

# Practical-HDR: A Simple and Effective Method for Merging High Dynamic Range Videos

Ezgi Akçora  
Department of Computer  
Engineering, METU, Turkey  
ezgiakcora@ceng.metu.edu.tr

Francesco Banterle  
Visual Computing Laboratory,  
ISTI-CNR, Italy

Massimiliano Corsini  
Visual Computing Laboratory,  
ISTI-CNR, Italy  
corsini@isti.cnr.it

Ahmet Oğuz Akyüz  
Department of Computer  
Engineering, METU, Turkey  
akyuz@ceng.metu.edu.tr

Roberto Scopigno  
Visual Computing Laboratory,  
ISTI-CNR, Italy  
roberto.scopigno@isti.cnr.it

## ABSTRACT

We introduce a novel algorithm for obtaining High Dynamic Range (HDR) videos from Standard Dynamic Range (SDR) videos with varying shutter speed or ISO per frame. This capturing technique represents today one of the most popular thanks to the availability and the low cost of the equipment required; i.e., an off-the-shelf DSLR camera. However, naively merging SDR frames into an HDR video can produce artifacts such as ghosts (when the scene is dynamic), and blurry edges (when the camera moves). In this work, we present a straightforward, easy to implement, and fast technique that produces reasonable results in short time. This is key for having quick previews of the captured videos without waiting for a long processing time. This is extremely important especially when capturing videos on modern mobile devices such as smart-phones and/or tablets.

## CCS Concepts

•Computing methodologies → Image and video acquisition; Computational photography; Image processing;

## Keywords

High Dynamic Range Imaging, Image Alignment, HDR Video

## 1. INTRODUCTION

Considering the real world luminance range and the capability of modern cameras to preserve scene details, the difference between human eye visualization and photographs has been a major problem in both computer graphics and signal processing communities. In its most basic form, the aim of HDR imaging is to get closer to the luminance range

that we observe with our visual system in everyday life [10, 1].

Following this idea, many methods for capturing HDR content have been proposed both with software and hardware solutions. The latter case is presently omitted due to two main reasons; the high costs for dedicated HDR cameras. Moving to software solutions for HDR videos, one of the most popular methods is to capture a SDR video with varying shutter speed or ISO; i.e., capturing a short and long exposure frames. Then, these two or more frames are merged together into a single HDR frame. However, a dynamic scene (i.e., moving people and/or objects) is challenging for merging because ghosts; i.e., faded moving objects, may appear.

Even though there are several methods for removing ghosts; i.e. deghosting, these solutions are typically extremely computationally expensive. As the primary motivation of our study, we set out to develop a straightforward to implement and integrate, fast, and at the same time an effective HDR video deghosting method. Considering the speed of the algorithm, the results are aimed to be acceptably free of artifacts. In other words, the speed and quality terms are tried to be kept in balance.

Our algorithm employs three basic steps: **global alignment** to stabilize handheld videos; as described in Section 3.1, **deghosting** to remove ghosting artifacts; as described in Section 3.2, and **merging**; as described in Section 3.3, to merge SDR frames aligned into a HDR frame. Comparing with the other deghosting methods in the literature the computational cost of this algorithm is substantially low; see [13] for a detailed review.

## 2. RELATED WORK

In last years, an enormous number of algorithms for both HDR image and video deghosting have been proposed. Recently, Tursun et al. [13] reviewed approximately 50 image deghosting methods for both images and videos. This survey classifies them into four major categories: global exposure registration methods, moving exposure removal methods, moving object selection methods, and finally moving object registration methods. Among these algorithms for images, Sen et al.'s patch-based HDR image synthesis framework [11] was found to produce the highest quality results according to a subjective experiment performed by the same

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CVMP '16 December 13–16, 2016, London, UK

