following technique is applied to propagate reliably disparity information from the right disparity map to the left disparity map.

For a pixel ϕ, with $T_{LR}(\phi) = 1$, the subset of pixels within radius 7 from ϕ, which at the same time belong to the same segment as ϕ, is defined. This subset is used to estimate a reliability metric $Rel_{LR}(\phi)$, whose value is given by the division of the number of pixels within this subset with $T_{LR} = 0$ towards the total number of pixels in this subset. Correspondingly, for a pixel ψ, which is the correspondence of pixel ϕ in the right image, is computed similarly the metric $Rel_{RL}(\psi)$. The disparity of pixel ψ is propagated to pixel ϕ in case that $d'_{LR}(\phi) > d'_{RL}(\psi)$ and $Rel_{LR}(\phi) < Rel_{RL}(\psi)$. Pixels ϕ,

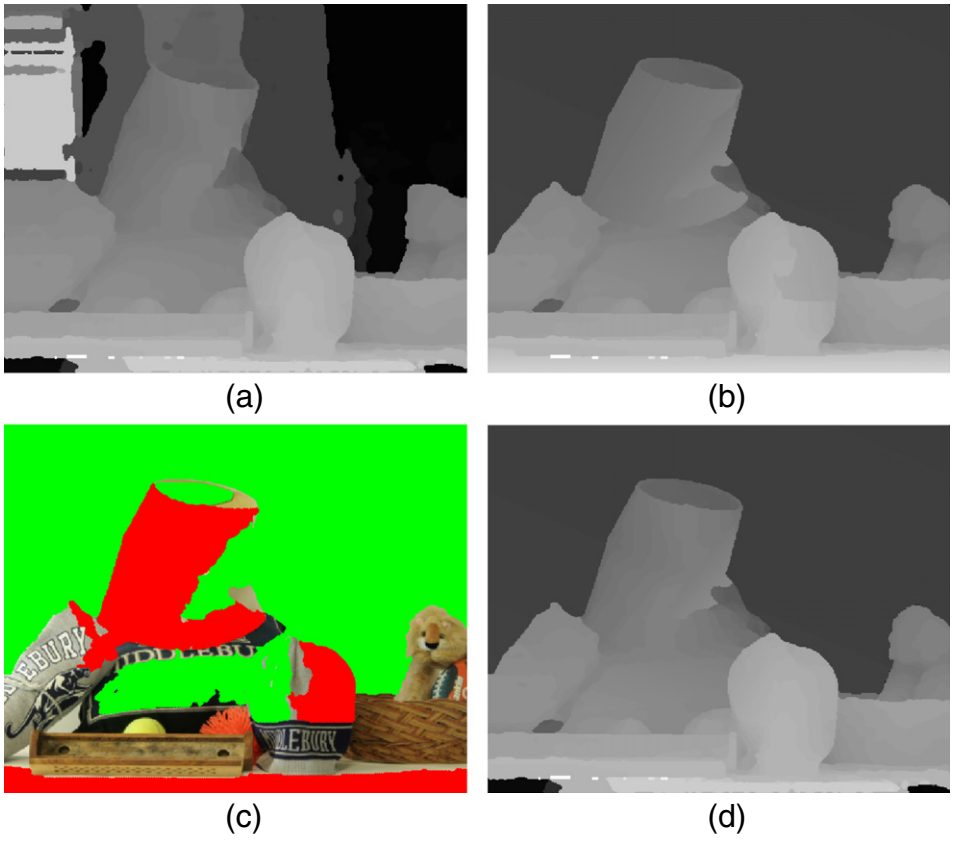


Fig. 17. (a) Disparity map without uniform area handling, (b) disparity map with uniform areas handling (without exploiting the MED_{fit} verification metric), (c) uniform areas where the disparity plane fitting is assumed as successful according to $MED_{fit} < 0.5$ are denoted with green, (d) disparity map after uniform area handling for the areas that satisfy $MED_{fit} < 0.5$.

Table 1
Computational time and the percentage of time spent on each algorithm's stage.

Image	Resolution	Disp. levels	Proposed (sec)	M.C. (%)	C.A. (%)	D.O. (%)	D.R. (%)
Tsukuba	384 × 288	15	24.33	4.30	88.93	0.93	5.84
Venus	434 × 383	20	49.25	3.71	89.72	0.92	5.65
Teddy	450 × 375	60	154.23	2.89	94.35	0.86	1.90
Cones	450 × 375	60	154.78	2.82	94.45	0.92	1.81

whose disparity has been propagated from their corresponding ψ pixels and the pixels that correspond to $T_{LR} = 0$ are considered as "unoccluded". These pixels are used in the application of the mean shift segmentation-based occlusion handling, as follows.

Initially, for each mean-shift segment the ratio of the unoccluded pixels inside this segment over the total number of segment pixels is evaluated. This ratio constitutes a reliability measure for the disparities inside this segment. Such a reliability map is illustrated in Fig. 11a. The warmer the color, the more reliable the disparities inside a segment are. A segment is considered as reliable if the ratio is over $T_{\mathbf{r}}$ (experimentally defined to be 0.3).

