Accurate Stereo Matching by Two-Step Energy Minimization

Mikhail G. Mozerov, *Member, IEEE*, and Joost van de Weijer, *Member, IEEE*

***Abstract*— In stereo matching, cost-filtering methods and energy-minimization algorithms are considered as two different techniques. Due to their global extent, energy-minimization methods obtain good stereo matching results. However, they tend to fail in occluded regions, in which cost-filtering approaches obtain better results. In this paper, we intend to combine both the approaches with the aim to improve overall stereo matching results. We show that a global optimization with a fully connected model can be solved by cost-filtering methods. Based on this observation, we propose to perform stereo matching as a two-step energy-minimization algorithm. We consider two Markov random field (MRF) models: 1) a fully connected model defined on the complete set of pixels in an image and 2) a conventional locally connected model. We solve the energy-minimization problem for the fully connected model, after which the marginal function of the solution is used as the unary potential in the locally connected MRF model. Experiments on the Middlebury stereo data sets show that the proposed method achieves the state-of-the-arts results.**

***Index Terms*— Stereo matching, energy minimization, bilateral filter, fully connected MRF model.**

I. Introduction

STEREO matching is one of the fundamental problems of computer vision [1], [2]. Stereo matching is important for a large variety of computer vision applications, such a intermediate view generation, 3D scene reconstruction, autonomous driving systems and robotics. Stereo matching can be divided into two categories: cost filtering methods (also known as cost aggregation methods) and energy minimization methods.

Early stereo methods predominantly use filters with local support windows and call this filtering cost aggregation. The main idea of most cost filter approaches is based on the assumption that all pixels in the matching neighborhood have similar disparities. The matching neighborhood can be defined in various spaces and with different norms. Early methods that used fixed or adaptive windows as the matching neighborhood were extended into algorithms which operated in the bilateral color-image space [3]–[7]. The fixed spatial extend of the cost filtering approaches limits the range in which information can be propagated. A problem which was later addressed by energy minimization approaches.

The energy minimization methods have lately attracted much attention in computer visions, especially in the context of image segmentation and optical flow estimation. The first implementations of the energy minimization methods such as belief propagation [8] and graph cuts [9] in stereo matching have provided a significant progress in disparity map estimation [10], [11]. However the occlusion problem introduces uncertainty in the choice of the constraint parameters: over-penalization of the smoothness term can help to overcome the ambiguity in occluded regions but this leads to errors for fine disparity structure [12].

A. Motivation

In stereo matching for each pixel we have to choose between several possible correspondences based on decision values (costs). There are three well known base problems in stereo matching, which make a naive pixel-wise correspondence search useless: noise, occlusion, and inherent matching uncertainty due to the possible local color sparseness of scenes. To overcome these problems the pixel-wise cost is augmented by a local support region (as proposed by the cost filtering methods) or disparity smoothness is imposed (as proposed by the energy minimization techniques).

Analyzing these main solutions to stereo matching we observe that:

- Cost filtering and energy minimization methods work in different ways while addressing noise, occlusion and uncertainty problems. The main assumption of most cost filtering methods (mostly based on bilateral filtering) is that the same color possesses the same disparity, which assumes smoothness primarily in the color space. This works reasonably well, especially when addressing considerable occlusion. However due to the large spatial extend of the bilateral filters required to overcome occlusion this may produce outliers due to violation of the disparity smoothness in the color space. On the other hand, energy minimization approaches produce errors especially in the occluded regions of scenes.
- Both cost filtering and energy minimization approaches under certain constraints can be considered as iterative filters over stereo matching cost volume with local support windows. Vice versa the two approaches under different constraints can be represented as the global energy minimization algorithms.

Manuscript received July 15, 2014; revised October 28, 2014 and January 16, 2015; accepted January 16, 2015. Date of publication January 22, 2015; date of current version February 11, 2015. This work was supported in part by the Spanish Ministry of Science under Project TIN2012-39051, and TIN2013-41751 and in part by the Catalan Project under Grant 2014 SGR 221. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Chang-Su Kim.

The authors are with the Computer Vision Center, Department of Informatics, Universidad Autònoma de Barcelona, Barcelona 08193, Spain (e-mail: mozerov@cvc.uab.es; joost@cvc.uab.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2395820

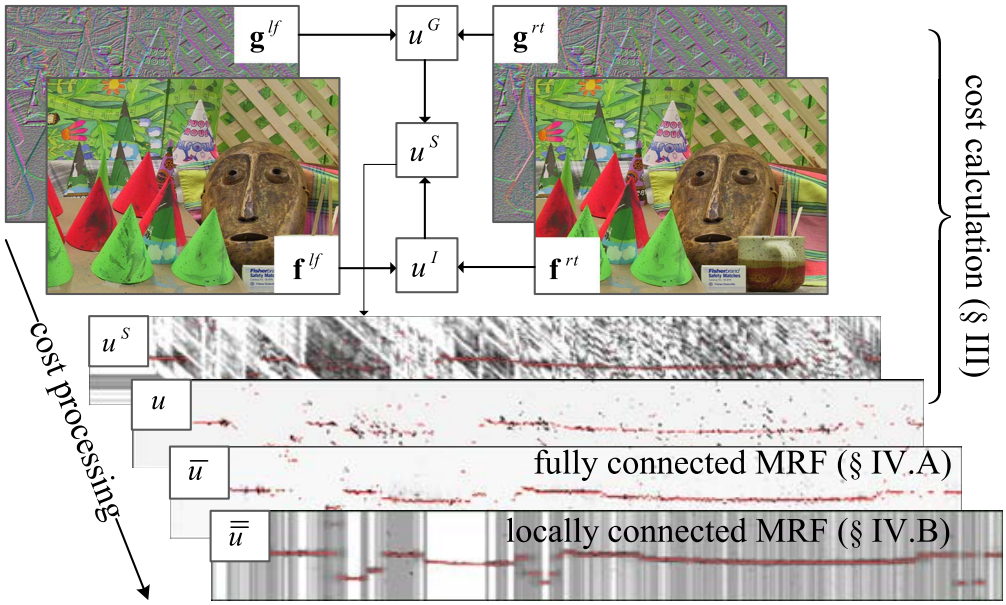


Fig. 1. The scheme of the two-step energy minimization strategy (or the cost-to-marginal processing in the correlation volume).

These observations give rise to the idea that the drawbacks of both approaches can be compensated by unifying cost filtering and energy minimization methods together into a single two-step energy minimization technique. First, we show that cost filtering can be rewritten as an energy minimization problem. The output of this minimization problem yields a more robust unary potential — especially in the presence of occlusion — which we will then apply as the input to a standard energy minimization algorithm to compute the final disparity map.

B. Contribution

We show that for certain constraints there exists exact equivalence between message passing and the cost filtering based on the bilateral filter equation. This allows us to reformulate cost filtering as an energy minimization problem. Based on this observation we propose to perform stereo matching as a two-step global optimization (see Fig. 1). In the first step we solve the energy minimization problem for the fully connected Markov random field (MRF) model. In the second step, the marginal function values are used as the likelihood in a locally connected MRF model. The second step of our global optimization can also be considered as the cost aggregation filtering in the correlation volume [3].

The core of our method is the two step energy minimization approach to stereo matching. We theoretically show that the bilateral filtering is the feasible solution of the functional minimization under specific cost constraints. In addition, we found experimentally that the required step-transform of the initial cost obtains significantly better results than generally used cost transforms.

We propose to perform stereo matching as a two-step minimization procedure: first a fully connected model (FCM) is used for cost filtering after which a locally connected model is applied to compute the final disparity maps. Consecutive application (without post processing) of the FCM and the LCM in our method considerably outperforms the *pure* analogues of the methods which are published on the Middlebury evaluation table (the average rank for the LCM more than 120 and for the bilateral filtering more than 80 in comparison with our 43). The main reason of the obtained result is that the combination of the two models reduces the errors which can arise in the separate processes. The cost calculation part of this scheme will be explained in Section III, and the fully connected MRF processing in Section IV.A, the locally connected MRF processing in Section IV.B. Post processing steps are explained in Section V. Experimental results are presented in Section VI and conclusions are presented in Section VII.

