
Novel Image Processing algorithms and methods for improving their robustness and operational performance

By

Ilya V. Romanenko

A Doctoral Thesis

Submitted in partial fulfilment of the requirements
for the award of

Doctor of Philosophy

of

Loughborough University

July 2014

© Ilya Romanenko 2014

Supervisor: Professor Eran Edirisinghe

Acknowledgment

I would like to express my special appreciation and thanks to my advisor Professor Dr. Eran Edirisinghe, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on the research has been priceless. Your guidance helped me to choose the right direction and progress with my research.

I would especially like to thank my colleagues at Apical LTD who helped me to run the experiments and process the results. All of you have been there to support me when I collected data for my Ph.D. thesis. With your support the research ideas became working devices. That help was priceless as allowed us all to see the value of my research and helped to prove the ideas. I would like also to say thank you to Michael Tusch for the opportunity given me at Apical LTD to work on my PhD.

I would like to express special thanks to my alma mater university Moscow Institute of Physics and Technologies for giving me the knowledge and the courage to become a researcher.

A special thanks to my family. Words cannot express how grateful I am to my parents for the warm support and continuous encouragements that never stopped from their side. And most of all, I say thank you to my loving, supportive, encouraging and patient wife Inessa, whose faithful support during the years of this PhD is so appreciated.

Abstract

Image processing algorithms have developed rapidly in recent years. Imaging functions are becoming more common in electronic devices, demanding better image quality, and more robust image capture in challenging conditions. Increasingly more complicated algorithms are being developed in order to achieve better signal to noise characteristics, more accurate colours, and wider dynamic range, in order to approach the human visual system performance levels.

The research presented in this thesis proposes a novel and efficient approach to improve the performance of image processing algorithms by modelling the image sensor characteristics. The proposed approaches allow not only achieve better operational performance but also a number of algorithmic optimizations, making their practical use feasible.

The fundamental aim of the research presented in this thesis is to review the traditional image processing algorithms and to find ways to use the information about image sensor characteristics efficiently in them by re-arranging the image processing pipeline and re-designing the algorithms. The re-design of the image processing pipeline requires the re-design of the main processing blocks. The results of the proposed research allow newly designed functional blocks to work reliably and improve their performance to levels, where their use becomes practical.

The results of the research presented in this thesis cover a number of important image processing areas.

Proposed spatial and spatial-temporal noise reduction techniques allowed achieving the performance on the level and above of the best known noise reduction algorithms. Due to a number of algorithmic optimizations and a novel approach of applying algorithms in the Bayer RAW domain, using sensor noise modelling, the proposed algorithms were efficiently implemented in hardware and used in a number of commercial products. Other algorithms of comparable performance are not known to be used commercially.

The proposed frame accumulation algorithm for de-noising of still images is shown to perform to a high standard. It is based on previously developed technique and was implemented in hardware. The ability of the proposed frame accumulation algorithm to compensate for large objects offsets and efficiently accumulate still images is unique and enables improved camera

performance in low light conditions. Other known techniques are not used in commercial products due to complexity and, more importantly, poor image quality in various conditions.

The proposed multi-exposure image fusion algorithm is based on the frame accumulation algorithm and allows multi-exposure fusion free from ghosting artefacts. The algorithm for multi-exposure image fusion enables the wide dynamic capture in standard cameras. The algorithm is implemented in hardware and allows wide dynamic range fusion in real time.

The proposed edge detection normalisation technique improves object detection reliability. The algorithm performs edge detection on sensor data directly, thus allowing object detection to be implemented on camera without the image processing pipeline. The object detection system, using the proposed approach, is implemented in hardware and demonstrates improved detection performance compared to traditional object detection system.

Ilya Romanenko, May 2014

Keywords: De-noising, Image Processing, Sensor Noise, Bayer RAW, Non-Local Means, FPGA, Real Time, Spatial-Temporal Noise Reduction, Optical Flow, Motion Compensation, HDR, Image Data Matching, WDR, Image Features Matching, Feature Vector Extraction, Object Detection.

Table of Contents

Acknowledgment	1
Abstract	2
List of Figures	7
List of Tables	9
Abbreviations	10
Chapter 1. Introduction	12
1.1. Research Problem statement	12
1.2. Aim and Objectives	13
1.3. Contributions	15
1.4. Organisation of Thesis	17
1.4.1. Part 1: Introduction, background theory and known methods	17
1.4.2. Part 2: Novel image processing algorithms and methods to solve difficult known problems in image processing.	17
Chapter 2. Background Theory and Related Work	19
2.1. Overview of image sensors and their characteristics	19
2.2. Data sampling	19
2.3. Image processing pipelines	19
2.4. Noise characteristics	21
2.5. Interpolation and statistical data accumulation	24
2.6. Muti-scale data segmentation	27
2.7. Temporal methods: image data accumulation using the Gaussian background model.	30
Chapter 3. Block matching de-noising method for photographic images, applied in Bayer RAW domain, optimized for real-time implementation	32
3.1. Introduction	32

Table of Contents

3.2. Block matching approach for Bayer RGB sensors.....	36
3.3. Experimental results.....	40
3.3.1. Simulated test.....	40
3.3.2. Real world test	45
3.4. Conclusion.....	48
Chapter 4. A Spatio-Temporal noise reduction method optimized for real-time implementation	50
4.1. Introduction	50
4.2. Block matching approach for Bayer RGB sensors.....	54
4.3. Temporal data accumulation using Gaussian background model.....	57
4.4. Experimental results.....	58
4.5. Conclusion.....	64
Chapter 5. Image Matching in Bayer RAW Domain to De-noise Low-light Still Images, Optimized for Real-Time Implementation.....	65
5.1. Introduction	65
5.2. Robust optical flow	68
5.3. Block matching approach for Bayer RGB sensors.....	72
5.4. Experimental results.....	74
5.5. Conclusion.....	80
Chapter 6. Image Matching in Bayer RAW Domain to Remove Ghosting in Multi-Exposure Image Fusion.....	81
6.1. Introduction	81
6.2. Proposed Image Fusion Method.....	82
6.2.1. Intensity Matching	83
6.2.2. Coarse and fine motion estimation and compensation	84
6.2.3. Image blending.....	84
6.2.4. Dynamic range compression.....	85

Table of Contents

6.3. Experimental results.....	86
6.4. Conclusion.....	92
Chapter 7. The use of sensor noise modelling in the segmentation and detection of objects.....	93
7.1. Introduction.....	93
7.2. A feature extraction model, utilizing histogram of oriented gradients.	94
7.3. Proposed feature normalization method.....	95
7.4. Experimental results.....	96
7.5. Conclusion.....	100
Chapter 8. Conclusions and Future Work.....	101
8.1. Conclusions.....	101
8.2. Future work.....	103
References.....	105
Appendix A List of Publications.....	110
Appendix B Sensors used in experiments.....	111

List of Figures

Figure 1: Traditional image processing pipeline.	20
Figure 2: Proposed Image processing pipeline organization.	21
Figure 3 Sensor noise at ISO100, ISO400 and ISO800.....	23
Figure 4: Sensor noise experimental data	23
Figure 5: Intra-frame accumulation	26
Figure 6: Inter-frame image data matching	27
Figure 7: Algorithm block diagram.	35
Figure 8: Block diagram of the simulated test process	40
Figure 9: Block diagram of the simulated test procedures for BM3D and Adobe Lightroom.	41
Figure 10: Kodak image (4) close-up.	43
Figure 11: Kodak image (23) close-up.	43
Figure 12: Images taken by Sony Nex-5 camera at ISO12800 and ISO200.	45
Figure 13: Block diagram of the real-world test procedure.....	46
Figure 14: Real world test results.	47
Figure 15: Algorithm block diagram.	52
Figure 16: Spatial-Temporal filter block diagram	53
Figure 17: Experimental results	58
Figure 18: Experimental results	59
Figure 19: The effect of motion compensation.....	60
Figure 20: Experimental results, Motion Compensation evaluation.	61
Figure 21: Ground truth images.....	62
Figure 22: Algorithm block diagram.	67
Figure 23: Image scale pyramid.....	71
Figure 24: Multi-scale optical flow calculation	71
Figure 25: Motion field.....	74
Figure 26: Experimental results, moving background.....	75
Figure 27: Experimental results, moving foreground.....	75
Figure 28: Experimental results, indoors scene	76
Figure 29: Experimental results, indoors scene	76
Figure 30: Experimental results, lab scene	77
Figure 31: Experimental results, lab scene low light.....	77

Table of Contents

Figure 32: Experimental results, lab scene low light.....	78
Figure 33: Experimental results, indoors low light.....	78
Figure 34: The block diagram of the proposed multi-exposure image fusion algorithm	82
Figure 35: The results of fusion.....	86
Figure 36: Example of motion and calculated motion field.	87
Figure 37: PSNR values, calculated on images with DRC applied.	88
Figure 38: PSNR values, calculated on images with DRC applied.	89
Figure 39: HDR image obtained as the result of the proposed HDR method.....	90
Figure 40: Edge segmentation functions.....	94
Figure 41: Object detection experimental results.	97
Figure 42: An example of noise model normalized edge segmentation.....	99

List of Tables

Table 1: PSNR values on a sub-set of Kodak images.....	42
Table 2: Spatial noise reduction synthesis results.....	48
Table 3: PSNR values comparison table.....	63
Table 4: Spatio-Temporal noise reduction block implementation details	64
Table 5: PSNR values comparison table.....	79
Table 6: Synthesis details for the image matching block	80
Table 7: PSNR values, comparison table.....	91
Table 8: Synthesis results for proposed pixel mapping block.	92
Table 9: Detection rates statistical data.	98
Table 10: Object detection system resource utilization	100
Table 11: Image sensors used in experiments.....	111

Abbreviations

ASIC	Application specific integrated circuit
BM3D	Block matching 3 dimensional
BPF	Band pass filter
CMOS	Complementary metal–oxide–semiconductor, technology commonly used for digital image sensors manufacturing
dB	Decibel
DCT	Discrete cosine transform
DRC	Dynamic range compression
FB	Frame Buffer
FPGA	Field programmable gate array
FPN	Fixed pattern noise
GM	Gaussian Mixture
HD	High definition
HDR	High dynamic range
HOG	Histogram of oriented gradients
HPF	High pass filter
HVS	Human vision system
ISO	International Standards Organisation, in photography refers to the norm for the sensitivity, originally used for emulsion based films, later a similar measure was used for digital image sensors
ISP	Image signal processing
LBP	Local binary patterns
LPF	Low pass filter
MC	Motion compensation
ME	Motion estimation

Abbreviations

MV	Motion vector
NLBM	Non Local Block Matching
NR	Noise Reduction
PCA	Principal component analysis
PSNR	Peak signal to noise ratio
RAW	Linear data obtained from the sensor directly
SAD	Sum of absolute differences
sRGB	Gamma corrected RGB image data
SVM	Support Vector Machine
VBM3D	Video BM3D
WDR	Wide dynamic range

Chapter 1

Introduction

1.1. Research Problem statement

The technology of manufacturing image sensors has been rapidly evolving in the recent years, making image sensors available for mobile device use, reducing power consumption, increasing image capturing resolution and frame rate. Although a significant progress has been achieved in many areas of electronic imaging, digital imaging systems are still significantly inferior, when compared with the human visual system. The human visual system outperforms digital imaging systems in many areas, such as: the dynamic range of a captured scene, amount of captured details, sensitivity, limitations due to the presence of noise, as the image sensors concerned, and finally the capture rate at a nominal resolution. In our research we will attempt to resolve some of the outstanding operational and performance issues of digital imaging systems. The proposed approach will study the image sensors characteristics and behaviour in different situations, and use the modelled sensor behaviour to improve the image processing algorithms, by making them more robust yet feasible for practical use in digital image processing.

There are known algorithms for images and video de-noising, such as BM3D and VBM3D described in [3], [4]. The performance of BM3D and VBM3D algorithms is one of the best among known algorithms, however due to complexity the algorithm is not used in camera systems as its practical implementation is not possible not in software nor in hardware. There are other algorithms [6],[11],[13] based on non-local means block matching, block accumulation techniques, or PCA [10]. However their performance is inferior comparing to the BM3D and VBM3D.

The algorithms for frame accumulation and video de-noising were proposed in [9],[24],[25], however it was discovered that large object displacements is difficult to compensate. The proposed algorithms were implemented in RGB domain, and were unable to use the information about sensor noise. The quality of frame accumulation is fully depends on the quality of motion estimation and compensation, which is known to be a very difficult problem,

therefore the method of frame accumulation is not very common in practice. Attempts to minimize the effect of motion estimation imprecision were taken in [25], however the requirements for memory bandwidth and algorithm complexity made this algorithm implementation not practical.

The problem of multi-exposure images fusion is very difficult and related to the problem of frame accumulation. Another complication in multi-exposure fusion is that images are taken at different exposures. Due to possible large object displacements the appearance of ghosting artefact is very common and unacceptable in consumer applications. The problem of ghost-free multi-exposure fusion was not solved yet. The attempts to resolve the ghosting artefacts appearance in multi-exposure image fusion were taken in [45],[46],[52]. Currently there are no known methods allowing performing multi-exposure image fusion without ghosting artefact suitable for practical use.

The area of object detection is developing rapidly, however existing approaches assuming the object detection algorithms to run on recorded video or still images. One of the most reliable object detection techniques is known as HOG-SVM and described in [40],[41],[42],[43],[44]. There are a number of issues with the mainstream approach to object detection. Firstly object detection always need image processing system to produce quality RGB image or video sequence, which in many cases means increased system complexity. Secondly the object detection algorithms assume no knowledge about image source, as the image processing settings are not known. The performance of object detection algorithms deteriorates quickly in low light conditions. The possibility of running object detection algorithms on sensor data directly and use the sensor characterization to improve object detection quality is not investigated. The possibility of implementation of object detection system on the sensor silicon was investigated in [39], however the sensor can offer very limited resources for object detection algorithm, so the idea of object detection on sensor is not very practical.

1.2. Aim and Objectives

The aim of this thesis is to propose the re-organisation of a standard image processing pipeline as well as to propose novel approaches to traditional image processing algorithms, where data related to ground truth is used to operate algorithms in a more reliable fashion by modelling the sensor characteristics. The increased algorithmic performance also enables the reduction of the complexity of the algorithms and makes their practical use in commercial and industrial

devices feasible. The focus is to develop algorithms allowing efficient implementation in hardware e.g. FPGA, ASIC devices, as hardware implementation is the preferred method of implementation that guarantees no compromise between algorithm quality and performance. The ultimate goal is to increase the quality and performance of the algorithms to a level at which their practical use will not be a concern, such as in multi-exposure, wide dynamic range, image stitching, image data accumulation.

The following research objectives, if met, can assure the achievement of the aforementioned goals:

- Study the existing approaches and identify their weaknesses.
- Investigate possible solutions by assuming that the image sensor can be used as a calibrated measurement instrument.
- Concentrate on algorithms' design suitable for efficient hardware implementation to eliminate the compromise between algorithmic quality, performance and power consumption.
- Concentrate on algorithms' hardware implementation to be within 1-1.5 million gates, when implemented in ASIC, consider algorithmic optimizations first as most efficient. Satisfying this requirement make algorithms implementation practical and suitable for commercial use.
- Actively re-design the image processing pipeline to access image data at a point where they are not affected by non-linear algorithms and can be accurately and directly related to the model of the image sensor.
- Develop a spatial noise reduction algorithm, allowing high level of optimization for hardware implementation, at the same time providing the details preservation and the efficiency of noise filtering on the level or above of the best known algorithms.
- Develop a spatial-temporal noise reduction algorithm, featuring local motion compensation and efficient data accumulation. The developed algorithm should allow high level of optimization for hardware implementation and minimization of memory bandwidth, at the same time providing the details preservation and the efficiency of noise filtering on the level or above of the best known algorithms.
- Develop a frame accumulation algorithm able to compensate large object offsets to allow photographic images accumulation to improve signal to noise ratios in a low light conditions. The developed algorithm should allow high level of optimization for

hardware implementation and minimization of memory bandwidth, at the same time providing the details preservation and the efficiency of noise filtering.

- Develop an algorithm for fusion of images taken at different exposures to achieve wide dynamic range capture and eliminate appearance of ghosting artefact. The developed algorithm should allow high level of optimization for hardware implementation.
- Re-design edge detection algorithm used in object detection system to improve detection rate and reduce false positives rate by using image processing techniques developed in this work.
- Allow re-designed object detection system to work with sensor data directly, thus eliminating the need for image processing sub-system in embedded object detection systems.

1.3. Contributions

A number of original contributions have resulted from the work conducted within the research context of this thesis.

- A spatial noise reduction algorithm was proposed to operate on Bayer RAW data obtained from the image sensor. An image scale pyramid and a block matching approach were used directly on the image sensor data. The sensor noise model was used to weigh the decisions made during the block matching process. A non-linear SAD filter was proposed to separate data auto-correlation from noise. The proposed algorithm implemented as a hardware block, achieved significant improvements in spatial noise reduction and has already being used in a number of commercial devices.
- The novel spatial noise reduction block has been used within a temporal noise reduction algorithm in order to match image data between different frames and perform precise pixel mapping. The original idea was to perform block matching in Bayer RAW data space. The sensor noise model was used to define the reference for the block matching algorithm, which enabled accurate local motion compensation. Another contribution is to use a Gaussian background model to achieve optimal data accumulation and suppress errors in temporal data matching. Performing data accumulation in Bayer RAW data domain enabled the use of the sensor noise model as a reference data variance in the Gaussian background model.

- The above Spatio-Temporal noise reduction algorithm was further extended by using a robust optical flow based approach (working in Bayer RAW data space), which enabled the improvement of the motion field estimation at relatively low computational cost. Proposed Spatial-Temporal noise reduction algorithm is able to compensate motion on a very large scale, as well as accurately match image data between different frames at a pixel level. A specific contribution made was to use the sensor noise model as a reference for any pattern matching process, making the decision more reliable, and the algorithm less complex.
- Subsequently the Spatial-Temporal noise reduction algorithm and the robust optical flow algorithm above are used in multi-exposure frame fusion. The novelty of this work is to reformulate the problem of multi-exposure image fusion into the problem of spatial-temporal image data matching, using the noise reduction framework. The result of this work is an algorithm, which allowed performing multi-exposure image data fusion, eliminating any local motion artefacts and producing wide dynamic range images, matching the human vision capabilities.
- Finally the methods of data processing developed in previous research are used to improve feature extraction quality in object detection algorithms. The specific contribution made by this research is to investigate and develop a method for the normalisation of edge detector responses, based on the sensor noise model. It will be shown that the results of a typical object detection task can be significantly improved based on this improvement.

1.4. Organisation of Thesis

This thesis is organized in two parts. The first part includes non-contributory chapters providing basic information about the research problem addressed in the thesis and fundamental background knowledge of the subject area. The second part of thesis includes five contributory chapters, each dedicated to a particular problem in image processing area and a novel approach to addressing that problem. A summary of each part/chapter can be represented as follows:

1.4.1. Part 1: Introduction, background theory and known methods

Chapter 1 provides an overview of the thesis, defines the research problem, states the research motivation and specifies the thesis aims and objectives. Finally it outlines the organisation of the thesis.

Chapter 2 provides an overview of an image processing system and fundamental elements of a typical image processing pipeline. This chapter also presents the known approaches to image processing, spatial and temporal data accumulation in particular, which we will employ in our proposed algorithms in an unusual way.

1.4.2. Part 2: Novel image processing algorithms and methods to solve difficult known problems in image processing.

Chapter 3 proposes a novel block matching noise reduction method, applied in Bayer RAW data space, algorithmically optimized for efficient hardware implementation. This chapter also includes a literature review and presents state of the art algorithms in the image noise reduction area, as well as an explanation of contributions and experimental results.

Chapter 4 provides details of the proposed Spatio-Temporal noise reduction method, applied in Bayer RAW data space. It is shown that the proposed algorithm is optimized for hardware implementation. The method of image matching in spatial domain is extended to perform a data accumulation in temporal domain. Sensor noise characteristics are used to normalize spatial and temporal variations. A literature review of state-of-the art algorithms in image noise reduction area, as well as an explanation of the original contributions and experimental results is also provided in this chapter.

Chapter 5 provides a description of a novel algorithm for image data accumulation based on optical flow, which is designed to de-noise high resolution photographic images. In this

research the robust optical flow algorithm was improved by using the information about the image sensor noise characteristics. Temporal data accumulation method described in Chapter 4 was adopted in this proposed algorithm. This chapter also includes a literature review and presents state of the art algorithms in image noise reduction, as well as explanation of the contributions made by the proposed algorithm.

Chapter 6 provides details of the application of the algorithm for Image Matching in Bayer RAW domain for ghosting removal in multi-exposure image fusion. In this research, previously developed reliable methods for image matching and data accumulation were used to match images taken at different exposures. A literature review, a presentation of state of the art algorithms in image noise reduction and an explanation of the contributions made by the proposed research is presented.

Chapter 7 presents details on research conducted in Sensor Noise modelling, which will be used in edge detection, to improve the performance of object detection algorithms. In this research the impact on object detection and false positive rate imposed by the sensor noise modelling has been investigated in the edge detection part of an object detection algorithm. A literature review and presentation state-of-the-art algorithms in the image noise reduction area, as well as explanation of the contribution made by the research presented in this thesis and relevant experimental results are also provided in this chapter.

Chapter 2

Background Theory and Related Work

2.1. Overview of image sensors and their characteristics

Image sensors perform a transformation of optical information into electrical signals. In order to acquire the image data, image sensor area is divided into a large number of individual photo-sites - pixels. Thus each pixel characteristics can be described as a single photo diode, whereas the whole device can be considered as a spatial grid of photo-diodes, performing spatial sampling of optical information. Such characteristics as spectral sensitivities and quantum efficiency are defined by the photo-diode, while noise generated by the image sensor is given by various noise sources: quantisation of photons count, random thermal processes in image sensor material, pixel data multiplexor circuits, analogue amplification circuits, analogue to digital converter quantization noise. It is important to note that various noise sources can be separated by the effect they produce on a final digital image.

2.2. Data sampling

A spatial grid of photo-diodes can be considered as a two-dimensional sampling array. We can consider that sampling theory is applicable to image sensors. Image sensors have common problems with optical crosstalk between adjacent pixels, electrical crosstalk between pixels. There is also a problem with spatial frequency aliasing, which is a result of spatial sampling without filtering in spatial domain in order to limit the bandwidth of signal. In order to capture colour information, the Bayer RGGB pattern is commonly used. In the proposed research different sensor array patterns such as RGBW, RGBIR or more exotic random pattern sensors will not be considered, as it would not affect our research, the results of which can be generalized for any alternative image data sampling methods.

2.3. Image processing pipelines

Image data, captured by the sensor is usually processed by a number of functional units, arranged in a chain of sequential processing blocks, named in literature as an Image Processing

Pipeline (IPP). Each stage of the processing is performed by its corresponding block. An example of a traditional IPP is presented in Figure 1:

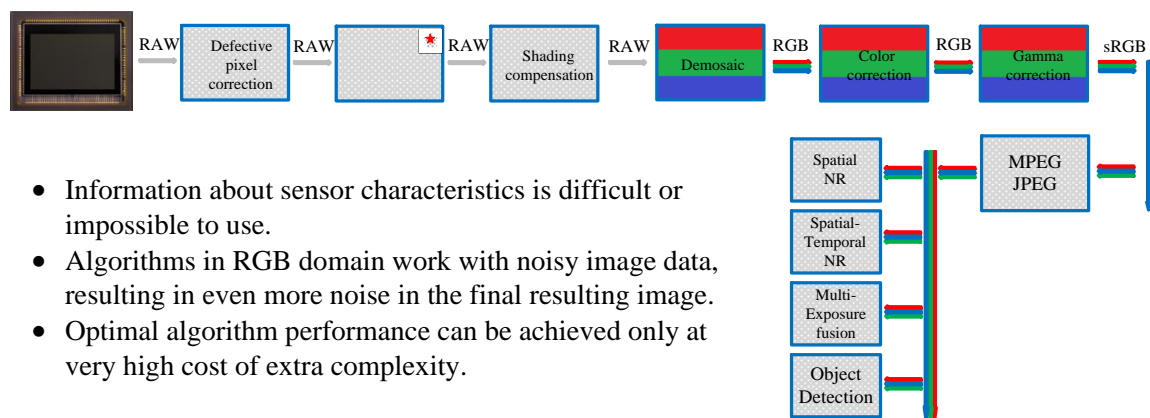


Figure 1: Traditional image processing pipeline.

In the above pipeline it is seen that some stages of processing are performed in the Bayer RAW data space, while some other processing is performed on RGB image data. It is important that starting from the de-mosaic block, processing of the data is performed by non-linear algorithms, making image intensity levels non-linearly distributed, thus breaking linear dependencies between different regions in the image. In this research an attempt is made to design image processing blocks, working in linear Bayer RAW data space, in order to benefit from predictable nature of data, enabling to perform effective sensor noise modelling. The estimation of the noise characteristics for each image region can drastically improve the reliability of most image processing algorithms, by providing a very reliable reference for any decision made by the algorithm's logic. However processing in the Bayer RAW data space will impose additional constraints and create some difficulties in algorithms design. The research conducted in this thesis will attempt to overcome these issues and propose reliable, robust but yet feasible solutions for algorithms that are practically implementable. The block scheme of proposed organization of IPP is presented in Figure 2:

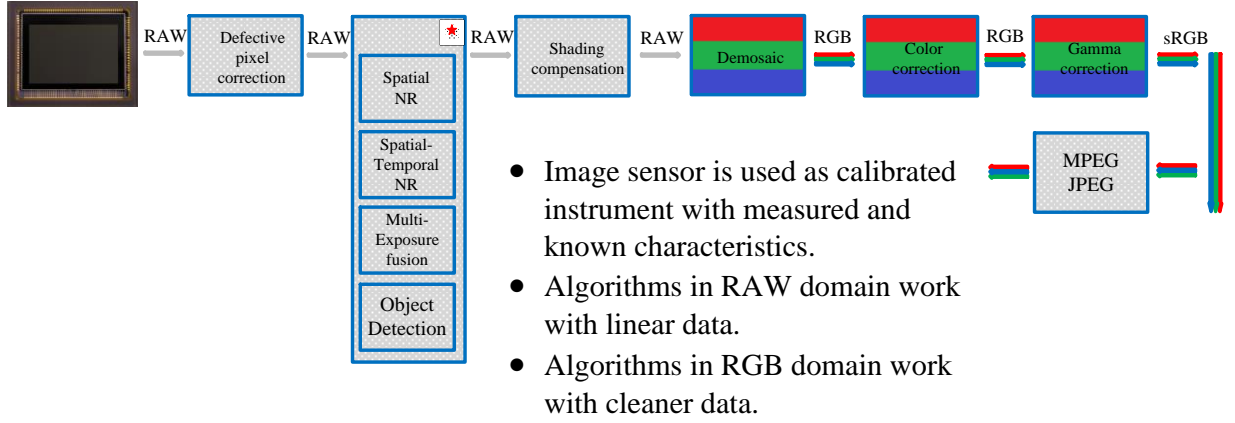


Figure 2: Proposed Image processing pipeline organization.

2.4. Noise characteristics

In the proposed research we will consider the effect of noise, added to the image. It has been previously investigated by other researches [12],[13],[16],[17] that the additive noise model is generally applicable for describing noise of an image sensor. It has been also proven that the actual sensor noise fits the Gaussian and Poissonian random processes model very well. The image data representing the actual scene image sampled by the sensor without noise added to the image is defined as $I_p(x, y, t)$. The ideal image data $I_p(x, y, t)$ is a function of coordinates x, y and t . In this research two dimensional coordinates x, y denoted as v for the compactness of equations, therefore the ideal image data to be defined as $I_p(v, t)$. Noise of different nature is assume: analogue noise $n_a(v, t)$, originating from analogue circuits and added to the image data, fixed pattern noise (FPN) $n_{fpn}(v)$, originating from multiplexors and sensor defects therefore not being a function of time, and photon noise $n_q(I_p(v, t))$, also known as a shot noise that is added to the image data $I_p(v, t)$, captured at time t and is sampled by the sensor as follows:

$$I_s(v, t) = I_p(v, t) + n_a(v, t) + n_{fpn}(v) + n_q(I_p(v, t)) \quad (1)$$

It is assumed that noise has a random nature and can be represented by a zero mean random process, therefore it can be removed by averaging data and noise. The expectation is that the signal and the noise are not correlated, and that image data are represented by some regular

patterns, so that correlation functions of image data between different parts of the image can be found. If data and noise are not correlated, the selection of averaging kernels, should allow us to preserve the details while reducing the amount of noise.

