RESEARCH

Backward compatible HDR stereo matching: a hybrid tone-mapping-based framework

Tara Akhavan[*] and Hannes Kaufmann

Abstract

Stereo matching under complex circumstances, such as low-textured areas and high dynamic range (HDR) scenes, is an ill-posed problem. In this paper, we introduce a stereo matching approach for real-world HDR scenes which is backward compatible to conventional stereo matchers. For this purpose, (1) we compare and evaluate the tone-mapped disparity maps to find the most suitable tone-mapping approach for the stereo matching purpose. Thereof, (2) we introduce a combining graph-cut based framework for effectively fusing the tone-mapped disparity maps obtained from different tone-mapped input image pairs. And finally, (3) we generate reference ground truth disparity maps for our evaluation using the original HDR images and a customized stereo matching method for HDR inputs. Our experiments show that, combining the most effective features of tone-mapped disparity maps, an improved version of the disparity is achieved. Not only our results reduce the low dynamic range (LDR), conventional disparity errors by the factor of 3, but also outperform the other well-known tone-mapped disparities by providing the closest results to the original HDR disparity maps.

Keywords: Tone mapping, Stereo matching, Disparity map, Low dynamic range, High dynamic range, Markov random field, Graph cut

1 Introduction

High dynamic range (HDR) images provide greater detail and larger brightness levels than conventional low dynamic range (LDR) ones. Even though capturing and displaying HDR images has been widely explored during the last two decades [1–5], they are not broadly used in image processing and computer vision applications such as stereo matching and 3D reconstruction, segmentation, alpha matting, and face recognition. Working in HDR space can lead to better results by using more detailed brightness information. The pixel values in HDR space are calculated using an estimated camera response function to fuse the multiple photographs into a single, high dynamic range radiance map whose pixel values are proportional to the true radiance values in the scene. Differently exposed images are used to estimate the camera response function [1]. It is not hard to predict that working in HDR space provides more informative disparities. This is especially true in challenging lighting conditions

and low-textured areas. However, two important challenges when switching calculations from LDR to HDR domain are *lack of data* and *backward compatibility*.

To generate a backward compatible solution to conventional stereo matchers, we use tone mapping operations (TMO) to compress the dynamic range into conventional range while preserving details of an HDR image. After comparing disparity maps achieved from different TMOs, we propose a fusion framework for achieving more informative disparity maps from tone-mapped HDR image pairs. In order to estimate the corresponding ground truth disparity maps, we make use of recently introduced HDR stereo images and an implementation of a customized stereo matching approach [6]. Based on these estimated disparities, we perform an objective evaluation of stereo matching approaches.

Why do we need a backward compatible HDR stereo matcher? Most of the available stereo matching software approaches are hard-coded to work with 8-bit input images, including the top-ranked methods in Middlebury [7], although a few methods including some hardware solutions implemented on microprocessors have the

*Correspondence: takhavan@ims.tuwien.ac.at
Institute of Software Technology and Interactive Systems, Vienna University of Technology, Favoritenstrasse 9-11, Vienna, Austria

advantage of running on floating point data [8]. Furthermore, most of the available stereo matching codes assume a fixed range of intensities for both left and right images while an HDR image pair most probably have different maximum luminance values for each image due to angular differences. See Section 5.1 for information regarding the maximum luminance values in the left and right images of our data set.

The described modifications are not hard to make in code level but will cause reviewing and re-factoring available stereo matching implementations (see Section 5.3 for an example). Stereo matching is an old research and industrial field, it is important for new stereo data sets to be runnable on conventional approaches. Therefore, we use tone-mapped image pairs with the standard interface to legacy stereo matchers and propose an algorithm to optimize the disparity maps to achieve the closest results to HDR stereo matching.

Even though we captured the HDR stereo data, we only used them for our ground truth disparity calculation and used HDR *tone-mapped* data in our proposed framework. The first reason for using tone-mapped images (not the original HDR image pairs) is described in the previous paragraph regarding backward compatibility. The second reason is, although our HDR stereo matching approach outperforms the tone-mapped and LDR approaches, it is not fully automatic. We used a manual in-painting post processing to fill-in some of the holes in the disparity maps as shown in Fig. 2. Manual post processing and in-painting of disparity maps for the purpose of ground truth generation (evaluation) is a common approach [9, 10] but usually not a scalable or real-time method. We used the original HDR disparity maps just as a reference for our evaluation method for the same mentioned reasons. The manual post processing method will be discussed in more detail in Section 5.3.

In this paper, we present a graph-cut based disparity map fusion framework using different tone-mapped stereo matching results in order to take into account the best features for stereo matching from several different TMOs. Stereo matching in HDR scenes introduces new challenges to the state-of-the-art matching. Computing the disparity on the tone-mapped image pairs is an approach to solve these challenges, but not so many tone-mapping operators are suitable to be applied on more than one frame or image while keeping the consistency of the images or frames. This problem has recently been addressed in video tone-mapping [11].

The remainder of the paper is structured as follows. The related work and background are discussed in the next section. We compare the disparity results obtained from different tone-mapped image pairs with a focus on edge-aware filtering based TMOs in Section 3. Our proposed framework to combine several computed disparity maps

is introduced in Section 4. Finally, Section 5 debates our experimental results, evaluation, and discussions. This Section discusses the HDR stereo image pairs as well as our reference disparity map generation approach used for the quantitative evaluation.

2 Related work

Even though there is a considerable amount of literature on the state-of-the-art in each of the HDR and stereo matching fields, not much work has been done on joining the two. Several approaches have been presented for constructing an HDR image from two differently exposed LDR stereo images by calculating the depth information of the scene [12–17]. More recently, Batz et al. [18] and Orozco et al. [19] proposed interesting approaches for HDR video reconstruction using depth information. Orozco et al. introduced a patch match-based method to generate 3D HDR video sequences using available hardware. The main goal in the mentioned articles is to generate better quality HDR image/video while our main focus is to use the available HDR content to achieve more informative disparity maps. Few approaches have been introduced for subjectively comparing tone-mapped stereo images with the focus on stereoscopic data generation [20] or disparity map calculation [6]. Recently, Aydin et al. [21] evaluated some of the TMOs based on edge-aware filters for HDR video tone-mapping taking into account the visual artifacts and temporal coherency in the tone-mapped video. The authors also introduced a faster and more efficient filter for high motion scenes. The key contribution of this filter is for achieving temporal stability without ghosting on high motion videos. Our work focuses on still stereo image pairs and does not contain motion.

Combining multiple TMOs. Combining results achieved from different methods is a common approach in image processing and computer vision research [22, 23]. More specifically, fusing several TMOs to achieve better quality images is addressed in [24]. The idea behind this fusion is that each TMO works better in a special image region and under some specific conditions and the best output can be calculated taking into account the suitable TMO for each image region. Mai et al. [20] explored that HDR tone-mapping can significantly enhance perceptual quality of 3D images. We go one step further and use tone-mapped stereo pairs to obtain better disparity information.