II. Related Work

Cost-volume filtering approaches were summarized in [13]. In this paper authors showed the relationship and similarity between the energy minimization techniques and local filtering methods, but they did not try to merge the approaches. To prevent the computational complexity of the bilateral filtering algorithm a guided filter approach was proposed in the paper. The overall performance of the algorithm was relatively high for the class of cost filter based methods; however it was achieved due to the additional post-processing. In [6] and [7] the classic bilateral filter (they call it cost aggregation) was applied. To speed up the filtering process authors restrict the label space and make the filter support region sparse. We can consider the method in [6] as the best of *pure* cost-volume filtering approaches. And it is opportunity to compare with our two-step process, because they did not apply post-processing.

As a *pure* energy minimization method we can consider two pioneer paper [9], [10]. In [9] they propose to minimize the energy function by an accurate graph cut algorithm called alpha expansion. In [10] the occlusion problem was solved

properly by introducing occlusion penalization term in the energy functional, thus this algorithm performs better than [9].

Generally cost aggregation and global optimization are considered as different techniques. Nevertheless, in two state-of-the-art works authors have proposed methods to combine local preprocessing within a global optimization framework. They combine segmentation based cost aggregation with BP energy minimization [14], [15]. Firstly, an initial image is divided into relatively small segments also called superpixels; and then an aggregated cost is used for the unary potential in the energy minimization problem. There are two problems that one faces implementing this approach: firstly, segmentation is not a trivial problem especially in context of stereo matching; secondly, the assumption that pixels inside the same segment belong to similar disparity values can possibly produce estimation errors, which are impossible to correct at the stage of disparity map post-processing.

In this paper we show that bilateral filtering can be written as an energy minimization problem on a fully connected MRF. A solution to fully connected MRF models is proposed in [16]. Unfortunately, the minimization algorithm, which they propose, is not optimal for the locally connected smoothness term: graph cut or sequential believe propagation provide noticeably better result in terms of minimization accuracy. On the other hand, the conventional stereo matching methods based on energy minimization are defined on a locally connected model and cannot be implemented on the globally connected model due to unsolvable computational complexity. Therefore, in this paper we propose to unify the approaches in one pipeline consecutively: firstly the energy min-marginals of the globally connected model are calculated, and then these values are used as an input cost for the second step of energy optimization with locally connected model.

The work in [17] can be considered as a prototype of our calculation scheme. The initial cost is filtered by a bilateral kernel and then is used in the data term of the energy functional. However, this functional is not directly related with the initial correlation volume as in our method. Furthermore, the mentioned scheme differs considerably from ours. The data term in [17] is not only the filtered cost but the merging result of the disparity map initialization by bilateral filtering, a post processing filter of this map and the filtered cost itself. Another important difference with our approach is the cost transform. We formulate our problem as an energy minimization and the bilateral filter is not a preprocessing step but a part of a message passing procedure under several constraints over cost values. In our experiment we show that the proposed cost transform contributes to the obtained state-of-the-art results.

III. Problem Definition

In the framework of the global approach the stereo matching problem is formulated in terms of energy minimization with the energy function of the following form:

$$E(l) = \sum_{p \in \mathcal{V}} u_p\left(l_p\right) + \sum_{(p,q) \in \mathcal{E}} B_{p,q}\left(l_p, l_q\right) \tag{1}$$

Where set $p \in \mathcal{V}$ corresponds to pixels and set $(p,q) \in \mathcal{E}$ to edges of an image graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$; l_p denotes the label of pixel p which belongs to some discrete set of disparities $l \in L$; $u_p(\cdot)$ defines a unary potential which corresponds to the conventional penalty cost; $B_{p,q}(\cdot,\cdot)$ is a binary potential which defines edge interaction between pixels (p,q).

A choice of the unary potential (cost) in stereo matching is very important sub-problem. There are two kinds of costs: a per-pixel matching dissimilarity measure and non-parametric transforms with a support region such as rank and census [18] or normalized cross correlation [19]. Using a combination of these two costs can significantly improve the result of stereo matching [20]. However, for our fully connected model the cost with a support region is redundant because dissimilarity measure calculation in this case is similar to cost aggregation. Instead we use a linear combination $u_p^S(l)$ of two per-pixel dissimilarities: $u_p^I(l)$ - between left and right stereo images $\mathbf{f}^{lf}$, $\mathbf{f}^{rt}$ and $u_p^G(l)$ - between two corresponding gradient images $\mathbf{g}^{lf}$, $\mathbf{g}^{rt}$:

$$u_p^S(l) = u_p^I(l) + \alpha u_p^G(l), \tag{2}$$

where α is an intrinsic parameter, which depends on the statistics of the data $u_p^I(l)$ and $u_p^G(l)$ (further explained in the Appendix).

Our choice is motivated by research papers in the optical flow area where such a combination demonstrates high robustness, see [21], [22]. The gradient image in our paper is a 6D vector function defined in the same domain as the stereo image itself and this function is calculated as follows:

$$\begin{aligned}&\begin{pmatrix} g_{c,x}(x,y) \\ g_{c,y}(x,y) \end{pmatrix} \\ &= \frac{\beta}{2}\begin{pmatrix} f_c(x+1,y) - f_c(x-1,y) \\ f_c(x,y+1) - f_c(x,y-1) \end{pmatrix} \\ &\quad + \frac{\beta}{4}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} f_c(x+1,y+1) - f_c(x-1,y-1) \\ f_c(x+1,y-1) - f_c(x-1,y+1) \end{pmatrix},\end{aligned} \tag{3}$$

where $c \in \{r,g,b\}$, $(x,y) \equiv \mathbf{x}$, and the digital value of a normalization factor β must equalize the standard deviation of two signals $\mathbf{f}$ and $\mathbf{g}$. The first term of sum in (3) is the conventional gradient operator and the second one is the same operator in the coordinate system rotated by $\pi/2$.

Then we calculate the cost $u_p^I(l)$ as truncated Birchfield-Tomasi dissimilarity [23] between stereo images:

$$\begin{aligned}&u_p^I(l) \\ &= \min\left(\min_{l-0.5 \le d \le l+0.5}\left(\sum_{c \in C}\left|f_c^{lf}(\mathbf{x}_p) - f_c^{rt}(\mathbf{x}_p - d)\right|\right), \tau^I\right),\end{aligned} \tag{4}$$

and the cost $u_p^G(l)$ as truncated dissimilarity between gradient images:

$$u_p^G(l) = \min\left(\left(\sum_{c \in C;\, \mathbf{x} \in \{x,y\}}\left|g_{c,\mathbf{x}}^{lf}(\mathbf{x}_p) - g_{c,\mathbf{x}}^{rt}(\mathbf{x}_p - l)\right|\right), \tau^G\right), \tag{5}$$

where τ^I and τ^G are two truncation thresholds with values 90 and 180 respectively.

IV. Stereo Matching by Two-Step Energy Minimization

In this section we incorporate cost filtering into the energy minimization approach to stereo matching. First we show that under certain constraints cost filtering can be written as an energy minimization problem. After solving this energy minimization problem, we will use the computed marginal as the unary potential of a standard energy minimization problem. The objective is to benefit from the desired behavior of cost filtering approaches in the presence of occlusions. An overview of the method is shown in Fig. 1.