The Gaussian noise, usually produced by analogue circuits $n_a(v, t)$ has a thermal nature and can be approximated by a zero mean Gaussian random process. Analogue noise does not depend on characteristics of light, and is added to the useful image data by analogue sensor components. In the proposed research a Gaussian distribution with a standard deviation of σ_a is used to characterize the analogue noise.

Further, sensor defects affect the level of resulting noise. Common sensor defects found in many sensors are namely, line, column and fixed pattern noise. Line and column noise can be characterized using a Gaussian noise distribution, applied in each dimension x and y with corresponding standard deviations σ_{ax} and σ_{ay} . Fixed pattern noise can be characterized by using a Gaussian noise distribution σ_{fpn} which is fixed over the time. Sensor defects can be considered as an addition to analogue noise $n_a(v, t)$.

Another source of noise present in a typical imaging sensor is photon noise $n_q(I_p(v, t))$, which increases as the light level increases, due to a larger numbers of photons captured by the sensor. This noise source can be described as a random process with a Poissonian distribution with standard deviation σ_q . It is assumed that $I_s(v, t) \cong I_p(v, t)$, which in practice means that the signal is stronger than noise. According to that assumption it can be put $n_q(I_p(v, t)) \cong n_q(I_s(v, t))$. The proposed system architecture can benefit from the knowledge of sensor noise characteristics. Sensor noise modelling was investigated in [12],[13],[16],[17] and standard deviation for sensor noise can be defined as follows:

$$\sigma^2(v, t) = \sigma_a^2 + \sigma_q^2 \times \frac{I_s(v, t)}{I_{\max}} \quad (2)$$

Where I_{\max} is a maximum level of intensity captured by the sensor. The standard deviation of sensor noise vs the intensity of the light captured by the sensor was calculated at different analogue gain values: 1, 4, and 8 times. The sensitivity of the sensors used in our experiments, corresponds to ISO100 at an analogue gain of 1, ISO400 at a gain of 4 and ISO800 at a gain of

8, as standard. The corresponding noise curves for the sensor AS3372, refer appendix B, are represented in Figure 3:

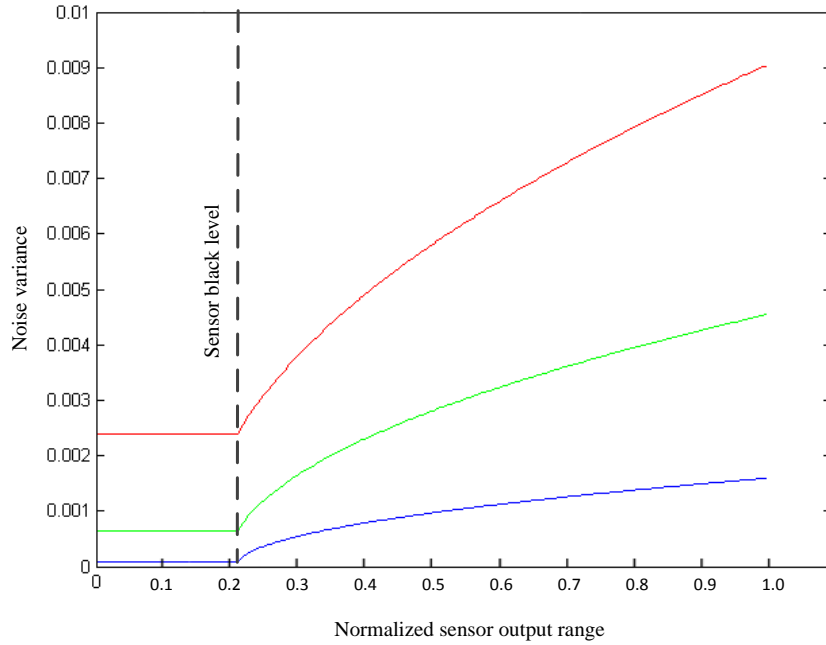


Figure 3 Sensor noise at **ISO100**, **ISO400** and **ISO800**

Further the precision of equation (2) can be illustrated by the scatter plot and the best fit graph illustrated in Figure 4:

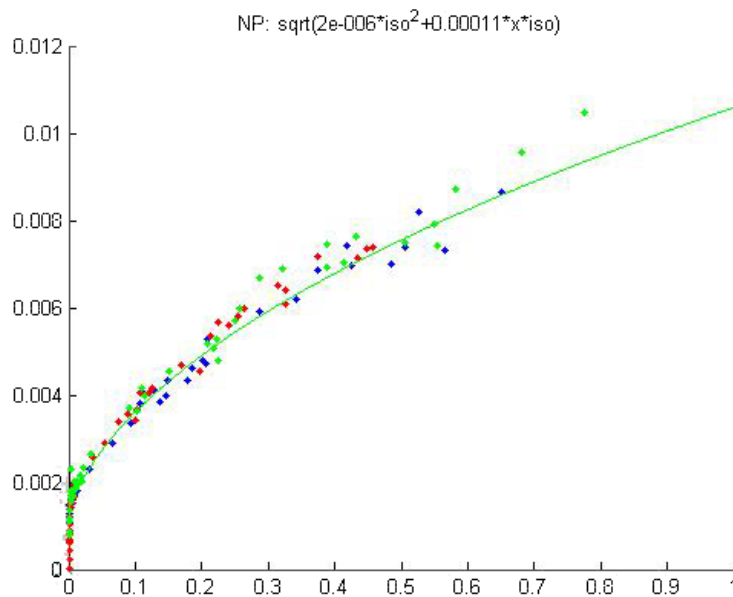


Figure 4: Sensor noise experimental data

In Figure 4, red, green and blue dots represent noise variances for the corresponding pixel colours, measured at different light intensities. The above graph is usually referred to as a sensor noise profile. For the experimental work conducted in this thesis different image sensors were used, refer Appendix B. The noise profile presented in Figure 4 was experimentally measured for a sensor AS3372 at ISO100, refer Appendix B for details. The values of σ_a^2 and σ_q^2 characterize the noise characteristics of a sensor and being used in equation (2) can provide the estimation of a noise for each pixel at given ISO settings.

2.5. Interpolation and statistical data accumulation

As we have seen above, the image data captured by the sensor cannot be represented directly as the image captured by the image sensor. Important stages of image reconstruction, usually referred to as image processing stages, would be required to separate the image data captured by the sensor from noise, normalize data levels, compensate for sensor defects. The captured image data has to be interpolated in order to obtain brightness, contrast and colour information and subsequently to restore the true colours of the image. Spatial filtering kernels can be used to significantly reduce the image data variance within the kernel radius. In a situation where several images are captured for the same scene and the sequence of images is available, temporal averaging can significantly improve signal to noise ratios of the image. At the same time the details in the image can be preserved if the correlation between the parts in the reference image and the accumulated image can be found. It is important to note that the sensor data acquisition time, usually called the integration time, can be increased in order to improve the signal to noise ratio in the data captured by the sensor. In the proposed research both spatial and temporal data accumulation techniques has been used. Both spatial and temporal accumulation techniques require respectively intra frame and inter frame image data matching to be performed, to enable data accumulation. The spatial data accumulation methods were proposed in [1],[3],[4],[5],[6],[11],[19],[20]. It was proved that the accumulation of image data from single image can significantly improve signal to noise ratio. The idea of temporal data accumulation (inter-frame) was investigated in [1],[18],[19],[24],[25]. It was proved that inter-frame image data block accumulation very efficient technique for image or video de-noising. Let $I(v, t)$ represent an image captured by the sensor at a discrete time t . Let us consider a pixel of the image $I(v, t)$ with coordinate x at a discrete time t and its neighbourhood $N(v, t)$ defined as a set of pixels of the image $I(v, t)$ within an area of size of $k \times k$ and a centre at coordinate v . In

order to perform data averaging, a limited search area $S(v, t)$ of size of $s \times s$ and a centre at v within the same image $I(v, t)$ can be used. In this work the values of s and k are constrained to satisfy the following rule $s > k > 1$ and both s and k are odd numbers. As the size of $S(v, t)$ was defined to be $s > k > 1$ we can define a set of pixels $I(v_j, t) \in S(v, t)$. For each pixel $I(v_j, t)$ at coordinates v_j a neighbourhood $N(v_j, t)$ can be defined, as a set of pixels of the image $I(v, t)$ within an area of size of $k \times k$ and a centre at coordinates v_j . In this research it is assumed that the image is formed by regular patterns and the self-correlation function is not singular, in that case spatial or temporal averaging of correlated data will reduce the amount of noise, given that the signal and noise are not correlated. Thus the neighbourhoods $N(v_j, t)$ of image $I(v, t)$ can be averaged with weights $w(v, v_j, t)$ to produce the de-noised image neighbourhood $N'(v, t)$, according to the equation (3):

$$N'(v, t) = \frac{\sum_{v_j \in S(v, t)} w_N(v, v_j, t) \times N(v_j, t)}{\sum_{v_j \in S(v, t)} w(v, v_j, t)} \quad (3)$$

A new image neighbourhood $N'(v, t)$ corresponds to a new (de-noised) image $I'(v, t)$, so that $N' \in I'$. Repeating the accumulation process defined in equation (3) for each pixel of $I(v, t)$ and corresponding neighbourhood $N(v, t)$ it can be seen that the accumulated image $I'(v, t)$ can be defines as:

$$I'(v, t) = \frac{\sum_{v_j \in S(v, t)} w(v, v_j, t) \times I(v_j, t)}{\sum_{v_j \in S(v, t)} w(v, v_j, t)} \quad (4)$$

In equation (4) weights $w(v, v_j, t)$ are obtained as averaging of weights $w_N(v, v_j, t)$ within the search window S . It can be noted that v_j can take $s \times s$ possible values. Further the t is omitted for the compactness of formulas as all data utilized in the accumulation process corresponds to the same time t . It can be seen that the weights $w(v, v_j, t)$ should have a higher value for correlated pixel neighbourhoods and lower or zero value for non-correlated pixel neighbourhoods. The efficiency of noise reduction and the preservation of image details thus fully depend on the correctness of weights $w(v, v_j, t)$ calculation. The process of intra-frame accumulation is represented in Figure 5:

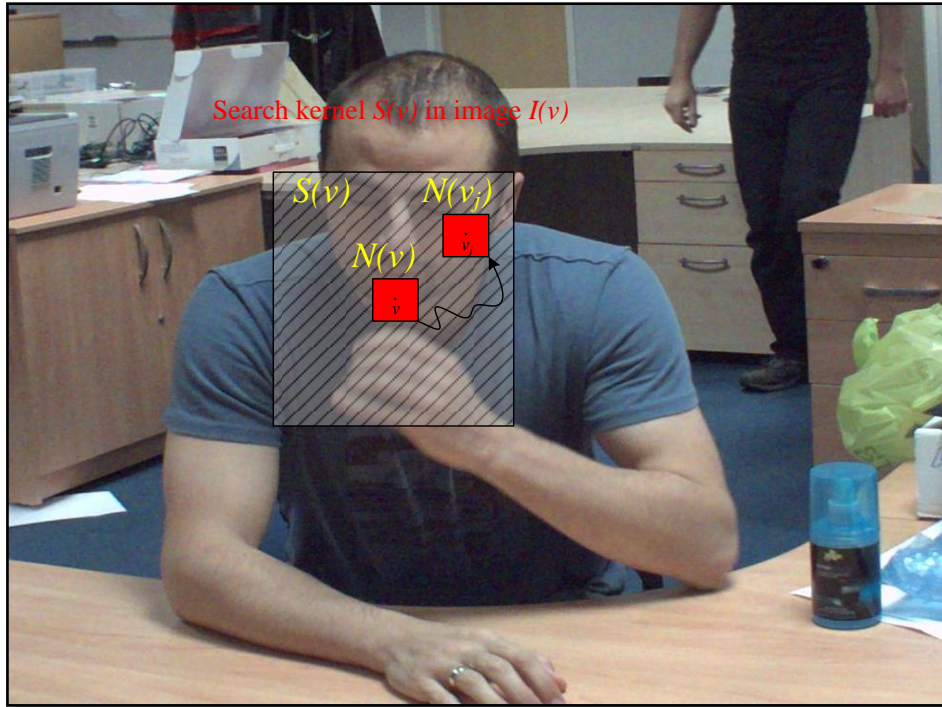


Figure 5: Intra-frame accumulation

In situations where the scene is captured by several images, images taken at different time can be accumulated in a frame buffer $I_{fb}(v)$. In this case search for the matching image data is performed between the reference image $I(v)$ and accumulated image in the frame buffer $I_{fb}(v)$. From the frame buffer image data, the matching current de-noised image is obtained according to the following equation (5):

$$I_{fb}'(v) = \frac{\sum_{v_j \in S(v)} w_{fb}(v, v_j) I_{fb}(v_j)}{\sum_{v_j \in S(v)} w_{fb}(v, v_j)} \quad (5)$$

The temporal difference between matching $I_{fb}'(v)$ and $I(v)$ is minimized, so that their linear combination can be used to update the content of the frame buffer and stored for the next frame as de-noised accumulated image. The process of inter-frame image data matching is presented in Figure 6:

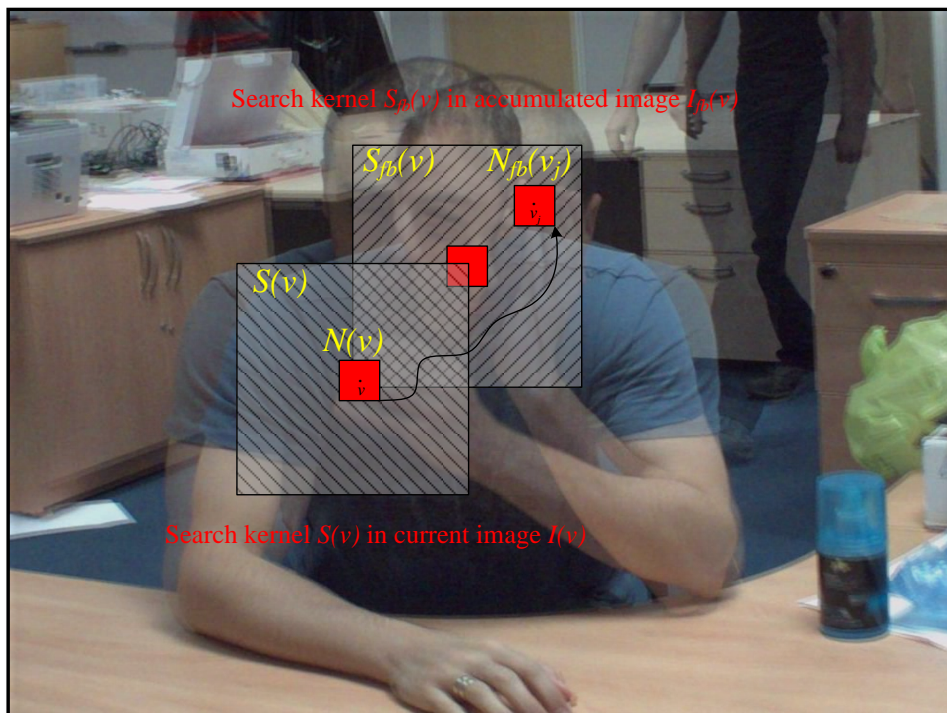


Figure 6: Inter-frame image data matching

Both data accumulation methods are widely used in [1],[3],[4],[6],[9],[10],[11],[14],[15],[20],[24] to reduce the amount of noise in captured images. However the challenge is to find optimal $w(v, v_j)$ and $w_{fb}(v, v_j)$ at a minimum cost and decide on the strategy when the data correlation matrix is singular.

2.6. Multi-scale data segmentation

In the proposed research the sensor noise distribution has been investigated in different spatial frequency bands. It was found that normally sensors would generate significant amount of noise in the high frequency band as well as in the relatively low frequency bands. Practically the frequency analysis of noise shows that the noise speckles of 1-2 pixels in size are as common as large noise patches of 30-60 pixels in size. In this work in order to efficiently filter the image data, a multi-scale approach has been used to achieve a large effective filtering kernel, while keeping resource usage minimised, and at the same time to improving the efficiency of filtering. The idea of data processing in frequency bands is not new and was used in [2],[3],[4],[6],[8],[11], however, in this research the challenge was to find the method of representation of image data by independent frequency bands, which is efficient for hardware implementation. A requirement was to keep the complexity of such a transformation low, in

order to allow real-time performance and low resource usage. In the proposed research the algorithms have been constrained to work in the linear Bayer RAW domain, which results in particular requirements for the transformation and filtering techniques. To this effect, the Gaussian kernels have been used to perform image data filtering.

The multi-scale approach we used can be described as a Gaussian pyramid of the image $I(x, y)$. It is computed for each colour plane according to equation (6):

$$G_k(x, y) = \begin{cases} I(x, y); & k = 0 \\ \iint G_{k-1}(x + x_g, y + y_g)g(x_g, y_g)dx_g dy_g; & k = 1,2,4,8 \end{cases} \quad (6)$$

As in the general case of this research we used the kernel $g = g'_i \times g_i$, where,

$$g_i = (1/16, 0, 4/16, 0, 6/16, 0, 4/16, 0, 1/16).$$

The difference between the previous Gaussian image G_0 and filtered image G_1 is named as the Laplacian image L_0 . This process is continued to obtain a set of band-pass filtered images, expressed as in equation (7):

$$L_k(x, y) = G_k(x, y) - G_{k+1}(x, y); \quad k = \{0, 1, \dots, K\} \quad (7)$$

From the above formula, it can be seen that the Laplacian pyramid is a set of band-pass images. It contains all of the image's textural features, at different scales. The bottom level of the pyramid contains the highest spatial frequency components such as the sharp edges, textures, high-frequency noise etc. The top level contains the lowest spatial frequency components. The intermediate levels contain features gradually decreasing in spatial frequency from high to low. A Laplacian pyramid has an important feature: the sum of Laplacian images will produce the original image:

$$I(x, y) = \sum_{k \in 0:K} L_k(x, y) \quad (8)$$

Since the filter parameters are related to the noise standard deviation of every level of the Laplacian pyramid, the noise characteristics of a Gaussian-Laplacian image pyramid needs to be investigated. If the standard deviation of the original noisy image is σ_0^2 , the noise variance σ_s^2 of the smoothed image, filtered with Gaussian kernel $g(x_g, y_g)$ is given by equation (9):

$$\begin{aligned}
 \sigma_s^2 &= \iint (G_s(x, y) - \overline{G_s(x, y)})^2 dx dy = \\
 &\iint ((\iint G_{s-1}(x+x_g, y+y_g)g(x_g, y_g)dx_g dy_g - \iint \overline{G_{s-1}(x+x_g, y+y_g)}g(x_g, y_g))^2 dx_g dy_g) dx dy = \\
 &\iint ((\iint g(x_g, y_g)^2 (G_{s-1}(x+x_g, y+y_g) - \overline{G_{s-1}(x+x_g, y+y_g)})^2 dx_g dy_g) dx dy = \\
 &\frac{1}{4\pi\sigma_g^2} \iint ((G_{s-1}(x, y)_i - \overline{G_{s-1}(x, y)})^2) dx dy = \frac{1}{4\pi\sigma_g^2} \sigma_{s-1}^2 \quad (9)
 \end{aligned}$$

Here, σ_g^2 is the standard deviation of Gaussian kernel and $\overline{G_s(x, y)}$ is a mean value of $G_s(x, y)$ by definition of standard deviation. In a particular case of image data processing, $\overline{G_s(x, y)}$ can be computed as temporal average of $G_s(x, y)$, when multiple observations of $G_s(x, y)$ are available, or as an average of $G_s(x, y)$ in a spatial kernel with centre at x, y assuming that the ground truth data variation is zero. In the proposed research, $\sigma_g^2=1$. Thus, if σ_0^2 denotes the noise variance of the level 0 image of an Gaussian pyramid, the noise variance of the level k image in the Gaussian pyramid is given by equation (10), below:

$$(\sigma_G^k)^2 = \frac{1}{(4\pi)^k} \sigma_0^2 \quad (10)$$

The noise variance of every L_k can be computed according to equation (11).

$$(\sigma_L^k)^2 = \frac{1}{(4\pi)^k} \sigma_0^2 - \frac{1}{(4\pi)^{k+1}} \sigma_0^2 \quad (11)$$

As the level of Laplacian pyramid increases, the amount of noise of Laplacian image decreases rapidly. In the proposed research the correlation between image parts is calculated by using of block matching technique with greatly reduced comparison window size for calculating the weights on scales $k = \{1, 2, \dots, K\}$, as the noise levels of Laplacian images decrease. For image data where the useful signal is well above the noise level, comparing similarity in small windows or even comparing pixel values gives reliable information for noise filtering.

2.7. Temporal methods: image data accumulation using the Gaussian background model.

In the proposed temporal noise reduction algorithms, temporal filtering is used to reduce the variations of image data from frame to frame, over time. In practice, temporal variations can produce large scale noise, thus the removal of temporal noise can reduce requirements for the maximum filtering kernel size used in spatial filtering. In the proposed research we assume that temporal differences can be described as a mixture of Gaussian zero mean process and random process with Poissonian distribution, which has being investigated in [1],[15],[17],[25],[26] and denoted as $\sigma^2(v,t)$, (refer equation (2)). It is assumed that moving objects typically found in video sequences will consistently keep moving in the same direction (i.e. will have salient movement) at least for some period of time. Though an attempt is made to minimize the temporal difference by performing temporal matching, it is expected that the temporal differences will be increased in the areas corresponding to the moving objects. As the measurements of motion will be affected by noise using a Kalman Filter as investigated in [53],[54],[55] can help to minimize the error of prediction of a new scene by using the knowledge about temporal differences from the previous scenes. Assume that the original image $I(v,t)$ and accumulated frame buffer image $I_{fb}(v,t)$ are functions of discrete time t and coordinate v , are the two signals that are input to the temporal filter (coordinate v is omitted from the further equations in this chapter for the purpose of clarity), the temporal difference $D(t)$ for each coordinate x would be calculated as:

$$D(t) = |I_{fb}(t) - I(t)| \quad (12)$$

It is noted that at any time interval t the new pixel value $I(t)$ of an image can belong to the moving object or the static background, with some probability. Further it is assumed that the temporal difference $D(t)$ gradually increases as the moving object enters the scene and the object moves consistently, i.e. not changing the direction of the motion randomly. In other words it is assumed that the image can be reconstructed by the transformation of the image stored in a frame buffer at time $(t-1)$ with error $D(t)$. When temporal difference is calculated, the expectation of the current image $\hat{I}(t)$ for each pixel can be found using a Kalman filter. In our case we assume that the error $D(t)$ is small and the content of the frame buffer is changing

slowly. Thus the expectation of the modified frame buffer image is $\hat{I}_{fb}(t) \cong I_{fb}(t-1)$. In the situations when our assumption is not valid, the efficiency of the accumulation will deteriorate and the effect of noise reduction will reduce. Assuming that the error at each discrete time t is $D(t)$ and the correction coefficient is $K(t)$, we get:

$$\hat{I}(t) = \hat{I}_{fb}(t) + K(t) \times D(t) \quad (13)$$

$$\hat{I}(t) = \hat{I}_{fb}(t) + K(t) \times (I(t) - I_{fb}(t-1)) \quad (14)$$

As $\hat{I}_{fb}(t) \cong I_{fb}(t-1)$, the resulting formula would be:

$$\hat{I}(t) = I_{fb}(t-1)(1 - K(t)) + I(t) \times K(t) \quad (15)$$

In our work we found $K(t)$ as the coefficient optimal for the Kalman filter to be defined as:

$$K(t) = K(t-1) \times \left(1 - \frac{1}{\sigma_I}\right) + D(t) \times \frac{1}{\sigma_I} \quad (16)$$

Function $K(t)$ in a real system should be limited by $K_{\min} < K(t) < K_{\max}$, where parameters K_{\min} and K_{\max} correspond to the maximum and minimum frame buffer image update rates, respectively. Parameter σ_I is proportional to the standard deviation of pixel intensity and is defined as a constant for each particular system. When both $K(t)$ and $\hat{I}(t)$ are calculated they are saved in the frame buffer to be used for the next frame. When the proposed algorithm is implemented in hardware, the parameter $K(t)$ can be quantized to 4 bit precision data. For the equation (16) to be operational, the parameter $K(t)$ can be calculated via lookup tables. At each discrete time t the content of a frame buffer: $I_{fb}(t)$ and $K(t)$ can be considered as temporal mean and variance of an image. Calculating mean and variance for the background and foreground parts of an image we found Gaussian mixture model applicable. Gaussian model of a background is very efficient method of data accumulation as it reduces memory bandwidth and helps to avoid appearance of motion blur artefact due to adaptive filtering and separation between static background and moving foreground.

Chapter 3

Block matching de-noising method for photographic images, applied in Bayer RAW domain, optimized for real-time implementation

3.1. Introduction

In any imaging system producing still or motion pictures, noise reduction is a very important component, which defines the resulting image quality of the image processing pipeline. Though noise reduction techniques are known for many years, practically their use in consumer electronics, video surveillance, professional photo and video devices is constrained and therefore rather limited.

The idea proposed in this work is to adapt block matching, block accumulation filters in a multi-scale system as investigated in [5],[8],[9], to de-noise photographic and video images in Bayer RAW data space as in [2], using sensor noise modelling as in [6],[9]. The idea of block matching block accumulation applied to Laplacian pyramid in Bayer RAW domain, using sensor noise modelling is new and was not investigated yet. There are a number of advantages in doing the processing on RAW data. One of the advantages is to have predictable noise characteristics, thus allowing making the decision about noise levels easier and more reliable.

The evaluation of existing CMOS sensors shows that the kernel size required for efficient noise reduction should not be smaller than approximately 31×31 pixels for full HD sensors. Using large sensors (e.g. 12 megapixels and above) the kernels of the size of 127×127 pixels are absolutely necessary, while even a 31×31 pixel size kernel is considered to be relatively large for a typical image processing pipeline.

The complexity of a block matching algorithm can be analysed as follows. If N is the number of pixels in an image, $k \times k$ is the number of pixels in a comparison window K (matching

block), the complexity of such algorithm is $O(k^2 \times N^2)$. For computational purposes, the simplified algorithm restricts the search range of similar windows in a search window S of size of $s \times s$ pixels. The final complexity of the algorithm is $O(s^2 \times k^2 \times N)$. By fixing the search window S at the size of 11×11 pixels and the comparison window K size at 5×5 pixels, the complexity of such algorithm would be $O(25 \times 121 \times N)$. The complexity of local algorithms with kernel of 11×11 pixels would be $O(121 \times N)$. Even the simplified algorithm still takes significant time to de-noise a full HD image on a general purpose PC. For a hardware implementation, performing block matching in a window of 11×11 pixels with block size of 5×5 pixels is feasible but not practical as the minimum kernel size of 31×31 pixels is required in most practical cases. It can be seen that the high computational complexity makes it not feasible to tackle with practical issues by applying non-local means de-noising approach directly.

In order to address the issue of algorithmic complexity a multi-scale approach for running non-local means on a raw sensor data is adopted. The simplification of filter design for higher filtering scales is also considered. Multi-scale approach enables modular filter design. Filters used on each scale, except for the first, can be the same. Another reason for use of a multi-scale architecture is to avoid specific banding artefacts produced by one large filter, seen in other implementations (Adobe Lightroom ©Adobe) on smooth gradients, while keeping the kernel sizes on each level of scale pyramid small. As investigated by other researchers de-noising in transform space (e.g. DCT, Fourier, Wavelet) has a number of advantages. The most significant advantage is that filtering does not necessarily lead to contrast and resolution loss and does not produce banding artefacts. In the chosen system architecture, firstly, image transformation cannot be performed on Bayer RAW data directly. Secondly the advantage of knowing the noise levels will be lost. Additionally image transform on its own is a computationally heavy task, increasing the computational cost in case of software implementation or requiring additional memory and logic when implemented in hardware. Finally the need to de-mosaic the image, prior to the transformation, would undermine the entire concept. Applying transforms on mosaic colour planes [3] is also possible but has other disadvantages. In the proposed algorithm an attempt is made to solve the abovementioned

known problems and optimize the algorithms to enable their implementation in commercial grade programmable logic devices.

