Real-Time HDR Imaging

Project Plan and Progress Report

*“Development of compact imaging system featuring expanded dynamic range for use on mobile robotic platforms”*

Author: Scott A. Rapson   
rapsa002@mymail.unisa.edu.au  
Student Number: 110094036

Supervisor: Dr. Russell Brinkworth russell.brinkworth@unisa.edu.au   
+61 8 830 23292

Co-Supervisor: Daniel Griffiths  
daniel.griffiths@unisa.edu.au

Contents

I. Abstract 3

II. Introduction 3

A. Literature Review 3

III. Project Methodology 3

A. Needs 3

B. Requirements 4

IV. Project Planning 4

A. Work Breakdown Structure 4

B. Resource Plan 5

C. Technical Risk 5

V. Progress Report 5

A. Pipeline Capabilities & Restrictions 5

B. Software Tests and Benchmarks 6

C. Hardware Research & Selection 7

VI. Conclusion 7

VII. References 7

# 

# Abstract

R

obotic systems are often required to operate in areas with uncontrolled incident lighting. Uncontrolled and outdoor environments present imaging challenges for camera systems due to the wide range of luminances present, often causing issues with accurately perceiving scene detail for human driven or autonomous navigation. Several approaches have been developed, however this project seeks to develop a software based frame blending and artifact correction tool leveraging commodity CMOS based webcams and single board computing platforms. The proposed system aims to provide a more useable image in situations with uneven incident light through the expansion of the dynamic range with multiple imaging sensors and is the focus of this project.

**Keywords**: HDR Imaging, Real-time processing, OpenGL, GPU Acceleration, robotics, telepresence.

# Introduction

Robotic platforms have seen a surge in popularity as their capabilities and versatility has expanded. As such, we now expect robotic platforms to navigate through and interact efficiently with complex uncontrolled environments [1, [2](#_ENREF_2)]. Conventional camera technologies have the ability to capture high resolution and high fidelity imagery, even at high frame rates if required, although most cameras are still limited by the dynamic range and bit depth. Research has shown improvements in imaging scenes with a high dynamic range by combining multiple exposures yet many of the existing techniques are expensive, suffer from restricted temporal resolution or aren’t designed for real-time video.

This project aims to develop a cost effective, mechanically simple system that is capable of real-time HDR imaging for use on untethered mobile robotic platforms. By utilising easily accessible commodity web cameras and a small single board computing platform with GPU accelerated software, image blending and artifact correction can be performed with significant performance gains over conventional CPU bound methods.

Although this project is self-standing, the proposed HDR camera system is being designed with the intent of integration into a full telepresence system developed by other students at the University of South Australia. The system will feature an immersive head-mounted display (HMD) with wireless IP based video streaming developed by Christopher Barr, and Anthony Sabino will develop an actuated camera mount designed to mimic the users’ head movements. When combined, the imaging improvements and headtracking will allow for intuitive wireless teleoperation for use in search and rescue applications on a range of ground and aerial platforms.

## Literature Review

HDR imaging systems have been a popular part of computer vision research dating back as early as the 1990’s [[3](#_ENREF_3)]. Continued research into HDR techniques has improved robustness, handling performance of local contrast and specular highlights whilst reducing artifacts like noise [[4](#_ENREF_4)].

To gain an increased range of illumination values, the most popular approach has been software based blending methods that integrate data from multiple source frames into a single HDR output frame [[4](#_ENREF_4)]. This approach is commonly used by photographers and can be found in newer smartphones however this technique typically relies on using the same sensor with varied exposures over time, sacrificing temporal resolution for imaging performance. Most software based blending solutions are designed for still photography and lack real-time video capabilities [[5](#_ENREF_5)].

To record HDR video with the same technique as for stills, the sampling rate of the sensor can be increased to gather additional frames whilst providing a HDR output video at smooth frame rates for video [[6](#_ENREF_6), [7](#_ENREF_7)], however suitable cameras with high resolution and high frame rates are expensive and incur light loss due to increased shutter speed. Alternative approaches have reduced reliance on high sampling rates, favouring hardware based solutions relying on specialised optics or custom sensors.

