



Vaasan yliopisto
UNIVERSITY OF VAASA

VELI-TAPANI PELTOKETO

Benchmarking of Mobile Phone Cameras

ACTA WASAENSIA 352

COMPUTER SCIENCE 16
TELECOMMUNICATION ENGINEERING

Reviewers

Professor Farag Sallabi
University of United Arab Emirates
Department of Information Technology
P.O. Box 15551
Al Ain, UAE

Vice President, Dr. Lasse Eriksson
Cargotec Oyj
P.O. Box 61
FI-00501
Helsinki, Finland

Julkaisija Vaasan yliopisto	Julkaisupäivämäärä Elokuu 2016	
Tekijä(t) Veli-Tapani Peltoketo	Julkaisun tyyppi Artikkeliväitöskirja	
	Julkaisusarjan nimi, osan numero Acta Wasaensia, 352	
Yhteystiedot Vaasan yliopisto Teknillinen tiedekunta Tietotekniikan yksikkö PL 700 65101 VAASA	ISBN 978-952-476-684-5 (painettu) 978-952-476-685-2 (verkkoaineisto)	
	ISSN 0355-2667 (Acta Wasaensia 352, painettu) 2323-9123 (Acta Wasaensia 352, verkkoaineisto) 1455-7339 (Acta Wasaensia. Tietotekniikka 16, painettu) 2342-0693 (Acta Wasaensia. Tietotekniikka 16, verkkoaineisto)	
	Sivumäärä 168	Kieli Englanti
Julkaisun nimike Matkapuhelinkameroiden suorituskykymittaus ja vertailu		
Tiivistelmä <p>Matkapuhelinten käytön kasvu on korostanut matkapuhelimien kameroiden kuvalaadun merkitystä. Viime vuosina on julkaistu useita uusia kuvalaatuun liittyviä standardeja ja vanhoja standardeja päivitetään jatkuvasti. Standardit määrittelevät kuvalaadun mittaukset kuitenkin vain tietyille ominaisuuksille ja yleisiä mittareita koko kameran laadulle ei ole olemassa. Tässä väitöskirjassa tutkitaan ja määrittellään vaatimuksia ja haasteita, jotka liittyvät matkapuhelin-kameroiden suorituskykymittauksiin ja vertailuun.</p> <p>Väitöskirjassa luodaan laaja katsaus digitaalisen kuvauksen laatuun ja tekijöihin, jotka heikentävät kuvan laatua. Samalla työssä paneudutaan virheisiin, joita tavataan digitaalisessa kuvauksessa. Väitöskirja käsittelee myös kameroissa esiintyviä viiveitä ja kameroiden nopeutta. Lisäksi kameroiden laadun ja suorituskyvyn mittauksessa käytettäviä mittareita ja metodeita esitellään kattavasti. Työ sisältää viisi aiemmin julkaistua artikkelia, jotka yhdessä väitöskirjan muun tekstin kanssa muodostavat tutkimuskokonaisuuden.</p> <p>Väitöskirja sisältää pohdintaa erilaisista kuvalaadun ja suorituskyvyn mittareista sekä miten niitä voidaan käyttää kameroiden vertailussa. Työ esittelee myös haasteet, jotka liittyvät useiden ominaisuuksien yhdistämiseen yhdeksi suorituskykyä mittaavaksi arvoksi kameroiden vertailun helpottamiseksi. Lisäksi väitöskirjassa käsitellään erilaisia ympäristömuuttujia ja niiden vaikutusta suorituskykymittaukseen.</p> <p>Tutkimuksen tuloksena väitöskirja esittelee uuden suorituskykymittausmetodin matkapuhelimien kameroille. Erittäin tärkeä osa työn tulosta on myös suorituskykymittaukseen liittyvien vaatimusten ja haasteiden määrittely.</p>		
Asiasanat Suorituskykymittaus, matkapuhelinkamera, kuvanlaatu.		

Publisher University of Vaasa	Date of publication August 2016	
Author(s) Veli-Tapani Peltoketo	Type of publication Selection of articles	
	Name and number of series Acta Wasaensia, 352	
Contact information University of Vaasa Faculty of Technology Department of Computer Science P.O. Box 700 FI-65101 VAASA FINLAND	ISBN 978-952-476-684-5 (print) 978-952-476-685-2 (online)	
	ISSN 0355-2667 (Acta Wasaensia 352, print) 2323-9123 (Acta Wasaensia 352, online) 1455-7339 (Acta Wasaensia. Computer Science 16, print) 2342-0693 (Acta Wasaensia. Computer Science 16, online)	
	Number of pages 168	Language English
Title of publication Benchmarking of Mobile Phone Cameras		
Abstract Great success of the mobile phone industry has highlighted image quality of mobile phone cameras. Several new standards are published during recent years and old standards are continuously updated. Nevertheless, each standard measures and validates a certain image quality feature or artefact and there is a lack of generic metrics to validate a whole camera system. This thesis investigates and defines the requirements and challenges of benchmarking when mobile phone cameras are compared. This thesis contains a comprehensive introduction of image quality factors, image quality distortions and artefacts, which are common in digital imaging. In addition, performance issues like delays and slowness of camera systems are investigated. Correspondingly, quality metrics and methods, which are used to validate image quality and performance of digital cameras are described. This work includes five previously published articles, which are an essential part of the thesis work. This thesis includes considerations of different image quality and performance metrics and how they should be used in benchmarking. The challenges of a single number benchmarking score is discussed. The environmental factors, like lightness are evaluated, too and their influence to benchmarking is discussed. The outcome of the research is a novel benchmarking method for mobile phone cameras which includes both quality and performance metrics. Even more important this thesis highlights the requirements and challenges of a mobile phone camera benchmarking.		
Keywords Benchmarking, mobile phone cameras, image quality, performance.		

PREFACE

It is contradictory to look back the time period which I have spent in University of Vaasa. Contradictory, because the time from the autumn 2009 is at the same time very short but contains so many exciting and rewarding moments. Moments, when I have been tested myself and qualified, how far I can force and motivate myself.

In 2009 I started to study again, after 16 years in the working life. From the very first steps, the spirit and extremely beautiful environment of Vaasa University fascinated and convinced me that I am the right place to learn more and try something new. When I graduated as a master of science in 2011, it was quite obvious to ask a permission to continue my studies as a postgraduate student.

However, a university, how beautiful it could be, is just a bunch of empty buildings without inspiring people. I would like to thank all the personnel and student colleagues of Vaasa University because they made the University of Vaasa a good place to be and study. Particularly, I am very grateful to Professor Mohammed Elmusrati, who has supported, advised, and guided me during the doctoral work.

I have had a privilege to work and study at the same time. The twofold role has given me perspective to focus to the essential areas in the studies and research. At the same time, my role in the working life has enabled to use learned things and research results at once in real products and vice versa, validate the research methods using real data got from the product development.

Therefore I have a great pleasure to thank the whole personnel of Sofica Ltd. The atmosphere of the company encouraged me to target higher than I would have never expected. Especially I would like to thank Mr. Marko Nurro, who partially forced but mostly motivated me to face new challenges. Without the exciting and rewarding moments in demanding customer meetings and negotiations, I would not have been ready to present my research and papers in several conferences.

I would also thank my colleagues in Nokia Technology. I have worked in Nokia Technology and particularly in Digital Media group, when I have been writing my doctoral thesis. Again, I have learned new facts of image quality and new approaches to tackle challenging problems.

I am also grateful to Mr. Graham Soundy, who has shown extremely patience to review and correct my English, which is – I afraid – far from the beautiful British English.

VIII

Every manuscript requires a proper review. Professor Farag Sallabi and Doctor Lasse Ericsson made a valuable work to review and comment my thesis. Especially I want to thank Doctor Lasse Ericsson, whose review was extremely comprehensive and denoted a great commitment to pre-examination work as well as great understanding of the research area.

Finally, it is difficult to highlight enough the importance of the most loved ones. The pages of this thesis would not be enough to express my gratefulness and love which I feel to my wife Arja and my sons Matias, Tuomas and Mikael.

All in all, the time from year 2009 has been very fun, which is the main motivator to make something new.

Nurmo, Finland, 1st June, 2016

Veli-Tapani Peltoketo

Contents

PREFACE	VII
1 INTRODUCTION	1
1.1 Background and motivation.....	1
1.2 Objectives and contributions	2
1.3 Methods	3
1.4 Structure of thesis	4
2 PRINCIPLES OF MODERN MOBILE PHONE CAMERA	5
2.1 Glance at history of photography.....	5
2.1.1 Film era	5
2.1.2 Digitalization.....	6
2.1.3 From CCD to CMOS	7
2.1.4 Mobile phone cameras	8
2.2 Generic structure of mobile phone camera	9
2.2.1 Image sensor	10
2.2.2 Camera module	11
2.2.3 Image processing pipeline.....	12
2.3 Future trends	13
2.3.1 Sensor innovations	13
2.3.2 New steps in lens systems.....	14
2.3.3 From one sensor to sixteen.....	15
3 IMAGE QUALITY, DISTORTIONS AND ARTEFACTS OF MODERN DIGITAL CAMERA	16
3.1 Image quality – problematic abstract.....	17
3.2 Image quality entities.....	20
3.2.1 Resolution	21
3.2.2 Color accuracy	22
3.2.3 Dynamic range	24
3.2.4 ISO speed	25
3.2.5 Image processing	26
3.2.6 Summary of image quality entities	26
3.3 Artefacts of digital imaging	27
3.3.1 Sensor based artefacts	28
3.3.1.1 Fixed pattern noise.....	28
3.3.1.2 Temporal noise	29
3.3.1.3 Banding.....	30
3.3.1.4 Green imbalance	31
3.3.1.5 Moiré	32
3.3.1.6 Blooming	32
3.3.1.7 Black sun	32
3.3.1.8 Rolling shutter	33
3.3.2 Camera module based artefacts.....	35
3.3.2.1 Lens aberrations.....	35

	3.3.2.2	Defocus	38
	3.3.2.3	Vignetting.....	39
	3.3.2.4	Color shading	39
	3.3.2.5	Short focal length issues.....	39
	3.3.2.6	Other lens artefacts.....	40
	3.3.3	Image processing pipeline based artefacts.....	40
	3.3.3.1	Compression.....	41
	3.3.3.2	Color inaccuracy	42
	3.3.3.3	Sharpening artefacts	42
	3.3.3.4	Noise removal artefacts.....	43
	3.3.3.5	Demosaicing.....	44
	3.3.3.6	Over processed images.....	45
	3.3.4	Summary of digital imaging artefacts	45
3.4		Video quality and artefacts.....	47
3.5		Is camera performance part of image quality?	48
4		IMAGE QUALITY MEASUREMENT METHODS AND METRICS OF MOBILE PHONE CAMERAS	51
4.1		Standardization and current tools	52
4.2		Traditional objective quality metrics.....	53
	4.2.1	Color measurements	53
	4.2.2	Noise measurements	56
	4.2.3	Dynamic range measurements.....	58
	4.2.4	Resolution measurements.....	59
4.3		Metrics for image quality artefacts.....	63
	4.3.1	Lens distortions	63
	4.3.2	Vignetting and color shading.....	65
	4.3.3	Flare and blooming.....	66
	4.3.4	Other artefacts.....	66
4.4		From objective to subjective metrics.....	66
4.5		New features and algorithms require new metrics	69
4.6		Video metrics	69
4.7		Performance metrics.....	70
5		FROM MEASUREMENTS TO BENCHMARKING	73
5.1		Benchmarking in general.....	73
5.2		Existing benchmarking metrics for digital cameras	74
5.3		Challenges of camera benchmarking	75
	5.3.1	Which metrics to select.....	76
	5.3.2	Metrics of different environments	77
	5.3.3	Perceptual benchmarking	78
	5.3.4	Several metrics to single score	79
	5.3.5	Practical issues of benchmarking	79
	5.3.6	Static benchmarking, compatibility requirement or trap? ..	80
5.4		Proposal for mobile phone camera benchmarking	81
6		INTRODUCTION TO ORIGINAL PUBLICATIONS.....	85

6.1	Article I: Objective verification of audio-video synchronization.....	85
6.2	Article II: Mobile phone camera benchmarking – Combination of camera speed and image quality	86
6.3	Article III: Evaluation of mobile phone camera benchmarking using objective camera speed and image quality metrics.....	87
6.4	Article IV: Mobile phone camera benchmarking in low light environment	88
6.5	Article V: SNR and visual noise of mobile phone cameras	92
7	CONCLUSIONS, DISCUSSION AND FUTURE.....	94
7.1	Future	96
	REFERENCES.....	98
	REPRINTS OF PUBLICATIONS.....	110

Figures

Figure 1	History of photography a) The first photograph, captured by Joseph Nicéphore Niépce in 1826 or 1827 and b) Boulevard du Temple by Louis-Jaques-Mandé Daguerre in 1838	5
Figure 2	Probably the first digital camera. (Photograph by Eastman Kodak)..	7
Figure 3	CMOS sensor a) Side view, picture by Samsung and b) Nayer filter, picture by Adimec	10
Figure 4	Camera module of modern mobile phone: a) Simplified example and b) Camera module of Lumia 1020 by Microsoft.....	12
Figure 5	Example of an image processing pipeline.....	13
Figure 6	Color differences between mobile phone cameras.	24
Figure 7	Noise in a picture captured in 30 lux	29
Figure 8	A maze pattern caused by green imbalance artefact	31
Figure 9	Black sun image artefact in the video stream of IAAF World Championships in Beijing 2015 (Youtube)	33
Figure 10	The principle of the rolling shutter defect. The image is based on paper by Sun et al 2012.....	34
Figure 11	Monochromatic aberrations a) Spherical aberration, b) Positive coma, c) Astigmatism, and d) Field curvature (Hecht 2002; Kingslake 1992)	36
Figure 12	Distortions a) Pincushion, b) Barrel, and c) Moustache	37
Figure 13	Chromatic aberrations a) Axial and b) Lateral.....	37
Figure 14	Image processing pipeline example: a) RAW image from sensor and b) Processed image.....	41
Figure 15	Sharpening artefacts	43
Figure 16	Noise removal artefacts, blurring and blockiness: a) Original scene and b) Aggressive denoising	44
Figure 17	Macbeth color chart	54
Figure 18	ISO 15739:2013 noise chart (Danes Picta)	57
Figure 19	MTF curves of three mobile phones captured from a low contrast slanted edge chart: (a) Very discreet sharpening, 8 mega pixels, (b) Over sharpening, 13 mega pixels (c) Poor resolution performance, 20 megapixels.	61
Figure 20	Examples of the resolution test charts: (a) High contrast slanted edge, (b) Low contrast slanted edge, (c) Detail of sinusoidal Siemens star and (d) Colored dead leaves. The image is based on paper by Peltoketo 2014.....	62
Figure 21	Lateral chromatic aberration and geometric distortion in a dot test chart.....	64
Figure 22	Hyperbolic zone plates of ISO 12233:2000 test chart.	68
Figure 23	Test scene of benchmarking.....	87
Figure 24	Measured devices in speed-quality coordinate system	88
Figure 25	Mobile phone camera parametrization in different illumination environments: a) ISO speed and b) Exposure time.....	89
Figure 26	Quality and speed metrics in different illumination environments: a) Spatial resolution and b) Focus time.....	90

Figure 27	Benchmarking in 30 and 1000 lux illumination environments: a) Speed score and b) Quality score	91
Figure 28	SNR based noise and visual noise in different illumination environments: a) 1000 lux and b) 30 lux.....	93

Tables

Table 1. Summary and a short description of image quality entities	26
Table 2. Summary and a short description of sensor based artefacts in digital imaging	45
Table 3. Summary and a short description of camera module based artefacts in digital imaging.....	46
Table 4. Summary and a short description of image processing pipeline based artefacts in digital imaging	46

Terms and abbreviations

3D	Three Dimensions
ANSI	American National Standards Institute
API	Application Programming Interface
BAPCo	Business Applications Performance Corporation
Bayer filter	Color filter array for red, green, and blue pixels
BM3D	Block-Matching and 3D filtering
Bokeh	The way the lens renders out-of-focus points of light
BSI	Back Side Illumination
CCD	Charge-Coupled Device
CFA	Color Filter Array
CIE	International Commission of Illuminance
CIEDE	International Commission of Illuminance, Delta E
CIPA	Camera & Imaging Products Association
CMOS	Complementary Metal Oxide Silicon
CPIQ	Camera Phone Image Quality
CTA	Consumer Technology Association
DCT	Discrete Cosine Transform
DSLR	Digital Single-Lens Reflex camera
DSNU	Dark Signal Non Uniformity
DSP	Digital Signal Processor
EBU	European Broadcast Union
EEMBC	Embedded Microprocessor Benchmark Consortium
Euro NCAP	The European New Car Assessment Programme
FPGA	Field-Programmable Gate Array
FPN	Fixed Pattern Noise
FR	Full Reference
GFS	Glare Spread Function
GPU	Graphics Processing Unit
H.26x	Series of video coding standards
HDR	High Dynamic Range
HVS	Human Vision System
I3A	International Imaging Industry Association
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
ISP	Image Signal Processor
ITU-T	International Telecommunication Union-Telecommunication
JND	Just Noticeable Difference
JPEG	Joint Photographic Experts Group
LCD	Lateral Chromatic Displacement
LGD	Local Geometric Distortion
MEMS	Micro Electro-Mechanical System

MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MTF	Modulation Transfer Function
NR	No Reference
OEFC	Opto-Electronic Conversion Function
OIS	Optical Image Stabilizer
P1858	IEEE Work group for Camera Phone Image Quality (CPIQ) work
PRNU	Photo Response Non Uniformity
RAW	Raw image data without image processing
RR	Reduced Reference
S-CIELAB	CIELAB with spatial extension
SMI	Sensitivity Metamerism Index
SMIA	Standard Mobile Imaging Architecture
SPEC	Standard Performance Evaluation Corporation
TC	Technical Committee
TOF	Time Of Flight
TPC	Transaction Processing Performance Council
VCM	Voice Coil Motor
VGI	Veiling Glare Index
VQEG	Video Quality Experts Group
WDR	Wide Dynamic Range
WoWCA	Workshop on Wireless Communication and Applications

Equations

(1)	38
(2)	38
(3)	39
(4)	54
(5)	55
(6)	56
(7)	65
(8)	83
(9)	83
(10)	83

List of publications

This thesis consists of a literature review of image quality issues and metrics, an introductory section about digital camera benchmarking and five published articles. The bibliographic data of the articles reprinted in this thesis are as follows:

- I. Peltoketo, V-T. (2012). Objective Verification of Audio-Video Synchronization. The 3rd Workshop on Wireless Communication and Applications, WoWCA.
- II. Peltoketo, V-T. (2014). Mobile phone camera benchmarking – Combination of camera speed and image quality. Proceedings of the Electronic Imaging conference. In *Image Quality and System Performance XI*. San Francisco, USA: SPIE 9016. DOI: 10.1117/12.2034348.
- III. Peltoketo, V-T. (2014). Evaluation of mobile phone camera benchmarking using objective camera speed and image quality metrics. In *Journal of Electronic Imaging*. 23:6. DOI: 10.1117/1.JEI.23.6.061102.
- IV. Peltoketo, V-T. (2015). Mobile phone camera benchmarking in low light environment. In *Image Quality and System Performance XII*. San Francisco, USA: SPIE 9396. DOI: 10.1117/12.2075630.
- V. Peltoketo, V-T. (2015). SNR and Visual Noise of Mobile Phone Cameras. In *Journal of Imaging Science and Technology*. 59:1. DOI: 10.2352/J.ImagingSci.Technol.2015.59.1.010401.

The articles are reprinted unchanged at the end of this thesis starting from page 110. The chapters of the thesis with the reprinted articles constitute an entity which defines the contributions of the work.

The included articles start at the following pages of this thesis:

Article I.....	110
Article II.....	116
Article III.....	125
Article IV.....	132
Article V.....	142

1 INTRODUCTION

The huge success of the mobile phone industry has pushed the use of digital imaging toward a skyrocketing growth. In year 2010, YouTube estimated that during every minute, 20 hours of video was uploaded to their server. Samsung estimates that in year 2015 880 billion digital images will be captured. The rate of still image capturing and video recording is not decreasing, rather vice versa.

Now that digital imaging has become an ordinary event for storing and sharing moments of everyday life, the importance of image quality and thereby the importance of camera quality has increased. Several new standards and image quality measurement methods have been published during recent years and old standards are continuously being updated. Still, each standard measures and validates a certain quality feature of a camera system and there is a lack of generic metrics to validate a whole camera system. Benchmarking is an approach to validate and compare whole camera systems and to help an end user to select the best camera for his or her purposes. If benchmarking is an independent metric, it could be used by mobile phone operators and vendors to advertise their products.

The thesis contains a comprehensive introduction to image quality factors, image quality distortions and artefacts which are common in digital imaging. In addition, performance issues like delays and slowness of the camera system are investigated. Correspondingly, quality metrics and methods, which are used to validate the quality of the digital cameras are described.

The work concentrates on benchmarking of mobile phone cameras and for that introduces a novel solution. The benchmarking includes different image quality metrics, performance metrics and methods that are used to create a straightforward score for comparing cameras. Also environmental factors are considered. The thesis is mainly focused on still image quality and performance though some video related metrics are used.

1.1 Background and motivation

The origins of the thesis are partially related to the work of the author in the Finnish startup company Sofica. The first product from the company was an automated test system for digital cameras and a logical step from this testing and measuring was a comparison of camera devices. At that time the mobile phone vendors were deeply involved in the mega pixel competition of mobile phone cameras and it was reasonable to expect that the use of mobile phone cameras would dramatically expand.

Though there were products in the markets that measured image quality, it was surprising that there were no companies, tools or standards making more comprehensive ranking or providing benchmarking of mobile phone cameras. The gap between the expectations of mobile phone cameras and the impossibility of properly comparing cameras encouraged the author to start developing a benchmarking system for mobile phone cameras.

Although this work is partly based on work of the author in the company, the proposed solution for the benchmarking is not a product of the company.

1.2 Objectives and contributions

In general, the objective of the work is to introduce a comprehensive benchmarking system which can be used to compare and rank mobile phone cameras. Since benchmarking is a combination of numerous image quality and camera performance factors, the work required a significant effort to inspect and validate the different quality elements of a mobile phone camera.

During the research following research questions emerged:

1. Which requirements should a comprehensive benchmark system of mobile phone cameras fulfill?
2. Which metrics should be included in a benchmarking system?
3. How should different environmental factors be taken into account in a benchmarking system?
4. How would the evolution of digital cameras, algorithms and testing methods affect the benchmarking system?

The research questions were answered during the years of the dissertation work. Certainly, numerous new questions were raised during the work and it is difficult or even impossible to give comprehensive answers to all the research questions. Still the work defines extensively the diversity and complexity of camera quality factors, considers the requirements of a benchmarking system and finally, introduces a solution for benchmarking mobile phone cameras. The thesis is a workflow of the investigations, considerations, trials and conclusions required to find answers to the research questions.

The main tasks and contributions of the thesis are:

- To create a comprehensive summary of image quality factors, image quality distortions, artefacts and performance issues, which are related to digital imaging,

- to collect and validate image quality metrics and methods available in different standards, papers and literature,
- to inspect the requirements a comprehensive benchmark system should fulfill, and
- to introduce a solution for a generic and public benchmarking method.

The published articles attached at the end of the work and the content of the thesis create an entity which answers the research questions, defines the tasks made during the work and finally, constitutes the contributions of the dissertation.

1.3 Methods

In general, evaluation of digital cameras can be divided to two main methods: objective and perceptual i.e. subjective methods. The objective methods are traditionally related to measurements and statistical analysis of image data whereas the perceptual methods use observers, which validate the quality and functionality of images and cameras.

Characteristics of objective methods enable to use automated measurements and calculations which make this approach very efficient. On the other hand, the correlation between objective methods and human inspection is not always good enough. For this reason, perceptual methods are used. However, perceptual methods require a significant amount of human work and make this method inefficient and time consuming.

