BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF INFORMATION TECHNOLOGY
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

GHOST-FREE HDR VIDEO USING FPGA

GHOST-FREE HDR VIDEO S VYUŽITÍM FPGA

SUMMARY OF DOCTORAL THESIS
ZKRÁCENÉ ZNĚNÍ DOKTORSKÉ PRÁCE

PHD THESIS
DISERTAČNÍ PRÁCE

AUTHOR — **Ing. MARTIN MUSIL**
AUTOR PRÁCE

SUPERVISOR — **Prof. Dr. Ing. PAVEL ZEMČÍK**
ŠKOLITEL

BRNO 2020

Abstract

This thesis proposes an algorithm for multi-exposure ghost-free HDR video acquisition for embedded devices. The Ghost-free HDR acquisition was evaluated on the state-of-the-art FPGA architecture and achieved more than real-time performance of 96FPS on FullHD resolution. The proposed Ghost-free algorithm produces output visually comparable to the state-of-the-art algorithms which are considerably more demanding or not implementable on embedded devices at all.

Abstrakt

Tato práce navrhuje algoritmus pro pořizování ghost-free HDR videa ze sekvence expozic, který je určený pro implementaci ve vestavěných zařízeních. Vlastnosti algoritmu byly ověřeny implementací ve state-of-the-art architektuře HDR kamery, kde je schopen zpracovávat HDR video s potlačením tzv. ghosting efektu rychlostí až 96 snímků za sekundu na FullHD rozlišení, což více než dostačuje pro zpracování v reálném čase. Navrhovaný ghost-free algoritmus produkuje výstup vizuálně srovnatelný s nejmodernějšími algoritmy, které jsou výpočetně řádově složitější a často je nelze na embedded zařízeních ani implementovat.

Keywords

HDR, HDR Acquisition, HDR Deghosting, Embedded Systems, FPGA, Real-time HDR processing

Klíčová slova

HDR, pořizování HDR, HDR Deghosting, vestavěné systémy, FPGA, zpracování HDR v reálném čase

Reference

MUSIL, Martin. *Ghost-free HDR video using FPGA*. Brno, 2020. PhD thesis. Brno University of Technology, Faculty of Information Technology. Supervisor Prof. Dr. Ing. Pavel Zemčík

Ghost-free HDR video using FPGA

Declaration

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Prof. Dr. Ing Pavla Zemčíka. Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

.....
Martin Musil
August 31, 2020

Acknowledgements

Tímto bych chtěl poděkovat svému vedoucímu prof. Dr. Ing. Pavlu Zemčíkovi za odborné vedení, konzultace a připomínky, které mi pomohly při řešení diplomové práce. Dále bych chtěl poděkovat kolegovi Ing. Svetozáru Noskovi za spolupráci na tvorbě HDR kamery, jejíž architektura byla několikrát publikována v odborných časopisech a na níž byl demonstrován vědecký přínos této práce.

Contents

Chapter 1

Introduction

In the real world, our human vision is capable of seeing and recognising objects in various light conditions, even when they mix in one scene, such as a view from dark room outside to the sunny street. In the contemporary digital world, we are also trying to get this real-looking images into digital form as photography, video etc. One of the current problem in digital image acquisition is very limited dynamic contrast that can be captured from the scene, because the current camera sensors have only limited and linear response to the light, unlike the human eye. This often leads to photos with some white (overexposed) and black (underexposed) sections.

An effort still exists to remove this bottleneck and capture a high dynamic range image (HDR). The first possible way is to assemble a chip with a non-linear response to lightning. They are currently available, but they are still in the early age of development and suffers from some bugs, they have small resolutions, etc. Currently, most spread way how to obtain an HDR image is by merging a sequence of low dynamic range images (LDR) captured by the ordinary camera into one HDR.

The algorithms that merge LDRs into HDR image are known for a quite long time, but they produce a good visual result only with static scenes. In case of any motion, either in the scene or by the camera itself, the ghosting artefacts occur in resulting HDR image. Quite many papers about deghosting techniques were proposed; however, it is still a challenge and a quite open problem, no universal method with reference „deghosted“ result exists.

This dissertation is motivated by a need of many surveillance, security, traffic monitoring systems, and industrial applications that can benefit from HDR video capture. These applications are typically cost-sensitive and so multi-exposure HDR acquisition is often the only feasible option. In these use-cases, the motion in the scene is inevitable and „ghosting“ in such systems, caused by the nature of image acquisition, troubles the applications. Therefore, I decided to develop a method of fast de-ghosting for such applications.

Applications in surveillance, security and industry require high performance in general – we cannot afford slow and demanding offline processing that the best state-of-the-art algorithms require. The essential goal is to capture HDR image fast, to be able to react to a certain situation very fast and or in a given time frame.

Image acquisition systems of this type are still being built on PC based systems; however, this approach is on the decline, since the PCs are expensive, they have large dimensions, and they consume a lot of power. Nowadays, the interest is turning towards compact embedded systems, which are breaking such limits. They often contain low power CPUs accompa-

nied by powerful, task tailored accelerators which require a fraction of power consumption comparing to CPU based systems, while they can deliver even much more performance.

The most efficient circuits are generally considered to be ASICs, which means Application-Specific Integrated Circuits. It is a collective name for single-purpose circuits/accelerators, tailored to provide specific functionality only. However, the manufacturing cost of such circuits is enormous; its manufacturing pays off only with high volumes of chips. The development processes of ASICs are taking place on large FPGAs (Field-programmable Gate Array), which are a completely customisable array of logic gates and registers, which can be interconnected in any desired way; therefore, they offer quite the same flexibility in design as ASICs, but with diametrically lower cost. Nowadays, FPGAs are very popular even in consumer electronics for their computing power, reliability, reprogrammability, low cost, and also low power consumption. These benefits are outweighed by designing time, which is still quite high. Also, not every task is implementable or convenient to accelerate on FPGA.

Some class of image processing algorithms are quite suitable for FPGA acceleration, at least when they uniformly process the image by pixels or blocks. For example, the HDR acquisition, as it was proposed by Debevec and Malik[4] is a typical example of a suitable algorithm. Unfortunately, this algorithm requires static images to produce a good-looking visual output. In case of motion in the scene, the ghost effects appear. As it is summarised later in this dissertation work, deghosting algorithms producing good visual output are very computationally demanding and quite often not even implementable on FPGA. The simpler algorithms are, on the other way, not very successful in deghosting and therefore, they are not suitable for applications in security, traffic monitoring, or industrial applications.

These circumstances led me to set the scientific contribution of this thesis to prove that a multi-exposure ghost-free HDR acquisition algorithm comparable to the state-of-the-art algorithms in quality can be designed for an embedded hardware device and achieves a real-time performance at high resolution.

The dissertation thesis begins with Chapter 2, which contains an overview of state-of-the-art algorithms related to the HDR acquisition and tonemapping. Chapter 2 further contains an overview of state-of-the-art deghosting algorithms, followed by selected deghosting algorithms feasible to be implemented in embedded devices. The thesis continues with Chapter 3 that contains an overview of hardware platforms suitable for implementation of deghosting algorithms, including an overview of embedded system-on-chip solutions. Chapter 3 is further focused on embedded platforms of for HDR acquisition, followed by an overview of existing embedded HDR deghosting solutions.

The proposal of ghost-free merging algorithm, which I developed to fulfil the goal stated in this thesis, is located in Chapter 4, which also contains algorithm evaluation, comparison to related algorithms, and also to the state-of-the-art. The chapter contains an evaluation of performance and power consumption, which demonstrates the engineering contributions of the proposed solution. The chapter ends with an evaluation of scientific contribution and by a summary of possible applications of the proposed algorithm.

Chapter 2

HDR acquisition and deghosting

This chapter contains an overview of state-of-the-art algorithms related to HDR acquisition and tonemapping. The chapter further contains an overview of state-of-the-art deghosting algorithms and also an overview of selected deghosting algorithms feasible to be implemented in embedded devices.

2.1 HDR acquisition

Two main approaches to HDR (High Dynamic Range) image capture exist. First of them is to build special cameras with HDR sensor. Some commercial products start to be available, such as SpheroCam HDR[1], or Panoscan MK3[2] In the academic world, Sakakibara et al. [37] introduced a High-Sensitivity CMOS sensor with gain adaptive column amplifiers and 14 bit analogue-digital converters. Zhao et al. [57] capture HDR using the modulo camera. All the above approaches require the availability of special HDR sensors or generally expensive and technologically demanding equipment. Regarding the HDR sensors, it is questionable whether some physical limit in a dynamic range will eventually be reached and what it will be.

The second and more frequently used approach is based on standard sensors/cameras which captures the high luminance range in the scene sequentially, by the acquisition of multiple images typically with varying exposure times [4, 28, 36, 25]; such sequence is then merged into one HDR image. The individual images can be captured simultaneously, e.g. using a beam splitter with several CCD/CMOS sensors [45], or, more often are gathered sequentially using a single image sensor which causes ghost effects by a motion of objects during the sequence acquisition. This approach is technologically less demanding and results in cheaper systems.

HDR acquisition algorithms

Two main approaches how to merge differently exposed standard images into an HDR image exist, the first and more efficient approach involves a combination of pixels in the image domain (direct merging of pixels). As an example, a method presented by Mertens et al. [25] combines multiple exposures directly without any knowledge of the camera response function(CRF). In this approach, only the best parts of frames from each exposure are

[1] https://www.spheron.com/
[2] http://www.panoscan.com/

exploited. A resulting HDR image is obtained as a weighted average of pixel values across the exposures:

$$I_C = \sum_{k=1}^{N} w\left(Z_k\right) Z_k \tag{2.1}$$

where I_C is a composite image, Z_k is a pixel value and $w\left(Z_k\right)$ is a weight of a pixel. This approach produces the HDR images which can be directly displayed on LDR (Low Dynamic Range) monitors.

The second approach is based on merging in the radiance domain, in the meaning of real illumination in the given scene. Algorithms using this approach are attempting to calculate the exact value of luminance in the scene. These methods require knowledge of the camera response function [4, 36, 28], which is the response function of the camera sensor to the incident light. The inverse function of CRF is then applied to obtain an image with approximately linear response to light. The CCD and CMOS technology generally do have a linear response function, but the image results are often affected by postprocessing algorithms, for example, by gamma-correction or by white balance. In general, RAW images are preferable for HDR composition because they contain data obtained directly from CCD/CMOS sensors without any postprocessing, and therefore it can be assumed that they have a linear response function. Unlike the merging in the image domain, this class of algorithms produces an image with higher bit-depth, which is not directly displayable on standard LDR devices. The HDR images have to be post-processed by algorithms commonly called tone mapping operators. The operators reduce the bit-depth of the HDR image while they preserve all important image details.

Debevec and Malik [4] proposed an algorithm which can fuse multiple photographs into a high dynamic range radiance map whose pixel values are proportional to the true radiance values in the scene. The contribution of each pixel is determined from the weight function [4]. Resulting pixel value p in HDR image is calculated as a weighted average of each pixel exposures:

$$L_p = \frac{\sum_{i=0}^{N} w(Z_{ip}) \frac{Z_{ip}}{t_i}}{\sum_{i=0}^{N} w(Z_{ip})} \tag{2.2}$$

where L_p is the resulting pixel value p in HDR image, N is the number of input images, Z_{ip} is the value of a pixel p in image number i, t_i is the the exposure time of image i.

This algorithm could also be applied with different weighting functions [36, 28].

2.2 HDR deghosting

The HDR merging algorithms [4, 28, 36, 25] summarized in Section 2.1 are suitable for static scenes only. Motion of objects during the image sequence capture causes adverse effects called *ghosting*. To reduce such effects, various methods to detect and remove ghosting from HDR images have been developed.

The problem of removing motion artefacts for sequential HDR imaging has been the subject of extensive research and has led to two major type of approaches. The first type assumes that the images are mostly static and that only a small part of the scene contains motion. These de-ghosting algorithms use the input frames to determine whether a given pixel is static or has motion and then apply different merging algorithms in each case. For static pixels, the traditional HDR merge can be used. For motion pixels, many algorithms use only a subset of exposures (in many cases only one) to produce a deghosted HDR. The

fundamental problem with these techniques is that they cannot handle scenes with large motion if the moving parts of the scene contain HDR content.

The second type of approaches attempts to align the input sources to a reference exposure before merging them into an HDR image. The most successful algorithms use optical flow to register the images, but even these methods are still brittle in cases of large motion or complex occlusion/dis-occlusion.

According to the goal of this dissertation, I focused on algorithms feasible of capturing ghost-free HDR images in real-time. Anyway, a short introduction into the demanding optical flow and patch-based algorithms is presented.

