A METHOD AND SPECIFIC CHALLENGES FOR OBTAINING A HDR HYPERLAPSE FROM STILL IMAGES

Vlad CODÎRLĂ, Aurelia CIUPE, Bogdan ORZA, Şerban MEZA
Technical University of Cluj-Napoca, Cluj-Napoca, Romania
Memorandumului Street Nr. 28, Tel. +4 0264 401 200

Abstract: Creating an in-motion time-lapse while moving the camera over significant distances (herein named as a hyperlapse) through an environment with high variance in illuminance poses multiple challenges. Normally, one would have to choose between changing the camera settings between frames, thus causing distorted artifacts, or dealing with the compromise of selection between under-exposed or over-exposed areas during movement, that results in lower fidelity of the captured scene. The current work proposes an approach according to which each frame is batch-captured using varying exposure levels. These exposures are then combined for each frame of the hyperlapse to create a HDR (high dynamic range) hyperlapse suited for display on LDR (low dynamic range) equipment, thus combining two traditionally separate areas of research. Multiple algorithms for combining the images have been experimentally tested and compared based on several metrics: blending workflow, computational requirements and empirical quality assessment.
Keywords: *HDR, high dynamic range, hyper lapse, time-lapse, experimental acquisition*

I. INTRODUCTION AND PROBLEM STATEMENT

Time-lapse photography has long been used in the presentation of scientific, commercial and artistic works by making evident events that take place so slowly so as to be imperceptible in real-time. Hyperlapses have proven to be an effective way of presenting a large amount of visual content quickly. They are currently being used in commercial applications to provide virtual experiences of widespread locations such as an entire city. The hyperlapse is being increasingly used in the consumer market by mobile device owners looking to take advantage of its unique way of presenting content.

When recording a time-lapse sequence, it is imperative to maintain constant capture settings (exposure, aperture, focus) between successive still images, otherwise the resulting sped-up sequence will exhibit flickering [1]. However, in the case of a hyperlapse, the recording settings that ensure optimal exposure for one frame might not match the following frames, especially when the camera is moved along a path. Even if these settings could be adapted, the environment exihibits a high degree of uncertainty in many situations, particularly when working outdoors. Since a time-lapse (or hyperlapse) requires a capturing duration of several hours, variations in the light levels are inevitable for most outdoor scenes. These changes can be gradual, such as the changing position of the sun, or sudden and repetitive, such as clouds moving in and out of the path of direct sunlight.

Another issue is the stability of the resulting video sequence. As the process involves a camera movement along a predefined path, minor framing variations are amplified when sped-up. The following work proposes a method for obtaining a moving HDR video sequence from still images captured with a DSLR camera. Image stills have been taken on a cobbled street, to address the above-stated challenges through making consistent framing much more difficult.

In section 2, current research work and commercial applications in the fields of hyperlapse photography and HDR video are presented, following 2 main directions: methods of constructing hyperlapses and HDR video rendering solutions.

Section 3 describes the proposed workflow for image acquisition and the dedicated environment (hardware, software) used to compute the hyperlapse rendering. The test medium is discussed in section 3.1. Section 3.2 presents a comparison of algorithms for obtaining HDR images while also addressing the issue of batch image processing, required when working with hundreds or thousands of still images. In Section 3.3 it is shown how, after the tone-mapping processes were applied, hyperlapse sequences were constructed and motion stabilization work was done to mitigate the inherent problems that appear during capture. A comparison between the computational loads of two different tone mapping algorithms is presented in section 4. Section 5 will showcase the final output and an input/output comparison of the hyperlapse sequence with proposed improvements. In section 6 are presented conclusions.

II. RELATED WORK

In-motion HDR content generates interest in multiple research and commercial areas. In creating hyper-lapse sequences with constant exposure, the general approach has been to record video sequences from which the hyperlapse sequence is then constructed. There have been two approaches used in this case, both of which have been implemented as Apps for mobile devices.

The concept proposed by the Microsoft Research division [2], [3] first constructs a virtual camera path and viewpoint, after which parts from successive frames are combined to create one frame of the hyper-lapse along the virtual path. Even within the case of extreme camera

Manuscript received February 15, 2017; revised March 16, 2017

movements, the proposal delivered accurate results. The other approach has been implemented by Instagram Engineering in their proprietary Hyperlapse© App [4]. In this case, the frames are warped using information from the mobile device gyroscope to create a smoothly moving sequence in the center of the frame, after which the edges are cropped out. In this case, significant camera movement will result in an excessively cropped sequence with a poor image resolution.