Although, our first objective was to report *the best* TMO regarding stereo matching, the outcome of our experiments is consistent to the result of many evaluations that have been done on TMOs [25–27]; there is no single TMO which performs the best in all conditions. Therefore, we combined the disparity maps from the tone-mapped images to maximize the quality of the disparity map taking into account the strong points of each TMO.

Combining multiple depth maps. Combining several depth maps achieved from different view points is studied by Schuon et al. [28]. Combining range information to generate a more accurate result is a well-known approach in the 3D society [29]. Another successful example was introduced by Izadi et al. [30] to fuse a sequence of depth maps generated by a kinect camera. We use information from several tone-mapped disparity maps and combine them to provide backward-compatible stereo matching results for HDR scenes.

Many different methods can be used for combining results. Some of the simple ways are to calculate the average, weighted average, or median of the candidate results. In most applications, using simple combination methods without taking into account any prior or statistical knowledge of candidate results does not achieve the best outcome. Akhavan et al. used a machine learning-based approach to combine results of color constancy in [31]. Markov random field (MRF)-based method was described and used is [29] to combine range information in 2005. In [23], a fuzzy integral method was introduced as a combination method which considers the dependencies between the candidate results. In this paper, we use Markov random field and more specifically graph-cuts to combine the disparity maps. Using MRF is proved to be effective in labeling problems such as stereo matching [32], since the output values of the matching algorithms are *cost* values that are suitable targets for energy minimization algorithms. In Section 5.5, the combination outcomes using average, median, and our method are compared.

3 Tone-mapped stereo matching

Tone-mapping is the approach of compressing the dynamic range into conventional range while preserving details of an HDR image. Therefore, tone-mapped images of HDR scenes contain more information compared to conventional LDR images. We compare and evaluate some TMOs specifically for the stereo matching purpose and provide an objective evaluation comparison to HDR disparity maps. By doing this evaluation, we find the most effective TMO for stereo matching which achieves the closest results to the HDR method. This will enable us to achieve as close disparity information as possible to disparities computed on HDR images without the need to customize stereo matching codes. From the stereo matching perspective, the tone-mapped image is treated like an LDR image and can be used easily as an input to any stereo matching method. Therefore, we call the stereo matching using tone-mapped image pairs, the backward-compatible stereo matching approach.

Choosing among so many available TMOs is a challenging task. It is obvious that comparing all of the available TMOs is not possible. In most of the evaluations of TMOs, a subset of five to ten different methods are chosen for comparison [25–27]. TMO's main purpose is for displaying HDR images on conventional devices, therefore most of the comparisons on TMOs used subjective evaluation methods. We evaluate two sets of TMOs. (1) Some highly ranked TMOs which are reported among the most effective ones in the evaluations [25–27] (Section 3.1), and (2) TMOs that are based on an edge-aware filtering approaches (Section 3.2). A big challenge in using TMOs for stereo image pairs is to keep the consistency between two image frames. This challenge is very well-known in video tone mapping [11]. Using TMOs which have less dependency to statistical information from an image or a frame, is more consistent to be used for stereo, multi, or video frames. According to our experiments, edge-aware filtering approaches give better results considering the consistency among left and right images. A reason for this is that these approaches do not try to estimate any curve depending on the image information or luminance range. But they divide the HDR images into two layers, keep the detail layer, and compress just the base layer.

Akhavan et al. [6] subjectively compared the disparity maps obtained from two of the TMOs with LDR and HDR disparities. In this paper, a broader range of TMOs are studied. Moreover, the effect of edge-aware TMOs are taken into account for a subjective as well as an objective comparison which is followed by proposing a new combining method for disparity estimation. To maximize the effect of our combining method, a diverse selection of TMOs is suggested. Since we aim to combine the best features for stereo matching (among TMOs), the bigger and more diverse our feature set is, the more effective the results will be. Due to the fact that there are many TMOs available, it is easy to choose several and apply our fast combined solution to achieve as informative disparity maps as HDR disparities.

3.1 Highly ranked TMOs

According to most of the comparisons on TMOs mentioned before, Reinhard TMO [33], Fattal TMO [34], and Drago TMO [2] are among the most effective ones. *Reinhard TMO* is a global TMO which took the inspiration from traditional wet-film photography techniques. *Fattal TMO* represents a local TMO which is based on gradient domain operators. *Drago TMO* is an example of a global TMO which is extending the logarithmic response curves to perform on a wider dynamic range. In the following subsections, we compare and discuss these TMOs from the stereo matching point of view.

3.2 Edge-aware TMOs

Most local tone-mapping operators use a decomposition of the image into different layers or scales to reduce the

contrast differently for each scale, and the final image is a recomposition of the various scales after contrast reduction [35]. A very common way to decompose the image into layers is using the edge-aware filtering approaches. Using edge-aware filters for tone-mapping was first introduced by Durand in [35] using bilateral filter and was evaluated as one of the best TMOs. Therefore, in this paper, we implemented and compared three TMOs using some well-known edge-aware filters such as guided filter [36], Farbman filtering approach [37], and domain transform filter [38]. Even though all of the mentioned filtering approaches introduced tone-mapping as one of their applications, the analysis of tone-mapping in these studies has been brief, since the main contributions of these works lie elsewhere. No thorough comparison of these edge-aware tone-mapping operations has been published (except for Aydin et al. [21], mentioned in Section 2 which focuses on high motion videos).

Our experiments show that edge-aware TMOs achieve more discriminative disparity maps since they are more robust to lighting changes between the left and right images.

Durand TMO: Based on the idea of an image being consisted of a high spatial frequency (LDR) and a low frequency (HDR).

Gastal, He, and Farbman TMOs: Based on the same idea as Durand TMO but using different edge-aware filtering approaches.

The tone-mapped disparity maps are calculated using the cost-volume filtering stereo matching [39] on every tone-mapped image pair. The TMOs which are used in this experiment are as follows: Reinhard, Durand, Fattal, Gastal, Drago, He, and Farbman TMOs.

Our main contribution is the introduction of the combination approach to use all the different characteristics of TMOs into account. The TMOs which we used in our benchmark are interchangeable.

TMO parameters: Most TMOs have tunable parameters. We used the default parameters suggested by the authors in most of the cases for TMOs. Here, we list the parameter values which we modified to tune the TMOs for tone-mapped disparity calculation according to our experiments. For more details of the parameter definitions, please refer to the references.