2.5.1.3.1. Reliable segments. For the outlier pixels inside a reliable segment S, a voting scheme that counts votes of the reliable pixels disparities is applied. In more detail, for each outlier pixel $\mathbf{x} \in S$, the inlier pixels that belong to S and lie within Euclidean distance R_S (radius of support region defined in Section 2.3.1) from $\mathbf{x}$ are taken into account in order to get the most frequent disparity. This disparity is propagated to $\mathbf{x}$, which is considered as reliable now. This process is repeated for all outliers inside a segment S.

2.5.1.3.2. Unreliable segments. For unreliable segments, the information from reliable neighboring segments is used to define their disparity. Reliable neighboring segments are the reliable segments that have common borders with the unreliable segments. For example, in Fig. 11b the unreliable segment is surrounded by the colored neighboring segments. The reliable neighboring segment that will propagate its prevalent disparity to the unreliable segment is the one that has the most similar color to the unreliable segment. Notice that the mean color of each segment was estimated during the mean-shift segmentation. The color similarity between two segments is defined as the mean Euclidean distance between their mean RGB colors and should be over T_S (experimentally defined to be 25). The disparity map, after handling occluded areas based on the mean-shift segmentation-based approach, is visualized in Fig. 11d. The red areas in Fig. 11d correspond to pixels that have not been handled using the mean-shift segmentation-based handling.

2.5.1.4. Combined occlusion handling. Finally, the occluded areas that have not been handled using the mean-shift based segmentation handling are filled with the disparities that have been estimated through the basic occlusion handling and in this way the combined disparity map of Fig. 11e is acquired.

2.5.2. Disparity edges refinement

Disparity edges, which correspond to depth discontinuities, may contain disparity errors [9]. Therefore, a two-step approach is used to refine the disparity information at the edges. The first step detects and handles the disparity edges at a coarser level and the second one at a finer level.

The pixels that belong to a disparity edge are assumed to have a difference greater or equal to 2 with at least one of their 4-adjacent pixels disparity. Otherwise, if the difference is below 2, then the surface varies smoothly and therefore one can assume that there is no depth discontinuity. Fig. 12b shows the disparity edges extracted from the disparity map of Fig. 12a. During the first step, around each pixel of the disparity edge, a circular region of radius 3 is defined. The disparities of the pixels that fall inside the circular region and at the same time belong to the same mean-shift segment, as the pixel of the disparity edge, are used to find the most frequent disparity value. This value is propagated to the edge pixel. The disparity result after the first step is depicted in Fig. 12c.

The second step handles discontinuities at a finer scale. Firstly, canny edge detection (see Fig. 12d) is applied to the disparity result of Fig. 12c. Canny can detect disparity edges at finer scale than the coarse previously-described step (this is evident when comparing Fig. 12b and Fig. 12d). Then a patch of size 3 × 3 is centered at each edge point and the disparity regions separated by the edge are found. Fig. 12d shows that the edge separates the patch into a yellow and green disparity region. The disparity region that contains the pixel with the greatest color similarity to the edge pixel (the color similarity is found according to the initial reference stereo image) gives its disparity to the considered pixel.

2.5.3. Uniform areas handling

Usually, images contain large uniform areas, where it is difficult to establish accurate pixel correspondences between two images. In order to deal with ambiguous matches in these areas, a new methodology is proposed in this paper.

2.5.3.1. Detection of uniform areas. Initially, large uniform areas on the image are detected. Large uniform areas are considered to be the mean-shift segments that contain over $2 \cdot {R_S}^2$ pixels (R_S is the radius of the support region as defined in Section 2.3.1). Then, each segment's "inlier" pixels are estimated and used for the uniform areas handling.

Table 2
Parameters testing.

	"Best"	$\beta = 0.25$	$\beta = 0.35$	$\lambda_{RGB} = 25$	$\lambda_{RGB} = 35$	$\gamma_c = 7$	$\gamma_c = 9$	$\lambda_{CEN} = 40$	$\lambda_{CEN} = 50$	$R_S = 17$	$R_S = 21$	No criterion
Avg. Rank	13.9	15.1	14.2	14.5	15.2	15.2	15.0	15.6	14.8	14.2	15.1	15.9
Nonocc	2.08	2.11	2.09	2.10	2.11	2.10	2.10	2.12	2.10	2.09	2.10	2.16
All	4.51	4.57	4.52	4.51	4.53	4.53	4.54	4.54	4.53	4.52	4.54	4.57
Disc	6.41	6.41	6.43	6.43	6.41	6.51	6.43	6.41	6.48	6.46	6.49	6.39