Larger improvements in real-time HDR imaging have been made through the use of multiple sensors allowing imaging without loss of temporal resolution or additional light loss. Examples has been demonstrated featuring the integration of several cameras allowing for the collection of multiple exposures simultaneously [[7](#_ENREF_7)], and this technique was improved though the use of a common optical path with beam splitter arrangements [[8](#_ENREF_8)].

Sensor technologies also allow for hardware HDR by inducing differences in exposure with neutral density filters [[8](#_ENREF_8), [9](#_ENREF_9)], split aperture systems [[10](#_ENREF_10)], or specialised gain control at the pixel level [[11](#_ENREF_11)]. These hardware-based approaches are the most compact and offer artifact free imaging but with significantly higher development and manufacturing costs.

Image processing and computer vision are computationally expensive due to large pixel count and the frequency at which image must be processed to maintain a real-time video output. Computer vision and image processing are well suited to the dedicated graphics architecture found on the GPU [[12](#_ENREF_12)] and similar alignment, blending and exposure equalisation techniques have been employed for mobile computer vision systems for panorama stitching [[13](#_ENREF_13)], image analysis [[14](#_ENREF_14)] and compression of HDR images to a LDR output [[5](#_ENREF_5), [15](#_ENREF_15)] for display on typical monitors.

# Project Methodology

To summarise the project and differences to prior-art, the needs and requirements for the system were defined, allowing further scope definition. These requirements will be evaluated against during the testing and performance validation stages.

## Needs

The system should allow the user to:

1. View a real-time video stream of typical outdoor environments without excessive clipping of highlight or shadow details.
2. Interpret and process the wide range of luminances from source data in preparation for human use.
3. Form factor suitable for use on mobile untethered platforms.

## Requirements

* Sample image data from multiple (>2) sensors or frames with different base exposures in order to acquire a wider range of luminances [N1].
* Image acquisition through to display is a continuous online process, characterised by low system latency under 200ms [N1].
* Output image can have a total dynamic range exceeding that of a single frame with no clipped highlights in naturally lit environments [N1].
* Operate with a minimum resolution useable for telepresence or simple computer vision tasks such as detection of shapes/objects and fiducial markers. A resolution of approximately 480x320 would be required for resolution of smaller features in such tasks. [N2].
* Has user selectable modes to tailor image acquisition and processing stages to best suit a given application [N2].
* Low mass (<250g for full system) and compact enclosure no larger than 150x150x150mm to simplify mounting and protect against damage. [N3].

To achieve the specified requirements whilst working with a restricted budget, the experimental apparatus will not feature a common optical path, opting instead for multiple discrete imaging sensors similar to the EyeTap viewing aid [[7](#_ENREF_7)]. To allow for more versatile use with near field objects, frame alignment and parallax error will be mitigated with software based correction using translate+warp techniques similar to those developed by Sawhney et al [[16](#_ENREF_16)], and the HDR data will be initially processed following the works of Smith & Krawczyk [[17](#_ENREF_17)] due to its superior output accounting for glare and luminance adaption for natural output [[18](#_ENREF_18)].

This implementation will be GPU based using similar techniques to the Akyüz GPU based rendering pipeline [[19](#_ENREF_19)], allowing for online processing due to accelerated correction and blending stages. Gradient domain compression [[15](#_ENREF_15)] will be used for display on computer monitors and the Oculus HMD. This combination of techniques will provide for a novel low cost solution for real-time HDR imaging for online uses such as tele-operation.

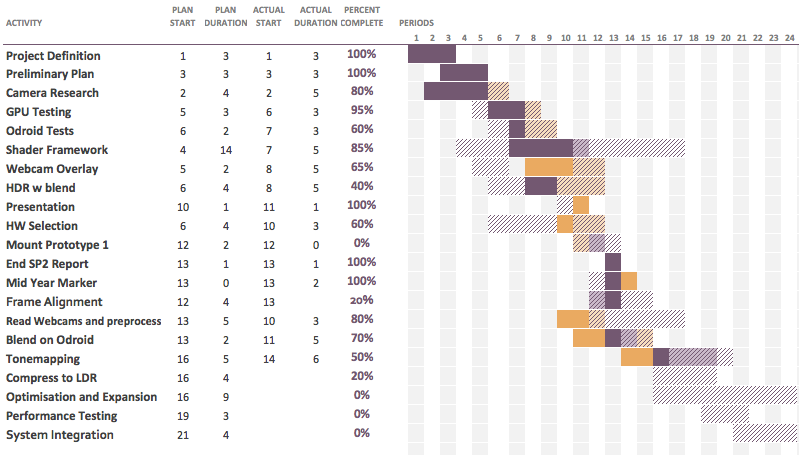


Fig 1: Proposed Schedule

# Project Planning

To successfully integrate with the other system components, the proposed camera system should be lightweight and compact whilst minimising power draw and cost. The software will be the most time consuming and important aspect of the project due to integration of several complicated computer graphics techniques, namely the frame alignment and distortion compensation stages, and compressing the HDR images into a LDR output for display on accessible monitors.