2.3 Motion object selection methods

This dissertation work focuses on embedded systems and real-time processing; therefore, only simple, computationally unpretentious methods, categorised by Tursun [46], Srikantha [41] and other authors as „motion object selection“ methods are reviewed in this subsection. The optical flow-based and patch-based algorithms are, due to their high computational demands, reviewed only for the coherence of the topic. Also, the global image registration is not addressed, as we assume only static cameras.

Gallo et al. [8] assumes a linear dependency between couples of pixels when they „see“ the same radiance levels, based on knowledge of exposure times. The following relation between the images is expected:

$$L_i = L_j \cdot \frac{t_i}{t_j} \tag{2.3}$$

Any image spot violating this linear relation is considered as containing a motion. All images are registered to the reference image L_{ref}; to suggest a good reference frame, they find the saturated pixels in each image of the stack, then they remove small saturated regions with morphological operators (erosion followed by dilation) because such area's neighbourhood usually contains enough information to avoid artefacts. Finally, they pick the exposure with the fewest remaining saturated pixels. [8]

The reciprocity assumption states that if the radiance of the scene does not change, the exposure time and the irradiance are linearly related through the exposure time Δt:

$$X = E \cdot \Delta t \tag{2.4}$$

To increase a robustness and prevent rising of such artifacts, the algorithm operates on relatively large rectangular patches (e.g. 40x40 pixels) rather than individual pixels. Patches with a large number of not corresponding pixels are omitted from merging, causing visible artefacts to occur at their boundaries; Gallo et al. [8] suggest their suppression by Poisson blending.

Raman et al. [34] extended the work of Gallo et al. [8] so that it does not require any knowledge of the CRF or exposure settings. They introduced an intensity mapping function (IMF) obtained from the static part of the scene – they assume that upper 5-10 image lines are usually static. The authors assume the motion is mostly confined to the ground plane of the scene. This assumption may be very limiting, and it can work only for certain scene compositions.

Grosch [9] proposed a simple method based on the estimation of pixel value from the known exposure time and CRF. In opposite to the most of the algorithms that require a

static scene and direct correspondence of pixels to obtain a CRF, Grosch uses the algorithm presented by Grossberg and Nayar [10] to recover a CRF from a non-aligned sequence with object motion. This algorithm calculates the response function based on cumulative histograms and is mostly unaffected by camera or object motion. [9]

With a known camera response function, they can predict the pixel colour from one image to another. For each pair of consecutive images, they test if the real colour in the second image is well approximated with the predicted colour from the first one. If the pixels at the certain position do not fit the estimation, the corresponding region is marked as ghosted into the *error map*. [9] To increase the robustness and eliminate the influence of the noise in the source images, the author uses a user-defined threshold for the pixel colour comparison.

Wu et al. [51] algorithm estimate the CRF from regions where RGB vectors remain fixed with respect to the changes of exposure. The algorithm refines motion detection by a combination of pixel order relation from Sidibe et al. [39] and pixel value estimation from Grosch [9].

Wang et al. [50] proposed the motion region detection method, that is motivated by the inter-frame difference method for video sequence that does subtraction to compute the difference between adjacent frames on the intensity domain. To enable it, the algorithm normalises all images L_i according to the reference image L_{ref}. For each pixel, if the corresponding difference value is bigger than a certain threshold, then the pixel is considered to be in a motion region. This method is commonly used on motion detection of video stream. [50]

Figure 2.1: The Figure shows the results of variance based deghosting method by Jacobs et al. [12]. The variance map (bottom left) is obtained from the image sequence (upper row) and used to generate the HDR image (bottom right). Figure obtained from [12].

The algorithm of Jacobs et al. [12] is calculating pixel variance over the exposures to detect the presence of motion. The *Variance Image* is created, storing pixel's variance over the exposures in a matrix with the same resolution as input images. Further, they ignore under and over-saturated pixels in *Variance Image*. The *Variance Image* is transformed into

binary map (equivalent of *ghostmap*, see on Figure 2.1)), with movement clusters, which are formed by comparing the *Variance Image* with fixed threshold. The *Variance Image* is supplied by *Uncertainty Image*, which is calculated using the local variance, obtained from a histogram of a small 2D window; 5x5 pixels in size [12].

Min et al. [26] improved method of Pece et al. [32] and introduced multi-level threshold map, where thresholds are selected to divide the image into multiple regions according to the pixel intensity, each region having the same number of pixels (see Figure 2.2). Any difference between the threshold maps of input images and the reference image, presented typically by the mid-exposure one, is marked as a motion-region. Introduction of multiple histogram regions, in opposite to Pece et al. [32], allows for the incorporation of a tolerance in which shifts of pixels within neighbouring regions are not evaluated as motion. The algorithm suffers from dependence on scene composition and image histogram layout. The above methods by Pece et al. [32] and Min et al. [26, 27] are using coarse morphological operators, such as erosion and dilatation, to suppress false detection rising on edges or by noise.

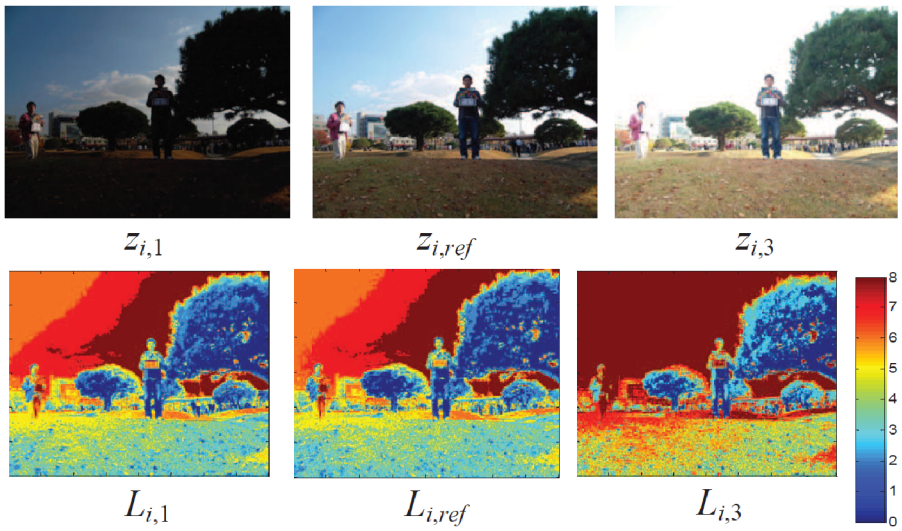

Figure 2.2: The figure presents the intermediate step of deghosting algorithm by Min et al. [26]. The source sequence is on the top, the bottom images shows the multi-level threshold maps for corresponding images on the top. Figure obtained from [26].

Bouderbane et al. [2] implemented simple ghost removing algorithm on FPGA based platform. They were inspired by the work of Sidibe et al. [39] and presented the algorithm based on the modification of Debevec [4] weighting function. The idea of the methods is to adjust pixel weights based on the deviation from the reference image [39]. The function gives a higher weight for pixels whose value are closed to the reference value and low weight for pixels whose value diverges considerably from a reference value. Consequently, they achieved the same performance as the Debevec and Malik [4] standard algorithm with a ghost removing in a radiance domain, right before HDR data generation. [2]

2.4 Motion object registration methods

The following algorithms are not suitable for real-time processing; however, I reviewed them for the coherence of the HDR deghosting topic and also because they are part of the state-of-the-art in terms of deghosting quality. Achieving good visual results comparing to such algorithms is also one of my side-goals.

Patch-based and Optical flow based algorithms

These approaches attempt to align the different LDR exposures before merging them into the final HDR image. Although the alignment of images has long been studied in image processing and vision communities (e.g. Zitová and Flusser [58]), its application to HDR imaging has special considerations. The quality of the HDR images produced by these techniques is fundamentally limited by the accuracy of the alignment. Furthermore, optical flow cannot typically synthesise new content and thus cannot handle disoccluded content that could be made visible when aligning one image to another. [38]

The algorithm proposed by Sen et al. [38] is a patch-based energy minimisation formula. The algorithm produces an HDR image from a set of LDR images captured with different exposures which is aligned to the reference image L_{ref} and which is also an LDR image that contains the best-exposed pixels. The resulting HDR image contains as much information as possible from the well-exposed pixels from the L_{ref} image (see Figure 2.3). In places where L_{ref} is not well exposed, every patch in the image H at a given exposure should have a similar patch in one of the LDR images after exposure adjustment (coherence). Also, every exposure adjusted patch in all L_k images should be contained in H at exposure k (completeness). The iterative approach performs joint optimisation of image alignment and HDR merge process until all the exposures are correctly aligned to the reference exposure, and a good quality HDR result is produced.

Figure 2.3: The figure shows the source sequence, images reconstructed by patch-based algorithm by Sen et al. [38] and the resulting HDR image. Image obtained from [38].

Ferradans et al. [7] find dense correspondence of input images in the radiance domain with respect to the reference image. In order to detect the mismatches in the estimated flow fields, the input images are warped using the estimated fields, and the absolute difference

map of each pixel is calculated. Instead of applying a fixed threshold to the difference map, its histogram is modelled as a mixture of Gaussians. The pixel intensities corresponding to the flow vectors causing the mismatch are assigned zero weight in HDR reconstruction. The information from the remaining pixels in each input image is fused in the gradient domain. Jinno and Okuda [13] use a novel weighting function which has significantly smaller overlap between the contribution of input LDR images to the radiance domain. The proposed method assumes that the global alignment is already performed. Displacement, occlusion, and saturation regions are modelled as Markov Random Fields. The optimal parameters are found by minimising the energy function (see [13]). [46]

CNN based algorithms

The latest published algorithms are based on popular Convolution Neural Networks (CNN). Kalantari et al. [14] based their approach on optical flow from Liu et al. [21] and merges images into HDR using CNN. At first step, the source images are normalised to the same level of luminance as the reference (middle) image – similarly to Wang et al. [50] and many others. Then, the optical flow algorithm of Liu et al. [21] is used to align the images. Such aligned set is merged using CNN network trained on their dataset containing ground truth sequences. The CNN is responsible for removing the ghosting artefacts appearing on the edges of motion regions. Yan et al. [54] proposed a similar approach; however, their proposed CNN uses not only surrounding information of a pixel as Kalantari et al. [14], but also considers the information from other frames.

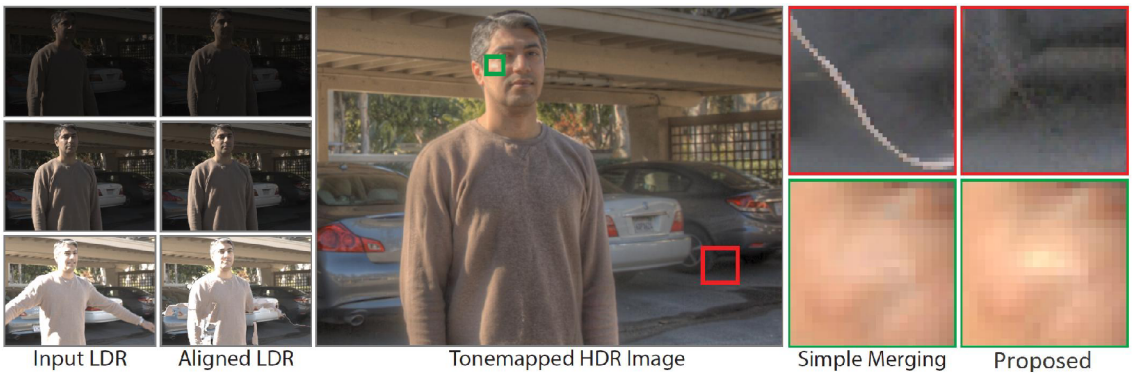

Figure 2.4: The figure presents the results achieved by Kalantari et al. [14]. From the left – the source sequence, images aligned by optical flow by Liu et al [21], resulting tonemapped images and the details of marked region merged by „simple“ merging (probably by Debevec and Malik [4]) and by proposed CNN based method by Kalantari et al. [14].

Chapter 3

Embedded HDR acquisition and deghosting

This chapter is devoted to a description of state-of-the-art implementations of HDR acquisition and deghosting on embedded devices and description of theirs, mostly custom based embedded platforms.

After consideration the features of target platforms, the FPGA was selected as a target platform, namely the SoC Xilinx Zynq, which is a powerful combination of FPGA and dual-core ARM processor on the same chip. This SoC allows the application of hardware-software codesign technique. It brings together the performance benefits of FPGA with the possibility of sequential execution of code - e.g. for driving the FPGA processing or to perform complex calculations, which acceleration in FPGA would be very demanding or not reasonable. Nowadays, the platforms with powerful embedded GPU, such as Nvidia Tegra, are starting to be concurrent at certain parameters. On the other hand, DSP platforms are slowly getting to the margins of interest.

3.1 State-of-the-art hardware solutions overview

Many research publications were published regarding the acquisition of HDR images; however, only a few of them are oriented on embedded devices. HDR merging itself is not a complex algorithm, but for real-time acquisition, it requires a high memory throughput and external memory buffer, which is not available on many embedded platforms.