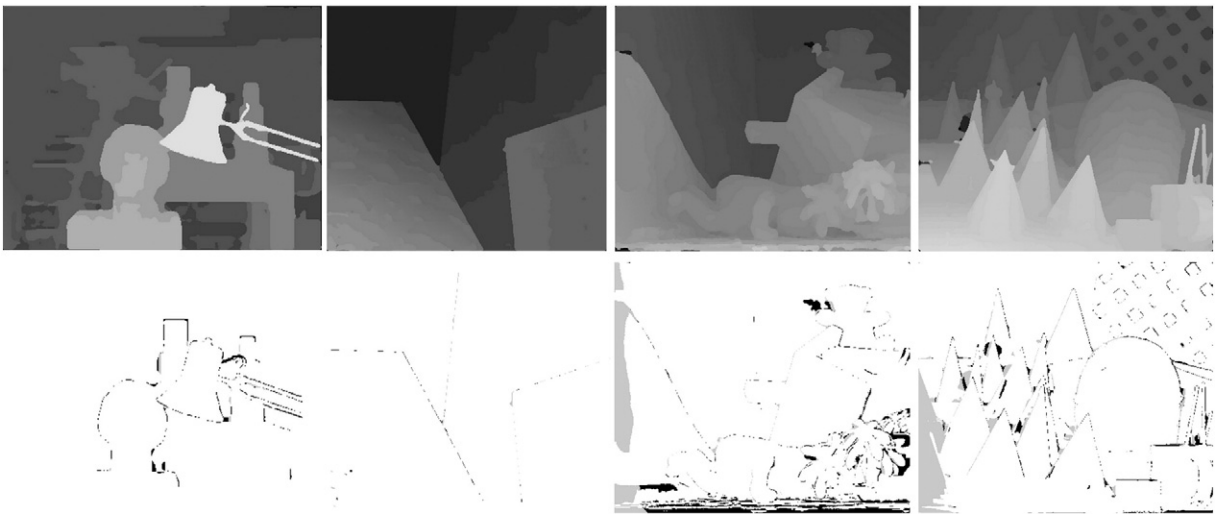


Fig. 18. Disparity maps generated with the proposed system and their corresponding disparity error maps for error threshold 1.

2.5.3.2. Inlier pixel regions. Inlier pixels $\mathbf{x}$ from $d_{LR}(\mathbf{x})$ (Section 2.3.2), are used for the uniform-areas handling. We determine inlier pixel regions as follows:

- The outliers map $O_1^{T_{LR}=1}(\mathbf{x})$ of Section 2.4.1 (see Fig. 8a), as well as the outliers map $O_2^{T_{LR}=0}(\mathbf{x})$ of Section 2.5.1 (see Fig. 10a) are considered in order to acquire their union, which defines the overall outliers map $O_U(\mathbf{x})$. In $O_U(\mathbf{x})$, outlier pixels are those that are outliers in either $O_1^{T_{LR}=1}(\mathbf{x})$ or $O_2^{T_{LR}=0}(\mathbf{x})$. Let X_{In} be the set of inlier pixels in $O_U(\mathbf{x})$.

A visual example is given in the first row of Fig. 13. Fig. 13a shows the overall outliers map that is generated after the union of the outliers maps acquired in Sections 2.4.1 and 2.5.1. The inlier pixels X_{In} are denoted with red color.

2.5.3.3. Extraction of a reliable pixels, based on histogram analysis. A histogram analysis, based on the inlier pixels' disparities $d_{LR}(X_{In})$ is applied in order to acquire a reliable subset of the pixels. For instance, for the mean-shift segment of Fig. 13b (marked with blue color), the histogram of the disparities of the inlier pixels inside this segment is depicted in Fig. 14a.

Theoretically, the disparities of the pixels in a segment S should vary continuously within a disparity range, since they belong to the same continuous surface. Based on this assumption, the employed approach is followed to get the subset of the reliable pixels.

Initially, the histogram of disparities is separated into parts (each part expresses a disparity range), as shown in Fig. 14a. To separate the histogram into parts, bins with a height below a "separation threshold" are ignored, so that they do not affect the separation process. This threshold is selected equal to: $\frac{\text{Number of inlier pixels in } S}{\text{Number of possible disparities}}$. The reliable subset of inlier pixels includes the pixels whose disparities belong to the histogram part with the most numerous population (3rd part of Fig. 14a).

2.5.3.4. Planar fitting. Afterwards, the reliable pixels and their disparities (red points in Fig. 15a) are used to fit a planar surface to the segment. The robust method of Principal Components Analysis (PCA) described in [43] is used to estimate the parameters of the plane. The two first principal components define the plane. Let the estimated plane be: $d_p(\mathbf{x}) = \mathbf{p}^T \cdot \mathbf{x}$, where $\mathbf{p} = [p_1, p_2]^T$. Then each $\mathbf{x} \in S$ is assigned the disparity $d_p(\mathbf{x})$. The new disparity values inside the segment are depicted with blue in Fig. 15a.