To combine the pros of both methods, conversion algorithms have been built to use efficient objective methods and convert, if required, results to perceptual ones. This is nowadays one of the main research area especially in image quality inspection.

Furthermore, evaluation of digital cameras can be divided according to the existence of original data. No-reference, reduced-reference and full-reference methods can be used. The full-reference method uses the original data, i.e. original image of a scene, whereas no-reference method has no information about original scene. The reduced-reference method uses certain pre-calculated characteristics of original data and compares them to corresponding ones of captured data.

This thesis is primarily based on objective quality and performance methods which are used by an automated measurement system. However, conversion algorithms have been used to certain image quality metrics to achieve better correlation towards perceptual inspection.

1.4 Structure of thesis

The thesis consists of an introductory section on image quality in general and different metrics, measurements and artefacts which could be related to mobile phone cameras. Moreover, benchmarking challenges are described and discussed. Finally, five publications are reprinted in their original form at the end of this thesis to describe the research into benchmarking of mobile phone cameras.

The first chapter introduces briefly the topic of the thesis, describes the background and motivation and specifies the research questions.

Chapter two describes the principles of a modern mobile phone camera starting from the early history of photography and following the technology steps through to recent models of mobile phone cameras. The generic structure of a mobile phone camera is discussed as well as future trends.

Chapter three concentrates on image quality features, distortions of image quality and quality artefacts. The chapter includes a broad literature review of image quality features like color, resolution, dynamic range and ISO speed. Also different image quality and video artefacts are described. Finally the chapter considers whether the performance of a camera should be part of the image quality.

The fourth chapter, Image quality measurement methods and metrics of modern mobile phone camera, defines how the features and artefacts of chapter three can be measured and which kind of metrics can be used. The chapter includes a view of current standards and tools. It also defines needs for new metrics due to new features of cameras.

Chapter five includes the challenges faced when individual metrics are combined into a benchmarking score. The chapter defines the tasks for suitable metric selection, environmental factors of a benchmarking, and how to create a benchmarking system when the features and requirements of mobile phone cameras are changing. Finally, the chapter introduces a solution for mobile camera benchmarking.

The sixth chapter, Introduction to the original publications, includes short summaries of the attached articles, plus the main objectives, tasks and results of each individual item of work.

Conclusions of the study and this thesis are finally drawn in chapter seven. The articles are reprinted unchanged at the end of the work.

2 PRINCIPLES OF MODERN MOBILE PHONE CAMERA

2.1 Glance at history of photography

2.1.1 Film era

Looking back at history gives a perspective on today's research. The oldest photograph, which has survived up to present times is an image captured by Joseph Nicéphore Niépce in 1826 or 1827 in the Burgundy region of France. The image is shown in Figure 1. Niépce used a pewter plate covered by bitumen in a camera obscura. The exposure time was at least eight hours. After the exposure, Niépce washed the plate using a mixture of oil of lavender and white petroleum and removed the bitumen which was not hardened by light. Thus the first image was a direct positive picture. The original pewter plate is held at University of Texas, Austin. (University of Texas; Tom A. 2014; Peres, M. 2007)

Niépce continued his photography development with Louis-Jaques-Mandé Daguerre and they created a method using a copper plate covered by silver and iodine. Daguerre managed to improve the method using mercury and finally captured very detailed pictures as shown in Figure 1b. (Tom A. 2014; Peres, M. 2007)



(a)

(b)

Figure 1 History of photography a) The first photograph, captured by Joseph Nicéphore Niépce in 1826 or 1827 and b) Boulevard du Temple by Louis-Jaques-Mandé Daguerre in 1838

The first real manufactured camera was built by Alphonse Giroux, who got a license from Daguerre and Niépce's son to use their technique in this camera. The camera was made using wooden boxes, and included a real 380 mm objective having an f-number between 14 and 15. The focus was adjusted by sliding the inner part of the camera in which the photography plate was mounted. Even though the price of the camera was notable high, 400 francs, the camera was still a great success. (Tom A. 2014)

At the same time British chemist, Henry Fox Talbot created a method using silver nitrate and captured the first negative images in 1835. He developed also a process where positive images were created from negative ones. The negative/positive method based on work of Talbot, dominated the photography industry more than 150 years. (Tom A. 2014; Peres, M. 2007)

Photography was an instant success, cameras were spread around the world and new inventions were made all the time. New film materials, new camera techniques and even 3D cameras were implemented. Wilhelm Rollman created a 3D camera technique as early as 1852 and the same technique is still in use today (Tom A. 2014). Finally, based on Hannibal Godwin's work, George Eastman created a roll film in 1888 and the regime of modern film based photography started. (Tom A. 2014; Peres, M. 2007)

It seems shameful to bypass the golden times of film photography but because it is not the topic of this work and it would require several books to highlight the importance of that time, we have to step directly into the late 20th century and into the era of the first digital cameras.

2.1.2 Digitalization

It is symbolic that the first known digital image was not captured from a real world scene but from a readymade photograph captured using an analog film. Russel Kirsch made a digital image which was scanned from a photograph of his son in 1957. The size of the image was 176x176 pixels. Probably the first digital camera was developed by Steve Sasson, an engineer of Eastman Kodak in 1975. A charge-coupled device (CCD) camera with 10 000 pixels was mounted on several circuit boards and the result was stored on a cassette tape as shown in Figure 2. (PetaPixel)

The CCD was invented by George E. Smith and Willard Boyle in 1969. The invention was the basis of modern digital imaging and CCD sensors are still used in astronomy and scientific imaging due to their superior noise characteristics. The

importance and revolutionary impact of the invention was highlighted by the Nobel Prize awarded to Smith and Boyle in 2009. (Nobel Prize)

The first digital cameras were published during the 1990-2000 decade. The first camera for consumer use was the Apple QuickTake 100 with 640x480 resolution. This camera was produced by Apple, though it was designed by Kodak. (Imaging resource)



Figure 2 Probably the first digital camera. (Photograph by Eastman Kodak)

2.1.3 From CCD to CMOS

In 1968, a year before CCD was invented, a complementary metal oxide silicon (CMOS) sensor was also published. However, CMOS based sensors suffered from poor fabrication processes and the fixed pattern noise of these sensors was extremely high (Wang 2008). It was more than 25 years before CMOS sensors were improved so much they could seriously thread the dominance of the CCD sensors. In year 1995, a CMOS sensor based camera-on-chip solution was published. The same chip contained several features like timing, control block, sampling and noise suppression logic (Nixon et al. 1995). The low power consumption of the CMOS

sensor and the potential to add more logic to the sensor chip and therefore lower the costs of the chip pushed the use of CMOS images to a rapid growth. According to the latest Yole's report, CMOS image sensor revenues were bypassing the revenue of CCD sensors in 2010 (Yole 2015).

The second significant performance step of CMOS cameras was the invention of the back side illumination (BSI) technique. Earlier, the photo sensitive elements, photodiodes, were located at the bottom of the chip whereas all the wirings were between the light and photodiodes. Swapping the chip upside down, the photodiodes were located at the right side, i.e. from where the light was coming. This technique increased significantly the amount of photons hitting the photodiodes and therefore also increasing the quantum efficiency of the sensor. Sony published the Exmor R sensor family in 2008 which included the first BSI sensors. Five years later, Yole reported that the revenue of BSI based CMOS sensors was more than 50% of all CMOS image sensors (Yole 2015).

During the development period of CMOS sensors, pixel sizes decreased and this enabled a higher and higher pixel count. When the sensor by Nixon et al. had pixel size of 20 μm , the latest CMOS sensors have reached 1 μm pixel size.

Low cost and low power consumption has made the CMOS technique very suitable for mobile phone cameras. Due to continuous research and new inventions, the image quality of CMOS sensors has also reached the quality of CCD sensors.

2.1.4 Mobile phone cameras

The first prototype of a mobile phone camera was introduced in Telecom 95 by Panasonic. Depending on the references, the honor of being the first commercial mobile phone camera goes to the Sharp Corporation J-SH04 model or to the Samsung SCH-V200 model. The former had a 0.1 mega pixel CMOS sensor and the latter a 0.35 megapixel CCD sensor and they were both launched in 2000. However, the first picture captured by a mobile phone and shared using the phone was taken in 1997, when Philippe Kahn captured an image of his newborn baby using a camera integrated into his phone. (EETimes; Sharp; Samsung).

Since then the evolution of the mobile phone camera has been breathtaking fast. The pixel count is only one feature of a phone, though the development of this feature gives a useful overview of the mobile phone camera development:

- 1.3 megapixel camera released in 2003 by Sprint, model PM8920, the same year that Sony Ericsson released the first phone model, Z1010 with a front face camera.

- 2.0 megapixel in 2004 by Nokia N90
- 3.2 megapixel in 2006 by Sony Ericsson K800i model
- 5 megapixels in 2007 by Nokia N95
- 8 megapixels in 2008 by Samsung i8510
- 12 megapixels in 2009 by Samsung M8910
- 13 megapixels in 2010 by Sony Ericsson S006

Finally, to underline the madness of the megapixel race, Nokia released a 41 megapixel camera in 2012, the Nokia 808 PureView model. (Digital Trends)

Obviously, pixel count is not the only feature revolutionized during the recent years. The latest mobile phone cameras may include optical image stabilizer, sensor based autofocus, new color filters, multiple cameras, a global shutter, a high dynamic range and several other new features. Especially, recent mobile phone cameras have new image processing algorithms making images even better.

Yole forecasts that in year 2015 the revenue of CMOS image sensors will be 10 billion dollars, and 60% of this revenue will come from mobile devices. (Yole 2015)

There has been a huge advance in photography since the days of Niépce. However, history seems to repeat itself, and direct positive images are being captured once again.

2.2 Generic structure of mobile phone camera

In general a mobile phone camera can be divided into three logical parts: the sensor itself, the camera module, the image processing pipeline, or image signal processor (ISP), and the flash system.

The quality and benchmarking of the flash system is not part of this work. When the flash system is used, it generates a whole new dimension to still imaging. A proper investigation of a camera system with flash would require several new measurements like color temperature, uniformity, and magnitude of the flash system as well as several different environment should be noted. Even if the flash system is nowadays an essential part of mobile phone cameras, the evaluation of the flash would complicate benchmarking significantly and should be investigated in a different research.

2.2.1 Image sensor

An image sensor is an essential part of a camera system. It gets light through the lens system and transforms light first into analog signal and afterwards into digital numbers. Since CMOS sensors dominate mobile phone cameras, this section concentrates on CMOS technology. Figure 3a shows the simplified inner structure of a CMOS sensor. The example is from Samsung ISOCELL technology, where photodiodes are isolated from each other (Samsung ISOCELL).

The topmost element of the sensor is a micro lens, which collects light and bends it onto a pixel below. Use of a micro lens reduces optical crosstalk and allows use of a wider field of view in a camera system. A color filter array (CFA) below the micro lenses filters the light into different components. Usually a Bayer filter with red, green and blue filters is used (Peres, M. 2007). Without the CFA, the sensor would take monochromatic images.

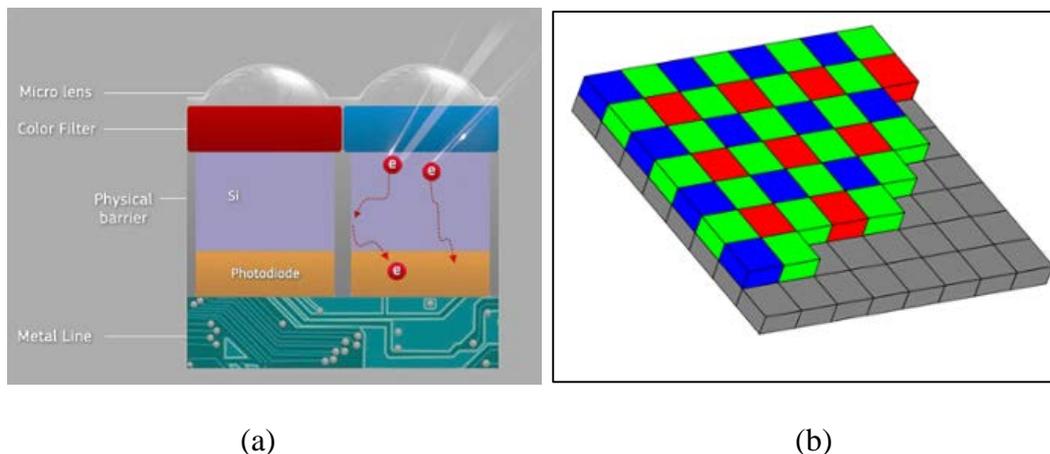


Figure 3 CMOS sensor a) Side view, picture by Samsung and b) Nayer filter, picture by Adimec

Figure 3b shows an example of the Bayer filter. The number of green pixels is double relative to other colors, and this correlates with the color sensitivity of the human vision system (HVS). Obviously, each color filter will absorb part of the incoming photons and will therefore decrease the quantum efficiency of the sensor. Several different studies are ongoing to replace the technique, but currently the Bayer filter is the main method (Business Wire; Sony; Invisage; Foveon).

When a photon hits to the silicon below the color filter array, it creates an electron-hole pair which can be electrically detected. To eliminate an electron leak between pixels i.e. electronic crosstalk, Samsung with other sensor vendors has made boundaries between pixels. Samsung calls this method the ISOCELL technique.

CMOS pixels are active pixels i.e. each pixel has its amplifier. Until now, the voltages of each pixel are read line by line, converted by an analog to digital converter and sent to the image processing pipeline. However, this rolling shutter method has weaknesses. When rows are read at different time, fast moving objects are distorted in the final image. Due to this, several global shutter CMOS sensors have been recently published (Sony IMX174LLJ, CMOSIS Global Shutter). The global shutter method requires more logic per pixel. While the first CMOS pixels included three transistors, a global shutter version now requires at least five.

Finally, the bottom level of the sensor contains metal wirings which transfer the information from a pixel.

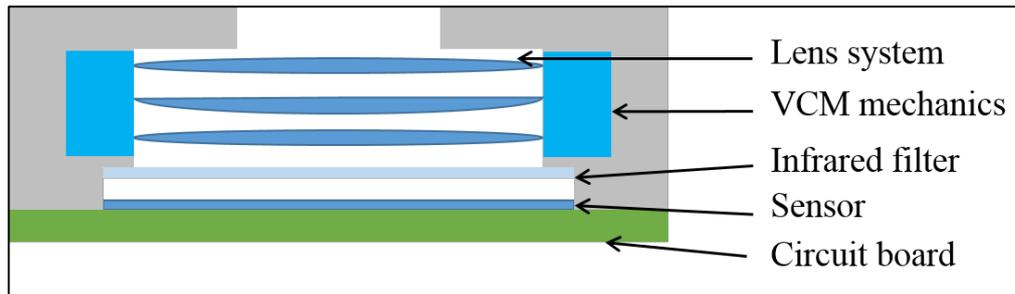
2.2.2 Camera module

A camera module packages the image sensor with the lens system and with mechanical parts which are required for features like auto focus, optical image stabilizer and aperture adjustments. It is also possible to integrate a digital signal processor into the camera module.

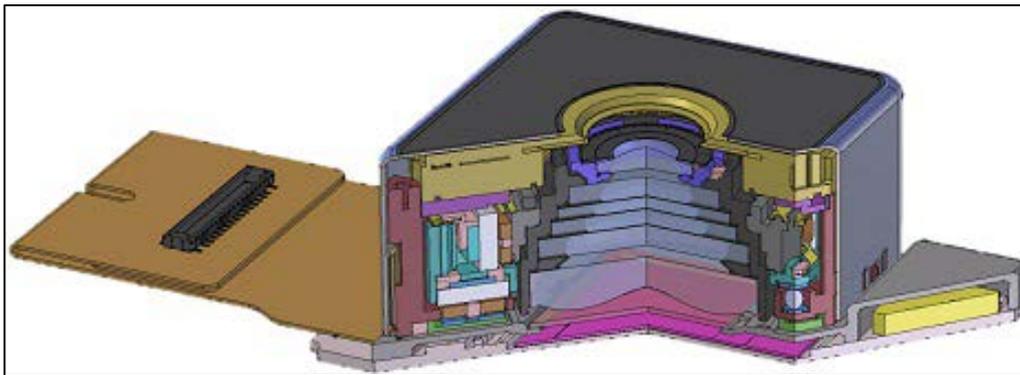
Figure 4a shows a simplified example of the camera module. Firstly, the package contains a lens system, nowadays mobile phone cameras with auto focus have 5-6 lens components. Secondly, the moving lens components have their own holders and controllers. Voice coil motor (VCM) is a widely used technique to adjust lenses, but new methods like micro electro-mechanical systems (MEMS) are coming to the markets. Thirdly, an infrared filter is mounted on top of the sensor to prevent saturation due to infrared light.

Finally, the sensor is wired and mounted onto a circuit board and the whole system is protected by a package. The module offers a connector which enables control of the camera and transfer of the image data.

Probably the most complicated mobile phone camera module, the camera module of Lumia 1020 phone is shown in Figure 4b. Among others, it includes a 41 mega pixel sensor, VCM based autofocus and optical image stabilizer where the whole lens system is resting on ball bearings. The size of the package is 25mm by 17mm and it contains over 130 individual components. (Microsoft)



(a)



(b)

Figure 4 Camera module of modern mobile phone: a) Simplified example and b) Camera module of Lumia 1020 by Microsoft

2.2.3 *Image processing pipeline*

An image processing pipeline has a significant role in modern mobile phone cameras. Unfortunately, the quality of an image without image processing (RAW image) is quite poor due to a small lens system, small pixel size and sensor artefacts. In practice, the image processing pipeline recreates the image using a large number of different algorithms.

The image processing pipeline can be implemented by a specific processor, digital signal processor (DSP) or graphics processing unit (GPU). Also field-programmable gate array (FPGA) are used in some cases. Moreover, the pipeline can be implemented in software and using the application processor of the phone. However, the image processing pipeline tends to be such a heavy process that it usually executes on a separate processor or chip.

Figure 5 gives an example of algorithms that the image processing pipeline may contain. The process can be divided into correction, conversion and controlling tasks, like denoising, demosaicing and auto focus correspondingly. The algorithms have many connections between each other and the actions of one quality algorithm may reduce quality of another feature. The parameterization of the algorithms is a trade-off between different quality features.

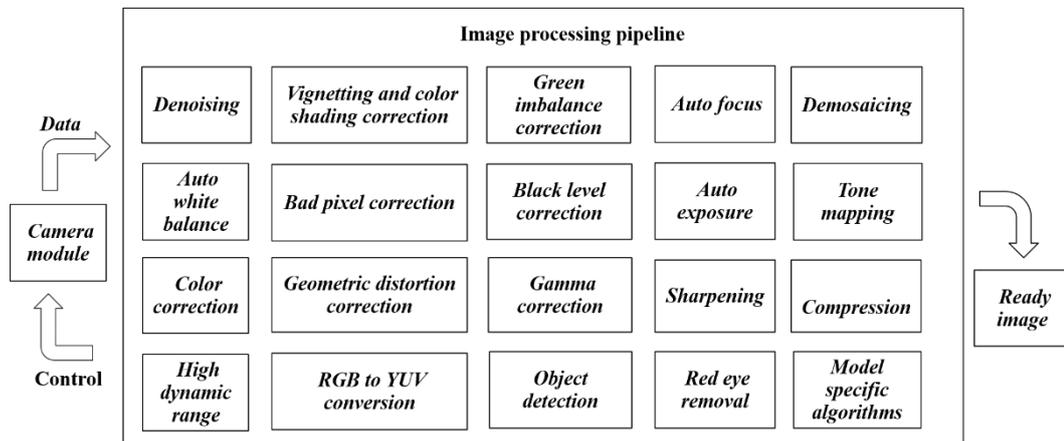


Figure 5 Example of an image processing pipeline

Auto focus and auto exposure especially, have critical roles because they control the camera functionality and they are very time critical processes. All in all the quality of the image processing pipeline defines largely the quality of the whole camera system.

2.3 Future trends

The future of digital imaging looks bright. Not only because of its own great success but due to the huge amount of new innovations. New methods and approaches in camera sensors and camera modules force engineers to implement new image processing algorithms that can comprehensively utilize new features. This section defines some of the trends, which can change the way images are captured and video is recorded.

2.3.1 Sensor innovations

In the sensor area alone, there are tens of different new methods which challenge current Bayer type sensors. Aptina and Sony have introduced their clear pixel sensors which have even replaced green pixels with white pixels, like Aptina or added extra white pixels to existing filters, like Sony (Business Wire, Sony). The

use of white i.e. unfiltered pixels increases the sensitivity of sensors. Sony has also published a patent which defines triangular and hexagonal pixels with seven different pixel types (Patent US 20130153748 A1).

Another approach is a quantum film invented by InVisage. The quantum film is a photosensitive layer which may replace the silicon from traditional sensors (Invisage). InVisage published a quantum film sensor with 13 mega pixels in late 2015. The sensor should provide a dynamic range which is three f-stops better than conventional CMOS sensors. Moreover, Foveon has published its X3 sensor with stacked photodiodes. The sensor is based on the fact that light with longer wavelengths penetrates silicon more deeply than light with shorter wavelengths. Using this phenomenon, Foveon has implemented a layered pixel, where blue light is detected on the top of the pixel, green in the middle and red wavelengths at the base of the pixel (Foveon). Obviously this kind of sensor does not need a color filter array at all and should be more sensitive than sensors which are using one.

On the other hand, Xerox PARC and IMEC develop multispectral and hyperspectral sensors mainly for industrial use, but they could also add interesting features to mobile phone cameras (GlobeNewswire, IMEC). Finally, very recently Panasonic published an organic CMOS sensor which dynamic range should be significantly better than any other conventional sensor (Panasonic).

2.3.2 New steps in lens systems

Sensors are not the only area, where innovations of new techniques occurs. In optics, LensVector has released a liquid lens. A single lens component contains electrically controlled liquid crystals and the focus can be adjusted not by moving the lens but controlling the crystals which makes the focus adjustments very fast (LensVector). On the other hand, the micro electro-mechanical system technology (MEMS) has superior performance features over current voice coil motor (VCM) methods, but manufacturing problems still prevent the approach from reaching greater success (DigitalOptics). Rambus has a technology called lensless smart sensor, which includes a spiral grating of diffractive optics and sophisticated algorithms to capture an image without lenses (Rambus<). Finally, Sony has released a market ready product with an optical variable low pass filter, where a user may control the filter to find the balance between resolution and aliasing artefacts like Moiré (Sony Optics).

2.3.3 *From one sensor to sixteen*

Multiple cameras and array imaging are related to three dimensional (3D) imaging but there are also other features, which can be made using multiple sensors. Pelican Imaging has introduced a compact sensor matrix containing sixteen sensors mainly targeting 3D imaging. However the technique offers also high resolution imaging by combining the information from the multiple sensors and post-capture refocus for still images and videos (Pelican Imaging). Altek has made a system containing two 13 mega pixel sensors where one is chromatic, the other is monochromatic. Altek advertise this as instant auto focus, high resolution, good low light performance with low noise and high dynamic range (Altek).

Light co. released in year 2015 perhaps the largest array imaging product: Their L16 camera with sixteen 13 mega pixel camera modules using three different focal length (Light). The product may challenge traditional digital single-lens reflex (DSLR) cameras. Also several time-of-flight (TOF) solutions are developed for 3D and depth imaging (Lytro, Heptagon).

Finally, what is the role of presence capture cameras? These cameras will record the whole environment, capturing 360 degrees or even 720 degrees in 3D including surround sound. The end user will be able to re-experience the original moment in a very new and comprehensive way. However, the solution requires new infrastructures for cameras, data transfer and displays.

All in all, digital imaging is rapidly changing. New products with astonishing features are already available or just around the corner and the markets and end users will decide the next successful trend. The rate of digital camera evolution challenges image quality metrics and measurements, too. When new techniques are taken into use, they will generate new features to validate and new artefacts to measure.