A different methodology to creating the effect of a hyperlapse is through the sourcing of images on the Internet [5]. This involves using a large number of pictures taken of, for example, a monument from different viewpoints and different time periods. These images are aggregated and a virtual path is constructed from the various viewpoints, which can then be viewed sequentially, simulating the hyperlapse effect of moving through space and a large time domain. The drawbacks to this method are the lack of control in choosing viewpoints, meaning a smooth path is difficult to achieve, as well as the lack of control upon the lighting conditions and, because pictures could be taken even years apart, unintended changes will likely appear in the captured scene [6].

A critical process in creating automated hyperlapses is video stabilization. Video stabilization algorithms designed specifically for hyperlapse sequences have been proposed with differing degrees of success [7]. The principal areas of difficulty that emerge are error accumulation when processing longer sequences, the difficulty of of accounting for imperfect capture conditions, particularly the distance between frames, and processing computational loads.

Another area of interest is the rendering of real-time HDR video. A fundamental question regarding HDR content, separate from its technical advantages, is about how the user perceives it subjectively, which leads to commercial viability considerations. In this sense, empirical studies concluded that participants showed a statistically demonstrable preference towards the HDR version of a video sequence as opposed to its LDR version [8]. The approaches taken in this field vary: one can use multiple cameras positioned side-by-side recording at alternating exposure levels and construct HDR footage from the overlapping sequences [9]. Another solution is to use multiple sensors to record incoming input at different exposure values with the help of beam-splitters [10], [11]. There is the approach used in the HDRx technology developed by the RED Camera manufacturers, to capture two frames with alternating exposure within the time frame of one [12]. Another approach is to use a single sensor and a single exposure which is then synthesized into a tone-mapped image [13],[14]. The software-based approach to this method, turning LDR content into HDR has been explored particularly with regards towards its commercial potential [15]. Remastering old content into HDR versions may not only give older content renewed interest in the consumer market, but also incentivize sales of devices capable of displaying true HDR content.

The work of Dario Cali represents a commercial alternative to the proposed solution, as can be seen in the video “This is Brisbane 2014 (HD Timelapse)”, available at www.vimeo.com/daduxio. The method consists of using video processing software on a traditional time-lapse to create the appearance of smooth motion through panning and zooming within the initial sequence, resulting in a simulated hyperlapse. The advantage of this solution is that these movements created using software do not require motion stabilization. On the contrary, working with pans and zooms on the captured images means a loss of resolution in the final sequence. The theoretical distance that the camera can travel is limited by the resolution of the captured images, while using the proposed solution in this article one can move the camera over any distance, creating motion without loss of image resolution.

As opposed to the automated solutions available separately for HDR and hyperlapse technology, the presented work proposes video sequence rendering from successively captured still images, thus being able to combine the HDR and hyperlapse features into a single sequence, a process for which as of yet no automated solution has been developed.

III. PROPOSED METHOD AND WORKFLOW

The setup used for the proposed experiment includes: a Canon EOS 500D camera, a Tamron 28-300mm lens and a Manfrotto 055XB tripod. Several software tools have been used for image processing and video rendering: Adobe Creative Suite (Photoshop, Premiere Pro, After Effects), EnfuseGUI, Performance Monitor, CPUID Hardware Monitor.

3.1 Test medium

Three sets of photos were taken along Potaissa Street in Cluj-Napoca, Romania, covering approximately 200m in length and 3 hours of time (*Figure 1*, *Table 1*). The distinctive feature of the location is the large medieval wall, built in the 15^{th} century, which provided an approximately even amount of shadow and sunlight to the scene, making it suitable for testing HDR aquisition. Another specific feature of the location is the cobbled pavement on the street, whose unevenness posed a challenge in obtaining a smooth and stable sequence of motion. Image sets for successive HDR frames were taken approximately 1 meter apart.

