A framework for HDR stereo matching using multi-exposed images

Tara Akhavan[1] and Hyunjin Yoo[1] and Margrit Gelautz[1]

[1]Institute of Software Technology and Interactive Systems, Vienna University of Technology, Austria

Abstract

Real world scenes that contain high dynamic range illumination present a special challenge for stereo matching algorithms due to a lack of texture in over- or under-exposed image regions. In this paper, we discuss possibilities for combining state-of-the-art stereo matching algorithms with High Dynamic Range (HDR) imaging techniques, in order to exploit a set of multi-exposed input images of both the left and right stereo view for high-quality stereo reconstruction. We sketch the overall concept of our HDR stereo matching framework and demonstrate some first steps of its implementation, including the acquisition of HDR stereo test data, stereo matching experiments on tone-mapped images, and ideas for combining disparity maps derived from different exposures.

Categories and Subject Descriptors (according to ACM CCS): High Dynamic Range, Stereo matching, Tone-mapping.

1. Introduction

During the last decades, stereo matching has become more and more important for many applications based on 3D reconstruction such as depth image based rendering, 3DTV, and multi-view autostereoscopic displays. The main goal of stereo matching is to reconstruct 3D real world information from a pair of images obtained from two different views. There are many stereo matching algorithms which perform well when there are no under-exposed or over-exposed areas in the scene. However, it is difficult to obtain a high quality disparity map using conventional photography, when a scene contains high dynamic range content. HDR (High Dynamic Range) imaging serves to represent a real world scene which contains a wide range of low to high luminances. By combining HDR with stereo techniques, the disparity map of a high dynamic range scene can be obtained. In an ongoing research project, we address the question in which way the HDR and stereo processing steps should be combined in order to maximize the quality of the disparity maps that can be obtained from a given set of multi-exposed stereo input images.

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 them. Sun et al. [SMW10] constructed an HDR image by using two stereo images with two different exposures. They first calculated an initial disparity map to obtain the camera response function. After refining the disparity map, an HDR image is constructed. Since they only used two images in the stereo matching process, if one of them is significantly over-exposed or under-exposed relative to the other, it may be difficult to calculate a reliable disparity map. Troccoli et al. [TKS06] also adopted multi-view stereo with different exposures to obtain both the depth map and HDR image. Rufenacht [Ruf11] created stereoscopic video content using different exposure times for the right and left cameras. He captured video frames, estimated the camera response function, aligned the image and estimated the HDR radiance map. In contrast to what we aim to achieve, his main focus was on the quality of HDR radiance maps and not the quality of disparity maps. Bonnard et al. [BLV*12] leveraged an eight view video camera to acquire HDR video content for autostereoscopic displays. They did not use any stereo matching, but built their pixel match list with the help of the camera properties with eight aligned, equally distributed objectives. Their main focus was on creating HDR content as well. Recently, Selmanovic et al. [SDBRC12] introduced a number of compression methods for Stereoscopic High Dynamic Range (SHDR). An HDR image requires floating point value storage per color channel per pixel, so SHDR needs twice of the storage. None of the aforementioned papers provides a systematic overview and evaluation of different strategies for obtaining disparity maps from multi-exposed stereo input data, which is the final goal of our work.

The rest of the paper is organized as follows: In Section 2 we present a theoretical framework for generating depth maps from multi-exposed stereo images as input. A brief explanation of the three proposed approaches are contained in this section. As part of our ongoing work, Section 3 illustrates some preliminary matching results. Finally, Section 4 gives a summary and outlook.

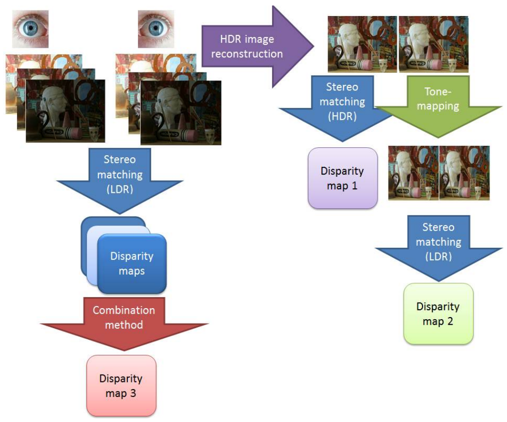

Figure 1: *Proposed framework for solving the multi-exposure stereo matching problem.*

2. Theoretical framework

In this section, we discuss our framework which contains three different ways of using multi-exposed left and right stereo input images to generate disparity maps.