3 IMAGE QUALITY, DISTORTIONS AND ARTEFACTS OF MODERN DIGITAL CAMERA

In a perfect world, a digital camera would reproduce exactly the photographed real world scene. The image would present all the smallest details, reproduce exact colors, without any noise and other artefacts, and in the case of consumer cameras, use the whole spectrum of the human eye. Also the dynamic range of the camera would be at least as good as the human eye and the image processing pipeline would mimic the brain's visual processing in a perfect way.

A quick glance at the endless image galleries of the Internet reveals that obviously we do not live in a perfect world. Limitations of cameras' hardware and software, manufacturing issues with camera sensors and lenses and problems in image processing pipelines cause different issues in images. The issues can be content destroying problems like out of focus adjustments and wrong exposure values or very small and even artistic faults like an unnatural bokeh or a slightly wrong color tint. Also nature itself creates final boundaries on image quality by limiting the performance of lenses and defining the smallest objects that can be observed using the wavelengths of the human vision system, for example.

When the concept of image quality is considered more closely, it can be seen problematic or even controversial. Though the image quality can be measured very comprehensively, in case of consumer products, images are ultimately judged by the human eye and by the human vision system. Although perceptual image quality and measured image quality correlates well, they definitely are not the same thing.

Strictly considering image quality as a measurable entity, image quality can be defined as an overall performance of the camera in reproducing the captured scene in an image. A quality distortion can be specified as a lack of performance and image artefacts are explicit errors in the images. However, image quality is not only an objective and measurable number, it is also a perceptual view of the image. There have been several attempts to bind objective and perceptual quality metrics together. For example Keelan defined a specific method and function, the integrated hyperbolic increment function (IIHF) to transform any objective metric into a perceptual one (Keelan 2002). Also many current image quality standards have been updated to measure the perceptual image quality, too.

This chapter defines the problematic concept of image quality in general. The content is not limited to mobile phone cameras, because the quality entities are generic to most of digital cameras. The chapter specifies the image quality entities, distortions of image quality and common artefacts of digital imaging. The purpose

of the chapter is not just to describe or list the quality issues and artefacts but highlight the diversity and number of quality problems in digital imaging and the challenges, which different issues present in quality measurement and benchmarking.

3.1 Image quality – problematic abstract

How can image quality be defined or quantified? The question is a fundamental one, when a camera system is investigated from the image quality point of view. The literature gives different approaches to the definition of image quality.

Keelan divides image quality to four attribute groups to help with the clarification of image quality (Keelan 2002):

- Artifactual attributes, like unsharpness and digital artefacts
- Preferential attributes, like color balance and contrast
- Aesthetic attributes, like composition
- Personal attributes, like how a person remembers certain cherished event

Obviously, artefactual attributes can be measured objectively by searching certain errors in captured images. Preferential attributes are still objectively measurable, but they also contain perceptual components like color saturation. Although aesthetic attributes are very perceptual or even personal attributes, still some evaluation can be made for example by investigating the usage of the golden ratio in captured images. Finally, personal attributes are so related to the history and emotions of a person that they cannot be measured by the image quality methods and rated as image quality attributes. However, personal attributes can be the most important factors when images are rated.

Specifically, Keelan defines image quality as follows: “The quality of an image is defined to be an impression of its merit or excellence, as perceived by an observer neither associated with the act of photography, nor closely involved with the subject matter depicted.” (Keelan 2002)

In his book, *Handbook of Image Quality*, Keelan defines an image quality unit, just noticeable difference (JND) to specify the smallest image quality difference which is noticeable to a human being. In practice, one JND is valid, if 75% of observers notices the difference (to get the specific definition of JND, see pages 35-45 from Keelan’s book). JNDs can be used separately for each quality attribute or a combination of attributes. Keelan defines also a method, where objective image quality measurement results can be transformed into JND units. (Keelan 2002)

Wang and Bovik concentrate strictly on objective image quality in their book *Modern Image Quality Assessment* (Wang and Bovik 2006). They specify a fundamental requirement of the image quality attribute: an image quality attribute is useless, if it does not correlate well with human subjectivity. Moreover, they define three uses for objective image quality measurements: They can be used to monitor the quality of the system, benchmark devices towards each other, and to optimize the camera system. (Wang and Bovik 2006)

Umbaugh defines a different objective image quality criteria in his book (Umbaugh 2005). The objective image quality is defined as an amount of error in a captured image compared with a known image, which is a logical approach. He defines several well-known statistical methods for the measurements: root mean square error, root mean square error signal to noise ratio and peak signal to noise ratio (Umbaugh 2005). Use of the equations reveals that the original images have to pre-exist and so called full-reference image quality method is used. In practice, the full-reference method is quite difficult to use in image quality measurements not only because the images are captured from a scene and exact reference image does not exist, but also because a modern image processing pipeline recreates the scene so fundamentally that a straight comparison at the pixel level is not sensible. In addition, measurements like mean square error do not always correlate with perceptual quality (Wang and Bovik 2009).

In the case of subjective image quality tests, Umbaugh relies on group of observers and how they rate the images. He divides the subjective image quality tests into three categories: an impairment test to rate images in terms of how bad they are, a quality test to rate how good they are, and a comparison test to evaluate images side by side. Surprisingly, he does not refer to known standards ITU-T P.800, ITU-T Rec. BT.500-11, and ITU-T Rec. P.910, which define very comprehensively the subjective image quality methods and environments.

According to the name of the book, *Perceptual Digital Imaging – Methods and Applications*, Lukac concentrates fully on subjective image quality (Lukac 2013). Like Wang and Bovik, Lukac divides subjective image quality into full reference (FR), reduced reference (RR), and no reference (NR) methods. However, it is notable that Lukac uses only FR and NR methods. The reduced reference method is completely omitted from the perceptual image quality assessments. The FR methods of the book are not based on the pixel level difference but more sophisticated algorithms like structural similarity and wavelet transform methods. On the other hand, the NR methods are extremely interesting ones as they evaluate the image without any information of the image content but fully rely on statistical analysis of the image data. The NR approach is recognized as Holy Grail of image

quality assessment. If it reaches only moderate reliability someday, it will revolutionize the whole area of image quality measurement. (Lukac 2013)

Finally, several standards define both objective and subjective image quality approaches. A mean opinion score (MOS) has been used to specify the subjective quality of images and videos. The origin of the MOS rating comes from telecommunications and quality observations of telephony networks. MOS has a five step validation for quality ranging from bad to excellent quality. MOS is an arithmetic mean of all scores given by observers. (ITU-T P.800) In addition, several perceptual video quality standards have been published by the International Telecommunication Union, Telecommunication Standardization Sector: ITU-T Rec. BT.500-11 and ITU-T Rec. P.910 in particular.

ISO standards specifically define several objective and also perceptual image quality methods for specific features of digital cameras. The methods are defined, for example, for features like color fidelity, noise and resolution. The quality entities and corresponding metrics of the standards are discussed later in the thesis.

As a summary, it can be said that division into subjective and objective image quality methods is widely accepted. Obviously perceptual or subjective image quality is the goal that should be pursued, because ultimately, consumer camera images are judged by the human vision system. However, there are several ways to measure subjective quality. One approach is to measure objective metrics and then convert results to perceptual ones (Keelan 2002). A group of observers can be used to rate images (ITU-T Rec. P.910). Also, image quality evaluation can mimic the human vision system and rate images accordingly (Wang and Bovik 2006). Finally, if the no-reference perceptual quality approach works reliably someday, it might replace all existing methods.

All methods have pros and cons. The objective measurements are easier and cheaper to make because they can be automated at least to some level, but they do not fully correlate with perceptual image quality even if conversion algorithms are used. The subjective measurements are definitely perceptual ones, but they are expensive and time consuming and the reliability of the measurements depends on the observers. A good example of a reliability problem of subjective measurements can be found in Winklers book *Digital video quality – vision models and metrics: Video Quality Experts Group (VQEG) ran several studies to find the best metric to measure subjective video quality. The methods were tested in a co-operation of several laboratories in identical environments. Finally, when the results were evaluated, it was noted that the test results between laboratories varied significantly (Winkler 2005).*

Moreover, subjective testing has always a variable called human being that may distort test results. Even though a large group of observers should reduce the effect of individuals, some collective phenomena can still happen. An example of a factor which may affect subjective image quality testing can be found in an article of *Current Biology* where it was noted that the human color perception may change between seasons (Welbourne et al. 2015). This kind of phenomenon may change the results of subjective image quality measurement.

As a conclusion it can be said that several different approaches have been developed in the image quality area. However, two main research paths can be derived from the numerous image quality books, articles and papers. Firstly, to find a reliable method for measuring the image quality from no-reference data and secondly, how to convert existing objective image quality metrics into perceptual ones.

The conversion between objective and perceptual metrics has been taken into account in this thesis. The latest color difference metrics as well as visual noise metrics are used in the benchmarking proposal of the research. Both the color difference and visual noise metrics represent the latest knowledge of objective image quality metric adjustment to perceptual one. However, the majority of metrics have been used in this thesis are still objective ones. Even if the conversion work is one of the main research path in image quality area, there are still comparably few acknowledged metrics which are acceptably converted.

Even if the no-reference methods are very interesting approach for image quality measurement, they are not mature enough to give comprehensive and reliable results. Therefore the methods are not used in this research.

3.2 Image quality entities

There are numerous image quality factors associated with modern digital cameras and each of them has some effect on the final quality. To manage the large number of factors, it is reasonable to make some classification. Keelan divides the device specific attributes into artefactual and preferential ones (Keelan 2002). An equivalent approach would be division to image quality artefacts and image quality performance of a camera system.

Image quality defines the ability of a camera system to produce high quality images whereas quality artefact defines an error which may limit and violate the image quality. This section defines image quality factors.

3.2.1 Resolution

When digital cameras and especially mobile phone cameras are advertised, the number of pixels seems to be the main attribute. This is understandable in advertising because a single number is easy to explain and it defines, at some level, the resolution of captured images.

Still, the number of pixels, even though it seems to be a very straightforward metric, can be noted in several different ways. According to the Camera & Imaging Products Association (CIPA) guideline, the term ‘number of effective pixels’ should be used when an image capture performance is clarified. Number of effective pixels is clearly a different metric to total number of pixels, because total number of pixels defines the maximum number of pixels in a camera sensor but number of effective pixels declares the number of pixels used to create an image. How can there be a difference between these metrics? For example, the mechanics of the camera system can be designed so that only part of the pixels receive light through the lens system. The Nokia 1020, of which the resolution was advertised as 41 mega pixels, the real maximum resolution of the image is 38.2 mega pixels or 33.6 mega pixels depending on the aspect ratio of the image (Nokia 2013).

However, there are several other factors which affect the final image resolution and pixel count is only one of them. Also the definition of resolution is not unambiguous as it can involve to some extent to the sharpness of the image.

According to the ISO 12233:2014 standard, the resolution is “an objective analytical measure of a digital capture device’s ability to maintain the optical contrast of modulation of increasingly finer spaced details in a scene.” Moreover, the sharpness, or acutance, is strictly separated from resolution and it is defined as the subjective impression of details and edges of the image. (ISO 12233 2014)

Like the ISO standard, DxO separates resolution and sharpness, too. According to the DxO, resolution defines the smallest detail a camera can separate while the definition of sharpness is identical to the ISO standard one. Moreover, DxO defines the acutance as an objective measure of sharpness. (DxO Sharpness)

In contrast, Imatest uses sharpness as a synonym for resolution defining it as the amount of details an imaging system can reproduce. (Imatest Sharpness)

As a summary, resolution can be defined as an objective metric which defines the level of details which a camera system may produce. Still, the factors of the resolution are not fully clarified.

The three main components of a camera system; camera module, sensor, and image processing pipeline have their own effects on resolution. Firstly, the lens system has a limiting resolution which can be smaller than the maximum resolution of the sensor. Moreover, the lens system has always aberrations which decrease the resolution. It is notable, that lens aberrations affect more areas far from the center of the lens (optical axis) and therefore corners and border area resolution of an image is usually poorer than the center area.

Secondly, the effective pixel count of the sensor limits the resolution. Even though the pixel count is the main characteristics of the sensor, artefacts like cross talk and noise reduces the maximum resolution. Thirdly, the image processing pipeline includes several algorithms that may affect the final resolution. Especially the autofocus algorithm has a crucial role when the final resolution is validated. If autofocus does not work correctly, the result is a blurry image whatever the resolution capabilities of other components. Moreover, algorithms like demosaicing, denoising and compression can be characterized as filtering algorithms which may filter out the smallest details from images. On the other hand, artificial sharpening algorithms may increase the subjective sharpness, even if they cannot improve objective resolution.

The final resolution of an image is definitely not the pixel count of the sensor but a combination of limiting the resolutions of each component of a camera system.

3.2.2 *Color accuracy*

The origins of color recreation in a digital camera are in camera sensor's color filter. A color filter array (CFA) filters the light on top of a monochromatic sensor and generates normally green, red and blue color channels and correspondingly colored pixels. A demosaicing algorithm interpolates the color of an individual pixel from the single colored pixel values around it. Finally, auto white balance and color correction methods of an image processing pipeline estimate the ambient light and correct the colors correspondingly. Also a lens system may change the colors by vignetting and color shading artefacts. The final color accuracy is a combination of all these factors.

The color accuracy, or fidelity, is an essential image quality feature of digital imaging and it can be defined as an ability of camera systems to reproduce colors as they exist in the original scene. In the case of objective color accuracy, the definition is quite clear, being the color difference between the scene and captured image. However, the perceptual color accuracy is a much more ambiguous metric, because it can vary between individuals, cultures or even seasons. Also it has been

noted that some amplification of color saturation gives the best perceptual color rate. The rate of the amplification varies between studies. Where Keelan et al. ended up with 10% amplification, the Camera Phone Image Quality (CPIQ) study does not recommend such a high value (Keelan 2012, CPIQ 2016).

Color itself can be divided in different components depending on the color space used. CIE XYZ or RGB can be defined as standardized color spaces whereas CIE $L^*a^*b^*$ or $L^*u^*v^*$ are perceptual ones (Lukac 2013). Since the most widely acknowledged color accuracy method is based on $L^*a^*b^*$ color space, it should represent perceptual color difference as discussed later in section 4.2.1. However, if observers prefer an image which does not replicate the colors exactly but has amplified colors, then color accuracy is probably the wrong method for measuring perceptual colors or at least, some weights should be added to match colorfulness requirements of observers.

When $L^*a^*b^*$ and $L^*u^*v^*$ color spaces are investigated, they have beside the chromatic components, the luminance (L^*) component. While a^* and b^* , or u^* and v^* components define the colorfulness and color balance, L^* defines the lightness of the image, correlating strongly with the exposure time and ISO speed. When the color accuracy is measured from $L^*a^*b^*$ color space, it also measures luminance accuracy expressing how well the captured image represents the brightness of the original scene.

The asterisks (*) are part of the color space names and they are used for historical reasons. In $L^*a^*b^*$ they have been used to distinguish them from the Lab presentation by Hunter (Hunter 1958). The origin of $L^*u^*v^*$ asterisks is harder to locate, they are probably used because $L^*u^*v^*$ color space is an improvement over CIE $U^*V^*W^*$ color space from year 1964.

Color accuracy is an even more problematic entity from a camera point of view, because the colors of the scene are combination of the ambient light and the original colors of the scene. The human vision system knows how to compensate the effect of ambient light, but for the camera system the task is difficult. In practice, the camera has to estimate the ambient light temperature or even its spectrum and adjust colors accordingly. The success of color correction can be judged in Figure 6 where four different mobile phone models have captured images in the same ambient light environment.



Figure 6 Color differences between mobile phone cameras.

The worst light environment is a situation where there are two or more different light sources, for example sunlight and fluorescent light and the camera system has to interpolate color correction factors between them. All in all, the color accuracy evaluation of a camera system requires measurements in several different ambient light environments.

3.2.3 *Dynamic range*

Dynamic range of a camera system represents the ratio between measured maximum and minimum light intensity in an image. In practice, the dynamic range defines how well the details are reproduced in the dark and bright areas in the same image. Normally the dynamic range is presented by decibels or f-stops (powers of two). Literature defines several values for dynamic range for a human eye, varying between 24-30 f-stops in situation, when the eye can adapt to the ambient light and 10-14 stops in a static light environment (Hoefflinger 2007; Cambridge in colour). The best DSLRs may have a dynamic range about 15 stops (DxO Mark) though the test results tend to vary between measurement software.

According to the ISO standard, dynamic range is: “ratio of the maximum exposure level that provides a pixel value below the highlight clipping value to the minimum exposure level that can be captured with an incremental signal-to-temporal-noise ratio of at least 1” (ISO 15739 2013). In practice, the dark end is reached when the temporal noise has same value as the signal.

Dynamic range can be artificially improved using high dynamic range (HDR), or wide dynamic range (WDR) techniques. The use of HDR and WDR terms vary a lot and they are also used as synonyms. Usually HDR is defined as a technique where several images are captured using different exposure times. The images are combined using dark end details of long exposure times and bright end details from short exposure images. In practice, this method can be used only in very static scenes, because any movement between images will ruin the result. WDR images are captured by using a nonlinear sensor where the differences in dark and bright areas are amplified (CMOSIS 2012). Finally, an image processing pipeline may include tone mapping algorithms which implement the same nonlinearity as the nonlinear sensor, but using software (Mantiuk, 2008).

3.2.4 ISO speed

Sensitivity of a camera, ISO speed, is an interesting feature especially in digital cameras because it is strongly related to the analog era of cameras. Originally ISO speed defined the sensitivity of an analog film towards light. At the same time when the sensitivity of the film increased the granularity of the film increased, too and the quality of images decreased. In practice, when the ISO speed changed, the physical composition of the film changed. During the analog film era, ISO speed was defined as a number, which was doubled when it increased, i.e. 50, 100, 200, 400 etc.

In the case of digital cameras, the ISO speed is purely a gain of the signal. Depending on the camera system, part of the gain can be added to the analog signal, before analog to digital conversion and rest to the digital signal. Since the ISO speed is only a coefficient, it affects the noise of an image significantly especially when it is added to the digital signal. The coefficient characteristics of the ISO speed in digital cameras has changed the traditional numbering of ISO speed. Quite often the ISO speed is handled as pure integer without the old rule of doubled values. In general, the ISO speed of a digital camera has quite similar characteristics to an analog film: it increases the sensitivity but decreases the quality.

Since the ISO speed is an adjustable parameter, like exposure time, one may ask if the ISO speed is a quality entity of a digital camera. However, a digital camera system has some native sensitivity. All components of the camera build up some generic base sensitivity which can be then amplified with an analog or digital gain and this base ISO, or native ISO, is definitely a quality factor of a digital camera.

To maintain the equivalence of ISO speed characteristics between analog film devices and digital cameras, ISO standard 12232 and CIPA DCC-004 define an

environment and equations to harmonize ISO speed ratings. Using the standards, the base ISO can be measured, too. The ISO speed can be calculated from a saturation based ISO speed or noise based ISO speed. The former is based on an exposure environment that produces an image, which has the maximum value, but is not saturated. The latter measurement is based on the signal to noise ratios (SNR), where an environment with SNR 40 defines the ISO speed. (ISO 12232 2006, CIPA DC-004 2004)

3.2.5 *Image processing*

As defined in section 2.2, the image processing pipeline of a digital camera has great number of algorithms which improve both objective and subjective image quality. Since the image processing pipeline may decrease the noise level significantly or increase the sharpness of images, it might be tempting to define the image processing efficiency as a quality entity. Particularly, in mobile phone cameras, the role of the image processing is crucial due to demanding environmental requirements of the sensor and lens system.

However, the qualification of the pipeline would be difficult, because it should measure the efficiency of the image processing. It would require an access to RAW images and in the case of mobile phone cameras, they are rarely available. On the other hand, image processing is a non-removable part of mobile phones and from a consumer point of view, the final quality is much more interesting.

In the case of digital single-lens reflex cameras, this kind of measurement would be reasonable, because they offer RAW images and image processing can be done using external image processing tools.

3.2.6 *Summary of image quality entities*

Table 1 gives a summary of image quality entities related to digital cameras and discussed in this section.

Table 1. Summary and a short description of image quality entities

Entity	Description
Resolution	A feature which defines the level of details which a camera system may produce.
Color accuracy	A camera ability to reproduce colors as they exist in the original scene.

Dynamic range	A feature which defines how well a camera can reproduce details both in dark and bright areas in a same image.
ISO speed	Analog or digital gain which amplifies an image data. On the other hand, base ISO speed or native ISO speed defines a native sensitivity of a digital camera without any amplification.
Image processing	A significant quality entity in digital cameras which includes several image quality improvement algorithms improving both objective and subjective image quality.

3.3 Artefacts of digital imaging

As discussed before, the concept of image quality is quite a difficult entity to specify accurately. Even if image quality can be measured in several ways, perceptual image quality always entails a problematic extension to the evaluation. An artefact of digital imaging is slightly easier to describe because the artefact is always an error in the image.

However, one may still ask what is an image artefact? Logically it would be a digression from a perfect image, a golden sample, which exactly represents the photographed scene. Still, here one may face a problem again, because image processing pipeline may boost colors a little bit or create high dynamic range images to increase the perceptual quality. A better description of the image artefact would then be an unwanted digression from the perfect image. And how it can be decided which change is an unwanted one? Again, we can observe that even image artefacts may have perceptual characteristics.

Like the imaging quality entities, imaging artefacts can be categorized in different ways. One approach is location based, which classifies artefacts by the location where the artefacts originate from (Imatest Image quality factors). As described in the section 2.2, a modern digital camera can be divided into camera sensor, camera module and image processing pipeline entities. A subset of imaging artefacts can be strictly assigned to specific camera parts, but usually an artefact is generated by a combination of several of them. However, the source based classification is straightforward and also pragmatic way to understand numerous sources of image artefacts.

3.3.1 *Sensor based artefacts*

A logical starting point for the artefact evaluation is the sensor of the camera, because it is the most essential part of a digital imaging. The sensor converts an analog photon flow to an electrical signal and finally to digital numbers, generating the first version of RAW image which is then processed by imaging pipeline.

3.3.1.1 *Fixed pattern noise*

One of the most obvious artefacts of a sensor itself is a bad pixel, or in more generic form, fixed pattern noise (FPN). Fixed pattern noise can be divided into two entities depending on the characteristics of the defective pixels. If the pixel has always a static value regardless of the input signal i.e. photon flow, the artefact is described as a dark signal non uniformity (DSNU). On the other hand, if the pixel value varies, but not according to the other pixels, the defect is categorized as a photo response non uniformity (PRNU).

The ISO 13406-2 standard defines artefacts of display panels, and the same definition of DSNU pixels is used in digital imaging. DSNU pixels can be categorized as hot, dead or stuck pixels, which always have the maximum, the minimum or a constant value, correspondingly. (ISO 13406-2 2001)

PRNU defects are more difficult to detect because the defective pixels do not have a static value. Typically for PRNU pixels, the error of the pixel depends on temperature, exposure time and ISO settings (Theuwissen PRNU). Obviously, more heuristics algorithms are needed for a PRNU pixel than a DSNU pixel.

The source of the fixed pattern noise is in the manufacturing process of the sensor, where the pixel construction in silicon is not always a perfect one. Quite often the sensor itself may remove DSNU pixels using calibration data got from the production line testing. Single bad pixels are not a major problem in a sensor with several million pixels, because they are almost impossible to detect in a non-zoomed image and they are easy to correct. However, several DSNU pixels can be located side-by-side creating a cluster, when the defect is more visible and more severe.

There are also several special cases of fixed pattern noise. A common hardware logic of pixel rows or columns may cause variation between rows and columns which cause column or row fixed noise. These can cause severe quality issues, since they create vertical or horizontal lines in the image and the human vision system is very sensitive to straight lines.