FPGA based platforms are more than suitable for such type of applications. Several papers focused on FPGA acceleration and related to our work were published [18, 20, 48, 49, 23, 33, 53, 43]. This section provides its overview and presents achieved properties.

Realtime HDR video for eyetap wearable computer by Mann et al.

Mann et al. [23] developed an FPGA based wearable HDR seeing aid designed for the electric arc welding (see Figure 3.1). The prototype consists of an EyeTap (electric glasses) welding helmet, with a wearable computer upon which are implemented a set of image processing algorithms that implement real-time HDR image processing together with applications such as mediated and augmented reality. The HDR video system runs in real-time and processes 120 frames per second, in groups of three or four frames. The processing method, for implementation on FPGAs (Field Programmable Gate Arrays), achieves real-time performance

for creating HDR video using the novel compositing methods, and runs on a miniature self-contained battery-operated head-worn circuit board, without the need for a host computer. The result is an essentially self-contained miniaturize hardware HDR camera system that could be built into smaller eyeglass frames. [23]

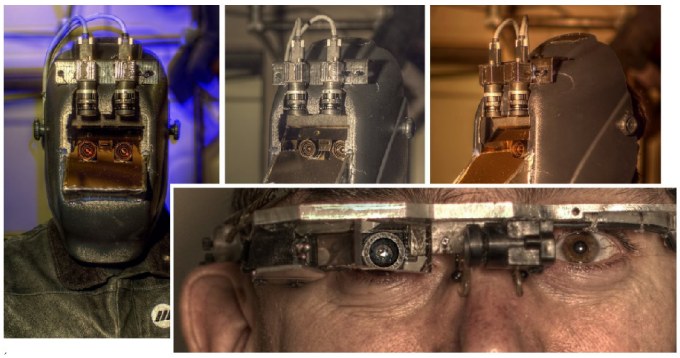

Figure 3.1: The "MannVis welding helmet" implements the EyeTap principle which causes each eye to, in effect, function as if the eye itself were both a camera and display. Image obtained from [23].

The HDR output values are precomputed for a full range of input pixel combinations and stored in lookup tables in BRAMs. Even after certain optimizations of memory consumption, the BRAM demands are very high, especially when more than two LDR images are used. The system is implemented on Spartan-6 LX45 FPGA and produces 720p video at 60 FPS while fusing two images.

Real-Time HDR Video Imaging on FPGA by Tao et al.

Tao et al. [44] extended the work of Mann [23] by introducing a lookup table compressed using quadtree structure, which saves the amount of BlockRAM resources. Tao replaced the weighted sum approach with the new quadtree-based compositing for high-quality HDR video production. The proposed compositing circuits are generated by the software, with parameters given by the user. It compresses and implements a 2D Lookup Table (LUT) on an FPGA, by bounding the error and space of quadtree representation of the original LUT according to the expected usage, so that the LUT is compressed to fit within the total amount of the block RAM resource available in a mid-sized FPGA. They also add the support for 1080p video at 60 FPS. [44]

HDR-ARtiSt: an adaptive real-time HDR smart camera by Lapray et al.

Lapray et al. [18, 19, 20] developed a complete FPGA-based smart camera architecture named HDR-ARtiSt (High Dynamic Range Adaptive Real-time Smart camera). This smart camera is able to provide a real-time HDR live video from multiple exposures capturing to display through radiance maps and tone mapping. The main contribution of their work is the generation of a new FPGA embedded architecture producing an uncompressed Black&White 1280×1024-pixel HDR live video at 60 FPS. [20]

According to the detailed description of these methodologies and the comparison of their real-time software implementations, they decided to use the Debevec's method [4] for HDR merging. The main advantage of this approach is that there is very little constraint

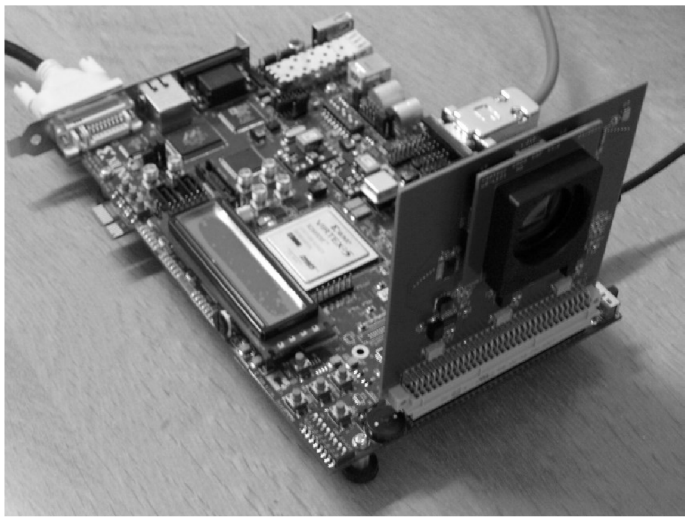

Figure 3.2: Xilinx Virtex-5 ML507 FPGA board equipped with 1.3MPix CMOS, where Lapray et al. [18, 19, 20] implemented the HDR image acquisition and tone mapping. Image retrieved from [20].

about the response function (other than its invertibility). Moreover, the proposed algorithm proved to be quite robust and easy to use due to the simplicity of Debevec's equation (see Equation 2.2). [20]

Regarding the tonemapping operators, Lapray et al. [18, 19, 20] implemented two the global tonemapping operators by Duan [5] and Reinhard [35]. Their implementations were published and described thorough their articles [18, 19, 20].

The HDR-ARtiSt platform [20] is a smart camera built around a Xilinx ML507 board, equipped with a Xilinx Virtex-5 XC5VFX70T FPGA (see Figure 3.2). The motherboard includes a 256 MB DDR2 SDRAM memory used to buffer the multiple frames captured by the sensor.

3.1.1 Real-time HDR video compression using an FPGA by Zemcik et al.

The architecture of the HDR camera proposed by Zemcik et al. [55] can capture 30 FPS FullHD with each frame formed from two exposures, or 20 fps FullHD video formed from three exposures. With sharing the expositions, the output can eventually reach up to 60FPS; however, the whole pipeline is limited by the capability of H.264 encoders, supporting 30FPS only. The main architecture highlight is the encoding of HDR video using two standard video codecs. The HDR camera designed by Zemcik et al. [55] is shown in Figure 3.3.

This architecture uses standalone 2K Flare[1] camera connected over 3G-SDI interface[2] (commonly used in TV studios). This camera is producing high quality FullHD RAW image at up to 60FPS.

[1] http://www.ioindustries.com/
[2] https://www.smpte.org/standards

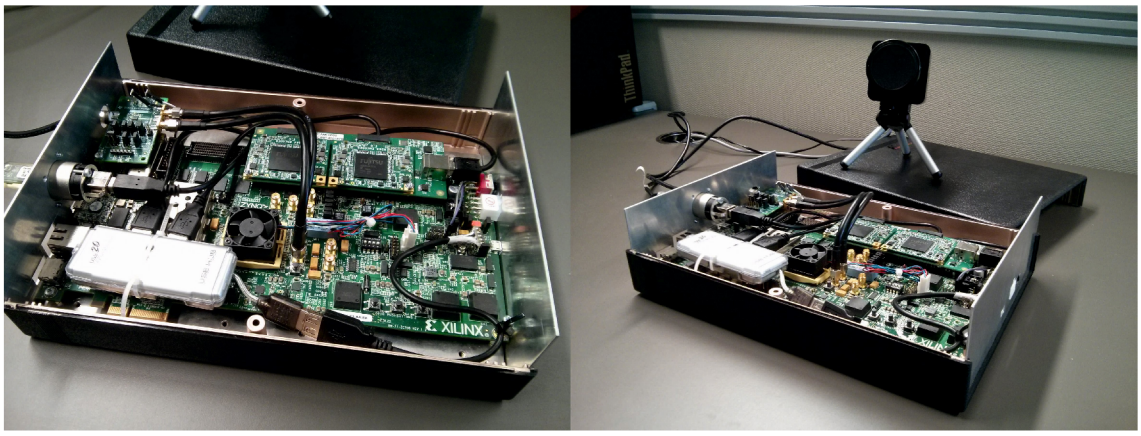

Figure 3.3: A photography of the HDR camera prototype. Note, please, the FPGA development board, the compression modules, and also the Flare camera connected by SDI interface. Image retrieved from Zemcik et al. [55].

Architecture by Zemcik et al. uses two or three images for HDR merging, depending on configuration, so there is implemented the equivalent number of framebuffers. The double buffering technique is used to avoid rising of image artefacts, which doubles the memory requirements but prevents rising of image artefacts.

The primary demand for the HDR merging algorithm was the capturing of the as-high-as-possible dynamic range, showing the benefits of HDR acquisition. Regarding that the architecture Zemcik et al. [55] use only a simple pixel selection algorithm, because the exposition times are set so far from each other (by multiples of eight), that the particular pixel is exposed well only in one exposition. The others are often under on overexposed; thus their contribution to computed HDR value would be marginal.

3.2 Ghost avoiding/removing solutions

The following section summarises the state-of-the-art HDR acquisition solutions, which either suppress and remove ghosting effect or prevent its occurrence.

HDR camera based on dual-gain CMOS by Tang et al.

Tang et al. [43] developed an HDR camera based on Altera FPGA and equipped with dual-channel CMOS GSENSE400BSI, which is able to apply different analogue gain to the same captured data (see the prototype on Figure [43]). The HDR camera can capture the wide dynamic range image of the nature scene without ghosting phenomenon, by combining the two images with different gain to an HDR frame up to 95 dB. Additionally, the frames are captured at the same moment by two channels with different gain, which reduces the interference between successive frames. However, the CMOS sensor has a rolling shutter, and the disruptive effects can still occur. [43]

In such way of HDR acquisition avoids rising of ghosting effect caused by sequential image acquisition; however, the CMOS sensor has a rolling shutter, and then another kind of image artefacts still occurs. The maximum frame rate of the camera is 60 FPS at a

resolution of 1920 × 1080. The camera uses a global tone mapping operator by Duan et al. [5].

Real-time ghost free HDR video using weight adaptation method by Bouderbane et al.

Bouderbane et al. [3] implemented a deghosting algorithm on the same platform as Lapray et al. [20]. Their method repose on the modulation of weights of the Debevec [4] algorithm, where they adjust pixel weights based on their deviation from pixels of the reference image using the weighting function given [2] (see Figure 3.4) which parameters are taken from Sidibe et al. [39].

To calculate final weighs (Figure 3.4 right) to be used in the high dynamic range reconstruction, they multiply the standard weights from Debevec [4] by the modulation factor (Figure 3.4 left).

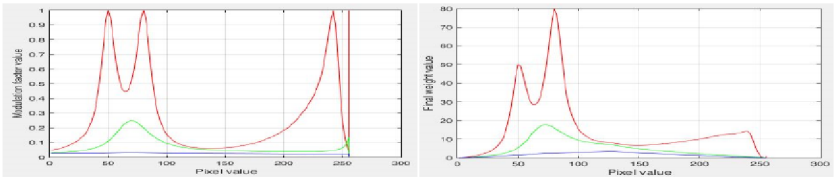

Figure 3.4: The weight modulation factor (left) and the final weight function(right) used in ghost removal HDR merging. Red curve is the factor for the closest radiance value of LDR images to the reference radiance value, the blue curve is the farthest value from the reference value and the green curve is for middle values. Image retrieved from [3].

True HDR camera with bilateral filter based tone mapping by Nosko et al. [30]

Nosko et al. [30] published the HDR architecture implemented on a custom camera platform based on SoC Xilinx Zynq XC7Z020 (see Figure 3.5). The platform is equipped by a low noise global shutter CMOS sensor Python2000 from ON Semiconductor with resolution of 1920 × 1280 pixels. The camera provides up to 30 FPS of grayscale HDR video with fixed f-stop range. However, the architecture itself is capable of processing up to 96 FPS. The architecture implements a high quality local tonemapping operator by Durand [6] based on the bilateral filter of 9 × 9 pixels. Resulting tonemapped image is streamed over the network in the form of MPEG2-TS stream.

The HDR camera architecture published by Nosko et al. [30] is based on the method by Debevec [4]. The exposition weights for individual images are calculated as follows: Given the image with shortest exposition t_1 time weight equal to one, the other images will be given the weights of $\frac{t_i}{t_1}$, where t_i is exposition time of ith image in sequence. The HDR pixel value is computed as follows:

$$H = \frac{\sum_{i=1}^{n} L_i \cdot w(L_i) \cdot \frac{t_i}{t_1}}{\sum_{i=1}^{n} w(L_i)} \tag{3.1}$$

where is the HDR pixel value, L_x is the x-th image in the sequence, t_i exposition time of i-th image and w the „plateau“ weighting function [1]).