The major artefact produced by filtering with a large kernel in the pixel intensity space is a contouring effect on smooth gradients. Appearance of this artefact can be diminished by reducing the filtering kernel size. Unfortunately, this contradicts with the main goal – the desire to increase the kernel size to deal with larger scale noise. In order to work around this contradictory requirement a multi-scale approach for Bayer RAW data is proposed. In the proposed system we decompose the image into 4 bands (in the case of a large kernel), and filter each scale with a relatively small kernel filter, 9×9 for the first scale and 7×7 for all other scales. Filters applied to the scales of Laplacian pyramid would correspond to a 15×15 pixels kernel for the first scale, 31×31 pixels kernel for the second scale, 63×63 pixels kernel for the third scale, and so on, in the final RGB interpolated data. Effective kernel size, achieved with this approach, reached 127×127 pixels in the interpolated final RGB image. Multi-scale approach however has an impact of increased residual noise, though the problem is less prominent when compared to what can be achieved with Adobe Lightroom (see the results section).

Applying noise reduction early in the image processing pipeline helps to improve signal to noise ratio for the rest of the pipeline, supplying cleaner data for the processing blocks such as, dynamic range compression, de-mosaic, colour correction and others. All these algorithms are sensitive to noise. The better the signal to noise ratio, the more reliable is the result that can be achieved from the complete image processing pipeline.

Due to the kernel size constraints and the requirement to implement the algorithm in hardware, it is highly desirable to perform processing in sensor RAW data space. The proposed algorithm was implemented in a real image processing pipeline to process both video and still images. The block diagram of the proposed algorithm that uses a 3 layer Laplacian pyramid is represented in Figure 7:

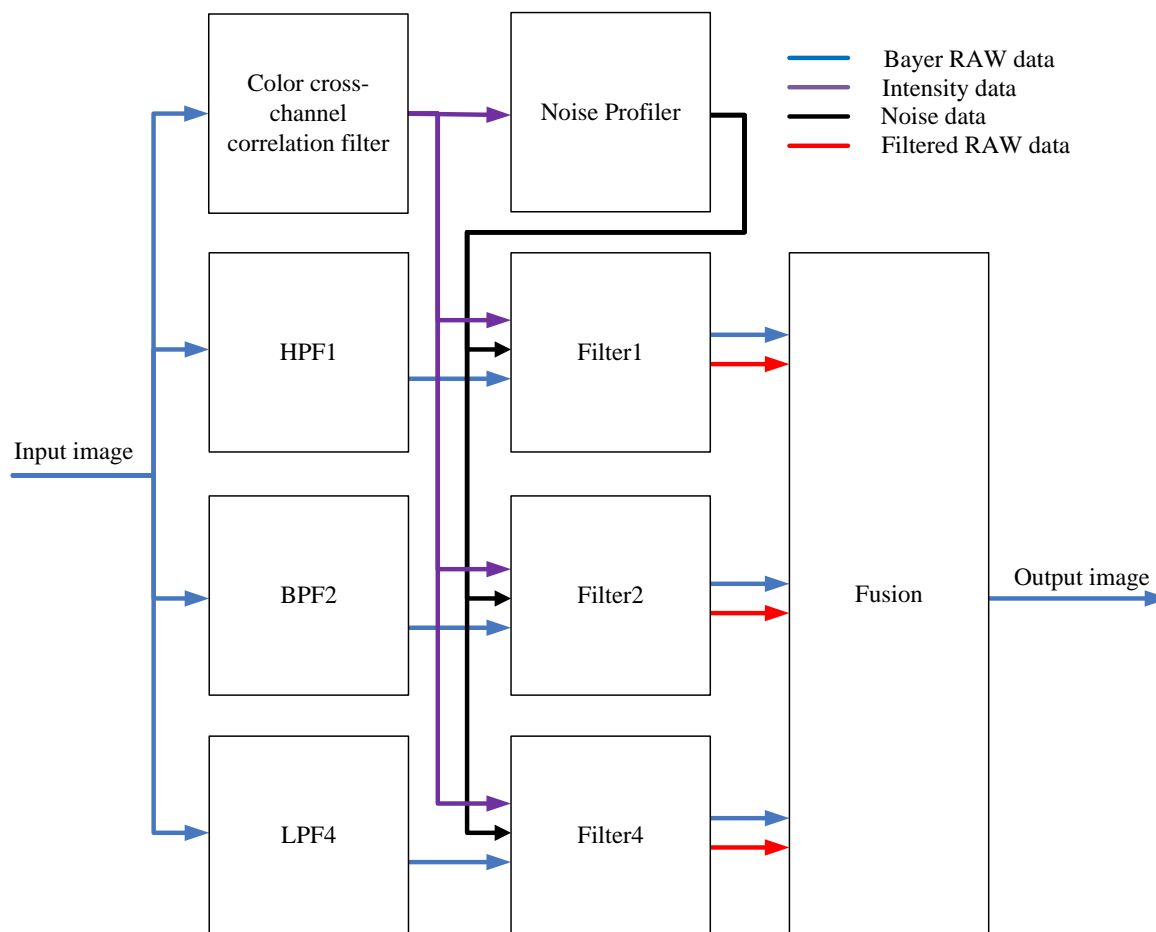


Figure 7: Algorithm block diagram.

The first step is to calculate the intensity for each pixel, including the information on colour cross- channel correlation. Further as mentioned previously the proposed algorithm will work in RAW data space. For the sensors with RGB Bayer pattern the intensity calculation is important for two reasons: an equal decision regarding filtering should be made for any colour when pixels of different colours belonging to the same object detail. Calculating intensity also helps to obtain an image with lower noise by eliminating colour cross-channel noise, thus helping to detect image details more reliably. Intensity calculation in multiple scales also removes the possibility of a checker pattern artefact appearance in the resulting RGB image, a fundamental drawback of some previously proposed algorithms [3].

The second step of the proposed algorithm is to calculate noise characteristics. This calculation is performed on a pixel basis. In parallel with intensity and noise profile calculations the image

is passed through the set of filters: High pass filter (HPF1), Band pass filter (BPF2), Low pass filter (LPF4), see Figure 7. When the filtering is completed, image data, intensity and noise profile data are used by non-local means filters (Filter 1), (Filter 2), (Filter 4), see Figure 7. The results of non-local means filtering are sent to the fusion block, i.e. the final processing block.

3.2. Block matching approach for Bayer RGB sensors

Non-local means approach [2],[3],[21],[25] is a technique, when the decision regarding similarity of different parts of the image is made based on evaluation of blocks of a certain size, and averaging is applied to a blocks of pixels. In the proposed algorithm this approach is applied in Bayer sensor data domain. In the non-local means technique a pixel neighbourhood $N(v)$ will be used to obtain the measure of similarity with another neighbourhood $N(v_j)$ as defined in Chapter 2, which then will be compared against the noise levels and a decision will be made, whether to use that neighbourhood $N(v_j)$ to average with the current pixel neighbourhood $N(v)$ and what averaging weight to use. The idea of RAW data filtering using non-local means technique was described in [3],[4], however in the proposed work several important improvements to the basic approach have been introduced. In the proposed work, two methods are developed to set appropriate thresholds for each block accumulation, dynamically, in contrast to previous attempts that used a fixed threshold value. The sensor noise was modelled to obtain an estimate of noise levels, refer Chapter 2, equation (2), to calculate the weights for block averaging. Further a method for non-linear data analysis has been developed to estimate the energy of image details, leading to the prediction of correlation values, which enables the guaranteeing of preservation of image details, while at the same time on image parts where no details can be found, the filtering strength is increased. In the proposed research a concept of spatial data accumulation is used, refer Chapter 2, equation (5). In the proposed algorithm the block matching and accumulation will be applied to the layers of Laplacian pyramid of image $I(v)$, which in its turn is obtained as an intensity of a sensor RAW data. Let us assume $N_k(v)$ to be a neighbourhood of a noisy Laplacian image $L_k(v)$ at coordinate v . In the proposed algorithm a limited search range $S(v)$ of size of $s \times s$ is used. Thus the neighbourhood $N_k(v_j) \in S(v)$ of Laplacian image $L_k(v)$ can be averaged with

weights $w_k(v, v_j)$ to produce a de-noised Laplacian image, $L'_k(v)$, (refer Chapter 2, equation (5)), according to the equation (17):

$$L'_k(v) = \frac{\sum_{v_j \in S(v)} w_k(v, v_j) L_k(v_j)}{\sum_{v_j \in S(v)} w_k(v, v_j)} \quad (17)$$

The weights $w_k(v, v_j)$ are calculated from block differences $d_k(v, v_j)$ for corresponding neighbourhoods $N_k(v)$ and $N_k(v_j) \in S(v)$ for each pixel coordinate v .

$$d_k(v, v_j) = \sum_{(v_j) \in S(v)} |N_k(v) - N_k(v_j)| \quad (18)$$

However each block difference $d_k(v, v_j)$ will be constructed by the sum of noise differences $d_{k\sigma}(v, v_j)$ and image autocorrelation functions $d_{k\alpha}(v, v_j)$. In the proposed algorithm no attempt is made to rotate or scale blocks to achieve better correlation. Thus, it is required to estimate the value of $d_{k\alpha}(v, v_j)$ to adjust $w_k(v, v_j)$ calculation. In this work $d_{k\alpha}(v, v_j)$ is estimated by the evaluation of the block difference $d_k(v, v_j)$ and expected noise level $\sigma(v)$, adjusted for the Laplacian scale according to the equation (8). For each block we estimate $d_{k\sigma}(v, v_j)$ according to equation (19):

$$d_{k\sigma}(v, v_j) = d_k(v, v_j) - d_{k\alpha}(v, v_j) \quad (19)$$

Resulting $d_{k\sigma}(v, v_j)$ is consequently converted into corresponding $w_k(v, v_j)$ by comparison with the expected variation of noise, derived from the sensor characterization as specified in equation (2). On the level k of Laplacian pyramid the result of averaging $L'_k(v)$ in each pixel location will be obtained as the result of averaging of data $L_k(v)$ within the search area $S(v)$ with weights $w_k(v, v_j)$. As discussed in Chapter 2 the number of weights $w_k(v, v_j)$ and block differences $d_k(v, v_j)$ associated with each pixel location will be $s \times s$, where s is a size of search area S . In this research the block differences $d_k(v, v_j)$ were calculated in linear Bayer

RAW data domain, this fact allows us to use the estimation of a noise level $\sigma(v)$ as a reference. In this work we propose to normalize the block differences according to equation (20), (21) and (22):

$$d_{k\sigma}(v, v_j)_{norm} = d_k(v, v_j)_{norm} - d_{k\alpha}(v, v_j)_{norm} \quad (20)$$

$$d_k(v, v_j)_{norm} = d_k(v, v_j) / \sigma(v) \quad (21)$$

$$d_{k\sigma}(v, v_j)_{norm} = d_k(v, v_j) / \sigma(v) - d_{k\alpha}(v, v_j)_{norm} \quad (22)$$

In this work we propose to find $d_{k\alpha}(v, v_j)$ as follows:

$$d_{k\alpha}(v, v_j)_{norm} = \min(d_k(v, v_j) / \sigma(v)) \quad (23)$$

The equation (23) is expected to be valid in the situations when the image data in the search area $S(x)$ contains neighborhoods $N_k(v)$ and $N_k(v_j)$ with good correlation, i.e. small value of $d_{k\alpha}(v, v_j)_{norm}$, presumably the minimum block difference will be achieved in a situations when the best match can be achieved and the normalized block difference is stayed within several sigma range. In the situations when good match in the search area is not possible the estimation of $d_{k\sigma}(v, v_j)_{norm}$ has to be forced to be a large number, which is finally result in details preservation at cost of reduced noise filtering efficiency. This corner case however is an important situation when logical decision can resolve the situations when the algorithm cannot be efficiently used. In this research it is proposed to introduce a coefficient Ca , which will force $d_{k\sigma}(v, v_j)_{norm}$ to be increased in a “no match” situations. The final equation for $d_{k\sigma}(v, v_j)_{norm}$ provided below:

$$d_{k\sigma}(v, v_j)_{norm} = (d_k(v, v_j) / \sigma(v) - \min(d_k(v, v_j) / \sigma(v))) \times Ca \quad (24)$$

In this work it was experimentally found that the optimum Ca can be calculated as:

$$Ca = \frac{Cn + \text{mean}(d_k(v, v_j) / \sigma(v))}{Cn + \min(d_k(v, v_j) / \sigma(v))} \quad (25)$$

The coefficient Cn is usually chosen from range 1:2. The physical meaning of equation (25) is that the coefficient $Ca \approx 1$ and has no effect on the final weights $w_k(v, v_j)$ in situations when normalized differences are not much different from 1, whereas in situations when the image contains texture and details, but reliable block matching is not possible the coefficient Ca will become greater than 1 as the variations in block matching results will increase. The equation (25) for the coefficient Ca is not unique, however in experiments it was seen that proposed formula performs well and easy for implementation. The conversion of normalized $d_{k\sigma}(v, v_j)_{norm}$ into $w_k(v, v_j)$ is performed according the equation (26):

$$w_k(v, v_j) = \exp\left(-\left(\frac{d_{k\sigma}(v, v_j)_{norm}}{Csigma}\right)^2\right) \quad (26)$$

In the proposed algorithm there are two constants: Cn and $Csigma$. The first constant Cn is used to adjust subjective parameter of non-regular image details preservation, which was set to 1.5 in all experimental results presented. The second constant $Csigma$ is used to adjust subjective parameter of noise-suppression aggressiveness. In all experimental results presented this parameter was set to 1 and never changed. The algorithm presented in this research automatically adopted to different sensors and lighting conditions, given that the sensor noise model was provided and was correct. The experimental results are provided in a next section.

3.3. Experimental results

In the experiments conducted three de-noising algorithms were compared: the proposed algorithm NLBM; BM3D and Adobe Lightroom (©Adobe Systems). Two individual testing procedures were adopted, a Simulated test and a Real world test.

3.3.1. Simulated test

In this test a reference ISP (©Apical LTD) within which the proposed noise reduction algorithm is included was used for the evaluation of results. The proposed de-noising was applied to the sensor RAW data, as illustrated in Figure 5.

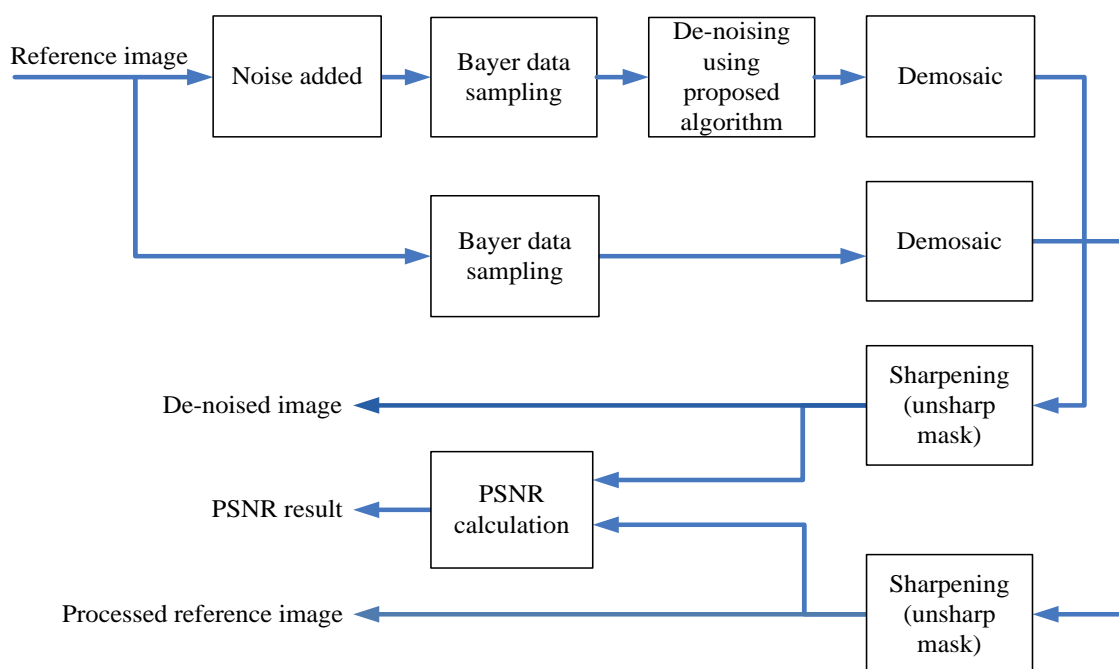


Figure 8: Block diagram of the simulated test process

For the purpose of comparison we applied BM3D and Adobe Lightroom algorithms on RGB data as these algorithms are designed for RGB data processing. All processing parameters were the same as used for the evaluation of the proposed NLBM algorithm:

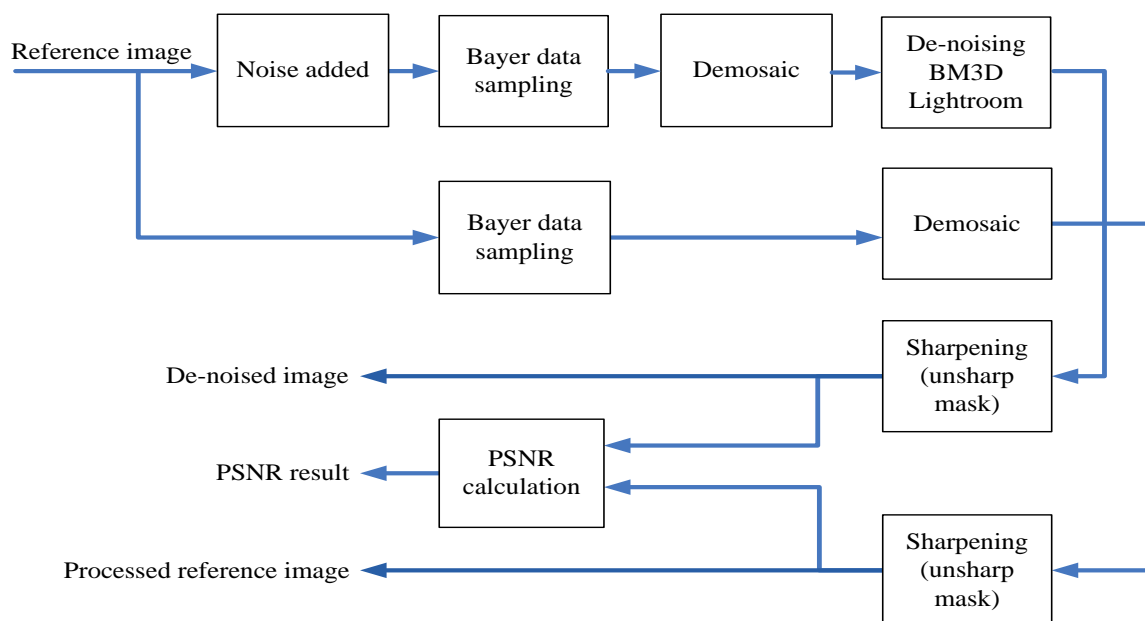


Figure 9: Block diagram of the simulated test procedures for BM3D and Adobe Lightroom.

In the above experiments a directional linear de-mosaicking algorithm, with a kernel of 5x5 pixels was used. The selection of the interpolation algorithm was based on attainable accuracy and predictable behaviour to minimize the effect of noise reduction on interpolation. An unsharp mask with an effective kernel of 1 pixel and strength 0.5 was applied to return the subjective sharpness of the output image to the level of the original image. The same sharpening algorithm (un-sharp mask) was used in all cases.

Three different levels of noise were added to the Kodak Test images $\sigma = 10/255, 20/255, 30/255$. The value of PSNR in dB, for each test image, calculated per colour and as an average of RGB (labelled as A in Table 1) values, for 3 different noise levels for each of 3 tested algorithms are presented in Table 1. The proposed noise reduction algorithm is referred as NLBM, Adobe Lightroom is referred as LR.

Chapter 3: Block matching de-noising method for photographic images, applied in Bayer RAW domain, optimized for real-time implementation





	σ		10/255			20/255			30/255		
			NLBM	BM3D	LR	NLBM	BM3D	LR	NLBM	BM3D	LR
(1)		R	30.97	29.37	30.04	27.80	26.77	27.43	25.98	24.99	26.08
		G	31.06	29.52	30.22	27.99	27.24	27.75	26.19	25.85	26.51
		B	31.21	29.49	29.69	28.11	27.18	27.30	26.34	25.75	25.97
		A	31.08	29.46	29.98	27.96	27.06	27.49	26.17	25.53	26.18
(4)		R	32.62	31.94	31.92	30.04	27.48	27.93	28.52	25.02	26.47
		G	34.01	33.76	33.02	31.28	30.82	30.24	29.83	29.28	29.67
		B	33.90	33.52	32.31	31.17	30.40	29.88	29.80	28.56	28.80
		A	33.51	33.07	32.41	30.83	29.56	29.35	29.38	27.62	28.31
(23)		R	36.16	35.90	34.83	32.49	30.65	30.10	30.20	27.18	28.15
		G	37.22	36.84	35.87	33.69	33.05	31.92	31.44	30.55	30.65
		B	36.81	36.48	34.16	33.22	32.50	30.86	31.17	30.15	29.33
		A	36.73	36.40	34.95	33.13	32.06	30.96	30.93	29.29	29.37
(24)		R	33.47	32.67	32.27	29.66	28.97	29.55	27.75	26.84	27.80
		G	33.44	32.83	32.56	29.76	29.24	29.63	27.73	27.33	28.03
		B	33.64	32.68	31.56	29.95	29.02	29.05	27.93	27.07	27.32
		A	33.51	32.72	32.13	29.79	29.07	29.41	27.80	27.08	27.71

Table 1: PSNR values for three different noise reduction algorithms on a sub-set of Kodak images.

It can be seen in Table 1 that the proposed algorithm produces better results compared to the benchmark algorithms in the case of all test images. The levels of improvement vary from 0.3dB to over 1dB for the listed average PSNR values. Further the subjective appearance of artefacts produced by the algorithm was considered. It is noted that the proposed algorithm does not emphasize on contrast preservation at cost of the quality degradation that results from artificial defects produced on the flat surfaces, when compared with the BM3D algorithm. However it is seen that the amount of image details is significantly more and the contrast is maintained better in the proposed algorithm when compared with images generated by Adobe Lightroom (see results for Kodak image 4 at noise 30/255 illustrated in Figure 10:

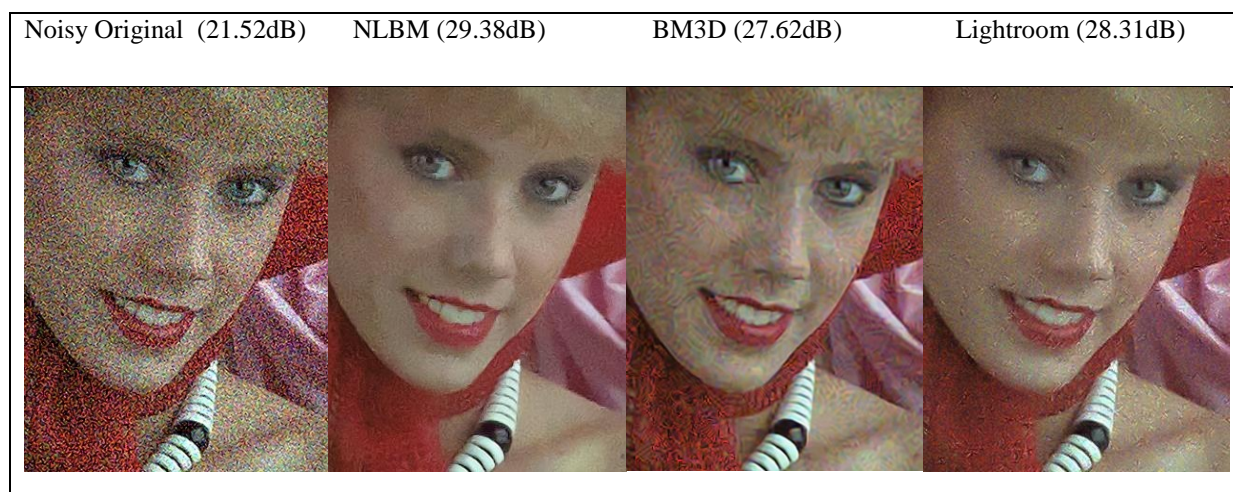


Figure 10: Kodak image (4) close-up.

The results of the noise reduction algorithms, applied for Kodak image (23) are presented in Figure 10:

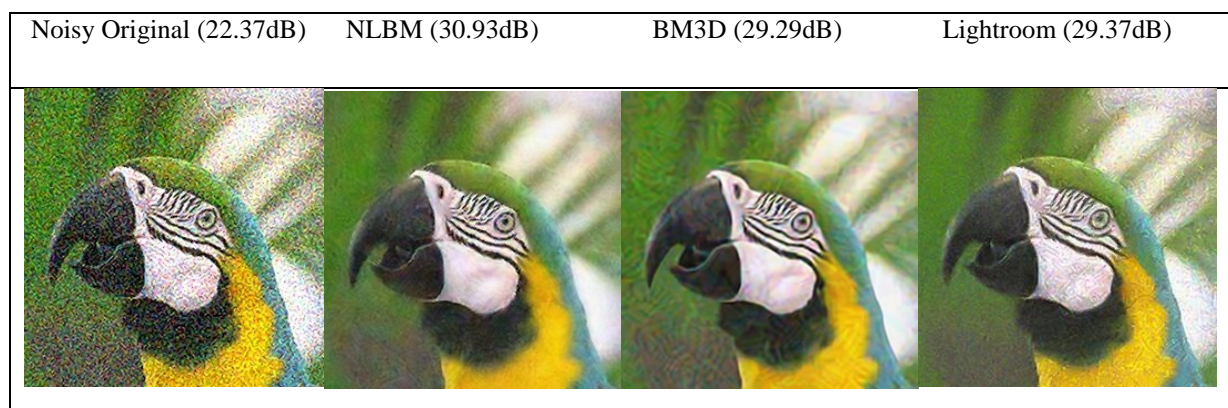


Figure 11: Kodak image (23) close-up.

In the experiments conducted the proposed algorithm's parameters were adjusted to produce the maximum PSNR values while keeping image details at a reasonable level and artefacts under control. As demonstrated by the results illustrated in Figure 10, Figure 11, a further reduction of residual noise in Lightroom or artefacts in BM3D is only possible at the cost of scarifying details, which would automatically lead to smaller PSNR values. Banding artefact on a smooth gradient produced by Adobe Lightroom in image 23 can be seen clearly in Figure 11. As seen in close-up images, the proposed algorithm provides the most natural looking images, even in extremely noisy conditions. However it is very important to evaluate the performance of noise reduction algorithm in real situations, as the noise generated by the image sensor is not exactly the simulated Gaussian random process. It is important to mention that the Poissonian noise found in actual imaging systems is not necessarily produce Gaussian noise in RGB domain. In practice noise distribution in RGB domain is significantly non-linear in most imaging systems, therefore the efficiency on noise reduction techniques assuming noise distribution to match the Gaussian distribution is compromised. In the forthcoming section the method of comparison of the proposed algorithm with industry leading Adobe Lightroom Spatial noise reduction approach is presented. Special arrangements have been made to perform a fair comparison of noise reduction algorithms.

3.3.2. Real world test

Experiments were conducted on real camera using a Sony Nex-5 camera. A series of images of the same scene were captured with different exposure times and ISO values. Images were taken in controlled lighting conditions with a fixed camera and a remote shutter release. Figure 12 samples taken at 200 ISO and 12800 ISO. During PSNR calculations special measures are taken to align images and eliminate any differences in pixel levels to avoid errors that emerge due to different intensity levels in images taken at different ISO.



Figure 12: Images taken by Sony Nex-5 camera at ISO12800 and ISO200.