3.3.1.2 Temporal noise

Unlike fixed pattern noise, a temporal noise varies over time and thus it is much more difficult to remove from images. The origins of temporal noise are mainly in the camera sensor even though the lens system may generate some. However, the image processing pipeline may affect the noise level in a significant way. Several algorithms in image processing add digital gain to the image, thus the gain of the noise component increases too and makes the noise more visible. On the other hand, denoising algorithms may reduce the noise significantly from the final image but too aggressive noise removal may reduce, for example, image resolution and sharpness.

Generally, noise is an unwanted variance in the image and affects the sensitivity and dynamic range of a camera system. Noise can be visible especially in low light images where a low signal level, a long exposure time and a high ISO value increases the noise as in Figure 7, which is captured in a 30 lux environment. The camera adjusted the exposure time to 63 milliseconds and the ISO speed was 1665. To visualize the noise pattern, an originally uniform gray patch is magnified.

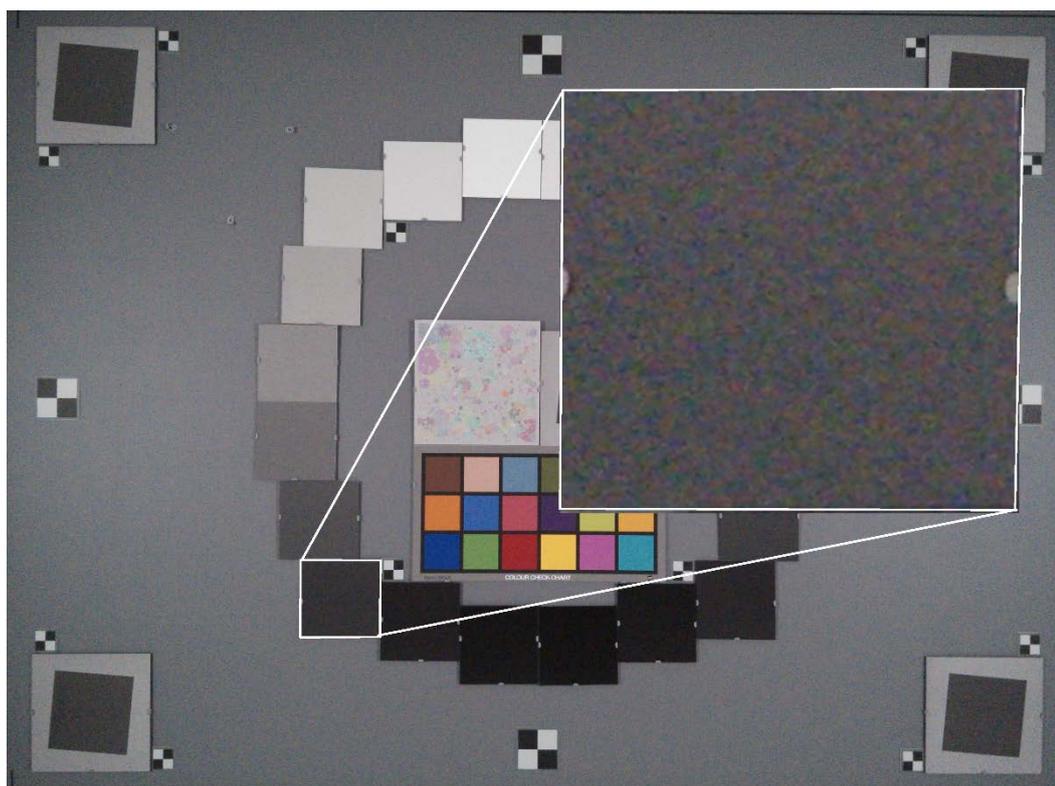


Figure 7 Noise in a picture captured in 30 lux

Roughly speaking, temporal noise can be divided into photon shot noise and read noise (Adimec Noise). More precisely, temporal noise can be divided into photon

shot noise, dark current shot noise, reset noise, and 1/f noise (Wang 2008). Even though the terminology for temporal noise varies, the read noise can still be defined as a combination of dark current shot noise, reset noise and 1/f noise. Also the quantization noise of the analog to digital converter can be defined as a form of temporal noise (Tian 2000).

The photon shot noise is related to the randomness of photons. The photon shot noise is a special noise, because it is a natural process of photons and it does not depend on the design of the sensor. There will be always photon shot noise in the RAW images and the photon shot noise follows the Poisson distribution. Thus the level of the photon shot noise is the square root of the mean signal level.

Dark current shot noise, or thermal noise, depends exponentially on the temperature and it can be partially controlled by design of the sensor (Wang 2008). The dark current defines the black level of the sensor. The black level is the mean value which a camera sensor generates without any light. The black level can be, for example, 5% of the maximum value of pixel, but it depends on the exposure time and temperature. The black level together with the white level affects the dynamic range of the sensor because they limit the true pixel value scale.

Reset noise, 1/f noise and quantization noise represent the rest of the read noise component, which can be reduced by good design of a sensor.

The noise characteristics of the sensor define in part the performance of the sensor by limiting the sensitivity and dynamic range of the sensor. Even when the denoising algorithms are efficient, they can still reduce other quality metrics of the image. All in all, a proper design of the sensor is essential for noise free and high quality images.

3.3.1.3 Banding

Every camera system has a certain bit depth, i.e. digital accuracy of a pixel. In the sensor, an analog to digital converter performs a quantization where analog signal i.e. electron flow, is changed to a digital number. Normally, a pixel has bit depth values from eight to sixteen meaning different pixel values from 256 to 65535 correspondingly.

If the bit depth is too small, the quantization may come visible in the image; this effect is called a banding or contouring artefact (Fenimore and Nikolaev 2003, Bhagavathy et al. 2007). Especially when an image contains an almost uniform area, small differences in the scene, for example in the sky, are not smooth but they generate visible edges in the image.

Bit depth is not the only variable to cause this artefact. Image processing algorithms like gamma correction and tone mapping may strengthen the banding artefact in bright and dark areas of images by stretching pixel value distances between corresponding illuminations.

3.3.1.4 Green imbalance

Even if green imbalance can be understood as a special case of photo response non uniformity PRNU, it is such a noticeable artefact that it should be discussed separately. Green imbalance origins are in a Bayer filter, where green has two different color channels: green in red rows g_r and green in blue rows g_b and in demosaicing algorithms. The green imbalance becomes visible when there is a mismatch between the green channels. Technically, the green imbalance is PRNU between two green channels and it is part of the noise entity of an image. The main reason for green imbalance is different cross talk between red and green rows (Guarnera et al. 2010) or an improper demosaicing method. Green imbalance causes a maze-type pattern in images as shown in Figure 8.

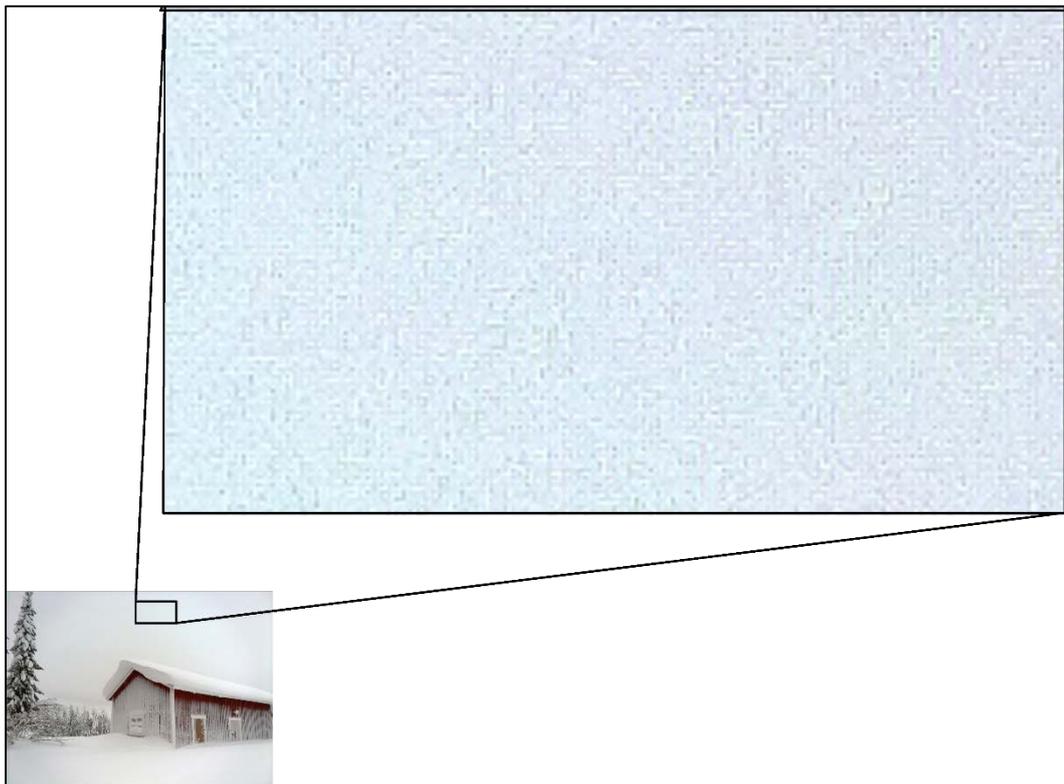


Figure 8 A maze pattern caused by green imbalance artefact

3.3.1.5 *Moiré*

Every sensor has its resolution limit specified by its pixel size and pixel pitch i.e. the distance between individual pixels and other limitations of the camera system. When the details of the captured scene are smaller than the resolution limit multiplied by two, according to the Nyquist law, the image sensor cannot reproduce the details of the image (Imatest Moiré). High frequency details, for example textiles, can produce stripes to captured image. These stripes are called as Moiré artefact. Quite often Moiré artefacts are avoided by using an optical low pass filters in the lens system. Especially in video broadcasting, high frequency details may cause flickering in the stream and be a very annoying issue. In the case of still imaging, Moiré causes stripes across an area originally containing high frequency details.

3.3.1.6 *Blooming*

Blooming is defined as an artefact which causes blurry borders in a highly exposed objects. In the worst case, the shape of the bright object will become unrecognizable and the saturated area will spread across the whole image. When blooming has occurred, pixels which have absorbed high number of photons and therefore have become saturated start to crosstalk i.e. spill electrons over to adjacent pixels. This may cause problems especially in outdoor imaging due to high illumination by the sun and on the other hand, in security systems where the low light performance is crucial, bright objects may corrupt captured images or a video stream.

Arganov et al. defines three different crosstalk components in a CMOS sensor: spectral crosstalk, optical spatial crosstalk and electrical crosstalk all of which cause different artefacts in images (Arganov et al. 2003). Even though electrical crosstalk is the main reason for blooming, it is not the only one. Theuwissen defines in his famous blog seven different mechanisms, which causes blooming (Theuwissen Blooming).

Fortunately, due to the design of CMOS sensors, the blooming is no longer such a severe problem as it is for the CCD sensors. In CCD sensors blooming may cause overflow of the whole vertical pixel line, which causes bright columns over the whole image (Adimec Blooming).

3.3.1.7 *Black sun*

In a black sun artefact, an extremely highly exposed object turns from white to black in the captured image. This often happens when the capturing scene contains

the sun and the circle of the sun becomes black in the captured image. One may think the artefact is due to overflow in the image processing pipeline, but the origins of the defect are inside the sensor's logic. When a pixel exposure starts, some sensors read the black level value of a pixel by exposing it for a very short time (CMOSIS 2012). This is done to reduce the black level noise by subtracting the black level from the real exposed value. However, if a certain pixel is illuminated by an extremely bright object i.e. the sun, the reset level may rise so high, that the final pixel value is subtracted to zero and therefore the pixel contains only black color. This is not so rare a problem as one may think. The issue was visible for example in the broadcast of IAAF World Championships in Beijing 2015, see Figure 9.



Figure 9 Black sun image artefact in the video stream of IAAF World Championships in Beijing 2015 (Youtube)

3.3.1.8 Rolling shutter

A rolling shutter defect is maybe the most surprising artefact of CMOS sensors because it can change the shape of an object. The origins of the artefact can be found in the implementation of the sensor itself. Currently, most CMOS sensors are exposed row by row. Due to the implementation, the readout period of each row cannot overlap with other rows, which means that the captured object may move or the exposure environment may change between the exposure time of each row.

The rolling shutter defect may cause three different artefacts. If an image is captured from an object which moves or rotates rapidly, for example the propeller of the airplane or fan, the captured object is skewed. The same phenomenon

happens when the object is stable but the camera is moving. Moreover, if the camera itself vibrates in a high frequency, for example when it is mounted in the car, the artefact may cause wobbling, also known as the jello effect, in the video and the recorded stream plays like a shivering jelly. (Baker et al. 2010)

Obviously, rapid exposure environment changes can cause the same kind of defects in the captured image, but now the illumination is changed between the rows of the image. The classical example of the artefact is weddings where there are numerous cameras with flashes. When several flashes are used simultaneously but not synchronously, the rows of the image are exposed differently and the result can be spoiled.

In general, rolling shutter defects are related to phenomena occurring during a very short time period. It also has to be remembered, that CCD sensors do not suffer from this defect. They use a global shutter which prevents the problem. Some of the latest CMOS sensors also use a global shutter.

Figure 10 shows the rolling shutter phenomenon using a very simple sensor with three rows. The white section represents the row which is exposed at the time t_1 , t_2 and t_3 correspondingly. As shown in the right most figures, the resulting images are skewed in various ways and the severity of the skew depends on the relative speed between the camera and object. If the relative movement is more shaking than linear movement, the result is wobbling, when the sensor is used for video recording.

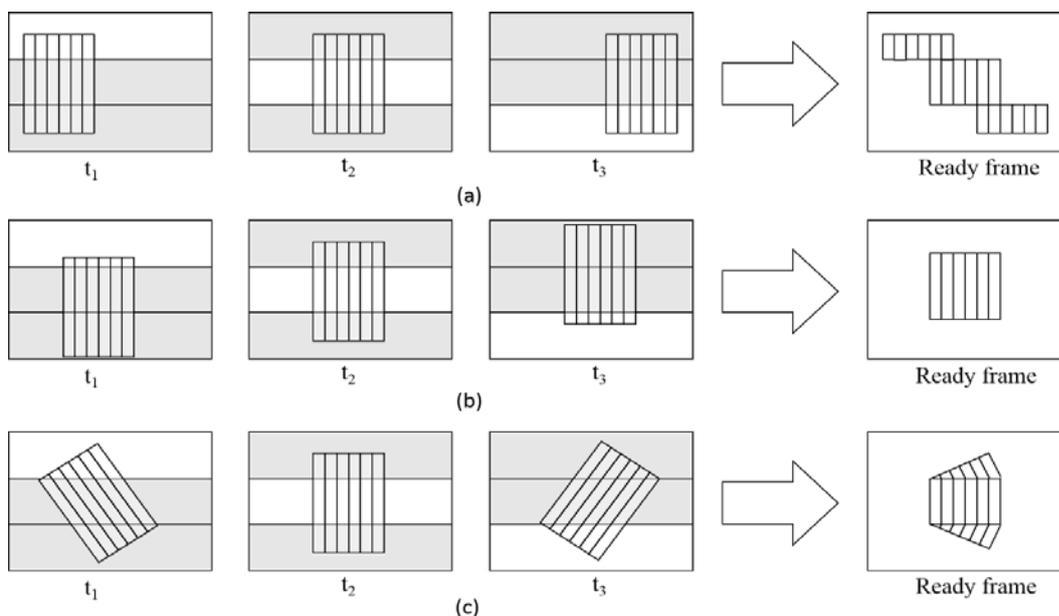


Figure 10 The principle of the rolling shutter defect. The image is based on paper by Sun et al 2012

3.3.2 Camera module based artefacts

A camera module can be divided roughly into two components: a lens system and a sensor. Moreover, camera modules may have an optical image stabilization feature, which may greatly affect the camera's performance. Nowadays, the camera module is no longer just a lens extension on top of the sensor.

The lens system can be defined as the second most important component of the camera system after the sensor itself. The lens system collects photons and focus them onto the sensor, it may have movable lenses for autofocus or even zoom and changeable aperture features. Usually, the lens system also includes an optical low pass filter to remove the Moiré artefact and an infrared filter to suppress red channel saturation due to infrared illumination. The quality of the lens system is crucial for generating high quality images.

3.3.2.1 Lens aberrations

Lens aberrations are classical optical artefacts which are clarified broadly in literature (Hecht 2002; Kingslake 1992; Walker 1998). They affect significantly the quality of the image starting from a correctly a focused image, defining the limiting resolution of the system together with the sensor and ending at a correctly drawn image without geometric distortions. The purpose of this section is to give an overview of the defects which affect especially the quality of mobile phone cameras.

Even though Heicht, Kingslake and Walker categorize lens aberrations in slightly different ways and using different terms, the main distortion types can be easily listed.

Starting with monochromatic aberrations, spherical aberration causes different focus points with different light rays which are parallel but have different distance from the optical axis. The light rays which are bent in the edge of the lens are refracted too much and they generate another focus points in front of the sensor. As a detail, it should be mentioned that the Hubble Space Telescope was suffering from the spherical aberration before it was fixed with extra mirrors. (Hecht 2002)

A coma, or comatic aberration, is the same kind of aberration as the spherical one, but it affects rays which are not parallel with the optical axis. Rays coming from the same point (negative coma) or that are parallel (positive coma) but traversing through the lens at different points are focused at different places in the focal plane. (Hecht 2002)

Astigmatism is a lens artefact where light rays, which are not parallel with the optical axis do not create a single focus point but two perpendicular focal lines where one is on a radial plane towards the lens field and the other on a tangential plane. The artefact is visible especially in the border areas of a lens. (Kingslake 1992)

Obviously all three aberrations mentioned above generate several focus points and thus cause blurry and defocused images.

A field curvature is an aberration, which affects especially the focus of an image border and corner areas. While a lens is a curved, it also tends to generate a curved focal plane. In practice, this means an optimally focused image is generated on a curved plane. However, image sensors are normally flat which means some parts of the image sensor can be always slightly out of focus (Hecht 2002). Obviously the center of the image is more important than the border areas of the image and therefore the center of the image sensor is located to the correct focus point.

Figure 11 gives an overview of the different monochromatic aberrations. The optimal lens would create a single focus point on focal plane (f) but different aberrations scatter the light rays and blurs the focus point.

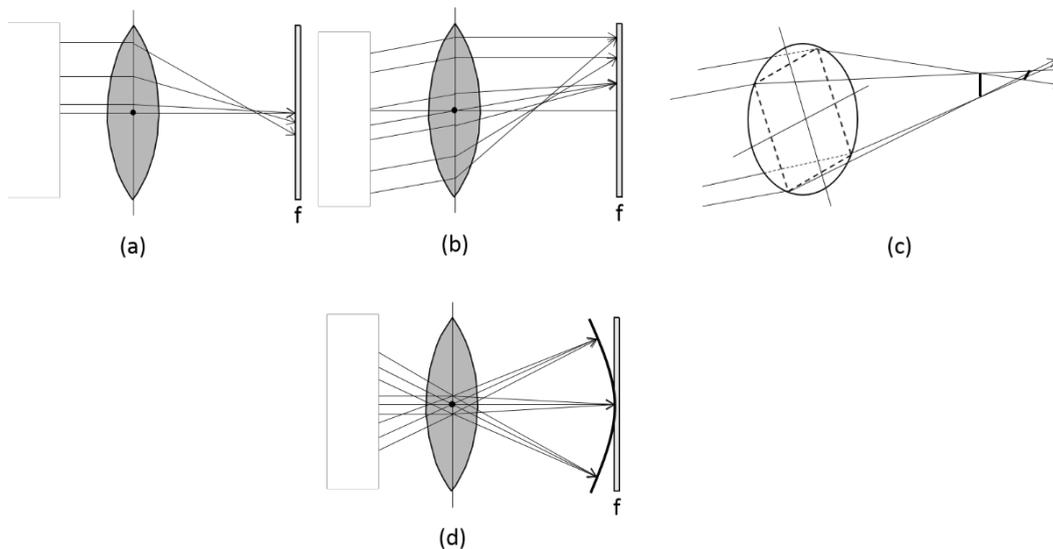


Figure 11 Monochromatic aberrations a) Spherical aberration, b) Positive coma, c) Astigmatism, and d) Field curvature (Hecht 2002; Kingslake 1992)

Monochromatic aberrations may also cause different geometrical artefacts in images. When the magnification factor of the lens system is not a constant but varies as a function of the optical axis distance, the distortions generate geometrical

errors in images (Hecht 2002). Usually, distortions are divided into positive or pincushion distortion, negative or barrel distortion or as a combination of these ones, moustache distortions. The names of the distortions are quite descriptive as Figure 12 shows.

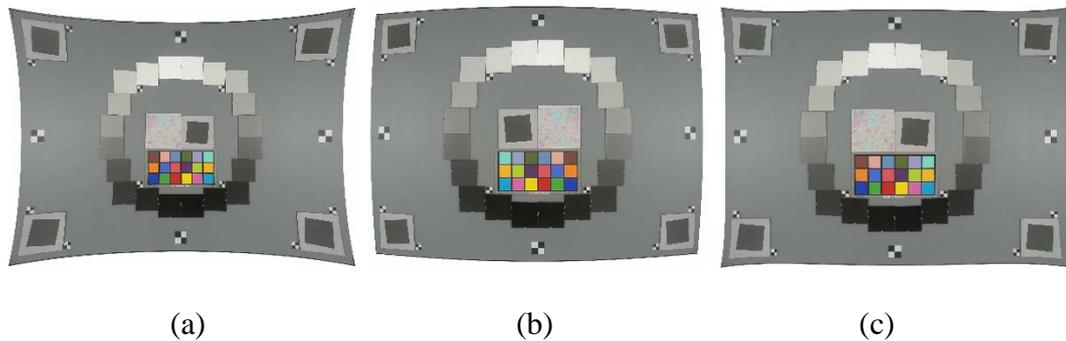


Figure 12 Distortions a) Pincushion, b) Barrel, and c) Moustache

The chromatic aberrations are almost equivalent to spherical and coma aberrations, but in case of the chromatic aberrations, different wavelengths of light generate different focus points. The reason for the chromatic aberration is quite simple, the wavelength of the light affects the refractive index of the lens and therefore also the focal length of the lens (Hecht 2002). The artefacts may cause color errors like halos near high contrast edges in images.

Chromatic aberrations can be divided into axial and lateral aberrations (Hecht 2002; Imatest Chromatic), where the former is functionally equivalent to spherical aberration and the latter to coma aberration.

In the case of axial chromatic aberrations, different wavelengths bend in the lens in different way and generate focus points at different distances in the focal plane, whereas lateral chromatic aberrations create focus points in different places in the focal plane as Figure 13 shows.

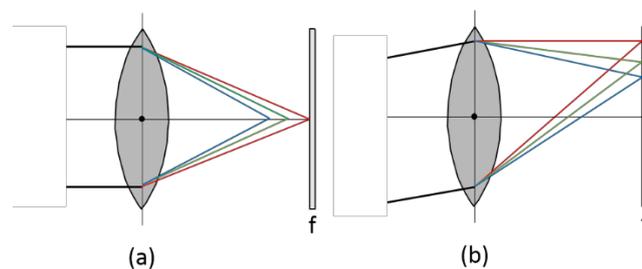


Figure 13 Chromatic aberrations a) Axial and b) Lateral

All lens aberrations originate from the optical design and the quality of the lens material. Although, there is always some kind of aberration, it can be partially reduced by image processing and proper design of the optical system.

3.3.2.2 Defocus

Defocus is one of the most crucial artefacts of the imaging process. In practice, a poorly focused image is nearly impossible to correct afterwards. The reasons for defocus are various: aberrations of a lens system, too aggressive image processing algorithms may blur the image causing same kind of effect as optical defocus, a motion blur caused by hand shaking or moving object or malfunction of an optical image stabilization.