2.1. Overview

We propose a framework for using stereo multi-exposed images to solve the stereo matching problem. Our data set contains three different-exposed images for each view (for the left view: under-, normal- and over-exposed images; for the right view: under-, normal- and over-exposed images). More details about the data set are discussed in Section 3. In the first and second approaches, as shown in Figure 1, HDR images of left and right views are first constructed from multi-exposed images for each view. The next step is to obtain the disparity map from the computed pair of HDR images. As an alternative, tone-mapped images can be used instead of the original HDR images as input for the stereo matching algorithm. Since more detail is represented in both the HDR and tone-mapped images in comparison to normal, low dynamic range (LDR) images, better disparity maps, especially in over-exposed and under-exposed areas, can be expected than when using a single initial pair of LDR stereo images as input.

The third approach applies the stereo matching algorithm on left and right images with same exposures. This results in three different disparity maps; one for each exposure. Then a combination method has to be used to combine the disparity maps. The processing chain is shown on the left side of Figure 1. In our implementation, a fuzzy measure and integral combination method is used as the combination method to fuse the disparity maps.

2.2. Methods

The three suggested approaches shown in Figure 1 are discussed in more detail in the following of this section. Each of them results in a different disparity map. A systematic comparison between these approaches should provide useful information for choosing the best, easiest and computationally least expensive way of dealing with the multi-exposed stereo problem.

2.2.1. HDR stereo images

In this approach, HDR images for each view are constructed first. There are many open source methods and tools that can create an HDR image from multi-exposed LDR ones. We use HDRshop [DM08]. By using an HDR image for each view and computing the disparity map between the two HDR views, better stereo matching results can be expected. However, one needs to generate HDR stereo matching algorithms for this purpose or revise the normal, LDR stereo matching methods in a way to work with HDR formats. Given the fact that a considerable number of state-of-the-art stereo algorithms are nowadays implemented on the GPU (Graphics Processing Unit), the change to the HDR format may not always be straightforward. In our ongoing experiments, we use a stereo matching algorithm based on cost volume filtering which was proposed recently by [HRB*13]. This approach leads to *Disparity map 1*, as shown in Figure 1.

2.2.2. Tone-mapped stereo images

To show an HDR image on a conventional display with low dynamic range, a tone-mapping process is needed. The importance of having a suitable method for reducing the dynamic range in HDR images has been increasing in the last couple of years and a large variety of tone-mapping methods have been introduced. Choosing a suitable tone-mapper among all of the available ones, according to the application, is not easy. A good comparison of tone-mapping operators has been presented in [CWNA08]. After constructing the HDR images for the left and right views, we tried a couple of the best tone-mappers suggested in [CWNA08] and tested them on our stereo HDR images. After choosing a suitable tone-mapping method, we run the stereo matching algorithm on the tone-mapped images of the left and right views. Contrary to the method from Section 2.2.1, no particular adjustment of the stereo matching algorithm - beyond usual parameter tuning - is necessary in this case. This approach leads to *Disparity map 2*, as shown in Figure 1.

2.2.3. Combination of disparity maps

In our third approach, a disparity map for each exposure is constructed, which results in three different disparity maps for three different exposures. The challenge here is to get the best combination of the disparity maps in order to gain the most accurate stereo matching result. For solving this issue, a fuzzy weighting system is proposed for fusing different parts of the disparity maps based on fuzzy measures and integrals. Fuzzy integrals and, more specifically, the Choquet fuzzy integral [Cho54] are widely accepted to fuse multi-classifier characteristics and express the importance of individual classifiers as well as considering the interactions among classifiers [FLC10].

Figure 2 shows an overview of the fuzzy integral method for combining disparity map results. *h(1)* to *h(3)* are the disparity maps for each exposure; *w1* to *w3* are weights for each disparity map and λ is the index indicating the interaction between disparity maps. This proposed framework can be used with more than three exposures. Obtaining accurate weights and interaction weights between disparities is very challenging. As in so many other fuzzy measure problems, useful and accurate weighting assignments are achieved with the help of an expert user input. This approach leads to *Disparity map 3*, as shown in Figure 1.