Unlike the algorithm by Debevec [4] they chose a plateau weighing function [1] as the one leading to the best visual experience; however, it can be easily customized.

Figure 3.5: Prototype of HDR camera by Nosko et al. [30, 31]

The exposition time of the middle image in the sequence is configurable, however, the mutual intervals between exposures are fixed to multiples of two, which leads to shift operations instead of multiplication. Only a middle exposition value is configurable [30].

The resulting HDR pixel is obtained by dividing the sum of pixels by sum of weights. The division is a time and resource-demanding operation, so Nosko et al. [30] decided to convert it into multiplication by a tabulated fractional value. The sum of weights, according to bit-widths of intermediate results, needs to be represented by 11 bits (sum of three 9 bit values fits into 11 bits), so the fraction value is tabulated on 2048 entries. The resulting HDR pixel is in 10.8 fixed-point representation.

Color HDR video processing architecture for smart camera by Nosko et al. [31]

This architecture further improves the architecture by Nosko et al. [30]. The architecture provides up to 30 FPS of colour HDR video with fully adjustable f-stops. However, the architecture itself is capable of processing up to 96 FPS. The architecture implements particularly a ghost removal algorithm and a high quality local tonemapping operator by Durand [6] based on the bilateral filter of 11×11 pixels.

The architecture is further enhanced by a colour support. They process individual pixels of colour Filter Array (CFA), in this case, a Bayer mask, in the same manner as the grayscale pixels [42]. The colourization of the HDR image is done later, during the tonemapping process.

HDR merging with ghost-free extension Nosko et al. [31] implemented the HDR merging algorithm from Debevec [4] with modification for Ghost removal. The proposed ghost-free HDR merging is based on a prediction of the pixel value. It is based on similar principles as the solutions of Grosch [9], Wu [51] and Wang [50].

Since the exposure time of each image is known, individual pixel values in image i can be predicted from reference image using values from j.

$$L_i \approx L_j \cdot \frac{t_i}{t_j} \tag{3.2}$$

where t_x and t_y are exposition times of images. If the pixel do not match predicted value, it is omitted from merging process. Certain tolerance is taken into account, since the sensor noise, quantization errors and CRF precision may influence the predicted value and thus cause the false ghost detections.

Chapter 4

Proposal of ghost-free HDR technique

This chapter contains the proposal of a novel ghost-free HDR merging algorithm, which is the core of my work during the pursuing of my Ph.D. The core of this chapter was published in Journal of Real-Time image processing as the article „De-Ghosted HDR Video Acquisition for Embedded Systems“ [29].

The scientific contribution of this thesis is the proof that:

A multi-exposure ghost-free HDR acquisition algorithm comparable to the state-of-the-art algorithms in quality can be designed for an embedded hardware device and achieves a real-time performance at high resolution.

The embedded hardware device should be based on FPGA technology with FullHD CMOS sensor onboard, at the same time be small in size and with low power demands to fit into the energy-efficient or battery-powered systems.

In this chapter, a novel architecture implementing the above idea in FPGA is proposed and its functionality and quality of output are experimentally proved. The chapter consists of the quality comparison to the related implementations and even state-of-the-art methods, that are too computationally demanding and even not feasible to implement and/or accelerate on FPGA. The aim is to show that proposed solution is simple, yet very powerful and providing good visual results at the same time. The performance and power consumption of algorithm implemented on various platforms is summarized at the end of this chapter.

4.1 Ghost-free merging algorithm

The proposed approach is based on pixel value matching, the idea being similar to the solutions proposed by Grosch [9], Wu [51], and Wang [50] but with quite different and improved processing. The exposure time of each image is known; therefore, it is possible to estimate and match pixel values in the adjacent images, except for the over or under-exposed patches where the pixel values will obviously not match. Such estimation is not very precise, the captured image data is affected by factors such as noise, sensor quantization errors, CRF, etc. The reviewed methods generally use fixed or user-guided thresholds which must be employed in order to introduce user-defined tolerance to these factors. These fixed

Figure 4.1: Figure obtained from real application of proposed ghost-free algorithm - traffic monitoring system with licence plate detection, which demonstrates the contribution of proposed method. Top left - stripes of original images with a significant car motion. Top middle and top right - Images representing coefficients used for the HDR merging (*certainty maps*, see Section 4.1). Bottom left - ghosted HDR image. Bottom right - HDR image merged using proposed method.

or user-defined thresholds often cause adverse effects in the final HDR images, such as visible transitions between static and motion areas etc. I propose a method to overcome such problems. [29]

4.1.1 Certainty map

In this approach, every image L_i is assigned a *Certainty map* C_i related to the reference image L_{ref}, which is generally considered to be the middle (exposure) image in the sequence. The *Certainty map* C contains values representing the estimated level of certainty that the individual pixels contain the same patch of the scene as the reference pixel, but obtained under a different exposure. Unlike ghostmaps, *Certainty maps* hold not only the patches containing motion, but rather all patches inappropriate for merging - such as under and over-exposed pixels. [29].

The probability distribution of low level value pixels is Poisson [22] due to the discrete nature of the incoming photons. With higher intensities, the distribution transforms into Normal (Gaussian). Therefore, I use the Gaussian function to derive the certainty (estimated probability) that the two luminance levels, estimated and measured, match. The *Certainty map* C_i (see Figure 4.2) replaces the binary *ghostmap* with soft assigned values, obtained using the information from the reference image L_{ref}, the estimated image $\overline{L_i}$, the exposure times t_i and t_{ref}, as well as the CRF. Note, please, that in this paper the inverse CRF was implicitly applied to all images L_i. Image $\overline{L_i}$ is estimated by the following equation:

$$\overline{L_i} = L_{ref} \cdot \left(\frac{t_i}{t_{ref}} \right) \tag{4.1}$$

Consequently, the estimated value for image i is processed along with the actual value of L_i to get the probability based *Certainty map* C_i as:

$$C_i = e^{-\frac{(L_i - \overline{L_i})^2}{2\sigma^2}} \tag{4.2}$$

where σ reflects the standard deviation of the pixel measurement (affecting the „softness“ weight). The lower σ is, the sharper or more strict the *Certainty* maps are, which results mainly in the dynamic range reduction. On the other hand, a high σ causes „softer“ *Certainty* maps, which may start to be ghosted. Ghost detection generally, and indeed inherently, cannot work well for the over and under-exposed spots of an image; thus the *Certainty* map algorithm contains a boundary condition: If the estimated value lies beyond the point of saturation, the Certainty is assigned at maximum value. [29].

Figure 4.2: Two *Certainty maps* (bottom) obtained from the sequence on the top. The *Certainty map* on the left was obtained from top left and top middle (reference) image, the *Certainty map* on the right was obtained from top middle (reference) and top right image.

4.1.2 Multi-exposure merging algorithm

Proposed modification of Debevec's [4] merging algorithm incorporates the weights from the *Certainty map*, obtained through Equation 4.2. The HDR image H is calculated as the weighted sum of pixels from n images using the following equation:

$$H = \frac{C_i \cdot w(L_i) \cdot L_i \cdot \frac{t_i}{t_{min}}}{\sum_{i=1}^{n}(C_i \cdot w(L_i))} \tag{4.3}$$

The C_i for reference image certainty is considered to be 1.

The $w(L_{ref})$ is considered to be 1, as the reference image is a „pattern“ with the desired object layout; it is not desirable to weight out the pixels, even if poorly exposed. A scheme illustrating the Equation 4.2 is shown in Figure 4.3. [29].

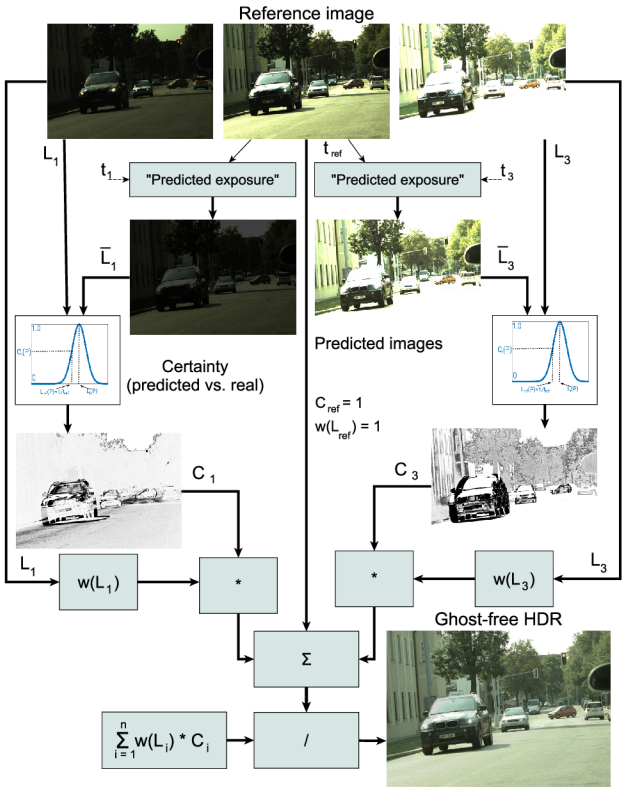

Figure 4.3: A scheme illustrating the proposed ghost-free merging of according to Equation 4.3 on a sequence of three images.

4.2 Implementation in HDR pipeline

We implemented the proposed algorithm into FPGA based HDR video acquisition pipeline proposed by Nosko et al.[31]. The proposed algorithm was designed to replace the original and very simple „Deghosting & merging“ block (please refer to Nosko et al. [31]). The Table 4.2 compares the resources consumed by such pipeline with pipeline from Bouderbane et al. [2]; unfortunately, they do not provide more detailed statistics. For detailed description regarding pipeline, please refer to the article by Nosko et al. [31]. Please note that proposed design is built on Xilinx Zynq and Bouderbane camera on Virtex-6 and also that in Nosko's pipeline, more than 1/3 of LUT and Register resources and most of BRAM and DSPs are occupied by local tone-mapping operator [31].

Table 4.1: FPGA Resource utilisation for merging 3 LDR images of 1920 × 1080 pixels. Design is routed for Xilinx Zynq Z-7020.

	LUT	LUTRAM	FF	BRAM	DSP
Certainty maps	3532	–	3339	4	4
HDR merging	893	–	2570	10	16
Total (HLS)	4425	–	5909	14	20
Total (Routing)	**1057**	**252**	**2052**	**2**	**16**
available	53200	17400	106400	280	220
utilisation [%]	1.99	1.45	1.93	0.72	7.27

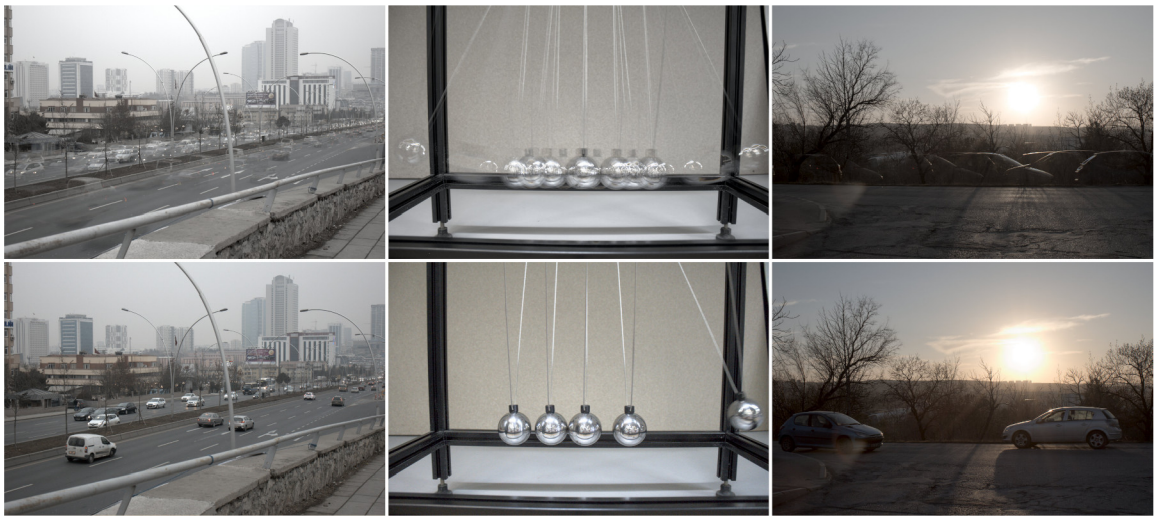

Figure 4.4: Ghosted HDRs (top line) and HDRs merged using proposed ghost-free method (bottom line) on sequences „Fast cars“ [46] (left), „105“ [47] (middle) and „117“ [47] (right). Datasets contains 9 LDR (Low Dynamic Range) images.