A. MRF Fully Connected Model

In the case of the fully connected MRF model each combination of pixels (p, q) of the image graph $\mathcal{G}$ belongs to the edge set $\mathcal{E}$. And the binary potential for the model must be expressed as:

$$B_{p,q}\left(l_p, l_q\right) = \omega\left(p, q\right)\varphi\left(l_p, l_q\right) \tag{6}$$

Where $\varphi\left(\cdot,\cdot\right)$ is a pairwise penalty which enforces smoothness of the estimated disparity function, $\omega\left(p, q\right)$ is a pairwise weight which defines some reciprocal influence between every node and depends on the distance between pixels (p, q) in the bilateral color-image space. In this paper we consider the classical form, which is usually used in bilateral filtering for cost aggregation methods:

$$\omega\left(p, q\right) = e^{-\frac{|\mathbf{x}_p - \mathbf{x}_q|^2}{2\sigma_x^2} - \frac{|\mathbf{f}_p - \mathbf{f}_q|^2}{2\sigma_f^2}} \tag{7}$$

Where $\mathbf{x}_p$ is a coordinate of a pixel p, $\mathbf{f}_p$ defines the image color vector value of the same pixel, σ_x and σ_f are intrinsic parameters of the bilateral kernel $\omega\left(p, q\right)$.

The problem of energy minimization in (1) can be solved in the framework of belief propagation (BP) approach. The aim is to find marginal function $\bar{u}_p\left(\cdot\right)$ and then estimate disparity in each pixel p by the simple winner take all (WTA) formula:

$$l\left(\mathbf{x}_p\right) = \arg\min_{l \in L} \bar{u}_p\left(l\right) \tag{8}$$

The core operation of BP is passing a message $m_{p\to q}$ from node p to node q for the directed edge $(p \to q) \in \mathcal{E}$. The marginal function then is expressed as follows:

$$\bar{u}_p^t\left(l\right) = u_p\left(l\right) + \sum_{p,q\in\mathcal{E}} m_{p\to q}^{t-1}\left(l\right) \tag{9}$$

Here t is a number of the iteration, and it is supposed that the related iterative message passing procedure is given [24]:

$$m_{p\to q}^t\left(i\right) = \min_{j\in L}\left(\bar{u}_p^t\left(j\right) - m_{p\to q}^{t-1}\left(i\right) + B_{p,q}\left(i, j\right)\right), \tag{10}$$

where indexes $(i, j) \in L$ belong to the discrete disparity domain L.

For the first step of our optimization we consider only the marginal function which is obtained after one iteration of message passing. Thus, taking into account that $m_{p\to q}^0\left(l\right) \equiv 0$ and $\bar{u}_p^0\left(l\right) \equiv u_p\left(l\right)$ finally we rewrite (10) as:

$$m_{p\to q}\left(i\right) = \min_{j\in L}\left(u_p\left(j\right) + B_{p,q}\left(i, j\right)\right) \tag{11}$$

Thus, to follow the BP paradigm for the fully connected model even with one message updating we can face intractable calculation problem: for the single vertex p it is necessary calculating numerous of messages by (11). Fortunately this problem can be simplified if we assume that the following constraints hold:

$$\begin{aligned} &u \in \{0, 1\} \\ &\varphi\left(i, j \in L\right) = 1 - \delta\left(i, j\right) \\ &\omega \le 1 \end{aligned} \tag{12}$$

Where $\delta\left(i, j\right)$ is the Kronecker delta function. Which means that unary potential can be only unity or zero, smoothness potential φ is equivalent to Potts potential, bilateral weights less or equal to unity. Then (11) can be rewritten as:

$$m_{p\to q}\left(i\right) = \omega\left(p, q\right) u_p\left(i\right) \tag{13}$$

Consequently, from (13) and (9) the marginal of the first message passing iteration is calculated:

$$\bar{u}_p\left(l\right) = \sum_{q\in\mathcal{V}} \omega\left(p, q\right) u_q\left(l\right) \tag{14}$$

Thus the initial optimization problem (1) is reformulated as the cost (unary potential) aggregation by the bilateral filtering in (14). To perform the bilateral filtering [25] in this paper we use the fast version proposed in [26]. For this step of our algorithm the digital value of σ_x and σ_f in (14) and (7) are equal to 14 and 1.55 respectively. Increasing the number of iteration at this stage of optimization does not improve the final result, thus we take marginal as the unary potential for the next step of our optimization.

In the paper [16] authors arrive to a similar transform by approximating the exact marginal distribution by the mean field and subsequently maximizing the energy exponent with KL-divergence. Here we show that there exists exact equivalence between message passing in (10) and the bilateral filter equation in (14) under constraint in (12). In addition, their model has not been applied to stereo matching.

To be consistent with the constraints in (12) we have to transform $u_p^S\left(l\right)$ of Eq. 2 into a unary potential $u_p\left(l\right)$ by a step function with a threshold θ. However, a simple threshold transform into $0, 1$ values is not robust and better to approximate the step function by error function:

$$u_p\left(l\right) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\vartheta\frac{u_p^S\left(l\right) - \theta}{\theta}\right)\right), \tag{15}$$

where ϑ and θ are two intrinsic parameters of the method (see Appendix).

The transform in (15) restricts the family of costs used in bilateral filtering of stereo matching. Fortunately, the proposed transform of the initial cost obtains better results when compared to generally used cost transforms, including the standard exponential cost and the untransformed cost form. We show this experimentally in Subsection VI A.

B. MRF Locally Connected Model

For the second step of our global optimization the energy functional can be expressed as the follows:

$$\bar{E}(l) = \sum_{p \in \mathcal{V}} \bar{u}_p(l_p) + \sum_{p,q \in \bar{\mathcal{E}}} \bar{B}_{p,q}(l_p, l_q) \tag{16}$$

Where the unary potential is the output marginal obtained in the previous step of the optimization $\bar{u}_p(l_p)$, see (14). The set of edges $p, q \in \bar{\mathcal{E}}$ includes only local neighborhoods (in a four connected graph model). $\bar{B}_{p,q}(l_p, l_q)$ is a binary potential, which enforces the local smoothness of the disparity map:

$$\bar{B}_{p,q}(l_p, l_q) = \bar{\omega}(p,q)\,\bar{\varphi}\left(|l_p - l_q|\right) \tag{17}$$

The energy minimization (16) with such a functional is well researched in the stereo matching literature [10]–[12]. Usually the function $\bar{\varphi}\left(|l_p - l_q|\right)$ is considered as a truncated linear or quadratic dependency of $|l_p - l_q|$ and the function $\bar{\omega}(p,q)$ divides the set of edges into two or more subsets corresponding to a high or a low image gradient [9]. We follow this paradigm and our binary potential consists of edge weights defined as a lookup table function:

$$\bar{\omega}(p,q) = \begin{cases} \lambda_1 & \textbf{if} \quad \left|\mathbf{f}(\mathbf{x}_p) - \mathbf{f}(\mathbf{x}_q)\right| < \mu_1 \\ \lambda_2 & \textbf{if} \quad \mu_1 \le \left|\mathbf{f}(\mathbf{x}_p) - \mathbf{f}(\mathbf{x}_q)\right| < \mu_2, \\ \lambda_3 & \textbf{if} \quad \left|\mathbf{f}(\mathbf{x}_p) - \mathbf{f}(\mathbf{x}_q)\right| \ge \mu_2 \end{cases} \tag{18}$$

where

$$\left|\mathbf{f}(\mathbf{x}_p) - \mathbf{f}(\mathbf{x}_q)\right| = \sum_{c \in \{r,g,b\}} \left|f_c(\mathbf{x}_p) - f_c(\mathbf{x}_q)\right|$$

The meaning of the weights in (17) is similar to the function $\omega(p,q)$ in (7): strong dependency between nodes when the colors are similar. In our experiments the constants in (18) are chosen: $\lambda_1 = 3.5$, $\lambda_2 = 0.6$, $\lambda_3 = 0.2$, $\mu_1 = 7$, $\mu_2 = 15$. The smoothness multiplier in (17) in our method is chosen as:

$$\bar{\varphi}\left(|l_p - l_q|\right) = \begin{cases} 0 & \textbf{if} \quad |l_p - l_q| = 0 \\ \beta & \textbf{if} \quad |l_p - l_q| = 1 \\ 1 & \textbf{if} \quad |l_p - l_q| > 1 \end{cases} \tag{19}$$

In stereo matching a nonzero value of $|l_p - l_q|$ is usually considered as the discontinuity of the disparity map. However this is not true in general and a value 1 for $|l_p - l_q|$ could be caused by discretization errors. Thus we choose the approximation to the truncated squared in (19) to stress the particularity of the value 1 and in our case the constant β is equal to $\frac{1}{6}$.