## Work Breakdown Structure

With a defined project scope, a proposed timeline was developed as shown in testing for feasibility is required to demonstrate the capabilities of the existing ODroid XU3 computers and webcams before purchasing more hardware. Additionally, evaluating restrictions in the OpenGL pipeline on the Mali T628 GPU will show likely bottlenecks early in the process to reduce unnecessary optimisation time.

The hardware will be purchased and multi-camera mount will be designed and produced in preparation of part arrival using spare hardware from the MechBot program and rapid manufacturing techniques including 3D printing. As mounting arrangement of the cameras will affect the parallax and optical error, multiple mounts will be evaluated to find the best compromise between horizontal & vertical displacement, and ease of calibration for alignment compensation.

Frame alignment and basic threshold based blending will be implemented once cameras have been mounted, quickly progressing to a demonstrable project with the remainder of time focussing on enhanced blending techniques for a human viewable output and optimisation with the aim of achieving higher resolution output. Most software stages are developed and completed around this time due to interlocking requirements (HDR image needed for LDR output, LDR output to evaluate the HDR results).

Testing and validation will be performed against the requirements defined in Project Methodology, performed in both controlled indoor and uncontrolled outdoor scenarios with a focus on latency reduction and improvements in special resolution.

Integration with IP based streaming solutions and headtracked gimbals will follow once a basic demonstration has occurred. Early integration of the camera system and gimbal will allow for mechanical optimisation and tuning for better stabilisation and headtracking performance. This integration process will occur in parallel to development for LDR output compression.

## Resource Plan

Hardware purchases will begin at the start of SP5, with development of mount brackets at the same time. The ODroid XU3 ($179 USD), 3 webcameras ($17 USD) and related peripherals (wifi, eMMC flash storage, power cable) cost approximately $283USD. All components are available from the same supplier to simplify the supply chain. It is estimated that the parts will take 2 weeks from purchase to arrive, with a goal purchase approval date of July 24th.

The frame mount prototypes will be printed with the Universities Zortrax M200 printers. It is likely that these printers will be heavily used during this process, so as a backup the parts can be printed on my personal printer with no backlog of jobs.

There are no additional purchasing or equipment requirements, as all other elements of the project are software based. If time permits, a finalised enclosure and mount can be constructed in the later weeks of September with a combination of CNC aluminium and routed carbon fibre sheet for a rugged and professional finish.

## Technical Risk

The proposed system is sufficiently complex as to present several areas of risk including:

1. Processing overheads creating a resolution ceiling (either spacial or temporal restrictions),
2. Unavoidable system latency; including time taken to sample image data from cameras via USB on a bus constrained platform, and post processing delays,
3. Compressed image data (MJPEG) from cameras may incur additional delays due to conversion to useable format,
4. Limited expanded dynamic range due to traditionally poor capabilities of attainable CMOS based cameras,
5. Parallax, distortion and frame positioning errors with the lens array due to lack of common optical elements.
6. Size and/or weight of the completed system exceed the capabilities of the actuated mount

To reduce the impact of likely risks (1 through 3) which may cause significant processing delays, early stage development will be performed with lower resolutions as to allow sufficient overhead prior to any optimisation or feature expansion. Additionally, bus limitations are unlikely to be met with smaller arrays of commodity cameras. This limitation is most likely to reduce modularity and additions to the camera array rather than affect the proposed system performance.

Issues with artifacts and optical errors caused by camera alignment are almost certain, and represent a major hurdle for the project as the alignment and parallax issues will cause viewing issues. The camera array will be rigidly fixed in a known state to allow for calibrated compensation for pose differences.