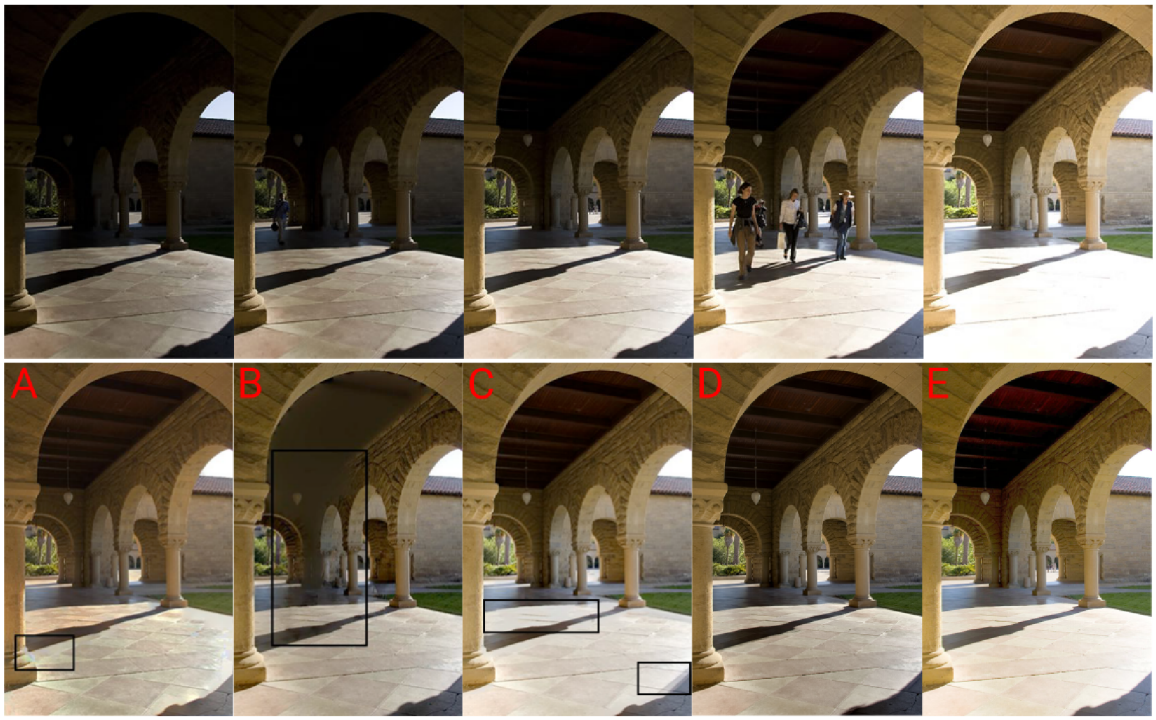

Figure 4.5: Output of proposed ghost-free merging method on the sequence of Gallo [8] (top). Previews of the various algorithm results are shown at the bottom: Gallo et al. [8] (A), Jacobs et al. [12] (B), Pece et al. [32] (C), Zhang et al. [56] (D) and proposed algorithm (E). The previews A to D are published online at http://www.vsislab.com/projects/IPM/HDR/project.html.

The proposed algorithm was implemented into FPGA based HDR video acquisition pipeline proposed by Nosko et al.[31]. The proposed algorithm was designed to replace the

original and very simple „Deghosting & merging“ block (please refer to Nosko et al. [31]). The Table 4.2 compares the resources consumed by such pipeline with pipeline from Bouderbane et al. [2]; unfortunately, they do not provide more detailed statistics. For detailed description regarding pipeline, please refer to the article by Nosko et al. [31]. Please note that proposed design is built on Xilinx Zynq and Bouderbane camera on Virtex-6 and also that in Nosko's pipeline, more than 1/3 of LUT and Register resources and most of BRAM and DSPs are occupied by local tone-mapping operator [31].

Table 4.2: Resource utilization of complete camera solution of Nosko et al. [31] enhanced by the proposed ghost-free merging block, comparing to Bouderbane [2].

	LUT	LUTRAM	FF	BRAM	DSP
Prop. pipeline	39145	3137	53592	51	58
Bouderbane [2]	49193	–	50399	35	20

4.3 State-of-the-art ghost removal evaluation

This section is focused on evaluation of proposed ghost-free HDR merging and provides the comparison to the state-of-the-art algorithms.

Figure 4.6: Sample outputs of related deghosting algorithms by Pece et al. [32] (left) and Min et al. [26] (right) on the scene from Figure 4.1. Our experiments revealed that listed algorithms should be successful only on images with convenient histogram distribution.

The proposed algorithm is evaluated on HDR datasets focused on evaluation of HDR deghosting methods [47, 46, 16], on the image sets retrieved from related articles [8, 38, 15] and also on the image sets captured by camera prototype by Nosko et al. [31] (see Section 3.2).

The results of the proposed ghost-free merging are presented in Figures 4.1, 4.4, 4.5, 4.9, 4.8, 4.10 and 4.12. Our method is suitable for almost any application with stationary cameras. Besides the evaluation of various generic datasets, the ghost removing capability was evaluated on a traffic monitoring task, where the main goal was to preserve the greatest possible level of detail so that the images can serve as evidence, with the readability of the licence plates of the vehicles in motion playing the most important part. Figure 4.1 contains a car approaching camera at approximately 50km/h. Still, six exposures ($\sim$ 66ms at 90FPS) were intentionally omitted between the images to show the capability of the ghost removing for e.g. faster moving objects.

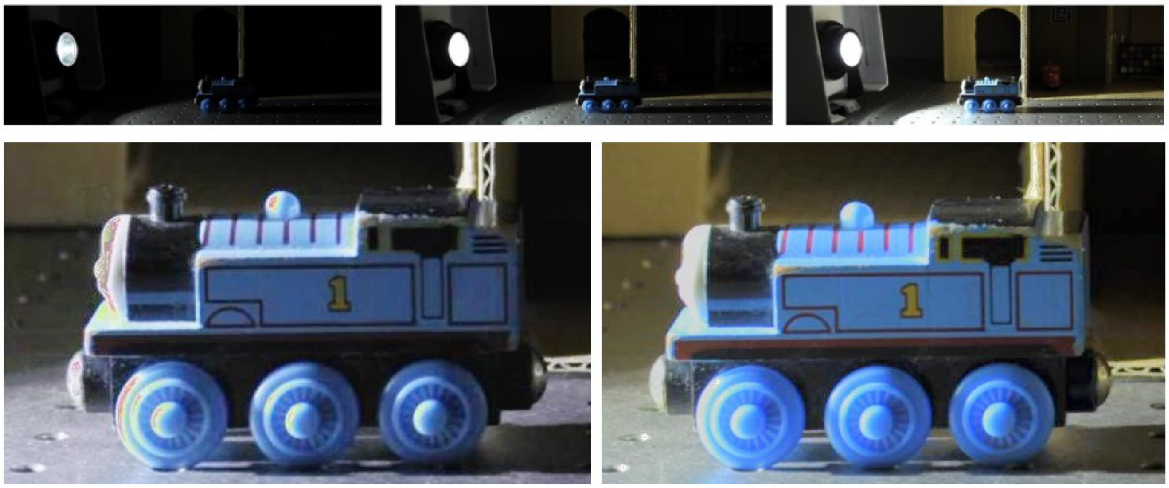

Figure 4.7: Figure shows the ghost-free HDR outputs of Bouderbane [3] (left) and proposed method (right), both tonemapped by Duan [5] operator. Bouderbane result and source images are retrieved from [3]. Please mind the color shift in very bright patches of Bouderbane result.

According to presented results, the visual outputs are comparable to the state-of-the-art; however, the proposed algorithm is capable of running in real-time, while state-of-the-art algorithms require long offline processing in terms of seconds or even minutes per image.

Probably only related work, which implements any ghost-free merging on embedded device, in this case on FPGA, was proposed by Bouderbane et al. [3]. They use method by Debevec and Malik [4], where they combined weighting function from Debevec with weight function proposed in their previous paper [2]; their method was inspired by the work of Sidibe et al. [39]. However, the ghost detection is based only on weak assumption, as Bouderbane use the weight function, which gives a higher factor for pixels whose recovered radiance value are closed to the recovered radiance of reference values and low factor for pixels whose radiance values diverge considerably from pixels radiance value of the reference image. On the image data supplied within the article, the method suppress ghosting quite well(see Figure 4.7) but the slight ghosting effect is still present (see results in the article [2]), also the dynamic range is quite reduced, even in parts with static background.

Algorithm by Gallo et al. [8] operates on relatively large rectangular patches (e.g. 40x40 pixels [8]) instead of individual pixels. If the patch contains large number of pixels not corresponding to patch from the reference image, the patch is omitted from merging. As the patches used in the algorithm are quite large, visible artifacts occur at their boundaries; the authors suggest their suppression by Poisson blending.

The methods based on histograms [32, 26] have a common issue, the scene has to be balanced from the point of histogram equalization. The method presented by Pece et al. [32] is marking pixels as ghosts based on decision, whether the pixel changes its relative position in histograms over all of the expositions. The position in histogram is acquired by comparison with median pixel value. If median is very low/high, for example if the scene has large large number of under/overexposed patches, the change of pixel position in histogram cannot be reliably detected. In the method proposed by Min et al. [26], one median threshold is replaced by eight percentiles and whole histogram is divided into nine

segments with equal number of pixels, but it only mitigates the same issue. The example outputs of the Pece et al. [32] and Min et al. [26] algorithms on data obtained by Nosko's camera [31] are shown on Figure 4.6.

In general, the existing methods are more or less using fixed or user-adjusted thresholds and binary ghost maps, which either includes the pixel into the merging process or omits it completely. Such approach negatively affects the merging process and appearance of the resulting HDR image, causes higher noise on the affected patches around the moving objects, and also on wrongly detected patches. Proposed approach does not have such limitations, it is more robust, and does not require user-guided tuning of parameters.

Figure 4.8: The source sequence (top left) is merged with (bottom) and without (top right) proposed ghost-free merging algorithm. Source images retrieved from Sing Bing Kang [15].

Figure 4.9: Output of the proposed HDR ghost-free merging method for *Complex Scene 1* of dataset [16] (left). Ghosted HDR image is shown on the right. Previews of various algorithm results are shown at the bottom. No de-ghosting (A), Silk et al. [40] (B), Sen et al. [38] (C), Photoshop (D), Photomatix (E) and proposed algorithm (F). The previews A to E are published as a part of a Karaduzovic dataset [16].

4.3.1 Dataset evaluation and comparison

I performed the evaluation on datasets [16, 46, 47], containing sequences of images of various scenes and different types of motion. The results provide a comparison of the proposed method with generally more precise and computationally demanding methods, commonly based on optical flow, which were not even included into the related work due to their complexity and high computational demands.

One of the datasets [16] contains multiple scenes with artificial objects movements. Its advantage consists in the existence of the ground truth image, which allows a comparison to the results as well as to many results of various published methods [11, 38, 40]. Figures 4.5 and 4.9 show the capabilities of the proposed method, showing that it provides results visually comparable to optical flow based methods.

Table 4.3: Results of the „Dynamic Region Dynamic Range“ metric proposed by Tursun [46] and evaluated on their dataset. The metric evaluates the resulting dynamic range within regions containing movement; the higher the value, the better.

Metric „DR“	[9]	[38]	[40]	none	**This work**
Cafe	**2.63**	2.61	2.60	2.47	2.42
FastCars	1.12	1.18	1.10	1.10	**1.38**
Flag	1.40	1.50	1.49	1.45	**1.59**
Gallery1	1.59	1.59	1.56	1.55	**1.70**
Gallery2	2.41	**2.56**	2.14	2.29	2.05
LibrarySide	1.78	1.93	1.60	1.76	**3.20**
Shop1	2.20	2.39	2.00	2.10	**2.42**
Shop2	2.68	2.72	**2.89**	2.55	2.42
WalkingP.	1.94	**2.07**	1.83	2.05	1.58

Tursun et al. [46, 47] published two datasets and proposed metrics for evaluation of HDR de-ghosting quality. The evaluated samples from the datasets are shown in Figure 4.4 and the HDR quality metric [46] is evaluated in Table 4.3. The metric evaluates the dynamic range achieved inside the motion regions, considering also the correctness of the de-ghosting. The image sets, in which we got worse results than other algorithms, were successfully de-ghosted anyway; however, the worse results were probably caused by losses in the dynamic range. Evaluation of the proposed method on these datasets also proves that the proposed method is generally usable for sequences larger than two/three images, commonly used in cameras. In all the referenced datasets [16, 46, 47], the proposed algorithms achieved results visually comparable or even better than more complex algorithms (see Figure 4.10). However, the proposed method and also many HDR de-ghosting methods may yield artifacts in regions where the moving objects in the reference image are poorly-exposed, as Tursun et al. concluded [46].