In general, the energy minimization problem in (16) is an NP-hard problem, thus approximate minimization algorithms have to be chosen to solve the problem. To make our choice we follow the analysis given in [24] and finally apply the TRW-S algorithm described in [27]. The TRW-S is the method, which was developed in the framework of the belief propagation paradigm. In the case of the functional (16) and the four connected neighborhood graph the iterative message passing procedure is given:

$$m^t_{p \to q}(i) = \min_{j \in L}\left(\frac{1}{2}\overline{\overline{u}}^t_p(j) - m^{t-1}_{p \to q}(i) + \bar{B}_{p,q}(i,j)\right), \tag{20}$$

TABLE I
IMPLEMENTATION OF POST-PROCESSING STEPS FOR THE TOP ALGORITHMS OF THE MIDDLEBURY EVALUATION TABLE. WE USE * TO INDICATE THAT THE AUTHORS OF AN ALGORITHM INCLUDE THIS POST PROCESSING STEP IN THE MAIN PART OF THE METHOD

Algorithm	Av. Rank	LRC check	Local filtering	Outlier suppression	Sub-pixel correction
Our method	8.8	yes	yes	yes	yes
ADCensus [20]	11.5	yes	yes	yes	yes
AdaptingBP [14]	14.7	no	yes*	yes*	no
CoopRegion [28]	15.3	yes*	yes*	no	no
RDP [29]	20.0	no	yes	yes	no
MultiRBF [30]	20.2	yes	yes	yes	no
DoubleBP [17]	20.8	yes*	yes*	yes*	yes

where $\overline{\overline{u}}^t_p(j)$ is the marginal function for this step of optimization, see calculation analogues in (9).

The sequential approach makes the TRW-S algorithm convergent and fast. For the truncated linear and quadratic priors the method usually reaches 0.3% approximation accuracy in a few iterations, thus outperforming the popular graph cut expansion algorithm both in accuracy and speed.

The second order marginal $\overline{\overline{u}}^t_p(j)$ uniquely defines the solution $l(\mathbf{x})$ (disparity map) of the two-step global optimization with (8). Thus this part of the method can be seen as a consecutive cost processing from u^S to $\overline{\overline{u}}$ as is illustrated in Fig. 1.

V. DISPARITY MAP POST-PROCESSING

Post-processing steps are an integral part of stereo methods and indispensable for good results. All state of the art algorithms in stereo use post processing or very similar elements in the body part [14], [17], [20], [28]–[30]. Table I provides an overview of the usage of post-processing steps in the state-of-the art methods in stereo. Several methods (e.g. AdaptingBP [14], DoubleBP [17]) merge some of the post-processing steps with core part of the algorithm. For example, in [17] authors make an iterative optimization over energy function, which strongly depends on of two disparity maps: initial-iterative firstly obtained by bilateral filtering approach and the same disparity map after plain fitting filter based on the mean shift segmentation. These methods are indicated from with a star in Table I. We will quantitatively analyze the impact of the post processing steps in our experiments.

We describe the post processing which we apply in this section. In our case post processing consists of a weighted median filtering a left-to-right disparity map cross-checking, outliers suppression, and a subpixel level correction. We follow recommendation from the papers [15], [17], [20], except for the subpixel level correction, where we apply a weighted median filter and Monte-Carlo strategy.

A. Left-to-Right Disparity Map Cross-Checking (LRC)

All left-right cross-checking algorithms are based on the assumption that the left disparity map $l^{lf}(\mathbf{x})$ is equal to the

unwrapped right disparity map $l^{rt}(\mathbf{x})$, namely:

$$l^{lf}(x,y) = l^{rt}\left(x - l^{lf}(x,y), y\right) \tag{21}$$

Consequently, all pixels where this assumption does not hold are marked as uncertain. Usually, disparity estimation in uncertain pixels is not a trivial problem. We propose to simplify this procedure by:

$$l^{lf}(x,y) = \min\left(l^{rt}\left(x - l^{lf}(x,y), y\right), l^{lf}(x,y)\right) \tag{22}$$

Surprisingly, we found that this simple algorithm works better than the interpolation algorithm proposed in [20].

B. Weighted Median Filter

Weighted median filter usually is a robust extension of the bilateral filter and widely used in stereo matching post-processing, for example in [17] and [31]. The main idea is to accumulate a weighted histogram in every pixel based on a previously estimated disparity map:

$$h_p(i) = \sum_{q\in\mathcal{V}} e^{-\frac{|\mathbf{x}_p-\mathbf{x}_q|^2}{2\bar{\sigma}_x^2} - \frac{|\mathbf{f}_p-\mathbf{f}_q|^2}{2\bar{\sigma}_f^2}} \delta\left(l_q, i\right) l_q \tag{23}$$

Where l_q is a disparity map obtained in the previous step, $\bar{\sigma}_x$ and $\bar{\sigma}_f$ are two intrinsic parameters of the weighted median filter. We take the digital value of $\bar{\sigma}_x$ equaling to 2, and the value of $\bar{\sigma}_f$ depends on a disparity map standard deviation that will be explained in the Appendix. Then the new value of the desired disparity map is estimated as follows:

$$l_p = \underset{i\in L}{\arg\operatorname{med}}\left(h_p(i)\right) \tag{24}$$

The two post-processing procedures: the left-to-right cross-checking and the weighted median filter in our method are implemented consecutively and twice. After which small area outlier suppression and subpixel correction are applied.

C. Small Area Outlier Suppression

To detect outliers the input disparity map must be segmented. In other words, the set of vertex $\mathcal{V}$ of the four connected graph $\bar{\mathcal{G}} = \left(\bar{\mathcal{E}}, \mathcal{V}\right)$ is divided into $k \in K$ subsets in such a way that:

$$\bigcup_{k\in K} \mathcal{V}_k = \mathcal{V}; \quad \bigcap_{k\in K} \mathcal{V}_k = \emptyset; \quad (p,q) \in \bar{\mathcal{E}}_k \Rightarrow \left|l_p - l_q\right| \le \Delta \tag{25}$$

This kind of segmentation can be achieved with strongly connected component analysis [32]. The threshold Δ usually is equal to 0, but we found that 1 produces more reasonable segmentation. Thus we use $\Delta = 1$ and mark all subsets, which satisfy the condition: $|\mathcal{V}_k| < 10^{-3}|\mathcal{V}|$ as the uncertain regions of the corrected disparity map. Next all the pixels in the uncertain regions (but only these pixels) are replaced by weighted median filter (23). Note, that in (23) the weights which correspond to uncertain pixels are zero.

We found that the small area outlier suppression algorithm is less critical in terms of final accuracy. For example, for the Middlebury data set stereo images Tsukuba, Venus and Teddy the accuracy change is negligible. For the image Cones the disparity estimation accuracy slightly improves. However, the computational cost of this part of our method is insignificant and we included this in our full process.

D. Subpixel Accuracy Correction

There are two kind of disparity subpixel enhancement methods:

1) Methods which interpolate the initial cost near the base disparity map by quadratic polynomial, and then to find a local minimum of the disparity in each pixel (see [17]);
2) Methods which extend the disparity (label) domain of the correlation volume by addition several intermediate disparity values divisible to $1/N$ (see [33]).

The first approach is rather coarse and needs an additional filtering. The second, in general, increases the volume of data and the computation time of full method by N.

Here we propose a method, which is a tradeoff between the first and the second subpixel accuracy approach. The main idea is to disturb randomly the disparity map obtained on the previews step of our process by rational extension of the disparity domain according to:

$$\tilde{l}_p = l_p + \frac{r_p}{N}, \tag{26}$$

where $\tilde{l}_p$ is the disturbed disparity map, N - a subpixel accuracy parameter, which is taken in our experiments as 4, $r_p \in \{-N+1, \ldots, 0, \ldots, N-1\}$ some uniformly distributed random variable.