- Guided filter (He TMO [36]): $\{r = 9, \epsilon = 0.0001\}$
- WLS filter (Farbman TMO [37]): $\{\alpha = 1.2, \lambda = 1\}$
- Domain transform filter (gastal TMO [38]): $\{\sigma_s = 60, \sigma_r = 0.33\}$

We use a state-of-the-art local stereo matching technique based on cost-volume filtering [39] for all of our disparity estimations. The matching cost calculation based on the color intensity values (Ic_i) and luma gradient information ($\nabla_x I_i$) is formulated in Eq. 1 [39]. The seven different tone-mapped image pairs are used as different inputs to our stereo matcher.

$$C_{i,d} = \alpha \cdot \min \left[||Ic_i - Ic'_{i-d}||, \tau_1 \right] + (1 - \alpha) \cdot \min \left[||\nabla_x I_i - \nabla_x I'_{i-d}||, \tau_2 \right]. \tag{1}$$

The cost-volume entry, $C_{i,d}$, determines how well a pixel i in the left image matches the same pixel in the right image shifted by vector (disparity) d in the x direction. Here, ∇_x is the gradient operator in the x direction. For weighting the color and gradient information, α is used, and τ values are truncation values. We apply this approach on different tone-mapped stereo inputs in the rest of the paper. The minimum cost disparity value (among the d disparities) is then estimated for each pixel i to be saved in f_i as the final disparity as in Eq. 2.

$$f_i = \underset{d}{\arg\min} \, C_{i,d}. \tag{2}$$

4 Combined tone-mapping approach for disparity map estimation

The discontinuities in a disparity map often co-occur with the color or brightness changes in the associated camera image [29] as illustrated in Fig. 3. Therefore, using tone-mapped image pairs which contain more accurate information of the brightness and color than LDR images helps the disparity estimation. Our results, discussed in Section 5, show that some of the TMOs provide better data for stereo matching, but here we propose an approach to take into account all the positive features of the different tone-mapped images. We tie together the available information from different tone-mapped image pairs. Each TMO is based on some specific image features. The probability distribution based on each disparity map (obtained from different TMOs) provides a practical platform to effectively obtain the most probable disparity value for each pixel.

4.1 Combination method

Markov random fields (MRFs) are being widely used in computer vision during the last two decades because of their enormous power in modeling the visual perception problems [40]. We model our combining problem using a pairwise MRF by defining the measurement and smoothness potentials.

To our best knowledge, this is the first time tone-mapped disparities are being combined to provide more discriminative disparity information in HDR, low-textured scenes. We apply a graphical model to the problem of fusing several disparity maps using a Markov random field approach for integrating the disparities. We propose a modified version of the MRF approach used in

[29] for integrating our seven different tone-mapped disparity maps. They used a maximum a-posteriori (MAP) which can be mapped to the usage of graph cuts to solve an energy minimization problem (see Eq. 8). Our MRF formulation of the problem works with seven layers of information, one layer per TMO disparity.

The target disparity value y is estimated using the prior seven disparity information $Z = \{z_1, z_2, \ldots, z_7\}$ and reference guided image (here, left image) x from the likelihood function $p(y|Z,x)$. Since the stereo images are available in high resolution, this insight is used to enhance the accuracy of the disparity estimation. We used the left image as a guided image in our approach. The MRF is defined in the form of

$$p(y|Z,x) = \frac{1}{C} \exp\left(-\frac{1}{2}(\psi + \phi)\right), \tag{3}$$

where C is a normalization factor. The disparity measurement potential ψ and the disparity smoothness potential ϕ are calculated as follows. The measurement potential is based on a quadratic distance between the estimated disparity value and the seven measured ones. The set of indexes for which different disparity values are available is shown by L. A constant weight of K can be placed on the depth measurements. In our calculations, we used $K = 1$.

$$\psi = \sum_{i \in L} K(y - z_i)^2, L = 1, 2, \ldots, 7. \tag{4}$$

The neighboring nodes to pixel i are considered in $N(i)$, and ϕ calculates the weighted quadratic distance between neighboring disparity information. Various numbers of neighbors can be used depending on the purpose of the application. We used values from eight neighboring pixels. The weighting values w_{ij} determine the correspondence between two adjacent pixels, using the constant c as a penalty for smoothing the edges in the image.

$$\phi = \sum_i \sum_{j \in N(i)} w_{ij}(y_i - y_j)^2. \tag{5}$$

The weights w_{ij} are calculated from the guidance reference image x, which in our case is the left view image since we calculate the disparity for the left view.

$$w_{ij} = \exp(-c \quad u_{ij}). \tag{6}$$

$$u_{ij} = ||x_i - x_j||^2. \tag{7}$$

Now that the MRF model is defined, there are many ways for solving the optimization problem. This problem can be solved using MAP or energy minimization (see Eq. 8). In [29], the conjugate gradient is used for solving the MRF. Here, we minimize the energy with the help of the graph cuts using α-expansion moves [41] as formulated in Eq. 10, since one of the classical usages of energy minimization is for assigning labels (here, disparity values) to pixels. Minimizing the energy/cost has gained a lot of popularity especially in solving low-level vision problems such as stereo matching.

$$E(y|Z,x) = -\log p(y|Z,x). \tag{8}$$

Equations 8, 9, and 10 show the relation between the energy and likelihood and their optimization approaches, where E is the energy function [40].

$$\hat{y} = \underset{y}{\operatorname{argmax}} \left\{ \exp \sum_p \in (p(y|z)) \right\}. \tag{9}$$

$$\hat{y} = \underset{y}{\operatorname{argmin}} \left\{ -\sum_p \ln (p(y|z)) \right\}. \tag{10}$$

5 Evaluation and experimental results

An assumption underlying much of image processing and computer vision is that image intensity values are proportional to scene radiance. This assumption breaks down especially in saturated regions, which can impact many vision tasks including stereo matching. There is usually an unknown, nonlinear mapping of scene radiance to image intensity values caused by several nonlinear functions that occur in the imaging process [1]. In particular, real-world scenes contain a high range of luminance levels which cause over- and under-exposed regions in the captured image. Accordingly, computing disparity in the mentioned regions is very challenging. Some examples of such scenes are shown in Figs. 1 and 2. Currently, there is a trend towards HDR imaging, which fuses several images, acquired with different exposures, into a single, HDR radiance map whose pixel values are proportional to true scene radiance [42].