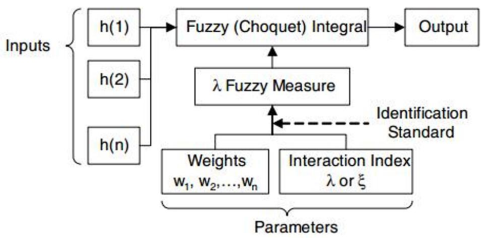

Figure 2: *Fuzzy measure and fuzzy integral.*

3. Experimental results

For our experiments, we use (1) available data sets provided by the scientific community and (2) new data sets that we generate in our lab. Each of our generated data sets contains eight different images with different exposures for the left view and eight images for the right view with the exact same exposures. The eight exposures calculated in seconds are: 1/3 s, 1/6 s, 1/13 s, 1/25 s, 1/50 s, 1/100 s, 1/200 s and 1/400 s. Figure 3 shows a sample from our self-recorded data set. In this figure we show three of the eight captured exposures, for each of the left and right views. In order to generate challenging data samples, we tried (1) to have a lamp or direct light source in our indoor images, or (2) to capture scenes containing a window, which is usually over-exposed in some areas. Conventional stereo matching methods would not be able to generate accurate disparity maps for under-exposed or over-exposed parts of an image. For instance, using the middle-exposed image from Figure 3, the bear's body (under-exposed), details inside the lamp (over-exposed) and the bear's left ear (over-exposed) would be challenging parts of the scene for the stereo matching.

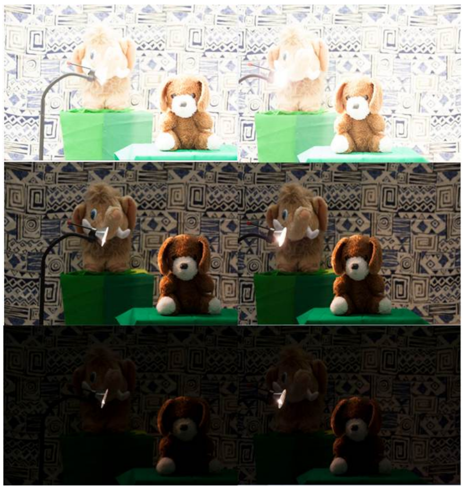

Figure 3: *A sample from a multi-exposed stereo data set. First row: over-exposed left and right views (exposure = 1/3 s); Second row: normal-exposed left and right views (exposure = 1/50 s); Third row: under-exposed left and right views (exposure = 1/400 s).*

For our first experiments, we used a simplified Matlab version of the stereo matching method introduced by Hosni et al. [HRB*13]. For comparison with ground truth, we used a multi-exposure stereo data set provided by the Middlebury stereo evaluation website [HS07].

According to Cadik [CWNA08], Reinhard's tone-mapper [RD05] works better than some other well-known tone-mapping operators with respect to some objective and subjective image attributes introduced by Cadik. The result we present here was obtained using Reinhard's tone-mapping method implemented in HDRtoolbox [HDr]. Figure 4 shows some of our preliminary matching results. On the top row of the figure, the left and right stereo Aloe images in three different exposures can be seen. Exposure values are reported in millisecond from [HS07]. The second row presents the middle-exposed Aloe image for the left and right views. Images **c** and **d** are view 1 and view 5 of the Aloe image, illumination 3, exposure 1, from the Middlebury stereo data set. On the third row, image **f** is the tone-mapped image using the three images shown in **b**. Image **e** is the available ground truth from Middlebury. The last row demonstrates the comparison of two disparity maps obtained by running the stereo