Then the weighted median filter is applied to the disturbed disparity to estimate the final disparity map:

$$h_p(i) = \sum_{q\in\mathcal{V}} e^{-\frac{|\mathbf{x}_p-\mathbf{x}_q|^2}{2\tilde{\sigma}_x^2} - \frac{|l_p-l_q|^2}{2\tilde{\sigma}_l^2} - \frac{\tilde{u}\left(\tilde{l}_p\right)}{\tilde{\sigma}_u}} \delta\left(\tilde{l}_q, i\right) \tilde{l}_q \tag{27}$$

Where the weight term $\frac{|\mathbf{x}_p-\mathbf{x}_q|^2}{2\tilde{\sigma}_x^2}$ defines reciprocal influence between pixels in the image space; the term $\frac{|l_p-l_q|^2}{2\tilde{\sigma}_l^2}$ defines influence between pixels in the disparity space (nearest values more important); the term $\frac{\tilde{u}\left(\tilde{l}_p\right)}{\tilde{\sigma}_u}$ defines importance of this disparity value according to its cost: a low cost dissimilarity is more likely to be the true value of the chosen disparity. The parameters of the filter $\tilde{\sigma}_x$, $\tilde{\sigma}_l$ and $\tilde{\sigma}_u$ are equaling to 5.5, 0.9 and 16 respectively. The dissimilarity cost $\tilde{u}\left(\tilde{l}_p\right)$ is calculated by:

$$\tilde{u}\left(\tilde{l}_p\right) = \sum_{c\in\{r,g,b\}} \left| f_c^{lf}(x_p, y_p) - f_c^{rt}\left(x_p - \tilde{l}_p, y_p\right)\right| \tag{28}$$

Note, the coordinate $x_p - \tilde{l}_p$ is not integer, thus the value $\mathbf{f}^{rt}\left(x_p - \tilde{l}_p, y_p\right)$ in our method is the result of bicubic interpolation of the neighbor pixels.

The one pixel accuracy in stereo matching is the classic and the base criterion to compare the ability of the methods. The subpixel accuracy is not so important for many stereo vision applications e.g.: object recognition, intermediate view

TABLE II
AVERAGE RANK ON THE MIDDLEBURY EVALUATION TABLE AND BAD PIXELS PERCENTAGE WITH THE ERROR THRESHOLD 1 FOR DIFFERENT COST TRANSFORMS. BOTTOM ROWS INCLUDE POST-PROCESSING IN COST TRANSFORM ANALYSIS

Cost transform function	Av. Rank	Tsukuba			Venus			Teddy			Cones		
		nocc	all	disc	nocc	all	disc	nocc	all	disc	nocc	all	disc
Our step function approximation	43.4	0.93	1.28	4.93	0.88	1.49	9.71	6.24	9.40	15.7	2.77	7.66	7.45
Exponential [17]	79.9	1.03	1.79	5.50	1.46	2.32	15.3	7.92	12.3	19.3	3.74	10.2	10.1
Exp. and filter kernel of [17]	87.5	1.07	2.17	5.78	1.88	2.79	18.0	8.26	12.6	20.1	3.85	10.7	10.6
Truncated linear (untransformed)	91.3	1.06	1.99	5.64	2.79	3.69	20.3	8.54	13.0	20.6	4.22	10.9	11.3
Our step function + PPS	8.8	0.86	1.13	4.65	0.11	0.24	1.47	5.61	8.09	13.8	1.67	6.16	4.97
Exponential [17] + PPS	33.9	0.93	1.55	5.03	0.16	0.30	2.08	7.79	10.9	17.3	2.76	8.16	7.44
Exp. and filter kernel of [17] + PPS	38.8	0.98	1.99	5.31	0.15	0.31	1.98	7.45	10.4	16.9	2.85	8.33	7.81
Truncated linear + PPS	43.8	0.89	1.67	4.79	0.19	0.34	2.32	8.33	11.4	18.3	3.05	8.62	8.5

TABLE III
AVERAGE RANK ON THE MIDDLEBURY EVALUATION TABLE AND BAD PIXELS PERCENTAGE WITH THE ERROR THRESHOLD 1 FOR EVALUATION OF THE TWO STEPS OF THE ENERGY MINIMIZATION AND COMPARISON TO SIMILAR METHODS

Algorithm	Av. Rank	Tsukuba			Venus			Teddy			Cones		
		nocc	all	disc	nocc	all	disc	nocc	all	disc	nocc	all	disc
LCM [9]	123	1.94	4.12	9.39	1.79	3.44	8.75	6.50	25.0	24.9	7.70	18.2	15.3
LCM our	97.2	2.11	3.11	7.18	1.49	3.14	10.1	8.68	12.5	18.1	6.28	11.3	10.2
FCM [6]	90.9	2.47	2.71	11.1	0.74	0.97	3.28	8.31	13.8	21.0	3.86	9.47	10.4
FCM our	81.3	2.07	2.33	6.05	1.81	2.45	9.95	9.89	13.0	18.8	6.51	8.23	11.3
FSM and :LSM our	43.4	0.93	1.28	4.93	0.88	1.49	9.71	6.24	9.40	15.7	2.77	7.66	7.45

generation, scene segmentation and understanding. In our paper we do not aim to achieve competitive results in this category of the accuracy. The subpixel accuracy step is applied only because it can improve the result in the base category while being computationally cheap.

VI. EXPERIMENTAL RESULTS

In this section we evaluate our two-step global optimization approach on the Middlebury data set. We start by considering experiments with the different cost transforms and evaluating the advantage of incorporating cost filtering into the energy minimization approach to stereo matching. Next we will compare our results to other state-of-the-art methods and we analyze the impact of the post processing steps.

A. Cost Transform Analysis

Here we analyze the performance of the cost transform in (15) and compare it to other generally used transforms. Usually, in stereo matching the cost can be applied directly in untransformed form (only simple truncation is applied), or with an exponential pre-transformation like in [17]: $u = 1 - \exp(-ku^s)$, where k is an intrinsic scale parameter. We compare these standard transforms to the proposed transform in (15). The result of the experiments is summarized in Table II. The first part of the table shows the impact of different cost transforms on the result of the two step energy minimization procedure. The second part represents the impact to the full procedure including post-processing. For the method described in [17] we also added an experiment which includes the bilateral kernel (for details see [17]). However, we found that this decreased overall performance in our system.

Experiments in Table II show that the proposed transform considerably improves the performance of the algorithm when compared to the other transforms. To explain this result let us consider the MAP model in (1). The general model supposes that a cost value is inversely propositional to the true solution probability. In fact this is not true. Clusters with small or great cost values have almost the same probability inside these clusters. Thus the distribution of costs is more like a Heaviside step function or even error function. As a result, the transform of (15) which approximates a step-function, makes the new cost less sensitive to the energy regularization term. (This is achieved by increasing the number of cost values which most probably belong or not belong to the true solution). Consequently, the energy minimization procedure requires less regularization and avoids over-smoothing.

B. LCM and FCM Analysis

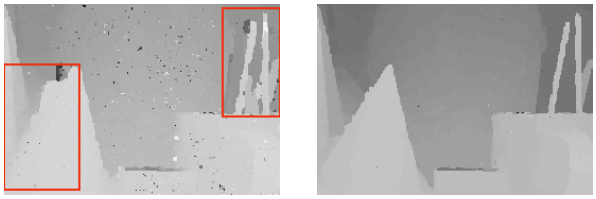
To demonstrate advantage of the proposed combination of the LCM and FCM (without PPS) we compare the combined version with the results of the *pure* LCM the *pure* FCM. In Table III we also refer the method [9] from the Middlebury evaluation table that can be considered as analogues of the *pure* LCM algorithm and the algorithm [6] can be considered as analog of the pure FCM algorithm. Our results for the *pure* LCM the *pure* FCM are slightly better than the analogues in [6] and [9], but it might be due to our robust cost calculation. In Table III the bad pixels error is the percentage of disparity map pixels for which the absolute difference between the ground truth values is more than the error threshold. The base threshold of the Middlebury table is 1 and in our evaluation we hold to this criterion.