© 2016 ACM. ISBN 123-4567-24-567/08/06...\$15.00

DOI: 10.475/123\_4

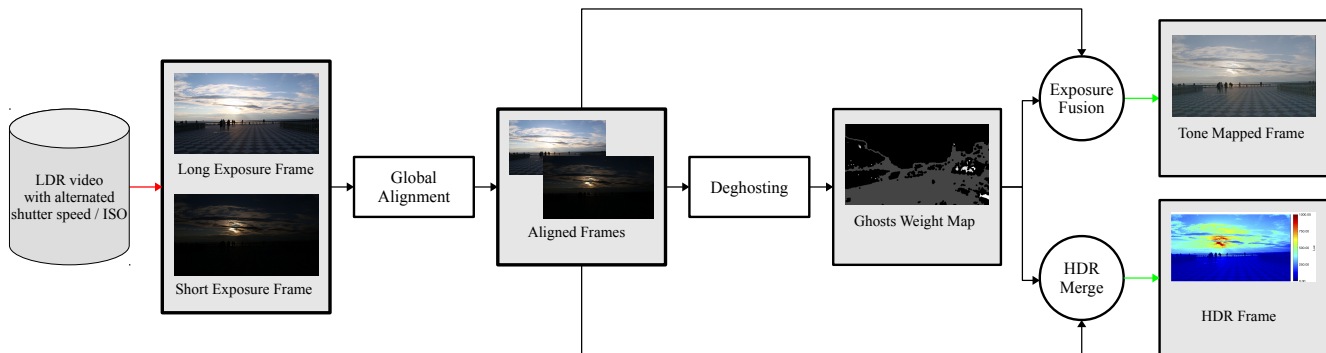


Figure 1: The pipeline of the proposed algorithm. The original images are copyright of Tomasz Sergej.

authors, a finding supported by other studies as well [4].

**Special Hardware.** Many researchers have worked on different hardware setups for obtaining HDR videos without alignment and ghost artifacts or minimizing these issues. They have proposed several solutions such as single-aperture and multiple sensors [?, ?], multiple cameras [2] (as in light-fields or stereo capture), or specialized HDR sensors [?]. These hardware setups produce high quality results, but they are very expensive and require accurate calibration.

**Enhancing SDR videos.** A different approach is to enhance SDR videos using reference HDR photographs. Bhat et al. [?] proposed a 3D reconstruction of the scene from the video using structure from motion and frame depth estimation for transferring HDR information from HDR images onto SDR videos. This method achieves high quality and allows the camera to be moved inside the scene, but it is limited to static scenes (no moving people/objects) and is computationally expensive. On the other hand, Banterle et al. [?] introduced a computationally efficient technique that replaces the background of a SDR video with an HDR background. However, the camera in this method, which allows a scene to be dynamic, has to be static.

**Spatially varying shutter speed/ISO.** A solution, which greatly reduces ghosts, is to vary shutter speed or ISO per pixel [?] similarly to the Bayer pattern for colors. In doing so, spatial resolution is traded for dynamic range resolution. However, we can still extract high quality images by employing sophisticated reconstruction techniques [?, ?, ?].

**Temporally varying shutter speed/ISO.** Alternatively, it is possible to obtain an HDR video by using an off-the-shelf digital camera that allows us to vary shutter speed or ISO between different frames. The pioneering works in this direction used optical-flow based solutions to register the individual pixels between neighboring frames [6, ?]. At the time of writing, the state-of-the-art method, on the other hand, is known to be Kalantari et al.’s patch-based HDR deghosting algorithm [5]. This algorithm combines Sen et al.’s patch-based framework with Kang et al.’s optical-flow based approach by limiting patch searches around the pixel positions indicated by the optical flow. Despite this optimization, the algorithm is still computationally very expensive and requires several hours to generate a short video at HD resolution. Mangiat and Gibson [7, 8] proposed an alternative technique, which is computationally cheaper, but is prone to more artifacts. The computational improvement

in this algorithm is due to a block-based search between two consecutive frames rather than pixel or patch-based searches as in Kang et al. [6] or Kalantari et al. [5]. However, the usage of blocks also limits the dynamic range within a block and the output may appear blocky in certain regions.