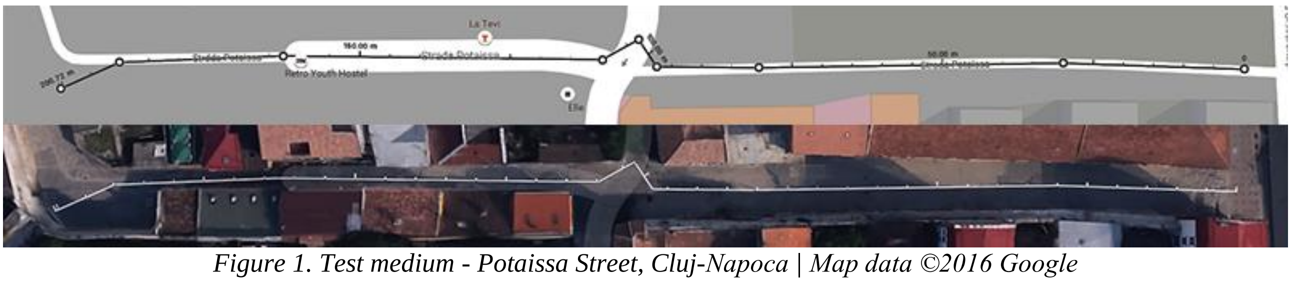

Figure 1. Test medium - Potaissa Street, Cluj-Napoca | Map data ©2016 Google

Table 1. Space-time characteristics of the captured image sets

Metric	Set1	Set2	Set3
Time of day	5:30PM to 6:30PM	11:00AM to 12:30PM	12:30PM to 13:00PM
Length of time covered	1 hour	1 hour	30 minutes
Distance covered	90m	95m	15m

The experiment will test the proposed workflow, summarized in *Figure 2.a, 2.b*, making note of possible improvements.

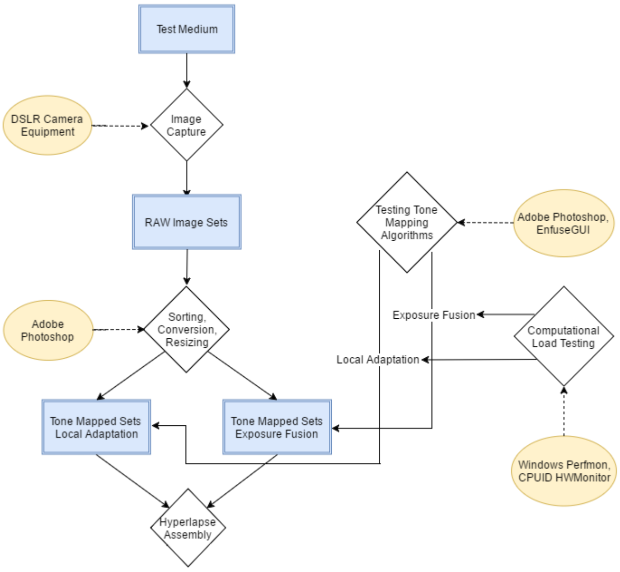

Figure 2.a. Project Workflow Diagram

3.2 Obtaining tone-mapped images

The captured images were first sorted, batch converted from RAW to JPEG format and resized. Several algorithms for tone-mapping HDR images have been empirically compared: Local Adaptation, Equalize Histogram, Exposure and Gamma, Highlight Compression (all implemented with Adobe Photoshop) and the Mertens-Kautz-Van Reeth Exposure Fusion algorithm [16] (implemented with EnfuseGUI). Parameter customization and batch processing capabilities have been considered the main criteria in selecting the algorithm to be furtherly applied.

Local Adaptation (*Figure 3.a*) and Exposure Fusion (*Figure 3.e*) have been selected, due to having the most customization options as well as there being batch processing capability. After configuring several parameter presets, the images were batch processed using Adobe Photoshop for Local Adaptation, making use of a script developed by David Milligan (available at www.github.com/davidmilligan) which allows the application of a tone mapping algorithm to multiple image sets, and EnfuseGUI for Exposure Fusion. A comparison between the computational loads of these processes is presented in *Table 2,3,4 and 5*. One challenge was the inadequacy, for Set1 and Set3, of a single Local Adaptation preset for batch processing, due to drastic light level changes within the image sets (*Figure 4*). Two sets of processed images have been selected for further use as sources for the hyperlapse sequence.