As an illustration of the advantage of incorporating cost filtering into the energy minimization we show the results for

TABLE IV

AVERAGE RANK ON THE MIDDLEBURY EVALUATION TABLE AND BAD PIXELS PERCENTAGE WITH THE ERROR THRESHOLD FOR THE TOP 7 ALGORITHMS AND THE ADDITIONAL POSITIONS FOR OUR RESTRICTED METHOD IMPLEMENTATION

Algorithm	Av. Rank	Tsukuba			Venus			Teddy			Cones		
		nocc	all	disc	nocc	all	disc	nocc	all	disc	nocc	all	disc
Our method	8.8	0.86 3	1.13 1	4.65 5	0.11 6	0.24 7	1.47 7	5.61 33	8.09 14	13.8 26	1.67 1	6.16 1	4.97 1
ADCensus [20]	11.5	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 [14]	14.7	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32
CoopRegion [28]	15.3	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01
RDP [29]	20.0	0.97	1.39	5.00	0.21	0.38	1.89	4.84	9.94	12.6	2.53	7.69	7.38
MultiRBF [30]	20.2	1.33	1.56	6.02	0.13	0.17	1.84	5.09	6.36	13.4	2.90	6.76	7.10
DoubleBP [17]	20.8	0.88 7	1.29 6	4.76 8	0.13 10	0.45 42	1.87 19	3.53 10	8.30 16	9.63 6	2.90 42	8.78 51	7.79 33
Our FCM+PPS	51.8	1.51 58	1.97 55	6.00 29	0.49 80	0.65 60	2.16 29	6.92 110	11.6 53	19.2 100	2.57 23	7.14 11	6.74 13
Our LCM+PPS	54.4	1.81 76	2.96 91	6.76 47	0.51 81	0.98 87	5.99 93	7.06 75	10.0 30	15.2 46	2.27 10	7.35 13	6.26 4



(a) (b)

Fig. 2. A part of the resultant disparity map Cones after implementation (a) - the *pure* LCM algorithm; (b) the proposed consecutive *pure* FCM plus *pure* LCM. Red bounding boxes focus on the false estimation regions in the presence of occlusion.

the *pure* LCM implementation and for the consecutive *pure* FCM plus *pure* LCM implementation in Fig. 2(a)-(b). We can see that the proposed combination compared with standard energy minimization approaches especially improves in the presence of occlusion: red bounding boxes in Fig. 2 (a) focus on the false estimation regions in the presence of occlusion that are the result of the implementation of the *pure* LCM. Some of those outliers could be corrected with the post-processing steps, but in general better initial disparity maps result in better final result.

The FCM part of our algorithm performs better in the category all and the LCM part performs better in *nonoccluded*, which means that the FCM part of the algorithm is less sensitive to occlusion. This mentioned trend results in a significant performance gain in the final result. Thus in general our approach performs better in comparison with other methods in the presence of occlusion rather than in nonoccluded regions.

C. Overall Results on Middlebury Data Set

Here we compare our method to state-of-the-art methods on the Middlebury data set. The results are summarized in Table IV and the full version can be seen on the Middlebury evaluation website.[1] Our algorithm outperforms all the other algorithms listed on the Middlebury evaluation table on average rank with the base error threshold 1. We obtain the best reported results for Tsukuba and Cones, and similar to state-of-the-art on Venus. On the Teddy test stereo image our overall rank is 24. The main reason is the abrupt inclination of the scene plain at the bottom of the image plus rich texture in the same region. Here the main assumption that the nearest colors in a neighborhood belong to the same disparity is violated because rich texture adds ambiguities in the cost values domain.

To show the importance of the two-step optimization, we also evaluated the results when excluding either the fully connected model or the locally connected model. In both cases we apply post-processing. The results are provided in Table IV. You can see that in both cases results decrease significantly. Best results are obtained with the fully connected model. This shows that combining both models is important for accurate stereo matching.

Note, the evaluation in Table IV is based on the Middlebury 2003 training dataset. However, in our experiments we also use the Middlebury 2005 and 2006 training datasets and several of resultant disparity maps obtained by our method from those datasets are illustrated in Figs. 3-.4 for visual evaluation. The rest of the results and demo code are available online.[2]

D. Analysis of Postprocessing Steps

Here we perform additional experiments to obtain a better understanding of the importance of the post-processing steps. We first run the system only based on the two-step global optimization approach. After which we add the post processing steps one-by-one. This allows us to show the contribution of the postprocessing steps. It should be noted that most state-of-the-art methods make extensive use of post-processing. However, rarely the relative gain obtained by the post-processing part is explicitly analyzed.

The results of this experiment are shown in Table V. One can see that the post-processing steps are crucial to obtain state-of-the-art results. Especially left-to-right cross-checking and weighted median filter contribute significant to the overall accuracy. We found that small area outlier suppression is of less importance on this data set.

In addition, we provide visual comparison of these steps in Fig. 5 (a)-(e), which confirms the importance of left-to-right

[1] http://vision.middlebury.edu/stereo/eval/

[2] http://www.cvc.uab.es/people/mozerov/Stereo/

TABLE V
AVERAGE RANK ON THE MIDDLEBURY EVALUATION TABLE AND BAD PIXELS PERCENTAGE WITH THE ERROR THRESHOLD 1 FOR DIFFERENT STEPS OF THE PROPOSED ALGORITHM

Step	Av. Rank	Tsukuba			Venus			Teddy			Cones		
		nocc	all	disc	nocc	all	disc	nocc	all	disc	nocc	all	disc
two-step global optimization	43.4	0.93	1.28	4.93	0.88	1.49	9.71	6.24	9.40	15.7	2.77	7.66	7.45
Left-to-right cross-checking	31.9	0.93	1.28	4.93	0.49	0.77	3.88	6.03	8.88	15.1	2.53	7.27	6.89
Weighted median filter	12.2	0.86	1.12	4.67	0.23	0.32	1.85	5.85	8.27	14.5	2.01	6.39	5.45
Small area outlier suppression	12.1	0.87	1.13	4.67	0.23	0.32	1.85	5.79	8.24	14.4	1.79	6.26	5.18
Subpixel accuracy correction	8.8	0.86	1.13	4.65	0.11	0.24	1.47	5.61	8.09	13.8	1.67	6.16	4.97

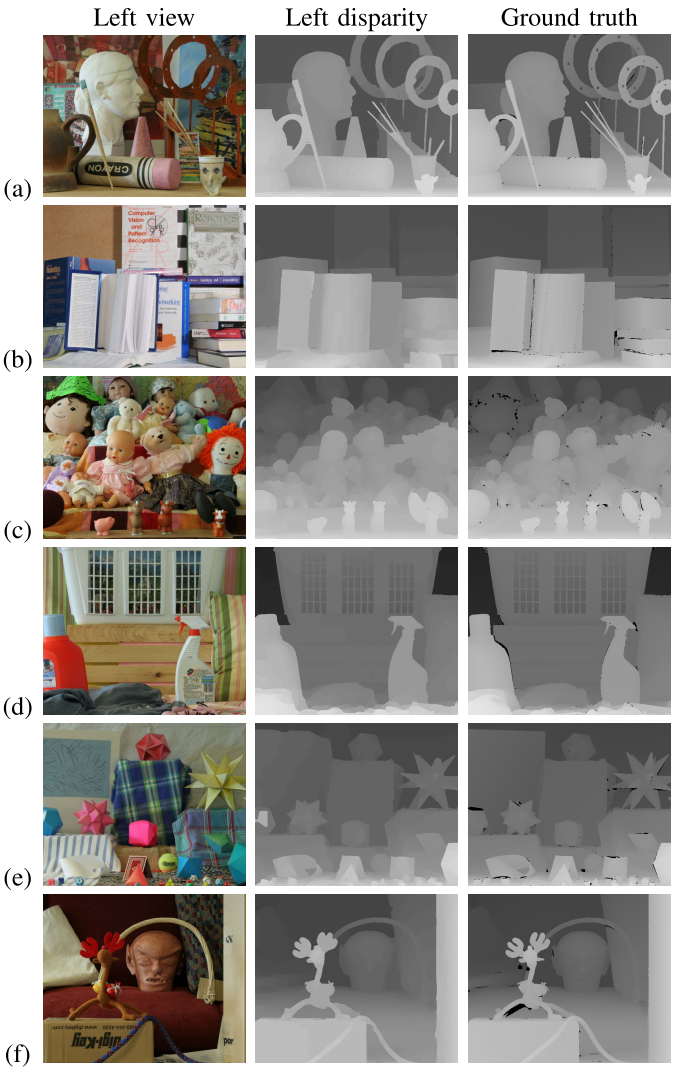


Fig. 3. The Middlebury 2005 datasets (a)- Art; (b) - Books; (c) - Dolls; (d) - Laundry; (e)- Moebius; (f) - Reindeer.

