HDRi2014 - Second International Conference and SME Workshop on HDR imaging,
Chalmers A., and Bouatouch K. (Editors)

Stereo HDR Disparity Map Computation Using Structured Light

Tara Akhavan[1] and Christian Kapeller[1] and Ji-Ho Cho[1] and Margrit Gelautz[1]

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

Abstract

In this paper, we present work in progress towards the generation of a ground truth data set for High Dynamic Range (HDR) stereo matching. The development and evaluation of novel stereo matching algorithms that are tailored to the characteristics of HDR images would greatly benefit from the availability of such reference data. We describe our laboratory setup and processing steps for acquiring multi-exposed stereo images along with a corresponding reference disparity map computed by a structured light approach. We discuss the special requirements and challenges which HDR test scenes impose on the computation of the reference disparities and show some preliminary results.

Categories and Subject Descriptors (according to ACM CCS): High Dynamic Range Imaging, Stereo Matching, Structured Light.

1. Introduction

Recent research in the joint area of High Dynamic Range (HDR) imaging and stereo matching [TKS06, CL08, SMW10, Ruf11, BLV*12, SDBRC12, AYG13, RLIA13, SDBRC13, SDBRC14, BRG*14] has demonstrated the need for HDR stereo data sets in order to develop and evaluate novel stereo matching approaches that are tailored to the peculiarities of HDR images. To the authors' knowledge, practically no HDR stereo data sets are currently publicly available. This is opposed to the rich amount of conventional - that is, Low Dynamic Range (LDR) - stereo images that are offered by various databases on the web. In particular, the well-known Middlebury Stereo Vision Page[†] has boosted stereo research in recent years by providing sets of stereo data along with their corresponding ground truth disparities. The disparity values denote the geometry-induced shift in location between corresponding pixels in the left and right stereo view and are inversely proportional to depth. An important requirement on such ground truth disparities is that their accuracy exceeds that of the algorithms to be evaluated. It is worth noting that the Middlebury data set includes also stereo image pairs that were acquired with different exposures. However, their differences in exposure are too small to cover a high dynamic range of illumination. Applying and evaluating stereo matching on HDR scenes requires HDR stereo image pairs as well as their matching ground truth disparity map. We discuss the first steps towards this requirement in this paper.

Figure 1: *Scene capturing process. Top left: experimental setup, top right: HDR image capturing, bottom left: example of a horizontal Gray-code image, bottom right: example of a vertical Gray-code image.*

[†] http://vision.middlebury.edu/stereo/data/

In the context of ongoing research towards the development of HDR stereo matching algorithms [AYG13], we seek to acquire a variety of HDR data sets along with corresponding reference disparity maps that can serve as ground truth for evaluating HDR stereo matching algorithms. When setting up the test scenes, we pay special attention to including scene characteristics that are representative of HDR scenes (for example, highly saturated regions) as well as features that are known to be challenging for stereo matching algorithms (for example, lack of texture or slanted surfaces).

We use a structured light technique for our reference disparity computation. In [SFPL10] a quantitative and qualitative survey on different structured light approaches is carried out. We follow the spatio-temporal Gray-code approach because (1) we prioritize accuracy over acquisition time, (2) we only capture still scenes, and (3) the capturing method used in the Middlebury data sets is well proven [SS03]. In this paper, we explain our experimental hardware and software setup (see Fig. 1, top left) for (1) the stereo HDR capturing process (see Fig. 1, top right) and (2) the disparity map computation process based on structured light (see Fig. 1, second row). In Section 3 the preliminary calculated reference disparity map for the HDR scene captured in Fig. 1 is shown and discussed.

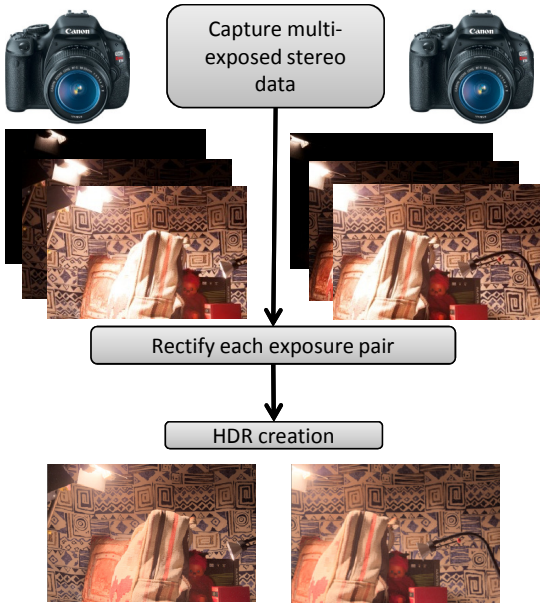

Figure 2: *Stereo HDR capturing process.*

2. Overview of our approach

In this section, we explain our stereo HDR data capturing and reference depth computation process following the structured light technique described in [SS03] using Gray-code patterns.