In our work, we tackle the problem of obtaining HDR videos starting from SDR videos with temporally varying exposure time or ISO. The premise of our work is that by combining simpler methods that are originally designed for images, one can obtain an HDR video production algorithm that is both fast and robust. To this end, we decided to combine Pece and Kautz’s bitmap movement detection (BMD) algorithm [9] with Ward’s image alignment method [14].

The BMD algorithm is based on the well-known Median Threshold Bitmap (MTB) descriptor by Ward [14]. It starts by creating an MTB for each input exposure (the input exposures must be globally aligned first). Then, the bit values (either 0 or 1) are accumulated for each pixel. If the result is different from 0 or  $N$ , where  $N$  is the number of input exposures, the corresponding pixel is marked as a potential ghost region. Through subsequent morphological operations, which eliminate noisy markings, the final motion map is obtained. These motion regions are then taken from the best single exposure.

### 3. ALGORITHM

The aim of our work is to provide a novel method for converting a high frame rate SDR video with varying either shutter speed or ISO per frame into an HDR video at a standard frame rate; e.g., 24 fps.

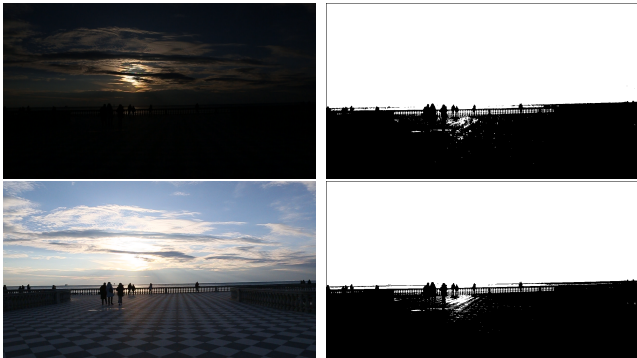
As previously stated, our main goal is to achieve an algorithm that is straightforward to implement, effective, and computationally fast. Previous works have shown high quality HDR reconstruction, which may take several minutes per frame [5] using MATLAB. This may be not ideal for real-time processing or for having a quick preview of the captured footage on a site.

The proposed algorithm is summarized in the following three steps:

- Global alignment;
- Deghosting;
- HDR merge and/or exposure merge.

Figure 1 shows the full pipeline of the proposed algorithm.

### 3.1 Global Alignment

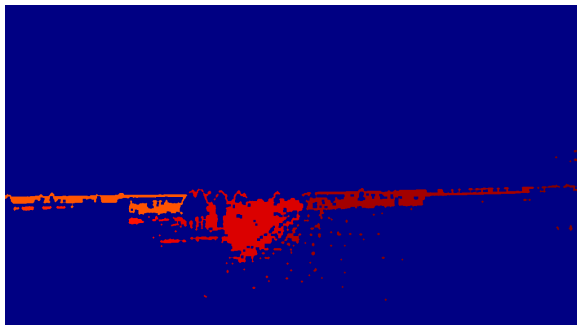


**Figure 2: An example of MTB: On the left side input frames at different shutter speeds. On the right side, the corresponding MTB results. The original images are copyright of Tomasz Sergej.**

The first step of the algorithm is to globally align consecutive frames at different exposures. In doing this, we assume that the 2-3 nearby frames do not change rapidly or they share many common features. The focus of this work is on videos where the shutter speed or the ISO can vary every two or three frames.

Many solutions [13] are available for the global alignment of SDR images at different exposure values. After testing different techniques, we found that the MTB algorithm [14] is the preferred solution in our context, see Figure 2. Although its primary goal is to correct translational misalignments, it can be easily extended to handle small rotations as well [10]. Furthermore, it is computationally efficient; it takes a few seconds for a full HD resolution image in MATLAB.

We also experimented using homographies computed using local features such as SIFT [12]. However, this approach failed for many cases, because the difference in f-stops between the short and long exposure frames is too large and local feature can not be detected in some cases. This issue usually happens every 10-20 frames, and it can be solved using tracking; e.g., a Kalman filter. Nevertheless, tracking increases complexity in our approach in an unnecessary way.