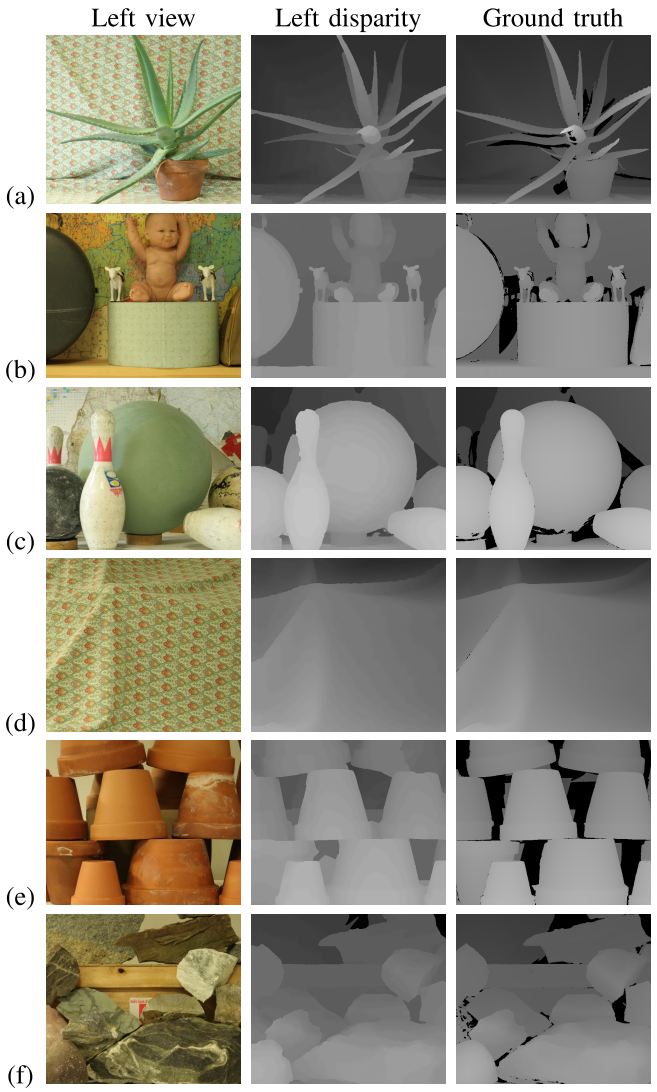


Fig. 4. The Middlebury 2006 datasets (a)- Aloe; (b) - Baby3; (c) - Bowling2; (d) - Cloth1; (e) - Flowerpots; (f) - Rocks1.

cross-checking and weighted median filter post-processing steps.

E. Computational Complexity Analysis

The computational complexity of the method is $O(NL)$. Where N is the number of pixels in the image plain and L is the number of labels. The memory requirement is also proportional to the size of the correlation volume NL and for the Teddy data set the minimal required size of the used RAM is approximately 0.3GB for the float 4-byte calculation version and 0.5GB for the double 8-byte calculation version. The overall running time for the Teddy data set is 20 seconds and 3 seconds for the Tsukuba data set on a computer containing an Intel Core i5-4300U 1.9-GHz CPU (using a single core only) and a 6-GB RAM. In principle, all components of the algorithm can be implemented in a parallel calculation

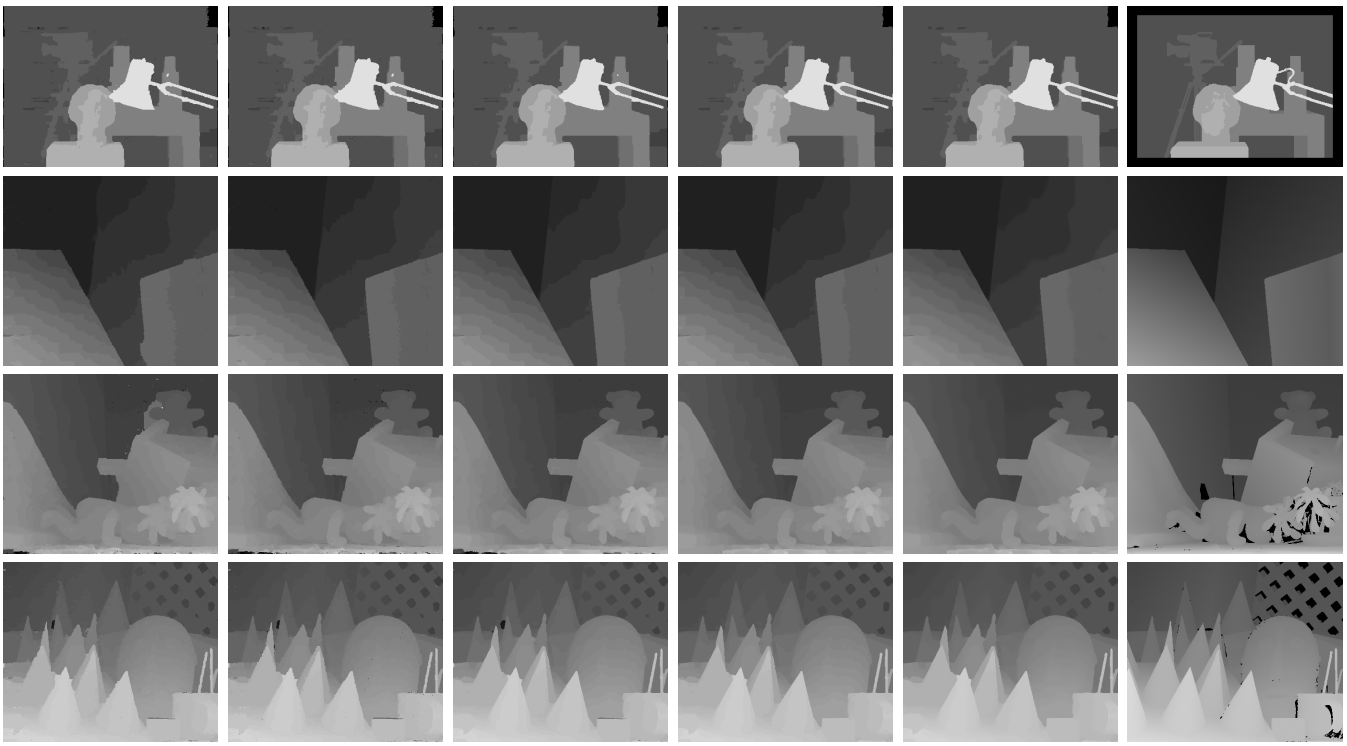


Fig. 5. (a)- two-step global optimization; (b) - left-to-right cross-checking; (c) - weighted median filter; (d) - small area outlier suppression; (e) - subpixel accuracy correction; (g) - ground truth.

scheme, but the bottleneck of the multiple core application is the TRW-S sequential message passing algorithm, which parallel processes are restricted by the number of strings or columns of the image grid.