5.1 Stereo image pairs

The well-known Middlebury stereo data set [7, 43] is not sufficient for our matching purpose, even though it contains multi-exposure views. Most of the scenes captured in Middlebury could be categorized as normal LDR indoor scenes of 10^2 cd/m^2 since they do not contain large brightness differences. These scenes are categorized in Table 1 as indoor scenes which most of the time do not need HDR capturing. On the other hand, sunny outdoor scenes can get as bright as 10^5 cd/m^2 which is a wide dynamic range of brightness. We constructed bright scenes of 10^3 cd/m^2 in our laboratory (using multiple illuminations in the scene) which contain just enough brightness to generate over-/under-exposed areas in conventional photography. These types of scenes are the ones who are called HDR scenes and need HDR capturing methods to cover the whole range of brightness in the image. We generated a data set of HDR scenes including highly exposed regions and low-textured areas for our experimental process (see Table 1).
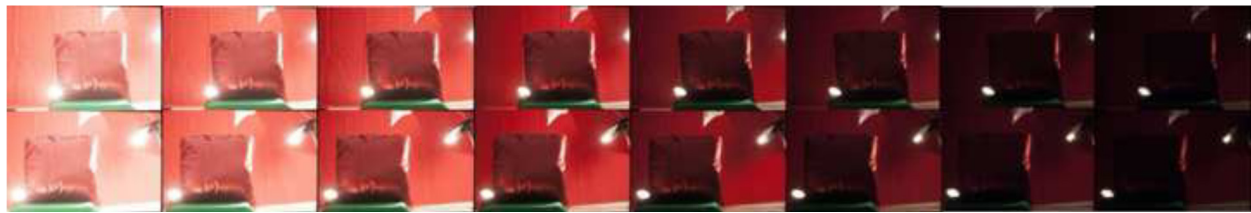
Fig. 1 A complete LDR sample from the multi-exposed stereo data. *First row*: left view, *second row*: right view. Exposure times from left to right: 1/15, 1/30, 1/60, 1/125, 1/250, 1/500, 1/1000, and 1/2000 s. Baseline: 150 mm

We captured normal (LDR) images in eight different exposures for each view. Figure 1 is a detailed example of our full exposure stack of images. It shows the eight exposures for both views which are used for creating HDR images from a scene containing low-textured regions captured under a non-uniform luminance in the range of 1 to 8000 cd/m^2. Table 2 illustrates the maximum luminance values per image for all of the scenes in our data set. The values are calculated from the original HDR images. As shown in the table, in all five scenes, the right images contain a higher brightness range since our lamps were located on the right side of the scene.

Our data set comprises two different baselines of 75 and 150 mm between the two stereo views. For both baselines, the images were separately rectified. The HDR images

were directly generated from raw image files of the eight LDR exposures following the approach of Debevec et al. [1]. Five different samples of our stereo LDR-HDR data set are shown in the first two rows of Fig. 2.

5.2 LDR disparity maps
In order to have a fair comparison of the LDR disparity maps with the tone-mapped and HDR competitives (see Fig. 4), we used the most informative disparity map among the eight captured exposures for each of the images in our data set. As shown in Fig. 1, we captured the left and right images in the following eight exposures: 1/15, 1/30, 1/60, 1/125, 1/250, 1/500, 1/1000, and 1/2000 s. As expected, the middle exposures are the ones which capture most of the information and the others
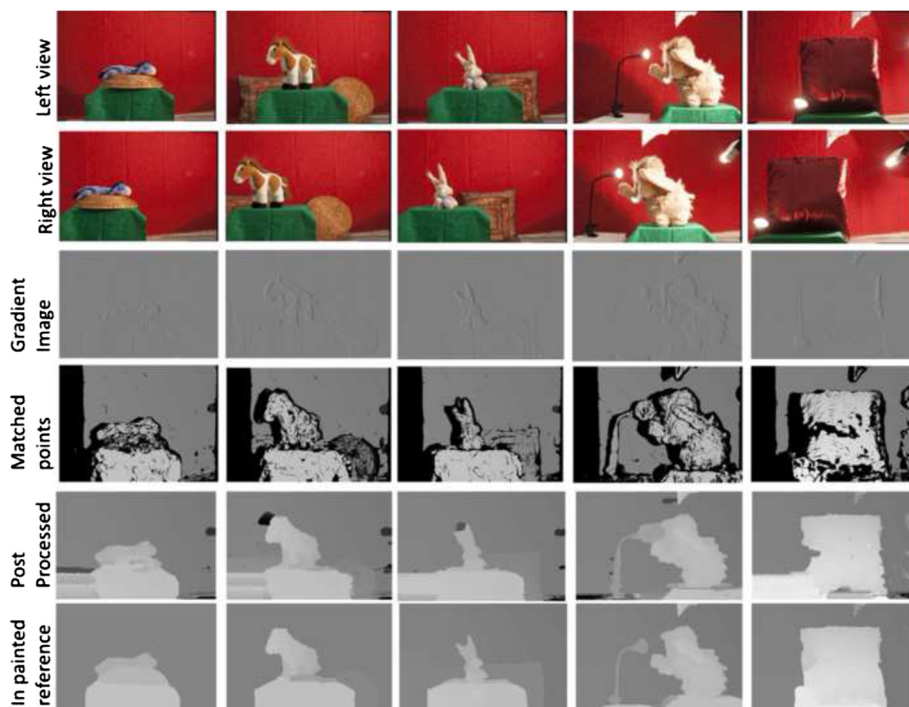

Fig. 2 HDR stereo data set and their corresponding reference disparity maps. *First two rows*: the left and right views of the images. *Third row*: the gradient images corresponding to the left view. *Forth row*: the matched points from two views using the customized HDR cost-volume stereo matcher. *Fifth row*: the post processed disparity maps. The *last row*: the in-painted reference disparity maps

Table 1 Maximum luminance value of the left and right images in our data set

Scene	Left, max. luminance (in cd/m^2)	Right, max. luminanc (in cd/m^2)
Donkey	5412	5536
Horse	5423	5543
Rabbit	5401	5570
Elephant	6802	7153
Pillow	8150	8221

The rows of this table are arranged with respect to the data set shown in Fig. 2 from left to right: Donkey, Horse, Rabbit, Elephant and Pillow

Table 3 Our combined tone-mapped-based approach in comparison with other approaches

Approach	Average RMSE
Reinhard	27.9986
Durand	26.1757
Fattal	24.4647
LDR	16.6513
Gastal	14.1403
Drago	11.9454
He	10.6072
Farbman	9.0224
Our combined approach	5.1754

Our results were compared to conventional LDR stereo matching and seven other well-known tone-mapped stereo matching approaches. The error is calculated as the average RMSE on the five introduced stereo image pairs

are over-/under-exposed. In Table 3, we illustrated our chosen exposures for each image in our data set. To choose the most informative LDR image, we ran the stereo matcher on all of the exposures and calculated the error using our ground truth. The image pair with the smallest error was chosen to be compared with other approaches.