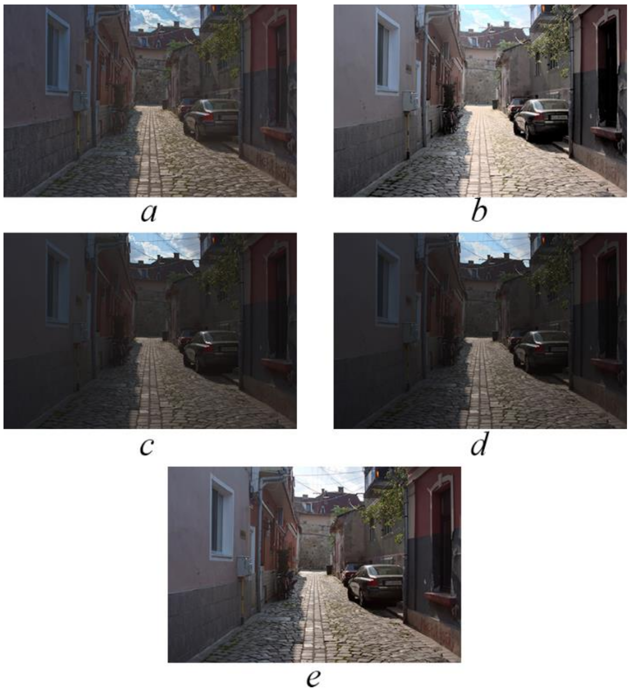

Figure 3. Tone mapping algorithms: a - Local Adaptation, b - Equalize Histogram, c - Exposure and Gamma, d - Highlight Compression, e - Exposure Fusion

Figure 4. Light level changes within same image set

3.3 Assembling the hyperlapse

Consecutive stills of the raw image sets have been stacked as PoCs (Proof of Concept) for 2 hyperlapse sequences. As the current work has an experimental purpose of testing the hyperlapse flow, Adobe Premiere Pro has been found adequate for merging. A two-frame duration was chosen for the still images, short enough to convey motion, while also long enough to allow proper analysis. It can be seen (*Figure 6.a*) that for these stage, the hyperlapses exhibit severely unstable framing. The use of a tripod during capture has been countered by the uneven surface on which the movement has been tested.

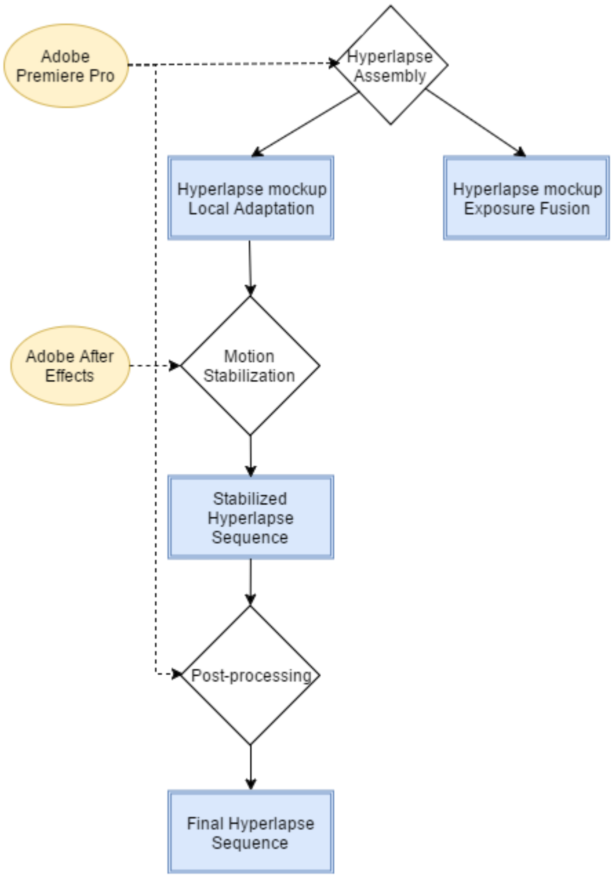

Figure 2.b. Project Workflow Diagram (continued)

Thus, the next step is to stabilize the motion to create smooth transitions between successive frames. As the process is identical, yet non-transferable between the two differently processed hyperlapse sequences, only one was chosen to be taken further forward through the, specifically the one using the Local Adaptation algorithm.

Motion stabilization work was done using Adobe After Effects, using the Stabilize Motion method from the Tracker panel. The effects can be seen in successive frames in *Figure 6*. An additional challenge at this stage was the deliberate panning of the camera up and down during capture of Set1 and Set2. This meant that tracking points had to be manually adjusted for successive frames in order to retain the panning movement.