The block diagram of the real world test procedure is presented in Figure 13:

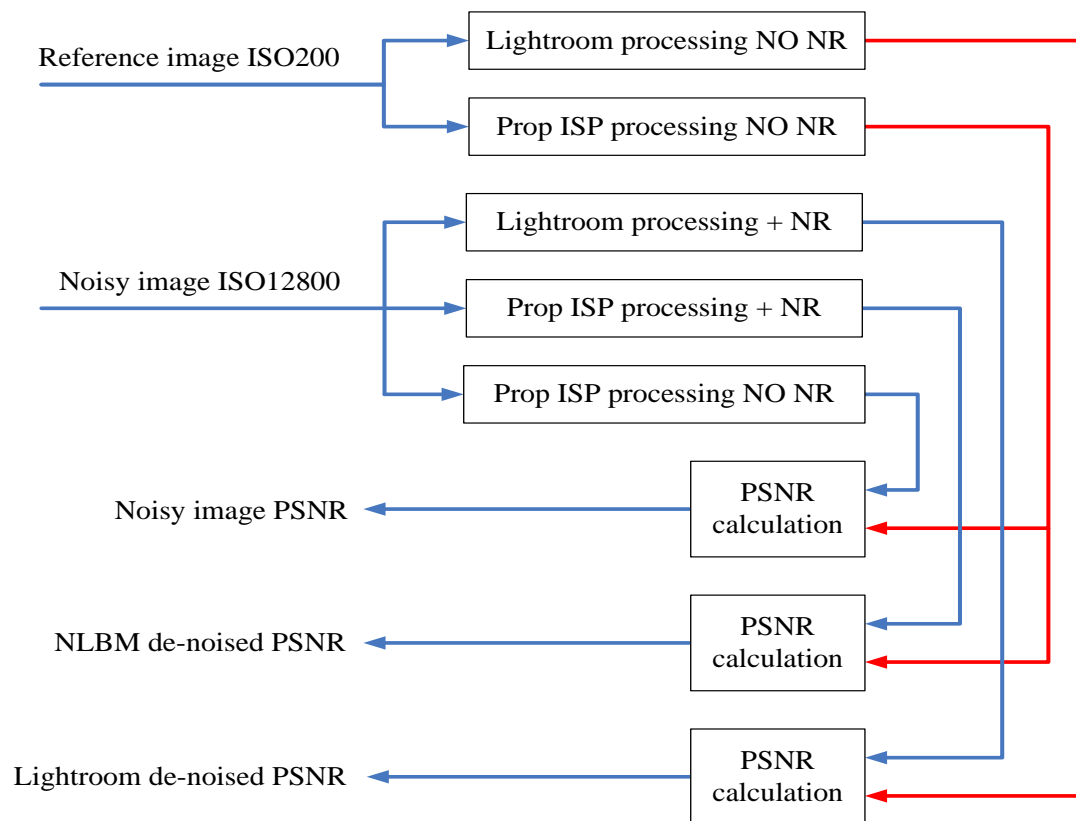


Figure 13: Block diagram of the real-world test procedure

In the experiments conducted the PSNR of an image when processed by the Lightroom image processing pipeline was calculated. Using an image of ISO12800 (i.e. the original image), the proposed NLBM approach is compared to a Lightroom image processing pipeline processed image (see Figure 14). The corresponding PSNR values are also indicated.

Chapter 3: Block matching de-noising method for photographic images, applied in Bayer RAW domain, optimized for real-time implementation



Figure 14: Real world test results.

The average PSNR values for R, G, and B colour channels were calculated, as seen from the previous set of tests. It is observed that the numbers across colour channels are not very different. It is also worth mentioning that we used the ISO200 reference image produced by our ISP to calculate the effect of de-noising of our algorithm and ISO200 reference produced by Lightroom to calculate the effect of de-noising of Lightroom as shown in a Figure 14.

3.4. Conclusion

In this chapter, a robust and a fast non-local de-noising algorithm has been proposed. The algorithm is based on a Laplacian pyramid and a modified non-local means noise reduction filter. The Laplacian pyramid is used to break up a noisy image into band-pass images. By performing a modified non-local means noise reduction algorithm on different levels of the Laplacian pyramid, with different sizes of comparison windows, both high and low-frequency noise are effectively removed, while preserving the image details (edges, textures, etc.) and keeping algorithm complexity low.

The proposed algorithm was implemented as a hardware block and used in Apical IPP. The results of algorithm implementation on Altera FPGA and in ASIC, using 65nm TSMC technology libraries are presented in Table 2:

	FPGA	ASIC
Logic elements (gate-count)	117K	750K
Effective kernel size	31x31	31x31
Number of scales	2	2
Multipliers	240	(included in gate-count)
Pixel clock frequency	150MHZ	350MHZ
Video performance	1080p 60fps	4k camera 60fps
Device (silicon area)	Altera FPGA EP4C150	1.12 mm ² using 65nm process.

Table 2: Spatial noise reduction synthesis results

In a hardware implementation of the proposed algorithm on an Altera FPGA EP3C120 a two layer Laplacian pyramid with an effective kernel size of 31×31 pixels and a bit-depth (precision of algorithm) 12 bits was used. It was possible to achieve a processing speed of 150Mpix/sec, which is sufficient to process HD video at 60 frames per second.

It is important to mention that due to the multi-scale architecture, the effective kernel of the proposed algorithm, implemented in hardware, can be increased to a very large size e.g. 127×127 at a very small increase in gate count.

It was shown that the proposed noise reduction algorithm shows improvements over many well-known noise reduction algorithms. The proposed algorithm is compared against Adobe Lightroom and BM3D. In most situations the proposed algorithm shows an advantage over competitive algorithms in PSNR, noise structure (spectral characteristics) and the natural look of images.

Chapter 4

A Spatio-Temporal noise reduction method optimized for real-time implementation

4.1. Introduction

In any imaging system producing still or motion pictures, noise reduction is a very important component, which defines the resulting image quality of the image processing pipeline. Though noise reduction techniques were known for many years, practically their use in consumer electronics, video surveillance, professional photo and video applications is constrained and therefore rather limited.

In this work an attempt is made to solve the complexity and performance issues with an optimized implementation of a practical spatial-temporal de-noising algorithm. Spatial-temporal filtering was performed in Bayer RAW data space, which enabled to benefit from predictable sensor noise characteristics and reduce memory bandwidth requirements. The proposed algorithm efficiently removes different kinds of noise in a wide range of signal to noise ratios. In our algorithm the local motion compensation is performed in Bayer RAW data space, while preserving the resolution and effectively improving the signal to noise ratios of moving objects.

The main challenge for the use of spatial-temporal noise reduction algorithms to de-noise video sequences is the compromise between the quality of the motion prediction and the complexity of the algorithm and required memory bandwidth. In photo and video applications it is very important that moving objects should stay sharp, while the noise is efficiently removed in both the static background and moving objects. Another important situation is when the background is non-static as well as the foreground where objects are moving.

The original aim of the proposed research is to combine block matching, block accumulation filters investigated in [3],[4],[6] and temporal noise reduction based on Gaussian background modelling described in [26],[27] to de-noise photographic and video images in RAW data

space, using sensor noise modelling investigated in [12] and covered in Chapter 2. The purpose of using block matching block accumulation filters was not just to do filtering in spatial domain, but also to find the best match between the current image of the video sequence and the accumulated image, thus performing the task of local motion compensation, to minimize the temporal difference. However there is a difficulty of matching current video data with accumulated video data in Bayer RAW data space. The Bayer pattern of modern RGB sensors has a structure with 2 pixel period, which means that simple matching of repetitive patterns may lead to loss of image details. In order to address this issue the block matching algorithm was modified to perform block matching of local neighbourhoods of red and blue pixels differently, from block matching within local neighbourhoods of green pixels. The proposed modification of the block matching technique seems viable as in most de-mosaic algorithms the green color planes normally have higher priority for details interpolation and, hence full precision in green color plane motion compensation is absolutely required, while the impact of reduced precision of motion compensation in red and blue color planes does not produce any significant degradation of the image quality during the de-mosaic interpolation. Another important modification to the block matching algorithm, that has been made, was concerned to the use of the non-linear weight filter. In Chapter 3 the use of a non-linear weight filter was proposed to reduce the amount of residual noise and grain. In our current work the logic of this filter has been altered to minimize the number of matches between accumulated image and the current image in a video sequence. These modifications made to the non-linear weight filter helped to maintain better sharpness in the output image.

There are a number of advantages in performing the processing in RAW data domain. One of the advantages is to have predictable noise characteristics, thus allowing making the decision about noise levels easier and more reliable. Another advantage is that the signal to noise ratios are greatly improved in the front end of the image processing pipeline, allowing other blocks to work more precisely and reliably with cleaner image data.

The evaluation of existing CMOS sensors shows that the kernel size required for efficient noise reduction should not be smaller than approximately 31×31 pixels for full HD sensors. On a large sensors (e.g. 12 megapixels and above) the kernels of the size of 127×127 pixels are absolutely necessary, while even a 31×31 pixel size kernel is considered to be relatively large for typical image processing pipeline. On the other hand temporal noise reduction in most cases allows reducing the kernel size down to 15×15 for full HD sensors.

Because of the kernel size constraints, memory bandwidth limitations and the requirement to implement the algorithm in hardware, it is highly desirable to perform processing in sensor RAW data space. Our algorithm was implemented in a camera image processing pipeline to process both video and still images. The block diagram of the proposed non-local means algorithm with a kernel of size of 15×15 pixels and Gaussian background modelling temporal filter is represented in Figure 15:

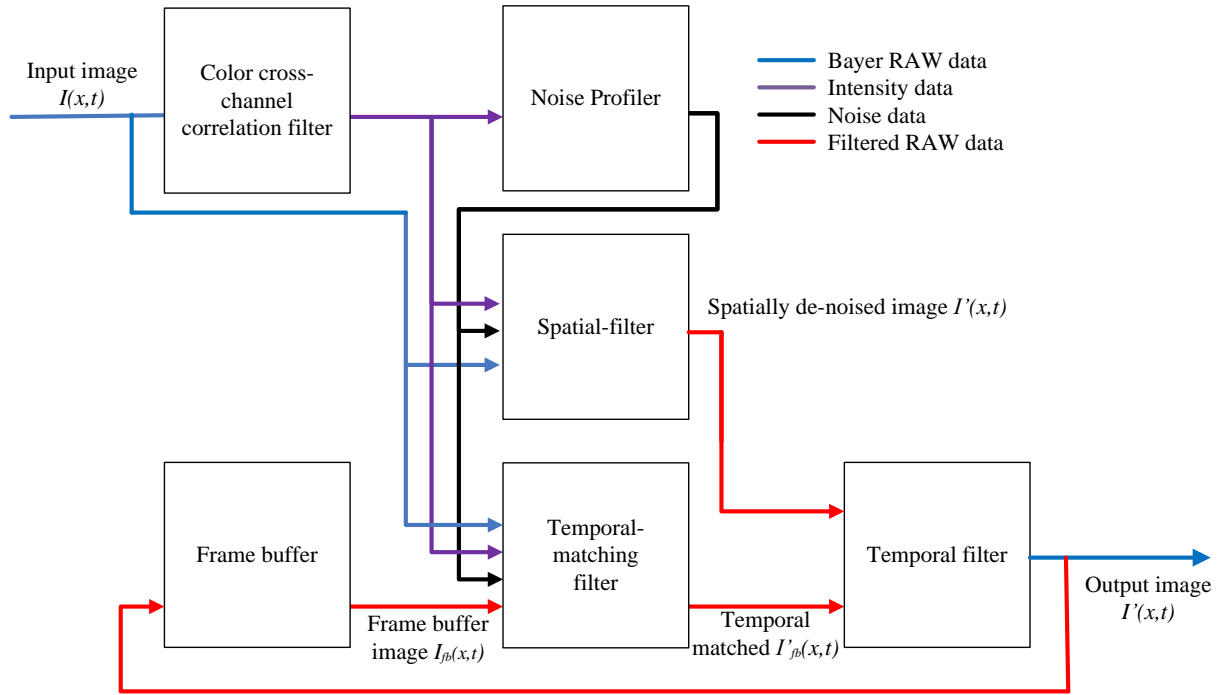


Figure 15: Algorithm block diagram.

The first step in our processing is to calculate the intensity for each pixel, including the information about colour cross-channel correlation. Further as mentioned previously our algorithm will work in RAW data space. For the sensors with RGB Bayer pattern the intensity calculation is important for two reasons: equal decision regarding filtering should be made for any colour when pixels of different colours belonging to the same object detail. Calculating intensity also helps to obtain an image with lower noise by eliminating colour cross-channel noise, thus helping to detect image details more reliably. Intensity calculation also allows us to avoid a checker pattern appearance in the resulting RGB image, unlike in some existing algorithms [3].

The second step in our proposed algorithm is to calculate noise characteristics. This calculation is performed for each pixel according to equation (2). In parallel with intensity and noise profile calculations the image is passed through the spatial filter to get rid of some noise, especially in a high frequency band. At the same time the original image and the data from the frame buffer passed to temporal matching block. The result of filtering along with data from temporal matching block then sent to temporal filter GM, which performs temporal filtering by computing whether image data belongs to the Gaussian background model or needs to be updated from the current video data. The result is stored in a frame buffer. The result of temporal filtering is sent to the output and stored in the frame buffer with updated variance estimation. In the research presented here we put $I(t-1)$ as the image captured at discrete time $t-1$, $I(t)$ as the image captured at discrete time t , $I'(t)$ as the image transformed to match the image $I(t)$. The variance of temporal noise of image $I'(t)$ is denoted as $\text{var}(I'(t))$. The variance of temporal noise of image $I_{fb}(t)$ is denoted as $\text{var}(I_{fb}(t))$. The operational block diagram of the proposed Spatial-Temporal filter presented in Figure 16:

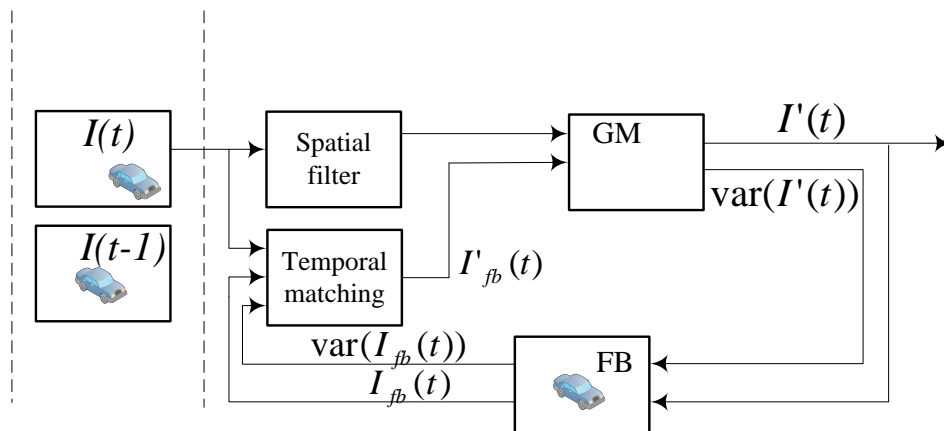


Figure 16: Spatial-Temporal filter block diagram

In the figure above $I'(t)$ has a meaning of temporal average of predicted image, i.e. obtained in a result of temporal matching and accumulation. The use of Gaussian background model enables recursion-like data accumulation, reducing the requirements for the memory bandwidth drastically. Comparing the requirements for memory bandwidth of the proposed algorithm with algorithm described in [25], it is likely that the memory bandwidth in the proposed algorithm is 5 times less, considering that in the proposed algorithm the image data has 12 bit precision and

variance estimation use 4 bit quantization, while in [25] filtering is applied to 8 bit YUV data and the matching is performed for 4 pairs of images.

4.2. Block matching approach for Bayer RGB sensors

Non-local means approach used in inter-frame data accumulation is a known technique described in [25], but in this research this method was applied for Bayer sensor data. The idea of Bayer RAW data filtering using non-local means technique was described in [3], however in our work we included several important improvements. In the proposed research two methods were developed to set appropriate thresholds for each block accumulation dynamically, in contrast to other researchers, who used a fixed threshold value. The sensor noise was modelled to get the estimate of noise levels, to calculate the weights for block averaging (see Chapter 2). Further a method of non-linear data analysis to estimate the details energy has been developed, leading to prediction of correlation values, which allowed us to guarantee details preservation while at the same time having increased strength of filtering on image parts where no image details can be correlated. In a discussion related to the image matching, using block matching algorithm, it is noted that both $I(v, t)$ and $I_{fb}(v, t)$ correspond to the same discrete time t , so the variable t will be omitted. Let us put $N(v)$ to be a neighbourhood of a noisy image $I(v)$ at coordinate v . In the proposed algorithm a limited search range $S(v)$ of size of $s \times s$ is used. Thus the neighbourhood $N(v_j) \in S(v)$ of image $I(v)$ can be averaged with weights $w(v, v_j)$ to produce a de-noised image, $I'(v)$ according to the equation (27), (see Chapter 2, equation (5)):

$$I'(v) = \frac{\sum_{v_j \in S(v)} w(v, v_j) I(v_j)}{\sum_{v_j \in S(v)} w(v, v_j)} \quad (27)$$

In case of temporal matching we would need to match the image data stored in a frame buffer $I_{fb}(v)$ and current image $I(v)$. Frame buffer data, matching current de-noised image obtained according to equation (28) (refer Chapter 2, equation (6)):

$$I'_{fb}(v) = \frac{\sum_{v_j \in S(v)} w_{fb}(v, v_j) I_{fb}(v_j)}{\sum_{v_j \in S(v)} w_{fb}(v, v_j)} \quad (28)$$

In case of intra frame accumulation, the weights $w(v, v_j)$ are calculated from block differences $d(v, v_j)$ for corresponding neighbourhoods $N(v)$ and $N(v_j) \in S(v)$ of pixels of image $I(v)$. The block differences $d(v, v_j)$ are given by the equation (29):

$$d(v, v_j) = \sum_{v_j \in S(v)} |N(v) - N(v_j)| \quad (29)$$

In case of inter-frame accumulation, the weights $w_{fb}(v, v_j)$ are calculated from block differences $d_{fb}(v, v_j)$ for corresponding neighbourhoods $N(v)$ of image $I(v)$ and $N_{fb}(v_j) \in S(v)$ of pixels of image $I_{fb}(v)$ with coordinates v and v_j respectively:

$$d_{fb}(v, v_j) = \sum_{v_j \in S(v)} |N(v) - N_{fb}(v_j)| \quad (30)$$

Each block difference $d_{fb}(v, v_j)$ will be constructed of a sum of noise differences $d_\sigma(v, v_j)$ and image autocorrelation functions $d_\alpha(v, v_j)$. In our algorithm we do not attempt to rotate or scale blocks to achieve better correlation, thus, we need to estimate the value of $d_\alpha(v, v_j)$ to adjust $w_{fb}(v, v_j)$ calculation. In our implementation we estimating $d_\alpha(v, v_j)$ by evaluation of the block difference $d_{fb}(v, v_j)$ and expected noise level $\sigma(v)$ adjusted for the image data according to the equation (2). In spatial filtering block, for each block we calculate the estimation of $d_\sigma(v, v_j)$ according to equation (31):

$$d_\sigma(v, v_j) = d_{fb}(v, v_j) - d_\alpha(v, v_j) \quad (31)$$

In the temporal matching algorithm it is more important to measure $d_\alpha(v, v_j)$, as it will represent the similarity between the best matching block in frame buffer image and the current image. The estimation of $d_\alpha(v, v_j)$ can be calculated according to the equation (32):

$$d_\alpha(v, v_j) = d_{fb}(v, v_j) - d_\sigma(v, v_j) \quad (32)$$

Resulting $d_\alpha(v, v_j)$ is consequently converted into corresponding $w_{fb}(v, v_j)$ by comparison with the expected variation of noise, derived from the sensor characterization and specified in equation (2). The result of averaging $I_{fb}'(v)$ in each pixel location will be obtained as the result

of averaging of data $I_{fb}(v)$ within the search area $S(v)$ with weights $w_{fb}(v, v_j)$. As discussed in Chapter 2 the number of weights $w_{fb}(v, v_j)$ and block differences $d_{fb}(v, v_j)$ associated with each pixel location will be $s \times s$, where s is a size of search area S . In this work, block differences $d_{fb}(v, v_j)$ are calculated in linear Bayer RAW data domain, this fact allows to use the estimation of a noise level $\sigma(v)$ as a reference. In this research we propose to normalize the block differences according to equation (33) and (34):

$$d_{\alpha}(v, v_j)_{norm} = d_{fb}(v, v_j)_{norm} - d_{\sigma}(v, v_j)_{norm} \quad (33)$$

$$d_{\alpha}(v, v_j)_{norm} = d_{fb}(v, v_j) / \sigma(v) - d_{\sigma}(v, v_j) / \sigma(v) \quad (34)$$

The equation (34) provides the estimation of block matching $d_{\alpha}(v, v_j)_{norm}$ and based on directly measured $d_{fb}(v, v_j)$ and predicted $d_{\sigma}(v, v_j)$. In this work $d_{\sigma}(v, v_j)$ can be put $d_{\sigma}(v, v_j) \cong \sigma(v)$. In this case the value of normalized block difference, reflecting the similarity of the blocks can be expressed as:

$$d_{\alpha}(v, v_j)_{norm} \cong d_{fb}(v, v_j) / \sigma(v) - 1 \quad (35)$$

The conversion of normalized $d_{\alpha}(v, v_j)_{norm}$ into $w_{fb}(v, v_j)$ is performed according the equation (36):

$$w_{fb}(v, v_j) = \exp\left(-\left(\frac{d_{\alpha}(v, v_j)_{norm}}{Csigma}\right)^2\right) \quad (36)$$

In the proposed algorithm the parameter of $Csigma$ is used as a threshold for a block accumulation and normally set to 1, the best matching blocks will be taken from $I_{fb}(v)$ and accumulated with higher weight. The algorithm presented in this research automatically adopted to different sensors and lighting conditions, given that the sensor noise model was provided and was correct. In the proposed algorithm matching $I_{fb}'(v)$ and $I(v)$ are used by Gaussian background model, which performs temporal accumulation of data, and consequently reduce the amount of noise in accumulated image.

4.3. Temporal data accumulation using Gaussian background model

In our system we use temporal filtering to reduce the variations of image data from frame to frame over time. In practice temporal variations can produce a very large scale noise, thus removal of temporal noise can reduce requirements to the maximum filtering kernel size. In our research we are assuming that the temporal differences have a distribution with variance $\sigma(v, t)$, described in Chapter 2, equation (2). In this research it is assumed that moving objects found in video sequence will consistently keep moving in the same direction at least for some period of time. Though we will attempt to minimize temporal difference by performing temporal matching, we expect that temporal differences will be increased in the areas corresponding to the moving objects. Let us assume that the original image $I(v, t)$ and accumulated frame buffer image $I_{fb}(v, t)$ are the functions of discrete time t and coordinate $v=(x, y)$, these two signals are the inputs of our temporal filter, further the coordinate v will be omitted for the compactness of the formulas. The formula for the temporal filtering and the recursive coefficient was derived in Chapter 2, equations (15), (16):

$$\hat{I}(t) = I_{fb}(t-1) \times (1 - K(t)) + I(t) \times K(t) \quad (37)$$

In our work we found $K(t)$ as the coefficient optimal for the Kalman filter to be defined as:

$$K(t) = K(t-1) \times \left(1 - \frac{1}{\sigma_I}\right) + D(t) \times \frac{1}{\sigma_I} \quad (38)$$

Parameter σ_I is proportional to the standard deviation of pixel intensity and is defined as a constant for each system. When both $K(t)$ and $\hat{I}(t)$ are calculated they are saved in the frame buffer to be subsequently used for the analysis of the next frame. In the proposed system $K(t)$ is quantized to 4 bit precision data, for the equation (38) to work we defined $K(t)$ calculations via lookup tables. Proposed temporal data accumulation method provided an efficient way of reducing noise in video image sequences and allowed to minimise the appearance of motion blur artefacts.

4.4. Experimental results

In the experiments conducted a custom camera system using image sensor AS3372 (see Appendix B) which provided access to raw data was used. Integration time was set to 1/60 sec. For outdoor scenes the lens aperture was set to F2.0, whereas for indoors scenes the lens aperture was set to F8.0. In order to expose the image correctly, the sensor gain was programmed to 30dB. The videos were processed through the full image processing pipeline. For comparison purposes we have compared our proposed algorithm and VBM3D at the same operational place in the image processing pipeline, making sure that the processing is done using the same image processing pipeline settings. For the VBM3D integration we have used the method described in [3]. See examples in Figure 17, Figure 18, Figure 19 and Figure 20 below. The effect of motion compensation can be seen in Figure 19 and Figure 20 below:



Figure 17: Experimental results

Experimental results illustrate that the proposed algorithm is able to remove noise efficiently in an image with static background and a moving foreground object. The appearance of residual

noise in (b) is smooth and very similar to one found in an images captured with low gain settings (corresponds to ISO-100), which looks aesthetically and more pleasing than the result in (c). Another example is presented in Figure 18:

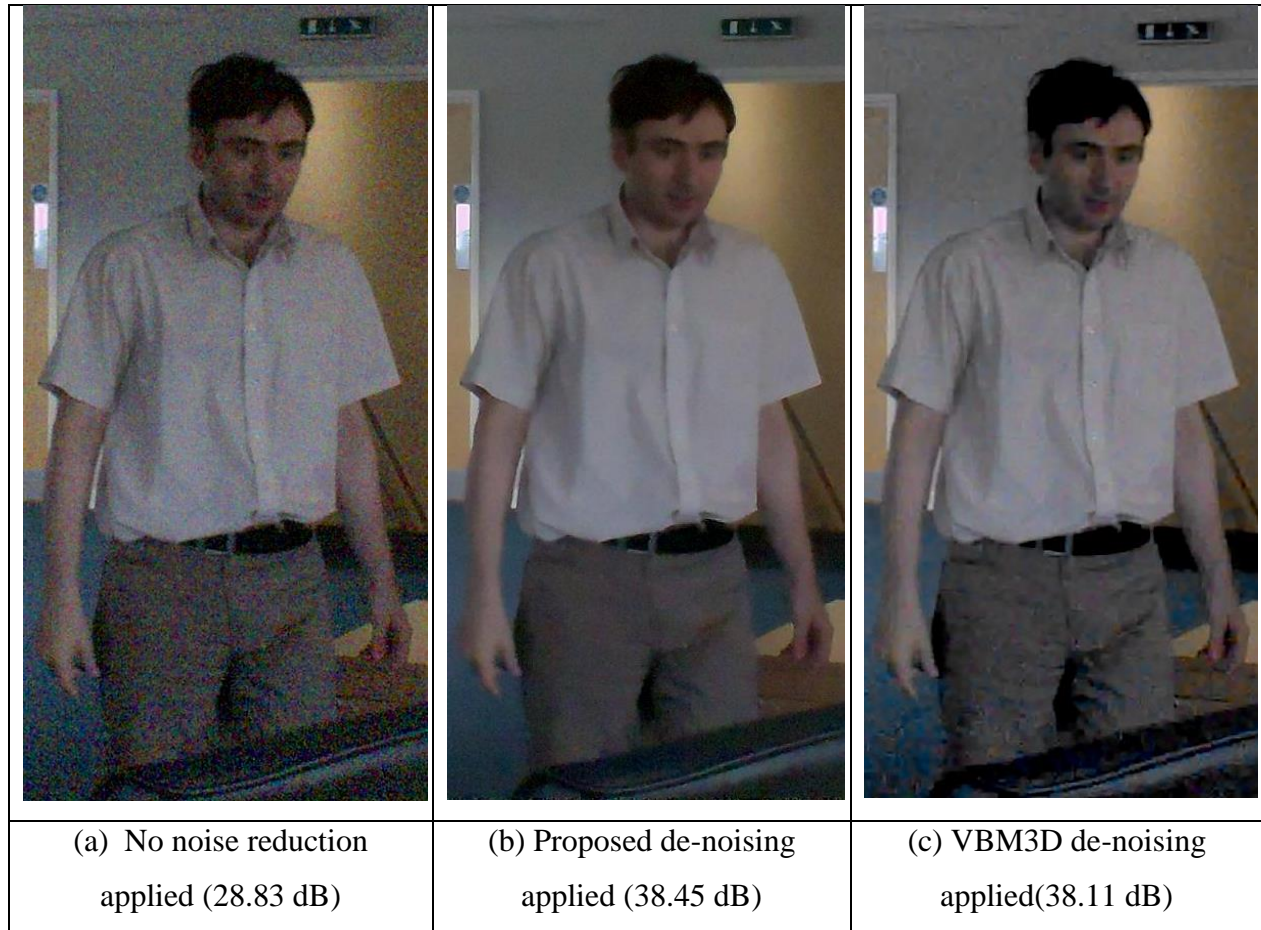


Figure 18: Experimental results

In Figure 18 it is observed that the colour noise suppression as well as the suppression of large scale noise is more efficient in (b) than in (c). Though the contrast in (c) is higher than in (b), it is worth mentioning that pixel levels and contrast in (b) is close to that of the original image (a). In the comparison conducted the executable model of VBM3D was used and an attempt was made to match noise and details. As VBM3D model was used as a black box it was not possible to match outputs precisely, though it is believed that the results achieved are good enough for a valid comparison.