5.3 Reference/ground truth disparity map

Disparity maps obtained from HDR stereo image pairs are post processed and used as our reference disparity information. We computed the HDR disparity maps by replacing the color intensity information and luma gradient values in Eq. 1 with radiance (R_i) and radiance gradient values ($\nabla_x R_i$), respectively [6], as shown in Eq. 11. These HDR disparity maps are shown as *post processed disparity maps* in Fig. 2.

$$C_{i,d} = \alpha \cdot \min \left[||R_i - R'_{i-d}||, \tau_1 \right] + (1 - \alpha) \cdot \min \left[||\nabla_x R_i - \nabla_x R'_{i-d}||, \tau_2 \right]. \quad (11)$$

In Fig. 2, the first two rows demonstrate the left and right views of the images which contain the low-textured background as well as tricky lighting conditions for stereo matching. The third row shows the *gradient images* corresponding to the left view, showing the lack of gradient information. In the fourth row, the *matched points* from two views using the customized HDR cost-volume stereo matcher [6] are shown. These

Table 2 Comparison of our combined graph-cut-based approach to several other combination methods

Combining approach	Average RMSE
Average	16.6087
Median	15.3736
Our graph-cut-based combination method combining four random tone-mapped disparities	10.5418
Our graph-cut-based combination method combining seven tone-mapped disparities	5.1754

disparity maps are not post processed. *The post processed disparity maps* are presented in the fifth row. In the last row, *the in-painted* improved reference disparity maps are depicted using the gradient information to fill in the wholes in the disparities presented in the fifth row.

We seek to get a deeper insight into the observed quality differences between the LDR and HDR matching results by comparing computed cost values in the intensity space versus HDR radiance space. Details of the cost values computed in the intensity (LDR) and radiance space (HDR) for a specific sample are provided in Fig. 3. Graphs (g_1), (g_2), (g_3), and (g_4) show the computed matching costs for specified pixels in e and f. In all graphs, cost values (first rows) are calculated from color/radiance information (middle row) and gradient information (bottom row) using the LDR and HDR cost computation algorithms as described before. Pixels x_1 and x'_1 in the LDR and HDR disparity maps (e and f), respectively, achieve a plausible disparity in both cases.

Contrarily, x_2 and x'_2 represent the same pixel in the LDR and HDR disparity map for which a wrong disparity in LDR but a correct disparity in HDR was achieved. The top row in graph (g_3) shows how working in conventional intensity space results in very similar matching costs inside low texture areas that are very likely to cause a mismatch. The top row in graph (g_4) presents the corresponding pixel matching costs in radiance space with only one minimum among matching costs. The same effect was observed on other image pairs too. Using scene radiance information instead of intensity values, we achieved one global minimum which is known to be a more reliable situation in optimization algorithms. Therefore, we use the HDR disparity maps (forth row in Fig. 2) with small modifications (last row in Fig. 2) as reference disparity information for our quantitative evaluation.
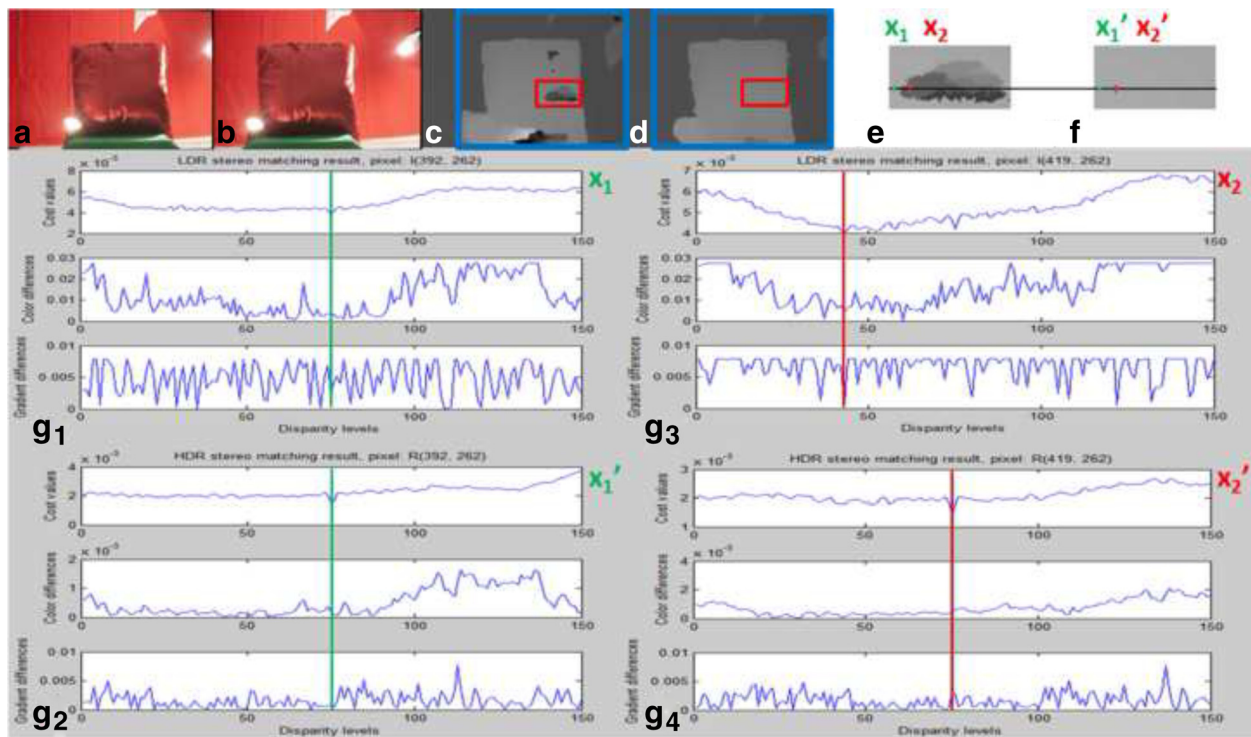
Fig. 3 LDR-HDR cost calculation comparison. **a**, **b** *Left* and *right* views; **c**, **d** LDR and HDR computed disparity maps; **e**, **f** zoomed areas from **c** and **d**; (g_1), (g_2), (g_3), and (g_4): from *top* to *bottom*: cost values, color/radiance-differences, and gradient-differences between pixels x_1, x_1', x_2, and x_2', and the pixel shifted by different disparity levels (1–150 pixels) in the *x* direction of the same line. (g_2) and (g_4) show how by working in HDR space we achieved a more pronounced minimum among costs

5.4 Qualitative evaluation

Figure 4 compares the disparity maps achieved from different input stereo data. The first row shows the HDR reference disparity maps calculated as described in Section 5.3. The following rows show the matching results on different tone-mapped stereo images as well as the LDR stereo input. The last row represents our combination approach disparities. Five HDR scenes with low-textured background are chosen for our experiments. The lighting and shadow situation in all of the scenes creates challenging situation for stereo matching using conventional LDR images.

As it is noticeable from the results, some TMOs perform worse than the traditional LDR approach. Reinhard TMO, Durand TMO, and Fattal TMO cause more artifacts and mismatched pixels than the traditional LDR matching. The reason for this is that the TM is applied on each image separately. Not so many TMOs are capable of keeping the consistency between several frames. Our results illustrate that Gastal, Drago, He, and Farbman TMOs are robust to some changes in the lighting condition and therefore achieve more informative disparity maps from stereo pairs. The combined approach outperforms all of the other calculated disparities.