Another metric that have been found useful is HDR-VDP2 by Mantiuk et al. [24]. The metric evaluates the visibility and quality differences in image pairs and represents a probability that an average observer will notice a difference in the images in the pair (see Figure 4.11). The essential problem for the metric evaluation is the absence of ground truth images. Applying this metric on image sets without ground truth reference seems useless, as even the state-of-the-art algorithms may fail in ghost detection and/or changes in the image quality e.g. by bluring of motion regions (see top of Figure 4.10). As a result, the metric output obtained on such data does not have any meaningful value.

Table 4.4: Evaluation of the HDR-VDP2 [24] metric on a „complex“ scene from Karaduzovic's [16] dataset.

	scene1	scene2	scene3	scene 4
Q	73.24	76.20	82.83	71.58

Karaduzovic's [16] dataset contains ground truth images, because it contains scenes with artificial object motion. The metric was evaluated on „complex“ scenes and used the HDR merged from the ground truth sequence as a reference. The ground truth sequence is processed also by our algorithm (with de-ghosting disabled) to eliminate the effect of unrelated image enhancements and enables the direct comparison of the resulting HDR images. Table 4.4 contains an overall „quality“ metric of the produced ghost-free HDR output according to HDR-VDP2 [24] metric. Figure 4.11 shows „scene 1“ with highlighted differences between ground truth HDR and ghost-free HDR.

Figure 4.10: Figure shows scene „Cafe“ from Tursun's [46] dataset processed with Sen [38](top) and the proposed ghost-free algorithm(bottom). Sen [38] produce a heavily blured image, which precludes the HDR-VDP metric [24]; moreover, the de-ghosting method fails (see marked areas, where objects are shadowed and blured).

4.4 Performance evaluation

The performance of the algorithm on the relevant platforms is summarised in Table 4.5. Only the core parts, the certainty map creation and HDR merging were benchmarked, without including any data preprocessing time – please assume that at least in the FPGA and GPU implementations, the images are transferred into the memory using DMA in the background, without any performance losses. With the proposed optimisations, the algorithm is single-pass only. Table 4.5 compares the performance of the proposed Ghost-free merging of three LDR images on FPGA, SoC GPU and CPU platforms. In the case of FPGA, the design achieves target frequency of 200MHz and is fully pipelined; therefore, it allows the production of result pixels every clock cycle. Unlike in the sequential CPU and GPU processing, increasing the amount of work that the FPGA pipeline performs leads to

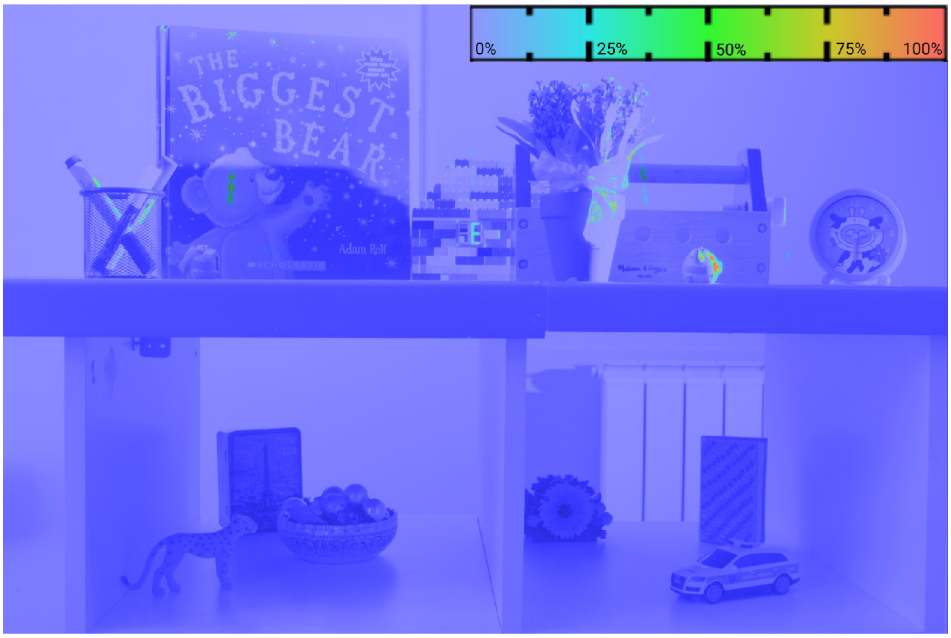

Figure 4.11: Figure shows „scene 1“ from Karaduzovic's [16] dataset processed by the HDR-VDP2 [24] metric. The colour bar reflects the probability that an average observer will notice a difference between ghost-free HDR and ground truth HDR. De-ghosted HDR visual quality, according to HDR-VDP2 metric [24] is 73.24 (see Table 4.4).

consumption of more resources and prolonging the processing pipeline, which has a negative influence on latency; however, the data throughput remains the same (see Table 4.5).

Table 4.5: The table compares the performance of the proposed ghost-free merging of 3 LDR images (Figure 4.2) with a resolution of 1920 × 1080 on following platforms: FPGA Xilinx Zynq, embedded CPU and GPU Nvidia Tegra TX2 and CPU Intel Core i7-3770 (single core).

	FPGA	TX2 GPU	TX2 CPU	CPU
Certainty map [ms]	10.3	1.59	45.9	16.6
Merging [ms]	10.3	4.58	112.3	23.0
Total [ms]	10.3	6.17	158.2	39.6
Overall FPS	96.45	162.07	6.32	25.25

I chose the HDR camera prototype by Nosko et al. [31] for the integration of the proposed ghost-free method. I chose this camera prototype due to its compact size, presence of a FullHD resolution CMOS (and optionally with even higher resolution) and presence of Xilinx Zynq SoC. Moreover, I participated on the development of Nosko's prototype as well. The proposed method and its implementation into architecture Nosko et al. [31] have not been published yet; however, the performance parameters are already known. Moreover, I have designed the new ghost-free merging block as a 1 to 1 replacement of previously published HDR merging with ghost removal[31], then the overall design shares all other features, such as advanced local tonemapping.

Figure 4.12: A car passing by the camera – ghosted HDR (left) and result of the proposed ghost-free merging algorithm (right).

The proposed algorithm should also be easily integrated into existing solutions of HDR acquisition devices by Popadic et al. [33], Lapray [20], Nosko [30] and probably others which are all based on pixel weighting, similar to Debevec and Malik [4].

The implementation on Nosko's platform allowed direct comparison with other architectures, however, probably only related work, which implements multi-exposure ghost-free HDR acquisition on an embedded device, in this case on FPGA, was proposed by Bouderbane et al. [3]. The Table 4.6 provides an overall comparison to Bouderbane solution.

Table 4.6: Comparison of main parameters of proposed solution to Bouderbane et al. [3].

	Proposed pipeline on [31]	Bouderbane
Platform	Zynq 7020	Virtex 6
Resolution	1920×1080	1280×1024
TMO	Durand (Local)	Duan (Global)
Arithmetic	Fixed point	Floating point
Maximum speed	$200Mhz$	$114.2Mhz$
Throughput	$200Mpix/s$	$114.2Mpix/s$
Framerate	$96FPS$	$60FPS$

As can be seen on Table 4.6, the design outperforms the Bouderbane architecture in all parameters. The design is fully pipelined, producing HDR pixel in every clock cycle. Target clocking frequency is 200MHz, which enables the acquisition of FullHD HDR images at up to 96FPS. The fixed point arithmetic has a positive contribution to clocking frequency, low resource requirement and low power consumption. At the same time, the calculations are still performed in high accuracy, which was summarized in Section 4.2. The platforms are both implemented in different FPGA family, where Zynq is part of 7th and Virtex part of 6th series from Xilinx. However, Virtex is a High end, while Zynq(Artix) only mid or low-end FPGA. The Table 4.2 shows the overall FPGA resource consumption for both complete camera solutions; Bouderbane does not provide separately the resources consumed by deghosting and HDR merging circuits. Please notice that local tonemapping operator with bilateral filter (on Nosko's platform) require more than 1/3 of overall LUT and Register resources and consumes most of BRAM and DSP resources.

The following part of the evaluation aims to compare proposed algorithm performance to related state-of-the-art implementations. The essential problem is those relevant algorithms are not generally available in the form of code or executable; therefore the performance comparison is rather limited to algorithms, where I managed to get source codes to run or where I get required information from relevant articles.

The comparison from Table 4.7 confirms that deghosting algorithms do not generally achieve real-time performance. Depending on the algorithm and desired deghosting quality (if available), the process can take from tens of seconds to more than ten minutes. Certain algorithms, such as Grosh [9] which have similar computation complexity as the proposed algorithm, do achieve relatively low processing time, however, the output is not deghosted very well, as shown in evaluation by Tursun et al. [46]. The Table 4.7 shows, that proposed algorithm is running much faster than any compared algorithm and is even twice faster than the algorithm by Grosch [9]. Please note, that average times from Table 4.7 are valid for sets of nine images with 4MPix resolution.

Table 4.7: Table results obtained from Tursun et al. [46]. The table shows the average processing time of deghosting algorithms on Tursun dataset [46]. Source image sets contain 9 images with 4MPix resolution (Sen and Khan merges three images with 1024×683 only). Tursun achieved these results on CPU Intel i7-3770; however, it is not specified whether the algorithms utilized all CPU cores or not. Proposed method result is benchmarked as single-core.

	Proposed	Grosch[9]	Khan[17]	Sen[38]	Silk[40]	Hu[11]	Tursun[46]
Avg. time[s]	0.48	1.04	616.45	209.78	14.33	230.36	7.09

The Table 4.8 compares the performance of proposed algorithm and algorithms by Pece et al. [32] and Min et al. [26]. I chose these algorithms to compare due to its possible easy implementation on FPGA; the author claims that the algorithm does not use the multiplication, division, and floating-point operations for object motion detection. Moreover, all operations, including the histogram calculation, are relatively easily implementable on FPGA. Also, the deghosting ability presented in the paper seemed to be very promising.

Table 4.8: Table compares the performance of the proposed algorithm with algorithms by Pece et al. [32] and Min et al. [26]. Algorithms were benchmarked on CPU Intel i7-3770 (single thread) on a scene from Figure 4.1. Source image set contains three images with FullHD resolution (1920×1080).

	Proposed	Pece et al. [32]	Min et al. [26]
Avg. time [ms]	39.6	193	206

I implemented the proposed algorithm and algorithms by Pece et al. [32] and Min et al. [26] in C++, so the performance presented in Table 4.8 should be comparable. As could be observed, the proposed algorithm is almost five-times faster. Moreover, the deghosting results are also better, as can be observed in Figure 4.6. The explanation of why these methods do not achieve good deghosting results is in the Subsection 4.3.

The Table 4.9 presents the performance results obtained by Yan et al. [54] and presented in their article. They compared a number of algorithms and benchmarked them on CPU

Intel i7 and GPU NVIDIA GeForce GTX 1080Ti. As can be observed, the proposed algorithm achieved better performance on CPU architecture (single-core) than others on even high-end GPUs. Yan's [54] and Wu's [52] CNN-based merging are relatively fast; however, they run on high-end GPU, which consumes much more energy than CPU (up to 280W).

Table 4.9: Table results obtained from Yan et al. [54]. Table shows average processing time of deghosting algorithms on three images with resolution 1000×1500. CPU used is Intel i7 (not further specified by Yan), GPU used is NVIDIA GeForce GTX 1080Ti.

Algorithm	Proposed	Yan [54]	Wu [52]	Kalantari [14]	Sen [38]	Wu [51]
Platform	CPU	GPU	GPU	CPU+GPU	CPU	CPU
Time [s]	0.022	0.31	0.24	29.14	61.81	79.77

4.4.1 Power consumption summary

This subsection provides a brief comparison of the power consumption of selected algorithms. The overall energy in Joules per one frame is estimated from processing time and TDP of a processor; alternatively, FPGA consumption is estimated by the design tools based on the amount of logic used. The Table 4.10 and Table 4.11 contains the energy requirements converted from performance summary in Table 4.7 and Table 4.8.

Table 4.10 presents that the proposed algorithm has the lowest power consumption among all of the measured algorithms. The results were achieved on the dataset published by Tursun[46], which contains sequences of nine images with 4MPix resolution. It is not specified by Tursun [46] whether the algorithms utilize single or multiple cores; therefore, I assume only single-core implementation as a lower estimate of the possible power consumption. Proposed method result is benchmarked as single-core.

Table 4.10: Table of energy consumption per HDR frame, derived from Table 4.7. The table shows average energy consumption for processing one HDR frame of deghosting algorithms on Tursun dataset [46].

	Proposed	Grosch[9]	Khan[17]	Sen[38]	Silk[40]	Hu[11]	Tursun[46]
Avg. energy [J]	12	26	15411	5244	358	5759	177

Table 4.11: Table compares the energy consumption for processing one HDR frame by proposed algorithm with algorithms by Pece et al. [32] and Min et al. [26]. Algorithms were benchmarked on CPU Intel i7-3770 (single thread) on scene from Figure 4.1. Source image set contains three images with FullHD resolution (1920×1080).