2.1. Stereo HDR image capturing

Our laboratory setup is shown in Fig. 1. The scenes are arranged on an optical table. For image capturing, we use a Canon EOS 1D Mark III DSLR with a 28.1 x 18.7 mm sensor capable of delivering RAW images of 3609x2699 pixel size. The images were shot with a focal length between 45 mm and 60 mm. The camera is mounted on a 150 mm motorized linear stage placed on one end of the table. We capture both left and right views with the same camera using a controller software. This configuration ensures that the intrinsic camera parameters of our vision system remain identical in all captures. We use several powerful direct light sources to generate highly saturated regions in our images as shown in the top right image of Fig. 1.

We acquire our stereo HDR data following the steps illustrated in Fig. 2. We capture nine different exposures for the left view, then using the controller software we move the camera to capture the same scene with the same exposures from the right point of view. The nine different exposure times calculated in seconds are: 1/15s, 1/30s, 1/60s, 1/125s, 1/250s, 1/500s, 1/1000s, 1/2000s and 1/4000s. For each exposure the left and right views are rectified [Bra00]. If the images are rectified to an epipolar geometry with corresponding horizontal lines, a 1D search for stereo matching can be performed, otherwise a 2D search is required. The HDR image of each stereo view is computed using the corresponding multi-exposed rectified images for that view following the approach in [DM97]. Fig. 3 presents the stereo HDR example as well as three of the nine captured exposures for the left and right views. HDR stereo matching could now be performed on the stereo HDR data to compute the disparity map of the HDR scene.

In view of HDR stereo matching, we seek to capture scenes which are interesting from both the HDR and stereo matching points of view. As mentioned before, we use an optical table with a fixed 150 mm motorized rail to move the camera. With this configuration we cannot achieve window scenes or sky HDR scenes in our current lab environment. To generate a variety of indoor HDR scenes, we use different light sources and objects with different sizes, material properties, and textures. Placing some objects close to strong light source causes under-exposed areas behind them. Capturing the light source as well as under-exposed regions in one image enables us to reach a high dynamic range of luminance. From the stereo matching point of view, we place highly textured as well as low textured objects at different distances from the camera to achieve variations in disparities. Some slanted objects such as the pillow in Fig. 3 will make the stereo matching process more challenging, since many stereo matching algorithms favor front-to-parallel surfaces [BRR11].

Figure 3: *Stereo HDR data set example. First row: left and right views of an HDR pair (tone mapped using Drago TMO [DMAC03] for better display) with a baseline of 150 mm and dynamic range of 8140:1. Second row: left view in the highest, middle and lowest exposures (1/15s, 1/250s, 1/4000s). Third row: same as the second row for the right view.*

2.2. Reference disparity computation using structured light

To generate the reference disparity maps, we use the structured light technique with Gray-codes as used for the Middleburry disparity map computation [SS03] described in Fig. 4. A Panasonic projector PT-AH1000AE is placed beside the camera (see Fig. 1, top left). It is capable of projecting resolutions up to 1920x1080 pixels with a dynamic range of 500000:1. We also use a desktop PC to control the camera, rail and projector during the capturing process. The projector is used to generate the binary Gray-code patterns that contain black and white (on/off) pixel values.

The first step according to Fig. 4 is to capture all patterned images for both views. To distinguish among n locations, $\log_2(n)$ patterns are required. Each pattern represents one bit plane, and all $\log_2(n)$ patterns together achieve one unique coded value for each pixel. As described in [SS03], using a projector resolution of 1024 x 768, 10 horizontal and 10 vertical patterns are sufficient to uniquely code each pixel in the scene. This sums up to 40 patterned images, 20 for each of the left and right views. To threshold the pixels more reliably into on or off categories, patterns and their *inverse patterns* are being used, which doubles the image numbers into 80. For more accurate decoding, we capture the same patterns with two different exposures, which results in 160 patterned images to be processed.