5.5 Quantitative evaluation

The root mean square error (RMSE) value indicating the difference between each calculated disparity map and the corresponding HDR reference disparity is shown below the results in Fig. 4. Table 4 contains the average RMSE values on all of the five data set shown in Figs. 2 and 4. Based on the information presented in the Table 4, we derive some conclusions:

- Not all of the TMOs perform well for stereo matching. Three out of seven tone-mapped stereo images used in our experiment resulted in more mismatched disparities compared to conventional LDR image pairs. Choosing an effective TMO for the stereo matching and 3D reconstruction purposes is a completely different task than evaluating TMOs for visual and display purposes.

- Local TMOs based on edge-aware filters perform better for stereo matching than some of the other TMOs which are based on a global curve estimation. Among the four edge-aware filtering approaches implemented and studied in this paper, the Farbman filter suits disparity estimation better than the others.

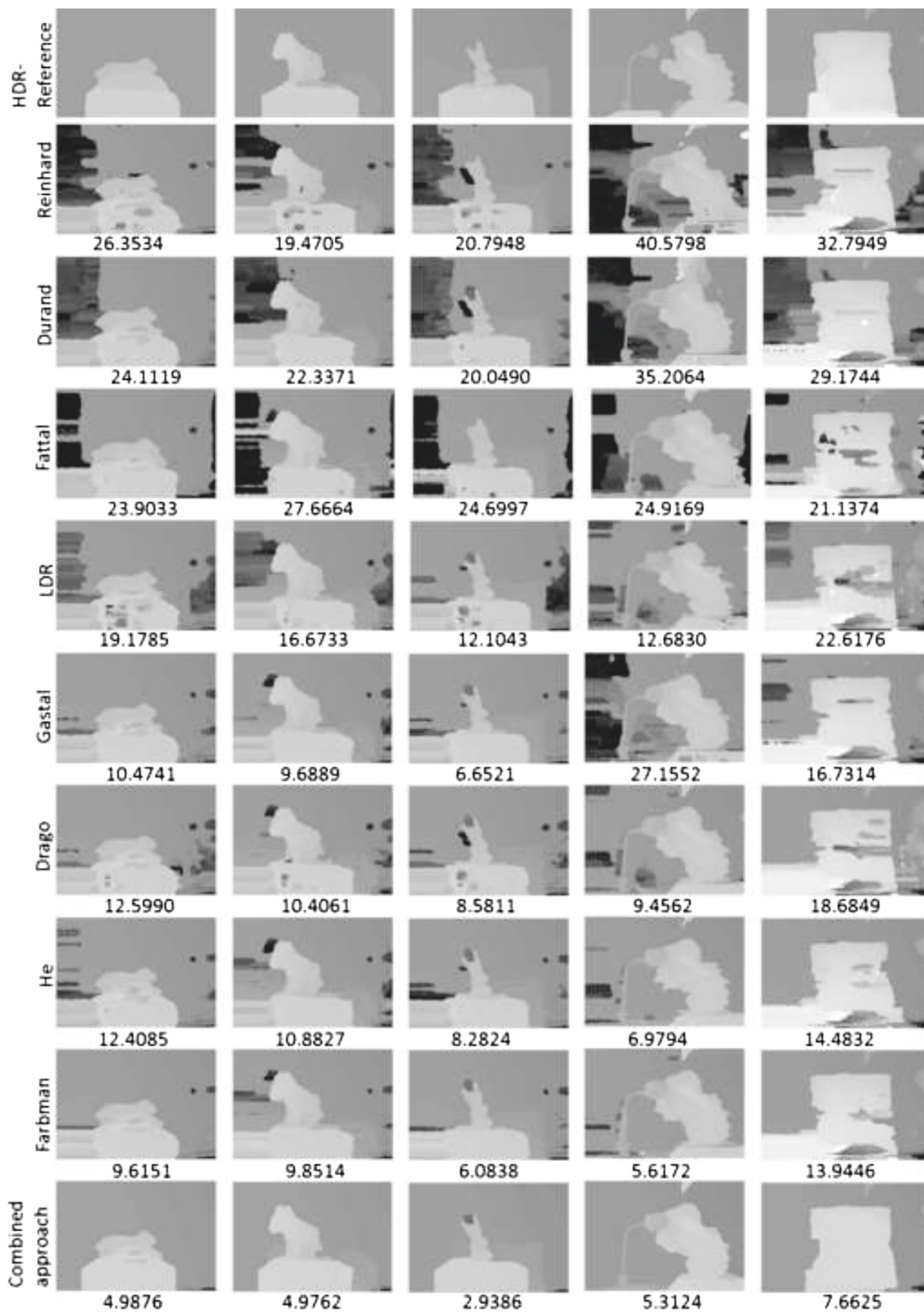- As expected, the combined disparities outperform all of the other tone-mapped stereo matchers. This is

Fig. 4 Disparity map comparison. The RMSE values are shown below each disparity. *First row*: HDR reference disparity maps illustrated in Fig. 2. Reinhard, Fattl, and Durand TMOs achieve disparities with bigger error than conventional LDR stereo matching. Gastal, Drago, He, and Farbman TMOs perform better than LDR stereo matching while our combined graph-cut approach outperforms all of the LDR and tone-mapped disparity maps

Table 4 Our data set luminance level in comparison to some common lighting environments [46]

Condition	Illuminance (in cd/m^2)
Starlight	10^{-3}
Moonlight	10^{-1}
Indoor lighting (controlled)	10^2
Sunlight	10^5
Our data set lighting	10^3
Max intensity of common monitors	10^2

based on the fact that the disparities with mismatched pixels in some regions might carry valuable information in some other specific areas of the image. Our experiments show that combining the four best performing tone-mapped disparities (Gastal, Drago, He, and Farbman) does not achieve as informative result as when we used all the seven disparities.

Table 5 shows the effectiveness of the MRF model. The simplest way to combine some proposed results is to use average or median. We averaged our seven disparity maps and achieved an error which is very close to the LDR result. We also show the average RMSE on choosing four random tone-mapped disparities among the seven in our MRF solution. Even though combining four disparities results in less RMSE, our experiments point out the fact that using more information to combine results in more informative results. One can use more than seven tone-mapped disparities; however, both of our qualitative and quantitative evaluations illustrate that seven proposed results are discriminative. There is a trade-off between the complexity and reducing the RMSE. We suggest to keep the proposed results for combination below ten to be able to achieve near real-time disparity estimation. However, the question of how much improvement one could gain using the proposed fusion approach remains open.