VII. Conclusions

The main contribution of the presented paper is the two-step global optimization. Its aim is to improve the robustness of energy minimization approach to stereo matching in the presence of occlusions. We do so by incorporating cost filtering, reformulated as a energy minimization problem on a fully connected MRF, into the global optimization. The proposed algorithm achieves state-of-the-arts results and outperforms all known methods with the experiments on the Middlebury stereo datasets. As future work we are interested to extend the developed techniques to image segmentation and optical flow estimation areas.

Appendix

Several intrinsic parameters depend on topology of the processed stereo image; they depend on the statistic of the cost in the correlation volume. Other parameters are functions of the standard deviation of the desired disparity map. Here we describe these intrinsic parameters.

If we omit the binary potential in the energy functional (1), we can call this variable as the cost energy. Then this energy reaches its minimum at the WTA solution l_p, which is completely defined by the cost function u_p. Let us define the value of this energy minimum as ε_u and calculated by:

$$\varepsilon_u(l) = \frac{1}{|\mathcal{V}|}\sum_{p\in\mathcal{V}} u_p\left(l_p\right) \tag{29}$$

It is also convenient to calculate what we call the relative deviation of the cost energy $\hat{\varepsilon}_u$:

$$\hat{\varepsilon}_u(l) = \frac{1}{L\,|\mathcal{V}|}\sum_{l\in L;\, p\in\mathcal{V}} \left|u_p(l) - u_p\left(l_p\right)\right| \tag{30}$$

Consequently, the parameter α in (2) can be calculated:

$$\alpha = 3.5\frac{\varepsilon^I}{\varepsilon^G}, \tag{31}$$

where ε^I is the cost energy in (29) with the cost u^I; and ε^G is the cost energy with the cost u^G. Thus the factor α balanced two costs in one superposition cost u^S.

Other method parameters in (15) θ and ϑ also depends on the cost energy and the relative deviation of the cost energy:

$$\begin{aligned}\theta &= \varepsilon^S\\ \vartheta &= 9.5\ 10^{-4}\left(\hat{\varepsilon}^S - \varepsilon^S\right)^2,\end{aligned} \tag{32}$$

where ε^S and $\hat{\varepsilon}^S$ are the cost energy and its relative deviation with the cost u^S.

Another useful characteristic is the standard deviation of the disparity map $v\left(l_p\right)$. If the value of v is small, most probable that the occlusion region is not important. Thus, the left-to-right cross-checking step of algorithm cannot improve the accuracy of estimation. On the other hand, the cross-checking step enforce to implement the base and the most

computational part of our algorithm twice: for the left and for the right disparity map. Consequently, for the value of $\nu < 3.5$ we omit the cross-checking procedure. For the same reason the subpixel accuracy level is not recommended for disparity maps with the same small value of the standard deviation.

Surprisingly, the optimal value of the parameter $\bar{\sigma}_f$ of the weighted median filter also depends on this characteristic:

$$\bar{\sigma}_f = 5\left(\frac{\nu\left(l_p\right)}{L}\right)^2 \tag{33}$$

For the coarse approximation this value can be put as 0.2.

REFERENCES

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

[2] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, Aug. 2003.

[3] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 650–656, Apr. 2006.

[4] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda, "Near real-time stereo based on effective cost aggregation," in *Proc. 19th ICPR*, Dec. 2008, pp. 1–4.

[5] Q. Yang, "A non-local cost aggregation method for stereo matching," in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 1402–1409.

[6] D. Min, J. Lu, and M. N. Do, "A revisit to cost aggregation in stereo matching: How far can we reduce its computational redundancy?" in *Proc. IEEE ICCV*, Nov. 2011, pp. 1567–1574.

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

[8] A. T. Ihler, J. W. Fisher, III, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *J. Mach. Learn. Res.*, vol. 6, pp. 905–936, Dec. 2005.

[9] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[10] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. 8th IEEE ICCV*, Jul. 2001, pp. 508–515.

[11] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 787–800, Jul. 2003.

[12] S. Roy and I. J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem," in *Proc. 6th ICCV*, Jan. 1998, pp. 492–499.

[13] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Proc. IEEE Conf. CVPR*, Jun. 2011, pp. 839–846.

[14] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proc. 18th ICPR*, 2006, pp. 15–18.

[15] C. Vogel, K. Schindler, and S. Roth, "3D scene flow estimation with a rigid motion prior," in *Proc. ICCV*, 2011, pp. 1291–1298.

[16] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Advances in Neural Information Processing Systems 24*. Red Hook, NY, USA: Curran Associates, 2011.

[17] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 492–504, Mar. 2009.

[18] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. 3rd ECCV*, 1994, pp. 151–158.

[19] J. P. Lewis, "Fast template matching, vision interface," in *Proc. Canadian Image Process. Pattern Recognit. Soc.*, vol. 95. Quebec City, QC, Canada, May 1995, pp. 120–123.

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

[21] L. Xu, J. Jiaya, and Y. Matsushita, "Motion detail preserving optical flow estimation," in *Proc. CVPR*, 2010, pp. 18–23.

[22] M. G. Mozerov, "Constrained optical flow estimation as a matching problem," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2044–2055, May 2013.

[23] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *Int. J. Comput. Vis.*, vol. 35, no. 3, pp. 269–293, 1999.

[24] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields," in *Proc. 9th ECCV*, 2006, pp. 16–29.

[25] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 16th Int. Conf. CVPR*, Jan. 1998, pp. 839–846.

[26] A. Adams, J. Baek, and M. A. Davis, "Fast high-dimensional filtering using the permutohedral lattice," *Comput. Graph. Forum*, vol. 29, no. 2, pp. 753–762, 2010.

[27] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.

[28] Z.-F. Wang and Z.-G. Zheng, "A region based stereo matching algorithm using cooperative optimization," in *Proc. IEEE Conf. CVPR*, Jun. 2008, pp. 1–8.

[29] X. Sun, X. Mei, S. Jiao, M. Zhou, and H. Wang, "Stereo matching with reliable disparity propagation," in *Proc. Int. Conf. 3DIMPVT*, May 2011, pp. 132–139.

[30] X. Zhou and P. Boulanger, "New eye contact correction using radial basis function for wide baseline videoconference system," in *Proc. 13th Pacific-Rim Conf. Multimedia (PCM)*, 2012, pp. 68–79.

[31] M. Mozerov, J. Gonzalez, X. Roca, and J. J. Villanueva, "Trinocular stereo matching with composite disparity space image," in *Proc. 16th IEEE ICIP*, Nov. 2009, pp. 2089–2092.

[32] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA, USA: MIT Press, 2001.

[33] Y. Mizukami, K. Okada, A. Nomura, S. Nakanishi, and K. Tadamura, "Sub-pixel disparity search for binocular stereo vision," in *Proc. 21st ICPR*, Nov. 2011, pp. 364–367.

Mikhail G. Mozerov is currently a Senior Scientist with the Computer Vision Center, Universidad Autònoma de Barcelona, Barcelona, Spain. He received the M.S. degree in physics from Moscow State University, Moscow, Russia, in 1982, and the Ph.D. degree in digital image processing from the Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow, in 1995. From 2006 to 2010, he was a Ramon y Cajal Fellow with the Universidad Autònoma de Barcelona. His research interests include signal and image processing, stereo and optical flow, pattern recognition, and digital holography.

Joost van de Weijer is currently a Senior Scientist with the Computer Vision Center, Universidad Autònoma de Barcelona, Barcelona, Spain, and a member of the LAMP Team. He received the M.Sc. degree in applied physics from the Delft University of Technology, Delft, The Netherlands, in 1998, and the Ph.D. degree from the University of Amsterdam, Amsterdam, The Netherlands, in 2005. From 2005 to 2007, he was a Marie Curie Intra-European Fellow with the LEAR Team, INRIA Rhone-Alpes, Grenoble, France. From 2008 to 2012, he was a Ramon y Cajal Fellow with the Universidad Autònoma de Barcelona. His main research is in the usage of color information in computer vision application.