However, the main reason for the defocus artefact in modern mobile phone cameras is the autofocus algorithm itself. Several methods are used in autofocus functionality, the algorithm may use separate phase detection pixels or calculate the status of focus from the scene (Toshiba PDAF). Also laser based methods are used in several mobile phones (Image Sensors World AF). According to this information, the algorithm adjusts the movable lens inside the lens system to get a perfectly focused image. If the algorithm fails to calculate the exact lens position or if the algorithm is too slow to react the changes of the scene, the result is usually a defocused image.

One may think that by using a perfect lens system and a sensor with an unlimited number of pixels, all details of the scene can be captured in the image. However, there is a physical limit on the focus called airy disk which specifies the minimum spot size a perfect lens and circular aperture (iris) can make (Hecth 2002). The specification of the airy disk is based on research by George Biddell Airy, 1835.

The size of the spot is limited by the diffraction of light of certain wavelengths and it cannot be improved according to current knowledge of optics. The limiting resolution can be calculated as specified in (1).

$$(1) \quad x = 1.22 \frac{\lambda f}{d}$$

Here λ is the wavelength of the light, f is the focal length of the camera, d is the aperture size and thus f/d is the f-number of the camera. If the limiting resolution is calculated for a typical mobile phone camera for average green light wavelengths, we will get

$$(2) \quad x = 1.22 \times 530 \times 10^{-9} \text{ nm} \times 2.2 \Leftrightarrow x = 1.42 \mu\text{m}$$

If the pixel size of the sensor is smaller than this limit, the limiting resolution of green light is due to the lens system, not the pixel size of the sensor. Still some mobile phones use sensors with smaller pixel sizes (Samsung Tomorrow). However, some other improvements can be made using a small pixel size and high pixel count, like oversampling and denoising (Nokia 2013).

3.3.2.3 Vignetting

Vignetting is an artefact that reduces the amount of light falling on the sensor area away from the optical axis. In practice, this means the border area and corners of images become darker than the center area. The artefact is widely discussed in different literature and the origins of the defect can be found in physics and the characteristics of a lens system itself. The amount of light falling on the sensor can be calculated from (3) (Kingslake 1992).

$$(3) \quad E_{\phi} = E_0 \cos^4 \phi$$

E_{ϕ} is the amount of illumination towards the sensor originating from direction ϕ related to the illumination E_0 which comes via the optical axis. In the case of a large field of view, as widely used in mobile phone cameras, the vignetting reduces dramatically lightness at the corners of the image and it has to be artificially increased in the image processing phase.

3.3.2.4 Color shading

Color shading causes large color artefacts in images. Quite often it can be visualized as a pink image center and greenish borders. Obviously, the artefact is more visible when the image contains large, uniformly colored areas, for example sky or snow. The primary reason for color shading is an infrared filter malfunction, but also some color filter array level crosstalk may cause part of the color shading (DxO Color shading; Agranov et al. 2003; Hsu et al. 2005).

Color shading is a problematic artefact, because it changes according to the wavelength of the light. This means the shading effect has to be fixed according to the ambient light environment, which requires more parameterization and intelligence in the image processing algorithms.

3.3.2.5 Short focal length issues

Due to very demanding space requirements of mobile phones, the focal length is small, between 25-30 mm calculated as the 35 mm film equivalent. The real focal length is notably small, at about 4 mm. The small focal length means a large field of view angle, about 60°. This allows light rays to fall on the sensor at a very steep

angle. The large angle may cause optical crosstalk in the color filter array or malfunction of the infrared filter on top of the sensor as defined in the previous sections. As well as these artefacts, the short focal length may also cause perspective errors: When an object lies near this kind of lens system, the lens system magnifies the center area much more than the border areas. In case of near portraits, for example in selfie images, this phenomenon causes more magnification in the center of the face and, in practice, generates a big nose in the image.

3.3.2.6 Other lens artefacts

A lens system may generate also some other artefacts due to improper manufacturing or poor materials. A blemish is defined an artefact which produces an area of slightly darker pixels in the image caused by a scratch in the lens or dust or other objects inside the camera module (Lepistö, 2009). On the other hand, if the lens system is not mounted correctly on top of the sensor, the perspective of the camera is distorted (Imatest Tilt). A slightly tilted lens system causes so called keystone effect, where the camera produces also perspective in objects which are perpendicularly oriented towards the camera.

An artefact called veiling glare, or flare, can decrease the quality of images. The flare occurs when a light ray scatters and reflects inside the lens system and makes bright circles or hazy light areas in images (Kingslake 1992). In some literature, flare and glare has been separated, where flare is a generic increase of the black level (hazy light) and glare defines bright circles or ghost images. The artefact can be partially removed by high quality lens materials, lens coatings or lens hoods.

3.3.3 Image processing pipeline based artefacts

The third link in the image creation chain is the image processing pipeline as defined in section 2.2. In modern mobile phone cameras, the significance of the pipeline is incontrovertible. Due to small sensor size, extremely small pixel size, and very demanding lens requirements, the quality of the RAW images from the sensor is quite often very poor. Images are dark, noisy, distorted, and colors of images are not correct. In practice, the image processing pipeline recreates the image by denoising, color corrections, gamma correction, tone mapping, and several other algorithms. An example of differences between the original RAW image from sensor and the processed image is shown in Figure 14.



(a)

(b)

Figure 14 Image processing pipeline example: a) RAW image from sensor and b) Processed image

It may even be said that severe overexposure, underexposure and defocus are the only artefacts the image processing pipeline cannot correct. However, the quality of the final image depends highly on the algorithms and severe errors may occur when the algorithms or parameterization of the algorithms are not correct. This section declares the most problematic artefacts the image processing pipeline may cause.

3.3.3.1 Compression

Usually all images and video streams are compressed before they are used, stored or broadcast. The majority of compression algorithms are lossy i.e. they remove information from the original image or stream. Obviously, compression artefacts are related to the compression method actually used. Artefacts of the most common compression methods used in mobile phone cameras are discussed in this section.

A block based compression method is used in many image and video compression algorithms, for example JPEG, MPEG-1, MPEG-2, and H.26x compressions use the method. In the case of the widely used JPEG compression, the block size is 8x8 pixels and a local discrete cosine transform (DCT) is executed in each block. Since the compression is based on blocks, a discontinuity between blocks is possible and can cause blocking artefacts. (Keelan 2002; Wang and Bovik 2006; ITU-T T.81 1992)

JPEG2000 compression is based on a wavelet compression, which transforms the whole image and does not suffer from blocking. However, the wavelet based compression may cause a ringing artefact, which causes faulty luminance or color

highlights near high intensity edges in a quite similar way to over sharpening. (Wang and Bovik 2006)

In general, compression tends to filter out high frequencies especially in chromatic channels because the human vision system is less sensitive to those. Too high compression rate may lead to the blurring artefact where high frequencies, small objects and texture, are filtered out.

In case of video compression, the artefacts can be more visible, as the compression is not made inside one video frame but between frames. This may cause also temporal artefacts. Video artefacts are discussed more in section 3.4.

3.3.3.2 Color inaccuracy

Color accuracy in image processing is based on two methods; estimating the light temperature of ambient light and color correction according to this estimation. Obviously, both methods can cause color errors in the final image.

If the ambient light is estimated falsely, wrong color correction factors are used and, for example, a scene captured in sunlight may turn bluish, if it is corrected using fluorescent correction factors. On the other hand, the color correction factors can be inaccurate, if the camera system is not calibrated correctly to each type of ambient light or the interpolation between color correction factors does not work correctly.

Moreover, the algorithm itself can fail to reproduce the colors of the scene. A good example is an old but widely used method called gray world which assumes that the mean color of the whole image is always gray and estimates the ambient color correction according to that assumption. The method works well until the scene includes a dominant color. In such a case, the colors of the images are biased according to the dominant color.

3.3.3.3 Sharpening artefacts

Sharpening is a method where the intensity of the edges in the image is artificially amplified by increasing the contrast of the edges. For example, the border between light gray and dark gray is amplified by darkening the dark gray area near the edge and lightening the light gray area correspondingly. The sharpening can be used to increase the perceptual sharpness of images, but it may easily generate various artefacts, too.

If the edges are amplified too much, the sharpening comes visible and causes a ringing artefact, a halo around edges. Correspondingly the dark side of the edges

may turn too dark causing visible dark lines. Wrongly parametrized sharpening starts to highlight particles in the image that are too small, it amplifies the noise and also may exaggerate textures or even filter high frequencies. (Caponigro) All in all, the image may turn unnatural looking.

Figure 15 illustrates the sharpening artefacts both in the edges and in texture areas.

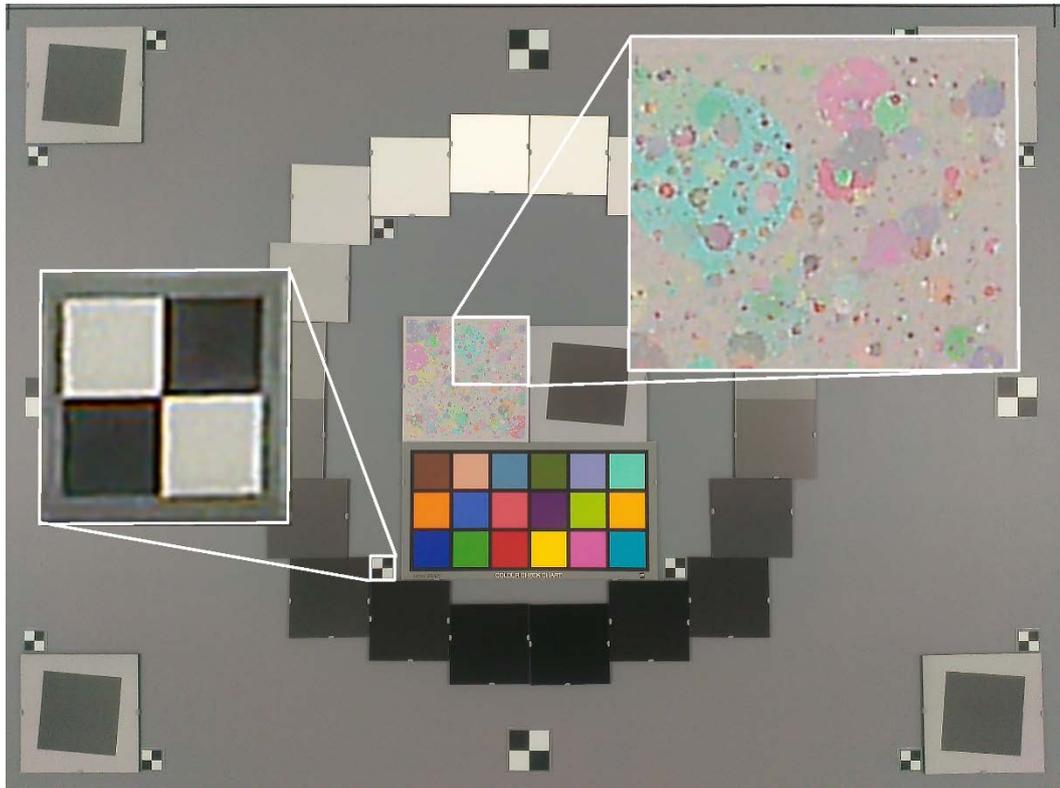


Figure 15 Sharpening artefacts

3.3.3.4 Noise removal artefacts

The main artefact related to noise removal is generic blurring. When the noise particles are removed efficiently, also edge areas tend to be smoothed. On the other hand, if the image contains small detailed natural texture, for example sand, the characteristics of the texture are quite similar to the noise generated by the camera system. It is very difficult for the noise removal algorithm to separate natural texture and artefactual noise. This is problematic especially in the texture parts of the picture where noise removal may cause texture loss (Artmann and Wueller 2012).

Equally, when denoising is too efficient, it may cause over smoothing in uniform areas of the images. This may lead to unnatural images which appear oil painted

(da Silva et al. 2013). Some block based denoising algorithms like block-matching and 3D filtering (BM3D) may cause also blockiness in the images (Dabov et al. 2006). An example of a poor image quality after too aggressive BM3D noise removal is shown in Figure 16.

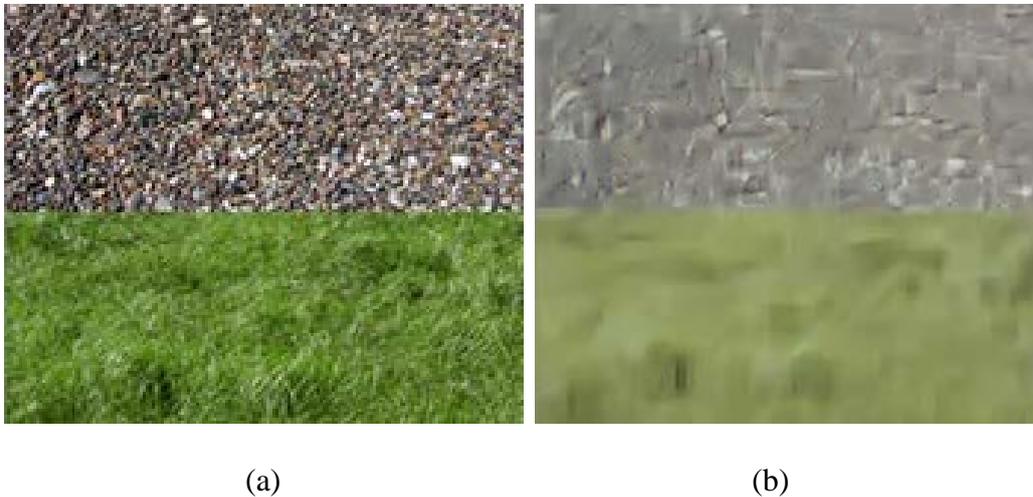


Figure 16 Noise removal artefacts, blurring and blockiness: a) Original scene and b) Aggressive denoising

3.3.3.5 Demosaicing

Since demosaicing interpolates the colored pixels of the Bayer filter to single colored pixel values, it may affect several quality features of the camera system; noise, colors and resolution. In addition, more specific errors like maze pattern artefact, moiré and zippering may occur. It has to be also remembered that the Bayer type sensor has two times more the green pixels than red or blue pixels, thus the resolution of green color channel is two times better than the other ones. The demosaicing algorithms have to allow this imbalance.

There have been a lot of research and suggestions for algorithms to be used in demosaicing. For example, nearest neighbor replication, bilinear interpolation or cubic spline interpolation can be used to calculate the colors of a pixel (Menon et al. 2006). Since demosaicing is always based on interpolation, it generates only an estimate of the missing two color components of a certain pixel. The estimation always generates noise in the image and the accuracy of the estimation defines the level of blurriness caused by demosaicing as well as the color accuracy of the final pixel value.

Inefficient demosaicing may cause a maze type pattern in the image, if the original Bayer filter structure is not filtered out properly. Finally some demosaicing types,

like plane-wise interpolation may distort object boundaries by generating zipper shaped edges (Hirakawa and Parks 2005).

3.3.3.6 Over processed images

Finally, an image can be processed too much. The reason for an over processed image may be poor quality of the RAW image or too aggressive parametrized image processing algorithms. Even if the algorithms like denoising, sharpening and tone mapping do not create any artefacts in final images, the final image may look unnatural. Obviously, the naturalness of the image is very perceptual quality feature and it is difficult to measure.

3.3.4 Summary of digital imaging artefacts

Since a digital camera system may have numerous artefacts based on several sources of a camera system, tables 2, 3 and 4 give a summary of digital imaging artefacts discussed in this section.

Table 2. Summary and a short description of sensor based artefacts in digital imaging

Entity	Description
Fixed pattern noise	A noise which is generated from faulty pixels which do not react correctly to a photon flow. Erroneous pixels may have always static values (DSNU) or they can react differently than majority of the pixels (PRNU).
Temporal noise	A noise which varies over time. It can be divided to photon shot noise, which is related to the randomness of photons and read noise, which is related to design of a sensor.
Banding	Visible quantization in images which generates edges in almost uniform areas in images.
Green imbalance	Imbalance between two green channels in a Bayer type sensor. Green imbalance generates maze type noise especially in uniformly colored areas.
Moiré	Moiré generates chromatic or monochromatic low frequency stripes on top of high frequency details in images.
Blooming	An artefact which generates blurry borders in a high exposure objects.

Black sun	A sensor based artefact which turns extremely highly exposure objects from white to black.
Rolling shutter	A sensor feature which may generate three types of artefacts: it may change the shape of a moving object, it may generate erroneously exposure images, and generate a vibration to a video recording.

Table 3. Summary and a short description of camera module based artefacts in digital imaging

Entity	Description
Lens aberrations	Optical distortions which can generate blurriness, geometrical errors, and color errors to images.
Defocus	Blurriness in images generated by lens system, image processing pipeline, or improper functionality of an autofocus algorithm.
Vignetting	A phenomenon where border areas and corners of images are darker than the center areas.
Color shading	Large color artefact in images. Generated by infrared filter malfunction or cross talk in a sensor.
Short focal length	A lens system with a short focal length and a large field of view enables color shading and may generate perspective errors.
Blemish	A scratch or dust in a lens system generating darker areas in images.
Keystone effect	Tilted sensor mounting towards the lens system generating perspective errors.
Veiling glare, flare	Light ray scattering inside a lens system generating bright circles or hazy light areas in images.

Table 4. Summary and a short description of image processing pipeline based artefacts in digital imaging

Entity	Description
Compression	A compression may generate blurriness, block artefacts, and faulty edge coloring.
Color inaccuracy	When a camera system estimates wrongly the ambient light temperature, it will use wrong color correction factors and cause faulty color tint over the whole image.

Sharpening	Too aggressive sharpening will amplify noise, generate halos around edges and generate unnatural images.
Denoising	Too aggressive denoising causes blurriness and texture loss to images. Some denoising algorithms may generate also block errors to images.
Demosaicing	Demosaicing algorithm may reflect to noise, colors and resolution of images. Moreover, it may generate maze type noise, Moiré, and zipper type pattern to edges.
Over processing	In general, too aggressive image processing pipeline causes unnatural images.

3.4 Video quality and artefacts

Even though this thesis is mainly concentrated on still image quality and its benchmarking, some video related quality factors should still be evaluated. Especially because the role of video is increasing all the time in mobile phone usage.

When video quality entities and artefacts are evaluated, most of the still image metrics are still valid. Resolution, color accuracy, dynamic range and noise levels build the base of the video quality. Also artefacts defined in section 3.3 can occur in a video recording or stream. However, acceptable values of the still image metrics can be different in a video environment. For example, spatial resolution requirements can be decreased, because the frame specific resolution is not as visible in a video stream as it would be in a single frame.

In case of video, the temporal performance and temporal related artefacts have to be highlighted. An essential part of video quality is also audio quality, for example improper synchronization between audio and video stream can cause severe quality regression in video (EBU R37 2007).

Winkler divides video artefacts into two logical entities, compression artefacts and transmission errors. Since transmission artefacts are not valid in a mobile phone camera system, they are not discussed here. However, a notable amount of compression artefacts were listed in the book: blocking effect, blur, color bleeding, DCT basis image effect, staircase effect, ringing, false edges, jagged motion, chrominance mismatch, mosquito noise, flickering and aliasing (Winkler 2005). Jagged motion, flickering and mosquito noise are clearly temporally based artefacts.

Jagged motion as well as motion blur are related to the frame rate and exposure time of the camera system as well as to the compression algorithms. They can be associated with the temporal resolution of the video i.e. how well the camera system can follow movement in the recorded scene.

Mosquito noise is equivalent to the still image ringing artefact, but it varies between frames and causes a local flickering. The flickering in general may be one of the most annoying artefacts in the video stream and it can be visible in several ways. The noise level may change between frames and cause changing noise patterns especially in the uniform areas. Auto exposure, auto focus and auto white balance algorithms may cause the flickering effect in lightness, sharpness and colors correspondingly.

Even if the camera system provides an error free transform from one scene to other, the smoothness of the convergence could also cause artefacts in the video. Nuutinen et al. have investigated auto exposure and auto white balance convergence metrics. Even though the study did not give unambiguous metrics, high saturation during the convergence was rated more annoying than the convergence duration being too short or too long. (Nuutinen et al. 2013)

Finally, optical image stabilization (OIS) may cause some issues in the video stream. OIS detects the movements of the camera and compensates the movements by moving either the camera sensor or certain part of optics to the opposite direction. Using OIS, a significant amount of hand shaking effects can be removed during image exposure and longer exposure times can be used. This especially helps low light imaging. However, there have been cases, where optical image stabilization has started to oscillate and generated a significantly distorted video stream (Business Insider).

3.5 Is camera performance part of image quality?

Camera performance, meaning the functional speed of the camera in general or quickness of a certain camera functionality is a very novel measurement area. As late as 2013, the first ISO standard for camera speed was published: ISO 15781. A year before, CIPA DGC-002 standard was translated from Japanese including also camera performance measurement guidelines (ISO15781 2013; CIPA DCG-002 2012). Before then, the speed of camera features were perceived as a generic usability of the camera or smoothness of the user interface.

It can be argued, if camera performance is part of image quality at all. The consideration can be started from exposure time which is one of the most critical

entity of photographing in general. Under or over exposure can destroy the captured image entirely. The correctness and accuracy of exposure timing are certainly quality features of digital imaging. On the other hand, auto-exposure feature is a standard feature in mobile phone cameras. The accuracy and speed of auto-exposure algorithm enables both correct illumination of an image and smooth capturing functionality.

ISO 15781 standard highlights the situation where pictures are taken from moving targets. Too long delay between pressing the exposure button and real image capturing may ruin to preserve the moment (ISO 15781 2013). Furthermore, different features like auto-focus, image post processing, image stabilization and video recording may generate own delays to an image capturing process. The delays do not prevent image capturing nor reduce traditional image quality but they still can prevent to capture the required moment. The usage of an image captured in wrong moment is quite same than an image with poor image quality, the captured image is deleted from mobile phones memory.

Bucher et al. describes an interesting performance feature of mobile phone cameras. In the research several cameras had a negative shutter lag. This means that a camera system is capable of storing frames during the whole capturing process and selecting required frame afterwards. (Bucher et al. 2014) Wrong functionality of the feature may generate a strange and unwanted phenomenon where a camera captures images too early.

Masson et al. notes other performance features which affect the functionality and quality of digital imaging. Speed of a rolling shutter affects significantly the rolling shutter artefacts. If the rolling shutter speed is slow, i.e. the delay between the first row exposure and last row exposure in an image sensor is long, it may cause distortion to moving objects or exposure errors to an image. The research includes also performance measurements of image stabilization. The research revealed how much exposure time can be extended when the image stabilization is active. (Masson et al. 2014)

It can be assumed, that the significant growth of video recording will highlight the performance of auto focus and auto exposure speed and the smoothness of these features because they are no longer pre-processing steps in image capture but they affect the real recording result. Moreover, auto white balance will have same kind of convergence delay and it should be investigated, too.

Until now, camera performance factors have been more like usability features than quality factors, because they did not affect traditional image quality. However, fast functionality of the camera is a feature which allows a user to capture an instant

moment. Conversely slow functionality could prevent this capture. According to the measurements taken during this study, a camera can generate delays of several seconds when an image is captured. It can be discussed, if camera performance is an image quality feature, but in the case of camera usability, the role of camera performance is incontrovertible.

4 IMAGE QUALITY MEASUREMENT METHODS AND METRICS OF MOBILE PHONE CAMERAS

If all the image quality standards and de facto standards are listed, numerous quality metrics can be found and the metrics can be classified in different ways. Keelan defines in his book, *Handbook of Image Quality*, following division into separate objective metrics: (Keelan 2002)

- Quality metric, a single number value correlating to a perceptual image quality
- Objective measurement, a function of at least one variable, for example modulation transfer function (MTF) of the slanted edge test chart
- Engineering parameter, a single number value describing a property of a camera system, for example the pixel count
- Benchmark metric, a single number variable combining usually several objective metrics to compare features of the cameras

On the other hand, Wang and Bovik define methods for image quality measurement as follows: (Wang and Bovik 2006)

- Full-reference, no-reference and reduced-reference image quality measurement. The division is used frequently when image quality measurements are defined. Obviously, the methods of the image quality measurements are very different depending on the availability of the reference data.
- General purpose and application specific image quality measurement. The application specific measurement concentrates on some specific quality feature or artefact of the image, for example lens distortion or video artefacts. On the other hand, the general purpose measurements give a generic score or result of the image quality.
- Bottom-up and top-down image quality measurement. When the image quality methods are defined, they have to simulate or mimic the human vision system (HVS). There are two ways to build up the simulation. The bottom-up method divides the HVS simulation into its relevant components and psychophysical features and builds the simulation by combining features together. The top-down procedure creates an overall model of the entire HVS and defines the simulation as a black box model.