	<p>(a) No noise reduction applied (21.83)</p>
	<p>(b) VBM3D noise reduction applied (32.48 dB)</p>
	<p>(c) NLBM3D noise reduction applied with motion compensation enabled (32.16 dB)</p>
	<p>(d) NLBM3D noise reduction applied with motion compensation disabled (32.78 dB)</p>

Figure 19: The effect of motion compensation

In Figure 19 it is demonstrated that in extremely noisy video the best results are produced by the proposed algorithm. Further using the block matching filter for local motion compensation helps to suppress noise better and recover more details (see (b) and (d)).



Figure 20: Experimental results, Motion Compensation evaluation.

As seen in Figure 20 (b) and (c), the motion compensation has a significant effect on the appearance of temporal ghosting artefacts, which can be seen around the hand and leg regions in Figure 20(c). In Figure 20 a sequence, was captured with a handheld camera, in which the background is not static. It can also be noticed that the local motion compensation improves the sharpness of background image.

Since it is not possible to use the ground truth image, for the purpose of estimating the noise levels, areas within the static background were used (see Figure 21 below):

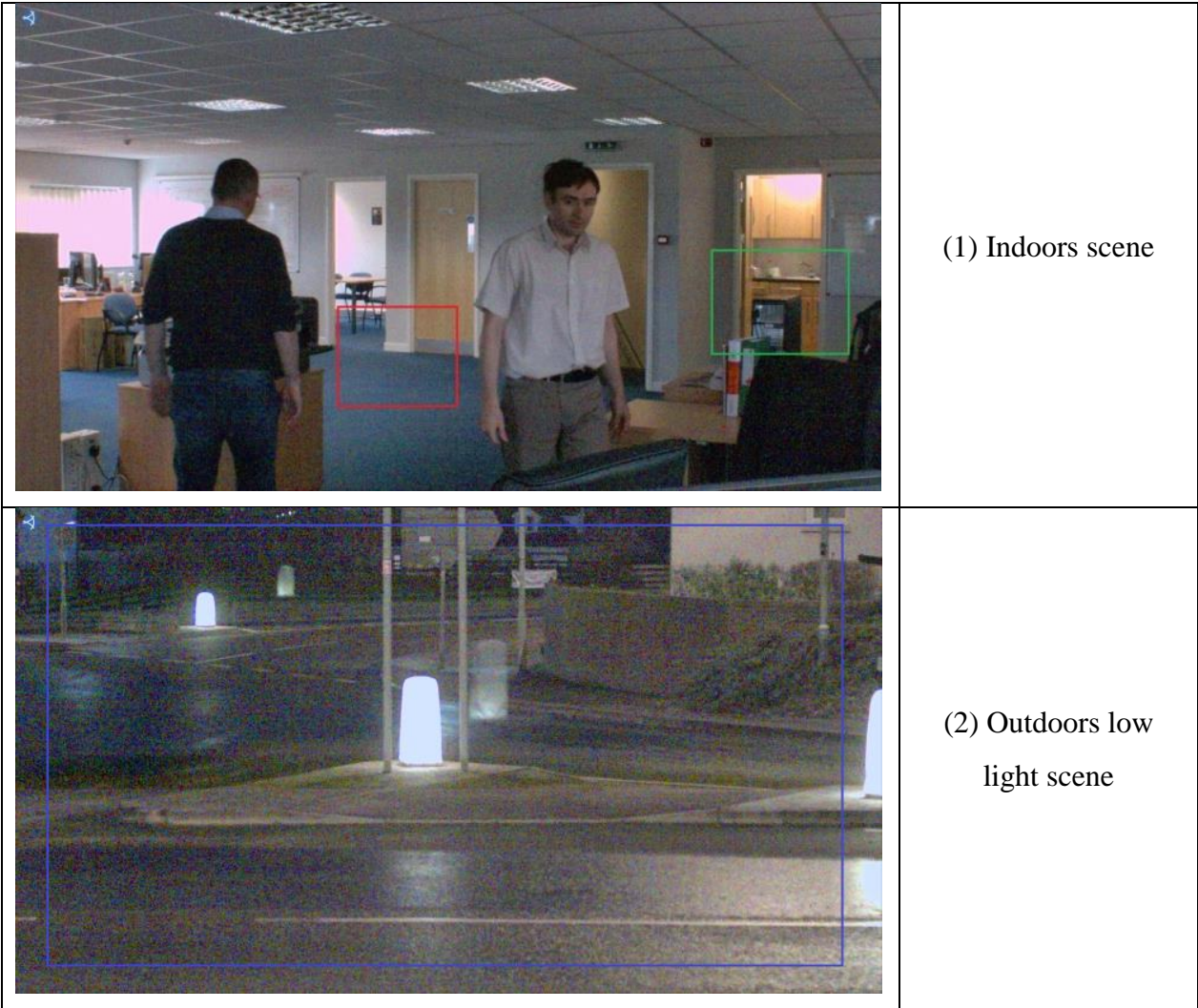


Figure 21: Ground truth images.

In selected areas the average PSNR values have been calculated for temporal variations over a number of frames. The results are presented in Table 3:

Scene (1) - Red Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
VBM3D $\sigma=16$	37.3881	42.4937	38.3489	39.4102
NLBM3D	39.6227	40.5834	39.3227	39.8429
No NR applied	25.4603	27.1247	26.7285	26.4378
Scene (1) - Green Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
VBM3D $\sigma=16$	38.2573	39.9474	36.1338	38.1129
NLBM3D	38.5209	39.1383	37.6942	38.4511
No NR applied	28.8421	29.3311	28.3259	28.8330
Scene (2) - Blue Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
VBM3D $\sigma=16$	38.8525	39.4941	37.7438	38.6968
NLBM3D	41.9110	42.1838	40.2060	41.4336
No NR applied	29.7782	29.5214	26.8559	28.7185

Table 3: PSNR values comparison table.

Settings for de-noising algorithms were chosen to produce a substantial amount of noise reduction, while preserving similar amount of details. However it can be seen that the efficiency of VBM3D algorithm reduces as the noise level increases, unlike in the proposed algorithm where details and noise suppression efficiency are maintained better at higher sensor gains (see Figure 17 and Figure 18 for details).

4.5. Conclusion

In this chapter a robust and efficient spatio-temporal de-noising algorithm was proposed. Due to a number of algorithmic optimizations, the proposed algorithm, when implemented in hardware can be compact and the memory bandwidth requirements can be reduced, compared to the spatio-temporal noise reduction algorithm described in [25]. The proposed algorithm was implemented on an Altera FPGA EP3C120 and synthesized for ASIC chip, using 65nm TSMC technology library. Synthesis figures are presented in Table 4:

	FPGA	ASIC
Logic elements (gate-count)	105K	680K
Effective kernel size	17x17	17x17
Number of scales	1	1
Multipliers	190	(included in gate-count)
Pixel clock frequency	150MHZ	350MHZ
Video performance	1080p 60fps	4k camera 30fps
Device (silicon area)	Altera FPGA EP4C150	0.96 mm ² using 65nm process.

Table 4: Spatio-Temporal noise reduction block implementation details

Using a commercial grade FPGA it is possible to achieve a processing speed of 150Mpix/sec, which is sufficient to process HD video at 60 frames per second. ASIC implementation can perform more than two times faster and able to process 4k video resolutions in real time. The proposed algorithm was compared with the VBM3D algorithm. In most situations the proposed algorithm shows an advantage over competitive algorithms in the efficiency of noise reduction, noise structure and the natural look of images. The efficiency of noise reduction in proposed algorithm will reduce when the object moving fast and the displacement is greater than the size of a search window S . The possibility to use the proposed algorithm for frame accumulation of photographic image, where large object displacement is possible, is investigated in a next Chapter 5.

Chapter 5

Image Matching in Bayer RAW Domain to De-noise Low-light Still Images, Optimized for Real-Time Implementation.

5.1. Introduction

Noise reduction is a very important component, which defines the resulting image quality of the IPP. Efficient and robust spatial-temporal noise reduction algorithm is especially important for cameras with small sensor and poor optics, with limited light capturing capabilities. Though noise reduction techniques were known for many years, practically their use in consumer electronics, video surveillance, professional photo and video applications is constrained and therefore rather limited.

Temporal accumulation is a known approach to improve signal to noise ratios of still images taken in low light conditions [25]. However the complexity of known algorithms often lead to high hardware resource usage, memory bandwidth and increased computational complexity, making their practical use impossible. In the proposed research an attempt is made to solve this problem with an implementation of a practical spatial-temporal de-noising algorithm, based on image accumulation. Image matching and spatial-temporal filtering was performed in Bayer RAW data space, which allowed one to benefit from predictable sensor noise characteristics. This enables the use of a range of algorithmic optimisations. Proposed algorithm accurately compensates for global and local motion and efficiently removes different kinds of noise in noisy images taken in low light conditions. Global and local motion compensation are conducted in the Bayer RAW data space, while preserving the resolution and effectively improving signal to noise ratios of moving objects. Proposed algorithm is suitable for implementation in commercial grade FPGA's and capable of processing 12MP images at capturing rate (10 frames per second).

The main challenge for still images matching is the compromise between the quality of the motion prediction and the complexity of the algorithm and required memory bandwidth. Still images taken in a burst sequence must be aligned to compensate for background motion and foreground objects movements in a scene. A high resolution of still images as well as significant time between successive frames produce significant displacements of the parts of an image and creates additional difficulty for image matching algorithms. In photographic applications it is very important that the noise is efficiently removed in both static backgrounds and moving objects and the resolution of the image is maintained. In the proposed algorithm the issue of matching the current image with the accumulated image data in Bayer RAW data space is resolved in order to efficiently perform the Spatial-Temporal noise reduction. In this chapter the proposed algorithm is compared with the state of the art noise reduction algorithms and subjective experimental results are provided to demonstrate the ability of the proposed method to match noisy still images in order to perform efficient de-noising and avoid motion artefacts in resulting still images.

The idea of accumulation of images taken in a burst sequence is not new. However there are a number of difficulties, preventing this method to be used in industry. In practice there are no spatial-temporal frame accumulation algorithms able to deliver acceptable image quality at reasonable cost of implementation. First of all, images are taken at very high resolutions, which automatically mean that the time interval between subsequent captures is significant. In the experiments within the research context of this thesis capture rates of 7-10 images per second were used at resolutions 8-16MP. It can be expected that during that interval the whole scene composition can significantly change. The experiments revealed that the parts of the scene can move by as much as 512 pixels. This means that the first step of processing should be motion estimation and compensation. Another objective difficulty is that the lighting conditions may also change between frames. This can be due to the environmental changes or even indoors in controlled light conditions. When the scene is lit by an artificial light source, the camera can produce a significant variation of image brightness and colour due to the interference between the light and the shutter. Considering the scene variation between successive frames a conclusion was reached that the motion estimation has to rely on the data, invariant to image brightness, rotation and scale variations. Another requirement for motion estimation and compensation is that it should be done in the Bayer RAW space, as the previously developed Spatio-Temporal noise reduction technique is to be used. The aim of this work is to develop

frame accumulation algorithm delivering good image quality and compact, when implemented in hardware to be suitable for practical use. The idea of the proposed research is to perform coarse motion estimation and compensation in Bayer Raw domain and use the pre-matched images as input data for the previously suggested Spatio-Temporal noise reduction algorithm (see Chapter 4). This algorithm was constructed as a combination of block matching, block accumulation described in [4],[8],[9] and temporal noise reduction filters based on Gaussian background modelling, investigated in [17],[18] to de-noise photographic and video images in RAW data space, using sensor noise modelling according to [8],[9].

There are a number of advantages in performing the processing in RAW data domain. One of the advantages is to have predictable noise characteristics, thus allowing making the decision about noise levels easier and more reliable. The other advantage is that the signal to noise ratios are greatly improved in the front end of the image processing pipeline, allowing other blocks to work more precisely and reliably with cleaner image data.

The block diagram of the proposed algorithm with non-local means filter and Gaussian background modelling temporal filter is represented in Figure 14:

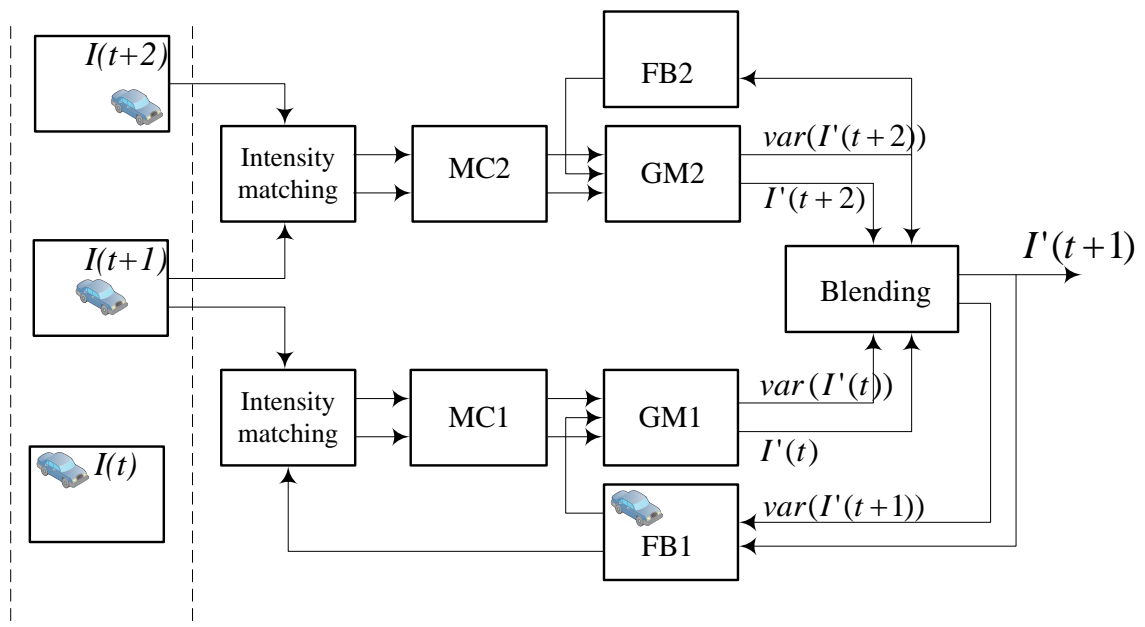


Figure 22: Algorithm block diagram.

Initially the sequence of images $I(t), I(t+1), I(t+2)...$ enters the algorithm. The first step in our processing is to match the intensity of input images, followed by motion compensation MC1 and MC2. Each motion compensation block MC is processing a pair of images: MC1 is compensating motion between the image $I(t+1)$ and the accumulated image $I_{fb}(t+1)$, which was stored in a frame buffer and corresponds to the image $I(t)$. MC2 compensates the motion between image $I(t+1)$ and $I(t+2)$. The implementation of bi-predictive scheme allowed the effective update of the accumulated image with objects that newly appear. A pairs of motion matched images are then processed through Gaussian Mixture temporal filters $GM1$ and $GM2$, which perform calculation of updated temporal mean $I'(t), I'(t+2)$ and variance $var(I'(t)), var(I'(t+2))$. Blending block chose the mean with minimum variance and updates $I'(t+1)$.

The use of Gaussian background model enables recursion-like data accumulation, reducing the requirements for the memory bandwidth drastically. Comparing the requirements for memory bandwidth of the proposed algorithm with the algorithm described in [25], it is likely that the memory bandwidth in the proposed algorithm is 4 times less, assuming that in our algorithm the data has a precision of 12 bit and the variance estimation has 4 bit quantization, while in [25] filtering is applied to 8 bit YUV data.

5.2. Robust optical flow

In the proposed algorithm a multi-scale sparse feature matching was adopted to perform coarse motion estimation. On each scale the image was transformed into a set of feature vectors, in order to achieve invariance to brightness change and improve robustness in the cases when objects in a scene rotate or change scale. In order to resolve the situation when a new object appears in a scene it is suggested to implement the bi-predictive scheme.

Modern optical flow estimation is usually posed as an energy minimization problem as investigated in [30],[29],[31],[33]. Let us consider two frames: $I_1(v)$ and $I_2(v)$ corresponding to the same scene, we denote two dimensional coordinate as $v = (x, y)$. In this research $I_1(v)$ and $I_2(v)$ represent an intensity of an image. Let us denote U as a flow field that represents the displacement vectors $u(v)$ between $I_1(v)$ and $I_2(v)$ for each pixel coordinate v , so that $u(v) \in U$. As the flow field is unknown we will attempt generate a set of U_n and evaluate them in order to choose the best one. The data term $E(v, U)$ can be defined as:

$$E_n(v, U_n) = \sum_{v_j \in S(v)} \alpha \|I_1(v_j + u_n(v_j)) - I_2(v_j)\| + \beta \|\nabla I_1(v_j + u_n(v_j)) - \nabla I_2(v_j)\| \quad (39)$$

Where ∇ is the gradient function, α and β are the weights balancing the costs of intensity matching and gradient matching and $S(v)$ is a spatial kernel of size $s \times s$ with centre at coordinate v . The matching field U is computed as the field U_n corresponding to minimum energy $E_n(v, U_n)$. The robustness of sparse feature matching multi-scale optical flow can be improved by introducing the estimation of data term remainder $R(v)$, which will bias the estimation of energy due to the noise factor. The definition of the remainder $R(v)$ is the energy estimation $E_n(v, U_n)$ obtained for the set of static image, where the ground truth translation field $U_n = 0$. It is proposed to implement the optical flow calculation in linear Bayer RAW space, which allows us to estimate the noise variation of $I(v)$ for every pixel location and, thus estimate the data term remainder $R(v)$ for the set of $E_n(v, U_n)$. Even for a sequence of images corresponding to a static scene the data term can be non-zero in the presence of noise. Furthermore the data term calculated for low contrast parts of an image can easily produce false minimums. It can be noted that the remainder $R(v)$ as it is defined for static image capture:

$$R(v) = \sum_{v_j \in S(v)} \alpha \|I_1(v_j + 0) - I_2(v_j)\| + \beta \|\nabla I_1(v_j + 0) - \nabla I_2(v_j)\| \quad (40)$$

The data term remainder can be represented as:

$$R(v) = \alpha \sum_{v_j \in S(v)} \|I_1(v_j) - I_2(v_j)\| + \beta \sum_{v_j \in S(v)} \|\nabla I_1(v_j) - \nabla I_2(v_j)\| \quad (41)$$

Considering that images $I_1(v)$ and $I_2(v)$ correspond to the same scene we can conclude that the remainder is actually represents the temporal variance of a random process of capturing images $I_1(v)$ and $I_2(v)$, $\nabla I_1(v)$ and $\nabla I_2(v)$. Thus the equation for the remainder can be reformulated as:

$$R(v) = \alpha \sigma(v) + \beta \sigma_g(v) \quad (42)$$

In this work $\sigma_g(v)$ considered to be proportional to $\sigma(v)$ as was proved in Chapter 2 for Laplacian operator. Thus we can finally define $R(v,0)$ as:

$$R(v) \cong \sigma(x) \quad (43)$$

When optical flow is calculated in gamma corrected image data domain it is believed that the different parts of an image have same noise characteristic. Although this is true to some extent and most cameras are designed to have constant noise characteristics. However when the ISO setting of a camera is increased, noise characteristics can vary significantly. Further it is worth mentioning that sensors with small pixel pitch, normally used in the mobile industry, do not satisfy uniform noise characteristics even at low ISO setting, being restricted to use a standard gamma. The use of standard gamma is required to match the inverse gamma of a standard display, where the image will be displayed. The noise characteristics of cameras using such sensors are significantly different from being uniform across the intensity range. However noise characteristics of a typical sensor remain constant, thus, being measured once, the sensor noise model can be used to normalize the response of a data term, making the search of a minimum energy precise and more reliable. As optical flow will be used for noise reduction applications, it has to be especially robust in the presence of significant amount of noise. In this research it is proposed to use the noise remainder $R(v)$ calculated locally to weight the local differences between $I_2(v_j)$ and $I_1(v_j + u_n(v_j))$. The following energy normalization scheme is proposed:

$$E_n^{norm}(v, U_n) = \sum_{v_j \in \mathcal{S}(v)} (\alpha \|I_1(v_j + u_n(v_j)) - I_2(v_j)\| + \beta \|\nabla I_1(v_j + u_n(v_j)) - \nabla I_2(v_j)\|) / R(v_j) \quad (44)$$

It can be noted that the proposed energy normalization method is similar to block difference normalisation. Effectively it normalizes the response from the local image features based on the measure of their reliability by comparison with predicted noise levels. In proposed experiments we prove that the introduction of noise remainder $R(v)$ improves the reliability of optical flow calculation to make its practical use feasible, thus improving the results of image matching. In practice the reliability of the proposed method made the algorithm suitable for the practical use. As the result, some of the experimental results provided without reference as the optical flow calculation without normalization technique does not meet minimal reliability expectations for motion field estimation. Multi-scale sparse feature matching in our algorithm is enabled by the Gaussian image pyramid discussed in Chapter 2. The use of Gaussian pyramid is explained by the ease of implementation, additionally research made in [25],[34],[35] proves that other than Gaussian multi-scale pyramid image representation does not provide significant advantage in motion estimation. Features for matching are calculated on each scale of a pyramid. The process of feature pyramid calculation is illustrated in Figure 23:

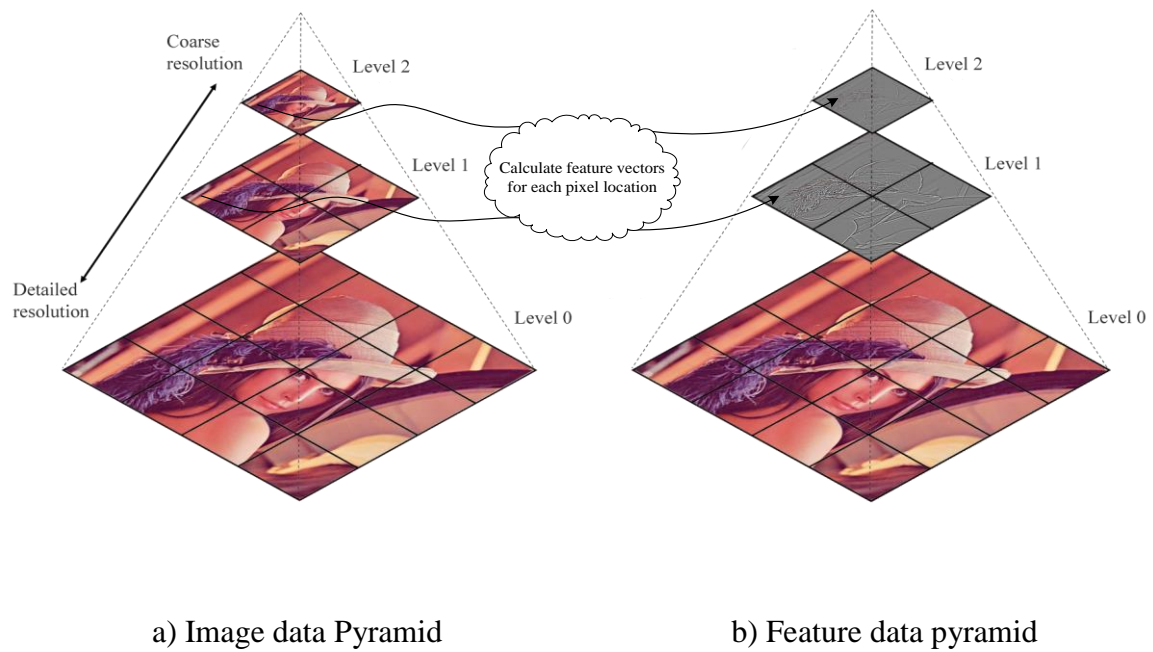


Figure 23: Image scale pyramid.

The search of displacement vectors between pairs of feature scales is performed by multi-scale sparse feature matching. On a most detailed scale, image matching is performed by a pixel mapping block. The process of motion compensation in an image pyramid is illustrated in Figure 24:

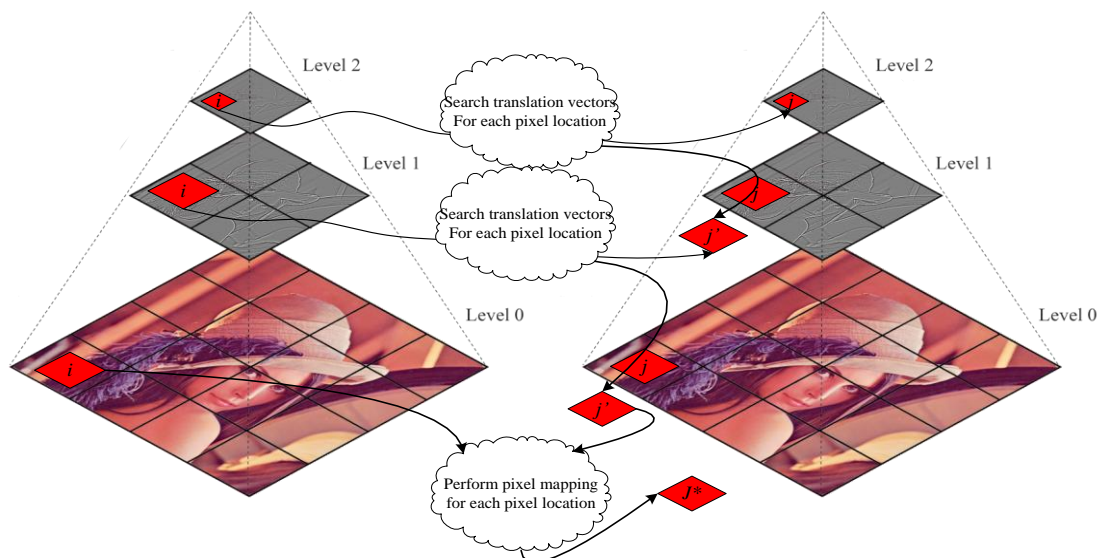


Figure 24: Multi-scale optical flow calculation

5.3. Block matching approach for Bayer RGB sensors

Non-local means is a known technique [3],[18],[19]. However within the research context of this thesis an attempt is made to apply this method for Bayer sensor data. In the non-local means technique a measure of similarity between pixel neighbourhoods $N(v)$ and $N_{fb}(v_j)$ will be calculated and then will be compared against the noise levels and the decision will be made, whether to use that neighbourhood $N_{fb}(v_j)$, to average with the current pixel neighbourhood $N(v)$ or not, see Chapter 2. The idea of RAW data filtering using non-local means technique was described in [3]. However in the proposed work several important improvements are introduced. Two methods are developed to set appropriate thresholds for each block accumulation dynamically, in contrast to methods proposed in literature that use a fixed threshold value. Sensor noise is modelled to obtain an estimate of noise levels, to calculate the weights for block averaging (see Chapter 2). Further a method of non-linear data analysis has been developed to estimate the details of energy, leading to the prediction of correlation values, which allows the guaranteeing of preservation of detail while at the same time having increased strength of filtering on image parts where no details can be found. In the proposed algorithm a block matching algorithm, previously suggested in Chapter 4 is used. For temporal matching it is needed to match the image data stored in the frame buffer $I_{fb}(v)$ and the current image $I(v)$. It is important to mention that the block matching algorithm developed previously allows accurate image matching between image $I(v)$ and de-noised accumulated image $I_{fb}(v)$ stored in a frame buffer, even in situations when correlation between images is difficult to find. Assume $N(v)$ to be a neighbourhood of a pixel with coordinate x of a noisy image $I(v)$. In the proposed algorithm a limited search area $S(v)$ of size of $s \times s$ is used. Thus the neighbourhood $N_{fb}(v_j) \in S_{fb}(v)$ of image $I_{fb}(v)$ can be averaged with weights $w_{fb}(v, v_j)$ to produce a de-noised image $I'_{fb}(v)$ according to the equation (45) (refer Chapter 2, equation (6)):

$$I'_{fb}(v) = \frac{\sum_{v_j \in S(v)} w_{fb}(x, x_j) I_{fb}(v_j)}{\sum_{v_j \in S(v)} w_{fb}(v, v_j)} \quad (45)$$

The process of image data matching between the frame buffer $I_{fb}(x)$ and the current image $I(x)$ is referred in this research as pixel mapping process. The weights $w_{fb}(x, x_j)$ can be considered as a synthesized aperture, which will perform data interpolation to produce the best match between the frame buffer image data $I_{fb}(x)$ and the current image data $I(x)$ on a pixel and sub-pixel level. The process of pixel mapping is based on inter-frame image data matching and explained in Chapter 2, Figure 6, equation (6).