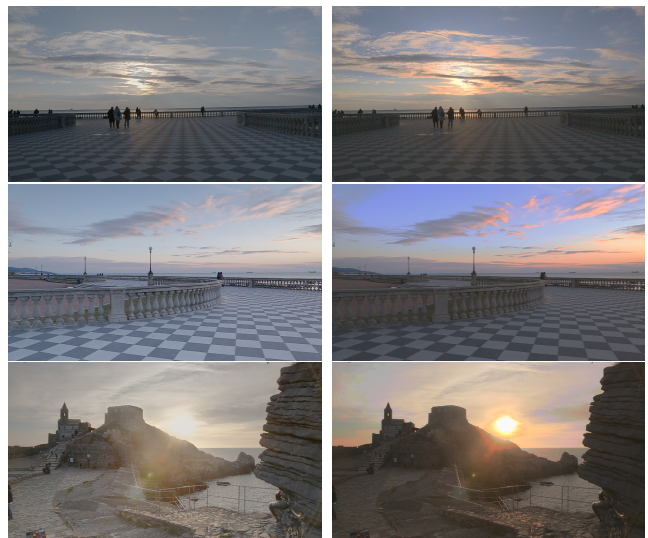
**Figure 3: An example of MDB using frames in Figure 2. Blue pixels are static. The original images are copyright of Tomasz Sergej.**

Typically, MTB algorithm produces high quality results

when the difference between exposure values is between 1-2 f-stops. However, our input videos typically have a difference around 3 f-stops. To improve robustness, we check if the translation magnitude and rotation angle of the alignment are over a certain threshold. If this happens, we clamp the values to the ones of the previous iteration. From our experiments, we found out that a value of  $\pm 48$  pixels for translation magnitude and 15 degrees for rotation work well as thresholds for 720p resolution videos.

### 3.2 Deghosting

After the global alignment step, the 2-3 nearby frames at different exposure values are globally aligned. If we reconstruct an HDR frame out of these frames, we may obtain some ghosts; i.e., moving parts of the scene which fade out. To solve this issue, we experimented with different deghosting techniques for still images [13]. From these experiments, we found out that the Pece and Kautz’s BMD method [9] typically provides reasonable quality results and it is computationally fast to compute, see Figure 3. This method is based on MTB, and it outputs a per-pixel weight map for each exposure, where a high weight is given to reliable pixels (i.e., no movement) in that frame. Note that this algorithm requires two thresholds to be set; i.e., the sizes of a dilation and an erosion filters. In our experiments, we found out that selecting global thresholds per scene cut produces satisfactory results. However, a temporally-varying automatic threshold could produce optimal results. We also tested other algorithms such as Gallo et al. [3] and one of its variants [10] (see Chapter 5, Section 9), which can also achieve similar quality results. We opted to keep the BMD algorithm in our work because it does not require a camera response function, and it is more straightforward to implement.



**Figure 4: Visual comparisons between our method (left) and Mangiat and Gibson’s algorithm (right) for the three captured sequences (please watch additional videos materials). The original images are copyright of Tomasz Sergej.**

### 3.3 HDR Merge

After global alignment using MTB and the computation of ghost weights for each exposure frame using the BMD method, we can finally obtain an HDR video. At this point, we can either merge frames at different exposure values using Debevec and Malik’s method [?] to obtain a radiance map or to apply exposure fusion to directly obtain a detail-rich SDR frame [?]. Note that the ghost weights can be plugged into both methods such that whenever a region is detected to be dynamic, the best exposed SDR frame is used for that region.

## 4. RESULTS

We compared our method against Mangiat and Gibson’s algorithm [8] using the default parameters reported in that paper. This method has similar goals as our work. Although Kalantari et al. [5] works on similar input videos and produces high quality results without noticeable visual artifacts, we have not tested against it or other high quality algorithms because the computational times is very high w.r.t our method and the Mangiat and Gibson’s algorithm; see also Section 2 for further details about processing time.