Once the motion stabilization of the hyperlapse was complete, the sequence was reimported into Adobe Premiere Pro, where post-processing work was done to obtain the final sequence. This consisted of color grading and local contrast enhancement.

IV. RESULTS

The hardware used to test the computational loads of the two tone mapping algorithms chosen at stage 3.2 includes: a Intel® Core™ i7-4770S CPU with a maximum frequency of 3.10GHz, 16GBytes of DDR3 RAM, a AMD Radeon™ HD 7650A 2GBytes Graphics Card as well as a HGST Travelstar™ Z7K500 500GBytes 7200 RPM hard disk for storage.

Memory usage during image processing with the Local Adaptation and Exposure Fusion algorithms is compared in *Table 2*. As shown, the Exposure Fusion algorithm makes 291% more page reads or writes per second compared to Local Adaptation. On the other hand, out of 16GBytes available memory, using Exposure Fusion commits 3.9Gbytes to the process, only 53% of Local Adaptation's 7,4GBytes. Thus, it can be concluded that the Exposure Fusion algorithm uses less of the RAM memory compared to Local Adaptation, but does so more intensively.

Table 2. Average Memory Usage for Set1

Metric	Local Adaptation	Exposure Fusion
Pages/sec	535	1,560
Available MBytes	8,542	12,112

With regards to the CPU, the processor runs at nearly its maximum frequency for both algorithms. However, Exposure Fusion batch processing uses the CPU for almost twice as long, 58% of the running time compared to 30% for Local Adaptation., as shown in *Table 3*.

Table 2. Average Processor Usage for Set1

Metric	Local Adaptation	Exposure Fusion
Processor Frequency (Hz)	2,924	3,040
% Processor Time	29.79	58.37

As in Table 4, Local Adaptation batch processing the hard-disk is in use 9% of the time compared to 13% for Exposure Fusion. The average size of the transfers to or from the disk during operations is 18.3% larger while using Local Adaptation compared to Exposure Fusion. As such, the strain on the hard drive is quite similar in both cases, with the Exposure Fusion algorithm making lower sized operations, but more frequently, compared to the Local adaptation algorithm which consequently makes use of the hard-disk resources over a shorter duration.

Table 3. Average Hard-disk Usage for Set1

Metric	Local Adaptation	Exposure Fusion
% Disk Time	8.87	13
Avg. Disk Bytes/Transfer	133,854	113,153

The total running time for batch processing the three image sets amounts to 14 minutes and 13 seconds for Exposure Fusion and 84 minutes and 30 seconds for Local Adaptation, a 594.4% increase (*Table 5*).

Table 4. Batch Processing Running Times

Tone mapping method	Set1 duration (mm:ss)	Set2 duration (mm:ss)	Set3 duration (mm:ss)
Local Adaptation	36 44	35:23	12:23
Exposure Fusion	5:57	6:06	2:10

Considering all these metrics, it can be seen that as far as computational strain on the hardware is concerned, using the Exposure Fusion algorithm while tone mapping images in batch is much faster but uses system resources more intensively compared to the Local Adaptation algorithm.

V. DISCUSSION AND PROPOSED IMPROVEMENTS

The final hyperlapse sequence exhibits significantly improved apparent dynamic range, despite the compromise of including regions of distorted content, resulted from HDR blending of individual frames 2 causes have been identified for such artifacts: 1) the "ghosting effect", meaning changes in the environment within a set of three differently exposed images used to assamble one HDR frame. 2) improper parameter calibration in the tone mapping algorithms. This is the result of compromising between the efficiency of batch tone mapping and the complexity of using multiple different parameter presets for sections of the hyperlapse with highly different lighting conditions. These results are visualized in *Figure 5.* A potential solution would consist of applying adaptive tone mapping based on detected light levels, color temperature or other metrics. As for the "ghosting effect", an algorithm has been proposed for the correction of HDR video through motion area detection [17]. This consists of detecting the areas of an image where movement occurs between frames and weighting these areas differently during processing.

Motion stability is also greatly improved compared to the

raw mockups, at the expense of image resolution loss_due to cropping. The effect of stabilization can be seen in *Figure 6.* The frames in *Figure 6.b* are presented before final cropping. One possible improvement of the motion stabilization process would be a method for grounding the horizontal axis of the images to the same line, since the current method, being centered around a single tracking path, produces a rocking side-to-side motion effect.