The weights $w_{fb}(x, x_j)$ are calculated from block differences $d_{fb}(x, x_j)$ for corresponding neighbourhoods $N(x)$ and $N_{fb}(x_j) \in S_{fb}(x)$ of pixels of images $I(x)$ and $I_{fb}(x)$ with coordinate x .

$$d_{fb}(x, x_j) = \sum_{x_j \in S(x)} |N(x) - N_{fb}(x_j)| \quad (46)$$

However each block difference $d(x, x_j)$ will be constructed of sum of noise differences $d_\sigma(x, x_j)$ and image autocorrelation functions $d_\alpha(x, x_j)$. In our algorithm we do not attempt to rotate or scale blocks to achieve better correlation, thus, we need to estimate the value of $d_\alpha(x, x_j)$ to adjust $w_{fb}(x, x_j)$ calculation. In our implementation we estimating $d_\alpha(x, x_j)$ by evaluation of the block difference $d_{fb}(x, x_j)$ and expected noise level $\sigma(x)$ adjusted for the image data according to the equation (2). In spatial filtering block, for each block we calculate the estimation of $d_\sigma(x, x_j)$ according to equation (47):

$$d_\sigma(x, x_j) = d(x, x_j) - d_\alpha(x, x_j) \quad (47)$$

Resulting $d_\sigma(x, x_j)$ is consequently converted into corresponding $w_{fb}(x, x_j)$ by comparison with the expected variation of noise, derived from the sensor characterization and specified in equation (2). The method for $d_\sigma(x, x_j)$ estimation and calculation of $w_{fb}(x, x_j)$ is described in Chapter 4. The efficiency of the proposed algorithm has been investigated and the experimental results are provided in a next section. Special attention has been paid to demonstrate that the results of the proposed frame accumulation are artefact free and the algorithm is efficient in all conducted tests.

5.4. Experimental results

In the proposed experiments a camera system able to capture high resolution raw data at 10 frames per second is used. In order to expose the image correctly, the camera system was programmed for a gain equivalent to ISO1600-ISO25600, enabling the capture of images at a fast shutter speed, thus eliminating motion blur. The burst sequences were processed through the full image processing pipeline ©Apical. In the first series of experiments the custom made camera system with OV8835 sensor was used. All burst sequences were taken with the camera being held by hand, thus significant amount of camera shake produced significant amount of motion in a static background captured by the camera. The amount of background motion comparable with the motion of a foreground object can be seen in Figure 25. The experiments proposed do not use any other algorithms for comparison, as an algorithm able to deal with realistic object displacements as in the case of the proposed, was not found. The result of the optical flow calculation on a coarse scale and two consecutive frames overlaid are illustrated in Figure 25 below:

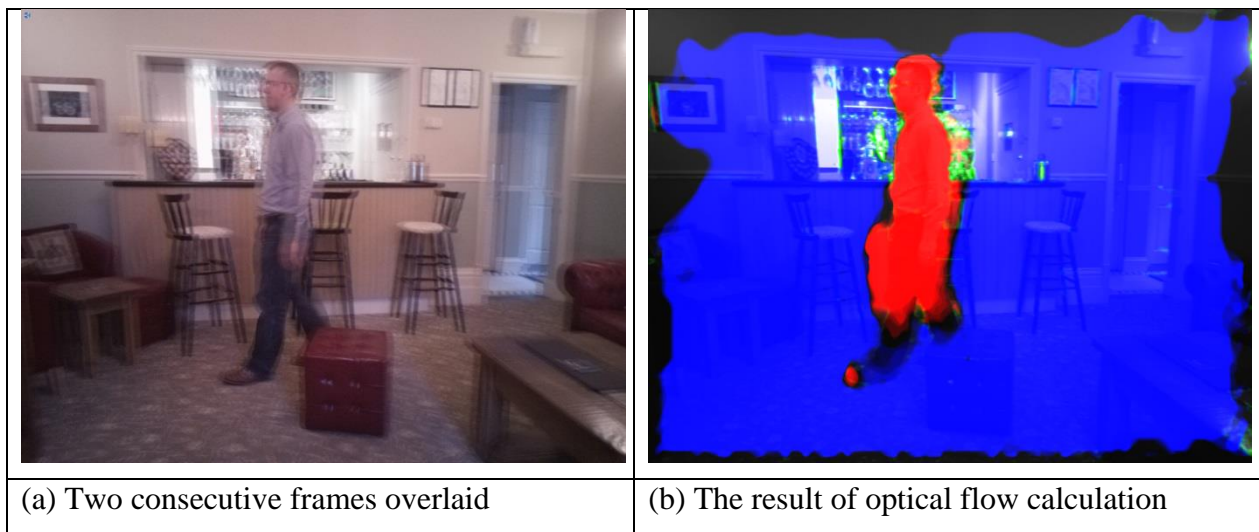


Figure 25: Motion field

The result of the noise reduction applied to noisy images taken at ISO1600 on a mobile sensor at 8MP resolution can be seen in Figure 27, Figure 28, Figure 29, Figure 30 below:

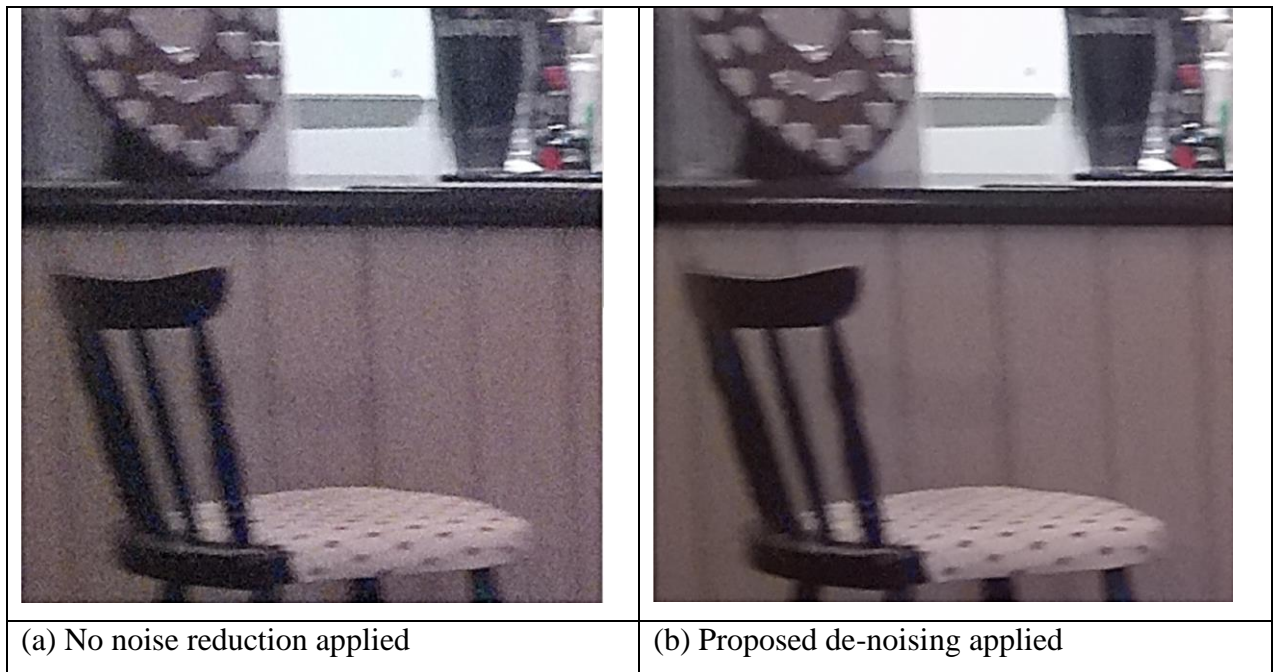


Figure 26: Experimental results, moving background.

Noise reduction effect in moving background demonstrated in Figure 26.



Figure 27: Experimental results, moving foreground.

Noise reduction effect in moving foreground object is demonstrated in Figure 27.

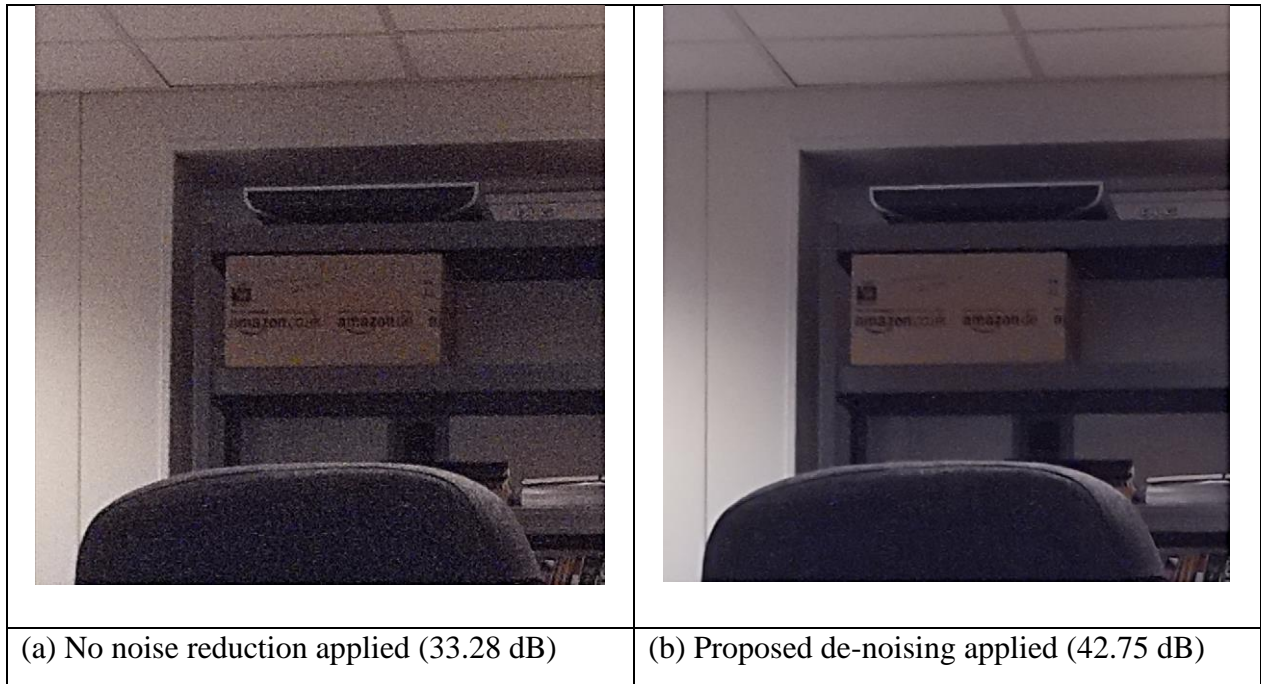


Figure 28: Experimental results, indoors scene

The PSNR values are calculated for the sample images in green, blue and red patches and provided as overlays and are also presented in Table 5.



Figure 29: Experimental results, indoors scene

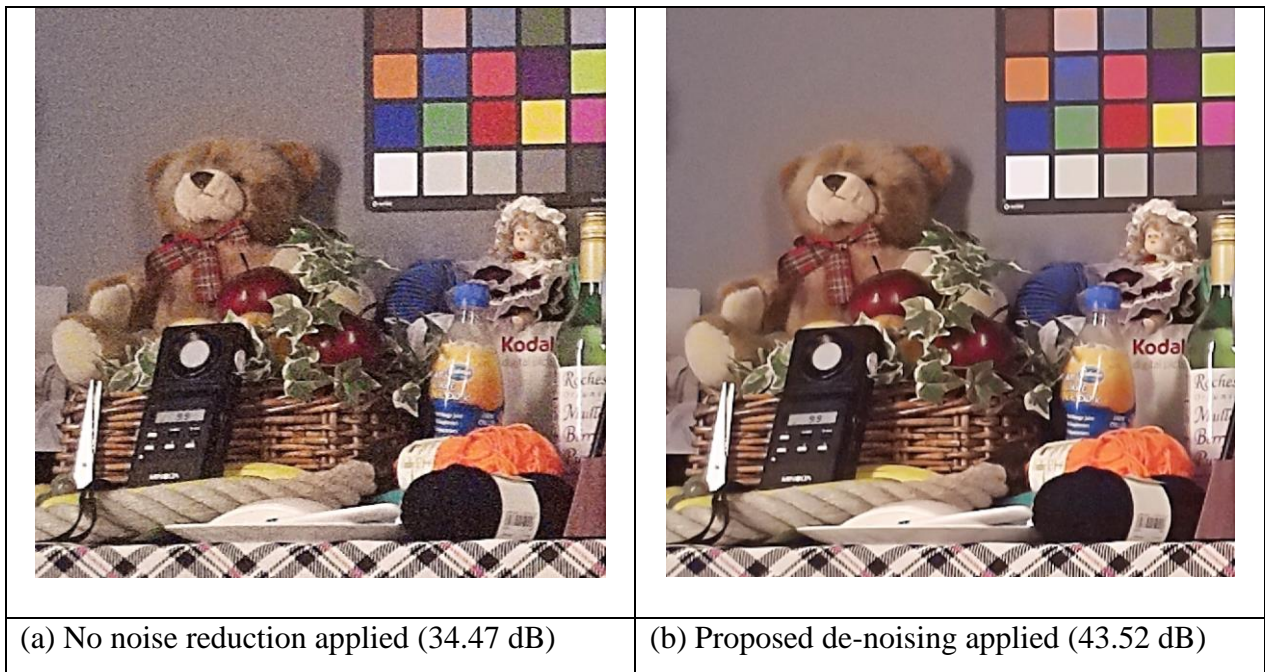


Figure 30: Experimental results, lab scene

The following samples were taken with a Sony Nex-6 camera at 16MP resolution at ISO25600:

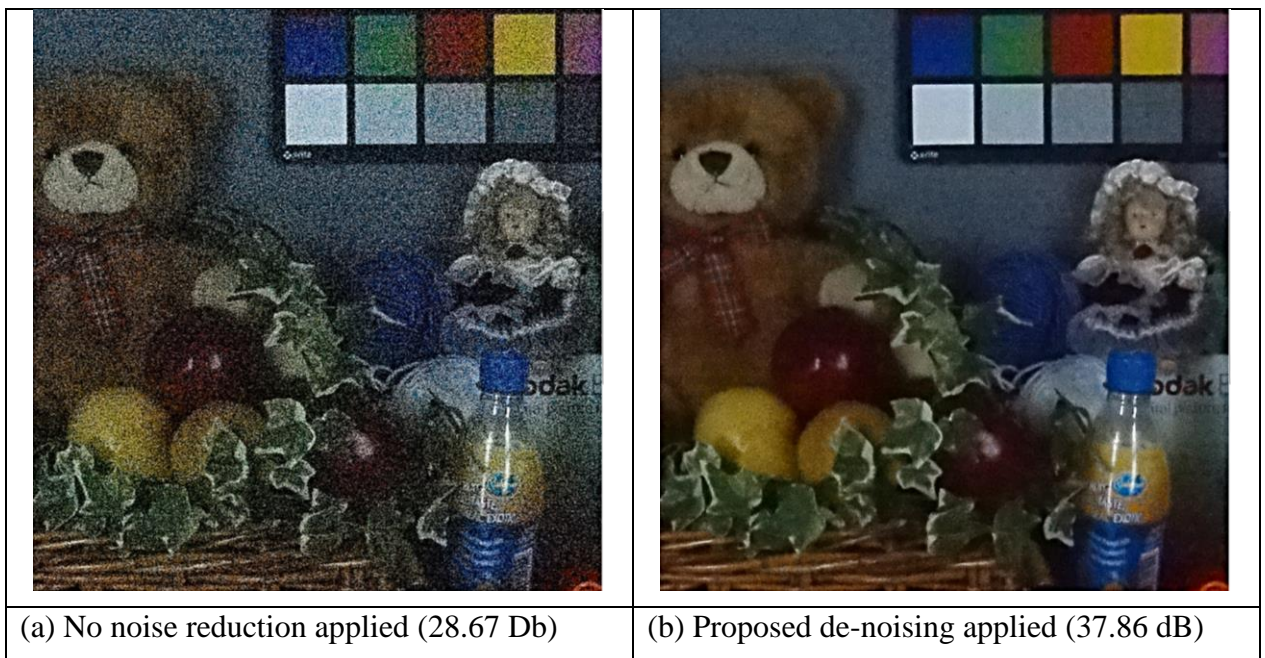


Figure 31: Experimental results, lab scene low light

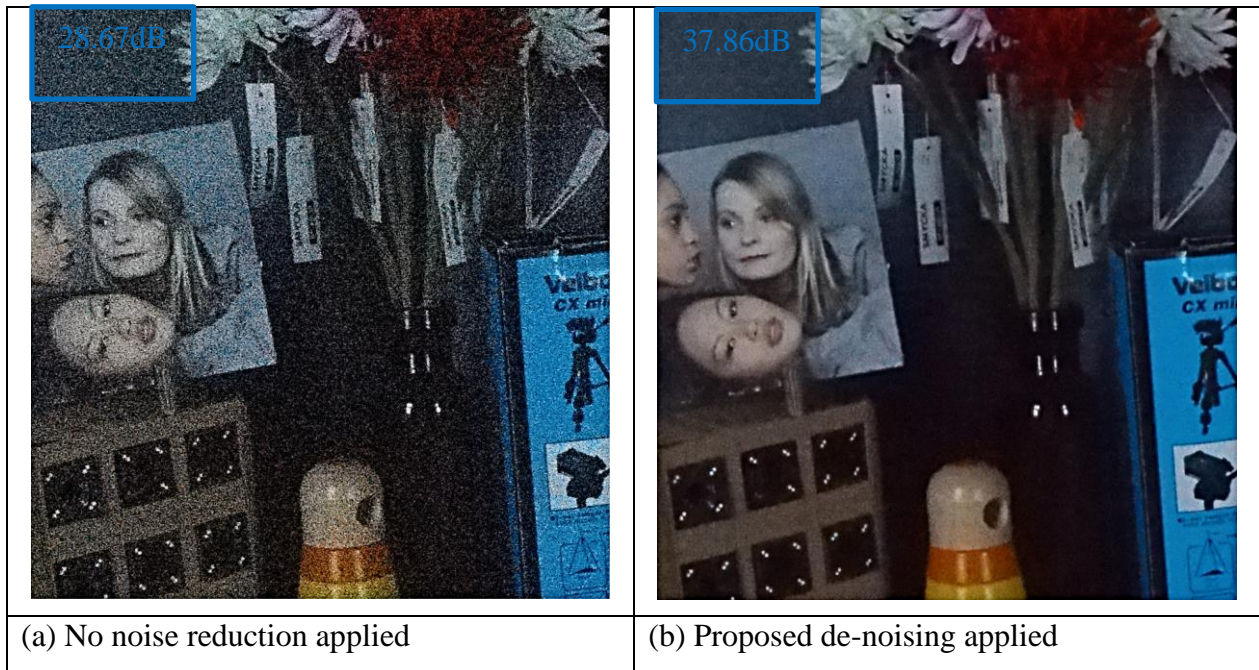


Figure 32: Experimental results, lab scene low light

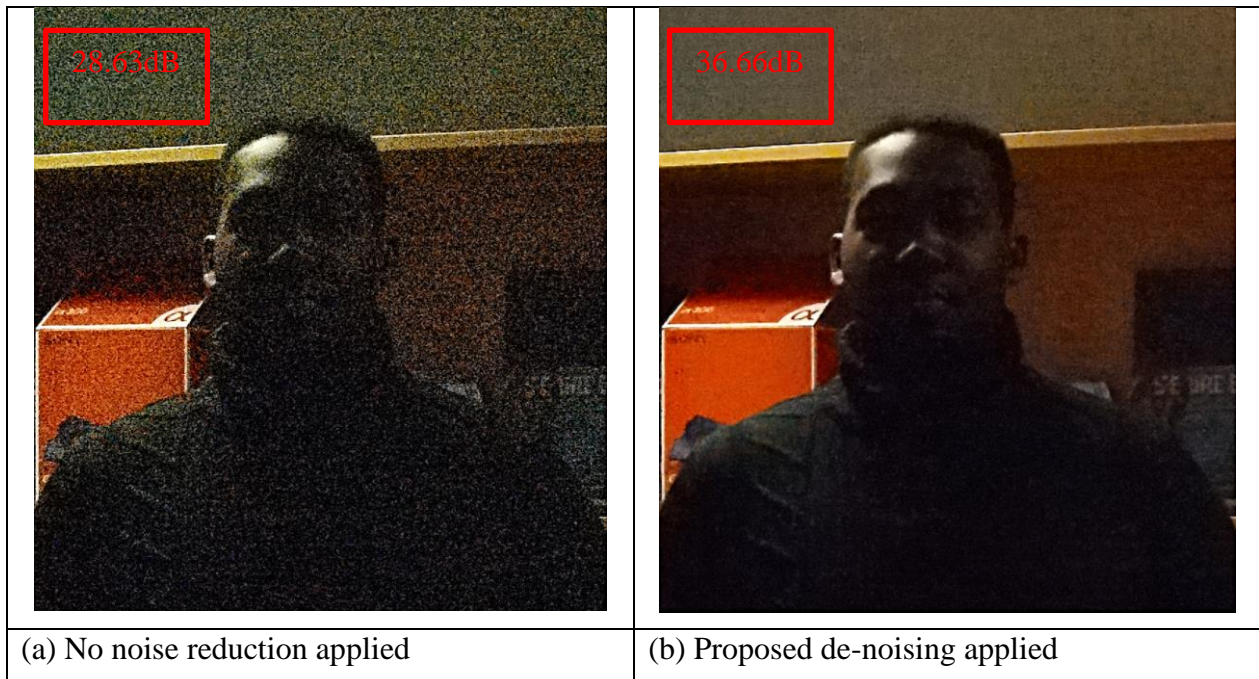


Figure 33: Experimental results, indoors low light

Since it is not possible to use the ground truth image, for the purpose of estimation of noise levels flat areas in the background were used. In selected areas the average PSNR was calculated. The results are presented in Table 5:

Scene (1) - Red Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
Proposed algorithm	42.06	42.09	41.92	42.0233
No NR applied	32.63	32.62	32.60	32.6167
Scene (2) - Green Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
Proposed algorithm	37.86	37.94	37.78	37.8601
No NR applied	28.69	28.66	28.65	28.6667
Scene (3)* - Blue Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
Proposed algorithm	36.63	36.71	36.65	36.6633
No NR applied	28.61	28.67	28.63	28.6367

Table 5: PSNR values comparison table.

Settings for de-noising algorithms were chosen to produce a good amount of noise reduction, while producing an increased amount of detail. In the proposed algorithm, as opposed to spatial only noise reduction techniques, it is seen that an increased amount of image details are obtained, at the same time achieving a remarkable amount of SNR improvement. In scene (c) we have the efficiency of the NR reduced by around 1 dB due to a significant amount of light variation (flicker).

5.5. Conclusion

In this research, a robust and efficient spatial-temporal de-noising algorithm has been proposed. Due to a number of algorithmic optimizations, the proposed algorithm, when implemented in hardware can be compact and memory bandwidth requirements can be substantially reduced compared to the spatial-temporal noise reduction algorithm described in [25]. The proposed algorithm was not compared against any existing algorithm such as VBM3D, as it would require defining a VBM3D search area of around 400 pixels, which will make the execution of the algorithm in any computer, unrealistic. The proposed algorithm has been implemented in hardware in an Altera FPGA EP3C120 and processing speed of 150Mpix/sec was achieved, which is sufficient to process HD video at 60 frames per second or 8MP images at capture rate of 15fps. Synthesis figures for the image matching block, implemented in Altera FPGA and ASIC 65nm TSMC library are presented in Table 6:

	FPGA	ASIC
Logic elements (gate-count)	93K	640K
Effective kernel size	15x15	15x15
Number of scales	1	1
Multipliers	160	(included in gate-count)
Pixel clock frequency	150MHZ	350MHZ
Video performance	1080p 60fps	4k camera 60fps
Device (silicon area)	Altera FPGA EP4C150	0.91 mm ² using 65nm process.

Table 6: Synthesis details for the image matching block

In most situations the proposed algorithm proved to be very competitive in efficiency of the noise reduction, noise structure and the natural look of images. As opposed to the previous implementation of an image matching algorithm proposed in Chapter 4, the spatial kernel was reduced as the image data was pre-matched by the Optical Flow and motion compensation. The proposed algorithm is compact when implemented in hardware and can be practically used digital camera systems to de-noise video and low-light photographic images.

Chapter 6

Image Matching in Bayer RAW Domain to Remove Ghosting in Multi-Exposure Image Fusion

6.1. Introduction

Multi exposure image fusion is a well-known approach adopted to create High Dynamic Range (HDR) images and emulate the Human Visual System (HVS) using Standard Dynamic Range (SDR) cameras. The main limitation of current multi-exposure image fusion techniques is their inability to compensate for moving objects in a scene and camera shake. Previous attempts to solve camera shake have been able to accurately align the multi-exposure images that have static backgrounds prior to their fusion. Nonetheless, image alignment cannot solve the issue of ghosting artefacts due to moving objects. In the proposed research local motion compensation technique, previously used for noise reduction purposes, is used to efficiently remove ghosting artefacts due to both, camera shake and object movement in the scene.

HDR photography can be achieved by capturing images at different exposures and then fusing them to produce a HDR image. However, in order to produce artefact free images, the fusion technique has to be able to compensate for motion caused by camera shake and object movements in the scene. During the fusion process, ghosting artefacts are generated due to the fractionally time difference instances of the objects' displacement within the multi-exposure images captured. An attempts were made by other researches to resolve ghosting artefacts in [45],[44], however practical solution to the problem of multi-exposure image fusion has not been found.

In Chapter 5 a practical method to perform spatial-temporal noise reduction in RAW images, and wide dynamic range images from SLR cameras was proposed. The spatial-temporal method relied on the idea of matching, blending, and recursive accumulation of image data into a frame buffer to improve signal to noise ratio. Errors due to motion were handled by the noise reduction engine. In this chapter, the previously developed spatial-temporal noise reduction

method has been extended, and utilised for the purpose of multi-exposure image fusion. It is noted that this is a practical application of the previously proposed spatial-temporal noise reduction method, in which accumulation of data is carried out by the sensor, and not by a frame buffer. Thus, the problem of matching images taken at different exposures is transformed into an already solved problem, which consists of matching a clean image and a noisy image in order to produce artefact free HDR images.

6.2. Proposed Image Fusion Method

In the proposed HDR method, all processes are performed in Bayer RAW domain as it allows more accurate calculations when fusing the images due to the linear nature of the data. The first step in the proposed approach is to match the intensities of the multi-exposure images prior to compensating for motion. The motion estimation and compensation is carried out in two stages. Firstly, a robust optical flow method described in Chapter 5 is used for coarse motion estimation and matching. Secondly, coarse image matching is followed by a block-matching process described in Chapter 5, allowing a sub-pixel image matching. Once the multi-exposure images are motion corrected, the images are fused through a blending process, and finally a dynamic range compression method is used to create the resulting HDR image. Figure 34 shows the block diagram of the proposed multi-exposure image fusion method:

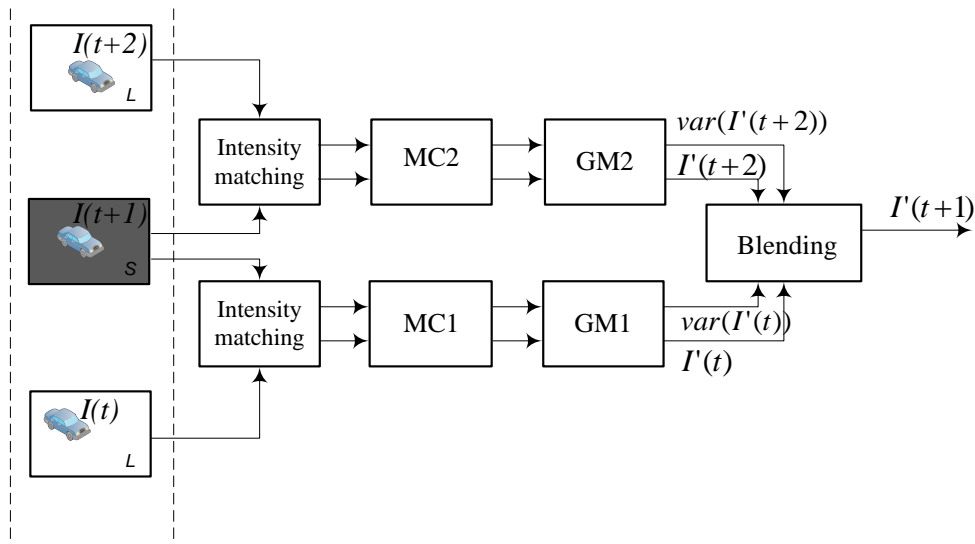


Figure 34: The block diagram of the proposed multi-exposure image fusion algorithm

The advantage of the proposed approach is that in the case of a failure in motion estimation, the resulting image will not have any warping distortions. In the worst case, the resulting image will appear as if motion estimation was never performed. Image matching algorithm is applied to the frame triplets with a frame order of Long-Short-Long (whenever available), where the Short-exposure image is considered to be a reference. Intensity matching is applied to match the global levels of the images intensity. Effectively Long-exposure images are divided by the exposure ratio value.