As in the stereo HDR capturing process, we first need to rectify the corresponding left and right views for all patterns and exposures [Bra00]. Using the decoded Gray-code patterns, we assign a unique label to each pixel of the scene. Corresponding pixels in the left and right view are found using the unique codes at each pixel (in 1D search, since images are rectified). The result of this correspondence process is called *view disparity*. Using view disparities and code labels, projection matrices for the illumination (pattern) sources can be determined. Reprojecting the code labels into the two view geometry results in *illumination disparities*. By combining all disparities, a final accurate disparity map is achieved. More details about the process can be found in [SS03].

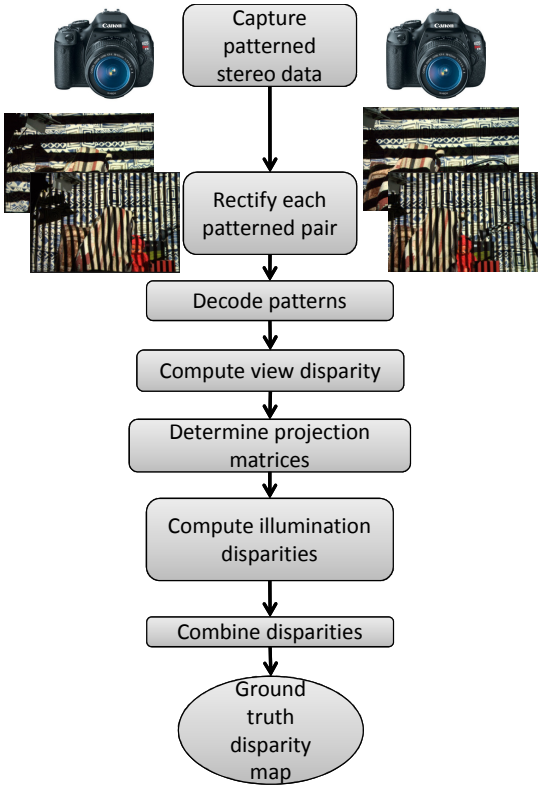

Figure 4: *Ground truth computation process [SS03].*

2.3. Post processing

Once the disparity map has been computed as described in 2.2, appropriate post processing is required. It is suggested in [SS03] to use some interpolation method to fill small holes in the ground truth disparity while no solution for recovering large holes is offered. Small holes are caused by noise or wrongly matched pixels which are usually easy to repair or fill in. Large holes appear in (1) areas that are shadowed under all illumination patterns and (2) specular surfaces.

We apply the post processing in two separate phases to, (1) fill small holes and (2) to cover big holes which are the result of shadows in the scene. It should be noted that in this context shadows refers to areas that do not receive any projection pattern, and not to shadowing effects related to other illumination sources, such as lamps, inside the HDR

scene. The latter type of shadows, however, may pose problems in stereo matching due to poor texture and noise. All holes have been marked as invalid disparities during a left-right consistency check to be post processed later. We repair invalid pixels by using their corresponding disparity value in a morphologically closed version of the current disparity map [Vin93]. Any other filtering method with the same purpose could be used as well.

3. Results and discussion

First results of our processing are shown in Fig. 5. Areas that are shadowed under all illuminated patterns create large holes which cannot be easily removed or filled in. In the bottom row of Fig. 1 a shadowed (non illuminated) area in the top left corner of the scene is visible, which is caused by the lamp. For better visibility, shadow regions are illustrated in the fully illuminated (using projector) image shown in Fig. 5 (a). In these regions a lack of illumination codes caused unknown disparity values (holes) in the disparity result (see Fig. 5 (b)). At the moment, we manually mark these regions and fill in the holes with values in the closest Euclidean distance. The final disparity map generated in the mentioned way from the left point of view is shown in Fig. 5 (d). It should be noted that this is just a preliminary result of our ongoing research.

For comparing the calculated reference disparity of the HDR scene to state of the art stereo matching results, we show the disparity map achieved from the fast cost-volume filtering approach [HRB*13]. Fig. 5 (e) shows the disparity calculated from the middle exposed stereo images and Fig. 5 (f) represents the disparity computed from the Drago tone mapped [DMAC03] stereo pairs. It is visible that the disparity information gained from the structured light technique is of better quality than the results of the stereo matching approach on LDR and tone mapped stereo images.