Table 5 The most informative exposure time which was used for LDR stereo matching results presented in Fig. 4 and Table 3

Scene	Chosen exposure time for LDR stereo matching (in seconds)
Donkey	1/125
Horse	1/125
Rabbit	1/125
Elephant	1/60
Pillow	1/60

The rows of this table are arranged with respect to the data set shown in Fig. 2 from left to right: Donkey, Horse, Rabbit, Elephant, and Pillow

5.6 TMO selection

As mentioned before, the TMOs used in this study are interchangeable. There are two important aspects of this combination approach which needs special attention. First is TMO selection and second is the number of the TMOs to be fused. There is no unique answer to these questions, but we share our insight from the trial and errors.

One can generate candidate results using different parameters for TMOs which is a similar approach as the one discussed in [44] to fuse optical flow solutions. The candidate solutions achieved from parameter modifications do not differ as much as when using totally different methods. Our proposed fusion method is choosing the most probable disparity value per pixel among the candidate disparity maps. A disparity value is considered as a robust and reasonable result if and only if all different approaches agree on that value. We strongly suggest using different TMOs based on completely different concepts to satisfy the diversity for the probabilistic treatment.

Table 5 shows the outcome of fusing four and seven TMOs disparity maps. Even though fusing four TMOs reduces the RMSE, we suggest using at least six or seven results. The maximum number of the disparities to be combined is an open question. The fusion approach is fast and works near real time, but it is obvious that adding more results to be combined will slow the approach down.

5.7 Post processing

In this section, we discuss the various parameter settings and post processing methods used in our research.

Stereo matching post processing/smoothing the cost-volume: Various smoothing approaches could be used for this purpose. In [39], guided filter [36] is found to be faster and more effective. We use the same filtering approach as the post processing step of our LDR and HDR stereo matching. The HDR modified formulas applied on the radiance channel are shown here. C' is the filtered cost volume.

$$C'_{i,d} = \sum_j W_{i,j}(R) C_{i,d}. \tag{12}$$

The filter weights $W_{i,j}$ (in pixel positions i and j) are chosen according to the weights of guided filter as used in [39]. Having the radiance guided image R, weights are defined as follows:

$$W_{i,j} = \frac{1}{|w|^2} \sum_{k:(i,j)\in w_k} (1 + (R_i - \mu_k)^T \tag{13}$$
$$\times \left(\sum k + \in U \right)^{-1} (R_j - \mu_k)),$$

where μ_k is the mean vector and $\sum_k$ is the co-variance vector calculated in a squared window w_k with dimensions of $(2r + 1) \times (2r + 1)$, centered at pixel k in radiance image R. Further details can be found in He et al. [36].

Manual in-painting for reference disparity generation: Most of the approaches for disparity ground truth calculation contain a step of smoothing or filling small wholes or noise-based artifacts to achieve a better quality disparity as the reference. It is suggested in [9] to use some interpolation method to fill small holes in the ground truth disparity. Akhavan et al. [10] manually filled in the holes in the reference disparity map caused by shadows. As shown in the last row of the Fig. 2, post processed disparity maps were manually enhanced using the (1) gradient information to estimate the edges and (2) neighboring disparity values. This is an additional step to enhance the reference disparities which can be omitted from the whole framework process. In other words, the post processed HDR disparities can be used as the ground truth.

5.8 Environmental settings

It takes approximately 1 s to process the combination of seven disparity maps using our proposed framework on an IBM Core i7 2.80GHz CPU using single-threaded Matlab code. This excludes the time of input/output operations and the time to calculate each of the disparity maps. Since most of the time is spent performing independent per-pixel operations, the algorithm is well suited for parallel processing.

5.9 Fusion moves

In addition to the proposed graph-cut combining approach, we tried using fusion moves for combining our candidate solution as introduced in [45]. The results were not as impressive as our demonstrated ones because the proposal set of solutions to be combined using fusion moves must satisfy two important constraints: (1) quality and (2) diversity according to [44]. Lempitsky et al. [44] used over 200 proposed solutions from different approaches and parameter settings for their fusion. Having enough diversity in the proposals is an important prerequisite to the fusion move approach. We only had seven candidate solutions for our combination which leads to not satisfactory disparity maps using fusion moves.

6 Conclusions

We proposed a novel framework for combining several tone-mapped disparity maps in order to reduce the number of incorrect matching points and improve the performance of image matching in the HDR scenes. We used TMOs to compress HDR stereo images in order to keep backward compatibility to conventional stereo matchers. Our obtained disparities from different tone-mapped stereo images are used along with our graph-cut implementation using α-expansion moves to select the minimum cost disparity value per pixel. To evaluate our results, we created the ground truth disparity map using original HDR stereo images and customized matching

calculations to radiance space. Our qualitative and quantitative evaluations of seven different tone-mapped disparity maps, LDR, HDR, and proposed combined disparities show that there is a lot of room for disparity improvement in challenging lighting environments. Our novel fusion approach reduced the average RMSE of the conventional LDR stereo matching by the factor of 3. Using more disparity maps in the combining framework might improve our results but for sure will slow the combination process down. Finding the maximum number of the disparities which can effectively improve the results is a challenge that can be investigated in the future. In this work, one specific stereo matching algorithm is used for all the disparity map calculations. An interesting future research project can be dedicated to investigate how the proposed framework affects other stereo matching methods.

References
1. P Debevec, J Malik, in *ACM SIGGRAPH*. Recovering high dynamic range radiance maps from photographs (ACM, New York, NY, USA, 1997), pp. 369–378
2. F Drago, K Myszkowski, T Annen, N Chiba, Adaptive logarithmic mapping for displaying high contrast scenes. Comput. Graph. Forum. **22**, 419–426 (2003)
3. H Seetzen, W Heidrich, W Stuerzlinger, G Ward, L Whitehead, M Trentacoste, A Ghosh, A Vorozcovs, in *ACM SIGGRAPH*. High dynamic range display systems (ACM, New York, NY, USA, 2004), pp. 760–768
4. Selmanović, K Debattista, T Bashford-Rogers, A Chalmers, Generating stereoscopic HDR images using HDR-LDR image pairs. ACM Trans. Appl. Percept. **10**(1), 3–1318 (2013)
5. C Wang, C Tu, A multi-exposure images fusion approach for very large dynamic range scenes. Int. J. Signal Process. Image Process. Pattern Recognit. **7**(5), 217–228 (2014)
6. T Akhavan, H Yoo, M Gelautz, in *22th European Signal Processing Conference (EUSIPCO 2014)*. Evaluation of LDR, tone mapped and HDR stereo matching using cost-volume filtering approach (IEEE, Lisbon, Portugal, 2014), pp. 1–6
7. D Scharstein, in *CVPR*. Learning conditional random fields for stereo (IEEE Computer Society, Minneapolis, Minnesota, USA, 2007), pp. 1–8
8. K Konolige, in *Proc. of the Intl. Symp. of Robotics Research (ISRR)*. Small vision system: hardware and implementation (Springer London, 1997), pp. 111–116
9. D Scharstein, R Szeliski, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2003)*. High-accuracy stereo depth maps using structured light, vol. I (IEEE Computer Society, Madison, WI, USA, 2003), pp. 195–202
10. T Akhavan, C Kappeler, J Cho, M Gelautz, in *HDRi2014 - Second International Conference and SME Workshop on HDR Imaging*. Stereo HDR disparity map computation using structured light (Eurographics Association and Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK, 2014)
11. G Eilertsen, R Wanat, RK Mantiuk, J Unger, Evaluation of tone mapping operators for HDR video. Comput. Graph. Forum. **32**(7), 275–284 (2013)
12. A Troccoli, SB Kang, SM Seitz, in *3DPVT*. Multi-view multi-exposure stereo (IEEE Computer Society, Chapel Hill, North Carolina, USA, 2006), pp. 861–868
13. H-Y Lin, W-Z Chang, in *ICIP*. High dynamic range imaging for stereoscopic scene representation (IEEE, 2009), pp. 4305–4308