### 4.1 Visual Comparisons using Real Data

We applied our algorithm and Mangiat and Gibson’s one to three  $1280 \times 720$  (720p) video sequences captured using a Canon 5D Mark III with Magic Lantern firmware at 48fps<sup>1</sup>. Figure 4 shows some frames from these reconstructed sequences (please review the additional material for video comparisons). Apart from color differences, both algorithms provide reasonable HDR reconstruction. While Mangiat and Gibson’s method can show some misalignments, our method typically has local flickering due to the fact that the deghosting happens per frame.

### 4.2 Objective Evaluation

We also compared our method with Mangiat and Gibson’s algorithm against ground truth using HDR-VDP2 [?, ?] to understand which method is closer to a ground truth. We modified Mangiat and Gibson’s algorithm [8] because it works (in the final stages) and outputs tone mapped images using the Reinhard et al.’s global operator [?]. However, we can easily obtain HDR frames by inverting Reinhard et al.’s global operator. Note that this process does not lead to loss of information because it happens before writing a quantized tone mapped image.

We generated multiple SDR videos from the HDR ground truth at varying shutter speeds per frame using the following procedure:

1. We selected two shutter speed values; i.e. short and long exposure values, for maximizing the dynamic range;
2. We created a stream with varying shutter speed per frame using the short and long exposure values with gamma encoding 2.2;
3. We ran our method and Mangiat and Gibson algorithm on the generated SDR video sequence with varying shutter speed per frame, obtaining two reconstructed HDR videos;

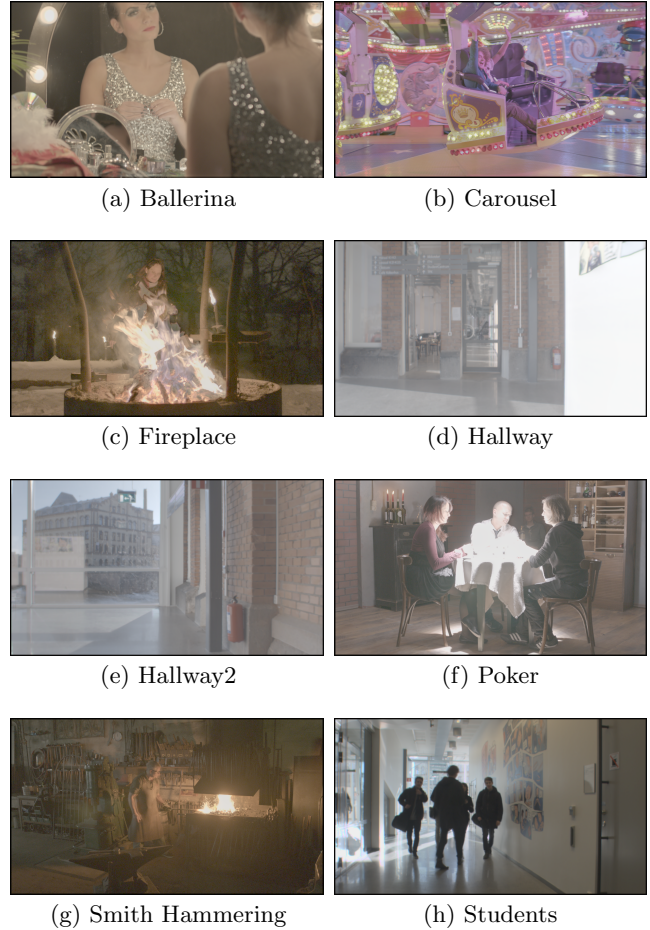


Figure 5: HDR video sequences (tone mapped for visualization) used in our experiments. Original frames in (d), (e), and (h) are copyright of Jonas Unger. Original frames in (a), (b), (f), and (g) are copyright of Jan Fröhlich.