A second example of uniform area handling is given considering the Cones stereo pair. In brief, Fig. 13c shows the overall outliers map. For the mean-shift segment of Fig. 13d (marked with blue color), the histogram of the disparities of the inlier pixels inside this segment is depicted in Fig. 14b. The reliable subset of inlier pixels includes the pixels whose disparities belong to the 1st histogram part of Fig. 14b. The reliable pixels and their disparities (red points in Fig. 15b) are used to fit a planar surface to the segment. The new disparity values inside the segment are depicted with blue in Fig. 15b.

Table 3
The rankings in the Middlebury benchmark.

Algorithm	Avg. rank	Tsukuba			Venus			Teddy			Cones		
		Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc
LCU [44]	12.0	1.06	1.34	5.50	0.07	0.26	1.03	3.68	9.95	10.4	1.63	6.87	4.82
TSGO [45]	12.1	0.87	1.13	4.66	0.11	0.24	1.47	5.61	8.09	13.8	1.67	6.16	4.95
Proposed	13.9	1.02	1.23	5.51	0.08	0.20	1.11	5.16	9.43	13.0	2.07	7.16	5.97
JSOSP + GCP [46]	14.1	0.74	1.34	3.98	0.08	0.16	1.15	3.96	10.1	11.8	2.28	7.91	6.74
ADCensus [9]	16.5	1.07	1.48	5.73	0.09	0.25	1.15	4.10	6.22	10.9	2.42	7.25	6.95
AdaptingBP [6]	20.4	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32

Table 4
Evaluation of the two-phase combination strategy.

	Init.	Phase1	Phase2	CENSUS	SIFT
Nonocc	8.81	7.91	6.43	18.5	15.0
All	14.4	13.6	12.2	23.1	19.8
Disc	15.9	15.6	14.6	27.3	27.8

Fig. 16 shows three examples of uniform region handling. In the first and second rows of Fig. 16, the disparity results before and after uniform region handling are visualized, respectively. The first and second columns include the result of handling the blue-colored segments of Fig. 13b and Fig. 13d, respectively. The third column shows an example for the Teddy stereo pair. The examples in the second and third columns show clearly the improvements in the disparity maps after applying the plane fitting process.

However, it is not always valid to assume that large areas with low texture are planar. Additionally, some large areas may have been wrongly segmented, leading to inaccurate plane fitting. Therefore, in this paper a specific metric is adopted, which is used to verify if the planar fitting is successful. This metric is the median of the absolute differences between the initial disparities of the reliable pixels and the disparities of the reliable pixels that are estimated after the plane fitting and is defined as: MED_{fit} (measured in disparity units). In this paper, the condition $MED_{fit} < 0.5$ has to be satisfied, in order to consider the planar fitting as successful.

In Fig. 17 an example is visualized. Uniform area handling is applied on the Midd1 stereo pair, which belongs to the extended stereo dataset, and contains large low-textured areas. In Fig. 17a the estimated disparity map without applying the uniform areas handling is depicted. It is obvious that disparity estimation is not reliable in low-textured areas. Fig. 17b shows the disparity map after applying uniform areas handling to all low-textured areas. Fig. 17c visualizes with green the low-textured areas with $MED_{fit} < 0.5$ and with red the low-textured areas with $MED_{fit} \geq 0.5$. In Fig. 17d the disparity map after applying uniform areas handling only for the green low-textured areas is depicted. The disparity error for the case of all regions and $\Delta d > 1$ for the Midd1 stereo pair is 40.65%, 14.88% and 9.69% for the disparity maps of Fig. 17a, Fig. 17b and Fig. 17d, respectively. Therefore, this example verifies the efficiency of the uniform areas handling to decrease the disparity estimation error.

A median filter using a 5×5 neighborhood is applied to the disparity result that is generated after executing all the steps of our approach, in order to remove spurious disparities before acquiring the final disparity map.

2.6. Computational cost

A non-optimized C++ implementation of the algorithm is used to report on the required computational time. The algorithm was executed on a desktop PC with Core i7-3770 3.40 GHZ CPU and 8 GB RAM. The total processing time, using as input the four stereo pairs of the Middlebury evaluation benchmark [26], is indicated in Table 1. The measured time is the average of 5 separate runs. Additionally, this table provides the percentage of the total time that is spent for each of algorithm's stages, which include: 1. the matching cost computation stage (M.C.) (Section 2.2), 2. the cost aggregation stage (C.A.) (Section 2.3), 3. the disparity optimization stage (D.O) (Section 2.4) and 4. the disparity refinement stage (D.R.) (Section 2.5). The cost aggregation is the most computational expensive stage (on average 91.84 % of the total processing time). Nevertheless, this stage can be parallelized since cost aggregation can be performed independently for non-overlapping parts of the image.

Concluding, most parts of the algorithm have low computational cost. The step of the algorithm with increased computational cost includes the adaptive support weight cost aggregation. However, this time consuming part can be implemented in Graphics Processing Units (GPUs) as can be verified in [47]. Additionally, there are works, such as [48,49] that propose approximations to derive fast implementations of the original adaptive support weight algorithm [12]. The drawback of these methods is that they sacrifice quality for high computational speed [47].