14. N Sun, H Mansour, RK Ward, in *ICIP*. HDR image construction from multi-exposed stereo LDR images (IEEE, Hong Kong, China, 2010), pp. 2973–2976
15. D Rüfenacht, *Stereoscopic High Dynamic Range Video*. (Master's thesis, EPFL, Lausanne, Switzerland, 2011)
16. F Lu, X Ji, Q Dai, G Er, in *ACCV*. Multi-view stereo reconstruction with high dynamic range texture (Springer, Queenstown, New Zealand, 2011), pp. 412–425
17. R Ramirez, C Loscos, IM Artusi, in *HDRi2013 - First International Conference and SME Workshop on HDR Imaging*. Patch-based registration for auto-stereoscopic HDR content creation (Eurographics Association and Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK, 2013)
18. M Bätz, T Richter, J Garbas, A Papst, J Seiler, A Kaup, High dynamic range video reconstruction from a stereo camera setup. Signal Process. Image Commun. **29**, 191–202 (2014)
19. R Ramirez Orozco, C Loscos, I Martin, A Artusi, in *Winter School of Computer Graphics*. Multiscopic HDR image sequence generation (Pilsen, Czech Republic, 2015)
20. Z Mai, C Doutre, P Nasiopoulos, RK Ward, Rendering 3D high dynamic range images: subjective evaluation of tone-mapping methods and preferred 3D image attributes. J. Sel. Topics Signal Process. **6**(5), 597–610 (2012)
21. TO Aydin, N Stefanoski, S Croci, M Gross, A Smolic, Temporally coherent local tone mapping of HDR video. ACM Trans. Graph. **33**(6), 196–119613 (2014)
22. P Paclik, RPW Duin, GMPV Kempen, R Kohlus, Segmentation of multi-spectral images using the combined classifier approach. Image Vis. Comput. **21**(6), 473–482 (2003)
23. T Akhavan, M Moghaddam, A color constancy method using fuzzy measures and integrals. Optical Rev. **18**(3), 273–283 (2011)
24. C Yaacoub, C Yaghi, C Bou-Rizk, in *ICASSP*. Fusion of tone-mapped high dynamic range images based on objective range-independent quality maps (IEEE, Florence, Italy, 2014), pp. 1195–1199
25. A Yoshida, V Blanz, K Myszkowski, H Seidel, in *Human Vision and Electronic Imaging X, SPIE*. Perceptual evaluation of tone mapping operators with real-world scenes (SPIE, San Jose, California, USA, 2005)
26. P Ledda, A Chalmers, T Troscianko, H Seetzen, in *ACM SIGGRAPH*. Evaluation of tone mapping operators using a high dynamic range display (ACM, 2005), pp. 640–648
27. M Čadík, M Wimmer, L Neumann, A Artusi, in *14th Pacific Conference on Computer Graphics and Applications*. Image attributes and quality for evaluation of tone mapping operators (Press, 2006), pp. 35–44
28. S Schuon, C Theobalt, J Davis, S Thrun, in *CVPR*. LidarBoost: depth superresolution for ToF 3D shape scanning (IEEE Computer Society, 2009), pp. 343–350
29. J Diebel, S Thrun, in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*. An application of Markov random fields to range sensing (MIT Press, Cambridge, MA, 2005), pp. 291–298
30. S Izadi, D Kim, O Hilliges, D Molyneaux, R Newcombe, P Kohli, J Shotton, S Hodges, D Freeman, A Davison, A Fitzgibbon. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera (ACM, 2011)
31. T Akhavan, ME Moghaddam, in *2nd International Conference on Image Processing Theory Tools and Applications*. A new combining learning method for color constancy (IEEE, 2010), pp. 421–425
32. SZ Li, *Markov Random Field Modeling in Image Analysis*, 3rd edn. (Springer, 2009). doi:10.1007/978-1-84800-279-1
33. E Reinhard, M Stark, P Shirley, J Ferwerda, in *SIGGRAPH. Volume 21 Issue 3*. Photographic tone reproduction for digital images (ACM, 2002), pp. 267–276
34. R Fattal, D Lischinski, M Werman, in *SIGGRAPH. Volume 21 Issue 3*. Gradient domain high dynamic range compression (ACM, 2002), pp. 249–256
35. F Durand, J Dorsey, Fast bilateral filtering for the display of high-dynamic-range images. ACM TOG. **21**, 257–266 (2002)
36. K He, J Sun, X Tang, in *ECCV*. Guided image filtering, (2010), pp. 1–14
37. Z Farbman, R Fattal, D Lischinski, R Szeliski, Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM TOG. **27**(3), 67–16710 (2008)
38. ESL Gastal, MM Oliveira, Domain transform for edge-aware image and video processing. ACM TOG. **30**(4), 69–16912 (2011)
39. A Hosni, C Rhemann, M Bleyer, C Rother, M Gelautz, Fast cost-volume filtering for visual correspondence and beyond. IEEE Trans. Pattern Anal. Mach. Intell. **35**(2), 504–511 (2013)
40. C Wang, N Komodakis, N Paragios, Markov random field modeling, inference & learning in computer vision & image understanding: a survey. Comput. Vis. Image Underst. **117**(11), 1610–1627 (2013)
41. Y Boykov, O Veksler, R Zabih, Fast approximate energy minimization via graph cuts. PAMI. **23**(11), 1222–1239 (2001)
42. E Reinhard, G Ward, S Pattanaik, P Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005)
43. H Hirschmüller, D Scharstein, in *CVPR*. Evaluation of cost functions for stereo matching, (2007), pp. 1–8
44. V Lempitsky, S Roth, C Rother, in *CVPR*. FusionFlow: discrete-continuous optimization for optical flow estimation, (2008)
45. V Lempitsky, C Rother, S Roth, A Blake, Fusion moves for Markov random field optimization. Technical Report MSR-TR-2009-60 (2009). Microsoft
46. BA Wandell, *Foundations of Vision*. (Sinauer Associates, Inc., Sunderland, MA, US, 1995)