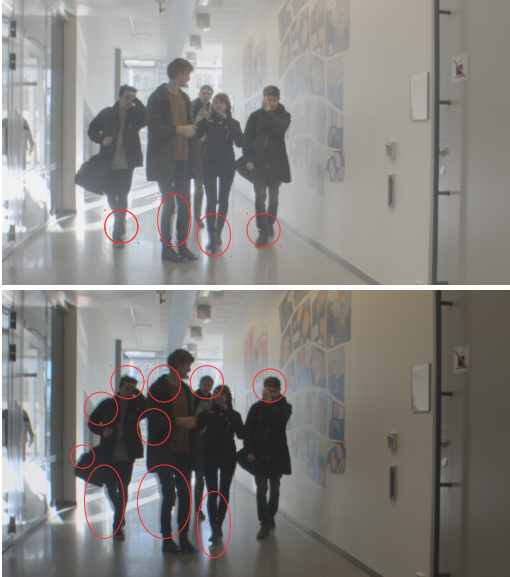
Sequence	Our	Mangiat and Gibson
Ballerina	<b>57.26</b>	54.79
Carousel	<b>58.93</b>	56.07
Fireplace	<b>91.21</b>	54.97
Hallway	54.90	<b>60.22</b>
Hallway2	56.98	<b>59.42</b>
Poker	<b>92.99</b>	64.20
Smith Hammering	<b>66.01</b>	55.58
Students	<b>68.00</b>	59.19
<b>Average</b>	<b>68.28</b>	58.05

Table 1: HDR-VDP2  $Q$  values for our method and Mangiat and Gibson’s one; bold means winner. A  $Q = 100$  means the best quality and gets lower for lower quality; i.e., the higher the better. Note that  $Q$  can be negative in case of very large differences.

<sup>1</sup>www.magicalantern.fm

- For each reconstructed video, we ran HDR-VDP2 (with standard parameters) per frame comparing the reconstructed current frame against the original corresponding HDR frame of the input video.

Note that the ground truth HDR videos were captured using an HDR videocamera<sup>2</sup> and a camera rig<sup>3</sup>.



**Figure 6: An example of comparisons between our method (top) and Mangiat and Gibson’s algorithm (bottom) for the *Students* sequence. Artifacts are circled, note that our method produces less artifacts in comparison. Original frames are copyright of Jonas Unger.**

Table 1 and Table 2 shows the results of this evaluation. Table 1 reports quality correlate values; the higher the better. From this table, we can notice that our method has a higher quality than Mangiat and Gibson’s algorithm in most cases. Table 2 reports a single valued probability of detection (i.e., to detect a change in the image), the lower the better. In this case, our method still performs better than Mangiat and Gibson’s method, but in some scenes they are equal. This may be due to the fact that those scenes have on overall higher absolute luminance values than others, which are pretty dark. Therefore, we may have less masking due to dark.

### 4.3 Timings

We timed the reconstruction time of both algorithms, which we implemented in MATLAB, using a 2013 iMac with a 3.1Ghz Intel Core i7, and 16GB of RAM. While our method takes 2.5 seconds on average for generating a frame, Mangiat and Gibson’s algorithm takes around 125-250 seconds on average (depending on the size of the search radius for motion estimation) for one frame. We noticed that around 90% of computations for this algorithm are spent in the motion estimation, which are the bottleneck of the algorithm.

<sup>2</sup>Linköpings University dataset - <http://www.hdrv.org>

<sup>3</sup>University of Stuttgart dataset - <https://hdr-2014.hdm-stuttgart.de/>

Sequence	Our	Mangiat and Gibson
Ballerina	0.9778	0.9778
Carousel	<b>0.8670</b>	0.9833
Fireplace	<b>0.0675</b>	0.9833
Hallway	0.9939	0.9939
Hallway2	0.9938	0.9938
Poker	<b>0.0016</b>	0.9793
Smith Hammering	<b>0.6861</b>	0.9714
Students	0.9920	0.9920
<b>Average</b>	<b>0.6975</b>	0.9844

**Table 2: HDR-VDP2  $P(X)$  values for our method and Mangiat and Gibson’s one; bold means winner, and green means equal.  $P(X) \in [0, 1]$  is a single valued probability of detecting changes; i.e., the lower the better.**

However, even though a very optimized motion estimation is employed we can safely state that our algorithm is still computationally the fastest between the twos. Note that, all steps of our method, which are based on MTB, can be implemented on the GPU [?].