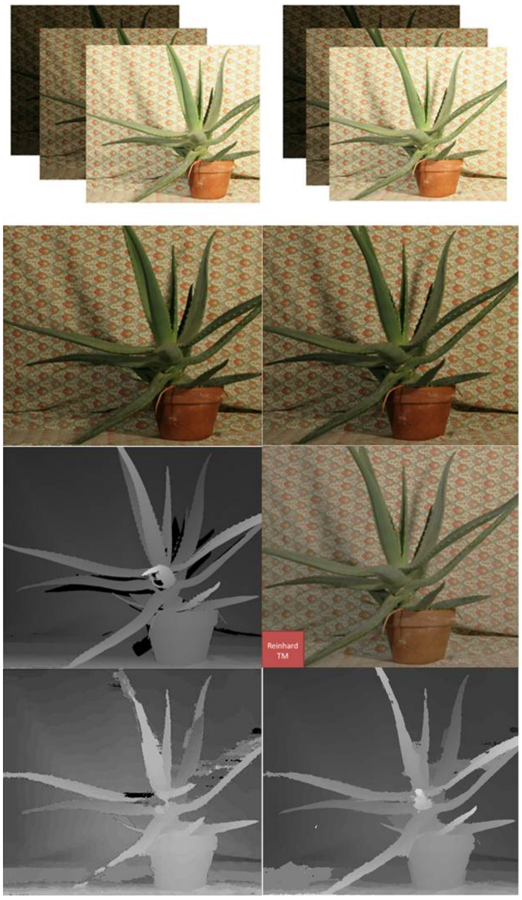

Figure 4: ***a**: Aloe left view (view 1) image with three exposures (125, 500, 2000 ms), **b**: Aloe right view (view 5) image with the same three exposures, **c** and **d**: Aloe left view (view 1) and right view (view 5) images, middle-exposure (500 ms), **e**: Middlebury disparity map ground truth for **c** and **d**, **f**: tone-mapped image from **b** using Reinhard's method, **g**: disparity map of images **c** and **d**, **h**: disparity map of tone-mapped images generated from **a** and **b**.*

matching method on (1) LDR views and (2) tone-mapped views of the Aloe image. Image **g** presents the disparity map of the LDR images **c** and **d** while image **h** is the disparity map of the tone-mapped LDR images. The disparity map derived from the stereo tone-mapped views preserves more detail of the edges.

4. Summary and outlook

The main contributions of our ongoing work as discussed in this paper can be summarized as follows:

1. A new multi-exposed stereo data set is generated. We focus on capturing high dynamic range stereo scenes which will provide the basis for evaluating different HDR stereo approaches.
2. A framework for solving the multi-exposed stereo problem has been introduced and three different approaches have been discussed. Preliminary results for the tone-mapped stereo approach have been presented.
3. A novel combination method, to achieve an accurate disparity map from different exposures, has been suggested, using a fuzzy measure and integral combination method.

The goal of our work is to explore possibilities for combining state-of-the-art stereo algorithms with HDR image processing techniques in such a way that the 3D scene reconstruction benefits as much as possible from the enhanced level of detail provided by the HDR stereo input.

Acknowledgements

This work was partially supported by COST Action IC1005. Tara Akhavan is funded by the Vienna PhD School of Informatics, and Hyunjin Yoo is supported by the Austrian Science Fund (FWF) under project M1383-N23.

References

[BLV*12] BONNARD J., LOSCOS C., VALETTE G., NOURRIT J.-M., LUCAS L.: High-dynamic range video acquisition with a multiview camera. In *SPIE* (2012). 1

[Cho54] CHOQUET G.: Theory of capacities. *Annales de l'Institut Fourier 5* (1954), 131–295. 3

[CWNA08] CADIK M., WIMMER M., NEUMANN L., ARTUSI A.: Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers and Graphics* (2008), 330–349. 2, 3

[DM08] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes* (2008), ACM, pp. 1–10. 2

[FLC10] FENG H.-M., LI X.-F., CHEN J.-F.: A comparative study of four fuzzy integrals for classifier fusion. In *International Conference on Machine Learning and Cybernetics (ICMLC)* (2010), pp. 332– 338. 3

[HDr] HDRTOOLBOX: http://www.mathworks.com/matlabcentral /linkexchange/links/2792-the-hdr-toolbox. 3

[HRB*13] HOSNI A., RHEMANN C., BLEYER M., ROTHER C., GELAUTZ M.: Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 2 (2013), 504–511. 2, 3

[HS07] HIRSCHMÜLLER H., SCHARSTEIN D.: Evaluation of cost functions for stereo matching. In *CVPR* (2007). 3

[RD05] REINHARD E., DEVLIN K.: Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 1 (2005), 13–24. 3

[Ruf11] RUFENACHT D.: *Stereoscopic High Dynamic Range Video*. Master's thesis, EPFL, Lausanne, Switzerland, 2011. 1

[SDBRC12] SELMANOVIC E., DEBATTISTA K., BASHFORD-ROGERS T., CHALMERS A.: Backwards compatible jpeg stereoscopic high dynamic range imaging. In *TPCG* (2012), pp. 1–8. 1

[SMW10] SUN N., MANSOUR H., WARD R. K.: Hdr image construction from multi-exposed stereo ldr images. In *ICIP* (2010), pp. 2973–2976. 1

[TKS06] TROCCOLI A., KANG S. B., SEITZ S. M.: Multi-view multi-exposure stereo. In *3DPVT* (2006), pp. 861–868. 1