Lens distortion is also likely to be present and is a minor issue unless the distortion differs greatly between cameras. If such differences are present between the lenses, it may induce additional artifacts during the blending stages. These issues should be minimised by using cameras of the same make and model, and software distortion compensation could be considered if the issue is severe.

Packaging size and mass are unlikely to exceed the stated dimensions in the requirements, as the hardware consists of the processor and mounted webcams. The total mass of electrical components is under 130g, and the enclosure will be developed with lightweight materials such as printed ABS or laser cut acrylic.

# Progress Report

Some video implementations like the Eyetap project [[7](#_ENREF_7)] require online, real-time processing that is performed on an FPGA to overcome bus and processing requirements. The proposed platform is required to operate in a similar manner to allow for head-tracked movements and fluent teleoperation of robotics, however FPGA development with USB based cameras would require a high level of technical skill and long development cycle.

With the recent improvements in SoC video capabilities, embedded platforms have access to a reasonably powerful OpenGL capable video card. As the GPU is designed for repetitive pixel based operations, shaders allow the developer access to the benefits of GPU. As a suitable alternative to FPGAs for image processing, GPU accelerated approaches also offer substantial improvement over CPU based processing. The evaluation of ARM based computing platforms and integrated GPU solutions in therefore important for the success of the project.

Current project progress has focussed on understanding and refining the design requirements, testing available hardware for required OpenGL call support, and selecting hardware to use for the test platform.

## Pipeline Capabilities & Restrictions

The GPU processing will be performed using the standard OpenGL feature set, and uses a combination of Vertex shaders and fragment shaders to handle all image manipulation and blending steps as shown in the proposed pipeline [Fig 2]. Additional computer vision may be implemented for exposure feedback control

Vertex shaders will be used to perform rough frame and perspective corrections, as operations such as scaling, transforms and warping can be performed with minimal overhead. The raw image data from each camera will be manipulated in Vertex shaders, followed with blending and filtering with Fragment Shaders.

Fragment shaders are pixel level programs used in a wide range of versatile applications including filtering, colour correction, procedural content generation and rendering pipelines. Fragment shaders have a few limitations that need consideration including the following:

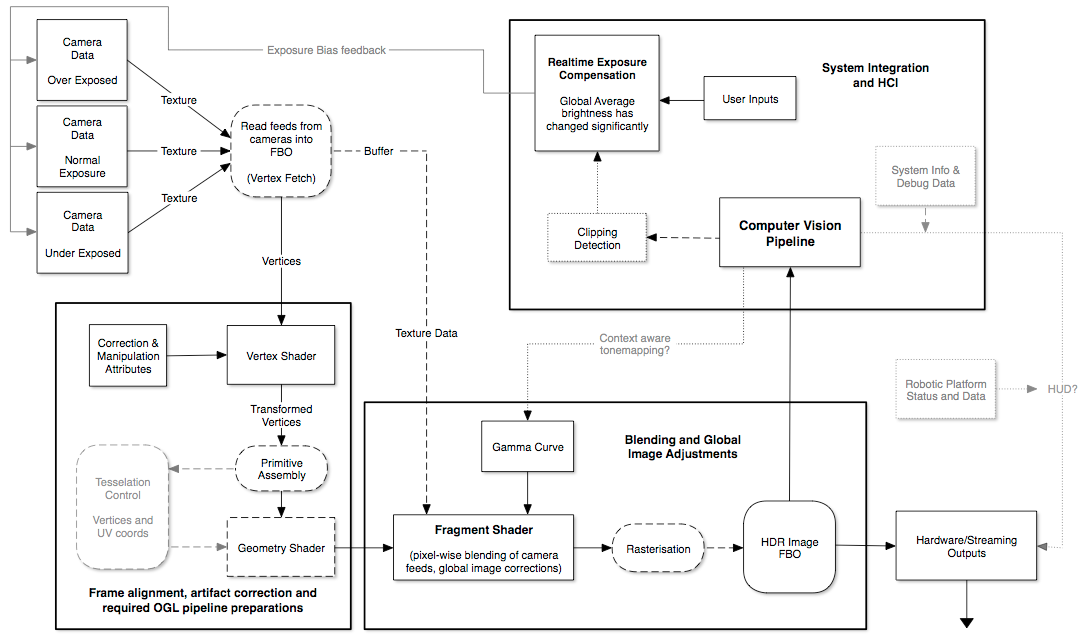


Fig 2: Graphics Processing Pipeline.