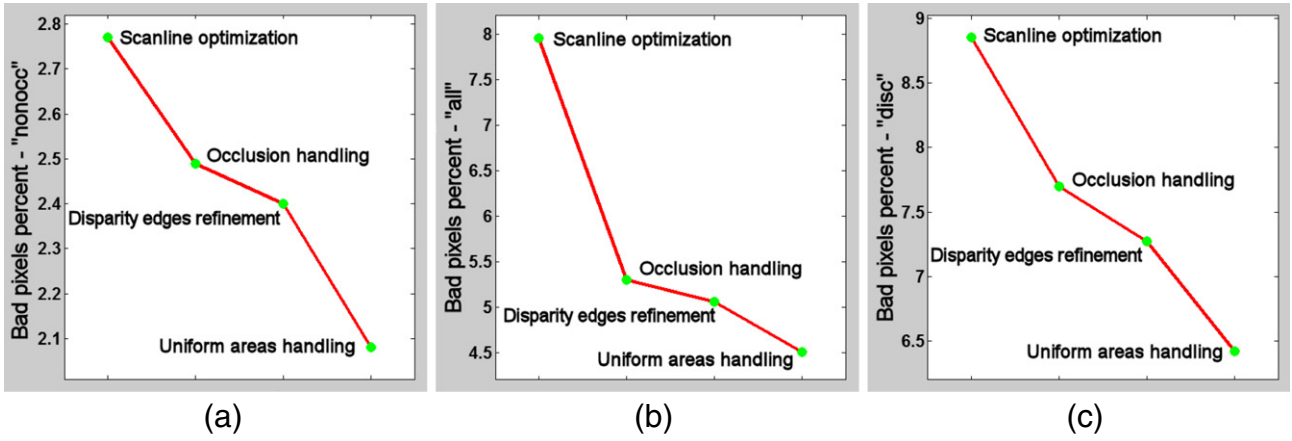


Fig. 19. Average percent of bad pixels after applying sequentially refinement steps for (a) non-occluded regions, (b) all regions and (c) near depth discontinuities regions.

3. Experimental results

In this section the experimental results on multiple datasets are presented. In more detail, the four stereo pairs of the Middlebury online stereo evaluation benchmark [26], except for the evaluation of this method, are used to select a set of optimum parameters and to evaluate the disparity refinement process. Furthermore, this section presents experimental results on 27 additional Middlebury stereo pairs in order to verify the efficiency of the proposed approach.

3.1. Set of optimum parameters

The parameters used for the experiments are the same for all tested stereo pairs. More specifically, β (defined in Section 2.2.1) is set equal to $\beta = 0.3$, while the parameters used for the cost functions (see Section 2.2.3) are $\lambda_{RGB} = 30$, $\lambda_{CEN} = 45$ and $\lambda_{SIFT} = 45$. The radius of the support area (see Section 2.3.1) is set equal to $R_S = 19$ and the adaptive weight parameters are $\gamma_c = 8$ and $\gamma_e = R_S$. Those values are selected based on experiments that were performed on the Middlebury Online Stereo Evaluation Benchmark.

In the column "Best" of Table 2, the numeric results from the Middlebury Stereo evaluation for the disparity maps extracted using these optimum parameters, are given. The results include the overall performance measure ("Avg. Rank"), the error in non-occluded regions ("Nonocc"), the error in all regions ("All") and the error near depth discontinuities ("Disc"). In Section 3.2.4 further parameters testing is performed.

3.2. Middlebury online stereo evaluation benchmark

3.2.1. Disparity results

The disparity results of the proposed method, for the optimum parameters set, accompanied with the disparity error maps as extracted by the Middlebury evaluation system are visualized in Fig. 18. Errors in non-occluded and occluded regions are marked in black and gray respectively.

The ranking results in Table 3 (reference period: November 2014), for error threshold equal to 1, indicate that the proposed method is 3rd out of 161 methods that are included in the Middlebury Stereo Evaluation. However, no information on the 1st [44] and 2nd [45] ranked methods is available, since they are currently under review. Therefore, the proposed method ranks 1st among already published methods. More specifically, the proposed method ranks: 9th for the "Tsukuba" image pair, 3rd for the Venus image pair, 32nd for the Teddy image pair and 5th for the "Cones" image pair.

The 32nd position in the ranking for the Teddy image pair is because of the very slanted surface at the bottom of the image, where the proposed method cannot handle well the very slanted surface. However, it can be deduced from the experimental results that the proposed method outperforms the rest of the published stereo methods, which are evaluated online in the Middlebury stereo evaluation benchmark, in image areas excluding very slanted surfaces.

3.2.2. Evaluation of the two-phase combination strategy

The improvement in the accuracy of the initial disparity map, which is achieved by using the two-phase combination strategy of Section 2.3.2, is evaluated according to the Middlebury online evaluation system. Table 4 depicts the average percent of bad pixels for the disparity maps generated using the four Middlebury image pairs. In specific, this table includes results for non-occluded regions ("Nonocc"), all regions ("All") and regions near depth discontinuities ("Disc").