Traditional image quality standards are mostly based on objective measurements according to the classification of Keelan and the reduced reference image quality measurements by Wang and Bovik. The combination is quite practical because the full reference method requires an exact digital reference which is not always

available. On the other hand, the no-reference method has not reached the level of reliability required to measure image quality as well as required.

This chapter defines different image quality metrics and methods starting with standardization in general, describing the division of color, noise, dynamic range, and resolution metrics, clarifying the needs of artefact measurements, new algorithms and perceptual quality metrics and ending with video and performance metrics.

4.1 Standardization and current tools

Unfortunately, several different organizations have independently developed their own image quality standardization. Different metrics and measurement approaches generate different results which make comparison between standards difficult.

The digital imaging standardization entity used most is probably the set of standards from the International Organization for Standardization (ISO). The technical committee of photography, TC42, includes 184 standards, though only part of them relate to digital imaging. Color standards are mainly based on work of International Commission of Illuminance (CIE). Nowadays, some of the color standards are defined as joint standards between ISO and CIE.

The Camera & Imaging Products Association (CIPA) is a Japan based organization including members mainly from Japanese camera companies. Part of the standards are translated into English and for example, CIPA has the only standard for optical image stabilization (OIS) testing.

The American National Standards Institute (ANSI) does not write its own standards, but it accredits organizations who develop their own. ANSI has published video related standards with the Consumer Technology Association, CTA.

The International Telecommunication Union-Telecommunication (ITU-T) has several standards especially for subjective quality assessment methods and for multimedia applications. Some video related measurements can be found in standards created by the European Broadcast Union (EBU) and the International Electrotechnical Commission (IEC).

A new approach to mobile phone camera standardization was launched in 2006, when the work of Camera Phone Image Quality (CPIQ) was started by the International Imaging Industry Association (I3A). In 2012 the standardization work

transitioned to the Institute of Electrical and Electronics Engineers (IEEE) P1858 working group. CPIQ phase 1 in 2007 and CPIQ phase 2 in 2009 have already published several proposals for mobile phone camera metrics and the official standard should be published during year 2016.

The offering of image quality tools is quite thin, there are only three tools which are used worldwide. Imatest and Image Engineering companies produce both image quality testing hardware and software tools for image quality measurements, whereas DxO offers comparison and benchmarking tools for DRLS, compact cameras, objectives and mobile phone cameras.

4.2 Traditional objective quality metrics

4.2.1 *Color measurements*

Color measurement is one of the most obvious image quality entities. There has to be a metric that defines how well a camera reproduces colors from the original scene. CIEDE colorimetry standards of the CIE organization have been acknowledged the most usable color difference metrics even though there has been competition between organizations like the Colour Measurement Committee (CMC) and ISO. However, the acknowledgment of the CIEDE colorimetry highlights the fact that the document is approved as a joint international standard between the ISO and CIE organizations (ISO/CIE 11664-6 2014; Habekost 2013).

Color measurements are usually made by capturing images from the Macbeth color chart, whose color values are known precisely (Figure 17). The images are captured in different ambient lights and corresponding color differences are calculated.

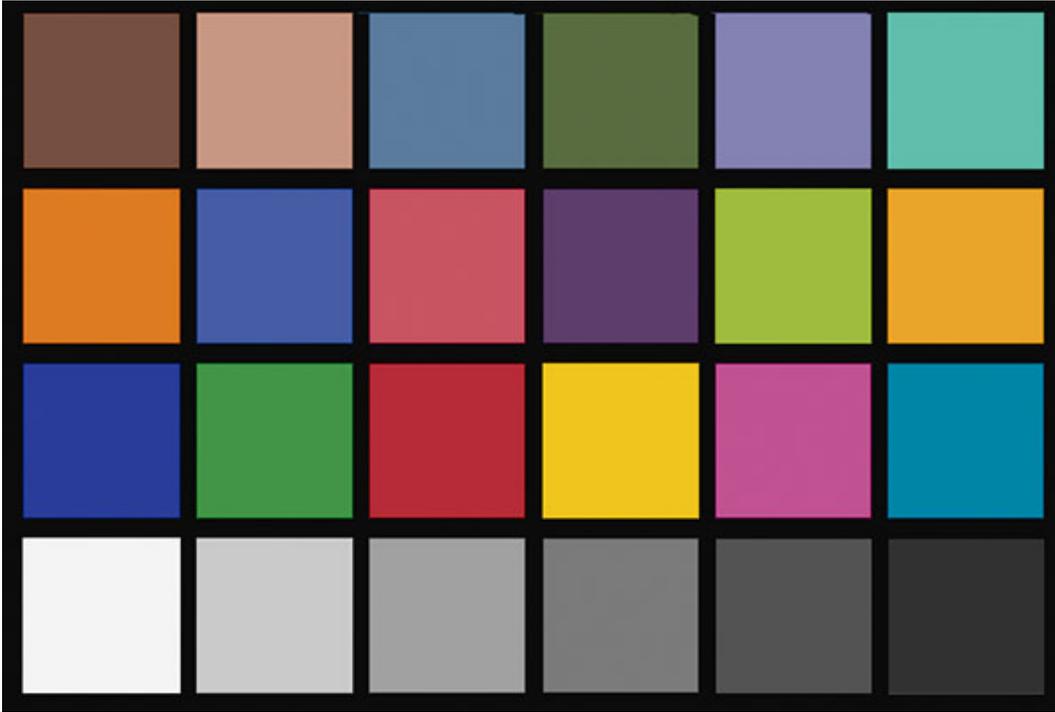


Figure 17 Macbeth color chart

The history of the CIE standardization starts as early as year 1931, but the first color difference standard was published in 1976. The first standard published a metric called ΔE , which defines a color difference between an original scene and a captured image in $L^*a^*b^*$ color space. $L^*a^*b^*$ color space was meant to be a color space which is perceptually uniform (Wyszecki and Stiles, 2000), later on it was revealed, that the approximation was not accurate enough (Mokrzycki and Tatol 2012). It is noticeable, that ΔE metric contains both lightness error ($\Delta L = L_2 - L_1$) and color errors as defined in (4). The equation is, in practice, the Euclidean distance in $L^*a^*b^*$ color space between captured image values Lab_2 and original values Lab_1 .

$$(4) \quad \Delta E_{ab} = \sqrt{(L^*_2 - L^*_1)^2 + (a^*_2 - a^*_1)^2 + (b^*_2 - b^*_1)^2}$$

Here L^* is a luminance value, a^* is a green-red chrominance and b^* blue-yellow chrominance. Normally the $L^*a^*b^*$ values are average values of uniformly colored test patches. The first version of the standardized color difference metric pointed the way to calculate color fidelity. Until now, the metric has been based on the exact difference between known reference values and captured values. However, the importance of the perceptual color quality is starting to change this method.

Since the first version of CIEs ΔE metric, the standard has been updated first by CMC in 1984 and then twice by CIE in 1994 and 2000. When new equations of ΔE

were published in 1994, also new metrics called chrominance error and hue error were established. The latest equation, ΔE 2000 or ΔE_{00} , compensates better for perceptual non-uniformities of the $L^*a^*b^*$ color space and thus correlates better with perceptual color difference than earlier equations (Mokrzycki and Tatol 2012). Equation 5 contains a ΔE_{00} calculation and shows the extent of its evolution since the first version of the CIEDE standard.

$$(5) \quad \Delta E_{00} = \sqrt{\left(\frac{\Delta L}{k_L S_L}\right)^2 + \left(\frac{\Delta C}{k_C S_C}\right)^2 + \left(\frac{\Delta H}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C}{k_C S_C}\right) \left(\frac{\Delta H}{k_H S_H}\right)}$$

ΔL , ΔC , ΔH are lightness error, chrominance error and hue error correspondingly. S_L , S_C , and S_H represent lightness-, chrominance-, and hue-dependent scaling functions. k values can be used to compensate experimental environments. However, in the reference conditions k values are set to value 1. Finally, R_T is a rotation function dependent on hue and chrominance and compensates the hue angle characteristics especially in case of blue color. If (5) is expanded to use only $L^*a^*b^*$ values and k parameters, it becomes extremely complicated.

Zhang and Wandell have suggested adding a spatial extension to the ΔE_{ab} color difference measurements and they named the result S-CIELAB or ΔE_s . The extension transforms an image into an opponent color space and each color space is filtered by a visual spatial sensitivity function of the color space. The visual spatial sensitivity function mimics the human vision system and highlights the color differences for frequencies to which the eye is most sensitive. (Zhang and Wandell 1997)

In year 2003, Johnson and Fairchild improved the S-CIELAB method to work with CIEDE2000 equations (Johnson and Fairchild 2003). Even though the spatial extension of the color difference measurement is not yet accepted to the color difference standards, the same approach is still used in visual noise measurements defined in section 4.2.2.

Changing ambient light makes the color measurement challenging. Whenever the light temperature of the captured scene changes, the spectrum of the luminated light (reflected light from the scene) also changes. This means that the camera system has to adapt to the ambient light and adjust the colors to be same, even if one picture is captured in sunlight and another in fluorescent light. The algorithm, auto white balance (AWB), is one of the most difficult feature to implement in the image processing pipeline. Human brains are extremely good at transforming the visual signal from the eyes according to the ambient light. If the camera system fails to adjust colors correctly, the error is very visible to the human vision system.

Even though ΔE_{00} is accepted by the ISO standardization organization, ISO still has its own standard to describe color accuracy, ISO 17321. This standard defines a sensitivity metamerism index (SMI), which measures color error of the image. The maximum value of SMI is 100 which means a perfect color accuracy, in practice this would mean the camera system mimics exactly the human vision system. As an example of the SMI scale, the standard defines value 50, which represent the difference for a certain color illuminated in daylight or in fluorescent light. (ISO 17321-1 2012)

The SMI is calculated as defined in (6).

$$(6) \quad SMI = 100 - 5.5\overline{\Delta E}_{abi}$$

Here ΔE_{abi} is the mean of the color differences calculated according to CIEDE ΔE_{ab} from year 1976, but using only eight color patches. As the formula 6 describes, the SMI can have also negative values. (ISO 17321-1 2012)

According to the DxO's measurements, DSLRs get SMI values between 75 and 85, whereas low-end cameras reach 40. DxO defines SMI as a not very discriminating metric and uses it as an informative value. (DxO Color sensitivity)

4.2.2 *Noise measurements*

Nowadays, the noise measurements of the digital cameras are mainly based on ISO 15739 standard. Indeed, the latest version of the standard is one of the most straightforward and informative standards in the digital imaging scene. The objective noise measurement is based on noise and signal to noise (SNR) calculations of uniform gray patches of a test chart as defined in Figure 18 (ISO 15739 2013). When the reference target is uniformly gray and correctly illuminated, all variations in the captured image can be judged as noise. Obviously, the measurement cannot isolate specifically the source of the noise, as it can be the sensor, lens system or image processing pipeline.

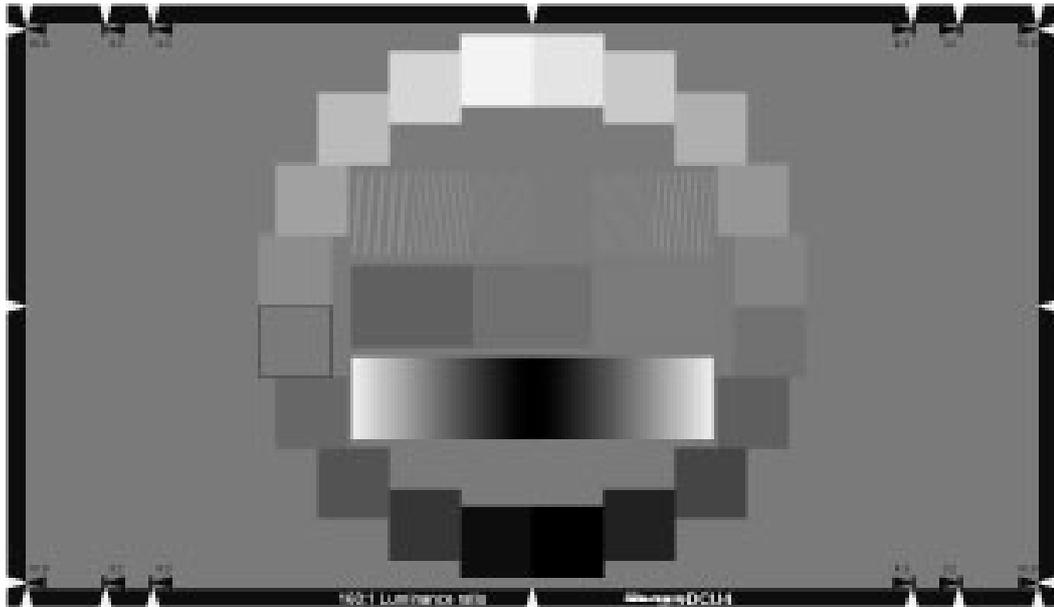


Figure 18 ISO 15739:2013 noise chart (Danes Picta)

However, the noise measurement can isolate fixed pattern noise and temporal noise and calculate corresponding SNR values. Some assumptions can be made according to this division, because the source of the fixed pattern noise is mostly the sensor and in some cases, the lens system. The fixed pattern noise can be isolated from the temporal noise by capturing several images, at least eight according to the standard, and calculating a mean image. If the mean image contains noise, it can be classified as fixed pattern noise (ISO 15739 2013).

ISO 15739 standard was updated in year 2013, the previous version was from year 2003 (ISO 15739 2003). As in several other standards, noise measurements have been changed due to the importance of perceptual quality. The latest version of the standard has specified visual noise measurements.

Visual noise measurements mimic the human vision system (HVS). Visual noise measurement is based on three main components. Firstly, the usage of an opponent color space AC_1C_2 , which includes luminance (A), green-red (C_1) and blue-yellow (C_2) color channels. Secondly, an evaluation of the opponent color channels in the frequency domain and filtering the channels by a contrast sensitivity function (CSF). The contrast sensitivity function is a commonly used model for the frequency response of HVS. The CSF of the HVS has a band-pass nature and the peak of the filter is about four cycles per degree of a spatial frequency. The noise particles which have this spatial frequency are the most visible ones. Thirdly and finally, a conversion back to the spatial domain and a color conversion to the CIE

$L^*u^*v^*$ color space. The visual noise is calculated using the standard deviation of each gray patch of the test chart. (Peltoketo 2015)

Also the corresponding CIE standard can be used for visual noise measurement, the measurement steps are equivalent to the latest version of ISO 15739, although the final calculation is made in CIE $L^*a^*b^*$ color space, whereas the ISO standard uses CIE $L^*u^*v^*$ color space and the visual noise result is calculated using ΔE (Kleinmann and Wueller 2007; Johnson and Fairchild 2003).

The author has made a comparison between SNR values and corresponding visual noise measurements in his article (Peltoketo 2015).

4.2.3 *Dynamic range measurements*

Dynamic range measurements have a surprisingly small role in the standards, though they are quite an important feature of a camera system. ISO 15739 standard, which was discussed in section 4.2.2, contains only one page for dynamic range calculations. Though it has to comply with the specification of the standard, dynamic range is quite easy to calculate: Two values are required to calculate dynamic range. Firstly, the highest luminance value, which does not generate saturated pixels and secondly, the lowest luminance value, which gives signal-to-temporal-noise ratio 1. Dynamic range is the ratio of these two values. The measurements are made using a test chart with twenty uniform gray patches as in Figure 18. (ISO 15739 2013)

Probably due to the modest definition of the standardized dynamic range, both Imatest and Image Engineering test tools state that they use their own, moderate version of dynamic range measurement, even if they are based on ISO 15739 standard.

It is notable, that even though the dynamic range specifies the tone scale of the camera system, it does not specify how the different luminance levels are distributed in the opto-electric conversion function (OECF). Algorithms like gamma correction and tone mapping may distort the function in such way that it causes, for example, a banding effect in the black and bright end or decreasing contrast in the middle tones. Currently, there is no specific metric for the tone representation, even though ISO 14524 defines the measurement methods of the OECF calculations (ISO 14524 2009).

4.2.4 Resolution measurements

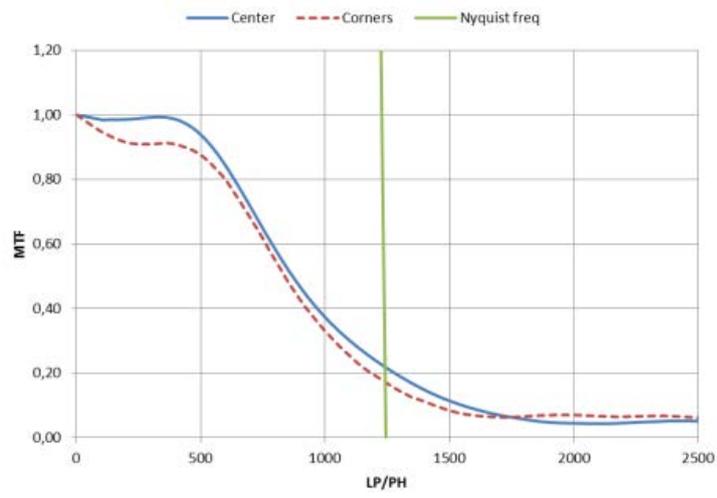
Quite often the resolution of the camera system is only equated with the pixel count of the sensor. Even though the pixel count has a strong correlation with final resolution of the camera system, it is only one factor in the resolution entity. The lens system has a significant impact on the resolution as well as the image processing algorithms. Since the final resolution is a product of the camera components, the evaluation of the result is more complicated.

It has been fascinating to follow the progress of resolution measurements and measurement standardization, because they represent well the race between image processing algorithms and quality measurement methods. The first resolution measurement standard for digital cameras, ISO 12233, was published in 2000, where the resolution measurement was based on a modulation transfer function (MTF) of a high contrast, slanted edge type test chart (ISO 12233:2000). The MTF method and slanted edge charts were very usable ways of measuring the resolution of digital cameras until the cameras began using artificial sharpening algorithms to improve the perceptual sharpness in images.

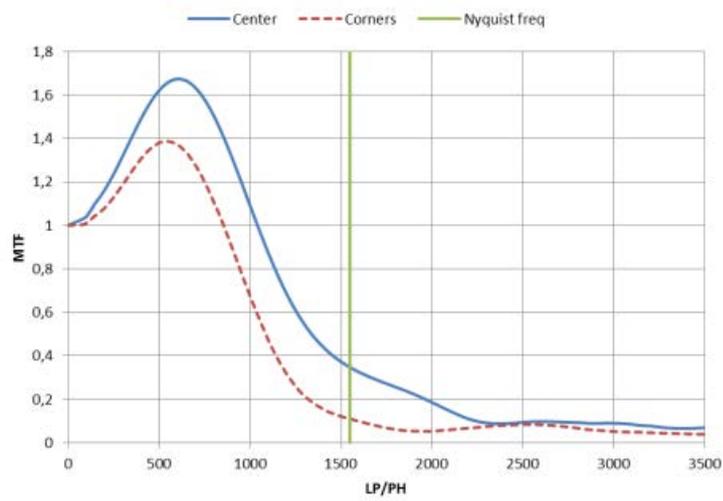
Artificial sharpening is a method where edge areas in the image are highlighted by adding artificial contrast to the edges. It is interesting to note that artificial sharpening is the same method the human vision system uses to separate outlines even in a poorly luminated environment, so called Mach Band effect (Umbaugh 2005). However, the use of artificial sharpening distorts the results of the MTF method when resolution is measured from high contrast slanted edges (Imatest Sharpening).

Figure 19 shows MTF examples from three mobile phone cameras with different pixel counts, sharpening methods, and overall resolution. Device (a) has a very discreet sharpening without any risk of over sharpening. On the other hand, device (b) has one of the strongest artificial sharpening (a bump in the MTF curve) of the measured devices. Finally, device (c) has significant problems with resolution, even though it has clearly the highest pixel count.

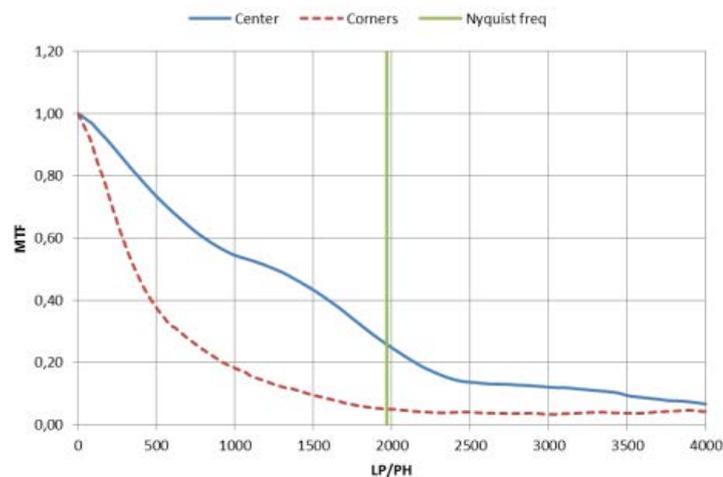
Some assumptions about the lens system quality can be made, when the center and corner resolutions are compared. Also the risk of aliasing and Moiré artefacts can be evaluated from the MTF level after Nyquist frequency.



(a)



(b)



(c)

Figure 19 MTF curves of three mobile phones captured from a low contrast slanted edge chart: (a) Very discreet sharpening, 8 mega pixels, (b) Over sharpening, 13 mega pixels (c) Poor resolution performance, 20 megapixels.

Artificial sharpening algorithms do not highlight low contrast edges as much as high contrast edges and therefore the test charts of the new standard are based on low contrast edges (ISO 12233 2014). Despite this the sharpening is still visible in low contrast slanted edge chart as shown in Figure 19. The new version of the standard defines two different test charts usable for the MTF calculation: a slanted edge type chart and Siemens star based one, see Figure 20. The sinusoidal Siemens star chart should be much more immune to artificial sharpening (Artmann 2015), even though contradictory measurements have been also published (Imatest Slanted-Edge versus Siemens Star). However, where the slanted edge method can measure only one resolution angle at a time, the Siemens star method can measure several.

Another notable change in the 2014 version is that it even contains three different test charts and also the old version of the test chart is kept as an informative annex. The reason for three different test chart can be found from the competition between different test algorithms and also competition between different test companies. The result, a standard with three different measurement methods, is quite a lamentable compromise.

At the same time as artificial sharpening issues were found, it was noticed that the sharpness of edges were not the only resolution metrics that should be measured (Artmann and Wueller 2009; Cao et al. 2009; CPIQ texture metrics 2009). When

aggressive denoising algorithms are used, they may corrupt texture areas of the images. For example, leaves, sand and other natural compositions which look like noise were filtered out by the denoising algorithms. To reveal and measure the texture artefact, so called dead leaves or spilled coins test chart was developed. The method was based on a statistically computed test chart which contains different sized circles, mimicking dead leaves on the ground. If the denoising algorithm is too aggressive, it starts to filter out the smallest elements of the chart and the filtering amount can be measured.