* Code runs per pixel with no access to any other pixel's data. This makes operation such as blurring or neighbour based filtering impractical.
* There are no libraries or helper tools to assist with maths or lookups.
* Are not capable of reading from memory, and must read textures from Frame Buffer Objects (FBO) stored prior on the GPU.
* Each instance has no time information or history complicating time domain operations requiring access to old pixel data (such as motion blur). Workarounds exist but the effectiveness of this technique is limited.

Despite these limitations however, fragment shaders are perfect for pixel based blending. Initial tests used thresholded pixel replacement to reduce blown highlight areas and clipped shadows, however a more intelligent approach will need to be implemented to allow for smooth transitions between frame data. I believe creating a camera response curve and blending gradually instead of a hard threshold likely achieve this.

By moving the bulk of calculations and corrections to the GPU with the OpenGL Shading Language (GLSL), the process is easily parallelised and will minimise CPU load to allow overhead for other tasks. As GPU capabilities improve in the future, the platform will easily scale to allow for higher resolutions and frame rates. This is one of the main attractions of the approach, and also features massive performance increases and scalability over frameworks like OpenCV.

## Software Tests and Benchmarks

To assess the graphics performance of the ODroid XU3 single board computers currently used by the MechBot program at the University of South Australia, some simple test programs were written to compare basic GPU functionality against other computers.

These test programs were designed to test the specific operations and calls that are most likely to limit performance, focussing on texture sampling and manipulation, draw call performance, and simple maths performance (implemented as a blend function). This configuration roughly represents the final product with respect to setup and texture read overheads. Average framerate during tests is plotted in [Fig 3] and shows the huge weaknesses of the Mali T628 GPU against desktop class cards when performing frequent texture reads and multiple draw calls. Texture reads are less efficient than expected, but this performance difference is less of an issue when the power consumption, size and cost are taken into consideration. I strongly feel that the T628 is capable of the required online processing, however the tradeoff between framerate and resolution will need to be considered if optimisation is insufficient. More powerful boards are discussed in the Hardware Selection topic on pg 7.

Fig 3: Texture Sample Performance. Tests ran for 10 seconds with unbound FPS with square textures. At a resolution of 512x512px, the XU3 achieved 18fps.

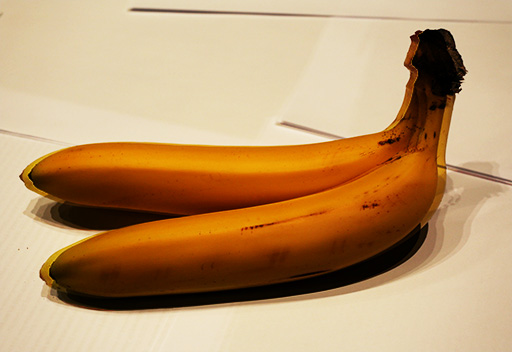


Fig 4: Two textures overlayed with no corrections. Note the semi-transparent artifacts on edges that show horizontal placement difference between the source frames

To evaluate the effects of overlayed images without correction or a common optical path, a test program was developed running on a desktop grade GPU. This test [Fig 4] shows the effect of uncorrected parallax in close field situations. As the object's distance to the lenses increases, this effect becomes less noticeable. Additional testing with different camera patterns is needed to determine how severe the issue is without correction, and if more invasive warping operations need to be considered.

## Hardware Research & Selection

The ODroid webcam was selected as most adequate for the application due to incredibly low cost ($20), small footprint and low mass, and well controlled optical distortion. These cameras will be removed from their standard housing for a further reduction in footprint, and mounted in a custom frame from which the camera's will be calibrated.

The ARM based ODroid XU3 from Hardkernel.com was selected as the development platform due to familiarity and pre-existing integration with the University of South Australia MechBot platform. The board features a powerful 4+4 core CPU, 2GB of RAM, flash storage and the Mali T628 GPU found on recent Android phones.

The XU3 has multiple USB2.0 ports and includes USB3 support. Although untested personally, it should be capable of handling the demanding IO requirements of multiple web cameras and has been demonstrated with 4 USB cameras in IP based security applications on the Hardkernel.com blog.