The evaluation results for the disparity map resulting via WTA from V_{R-C} are given in the "Init." column. The evaluation results for the disparity map resulting via WTA from $V_{R-\zeta}$ (which is acquired after applying first combination phase) are given in the "Phase1" column. The average numeric results for the disparity map d_{LR} resulting via WTA from V''_{R-C} (which is acquired after applying second combination phase) are given in the "Phase2" column. Obviously, each combination phase assists in improving the accuracy of the generated disparity map.

Additionally, Table 4 includes in the "CENSUS" column the evaluation results for the disparity map resulting via WTA from V_{CEN} and in the "SIFT" column the evaluation results for the disparity map resulting via WTA from V_{SIFT}. Though, the results in "CENSUS" and "SIFT" are worse that the results in "Init." the efficient exploitation of V_{CEN} and V_{SIFT} in the two-phase combination strategy improves the disparity estimation accuracy.

3.2.3. Evaluation of the disparity refinement process

Furthermore, the Middlebury online benchmark is exploited in order to examine the improvement introduced by the proposed disparity refinement steps. Fig. 19 depicts how the average percent of bad pixels decreases after applying sequentially each of the disparity refinement steps, which include occlusion handling, disparity edges refinement and uniform regions handling. Fig. 19 includes results for non-occluded regions (see Fig. 19a), all regions (see Fig. 19b) and regions near depth discontinuities (see Fig. 19c). As it is expected, the occlusion handling decreases the bad pixels percent more than the rest refinement steps, since it handles large outlier areas. Disparity edges refinement and uniform regions handling improve further the accuracy, so that the proposed framework becomes the top ranked published method in the Middlebury stereo evaluation.

3.2.4. Further parameters testing

As mentioned in Section 3.1, the column "Best" of Table 2 gives the numeric disparity estimation results using optimum parameters. In the rest columns of Table 2, we provide the results in the case that all parameters are kept the same as the optimum ones, except for the parameter in the top of the column. For each parameter a smaller and a larger value than the optimum one are tested. Table 2 verifies that the optimum parameters give the best results.

The last column of Table 2, with the annotation "No criterion", gives the results of this method for the best set of parameters, with the difference that in this case the new criterion for the definition of the smoothness terms in Eq. (21) is not used. The results prove that without the exploitation of the new criterion the disparity accuracy decreases.

The segmentation maps are exploited in different stages of this method. Therefore, it is important to verify that small variations to the optimum parameters $(h_s, h_r) = (3, 3)$ that adjust the segmentation result (see Section 2.1.2) do not affect significantly the performance of this method. Table 5 exhibits the error results for different values of the spatial radius and space feature radius. The rest of parameters are set to their optimum value.

For all parameter tests, the proposed method ranks in the top five ranking positions though the disparity accuracy decreases. This fact proves that this approach maintains its good disparity estimation accuracy even with changes to the optimum parameters.

Table 5
Segmentation parameters testing.

	$(h_s, h_r) = (2, 3)$	$(h_s, h_r) = (3, 4)$	$(h_s, h_r) = (4, 4)$
Avg. rank	16.1	16.9	16.0
Nonocc	2.16	2.14	2.12
All	4.77	4.69	4.69
Disc	6.51	6.54	6.59

3.3. Extended comparison

Many of methods are evaluated on just the four stereo pairs from the Middlebury online stereo database, which are mentioned in Section 3.2. However, evaluation on limited data is not sufficient to give a clear picture of the overall performance of an algorithm, since the average error rates of the best performing techniques are close to each other. For this reason, except for the four stereo pairs from the Middlebury online stereo evaluation benchmark, evaluation is performed on two additional Middlebury datasets in order to assess more rigorously the performance of the proposed methodology. The 2005 and 2006 datasets, presented in [50], include 27 stereo pairs with their ground truth. The error percentage is measured for both non-occluded and all regions.

Table 6 shows the results for the percentage of erroneous pixels having 1 or 2 disparity level difference with respect to ground truth. The results regarding the rest of methods in Table 6 are copied from the very recent work of [8]. The column "All" refers to case where all pixels on the disparity map are considered to estimate the percentage of erroneous pixels, while the term "Visible" refers to the case where only the pixels on the disparity map that correspond to unoccluded regions are considered to estimate the percentage of erroneous pixels. In general, the proposed work gives better results for the case of "All" regions than the rest of the methods that are evaluated in [8]. More specifically, for the case of "All" regions and $\Delta d > 1$, $\Delta d > 2$ the disparity errors of our approach are 2.02% and 2.7% less than the second best method, respectively. The low error for $\Delta d > 2$ indicates that the estimated disparity for some pixels is very close to their ground truth disparity and differs just 2 disparity levels.

The disparity maps for the 27 stereo pairs, with their respective disparity error maps for $\Delta d > 1$, can be found in slides 4–17 of the supplementary material that accompanies this paper.