4. Conclusion and future work

As part of ongoing work towards the generation of stereo HDR data sets and ground truth, we explained the hardware setup and visualized the data set generation and disparity computation approach step by step on an example scene. In principle, the chosen structured light approach is capable of delivering high-quality disparity maps that can be used as reference for stereo matching. However, our HDR test scene has also demonstrated the problem of missing disparity information caused by lamps placed in the scene foreground, which block the light pattern emitted by the projector. Since such light sources constitute an important component of meaningful HDR indoor test scenes, suitable ways for alleviating this problem (for example, by using a second structured light projector) should be explored. As future work we plan to generate more data sets with improved reference disparity maps and publish them to provide researchers involved in the HDR stereo area with standard HDR stereo ground truth disparity maps to evaluate their methods. Changing the hardware setup, we could also capture stereo HDR images of sky scenes or window scenes.

Figure 5: *Reference disparity map computation. (a): Full projected scene, with non-illuminated regions (shadows are marked black) specified; (b): computed reference map before manual post processing (shadow holes); (c): left LDR mid-exposed image; (d): calculated reference disparity after post processing; (e): LDR stereo matching result corresponding to (c); (f): tone mapped stereo matching result corresponding to top row of Fig. 3.*

Acknowledgements

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

References

[AYG13] AKHAVAN T., YOO H., GELAUTZ M.: A framework for HDR stereo matching using multi-exposed images. In *HDRi2013 - First International Conference and SME Workshop on HDR imaging* (2013), pp. 1–4. 1, 2

[BLV*12] BONNARD J., LOSCOS C., VALETTE G., NOURRIT J., LUCAS L.: High dynamic range video acquisition with a multiview camera. In *SPIE* (2012), vol. 8436, pp. 84360A–1 – 84360A–11. 1

[Bra00] BRADSKI G.: The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000). 2, 3

[BRG*14] BÄTZ M., RICHTER T., GARBAS J., PAPST A., SEILER J., KAUP A.: High dynamic range video reconstruction from a stereo camera setup. *Signal Processing: Image Communication 29* (2014), 191–202. 1

[BRR11] BLEYER M., RHEMANN C., ROTHER C.: PatchMatch

stereo - stereo matching with slanted support windows. In *BMVC* (2011), pp. 1–11. 2

[CL08] CHOI J., LEE K. M.: Accurate stereo matching using pixel response function. In *IPIU Workshop* (2008), pp. 1–16. 1

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH* (1997), pp. 369–378. 2

[DMAC03] DRAGO F., MYSZKOWSKI K., ANNEN T., CHIBA N.: Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum 22* (2003), 419–426. 3, 4

[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. 4

[RLIA13] RAMIREZ R., LOSCOS C., IGNACIO M., ARTUSI A.: Patch-based registration for auto-stereoscopic HDR content creation . In *HDRi2013 - First International Conference and SME Workshop on HDR imaging* (2013), pp. 1–4. 1

[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

[SDBRC13] SELMANOVIĆ E., DEBATTISTA K., BASHFORD-ROGERS T., CHALMERS A.: Generating stereoscopic HDR smages using HDR-LDR image pairs. *ACM Trans. Appl. Percept. 10*, 1 (2013), 3:1–3:18. 1

[SDBRC14] SELMANOVIĆ E., DEBATTISTA K., BASHFORD-ROGERS T., CHALMERS A.: Enabling stereoscopic high dynamic range video. *Signal Processing: Image Communication 29* (2014), 216–228. 1

[SFPL10] SALVI J., FERNANDEZ S., PRIBANIC T., LLADO X.: A state of the art in structured light patterns for surface profilometry. *Pattern Recognition 43*, 8 (Aug. 2010), 2666–2680. 2

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

[SS03] SCHARSTEIN D., SZELISKI R.: High-accuracy stereo depth maps using structured light. In *CVPR* (2003), pp. 195–202. 2, 3

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

[Vin93] VINCENT L.: Grayscale area openings and closings, their efĭňĄcient implementation and applications. In *EURASIP Workshop on Mathematical Morphology and its Applications to Signal Processing* (1993), pp. 22–27. 4