Optical flow is calculated for the intensity matched Short/Long-exposure image pairs using sparse image feature vectors matching technique previously proposed in Chapter 5. Pixel mapping is performed for the areas where motion error is less than the estimated noise, elsewhere the pixel data is taken directly from the Short-exposure image.

Resulting motion-compensated estimates are blended according to the least error that is obtained during motion compensation.

6.2.1. Intensity Matching

In the stage of intensity matching, the short exposure and long exposure images $I(t+1)$ and $I(t)$, $I(t+2)$ with exposures E_s and E_l are matched in the intensity domain. Further, images $I(t)$, $I(t+2)$ will be referred as I_{l1} , I_{l2} and the short exposure image $I(t+1)$ as I_s . In order to match the intensity of the multi-exposure images, the exposure ratios E_{r1} and E_{r2} given by equation (48) are calculated, if not known from the camera system, and I_{l1} , I_{l2} are matched to I_s using equation (49), where, \hat{I}_{s1} and \hat{I}_{s2} are the intensities of matched images.

$$E_{r1} = \frac{E_{l1}}{E_s}; \quad E_{r2} = \frac{E_{l2}}{E_s} \quad (48)$$

$$\hat{I}_{s1} = \frac{I_{l1}}{E_{r1}}; \quad \hat{I}_{s2} = \frac{I_{l2}}{E_{r2}} \quad (49)$$

The intensity matching process is performed to allow the image fusion process to have the least error. Otherwise, it would not be possible to blend parts of images that have the same content but non-matching pixel values. The intensity matching process can be very accurate since it is performed in linear Bayer RAW domain and the exposure ratio is known or controlled.

6.2.2. Coarse and fine motion estimation and compensation

The motion estimation and compensation process (*MC*) that matches \hat{I}_{s1} and \hat{I}_{s2} to I_s prior the fusion is accomplished in a two stage process represented as $\hat{I}_{smc1} = MC(I_s, \hat{I}_{s1})$ and $\hat{I}_{smc2} = MC(I_s, \hat{I}_{s2})$. In the first stage, coarse motion estimation and compensation is performed to remove possible artefacts due to large object movements and camera shake. This is achieved by adopting an extended version of the multi-scale motion estimation technique described in Chapter 5, which is based on linear matching of image features in a transform space. The extension of the multi-scale approach in this research included algorithmic optimisations and modifications to operate in linear Bayer RAW domain. In the second phase of motion estimation, a non-local means filter described in Chapter 4 is used in order to achieve sub-pixel precision matching of the images. The use of the non-local means filter allows accurate removal of artefacts due to possible objects' local motion in scenes and inability of coarse motion estimation to compensate local motion with pixel precision. An example of the motion map calculated during the proposed motion estimation process is presented in Figure 36. The combination of the two methods allowed achieving motion estimation and compensation for a wide range of object displacements and camera shake, while producing a high level of precision in image matching.

6.2.3. Image blending

In order to fuse the multi-exposure images, the image temporal filter proposed in Chapter 4 is adopted to perform images blending. In this stage two predicted images \hat{I}_{smc1} and \hat{I}_{smc2} are blended, based on the evaluation of its differences from the reference I_s , producing a wide dynamic range image, I_{WDR} . In this stage of processing, one can benefit from predictable noise levels, which can be used as an absolute reference for the quality of matching. In reference to the Figure 34, local variances: $\text{var}(I'(t))$ and $\text{var}(I'(t+2))$ can be normalised by the local noise variance expectation σ according to equation (50), (51):

$$d_t = \text{var}(I'(t)) / \sigma \quad (50)$$

$$d_{t+2} = \text{var}(I'(t+2)) / \sigma \quad (51)$$

Being normalised, matching differences can be linearly related to each other, making it possible to blend the resulting image on a pixel level according to formula (52):

$$I'(t+1) = \frac{I'(t+2) \times d_{t+2} + I'(t) \times d_t}{d_{t+2} + d_t} \quad (52)$$

The image $I'(t+1)$ will be referred further as I_{wdr} . The proposed method allows artefact free precise image fusion with local motion compensation. The proposed algorithm's complexity does not prohibit its practical implementation. Furthermore based on the algorithms described in previous chapters, it can be implemented as a hardware block.

6.2.4. Dynamic range compression

In order to be able to visualise the contents of the wide dynamic range image produced during the image blending stage, a local histogram equalization technique is applied to I_{wdr} in order to make the shadow part of the image as visible as the highlighted parts of the image. The dynamic range compression algorithm ©Apical was used in experiments. This dynamic range compression algorithm is an important part of the evaluation of the system, though not the subject for the proposed research. The objectives for the inclusion of the dynamic range compression algorithm are: the ability to represent the shadow parts of the resulting I_{wdr} image, matching the look of corresponding parts in the I_t , and the ability to represent the highlight parts of an image in the resulting I_{wdr} image, matching the look of corresponding parts in the I_s , and the minimisation of any low spatial frequency artefacts. The objectives are quite difficult to formalize and describe, using quantitative metrics. Thus such issues are not investigated assuming that the dynamic range compression algorithm serves its purpose. Thus the work concentrates on the noise measures of the resulting image I_{wdr} . The objectives discussed above are demonstrated in Figure 35 and can be found in each experimental result presented in this chapter.

6.3. Experimental results

Experiments were carried out to evaluate the performance of the proposed approach and its ability to remove ghosting artefacts in multi-exposure image fusion. Figure 35 illustrates the short and long exposure images taken with exposures that differ by a factor of 8 or 16 and processed by a conventional image-processing pipeline, compared with the result of proposed fusion technique followed by dynamic range compression:

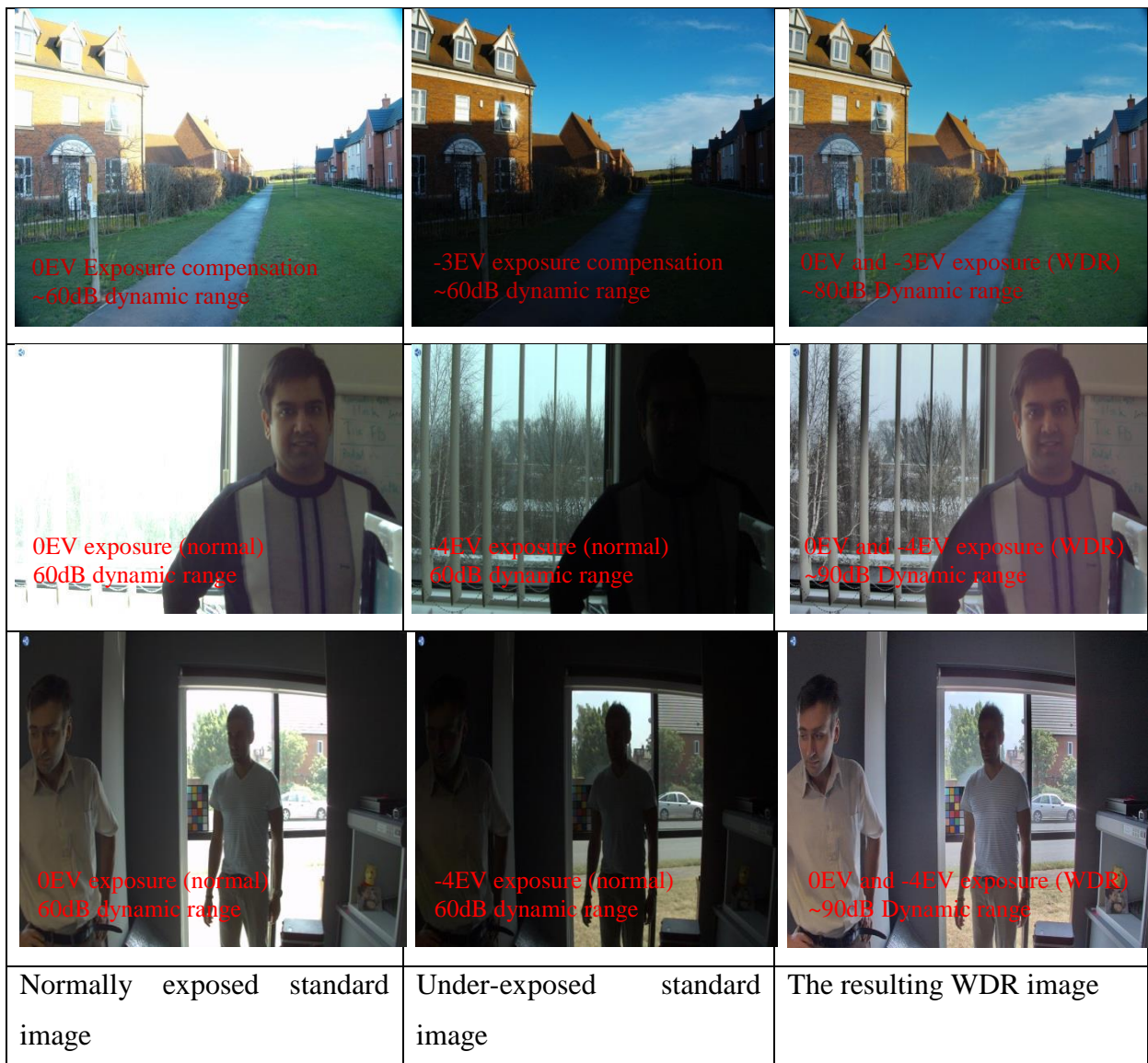


Figure 35: The results of fusion.

It can be seen in Figure 35 that the shadow parts of WDR image visually equivalent to corresponding image parts in normally exposed image, while the highlight parts of WDR image are taken from the under-exposed images.

The example of coarse motion calculation presented in Figure 36 below, image in the left column represents 3 video frames stacked. Image in the right column represents the motion fields calculated, overlaid on the reference image.



Figure 36: Example of motion and calculated motion field.

It is observed how fast moving objects are best represented partially in a short exposure image, and partially in the long exposure image. This scenario is a common situation where fusion techniques fail to deal with object displacements, and therefore ghosting artefacts are produced.

Chapter 6: Image Matching in Bayer RAW Domain to Remove Ghosting in Multi-Exposure Image Fusion

In the examples provided, the long exposure image has clipped highlight parts, while the shadow area is dark, but contains some detail. In contrast the short exposure image has well-exposed highlights but the shadow areas are clipped. The DRC algorithm applied to the short exposure image would reveal large amounts of noise in the shadow area, while long exposure image would provide reasonably clean shadow areas and clipped highlights. The proposed multi-exposure image fusion is able to produce a ghost free HDR image, and preserve the noise levels in shadow parts of the image, matching the long exposure image. The results are provided in Figure 37 and Figure 38.

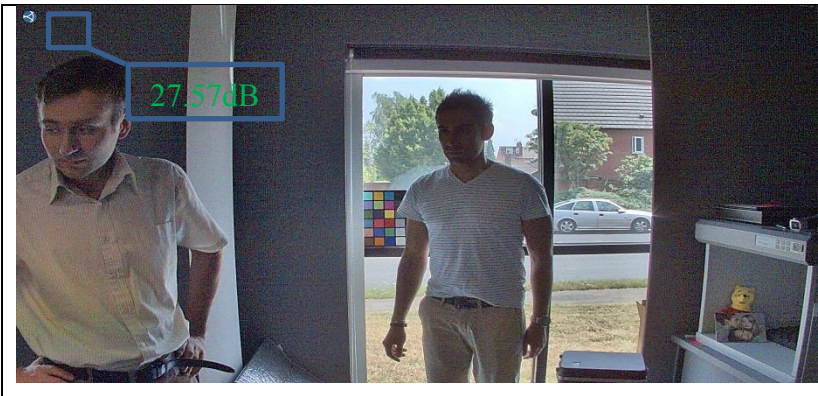
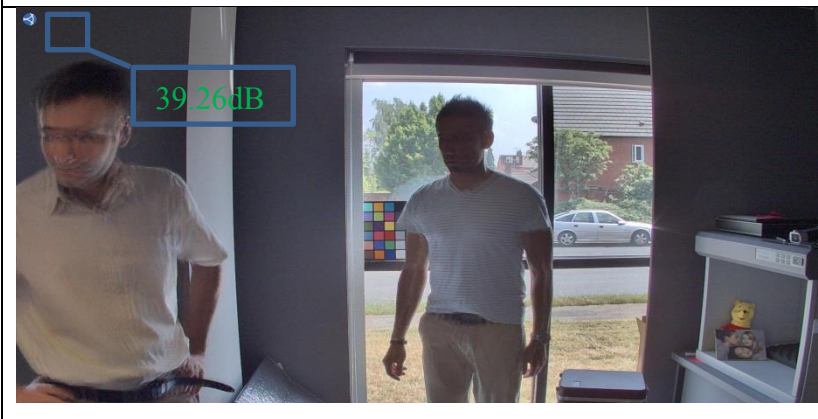
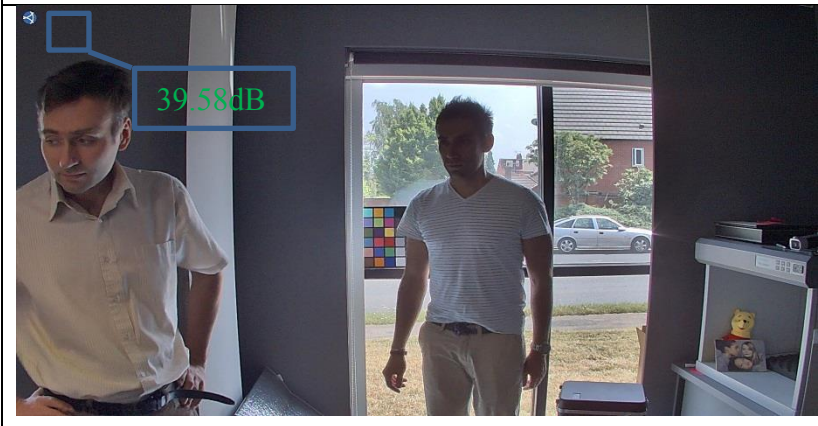
	<p>a) DRC applied to I_s</p>
	<p>b) DRC applied to I_{wdr} without motion compensation</p>
	<p>c) DRC applied to I_{wdr} with motion compensation</p>

Figure 37: PSNR values, calculated on images with DRC applied.




	<p>a) DRC applied to I_s</p>
	<p>b) DRC applied to I_{wdr} without motion compensation</p>
	<p>c) DRC applied to I_{wdr} with motion compensation</p>

Figure 38: PSNR values, calculated on images with DRC applied.

Another example of the use of the proposed approach is presented in Figure 32. In this case images were captured with a Sony NEX-6 camera at 16MP resolution and 8 times exposure ratio:

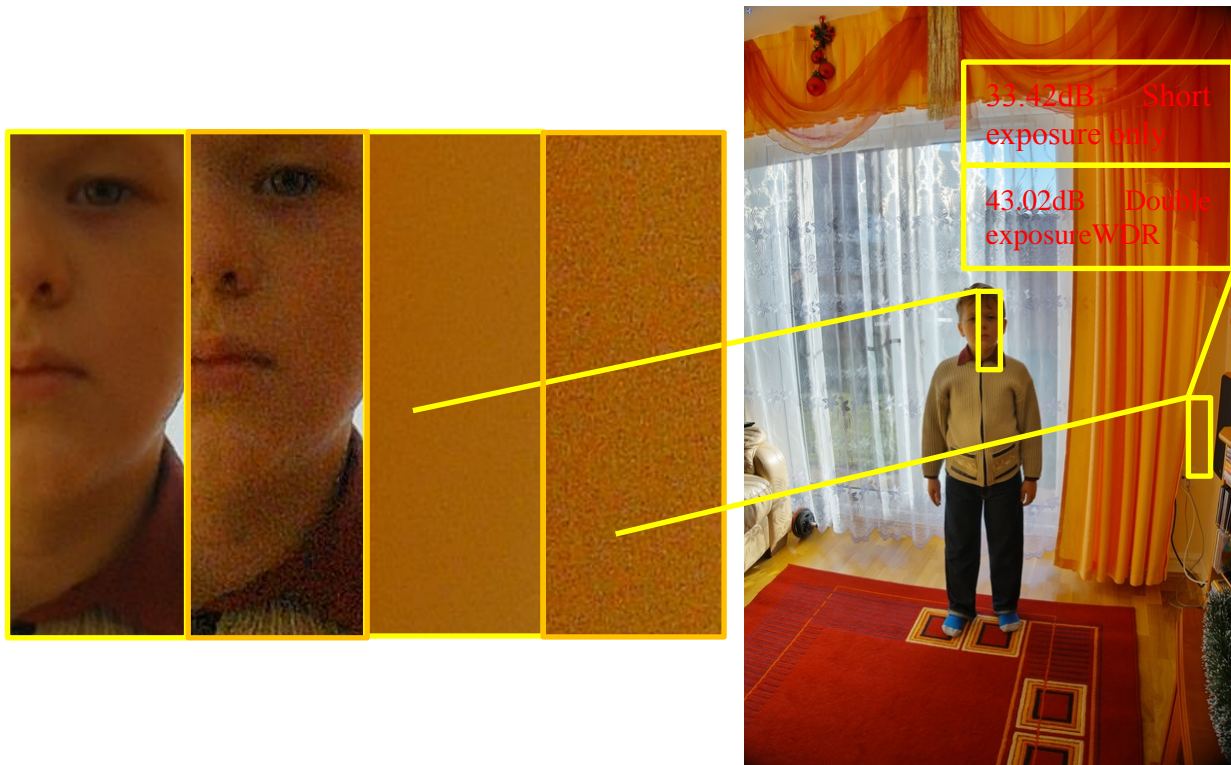


Figure 39: HDR image obtained as the result of the proposed HDR method.

The limitations of the proposed algorithm is similar to that of the previous proposed spatio-temporal noise reduction algorithm, i.e. in order to match objects in a pair of images, the objects should be present in both scenes and should be captured in both images. This limitation sets a limit on the range of optimal exposure ratios usable when capturing multi-exposure images. If images are taken at very different exposure levels, dark objects may have no details captured in a short exposure image, thus making the multi-exposure matching task impossible. In the experiments conducted an exposure ratio values of up to 16 was successfully used for this purpose.

The effect of dynamic range compression in the proposed experiments was balanced to achieve the visibility of shadow parts to match corresponding image sections of the long exposure image. The effectiveness of proposed image fusion algorithm was evaluated by measuring the PSNR values in shadow parts of resulting WDR image. The closer the PSNR values between

the long exposure image and the WDR image, the better is the quality of fusion, given that the ghosting artefacts were not found in the resulting WDR image.

Scene (a) - Red Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
Long exp	42.056	43.258	43.754	43.0233
Short exp. compressed	33.453	32.682	34.142	33.4257
Double exp. compressed	42.057	43.780	44.482	43.4397
Scene (b) - Green Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
Long exp	41.5630	41.1813	41.7133	41.4859
Short exp. compressed	27.4847	27.6332	27.6023	27.5734
Double exp. compressed	39.3774	39.7915	39.5819	39.5836
Scene (c) - Blue Patch PSNR				
	Red (dB)	Green (dB)	Blue (dB)	Average (dB)
Long exp	41.8070	42.3748	41.5054	41.8957
Short exp. compressed	29.8470	29.9396	29.8507	29.8791
Double exp. compressed	41.1211	41.3653	40.9409	41.1424

Table 7: PSNR values, comparison table.

6.4. Conclusion

A multi-exposure image fusion algorithm suitable for practical implementation in hardware has been proposed in this chapter. The proposed algorithm is able to perform images stitching, on images taken at different exposures, to allow dynamic range compression. A part of proposed image fusion algorithm referred to pixel mapping was implemented in hardware. The details of pixel mapping block implementation in Altera FPGA and ASIC 65nm are presented in Table 8:

	FPGA	ASIC
Logic elements (gate-count)	119K	750K
Effective kernel size	17x17	17x17
Number of scales	1	1
Multipliers	240	(included in gate-count)
Pixel clock frequency	150MHZ	350MHZ
Video performance	1080p 60fps	4k camera 60fps
Device (silicon area)	Altera FPGA EP4C150	1.12 mm ² using 65nm process.

Table 8: Synthesis results for proposed pixel mapping block.

The proposed algorithm proved to be efficient in different lighting conditions and scenes and was proven to work well with different sensors. It was shown that the shadow areas taken at longer exposures have better contrast and contain more details than the same areas processed through Spatio-Temporal NR. Proposed algorithm can successfully absorb mismatches between images being matched. In the situations when the successful images matching is not possible, the proposed algorithm demonstrates reduced de-noising effect, however affected image areas look natural and artefact free. Image areas, where pixel-mapping was successful, were reproduced by details captured in long-exposure image precisely.

Chapter 7

The use of sensor noise modelling in the segmentation and detection of objects.

7.1. Introduction

Most of the existing object detection algorithms are based on machine learning classifiers, which in their turn use features extracted from an image. The research being conducted at present in the object classification area is very intense, one of the most successful object detection techniques is known as HOG-SVM described in [24],[25],[26],[27],[28]. The results produced by object detection algorithms are continuously improving. Fundamentally there are two approaches to enhance the results of an object detection algorithm. The first approach is an enhancement of classification methodology, where many techniques have been proposed in literature (Linear classifiers, Neural networks etc.). The second approach is an enhancement of features used. Researches who focus their work on the enhancement of features extracted from an image mostly concentrate on finding the set of discrete primitives, describing the image content. The process of feature extraction is usually related to filtering of the image data and normalisation of the filter's response. However, there is one common flaw in most feature extraction techniques, i.e., during the normalisation and accumulation of image features the assumption is made that the filters producing stronger response represents stronger image features. In practice all researches work with digital video or photographic images, that are products of image processing pipelines, processing the image sensor data with unknown settings. As previously discussed, such processing can significantly alter image data, breaking linear dependencies between parts of an image and unbalancing the appearance of different image elements. This chapter investigates how feature extraction can be made more robust by taking sensor characteristics into account during feature extraction.

7.2. A feature extraction model, utilizing histogram of oriented gradients.

The first step in the calculation of Histogram of Oriented Gradients is edge detection, As opposed to the standard approach presented in [40],[41],[42],[43],[44] edge kernels will be applied to linear data and the output will be normalised according to the expected noise. Gabor edge kernels with 6 different orientations were used in our experiments. The Gabor functions for an orientation of 90 degrees are presented in Figure 40:

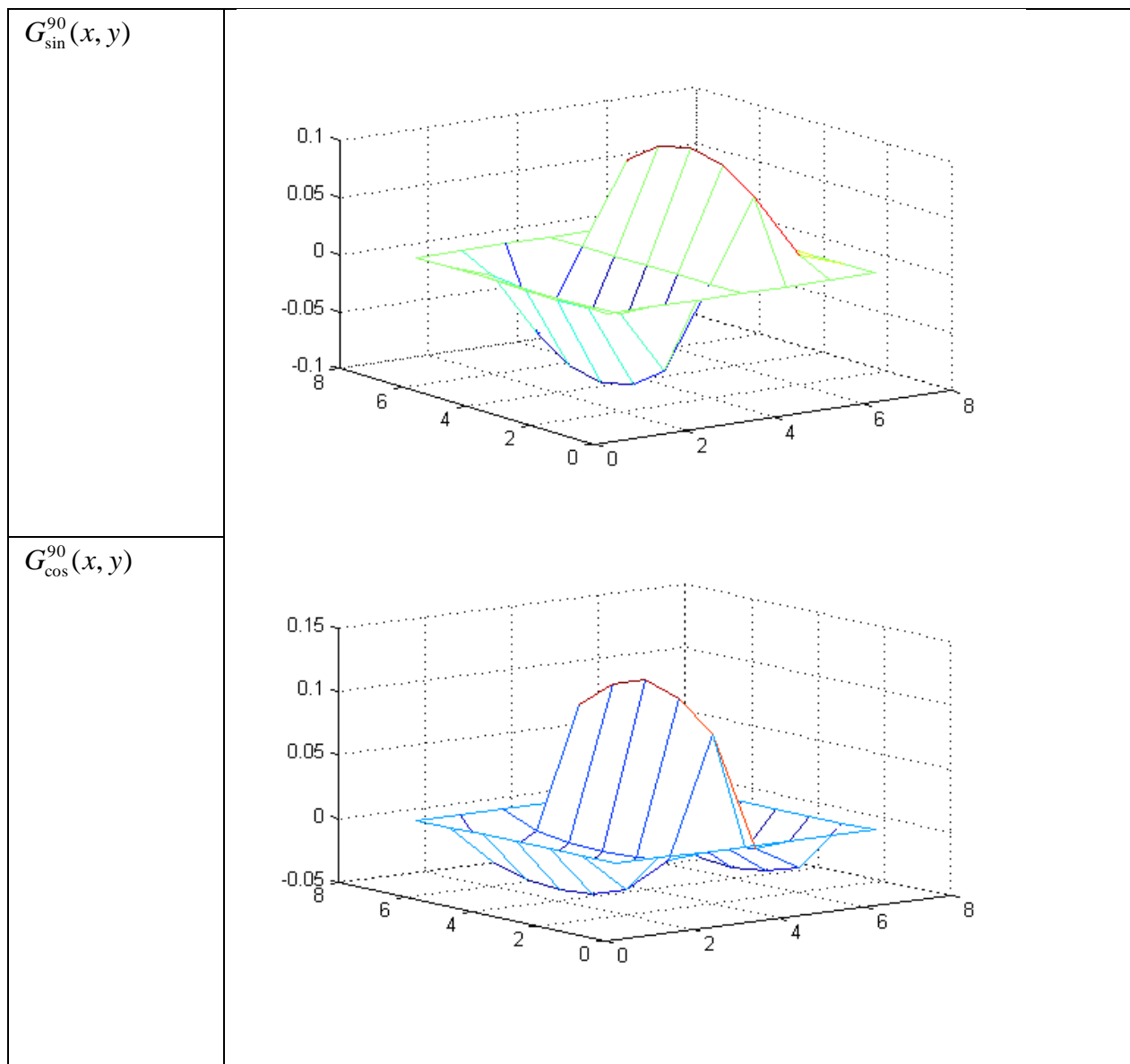


Figure 40: Edge segmentation functions.

The response for one edge orientation would be calculated according to the equation (53):

$$E^\alpha(x, y) = \left| \sum_{i,k \in K} G_{\sin}^\alpha(x+i, y+k) \times I(x+i, y+k) \right| + \left| \sum_{i,k \in K} G_{\cos}^\alpha(x+i, y+k) \times I(x+i, y+k) \right| \quad (53)$$

Where $K(i, k)$ is a spatial kernel of the Gabor function. Assuming that local details of an image $I(x, y)$ at each coordinate were illuminated with different intensity, the response $E^\alpha(x, y)$ will significantly differ for bright and dark parts of the image. In the proposed object detection system, however the interest is in some measure of reliability of detected edges. As the edge response was calculated in linear RAW data space, the response can be normalised by the expectation of noise at each pixel with coordinate x, y in the image.

7.3. Proposed feature normalization method

In reference to the equation (2), the expectation of noise variance $\sigma(x, y)$ for each image area $I(x, y)$ can be matched. Further it should be considered that the edge detection kernels $G_{\cos}^\alpha(x, y)$ and $G_{\sin}^\alpha(x, y)$ are constructed as a linear combination of Gaussian function $G_K(x, y)$ and functions of $\sin(x)$ and $\cos(x)$. Thus the normalization of the edge response is performed according to the following equation (54):

$$E_{norm}^\alpha(x, y) = \frac{E^\alpha(x, y)}{\sum_{i,k \in K} \sigma(x+i, y+k) \times G_K(x+i, y+k)} \quad (54)$$

For the purpose of comparison the Edge response $E_{gamma}^\alpha(x, y)$ was calculated according to the formula (55):

$$E_{gamma}^\alpha(x, y) = \left| \sum_{i,k \in K} G_{\sin}^\alpha(x+i, y+k) \times I_g(x+i, y+k) \right| + \left| \sum_{i,k \in K} G_{\cos}^\alpha(x+i, y+k) \times I_g(x+i, y+k) \right| \quad (55)$$

Where $I_g(x, y)$ is a non-linear representation of $I(x, y)$, obtained by the application of the non-linear standard gamma function sRGB. $E_{norm}^\alpha(x, y)$ and $E_{gamma}^\alpha(x, y)$ were used for the comparison of the object detection algorithm's performance. The proposed edge response normalization approach demonstrates improved performance of object detection, operating in non-standard conditions, such as low-light, which is also important for sensors with non-standard noise characteristics. It is important to note also that the proposed scheme makes object detection independent from the settings of the image processing pipeline, which guarantees the best performance in embedded and mobile devices.