4. Conclusions

In this paper a method that produces very accurate disparity results for stereo image pairs is presented. In order to achieve increased accuracy, the proposed method uses efficiently three cost metrics to acquire a reliable combined cost volume. The optimization of the cost volume is performed using a semi-global matching method, where a new criterion is introduced for the definition of the smoothness penalty terms that improves the disparity results. Outliers handling is performed combining a simple scheme and mean-shift segmentation base occlusion handling. Another innovative aspect of this paper is the way disparities are filtered based on histogram analysis in order to be used in uniform regions handling. The remarkable performance of the proposed method is verified experimentally using the Middlebury evaluation benchmark and an extended stereo dataset. The ideas introduced in this paper could be used or extended by future stereo algorithms in order to boost their accuracy performance.

Future work will focus on improving the disparity estimation for very slanted surfaces. Probably, the employment of a different cost aggregation approach is needed to achieve this objective.

Table 6
The error results for the extended stereo datasets.

Error%	$\Delta d > 1$	$\Delta d > 1$	$\Delta d > 2$	$\Delta d > 2$
	All	Visible	All	Visible
Proposed	12.13	8.26	7.64	4.74
Inf. Permeability [8]	14.15	7.98	10.34	6.46
Guided Filter [14]	15.06	8.40	11.82	6.80
Geodesic Support [17]	16.49	9.85	11.76	8.04
Var. Cross [16]	17.13	8.81	12.69	7.04
Adapt. sup. [12]	16.94	9.54	13.10	7.42

Acknowledgment

This work was supported by the EU funded IP project REVERIE under contract 287723.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.imavis.2014.12.001.

References

[1] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, IJCV 47 (1–3) (2002) 7–42.
[2] R. Zabih, J. Woodfill, A Non-parametric Approach to Visual Correspondence, IEEE TPAMI, 1996.
[3] W. Fife, J. Archibald, Improved census transforms for resource-optimized stereo vision, IEEE TCSVT 23 (2013) 60–73.
[4] N.Y. Chang, T. Tsai, B. Hsu, Y. Chen, T. Chang, Algorithm and architecture of disparity estimation with mini-census adaptive support weight, IEEE TCSVT 20 (6) (2010) 792–805.
[5] M. Humenberger, C. Zinner, M. Weber, W. Kubinger, M. Vincze, A Fast Stereo Matching Algorithm Suitable for Embedded Real-time Systems, vol. 114, no. 11, Elsevier CVIU, 2010, pp. 1180–1202.
[6] A. Klaus, M. Sormann, K. Karner, Segment-based Stereo Matching Using Belief Propagation and a Self-adapting Dissimilarity Measure, ICPR, 2006, pp. 15–18.
[7] X. Sun, X. Mei, S. Jiao, M. Zhou, H. Wang, Stereo Matching with Reliable Disparity Propagation, 3DIMPVT, 2011.
[8] C. Cigla, A.A. Alatan, Information Permeability for Stereo Matching, Elsevier Signal Processing: Image Communication, 2013.
[9] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, X. Zhang, On Building an Accurate Stereo Matching System on Graphics Hardware, ICCV Workshop on GPU in Computer Vision Applications, 2011.
[10] M. Gong, R.G. Yang, W. Liang, M.W. Gong, A performance study on different cost aggregation approaches used in real-time stereo matching, IJCV 75 (2007) 283–296.
[11] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, Classification and Evaluation of Cost Aggregation Methods for Stereo Correspondence, CVPR, 2008, pp. 1–8.
[12] K.-J. Yoon, I.S. Kweon, Adaptive support-weight approach for correspondence search, IEEE TPAMI 28 (2006) 650–656.
[13] L. Di Stefano, F. Tombari, S. Mattoccia, Segmentation-based adaptive support for accurate stereo correspondence, Proc. IEEE Pacific-Rim Symp. Image and Video, 2007.
[14] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, M. Gelautz, Fast cost-volume filtering for visual correspondence and beyond, IEEE TPAMI 35 (2) (2013) 504–511.
[15] Q. Yang, P. Ji, D. Li, S.J. Yao, M. Zhang, Fast Stereo Matching Using Adaptive Guided Filtering, vol. 32, no. 3, Elsevier Image and Vision Computing, 2014, pp. 202–211.
[16] K. Zhang, J. Lu, G. Lafruit, Cross-based local stereo matching using orthogonal integral images, IEEE TCSVT 19 (7) (2009) 1073–1079.
[17] A. Hosni, M. Bleyer, M. Gelautz, C. Rhemann, Local Stereo Matching Using Geodesic Support Weights, IEEE ICIP, 2009, pp. 2093–2096.
[18] V. Kolmogorov, R. Zabih, Computing Visual Correspondence with Occlusions Using Graph Cuts, ICCV, 2, 2001, pp. 508–515.
[19] Z.F. Wang, Z.G. Zheng, A Region Based Stereo Matching Algorithm Using Cooperative Optimization, CVPR, 2008, pp. 1–8.
[20] J. Kim, K. Lee, B. Choi, S. Lee, A Dense Stereo Matching Using Two-pass Dynamic Programming with Generalized Ground Control Points, CVPR, 2005, pp. 1075–1082.
[21] H. Hirschmller, Stereo processing by semiglobal matching and mutual information, IEEE TPAMI 30 (2008) 328–341.
[22] Q. Yang, L. Wang, R. Yang, H. Stewenius, D. Nister, Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling, IEEE TPAMI 31 (2009) 492–504.
[23] D. Min, K. Sohn, An Asymmetric Post-processing for Correspondence Problem, SPIC, vol. 25, no. 2, 2010, pp. 130–142.
[24] S. Mattoccia, F. Tombari, L.D. Stefano, Stereo Vision Enabling Precise Border Localization Within a Scanline Optimization Framework, Proc. ACCV, 2007, pp. 517–527.
[25] M. Humenberger, T. Engelke, W. Kubinger, A census-based stereo vision algorithm using modified semi-global matching and plane-fitting to improve matching quality, CVPR ECV Workshop, 2010.
[26] Middlebury stereo evaluation, http://vision.middlebury.edu/stereo/.
[27] H. Bay, T. Tuytelaars, L. Van Gool, SURF: speeded up robust features, Proc. ECCV, 2006.
[28] Y.S. Heo, K.M. Lee, S.U. Lee, Joint depth map and color consistency estimation for stereo images with different illuminations and cameras, IEEE TPAMI 35 (5) (2013) 1094–1106.
[29] G. Saygili, L.J.P. van der Maaten, E.A. Hendriks, Feature-based Stereo Matching Using Graph Cuts, ASCI, 2011.
[30] L. Hong, G. Chen, Segment-based Stereo Matching Using Graph Cuts, CVPR, 2004, pp. 74–81.
[31] L. Xu, O.C. Au, W. Sun, Y. Li, J. Li, Hybrid Plane Fitting for Depth Estimation, APSIPA, 2012.
[32] M. Fischler, R. Bolles, Random sample consesus: a paradigm for model fitting with applications to image analysis and automated cartography, CACM 6 (1981) 381–395.
[33] S. Kumar, C. Micheloni, C. Piciarelli, G.L. Foresti, Stereo Rectification of Uncalibrated and Heterogeneous Images, vo.31, no. 11, Elsevier Pattern Recognition Letters, 2010, pp. 1445–1452.