	Proposed	Pece et al. [32]	Min et al. [26]
Avg. energy [J]	0.99	4.82	5.15

The average consumption is 12J per one HDR frame, which is 46% of the second least demanding algorithm by Grosh [9]. Moreover, proposed algorithm demands are measured for single-core processing only, whether the data provided by Tursun [46] are not specified whether were achieved on single-core only; however, the results in Table 4.7 and Table 4.10 assumes they are.

Table 4.12: The table compares the power consumption of the proposed algorithm on the CPU and FPGA platform and shows the estimated energy required for the ghost-free merge of the sequence of three FullHD images. Please note that consumption of Camera Nosko [31] includes camera as a whole.

	Consumption [W]	Energy per frame [J]	comp. to CPU[%]
CPU Intel i7-3770	25W (single core)	0.99	—
Camera Nosko [31] (30FPS)	∼8W	0.266	26.9
Camera Nosko [31] (96 FPS)	∼8W	0.083	8.3
Tegra TX2 - GPU only	15W	0.093	9.3
Proposed - FPGA only	1,1W	0.0115	1.16

Table 4.13: The table compares performance of proposed Ghost-free merging of three LDR images (Figure 4.2) of resolution 1920×1080 on FPGA and CPU platforms. Data are selected from Table 4.5.

	FPGA Xilinx Zynq	CPU Intel Core i7-3770
Ghost det. [ms]	10.3	16.6
Merging [ms]	10.3	23.0
Total [ms]	10.3	39.6
Overall FPS	96.45	25.25

The HDR camera by Zemcik et al. [55] achieved overall power consumption of 12W and the HDR cameras by Nosko et al. [30, 31] even less, total 8W. Based on the performance summarized in Table 4.5 and assuming the maximum speed of 96.4FPS, the camera Nosko et al. [31] with proposed algorithm consumes $0,083J$ per frame. The power consumption of CPU Intel Core i7-3770 was measured in single-core load (running proposed algorithm) and achieved 25W. CPU achieved framerate of 25.25 FPS, which results in consumption approximately $0.99J$ per frame; note, please, that the difference between standby and the full load was power consumption measured, which shows only the desired dynamic part of power consumption.

In summary, the HDR camera by Nosko et al [31] with proposed algorithm consumes only 8.4% comparing to the CPU implementation. Moreover, most of the power consumption of HDR camera is spent on camera hardware, including CMOS chip and H.264 encoder, while the consumption of the FPGA itself consumes approx. $1,1W$ only (estimation by Xilinx Vivado tool). This result is much more favourable for FPGA, but the comparison is fairer because it compares only the „computing“ elements. The energy spent on one frame drops to approx. $0.011J$, which is little above 1% of the energy consumed by CPU.

4.5 Validation and scientific contribution

at the beginning of this Chapter 4 it was stated that the scientific contribution of this thesis should be the proof of the following hypothesis: *A multi-exposure ghost-free HDR acquisition algorithm comparable to the state-of-the-art algorithms in quality can be designed for an embedded hardware device and achieves a real-time performance at high resolution.*

In Section 4.1, I proposed a **Ghost-free** HDR acquisition algorithm implementable on **FPGA**. This method was implemented and its description is included in Section 4.2.

The proposed Ghost-free algorithm produces a visual output **comparable to the State-of-the-art** as evaluated in Section 4.3. Finally, the proposed design achieves more than **real-time performance** of 96FPS on fullHD resolution as summarized in Section 4.4. Therefore, I consider the hypotheses validated.

In more detail, the proposed novel ghost-free HDR merging algorithm is suitable for real-time implementation in embedded devices. The algorithm is well suitable for implementation on many platforms, including the CPU and GPU based platforms. However, the aim of contribution was a successful implementation of such an algorithm into FPGA, which was experimentally proved in Section 4.2. Also, the target performance, which is real-time processing on FullHD resolution, was fulfilled, since the proposed solution is able to run on up to 96 FPS (Table 4.5). At the same time, the proposed solution outperforms the FPGA solution of only state-of-the-art FPGA implementation of Bouderbane et al. [3], which achieved only 60 FPS on HD resolution (1280×1024).

The performance comparison with most of the state-of-the-art algorithms requires a CPU reference implementation. The performance evaluation in Section 4.4 shows that the algorithm performs well even on CPU; single-core implementation achieves up to 25.25FPS, as shown in Table 4.5, which is the best result. The second least demanding state-of-the-art algorithm, according to Table 4.7 is from Grosh [9]. Under the same conditions and on the same dataset[46], the proposed algorithm is faster by approx. 54%. Table 4.9 further compares the proposed algorithm with the latest and GPU accelerated state-of-the-art algorithms. Finally, Table 4.11 compares the CPU performance of related algorithms which I reviewed to be suitable for FPGA implementation. The proposed algorithm achieved the best result and is 4.8 times faster than the fastest FPGA implementable algorithm.

Table 4.7 and Table 4.11 compares the CPU power consumption to state-of-the-art algorithms and, linearly with performance, requires only 46% of power comparing to the second least demanding algorithm by Grosh [9].

The CPU implementation itself is so fast that almost accomplished the real-time requirement; however, the real benefits of the method stand out along with FPGA acceleration, which fundamentally affects the performance and power consumption. The Table 4.12 and Table 4.13 shows the effectiveness and benefits of FPGA acceleration of proposed algorithm. While the FPGA implementation offers almost 4-times higher performance comparing to CPU (25.25 FPS) and reaches the 96FPS, the energy consumption drops by 98,86% per frame. These parameters should be even much better in case of ASIC chip production (or integration into an existing chip, e.g. as an accelerator block), for which the FPGA reference implementation is necessary. However, I did not have such funding and contacts to ASIC manufacturing facility.

The comparison to the state-of-the-art algorithms (Section 4.3) and evaluation of HDR datasets (Section 4.3.1) shows, that proposed algorithm is performing ghost-free HDR merging well and the ghost effect is removed, at the same time have better results and is much more robust than related algorithms. The results are even comparable to the state-of-the-art optical flow-based algorithm, which belongs to the class of performance demanding, offline processing algorithms.

Chapter 5

Conclusion

In this dissertation work, I focused on the HDR acquisition on embedded devices. The main goal of this thesis was the proof that a multi-exposure ghost-free HDR acquisition algorithm comparable to the state-of-the-art algorithms in quality can be designed for an embedded hardware device and achieves a real-time performance at high resolution. This hypothesis I considered as validated, which was stated in Section 4.5.

I experimentally proved the hypothesis by the successful implementation of proposed ghost-free HDR merging algorithm (Section 4.1) on FPGA based embedded design (Section 4.2). The proposed implementation achieved the expected parameters and is capable of running faster than real-time, up to 96FPS at FullHD resolution (Section 4.4). At the same time, the algorithm produces visual results comparable to the state-of-the-art, as evaluated in the Section 4.3.

The performance evaluation in Section 4.4 shows, that the algorithm performs well even on CPU; single core implementation achieves up to 25.25FPS, which is very fast and multicore CPU could achieve real-time performance as well. Achieved results shows, that even CPU implementation outperformed all the related algorithms.However, essential benefit of this method stand out along with FPGA implementation, which fundamentally affects the power consumption, which is only approx. 1,1% of power comparing to the CPU, as summarized in Section 4.4.1.

The comparison to the state-of-the-art algorithms (Section 4.3) and evaluation of HDR datasets (Section 4.3.1) shows, that proposed algorithm is performing ghost-free HDR merging well and the ghost effect is removed, at the same time have better results and is much more robust than related algorithms. The results are even comparable to the state-of-the-art optical flow-based algorithm, which belongs to the class of performance demanding, offline processing algorithms.

In the future, I would like to continue in this topic and work on the applicability of the proposed solution in practice and in the commercial field, which already started within the Czech and EU research projects.

Bibliography

[1] *Luminance HDR* [http://qtpfsgui.sourceforge.net/]. 2017.

[2] Bouderbane, M., Dubois, J., Heyrman, B., Lapray, P.-J. and Ginhac, D. Ghost removing for HDR real-time video stream generation. In: Kehtarnavaz, N. and Carlsohn, M. F., ed. *Real-Time Image and Video Processing 2016*. SPIE, 2016, vol. 9897, p. 112 – 117. DOI: 10.1117/12.2230313. Available at: https://doi.org/10.1117/12.2230313.

[3] Bouderbane, M., Lapray, P.-J., Dubois, J., Heyrman, B. and Ginhac, D. Real-time ghost free HDR video stream generation using weight adaptation based method. In:. September 2016, p. 116–120. DOI: 10.1145/2967413.2967439.

[4] Debevec, P. E. and Malik, J. Recovering High Dynamic Range Radiance Maps from Photographs. In: *ACM Trans. Graph.* 1997. SIGGRAPH '97.

[5] Duan, J., Bressan, M., Dance, C. and Qiu, G. Tone-mapping High Dynamic Range Images by Novel Histogram Adjustment. *Pattern Recogn.* New York, NY, USA: Elsevier Science Inc. may 2010, vol. 43, no. 5, p. 1847–1862. DOI: 10.1016/j.patcog.2009.12.006. ISSN 0031-3203. Available at: http://dx.doi.org/10.1016/j.patcog.2009.12.006.

[6] Durand, F. and Dorsey, J. Fast Bilateral Filtering for the Display of High-Dynamic-Range Images. In: *ACM Trans. Graph.* ACM, 2002. SIGGRAPH '02.

[7] Ferradans, S., Bertalmío, M., Provenzi, E. and Caselles, V. Generation of HDR Images in Non-static Conditions based on Gradient Fusion. In: *VISAPP (1)*. 2012, p. 31–37.

[8] Gallo, O., Gelfandz, N., Chen, W.-C., Tico, M. and Pulli, K. Artifact-free High Dynamic Range imaging. In: *2009 IEEE International Conference on Computational Photography (ICCP)*. April 2009, p. 1–7. DOI: 10.1109/ICCPHOT.2009.5559003.

[9] Grosch, T. Fast and robust high dynamic range image generation with camera and object movement. 2006.

[10] Grossberg, M. D. and Nayar, S. K. Determining the camera response from images: what is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Nov 2003, vol. 25, no. 11, p. 1455–1467. DOI: 10.1109/TPAMI.2003.1240119. ISSN 0162-8828.

[11] Hu, J., Gallo, O., Pulli, K. and Sun, X. HDR Deghosting: How to Deal with Saturation? In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. 2013.

[12] Jacobs, K., Loscos, C. and Ward, G. Automatic High-Dynamic Range Image Generation for Dynamic Scenes. *IEEE Computer Graphics and Applications*. March 2008, vol. 28, no. 2, p. 84–93. DOI: 10.1109/MCG.2008.23. ISSN 0272-1716.

[13] Jinno, T. and Okuda, M. Multiple exposure fusion for high dynamic range image acquisition. *IEEE Transactions on Image Processing*. IEEE. 2011, vol. 21, no. 1, p. 358–365.

[14] Kalantari, N. K. and Ramamoorthi, R. Deep High Dynamic Range Imaging of Dynamic Scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*. 2017, vol. 36, no. 4.

[15] Kang, S. B., Uyttendaele, M., Winder, S. and Szeliski, R. High Dynamic Range Video. *ACM Trans. Graph.* New York, NY, USA: ACM. july 2003, vol. 22, no. 3, p. 319–325. ISSN 0730-0301.

[16] Karaduzovic Hadziabdic, K., Hasic, T. J. and Mantiuk, R. K. Multi-exposure image stacks for testing HDR deghosting methods. 2017.

[17] Khan, E. A., Akyuz, A. O. and Reinhard, E. Ghost removal in high dynamic range images. In: IEEE. *2006 International Conference on Image Processing*. 2006, p. 2005–2008.

[18] Lapray, P. J., Heyrman, B., Rossé, M. and Ginhac, D. HDR-ARtiSt: High dynamic range advanced real-time imaging system. In: *2012 IEEE International Symposium on Circuits and Systems*. May 2012, p. 1428–1431. DOI: 10.1109/ISCAS.2012.6271513. ISSN 0271-4302.

[19] Lapray, P.-J., Heyrman, B. and Ginhac, D. HDR-ARtiSt: A 1280x1024-pixel Adaptive Real-time Smart camera for High Dynamic Range video. In: *SPIE Photonics Europe*. Brussels, Belgium: [b.n.], April 2014. Available at: https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01196636.

[20] Lapray, P.-J., Heyrman, B. and Ginhac, D. HDR-ARtiSt: an adaptive real-time smart camera for high dynamic range imaging. *J. Real-Time Image Process.* 2016, vol. 12, no. 4, p. 747–762. ISSN 1861-8219.