VI. CONCLUSIONS

A solution has been proposed for obtaining a HDR hyperlapse video sequence from still images. The specific challenges of the method and remaining issues have been highlighted together with possible improvement solutions based on experimental evaluation. Manual intervention and time consumption represent the main drawbacks in proposing this method for commercial implementations. However, if optimized, an automated solution to combining HDR and hyperlapse video features would attract much interest for the future.

a *b* *c* *d*

Figure 5. Tone Mapping: a, b - dynamic range expansion; c - ghosting artifacts; d - tone mapping parameter artifacts

Figure 6. Motion stabilization: a - raw sequence, b - stabilized sequence

VII. REFERENCES

[1] M. Rubinstein, C. Liu, P. Sand, F. Durand, and W. T. Freeman, "Motion denoising with application to time-lapse photography," *Proc. 2011 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR '11*, no. 2011, pp. 313–320, 2011.

[2] J. Kopf, M. F. Cohen, and R. Szeliski, "First-person hyper-lapse videos," *ACM Trans. Graph.*, vol. 33, pp. 1–10, 2014.

[3] N. Joshi, W. Kienzle, M. Toelle, M. Uyttendaele, and M. F. Cohen, "Real-time hyperlapse creation via optimal frame selection," *ACM Trans. Graph.*, vol. 34, no. 4, p. 63:1-63:9, 2015.

[4] I. Engineering and V. Stabilization, "The Technology behind Hyperlapse from Instagram," *Instagram Engineering*, 2015. [Online]. Available: http://instagram-engineering.tumblr.com/post/95922900787/hyperlapse.

[5] N. Snavely, R. Garg, S. M. Seitz, and R. Szeliski, "Finding paths through the world's photos," *ACM Trans. Graph.*, vol. 27, no. 3, p. 1, 2008.

[6] R. Martin-Brualla, D. Gallup, and S. M. Seitz, "3D Time-Lapse Reconstruction from Internet Photos," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1332–1340, 2015.

[7] K. T. Zetterman, "Video stabilization algorithm from low frame rate video for hyperlapse applications Video Stabilization Algorithm from Low Frame Rate Video for Hyperlapse," 2015.

[8] R. Mukherjee, K. Debattista, T. Bashford-Rogers, B. Waterfield, and A. Chalmers, "A study on user preference of high dynamic range over low dynamic range video," *Vis. Comput.*, vol. 32, no. 6–8, pp. 825–834, 2016.

[9] V. Popovic, K. Seyid, E. Pignat, Ö. Çogal, and Y. Leblebici, "Multi-camera platform for panoramic real-time HDR video construction and rendering," *J. Real-Time Image Process.*, 2014.

[10] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile hdr video production system," *ACM Trans. Graph.*, vol. 30, p. 41, 2011.

[11] S. G. Bernal and G. Qiu, "A Portable Real-time High-Dynamic Range Video System," *Int. Conf. Image Process.*, pp. 2163–2164, 2014.

[12] Red.com Inc., "High Dynamic Range Video With HDRX." [Online]. Available: http://www.red.com/learn/red-101/hdrx-high-dynamic-range-video.

[13] S. Vargas-Sierra, G. Liñán-Cembrano, and Á. Rodríguez-Vázquez, "A 151 dB high dynamic range CMOS image sensor chip architecture with tone mapping compression embedded in-pixel," *IEEE Sens. J.*, vol. 15, no. 1, pp. 180–195, 2015.

[14] S. Vargas-Sierra, G. Linan-Cembrano, and A. Rodriguez-Vazquez, "Live demonstration: Real-time high dynamic range video acquisition using in-pixel adaptive content-aware tone mapping compression," *Proc. - IEEE Int. Symp. Circuits Syst.*, vol. 2015–July, p. 1900, 2015.

[15] F. Banterle, K. Debattista, A. Artusi, S. Pattanaik, K. Myszkowski, P. Ledda, and A. Chalmers, "High dynamic range imaging and low dynamic range expansion for generating hdr content," *Comput. Graph. Forum*, vol. 28, no. 8, pp. 2343–2367, 2009.

[16] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[17] S. Huang, Q. Liu, H. Wang, and T. Ikenaga, "Motion Area based Exposure Fusion Algorithm for Ghost Removal in High Dynamic Range Video Generation," pp. 1–4, 2014.