[34] M. Bleyer, C. Rother, P. Kohli, Surface stereo with soft segmentation, CVPR, 2010, pp. 1570–1577.
[35] T. Liu, P. Zhang, L. Luo, Dense stereo correspondence with contrast context histogram, segmentation-based two-pass aggregation and occlusion handling, Proc. IEEE Pacific-Rim Symp. Image and Video Technology, 2009.
[36] Edge detection and image segmentation (EDISON) system, http://coewww.rutgers.edu/riul/research/code/EDISON/doc/ref.html.
[37] D. Comanicu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE TPAMI 24 (2002) 603–619.
[38] P. Meer, B. Georgescu, Edge detection with embedded confidence, IEEE TPAMI 23 (2001) 1351–1365.
[39] C. Christoudias, B. Georgescu, P. Meer, Synergism in low-level vision, ICPR, 4, 2002, pp. 150–155.
[40] J.R.R. Uijlings, A.W.M. Smeulders, R.J.H. Scha, Real-time visual concept classification, IEEE Trans. Multimedia 12 (2010) 665–682.
[41] J.M. Geusebroek, A.W.M. Smeulders, J. van de Weijer, Fast anisotropic gauss filtering, IEEE Trans. Image Process. 12 (8) (2003) 938–943.
[42] X. Hu, P. Mordohai, Evaluation of stereo confidence indoors and outdoors, CVPR, 2010, pp. 1466–1473.
[43] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: a new approach to robust principal components analysis, Technometrics 47 (2005) 64–79.
[44] Anonymous, Using Local Cues to Improve Dense Stereo Matchingsubmited to IEEE CVPR 2015.
[45] M. Mozerov, J. van Weijer, Accurate Stereo Matching by Two Step Global Optimizationsubmitted to IEEE TIP 2014.
[46] J. Liu, C. Li, F. Mei, Z. Wang, 3D Entity-based Stereo Matching with Ground Control Points and Joint Second Order Smoothness Prior, SPRINGER, The Visual Computer, 2014.
[47] A. Hosni, M. Bleyer, M. Gelautz, Secrets of Adaptive Support Weight Techniques for Local Stereo Matching, vol. 117, no. 6, Elsevier CVIU, 2013, pp. 620–632.
[48] S. Mattoccia, S. Giardino, A. Gambini, Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering, ACCV, 2009, p. 371382.
[49] K. Zhang, G. Lafruit, R. Lauwereins, L. Gool, Joint integral histograms and its application in stereo matching, ICIP, 2010, p. 817820.
[50] H. Hirschmuller, D. Scharstein, Evaluation of cost functions for stereo matching, CVRP, 2007, p. 18.