[21] Liu, C. et al. *Beyond pixels: exploring new representations and applications for motion analysis*. 2009. Dissertation. Massachusetts Institute of Technology.

[22] Mandel, L. Fluctuations of photon beams: the distribution of the photo-electrons. *Proceedings of the Physical Society*. IOP Publishing. 1959, vol. 74, no. 3, p. 233.

[23] Mann, S., Lo, R. C. H., Ovtcharov, K., Gu, S., Dai, D. et al. Realtime HDR (High Dynamic Range) video for eyetap wearable computers, FPGA-based seeing aids, and glasseyes (EyeTaps). In: *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. April 2012, p. 1–6. DOI: 10.1109/CCECE.2012.6335012. ISSN 0840-7789.

[24] Mantiuk, R., Kim, K. J., Rempel, A. G. and Heidrich, W. HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions. *ACM Trans. Graph.* New York, NY, USA: ACM. july 2011, vol. 30,

no. 4, p. 40:1–40:14. DOI: 10.1145/2010324.1964935. ISSN 0730-0301. Available at: http://doi.acm.org/10.1145/2010324.1964935.

[25] Mertens, T., Kautz, J. and Reeth, F. V. Exposure Fusion. In: *Computer Graphics and Applications, 2007. PG '07. 15th Pacific Conference on*. Oct 2007, p. 382–390. ISSN 1550-4085.

[26] Min, T.-H., Park, R.-H. and Chang, S. Histogram based ghost removal in high dynamic range images. In: IEEE. *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*. 2009.

[27] Min, T.-H., Park, R.-H. and Chang, S. Noise reduction in high dynamic range images. *Signal, Image and Video Processing*. 2011, vol. 5, no. 3. DOI: 10.1007/s11760-010-0203-7. ISSN 1863-1711. Available at: https://doi.org/10.1007/s11760-010-0203-7.

[28] Mitsunaga, T. and Nayar, S. K. Radiometric self calibration. In: *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*. 1999, vol. 1, p. 380 Vol. 1. ISSN 1063-6919.

[29] Musil, M., Nosko, S. and Zemcik, P. De-ghosted HDR video acquisition for embedded systems. *Journal of Real-Time Image Processing*. Springer. 2020, p. 1–10.

[30] Nosko, S., Musil, M., Musil, P. and Zemcik, P. True HDR Camera with Bilateral Filter Based Tone Mapping. In: *Proceedings of the 33rd Spring Conference on Computer Graphics*. New York, NY, USA: ACM, 2017, p. 15:1–15:9. SCCG '17. DOI: 10.1145/3154353.3154367. ISBN 978-1-4503-5107-2. Available at: http://doi.acm.org/10.1145/3154353.3154367.

[31] Nosko, S., Musil, M., Zemcik, P. and Juranek, R. Color HDR video processing architecture for smart camera. *Journal of Real-Time Image Processing*. Jul 2018. DOI: 10.1007/s11554-018-0810-z. ISSN 1861-8219. Available at: https://doi.org/10.1007/s11554-018-0810-z.

[32] Pece, F. and Kautz, J. Bitmap movement detection: HDR for dynamic scenes. In: IEEE. *Visual Media Production, 2010 Conference on*. 2010, p. 1–8.

[33] Popadić, I., Todorović, B. M. and Reljin, I. Method for HDR-like imaging using industrial digital cameras. *Multimedia Tools and Applications*. May 2017, vol. 76, no. 10, p. 12801–12817. DOI: 10.1007/s11042-016-3692-8. ISSN 1573-7721. Available at: https://doi.org/10.1007/s11042-016-3692-8.

[34] Raman, S., Kumar, V. and Chaudhuri, S. Blind De-ghosting for Automatic Multi-exposure Compositing. In: *ACM SIGGRAPH ASIA 2009 Posters*. USA: ACM, 2009. SIGGRAPH ASIA '09.

[35] Reinhard, E., Stark, M., Shirley, P. and Ferwerda, J. Photographic Tone Reproduction for Digital Images. *ACM Trans. Graph.* New York, NY, USA: ACM. july 2002, vol. 21, no. 3, p. 267–276. ISSN 0730-0301.

[36] Robertson, M. A., Borman, S. and Stevenson, R. L. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *J. Electronic Imaging*. 2003, vol. 12, no. 2, p. 219–228.

[37] Sakakibara, M., Kawahito, S., Handoko, D., Nakamura, N., Satoh, H. et al. A high-sensitivity CMOS image sensor with gain-adaptive column amplifiers. *J. of Solid-State Circuits*. May 2005, vol. 40, no. 5. ISSN 0018-9200.

[38] Sen, P., Kalantari, N. K., Yaesoubi, M., Darabi, S., Goldman, D. B. et al. Robust Patch-Based HDR Reconstruction of Dynamic Scenes. *ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH Asia 2012)*. 2012, vol. 31, no. 6, p. 203:1–203:11.

[39] Sidibe, D., Puech, W. and Strauss, O. Ghost detection and removal in High Dynamic Range Images. In: *2009 17th European Signal Processing Conference*. 2009, p. 2240–2244.

[40] Silk, S. and Lang, J. Fast High Dynamic Range Image Deghosting for Arbitrary Scene Motion. In: *Proceedings of Graphics Interface 2012*. Toronto, Ont., Canada, Canada: Canadian Information Processing Society, 2012, p. 85–92. GI '12. ISBN 978-1-4503-1420-6. Available at: http://dl.acm.org/citation.cfm?id=2305276.2305291.

[41] Srikantha, A. and Sidibé, D. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*. 2012, vol. 27, no. 6, p. 650 – 662. DOI: https://doi.org/10.1016/j.image.2012.02.001. ISSN 0923-5965. Available at: http://www.sciencedirect.com/science/article/pii/S0923596512000306.

[42] Tamburrino, D., Alleysson, D., Meylan, L. and Süsstrunk, S. Digital camera workflow for high dynamic range images using a model of retinal processing. In: *IST/SPIE Electronic Imaging: Digital Photography IV*. 2008, vol. 6817, LCAV-CONF-2007-030.

[43] Tang, X., Qian, Y., Kong, X. and Wang, H. A high-dynamic range CMOS camera based on dual-gain channels. *Journal of Real-Time Image Processing*. may 2019. DOI: 10.1007/s11554-019-00877-8.

[44] Tao Ai, Ali, M. A., Steffan, G., Ovtcharov, K., Zulfiqar, S. et al. Real-time HDR video imaging on FPGA with compressed comparametric lookup tables. In: *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*. 2014, p. 1–6.

[45] Tocci, M. D., Kiser, C., Tocci, N. and Sen, P. A Versatile HDR Video Production System. In: *ACM SIGGRAPH 2011 Papers*. USA: [b.n.], 2011. SIGGRAPH '11. ISBN 978-1-4503-0943-1.

[46] Tursun, O. T., Akyüz, A. O., Erdem, A. and Erdem, E. The State of the Art in HDR Deghosting: A Survey and Evaluation. *Computer Graphics Forum*. 2015, vol. 34, no. 2. DOI: 10.1111/cgf.12593. ISSN 1467-8659.

[47] Tursun, O. T., Akyüz, A. O., Erdem, A. and Erdem, E. An Objective Deghosting Quality Metric for HDR Images. *Comput. Graph. Forum*. Chichester, UK: The Eurographs Association & John Wiley & Sons, Ltd. may 2016, vol. 35, no. 2, p. 139–152. DOI: 10.1111/cgf.12818. ISSN 0167-7055. Available at: https://doi.org/10.1111/cgf.12818.

[48] Ureña, R., Martínez Cañada, P., Gómez López, J. M., Morillas, C. and Pelayo, F. Real-time tone mapping on GPU and FPGA. *EURASIP Journal on Image and Video Processing*. 2012, vol. 2012, no. 1, p. 1. ISSN 1687-5281.

[49] Vytla, L., Hassan, F. and Carletta, J. E. A Real-time Implementation of Gradient Domain High Dynamic Range Compression Using a Local Poisson Solver. *J. Real-Time Image Process.* Secaucus, NJ, USA: Springer-Verlag New York, Inc. june 2013, vol. 8, no. 2, p. 153–167. ISSN 1861-8200.

[50] Wang, C. and Tu, C. An exposure fusion approach without ghost for dynamic scenes. In: *2013 6th International Congress on Image and Signal Processing (CISP)*. Dec 2013, vol. 2, p. 904–909. DOI: 10.1109/CISP.2013.6745293.

[51] Wu, S., Xie, S., Rahardja, S. and Li, Z. A robust and fast anti-ghosting algorithm for high dynamic range imaging. In: *2010 IEEE International Conference on Image Processing*. Sept 2010, p. 397–400. DOI: 10.1109/ICIP.2010.5654196. ISSN 1522-4880.

[52] Wu, S., Xu, J., Tai, Y.-W. and Tang, C.-K. Deep high dynamic range imaging with large foreground motions. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, p. 117–132.

[53] Xie, D. and Wang, Y. High definition wide dynamic video surveillance system based on FPGA. In: *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. 2017, p. 2403–2407.

[54] Yan, Q., Zhang, L., Liu, Y., Zhu, Y., Sun, J. et al. Deep HDR Imaging via A Non-Local Network. *IEEE Transactions on Image Processing*. 2020, vol. 29, p. 4308–4322.

[55] Zemcik, P., Musil, P. and Musil, M. Real-time HDR video processing and compression using an FPGA. *High Dynamic Range Video: Concepts, Technologies and Applications*. 2016, p. 145–154.

[56] Zhang, W. and Cham, W.-K. Gradient-directed composition of multi-exposure images. In: IEEE. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, p. 530–536.

[57] Zhao, H., Shi, B., Fernandez Cull, C., Yeung, S.-K. and Raskar, R. Unbounded High Dynamic Range Photography using a Modulo Camera. In: *ICCP*. 2015.

[58] Zitová, B. and Flusser, J. Image registration methods: a survey. *Image and Vision Computing*. 2003, vol. 21, no. 11, p. 977 – 1000. DOI: https://doi.org/10.1016/S0262-8856(03)00137-9. ISSN 0262-8856. Available at: http://www.sciencedirect.com/science/article/pii/S0262885603001379.

Author's Bibliography

Publications Related to Thesis

De-Ghosted HDR Video Acquisition for Embedded Systems

MUSIL Martin, NOSKO Svetozár and ZEMČÍK Pavel. De-Ghosted HDR Video Acquisition for Embedded Systems. Journal of Real-Time Image Processing, vol. 2020, no. 1, pp. 1-9. ISSN 1861-8200.

Color HDR video processing architecture for smart camera

NOSKO Svetozár, MUSIL Martin, ZEMČÍK Pavel and JURÁNEK Roman. Color HDR video processing architecture for smart camera. Journal of Real-Time Image Processing, vol. 2018, no. 1, pp. 1-12. ISSN 1861-8200

Real-Time HDR Video Processing and Compression Using an FPGA

ZEMČÍK Pavel, MUSIL Petr and MUSIL Martin. High Dynamic Range Video; Concepts, Technologies and Applications. High Dynamic Range Video, 1st Edition. London: Elsevier Science, 2016, pp. 145-154. ISBN 978-0-12-809477-8.

True HDR camera with bilateral filter based tone mapping

NOSKO Svetozár, MUSIL Martin, MUSIL Petr and ZEMČÍK Pavel. True HDR camera with bilateral filter based tone mapping. In: SCCG '17: Spring Conference on Computer Graphics 2017. Mikulov: Association for Computing Machinery, 2017, pp. 1-9. ISBN 978-1-4503-5107-2.

Publications Partly Related to Thesis

Cascaded Stripe Memory Engines for Multi-Scale Object Detection in FPGA

MUSIL Petr, JURÁNEK Roman, MUSIL Martin and ZEMČÍK Pavel. Cascaded Stripe Memory Engines for Multi-Scale Object Detection in FPGA. IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 1, pp. 267-280. ISSN 1051-8215.

High performance architecture for object detection in streamed video

ZEMČÍK Pavel, JURÁNEK Roman, MUSIL Martin, MUSIL Petr and HRADIŠ Michal. High Performance Architecture for Object Detection in Streamed Videos. In: Proceedings of FPL 2013. Porto: IEEE Circuits and Systems Society, 2013, pp. 1-4. ISBN 978-1-4799-0004-6.

Single-Loop Approach to 2-D Wavelet Lifting with JPEG 2000 Compatibility

BAŘINA David, MUSIL Martin, MUSIL Petr and ZEMČÍK Pavel. Single-Loop Approach to 2-D Wavelet Lifting with JPEG 2000 Compatibility. In: IEEE 27th International Symposium on Computer Architecture and High Performance Computing Workshops. Florianopolis: IEEE Computer Society, 2015, pp. 31-36. ISBN 978-1-4673-8621-0.