We also want to highlight that our method is also significantly faster than other state-of-the-art algorithms; for example the one of Kalantari et al. [5] requires more than three minutes per frame.

### 4.4 Limitations

The main limitation of our method is that it cannot remove all ghosts, but most of them in a reasonable time. If we compare it to Mangiat and Gibson’s method, we can notice that our method produces less ghosts artifacts; as shown in Figure 6. The main reason why we have ghosts is due to the fact we use global parameters for Pece and Kautz’s method [9] instead of per frame parameters.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a novel method for reconstructing HDR videos from a SDR sequence with varying shutter speed or ISO per frame. We have shown that our method is computationally fast and can achieve high quality results given the fractional computational time compared to a state-of-the-art algorithm with similar goals. Furthermore, all building blocks of the algorithm such as the global alignment and deghosting part can be implemented on the GPU with ease, which may pave the way to a real-time implementation.

For future work, we would like to implement our method on the GPU; the main building block; i.e., MTB can be easily implemented on graphics hardware [?]. Furthermore, we believe that we could greatly reduce ghosts using adapting parameters for BMD.

### Acknowledgements

We thank Jonas Unger, Radek Mantiuk, Tomasz Sergej, and Jan Fröhlich for providing datasets of HDR videos. We also thank the anonymous reviewers for their feedback and constructive comments.

## 6. REFERENCES

- [1] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced High*

*Dynamic Range Imaging: Theory and Practice*. CRC Press, (AK Peters Ltd), 2011.

- [2] Michel Bätz, Thomas Richter, Jens-Uwe Garbas, Anton Papst, Jürgen Seiler, and André Kaup. High dynamic range video reconstruction from a stereo camera setup. *Signal Processing: Image Communication*, 29(2):191–202, 2014.
- [3] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. *IEEE International Conference on Computational Photography (ICCP)*, April 2009.
- [4] Kanita Karaduzovic Hadziabdic, Jasminka Hasic Telalovic, and Rafal Mantiuk. Comparison of deghosting algorithms for multi-exposure high dynamic range imaging. In *Proceedings of the 29th Spring Conference on Computer Graphics, SCCG '13*, pages 021:21–021:28, New York, NY, USA, 2013. ACM.
- [5] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B Goldman, and Pradeep Sen. Patch-based high dynamic range video. *ACM Trans. Graph.*, 32(6), 2013.
- [6] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Trans. on Graphics*, 22(3):319–325, 2003.
- [7] Stephen Mangiat and Jerry Gibson. High dynamic range video with ghost removal. In *SPIE Optical Engineering+ Applications*. Int. Society for Optics and Photonics, 2010.
- [8] Stephen Mangiat and Jerry D. Gibson. Spatially adaptive filtering for registration artifact removal in HDR video. In *18th IEEE International Conference on Image Processing, ICIP 2011, Brussels, Belgium, September 11-14, 2011*, pages 1317–1320, 2011.
- [9] Fabrizio Pece and Jan Kautz. Bitmap movement detection: HDR for dynamic scenes. In *Visual Media Production (CVMP), 2010 Conference on*, pages 1–8. IEEE, 2010.
- [10] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, second edition, 2010.
- [11] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6):203, 2012.
- [12] Anna Tomaszewska and Radoslaw Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *15th Int. Conf. Central Europe on Computer Graphics, Visualization and Computer Vision*, 2007.
- [13] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. The state of the art in hdr deghosting: A survey and evaluation. In *Computer Graphics Forum*, volume 32, pages 348–362, 2015.
- [14] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of graphics tools*, 8(2):17–30, 2003.