Again, the first version of the texture resolution measurement was revealed to be inaccurate. When the captured image has a significant amount of noise, the noise particles were recognized as the smallest circles and the corresponding texture resolution result was too good (Artmann and Wueller 2012). Artmann and Wueller suggested measuring the noise of the image from a uniform gray area of the test chart and suppressing the noise from the dead leaves chart result accordingly. This method was acknowledged for a while, until it was noticed that some noise removing algorithms remove noise much more efficiently from the uniform areas than other parts of the image. When the noise of the gray area was lower than the noise of the dead leaves chart, the method still gave too good results. Finally, the latest suggestion of measuring the texture sharpness originated from the Kirk et al. paper, where the noise is calculated from the dead leaves test chart itself and a cross-correlation is calculated between the captured image and the original test chart data (Kirk et al. 2014). This seems to be a very good approach, but it requires a full references based approach, which is a very demanding testing method. Figure 20 describes part of the evolution of resolution measurement during recent years.

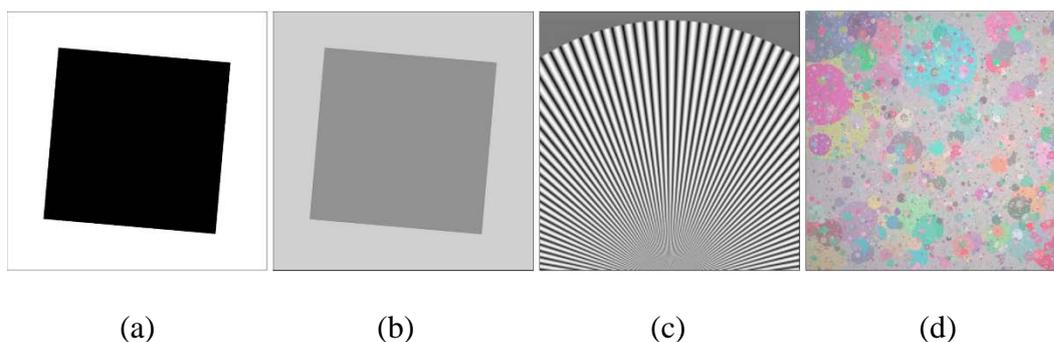


Figure 20 Examples of the resolution test charts: (a) High contrast slanted edge, (b) Low contrast slanted edge, (c) Detail of sinusoidal Siemens star and (d) Colored dead leaves. The image is based on paper by Peltoketo 2014

It is especially notable that several lens based artefacts affect sharpness differently depending on the distance from the optical axis. Thus it is reasonable to measure resolution at least from the center of the image and from the corners of the image.

The first version of the ISO 12233 standard defined a limiting resolution where the resolution response drops to 5% towards a reference response measured from the line width/picture height MTF curve. The most recent version of the standard does not define any limiting resolution but repeats the old version by highlighting the importance of the whole MTF curve. This is reasonable, because the MTF curve reveals much more data than a single resolution value, like sharpening and probability of aliasing. Due to the lack of the exact limiting threshold specification, several different MTF values are used to specify resolution using a single number. *MTF50*, *MTF10* and *MTF5* are used where the number (50, 10, 5) represents the value where the contrast is decreased to that specific percentage. Also peak values are used i.e. *MTF50P* where the decrease is not calculated from the initial value but from the peak of the MTF curve.

As a summary, two different resolution measurement metrics are mostly used and acknowledged nowadays. The MTF metrics, which are based on slanted edge charts or Siemens stars and the texture resolution measurement based on the dead leaves chart. The texture resolution metric is not yet part of any official standard, even though the CPIQ group has proposed it.

4.3 Metrics for image quality artefacts

The standardization work of formal metrics for digital imaging artefacts is not so mature than the traditional image quality metrics. Camera Phone Image Quality (CPIQ) group published Phase 1 in 2007 which included proposals for a color uniformity and flare testing (CPIQ Phase 1 2007) and Phase 2 in 2009 with geometric distortion and lateral chromatic aberration test proposals (CPIQ Phase 2 2009). Several ISO standard proposals are currently under development or have been published during recent years and the CPIQ group plans to publish the first official version of their standard in 2016 including many metrics for the imaging artefacts.

4.3.1 Lens distortions

There are several metrics to validate the geometric distortion of a lens system. European Broadcast Union (EBU) defined a picture height distortion metric in 1995, the work was based on ISO 9039 standard and uses same metric (EBU Tech

3249–E 1995; ISO 9039 1994 and 2008). Standard Mobile Imaging Architecture (SMIA) organization has defined a slightly different metric. The ISO version is based on the ratio between the vertical distortion error of the corners towards a non-distorted image, whereas the SMIA version defined a ratio between the biggest vertical distortion and the smallest vertical distortion. A slightly moderated version of the SMIA metrics is used in CIPA DCG-002 standard.

CPIQ group has defined another approach to the geometric distortion metric, which is strongly based on the test chart used: a white test chart filled with black dots whose locations are known precisely as shown in Figure 21. There are several benefits from the approach; the geometric distortion can be calculated from several locations of the image and therefore it can be modelled by a polynomial. Using the polynomial the distortion can be defined as a function of the distance from the image center point. The function can be used to correct the artefact. Moreover, the same approach and chart can be used for measuring lateral chromatic aberration. CPIQ group describes a metric called Local Geometric Distortion (LGD) which can be calculated from every dot of the test chart. (CPIQ 2016)

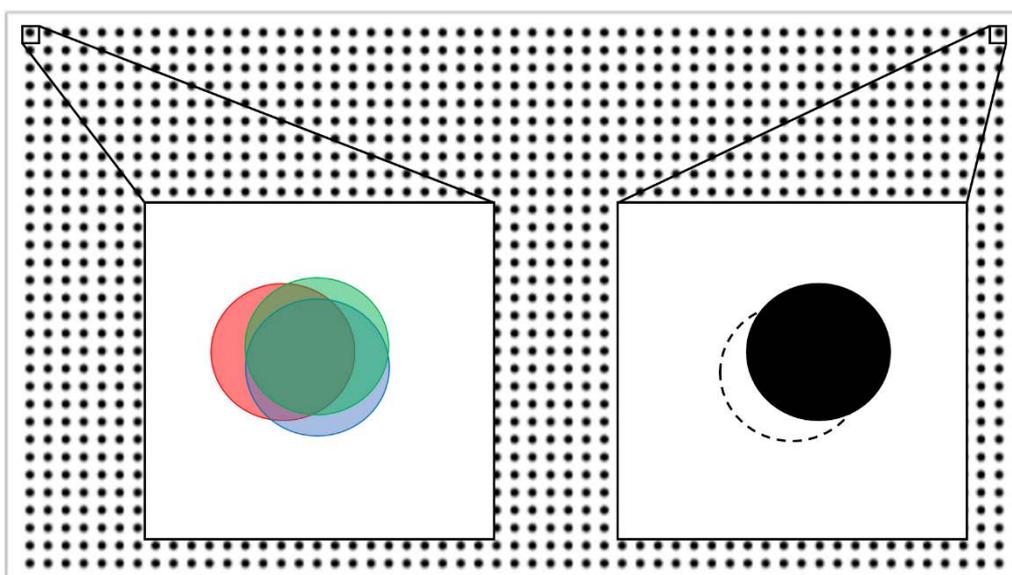


Figure 21 Lateral chromatic aberration and geometric distortion in a dot test chart.

Moreover, the ISO organization has been published very recently a new standard for digital cameras called ISO 17850, Photography - Digital cameras - Geometric distortion (GD) measurements. The standard follows the metrics of ISO 9039 but instead of a single lens system, it takes into account the whole camera system. It is obvious that CPIQ standard and ISO 17850 have been written simultaneously,

because several chapters of both documents are almost equal. In addition to LGD metric, ISO 17850 defines another geometric distortion metric called line geometric distortion, which is not part of the CPIQ standard. The line geometric distortion is the same as in SMIA standard, it describes the ratio between minimum and maximum vertical or horizontal geometric distortion error (ISO 15780 2015).

In case of lateral chromatic aberration, the positions of red and blue color channels are measured from the green channel, and the difference is modelled again as a function of the distance from the image center point. The worst distortion between the channels proportion to image height is defined as the lateral chromatic displacement metric (LCD). (CPIQ Phase 2 2009)

On the other hand, the ISO organization has been published very recently a standard for the lateral chromatic aberration: ISO 19084, Photography - Digital cameras - Chromatic displacement measurements. ISO 19084 defines chromatic displacement and radial chromatic displacement metrics, where the chromatic displacement equals to CPIQ LCD metric. The radial chromatic displacement metric is equal for chromatic displacement but it is described as a function of distance of the green channel to the image center. Moreover ISO 17850 introduces another test chart, a V pattern chart. (ISO 17850 2015) Again, a significant equality between CPIQ standard and ISO 17850 can be noted.

4.3.2 Vignetting and color shading

Vignetting and color shading metric examples were defined already in CPIQ Phase 1 documentation (CPIQ Phase 1 2007). Since then, they have acquired more details in Phase 2 and finally in ISO 17957 standard. Vignetting and color shading are calculated from a neutral gray chart using several light sources. Whereas the latest CPIQ version proposes two light sources, an outdoor light and incandescent light, ISO 17957 adds a fluorescent light to the list. Both standards divide the captured image into 18-20x15-32 blocks, depending on the aspect ratio of the image and the color shading is calculated as defined in (7).

$$(7) \quad D(i) = \sqrt{(a(i) - \bar{a})^2 + (b(i) - \bar{b})^2}$$

Here $D(i)$ is a deviation of a block, \bar{a} and \bar{b} are averages of the whole image and $a(i)$ and $b(i)$ are average values of blocks using the $L^*a^*b^*$ color space. The maximum color shading of all blocks is calculated and the maximum values is reported as the color shading metric of the image. The same method can be used for vignetting measurement by using the L^* component of the $L^*a^*b^*$ color space (ISO 17957 2015).

4.3.3 *Flare and blooming*

Two different flare metrics are defined in ISO 9358 standard. A veiling glare index (VGI) and glare spread function (GFS). The veiling glare index is captured using a white test chart which has a small, complete black element in the middle of the chart. The veiling glare is calculated from the ratio of the black element's luminance value towards the luminance value of the white area. Both values should be normalized by removing the gamma correction. (ISO 9358 1994)

The VGI test is quite simple and straightforward to build and execute, for example Imatest and Image Engineering and CPIQ Phase 1 use variations of the VGI test.

The glare spread function is a more complicated metric, requiring moving test equipment and a sensor, with a dynamic range of 13-19 f-stops. The method is based on a small, moving light source directed towards the camera system. When the light hits the camera system at different angles, even outside the camera's field of view, the flare effect can be calculated from the luminance values, which are a function of the light source's angle. (ISO 9358 1994)

CIPA DCG-002 standard defines also a flare metric, of which the test environment is identical to the VGI metric but measured using several exposure times (CIPA DCG-002 2012).

There seems to be a lack of blooming metrics. Theuwissen has proposed a metric based on a calculation of light spread across the image, where the scene contains a bright object (Theuwissen Blooming). However, no such standardized metrics can be found for still imaging. International Electrotechnical Commission (IEC) has defined blooming measurements for video cameras, however, the standard dates from 1997 and has not been updated since (IEC 61146-2 1997).

4.3.4 *Other artefacts*

There are no specific metrics for noise related artefacts like bad pixels, green imbalance and maze patterns. They are part of the noise results defined in section 4.2.2. Moreover, artefacts causing defocus and blur are validated in the resolution measurements as well as the over sharpening artefact.

4.4 From objective to subjective metrics

Image quality standards have been based on the objective quality metrics and measurements as Keelan defines the objective quantities (Keelan 2002). However,

a good correlation between objective and subjective, or perceptual, metrics is essential. Therefore several standards have been updated to correlate better with perceptual quality or transform objective metrics into subjective ones.

Keelan's book defines a conversion method, integrated hyperbolic increment function (IIHF) and the function has been used as an important reference in several standard updates. Also a generic perceptual metric, just noticeable difference (JND) is used more and more, when subjective image quality metrics are defined. This section summarizes the latest changes to the standards.

Since CIE $L^*a^*b^*$ color space was originally developed to be a perceptually uniform color space, the corresponding color metrics have the longest history of the perceptual metrics. All CIE color difference metrics were supposed to be perceptual metrics, but the accuracy of the first metrics were not sufficient. As defined in section 4.2.1., the color difference metrics have been updated several times since the first version of CIEDE in 1976. Still the latest version of the standard calculates only the difference from the original scene and does not take into account the possibility of improving the original colors to get more pleasing results. CPIQ group has proposed a color metric, which allows a camera system to increase the chrominance level of images.

There has been also a proposal to add a spatial extension to the CIEDE calculations (Zhang and Wandell 1997). The addition would use the contrast sensitivity function of the human vision system to highlight the faulty color particles which are most visible. However, the addition is not yet part of any color standard.

The visual noise was accepted as a standardized metric in year 2013, when ISO 15739 standard was updated. Visual noise measurements are based on the opponent color space and contrast sensitivity function of the human vision system and it is used as such, for example, in CPIQ proposals.

The first version of the ISO resolution standard, ISO 12233 from year 2000, already included a visual resolution part, a hyperbolic zone plate. Since the slanted edge method, or later on, Siemens star are quite difficult methods to validate the resolution without a significant amount of calculation, a visual check of the hyperbolic zones provides an integer value, which can be used as a resolution reference as shown in Figure 22. The 2014 version of the standard defines a specific test chart for the visual resolution test and CIPA DC-003 standard introduces the algorithms to validate the chart. (ISO 12233 2000; ISO 12233 2014; CIPA DC-003 2003)

Even though ISO 12233:2014 defines sharpness and acutance metrics as a subjective impression of the resolution it does not specify any transform functions. On the other hand, ISO 20462 part 3 provides ready functions between modulation transfer function curves (MTFs) calculated from slanted edge or Siemens stars and just noticeable difference (JND) values. The functions are used in CPIQ phase 2 documentation and later work of the CPIQ group. (ISO 20462-3 2012; CPIQ Phase 2 Acutance 2009)

As a detail, DxO has its own subjective resolution metric called Perceptual MPix. The metric informs how much the resolution decreases due to lack of quality and due to camera system artefacts affecting the maximum resolution of the sensor.

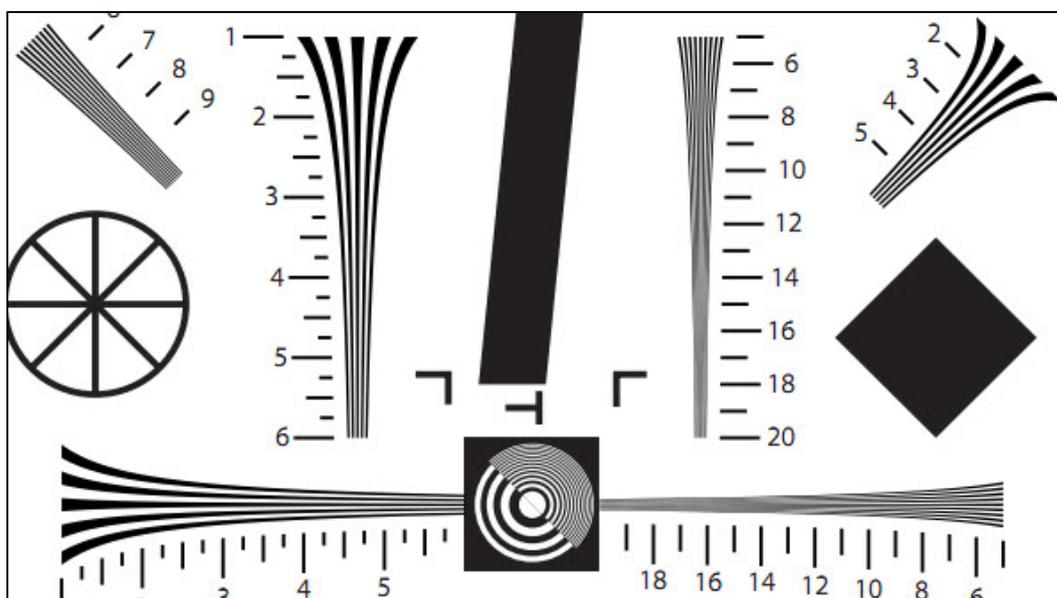


Figure 22 Hyperbolic zone plates of ISO 12233:2000 test chart.

In general, the CPIQ group has done significant work investigating the conversions of objective metrics into perceptual ones. Several former objective metrics have been studied and conversion functions have been implemented to change objective metrics into just noticeable difference (JND) values. Even though the upcoming standard is targeted as a benchmarking standard, it is also a notable step towards perceptual image quality measurement.

Though the trend seems to be to create perceptual image quality metrics, there is still a strong need for objective ones, too. Objective metrics help designers adjust and parametrize camera systems better than the perceptual ones because they correlate better with features and parameters of the camera hardware and software. Objective measurement results are an important tool when a new camera system is being designed and implemented.

4.5 New features and algorithms require new metrics

As defined in section 4.2.4, the race between camera algorithms and test methods is continuous and, perhaps, endless. New algorithms improving some quality entity may reduce another one and create new artefacts which have to be evaluated.

Also new features of camera systems force standardization organizations to create new metrics. The imaging industry creates new features and additions to existing features at a breathtaking speed. Inventions like new color filter arrays, stacked and quantum sensors, lensless cameras, liquid lenses, multiple cameras and 3D cameras, hyperspectral imaging, and curved sensors require old metrics but also very new ones.

The challenge of standardization is the delay between the arrival of new features as above and the adoption of algorithms for corresponding standardized metrics. For example, the first optical image stabilization standard, CIPA DC-X011, was accepted 18 years after the first product including the feature (CIPA DC-X011 2012). Moreover, the first white paper of CPIQ group was introduced in year 2007 and the first official standard of the group will be published in 2016. The development of camera systems has been huge during those nine years. Still, it has to be admitted that most of the metrics defined in the first white paper of the CPIQ group are still valid.

4.6 Video metrics

The role of video recording and live video streaming is increasing all the time in mobile phone usage. It can be foreseen that the video quality measurement will have more and more important role in mobile phone camera evaluation and benchmarking. Even if the video quality is not the main task of this work, it is reasonable to do a brief glance to the video quality metrics.

Skype has defined a comprehensive quality measurement specification for videos. The specification uses parts of the requirements of the ISO, ANSI, ITU-T, VQEG standards and CPIQ studies, but also some of its own measurements and metrics are defined, too. The document includes obvious metrics like spatial resolution, texture sharpness, exposure accuracy, noise measurements, dynamic range, color shading, geometric distortion, color accuracy and frame rate. The acceptance thresholds of the metrics are Skype specific. (Skype 2013)

However, the documentation defines several interesting new metrics for video, though part of them are valid also for still imaging: Temporal noise and SNR are

calculated using consecutive frames and autofocus performance is measured by MTF curves when a camera is forced out of focus and then adjusted back to the focus location. Moreover, video capture delay metrics are defined to validate, how quickly the camera starts to stream video and audio-video synchronization metric measures the delay between the audio and video components. Depth of field, field of view and pixel aspect ratio are metrics, which are valid also for still imaging, but Skype has defined its own threshold values for those measurements. Finally, the document lists several encoding artefacts but does not specify corresponding metrics. (Skype 2013)

Video Quality Experts Group (VQEG) has published a test plan for perceptual quality methods for digital videos. However, it concentrates mostly on the encoding, decoding and transmission issues of videos (VQEG 2008). ITU-T organization has several standards for perceptual video measurements and especially P.910 and P.913 standards define methods and environments for the perceptual quality measurements for video (ITU-T P.910 2008; ITU-T P.913 2014).

Wu and Rao's book *Digital Video Image Quality and Perceptual Coding* provides a breathtaking survey of video compressions, artefacts, quality testing and especially of video perceptual quality measurements. (Wu and Rao 2006).

All in all, there are a significant number of studies, standards and literature which concentrate especially on the perceptual and subjective verification of videos. In contrast, objective measurements seem to have a minor role in video validation.

4.7 Performance metrics

ISO 15781 is a recently published standard defining metrics for evaluating time delays in a camera system. The standard defines performance metrics as shooting time lag, shutter release time lag, start-up time and shooting rate. (ISO 15781 2013)

Start-up time is quite clear and self-explanatory. It defines delay between switching a camera on and the moment when the camera is ready for capturing images. Image shooting rate describes how fast a camera can capture images in a row. The delay is defined as time between beginning of exposure of the first image and beginning of exposure of the next image. The obvious limit for shooting rate is used exposure time, the shooting rate cannot be smaller than the exposure time and thus the limit creates a dependency between these metrics. (ISO 15781 2013)

According to the standard, shooting time lag is the time delay between pressing the exposure button and beginning of the exposure. It is notable that the delay contains

all the camera adjustments, meaning that the delays of auto exposure and auto focus are part of this delay (ISO 15781 2013). In particular auto focus delay can be a significant part of the image capturing time (Peltoketo 2015). The shutter release time lag is the delay from when the exposure button is fully pressed to when the exposure starts. In some cameras this delay can be zero or even negative.

Bucher et al. has made research especially in case of shutter release time of mobile phone cameras. According to the research, cameras predict image capturing by storing frames beforehand and, in practice, several camera models have a negative shutter release time. The result shows that the negative shutter release time can be as big as 250 milli seconds. Also a correlation between exposure time and negative shutter release time was noted (although the corresponding test was done only to single phone model): when the exposure time expanded, also the negative shutter time was increased. Even if the paper concentrated mainly to shutter release time, it did not give any recommendation of acceptable limits of positive or negative shutter release time values. (Bucher et al. 2014)

Bucher et al. highlights also the importance of statistical analysis of results. Measurement of 500 captures in same environment revealed that there can be a significant variance in the results. The shutter release time of one mobile phone camera model varied between 130 and 600 milli seconds which notes that a single measurement is not enough. (Bucher et al. 2014) Surprisingly ISO 15781 does not mention or suggest this approach.

The variance of camera performance metrics has been taken into account in this thesis by averaging several results to final benchmarking metric.

Masson et al. introduces a rolling shutter metric, which correlates straight to the rolling shutter artefacts. A longer rolling shutter delay, a greater possibility to have rolling shutter artefacts. On the other hand, the research reveals dependency between image stabilization functionality and exposure times. This dependency should be taken into account when exposure times are measured and reported. (Masson et al. 2014)

Finally CIPA DGC-002 includes some camera performance metrics. Shutter release time lag and shooting time lag, and shooting rate are same as in ISO 15781 standard. However, CIPA standard includes couple of more metrics to use. A focus speed is separately introduced and it defines how quickly a camera adjust focus before exposure can be started. Furthermore, a shooting interval metric is reserved to situation where a continuous shooting mode is not activated in a camera. (CIPA DCG-002 2012)

The challenge of performance metrics is that they are relative novel metrics. Even if the first standards are revealed and some research have been done, comprehensive user validation seems to be missing. It can be claimed that smaller delay or greater performance is always better, but there is always a limit when the delay is indistinguishable and does not affect to the user experience. Moreover, features like negative shutter release time may cause unwanted results.

5 FROM MEASUREMENTS TO BENCHMARKING

Using current measurements and metrics, mobile phone cameras can be compared feature by feature: colors using ΔE , noise from signal to noise ratio or from visual noise, and resolution from modulation transfer function, for example. The comparisons give various results and ratings when different features are compared. But when a general comparison or rating is made for mobile phone cameras, a more comprehensive tool is required. A tool which would combine different features of a camera and create a more straightforward answer to a question: What would be the best mobile phone camera for me?

The chapter clarifies benchmarking in general and existing benchmarking systems. Moreover, the issues of mobile phone camera benchmarking are discussed and finally, an example solution example for this benchmarking is defined.

5.1 Benchmarking in general

Benchmarking is a term, which varies a lot depending on the framework. However, a comparison against other products or companies has always been an essential part of benchmarking. Probably the first benchmarking process using that name was invented by Xerox to improve their photocopiers in the late 1970s. Xerox had severe quality problems with their X3300 device and they visited their competitors to compare key data. They created two concepts for benchmarking:

- Find the best products and companies (benchmark the data)
- Find out, how the best products are made (learn) (Stapenhurst 2009)

Obviously, from the company and business point of view benchmarking always includes both the comparison and learning parts. The goal is to reach and bypass the competitors or maintain a superior level over competitors. If benchmarking is used as a tool by an independent organization, which does not have its own products in that area or as a tool by an end user, the role of benchmarking is different: the comparison part is emphasized and learning is less important.