7.4. Experimental results

In the experiments conducted a RGB sensor with a Bayer pattern was used. The sensor is typical for use within security, automotive and computer vision systems. The setup of the experiment consisted of the custom made camera system, allowing video capture in Bayer RAW format at full HD resolution and 25 frames per second. A firmly mounted camera system was used to record video in indoor conditions. The computer vision algorithm, trained to detect people was used for object detection. In one scenario feature extraction was done traditionally, i.e., without any knowledge about image sensor. In the second scenario extracted features were locally normalized by the sensor noise variance expectation. To evaluate the effectiveness of the proposed scheme, a number of experiments were conducted, capturing video sequences at different lighting conditions. As expected, the detection rate deteriorates as the noise within the image increases. Another observation is that the detection rate was higher in a system, where sensor noise characteristics were taken into account. Examples of individual detections are presented in Figure 41 in which, the first row presents the original video frames, the second row presents the objects detected using proposed method and the third row presents the objects detected using the standard gamma method.

The statistics of the results of detections are presented in Table 9. People detection was performed under 2 categories: head and upper body (UB). Heads were detected using 3 classifiers, trained for 3 different poses. Upper body was detected using 5 classifiers, trained for 5 different poses, respectively. Strong detections refer to positive classifier responses larger than 0.4 and weak classifier responses refer to positive classifier responses between 0.1 and 0.4. People detections refer to combined response from either of two categories. Formal detection rate is counted as the number of strong detections over the number of possible detections. A human object is considered to be detected if it has a strong detection in either category. The formal false positive rate is based on the ratio of strong incorrectly classified objects to the total number of objects to be detected.



Figure 41: Object detection experimental results.

	ISO-100		ISO-1600	
	Sensor normalized	Gamma normalized	Sensor normalized	Gamma normalized
heads detected	635	634	593	565
heads strong	620	607	568	524
heads weak	15	27	25	43
heads missed	0	1	31	57
UB detected	631	631	612	598
UB strong	617	602	581	561
UB weak	14	29	31	37
UB missed	0	0	2	16
False positives	0	0	4	21
Missed people	0	0	2	16
Track errors	0	1	1	1
Formal people detection rate	100.00%	99.84%	99.76%	97.06%
Formal heads detection rate	97.64%	95.59%	91.02%	83.97%
Formal UB detection rate	97.78%	95.40%	94.62%	91.36%
Formal false positives rate	0%	0%	0.32%	1.69%

Table 9: Detection rates statistical data.

It can be seen that the normalization according to sensor noise model significantly improves detection rate and reduces false positives rate, which is more prominent at higher ISO settings. The visualization of the first three components of the edge detectors at 90, 60 and 30 degrees are presented in Figure 42. It is noted that the original image was taken from a camera system running at ISO-1600 sensitivity.

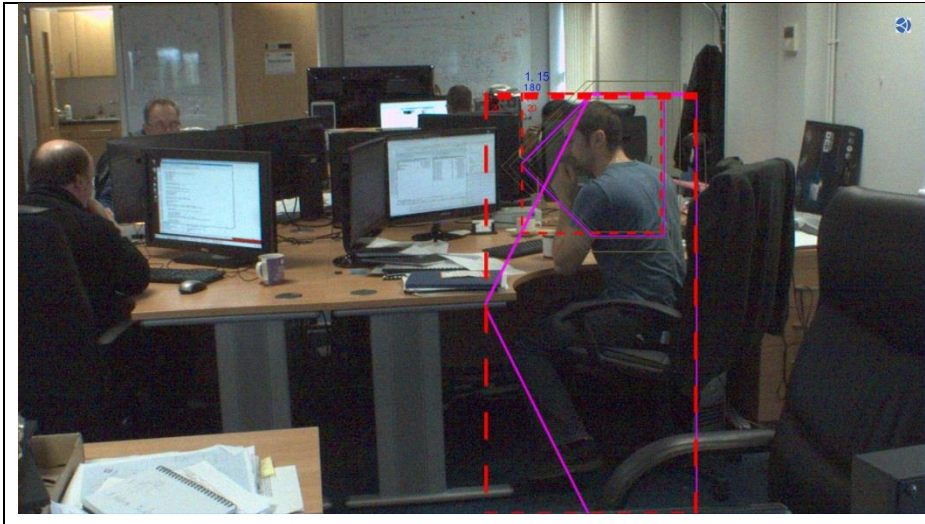
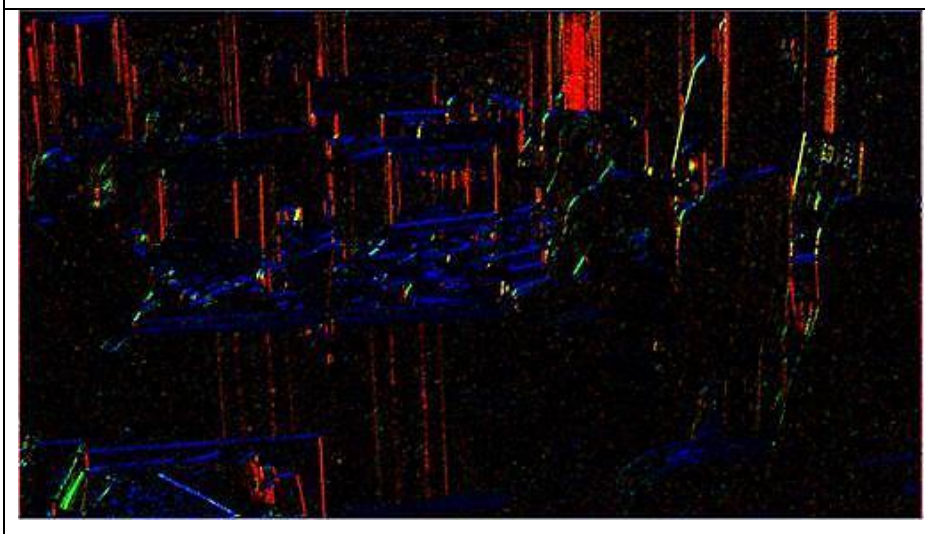
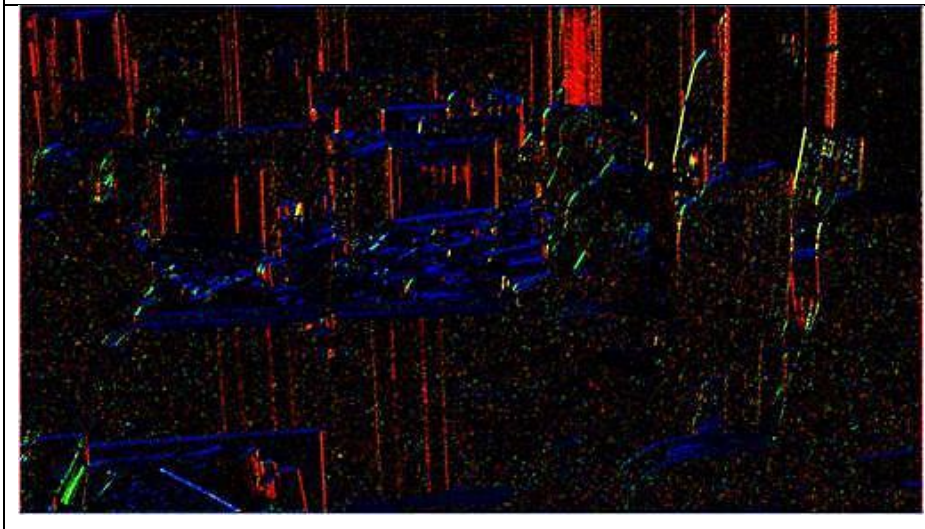
 The image shows a computer lab with several people at desks. A red dashed rectangle highlights a person sitting at a desk. A magenta solid line outlines a more complex shape around the person, representing a detection overlay.	<p>a) Original image with detection overlay</p>
 The image shows the edge detection results for the proposed scheme. The edges are highlighted in red, blue, and green, forming a complex network of lines that outline the objects in the scene.	<p>b) Visualized edge detection results using the proposed scheme</p>
 The image shows the edge detection results for the standard normalization method. The edges are highlighted in red, blue, and green, forming a complex network of lines that outline the objects in the scene.	<p>c) Visualized edge detection results using the standard normalization method</p>

Figure 42: An example of noise model normalized edge segmentation.

In Figure 42 it is seen that the response from the edge detector is cleaner after the normalization, comparing to the response from the standard edge detector, which was running on gamma corrected data. The improved system response in the presence of the noise, improves the results of object detection, the results of which are presented in Table 9. It is a fair statement that a similar effect can be achieved by using noise reduction techniques. However achieving a noise free output from the edge detectors by performing the sensor noise modelling is a more efficient way of improving the system performance.

7.5. Conclusion

The proposed method of edge detectors response normalization was successfully used in object detection engine, implemented in hardware. The details of object detection engine implemented in Xilinx Zynq 7045 FPGA are presented in Table 10 below:

Resource	Utilization	Available	Utilization %
FF	151193	437200	34.6
LUT	103865	218600	47.5
Memory LUT	259	70400	0.4
BRAM	197	545	36.1
DSP48	734	900	81.6
BUFG	5	32	15.6
MMCM	0	8	0.0

Table 10: Object detection system resource utilization

The proposed edge detection improvements allowed achieving better object detection performance comparing to other known systems. The proposed method of edge detectors response normalization also allowed running object detection engine on sensor RAW data. It can be noted that the proposed method can allow object detection system implementation without the ISP being involved, which reduces the overall cost of the system and can be beneficial when the actual image from the object detection system is not required for a privacy reasons.

Chapter 8

Conclusions and Future Work

8.1. Conclusions

This chapter summarizes contributions made to the research area of image processing algorithms by the work presented in this thesis. It further concludes the findings of the contributory chapters (3-7) based on experimental results and evaluations.

This thesis made a number of original contributions in further improvements of image processing algorithms, and image processing systems. In particular it was proved that image processing algorithms, using images sensors as calibrated instruments, can benefit from predictable noise characteristics and demonstrate better performance, while allowing significant algorithmic optimizations.

1. It was proposed that multi-scale block matching algorithms would perform better, preserve more details and suppress noise more efficiently, when applied in Bayer RAW data space. Despite additional complexity associated with processing in RAW domain, the complexity of the algorithm can be reduced due to the reduction in number of color components to process. Due to access to linear sensor data it was possible to perform sensor noise modelling, providing a reliable reference for a block matching algorithm. A non-linear SAD filter was designed to separate data auto-correlation from sensor noise. The non-linear SAD filter allowed the preservation of additional details and boosted noise suppression. It can be concluded that the proposed spatial noise reduction algorithm can outperform the best known spatial noise reduction algorithms, while its complexity remains at acceptable level. The algorithm has been included in a number of commercial products and appears to be one of the highest valued algorithms in Apical LTD's portfolio.
2. The multi-scale block matching algorithm designed to work within Bayer RAW data was subsequently modified to perform image matching in a video sequence. It was shown that inter-frame block matching proved to be more reliable when operated in Bayer RAW domain. Temporal data accumulation was implemented recursively,

Kalman filter coefficients were used to predict the aggressiveness of the accumulation, Due to access to linear sensor data, it was possible to perform sensor noise modelling, and use noise estimation as a reference level for accumulation aggressiveness calculation. Proposed algorithmic solutions allowed achieving one of the best known noise reduction performance characteristics commercially, while the resource usage and bandwidth were kept at a reasonable level, making practical use of the developed algorithm feasible.

3. The further development of Spatio-Temporal noise reduction algorithm was done by using the Optical flow algorithm to estimate the motion field. An introduction of Optical flow allowed reduction of the requirements for block matching and compensated for large object displacements. Aforementioned features allowed large resolution image processing, making the algorithm suitable for still image accumulation to perform image de-noising on a burst of still images. The quality and performance of widely used Optical flow algorithms would not allow us to propose any practical solution. In order to improve the precision of the optical flow and introduce algorithmic optimizations, it was proposed that optical flow is calculated in the Bayer Raw domain and use predicted sensor noise levels as a normalization factor in energy minimization framework. The noise-related normalization factor introduced, enabled the improvement of the precision of motion field prediction and to avoid more complex motion field filtering. Proposed algorithmic optimizations enabled the practical use of the proposed algorithms in commercial imaging devices.
4. Previously developed image matching algorithms were proved to perform reliably and being feasible for practical use in commercial devices. However a further key challenge that remains an open research problem is that these algorithms have not yet reached an acceptable level of quality to become useful is the problem of multi-exposure image fusion. This problem was traditionally approached from the side of post processing. In our research we proposed to attempt to solve this difficult problem from the point of view of noise reduction. The problem of image fusion was reformulated into a problem of image matching, using a spatio-temporal noise reduction framework. In the proposed approach we adopted previously developed algorithms to perform image matching in a linear data space, which allowed us to match images more precisely, in addition to more reliable image matching, performed by inter-frame block matching, Gaussian accumulation and optical flow. Being built on the base of previously developed

algorithms, multi-exposure image fusion was performed by the image processing blocks implemented as hardware blocks. This work makes an important use case of the image processing algorithms served by Apical's image processing pipeline.

5. Methods and algorithm developed in previous research were used to improve the quality of feature extraction functional block in an object detection system. The edge features used by the Histogram of Oriented Gradients were extracted in Bayer RAW data domain, whilst sensor noise model was used to perform the filter's response normalization. In our experiments we proved that the quality of edge feature extraction can be improved, compared to traditional methods of HOG feature extraction. Another important outcome of the proposed solution is that the object detection system proposed does not require an image processing pipeline that makes the whole system more compact and allows building an object detection camera, which does not produce video. The results of our research were used to produce an embedded object detection system, implemented as a hardware block.

8.2. Future work

In the research conducted within the scope of this thesis investigated how knowledge about the image sensor can be successfully used to improve traditional algorithms, related to image processing and object detection. In general any algorithm working with the intensity of color gradients and features can operate more reliably, when estimated noise levels can be predicted. In practice it means that the whole image processing pipeline has to be redesigned to allow algorithms to access linear data from the sensor. However as with any research work there is room for further development of the ideas presented in this thesis.

In the proposed research it was assumed that the sensor noise can be successfully modelled by the Gaussian Poissonian noise model. As the image sensors become more and more advanced the noise generated by such sensors does not match our simple model sufficiently well. In the future we are looking at the possibility of adding PCA to train the filter to learn the sensor noise and hence to detect sensor specific noise patterns and enforce their removal from the output image.

Our proposed block matching and Optical flow algorithms were designed to be robust to image brightness variations. However there is certainly an effect imposed by the varying image brightness on the quality of block matching and Optical flow calculation. Further the effect of

local brightness variation has a negative impact on multi-exposure image fusion. In the experiments conducted global brightness correction was adopted, but better quality of image processing could have been expected by introducing local image brightness matching.

The algorithm proposed in this thesis to perform still image matching and accumulation was implemented to improve the signal to noise characteristics of an image. However it can be successfully used to combine images taken with different focus settings, providing an image with infinite depth of field. Normally such an effect is achieved by reducing the lens's aperture, which has an impact of increased noise. It was proposed to use previously developed techniques to fuse images taken at different times at different focuses of lens. The image obtained would have all areas in a good focus and demonstrate a level of noise reduction. The resulting image can be consequently processed by algorithms, simulating shallow depth of field, in order to achieve DSLR like images. The contribution of this improvement will be to provide a better quality image, revealing more details and focused better in general. Out of focus effect can be simulated by using thin lens model.

Proposed in our work method of edge filters output normalization, according to modelled sensor noise can be generalized and used to improve the response of local binary pattern feature extraction algorithms. It is a known issue with local binary patterns (LBP) that their resilience to noise is weaker than in edge segmentation methods. The modelling of the sensor noise can improve the reliability of local pattern detection and consequently the quality of machine vision algorithms.

The work presented in this thesis result in a number of practical algorithmic implementations, making products where these algorithms were included, more competitive and of better quality practically. One of the philosophical outcomes of this work was the finding that understanding the nature of a data is a benefit to any algorithm, where this knowledge can be effectively used. Proposed methods require considerable efforts to be implemented in image data processing algorithms, but offer high practical value.

References

- [1] Hao Deng, "Mathematical approaches to digital color image denoising", Georgia Institute of Technology 3394373, (2009).
- [2] Peter J. Burt, Edward H. Adelson, "The Laplacian Pyramid as a Compact Image Code" IEEE Transactions on Communications (1983)
- [3] Aram Danielyan, Markku Vehvilainen, Alessandro Foi, Vladimir Katkovnik and Karen Egiazarian, "Cross-color BM3D filtering of noisy RAW data", IEEE International Workshop, 125-129 (2009).
- [4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik and Karen Egiazarian, "Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space", IEEE International Conference on Image Processing Vol. 6, (2007).
- [5] Priyam Chatterjee, Peyman Milanfar, "Practical bounds on image denoising: From estimation to information", IEEE Trans. Image Processing, VOL. 20, NO. 5, (2011).
- [6] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," IEEE Trans. Image Process., vol. 12, no. 11, pp. 1338–1351, Nov. (2003).
- [7] Priyam Chatterjee, and Peyman Milanfar "Is Denoising Dead?", IEEE Transactions on Image Processing Vol. 19, NO. 4, (2010).
- [8] Siwei Lyu, Simoncelli E.P., "Modeling Multi-scale Subbands of Photographic Images with Fields of Gaussian Scale Mixtures", IEEE Transactions on pattern analysis and machine intelligence Vol. 31, 693-706, (2008).
- [9] Paliy D., A. Foi, R. Bilcu, Katkovnik V., "Denoising and Interpolation of Noisy Bayer Data with Adaptive Cross-Color Filters", SPIE-IS&T Electronic Imaging Visual Communications and Image Processing Vol. 6822, (2008).
- [10] Zhang L., R. Lukac, X. Wu, D. Zhang, "PCA-based Spatial Adaptive Denoising of CFA Images for Single-Sensor Digital Cameras", IEEE Transactions on Image Processing Vol. 18, 797-812 (2009).
- [11] Alexey Lukin, "A multiresolution approach for improving quality of image de-noising algorithms", IEEE International Conference ICASSP Vol. 2, II857-II860 (2006).

- [12] Foi A. M., Trimeche V., Katkovnik, K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single image raw-data", *IEEE Transactions on Image Processing* Vol. 17, (2008).
- [13] Liu, C., W.T. Freeman, R. Szeliski, and S.B. Kang, "Noise estimation from a single image", *Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR 2006*, pp. 901-908, (2006).
- [14] Ce Liu, Richard Szeliski, Sing Bing Kang, C. Lawrence Zitnick, and William T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299-314, February (2008).
- [15] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Computer Vision*, 2(1):1-104, December 2006.
- [16] Blanksby, A.J., M.J. Loinaz, D.A. Inglis, and B.D. Ackland, "Noise performance of a color CMOS photogate image sensor", *IEEE Int. Electron Devices Meeting 97 Tech. Dig.*, pp. 205-208, 1997
- [17] Vladimir S. Petrovic, Costas S. Xydeas, "Sensor noise effects on signal-level image fusion performance", *Information Fusion - INFFUS* , vol. 4, no. 3, pp. 167-183, 2003
- [18] Bovik A., [Handbook of Image and Video Processing], Academic Press, New York, (2000).
- [19] C. Kervrann, J. Boulanger, P. Coupe, "Bayesian Non-Local Means Filter, Image Redundancy and Adaptive Dictionaries for Noise Removal", *SSVM Proceedings of the 1st international conference*, 520-532 (2007).
- [20] Hua Zhong, Yongwei Li, Jiao L.C., "SAR Image Despeckling Using Bayesian Nonlocal Means Filter With Sigma Preselection", *IEEE Geoscience and Remote Sensing Letters* Vol. 8, 809 - 813 (2011).
- [21] Bayer B.E., "Color imaging array", U.S. Patent 3971065, (1976).
- [22] I. V. Romanenko, E. A. Edirisinghe and D. Larkin, "Block matching noise reduction method for photographic images applied in Bayer RAW domain and optimized for real-time implementation," *SPIE Proceedings 8437*, (2012).
- [23] Ilya V. Romanenko, Eran Edirisinghe, Daniel Larkin, "Spatial-Temporal noise reduction method optimized for real-time implementation," *SPIE Proceedings 8655*, (2013).

-
- [24] Antoni Buades, Bartomeu Coll and Jean-Michel Morel, “Nonlocal image and movie denoising” *International Journal of Computer Vision* Vol. 76, Number 2, 123-139, (2008).
- [25] Ce Liu, William T. Freeman “A High-Quality Video Denoising Algorithm based on Reliable Motion Estimation”, *European Conference on Computer Vision* (2010).
- [26] Ren Ying, Chua Chin-Seng, Ho Yeong-Khing, “Statistical background modeling for non-stationary camera”. *Pattern Recognition Letters*, 24, 183-196. (2002).
- [27] Sheikh Y., Shah M., “Bayesian modeling of dynamic scenes for object detection”. *IEEE transactions on pattern analysis and machine intelligence*, 27(11), 1778-92, (2005).
- [28] I. V. Romanenko, E. A. Edirisinghe, D. Larkin, "Block matching noise reduction method for photographic images applied in Bayer RAW domain and optimized for real-time implementation", *Proceedings of SPIE* Vol. 8437, 84370F (2012).
- [29] J.Konrad and E. Dubois, “Bayesian estimation of motion vector fields,”*IEEE Transactions. Pattern Analysis. Mach. Intell.*, vol. 14, no. 9, pp. 910–927, Sep. 1992.
- [30] Li Xu, Jiaya Jia, Yasuyuki Matsushita, “Motion detail preserving optical flow estimation” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1293 – 1300, (2010).
- [31] H W Haussecker and D J Fleet, “Computing optical flow with physical models of brightness variation” *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(6):661–673, (2001).
- [32] B K P Horn and B G Schunck, “Determining optical flow”. *Artificial Intelligence*, 17(1-3):185–203, (1981).
- [33] C Lei and Y H Yang, “Optical flow estimation on coarse-to-fine region-trees using discrete optimization” *ICCV*, (2009).
- [34] M. L. Gong, Y. H. Yang. “Estimate large motions using the reliability-based motion estimation algorithm” *International Journal of Computer Vision* 68(3): 319-330 (2006).
- [35] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr. “Variational optical flow computation in real-time” *IEEE Transactions on Image Processing*, 14(5):608–615, May (2005).
- [36] A. G. Bors and I. Pitas, “Optical flow estimation and moving object segmentation based on median radial basis function network,” *IEEE Trans. Image Process.*, vol. 7, no. 5, pp. 693–702, May (1998).

-
- [37] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application in stereo vision," in Proc. 7th Int. Joint Conf. Artif.Intell., pp. 674–679, (1981).
- [38] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.
- [39] Alexey Lopich, David R. W. Barr, Bin Wang, Piotr Dudek, "Real-time image processing on ASPA2 vision system" IEEE International Symposium on Circuits and Systems - ISCAS , pp. 1989-1989, 2011
- [40] Moctezuma, D.; Conde, C.; de Diego, I.M.; Cabello, E., "Person detection in surveillance environment with HoGG: Gabor filters and Histogram of Oriented Gradient," Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on , vol., no., pp.1793,1800, 6-13 Nov. (2011)
- [41] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", in: CVPR, pp. 886–893, (2005)
- [42] W. Jiang, K.-M. Lam, T. Shen, "Efficient edge detection using simplified Gabor wavelets", IEEE Trans. Syst. Man Cybern. Part B 39 (4) 1036–1047, (2009)
- [43] V.Kruger, G.Sommer, "Gabor wavelet networks for efficient head pose estimation", Image Vision Comput.20 (9–10) 665–672, (2002).
- [44] Q. Zhu, M.-C. Yeh, K.-T. Cheng, S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients", in: CVPR, pp. 1491–1498, (2006).
- [45] Lluís-Gómez A.; Saravi S.; Edirisinghe E. A., "Subjectively optimised multi-exposure and multifocus image fusion with compensation for camera shake" Proc. SPIE 8436, Optics, Photonics, and Digital Technologies for Multimedia Applications. (2012).
- [46] Jinno, T, Okuda, M. "Multiple Exposure Fusion for High Dynamic Range Image Acquisition" Image Processing, IEEE Transactions on (Volume:21 , Issue: 1), Jan (2012).
- [47] Kotwal, K, Chaudhuri, S, "An optimization-based approach to fusion of multi-exposure, low dynamic range images", Information Fusion, 2011 Proceedings of the 14th International Conference Jul (2011).
- [48] E. Reinhard, S. Pattanaik, G. Ward, and P. Debevec, "High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting", ser. Morgan Kaufmann Series in Computer Graphics and Geometric Modeling. San Mateo, CA: Morgan Kaufmann, (2005).

References

- [49] P. Debevec, "Image-based lighting," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 26–34, Mar. (2002).
- [50] F. Mccollough, "Complete Guide to High Dynamic Range Digital Photography", China: Lark Books, (2008).
- [51] B. Hoefflinger, *High-Dynamic-Range (HDR) Vision*, ser. Springer Series in Advanced Microelectronics. New York: Springer-Verlag, (2006).
- [52] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 2005–2008.
- [53] Kalman R.E., "A new approach to linear filtering and prediction problems", *Transactions of the ASME-Journal of Basic Engineering*, 82 (Series D), pp 35-45, (1960).
- [54] Nikaido, M., Tamaru, N., "Noise reduction for gray image using a Kalman filter", *SICE 2003 Annual Conference*, (Vol. 2), Aug., (2003).
- [55] Turney, R.D., Reza, A.M., Delva, J.G.R., "FPGA implementation of adaptive temporal Kalman filter for real time video filtering", *Acoustics, Speech, and Signal Processing*, 1999. *Proceedings., IEEE International Conference on (Volume:4)* (1999)

Appendix A. List of Publications

Published Conference Papers:

Ilya V. Romanenko, Eran Edirisinghe, D. Larkin, "Image matching in Bayer raw domain to de-noise low-light still images, optimized for real-time implementation", in Computational Imaging XII, Charles A. Bouman; Ken D. Sauer, Editors, Proceedings of SPIE Vol. 9020 (SPIE, Bellingham, WA 2014), 90200Y.

Ilya V. Romanenko, Eran Edirisinghe, Daniel Larkin, "Spatial-temporal noise reduction method optimized for real-time implementation", in Image Processing: Algorithms and Systems XI, Karen O. Egiazarian; Sos S. Agaian; Atanas P. Gotchev, Editors, Proceedings of SPIE Vol. 8655 (SPIE, Bellingham, WA 2013), 86550L.

Ilya V. Romanenko, E. A. Edirisinghe, D. Larkin, "Block matching noise reduction method for photographic images applied in Bayer RAW domain and optimized for real-time implementation", in Real-Time Image and Video Processing 2012, Nasser Kehtarnavaz; Matthias F. Carlsohn, Editors, Proceedings of SPIE Vol. 8437 (SPIE, Bellingham, WA 2012), 84370F.

Ilya V. Romanenko, Alexis L Lluís-Gomez, Eran Edirisinghe "Image Matching in Bayer RAW Domain to Remove Ghosting in Multi-Exposure Image Fusion", in International Conference on Consumer Electronics (ICCE), proceedings 978-1-4799-1291-9-14 (2014)

Conference presentations: I. Romanenko, M. Tusch, October 2013 Embedded Vision Summit East.

Appendix B. Sensors used in experiments

Sensor model:	AS3372	OV8835	Sony NEX-5	Sony NEX-6
Resolution	2M	8M	13M	16M
Capture rate	60fps	12fps	7fps	12fps
Data bits	12	10	14	14
Sensor active area size	4.6x3.4 mm	4.6x3.4 mm	23.6x15.6 mm	23.6x15.6 mm
Pixel size	2.7u	1.4u	3.5u	3.3u
Max S/N ratio	59dB	36.6dB	Not available	Not available

Table 11: Image sensors used in experiments