The board will run Lubuntu 14.04 as it requires fewer system resources than a standard Ubuntu install whilst allowing for near full software compatibility. The main application software will be written in C++, with the bulk of the graphics pipeline using native OpenGL functionality and the OpenGL Shading Language which is similar to C.

Once a test system is fully operational and optimised, the platform could easily be moved to one of the upcoming NVidia Jetson based development boards such as the TK1, as it quotes a significant graphics improvement. The TK1 is capable of running a near identical software stack to the XU3 boards however acquiring the hardware is still difficult due to supply chain problems.

# Conclusion

With a more concrete understanding of the project's direction and requirements, the hardware and software can be finalised for the implementation of the project. Preliminary research and testing has demonstrated basic feasibility of the project, where the compromise between visual artifacts and hardware complexity has been assessed pending further investigations.

SP5 will see the full implementation of the proposed GPU accelerated HDR system, and ideally show integration with other concurrent projects. The project will proceed with the use of the ODroid XU3 and ODroid low cost webcams, where familiarity with the platform and simplified integration with the University of South Australia MechBot and MERC’s Freddy the Fire Truck is beneficial to the adoption of the project.

# References

[1] T. Tomić, K. Schmid, P. Lutz, A. Dömel, M. Kassecker, E. Mair*, et al.*, "Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue," *Robotics & Automation Magazine, IEEE,* vol. 19, pp. 46-56, 2012.

[2] J. Casper and R. R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 33, pp. 367-385, 2003.

[3] B. C. Madden, "Extended intensity range imaging," 1993.

[4] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Computer Vision, 1993. Proceedings., Fourth International Conference on*, 1993, pp. 173-182.

[5] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "High dynamic range image and video compression-fidelity matching human visual performance," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 2007, pp. I-9-I-12.

[6] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Transactions on Graphics (TOG),* vol. 22, pp. 319-325, 2003.

[7] S. Mann, R. C. H. Lo, K. Ovtcharov, S. Gu, D.-Q. Dai, C. Ngan*, et al.*, "Realtime hdr (high dynamic range) video for eyetap wearable computers, fpga-based seeing aids, and glasseyes (eyetaps)," in *Electrical & Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on*, 2012, pp. 1-6.

[8] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," in *ACM Transactions on Graphics (TOG)*, 2011, p. 41.

[9] J. Stumpfel, C. Tchou, A. Jones, T. Hawkins, A. Wenger, and P. Debevec, "Direct HDR capture of the sun and sky," in *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, 2004, pp. 145-149.

[10] H. Wang, R. Raskar, and N. Ahuja, "High dynamic range video using split aperture camera," in *IEEE 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Washington, DC, USA*, 2005.

[11] J. Döge, G. Schönfelder, G. T. Streil, and A. König, "An HDR CMOS image sensor with spiking pixels, pixel-level ADC, and linear characteristics," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on,* vol. 49, pp. 155-158, 2002.

[12] J. Fung and S. Mann, "Using graphics devices in reverse: GPU-based image processing and computer vision," in *Multimedia and Expo, 2008 IEEE International Conference on*, 2008, pp. 9-12.

[13] Y. Xiong and K. Pulli, "Fast panorama stitching for high-quality panoramic images on mobile phones," *Consumer Electronics, IEEE Transactions on,* vol. 56, pp. 298-306, 2010.

[14] M. B. López, H. Nykänen, J. Hannuksela, O. Silvén, and M. Vehviläinen, "Accelerating image recognition on mobile devices using GPGPU," in *IS&T/SPIE Electronic Imaging*, 2011, pp. 78720R-78720R-10.

[15] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *ACM Transactions on Graphics (TOG)*, 2002, pp. 249-256.

[16] H. S. Sawhney and R. Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 21, pp. 235-243, 1999.

[17] K. Smith, G. Krawczyk, K. Myszkowski, and H. P. Seidel, "Beyond tone mapping: Enhanced depiction of tone mapped HDR images," in *Computer Graphics Forum*, 2006, pp. 427-438.

[18] J. M. DiCarlo and B. A. Wandell, "Rendering high dynamic range images," in *Electronic Imaging*, 2000, pp. 392-401.

[19] A. O. Akyüz, "High dynamic range imaging pipeline on the gpu," *Journal of Real-Time Image Processing,* vol. 10, pp. 273-287, 2012.