There are several standardized benchmark tools which are implemented by non-profit organizations and for which the goal is to offer objective information about different devices. The standardized methods are especially used, when the quality and performance of microprocessor controlled devices are compared and ranked. Embedded Microprocessor Benchmark Consortium (EEMBC), contains several benchmarking metrics for comparing processors, multicore systems, memory, and also embedded systems, like mobile phones. Business Applications Performance

Corporation (BAPCo) has metrics for personal computers like desktops, laptops, and tablets. Moreover, Standard Performance Evaluation Corporation (SPEC) is concentrated on server computers and has also energy saving metrics and Transaction Processing Performance Council (TPC) has benchmarking for databases, big data, and transaction processing. In addition to these, there are dozens of different benchmarking tools for mobile devices, phones and computers.

Usually the device benchmarking methods offer a single number results, which makes device comparison easy and straightforward. Quite often the methods contain also more detailed information giving some background on the metrics used in the final score.

Probably the best known, end user benchmarking product is the Euro NCAP collision and safety test for cars. It offers generic and user friendly five star safety ratings, but also more specific information for five different safety categories (Euro NCAP). CPIQ standardization group plans to create a similar five star comparison for mobile phone cameras.

5.2 Existing benchmarking metrics for digital cameras

Currently, there are three world wide companies making image quality measurement tools for camera systems: Imatest, Image Engineering and DxO. The situation for camera benchmarking tools is even more limited. DxO has DxOMark for DSLRs and optics and DxOMark Mobile which is a dedicated benchmarking metric for mobile phone cameras. Moreover, Finnish startup company Sofica has released its own benchmarking system called Sofica Benchmarking Report.

Up to now, there has not been an accepted method for comparing digital cameras in general and mobile phones specifically. Even though DxOMark Mobile contains seven different measurements for still images and videos expanded with separate visual inspection tests (DxOMark Mobile), none of these tests are public ones and the tests cannot be reproduced outside DxO premises. The individual test results are single integers and they are not standardized image quality metrics like ΔE or signal to noise ratio. Even if the results are undoubtedly correctly measured, there are still unknown weighting factors in results.

Even if digital image quality is widely investigated, lack of scientific articles and research for mobile phone benchmarking is notable. Until now, there has not been lot of interest to make research which would have been dedicated to the area. Even if the research area is quite narrow, the interest and huge usage of mobile phone cameras should emphasize to research this area, too. However, Camera Phone

Image Quality (CPIQ) group has planned to publish the first mobile phone camera related benchmarking standard in 2016. Nowadays the CPIQ standardization work is done by the Institute of Electrical and Electronics Engineers (IEEE) P1858 working group. The goals of the work are: (P1858 2015)

- “Standardize image quality test metrics and methodologies across the industry”
- “Correlate objective results with human perception”
- “Combine the data into a meaningful consumer rating system”

In addition to the image quality metric standardization and perceptual conversion, the CPIQ group will create and manage a mobile phone camera certification program which gives CPIQ certificates to the imaging laboratories. The certified companies are allowed to test and benchmark mobile phone cameras and publish CPIQ certified results. For the final benchmarking or rating is planned to use a five star rating similar to Euro NCAP tests.

After the release of the CPIQ standard, the markets and end users will decide whether the benchmarking will be accepted and taken into use.

It is reasonable to mention that the articles of this work were written without a detailed knowledge of P1858 standardization work, even though CPIQ phase 1 and 2 documentation has been available. During the writing work of the thesis, the author became a member of the P1858 working group and had an access to the archives of P1858. Even though the author has not done any contributions to the P1858 standardization work, he has followed closely the development of the standard and shared the research results with the members of the working group.

5.3 Challenges of camera benchmarking

In case of a simple benchmarking of a processor or system, the result is based on the performance of the system in executing a certain item of test software. In practice, the time and memory usage the software takes to execute a certain algorithm is defined as the performance of the system. When a camera system is benchmarked, the situation is more complicated. There are numerous different quality and speed metrics available and selection and combination of the metrics can be problematic. Also different environments may change the ranking between cameras.

5.3.1 Which metrics to select

The first challenge is to select the metrics to be used in the benchmarking score. The CPIQ group defines the fundamental objective measurements as follows: spatial resolution, tone and color reproduction, sensitivity, noise, and geometric fidelity (CPIQ Phase 2 Introduction 2009).

In case of color accuracy, the selection of the standard is quite straightforward, because the CIEDE metrics are widely acknowledged and used. However, several metrics are still available. Chrominance, hue and luminance differences can be used as well as the generic color difference ΔE . Also the older versions of the CIEDE metrics are still used to some extent to keep the compatibility with old measurements and probably due to the complexity of the latest ΔE_{00} equations. A tone reproduction could be measured from the color difference values of gray patches of the test chart, but it is not defined as a standardized metric. However, some shortcomings can be faced when the ΔE based measurements are made. Some mobile phone vendors add an extra color tint to the images to get a better perceptual impression. Even if this color change is wanted, it increases measured color error. The phenomenon was noticed, when the color measurements of the attached articles were made.

Spatial resolution offers many more metrics to choose from. Firstly, the latest standard defines three different methods and test charts to calculate the spatial resolution. Even though slanted edge and Siemens star methods are based on modulation transfer function (MTF) curves, they do not end up with identical results. Secondly, the standard no longer specifies a limiting resolution, which could be referred to, but describe the whole MTF curve as a result of the spatial resolution. This gives a questionable freedom to different tools when specifying the single number result for resolution. Imatest defines *MTF50* or *MTF50P* as a good reference value. Image engineering defines the limiting threshold as 10% of the initial value and DxO defines the limiting threshold as 5% of the initial value. Finally, Skype video test documentation mentions *MTF30* as a good metric.

Moreover, texture resolution can be defined as a part of the spatial resolution, but there are no standardized metrics for those measurements. The latest de-facto standard, which is used in several tools, is so called dead leaves method defined in section 4.2.4, but the noise compensation methods are still investigated and they vary between tools. Also the resolution related artefacts like over sharpening and aliasing do not yet have specific metrics.

The noise standard ISO 15739 has several metrics for selection: total noise separated to temporal and fixed pattern noise, signal to noise ratio and the latest

version has visual noise metrics. The sensitivity of a camera system can be derived from the dynamic range measurement of ISO 15739 documentation or ISO speed measurements of ISO 12232.

To cover all the fundamental measurements of the CPIQ group, the geometric fidelity can be measured according SMIA metric, using height distortion of ISO 9039 or CPIQ Phase 2 proposal for the distortion.

Unfortunately, there are still several metrics left, even though the fundamental objective measurements of CPIQ have already been discussed: Section 4.3 defines lens distortion metrics like chromatic aberration, vignetting, lens shading and glare which can be measured using ISO standards and CPIQ Phase 2 proposals. Camera speed metrics are discussed in section 4.7 and defined in the ISO 15781 standard and they could be a very valuable addition to the benchmarking. Finally, video related metrics in section 4.6 may, at least, double the number of quality metrics.

Clearly, the number of the metrics is so big and the characteristics of the metrics are so different that some selection or weighting has to be done. Roughly, the number of all metrics including those for video would be more than fifty and it is difficult to imagine a single number score which could include and combine such number of values. The selection of the metrics is a tradeoff between the coverage and complexity of the benchmarking.

5.3.2 Metrics of different environments

The selection of the metrics is only one dimension of the benchmarking challenge. The imaging environment used, the photospace, will affect significantly the quality and performance of the camera systems. Keelan defines a photospace, which has illumination and object distance parameters (Keelan 2002). In case of mobile phone cameras, the most important environment parameter is the illumination. Due to small pixel size and demanding lens requirements, mobile phone cameras are very vulnerable to a low light environment. Moreover, different phone types react to the light changes in various ways as defined in the author's article (Peltoketo 2015). To get a comprehensive benchmarking result, camera systems should be tested in several light environments including both illumination and color temperature changes. ANSI organization has defined low light measurements for video recorders (ANSI/CEA-639 2010), but a corresponding still image standard is not yet available, even if there has been a proposal to create a similar metrics for still image cameras (Wueller 2013).

Flash usage will create another use case for low light imaging. The color of the flash light, luminous power, uniformity, and flash synchronization with the image capturing are elements which affects the flash supported low light imaging. The flash may also generate its own artefacts, like red eyes.

Another dimension of the photospace, the distance to the object, will affect the lens distortions artefacts, focus performance, and depth of focus features of the camera. In particular, near objects are challenging for the mobile phone cameras.

Finally, the movement of the photographed object or the movement of the camera will affect image quality. Camera parameters and features like exposure time, ISO speed, video frame rate, autofocus speed and possible image stabilization will improve or worse the final quality. Säämänen et al. have proposed defining a videospace, similar to the photospace but which also includes movement of the object (Säämänen et al. 2010). Even though the videospace is intended for specifying the quality of a video recording system, it might also be usable for still image testing.

All in all, there are several environmental factors which affect the quality and performance of a camera system and they should be considered part of the benchmarking process.

5.3.3 Perceptual benchmarking

As well as the image quality measurements, the benchmarking result should correlate with the perceptual judgement of a mobile phone camera. This would mean that every metric is either already perceptually adjusted or there is a conversion function which transforms the objective metric into a perceptual one. The final perceptual benchmarking score should be calculated from the perceptual metrics weighted so that the weight factors between metrics should be also perceptually adjusted.

Even though part of the image quality metrics are already perceptually adjusted, for example color differences and visual noise, this work concentrates mainly on objective benchmarking. A true perceptual benchmarking requires another approach to the measurements and benchmarking score equations. The coming CPIQ standard will publish the first perceptual benchmarking system, where every quality metric has been separately converted into a perceptual one. The future will reveal, if the conversion functions are accurate enough. However, the basic assumption of every objective quality metric is that it correlates, at some level, with perceptual quality. Thus, the objective benchmarking score calculated from

objective metrics should correlate also, at some level, with the perceptual benchmarking.

5.3.4 *Several metrics to single score*

To get a straightforward and comparable benchmarking score, several systems use a single number score to express the performance of a device. In case of a camera system several different metrics should be combined into one value, which could be a problematic task.

When the benchmarking includes metrics, which have clearly different effects on the final camera quality, the metrics have to be weighted. For example color accuracy and chromatic aberration cannot be treated as equal sources of a generic benchmarking score. Extreme care should be taken, when the weights are selected and evaluated. Since the final results can be totally manipulated using inappropriate weights, all equations and used metrics should be public.

With or without weighting factors, the individual metrics have to be combined and calculated to get a single number score. Several solutions can be found from literature to solve this issue and arithmetic, harmonic, geometric, and weighted means have been proposed. However, they are related to a system, where the metrics can be normalized or they have the same unit of measure (Fleming and Wallace 1986; Smith 1988; Lilja 2005). Since all averaging methods are misleading at some level, there is no unambiguous solution to the problem. Whichever method is used, it is necessary to reveal all the equations and measurement values that are used in the calculations.

CPIQ standardization group has planned to tackle this problem by transforming all individual metrics to perceptual, just notable difference (JND) values. The JND values represent the quality loss of each metric. Finally, the final benchmarking score, which represents the total quality loss of the system, is calculated using a multivariate equation expressed by Keelan. (P1858 2015; Keelan 2002)

5.3.5 *Practical issues of benchmarking*

The measurements and benchmarking may face several practical difficulties, which are not related to the standardized metrics as such. It is obvious that results will differ between phone models of a phone vendor, but there can be clear differences, even if the model name is the same. A good reference for this issue was Samsung Galaxy S4, where two different models were sold using the same name, S4, but S4

GT-I9500 had Samsung's own Exynos chipset whereas S4 GT-I9505 was powered by a Snapdragon chipset. Even though the camera modules were same, the performance of the Exonys version was significantly better than the other.

Even if the hardware content is exactly same, the software version may yet affect the camera quality. During the measurements for the third article, two different software versions of the Lumia 1020 model gave clearly different results (Peltoketo 2014). Moreover, there can be differences between individuals. For example, slightly different mounting of a lens system towards a sensor may generate clear problems in the resolution. Even if the worst cases are removed during the factory validation, some variance will remain.

If measurements are made automatically, i.e. a test system both captures required images or videos and calculates the results, the automation requires a software interface with the camera. Mobile phone vendors offer a public interface, but sometimes they have also their own, proprietary and hidden interfaces, which work better with their own applications. This leads to a situation, where third party benchmarking will have different results than a method, which could use proprietary interfaces. Especially in speed and performance tests, the optimized and proprietary interface may improve results significantly.

5.3.6 Static benchmarking, compatibility requirement or trap?

When mobile phone cameras are compared with previous versions, the benchmarking score and corresponding metrics should be compatible. The easiest way to guarantee this requirement is to make a static benchmarking system, which has constant metrics and the final score is always calculated in the same way.

However, this can be also a trap. If the benchmarking score and corresponding metrics are not updated, the score will become sooner or later out of date and it will no longer be valid. Even if the main quality metrics like colors, noise and sharpness do not change, the image processing algorithms are changing. The result may show that the quality fundamentals are in a good shape, but some other artefacts have appeared. Good examples of the phenomenon are the denoising and over sharpening artefacts. New camera models will also offer new features, which have to be validated and compared between models.

According to the Recon Analytics, the average lifespan of mobile phones varies between countries. In the United States it is as short as 22 months, whereas in India it is over seven years (Recon Analytics 2011). The variance forces benchmarking

systems to be flexible, new features have to be taken into account but also the compatibility with old models has to remain.

5.4 Proposal for mobile phone camera benchmarking

The reprinted articles II-V include evaluation and define a proposal for mobile phone camera benchmarking. The main findings, evaluations and proposals of the articles are described in this section.

Since the benchmarking score is a selection and combination of several metrics, both selection and combining have to be fair. Not only can weighting of the metrics skew the result, but also selection of or exclusion of certain quality and performance metrics may unfairly treat different camera models. For example, using the pixel count of the sensor as a resolution metric would raise the rank of the Lumia 1020 phone significantly, even though the pixel count does not describe the real resolution of the camera system. On the other hand, excluding speed metrics from the benchmarking would reduce the score of iPhone5 phone, which has notable performance features.

The initial approach of the work was to combine both quality and performance metrics to get a more comprehensive benchmarking score. Also it was noted, that weighting of the metrics was a very tricky task and, in practice, it would require a perceptual validation to adjust weight correctly. These thoughts led to an idea about a way of selecting such metrics, which would not require weights when they are combined.

As discussed before, the CPIQ group defines that the fundamental metrics of the image quality are spatial resolution, tone and color reproduction, sensitivity, noise, and geometric fidelity. In general, colors, resolution and noise are defined as the main image quality features. On the other hand, currently the only existing camera speed standard, ISO 15781 defines metrics which could be used to represent the performance of a camera system.

Two metrics were selected for the resolution, *MTF50P* values of the spatial resolution and texture sharpness. Using the peak values, the impact of the artificial sharpening can be reduced. The artefacts of the denoising can be detected using the texture sharpness. The 50% value is not an unambiguous result for the spatial resolution. However, when evaluating the results, the threshold values like 10% and 5% were located beyond Nyquist frequency and therefore they represented the possibility of Moiré artefact more than the resolution of the camera system.

According to the ISO 15739 standard, signal to noise ratio (SNR) and visual noise were selected to represent noise characteristics of the system. However, the fifth article shows that the visual noise metric follows well the SNR based metric but better represents the visual perception of the noise. According to this, the visual noise could be enough, when the noise of the camera system is benchmarked.

The ΔE_{00} metric was used to validate the color accuracy of a camera system. Since the metric contains also the luminance difference, the exposure accuracy was measured at the same time. To highlight the importance of the auto white balance and color shading correction, an extra color metric was added. The chrominance error of gray patches of the Macbeth test chart was calculated and used as another color fidelity metric.

The ISO 15781 standard defines five performance metrics for a camera system. The metrics can be combined into two metrics to get usual use cases of a mobile phone camera. Firstly, a capture time of a single image including the startup time of the camera and secondly, shooting rate, which defines how fast several images can be captured. Moreover, the audio-visual synchronization delay was selected to the delay metrics to represent the performance of the video area. Still, introducing only one video metric is definitely not a sufficient way to include video functionality in the benchmarking score.

To reveal outliers in camera performance measurements, every measurement was made at least five times and an averaging value was used as a result. If clear outliers were detected, they were carefully inspected to be real measurements and not spurious results produced by the testing environment. According to the research by Bucher et al. greater number of measurements would be reasonable (Bucher et al. 2014). However, Bucher et al. do not specify when the results start to stabilize in statistical point of view. It requires more research to find the optimal amount of measurements which fulfill statistical requirements.

The combination of different metrics into a single score was a tricky task and there is no ambiguous equation for that. Obviously, an arithmetic mean cannot be used because metrics do not have same scale. Arithmetic mean with normalization towards the maximum values of each metric could be a solution, but this approach generates own problems. When a new camera model has some superior feature and offers a new maximum value to the equations, scores of all other devices will change which would be a very confusing situation.

Especially Fleming and Wallace suggest to use geometric mean to combine different metrics to one score (Fleming and Wallace 1986). Even if geometric mean (8) looks very simplified solution for combining several different metrics together,

evaluation using the geometric mean revealed that balance between used metrics and, on the other hand, balance between quality and performance metrics can be achieved.

$$(8) \quad Score = \left(\prod_{i=1}^n a_i \right)^{1/n}$$

In the (8) n is the number of metrics and a is the specific metric. (9) and (10) describe the geometric means when they are expanded to calculate image quality score $Score_q$ and camera performance score $Score_p$ correspondingly. The total benchmarking score collects both image quality metrics and camera performance metrics and calculates a geometric mean using all metrics of (9) and (10).

$$(9) \quad Score_q = \sqrt[6]{MTF50P_{edge} \times MTF50P_{deadLeaves} \times SNR \times \frac{1}{VN} \times \frac{1}{\Delta E_{00}} \times \frac{1}{satErr}}$$

$$(10) \quad Score_p = \sqrt[3]{\frac{1}{t_{single}} \times \frac{1}{t_{five}} \times \frac{1}{t_{AV}}}$$

$MTF50P_{edge}$ and $MTF50P_{deadLeaves}$ represent spatial resolution and texture sharpness correspondingly. SNR and VN characterize the noise of a camera defining the signal to noise ratio and visual noise. ΔE_{00} and $satErr$ are color metrics describing luminance and chromatic error (ΔE_{00}) and white balance error ($satErr$). Finally t_{single} , t_{five} , and t_{AV} represent capture time of single image, capture time of five consecutive image, and audio-video synchronization delay.

An obvious weakness can be seen when this approach is used: the metrics in denominator cannot be zero.

Probably, the perceptual transformation and multivariate equation presented by CPIQ might result in a better perceptual combination of the metrics. However, perceptual benchmarking is outside the scope of this work.

When evaluating different environments in the fourth article, it was noted that the low light environment affects the benchmarking results. It should be considered, whether different environment factors like object distance and movement of the object should be measured, too. Having a single benchmarking score including several environments was judged to be too complicated. The interpretation of such a score would be very difficult. A better approach would be to have separate benchmarking for each environment.

Validation of any benchmarking metric is a tricky task because used metrics and algorithms varies. There is only one single score benchmarking method for mobile

phone cameras in the market, DxOMark Mobile, which is not a public one but the equations are proprietary. Therefore it is almost impossible to compare how different metrics will affect to the final score. If the ranking order between DxOMark Mobile and this research is compared, results are contradictory. When the rank order of five selected mobile phone model in the third attached article is compared towards DxOMark, the ranking is almost equal. On the other hand, rank of single mobile phone models can vary significantly between these methods. (Peltoketo 2014; DxOMark Mobile)

The final judgement of any benchmarking metric should be done with consumers. If the consumers' experiences correlate to the benchmarking result, the benchmarking system can be considered to be viable. A milestone of mobile camera benchmarking will be reached during year 2016 when CPIQ will release own standard of benchmarking and markets will decide, if it will be used globally.

The evaluations and trials over number of years highlighted several conclusions and requirements of a benchmarking system for mobile phone cameras. The items are discussed more detailed in chapter 7.

6 INTRODUCTION TO ORIGINAL PUBLICATIONS

This chapter includes introductions to the five articles, which are reprinted in the end of the thesis. The publications were published between years 2012 and 2015 in conferences and international journals. All articles were peer-reviewed before publishing. Even though many references in the articles are the same as the references in this thesis, they are not explicitly included here.

6.1 Article I: Objective verification of audio-video synchronization

The article was published in the Workshop on Wireless Communication and Applications, WoWCA in 2012. The article defines the challenges of audio-video synchronization in video streams and how to measure the delay between audio and video components.

Audio-video delay, so called lip synchronization, is one of the video recording and broadcasting artefacts. A human being is very sensitive to the delay between audio and its corresponding visual signal. A delay of 20 milliseconds can be detected when the audio leads the video signal. The paper describes an implementation and a validation of the audio-video synchronization measurements based on existing methods.

The algorithms used were based either on a full-reference measurement, where the original video stream was available or on a reduced-reference method, where hash signatures were calculated beforehand. The hash signatures were calculated from a certain amount of video and audio samples from the reference and processed stream. Using hamming distance calculations, the corresponding hash pairs were detected between reference and processed data. The information was used to measure the audio-video delay from the processed stream.

The most challenging part of the work was to find reliable parameters for the hash calculations. The first algorithm did not generate a uniform hash space and therefore several erroneous hash pairs were found. The final implementation was a combination of two different methods added by a control loop, which optimized the original algorithms. The implementation was validated using video streams with known audio-video synchronization artefacts.

The results of the article were not used in the benchmarking, but the experience of the work was useful when the audio-visual synchronization measurements and metrics were developed and used in the benchmarking.

6.2 Article II: Mobile phone camera benchmarking – Combination of camera speed and image quality

The article was published at the Image Quality and System Performance of Electronic Imaging conference in San Francisco 2014 and it is part of the proceedings of the conference. The paper includes the base work of the following papers defining the first version of the benchmarking score of mobile phone cameras.

The first task of the work was to collect, analyze, and summarize existing quality and speed metrics for mobile phone cameras. The work was based on existing standards and papers. Also new metrics were suggested to measure the speed of the camera systems. Secondly, suitable image quality metrics were selected: *MTF50P* from spatial resolution and texture sharpness measurements, *SNR* from noise, ΔE_{00} and *saturation error* from color fidelity. Also speed related metrics were selected: *capture time* for one image including camera startup time, *time for five consecutive images* and *audio-video synchronization delay*.

The work included considerations of how to combine different metrics into a user friendly, single number score. Since the measurements were based on an automated test system, there was no possibility to include a systematic perceptual inspection to the measurement process. When all selected image quality metrics were acknowledged ones and also characterized as fundamental objective measurements by CPIQ group, it was reasonable to target to a score, where each metric had an equal influence on the final score. On the other hand, to evaluate how image quality features and performance features would influence to the final score, a balance was maintained between image quality and speed metrics.

Since the metrics have different scales and some of the metrics defines the superiority of a features whereas others the severity of artefacts, some equations were needed. After evaluation, a geometric mean was used to combine the different metrics. Three different benchmarking scores were calculated: *speed score*, *image quality score* and *total score*. During the evaluation of different score calculation methods, it was noticed that a single score is very easy to manipulate and it is essential to also publish the metrics from which the final scores are calculated.

The benchmarking measurements were made using an automated test system, where the testing software used an application programming interface (API) of the mobile phone cameras. The graphical user interface was not used during the tests. All quality measurements were made towards a large test scene containing several test patches: 20 gray patches for opto-electrical conversion function and noise

measurements, Macbeth color chart for color fidelity, five low contrast slanted edge charts for spatial resolution and dead leaves chart for textures sharpness. The test scene is shown in Figure 23.

The work contained test results for five unnamed mobile phone cameras selected from a larger test database. The results revealed that the variance of the speed results was significantly bigger than the variance of the image quality results. The difference skewed the total benchmark score towards the order of the speed score.

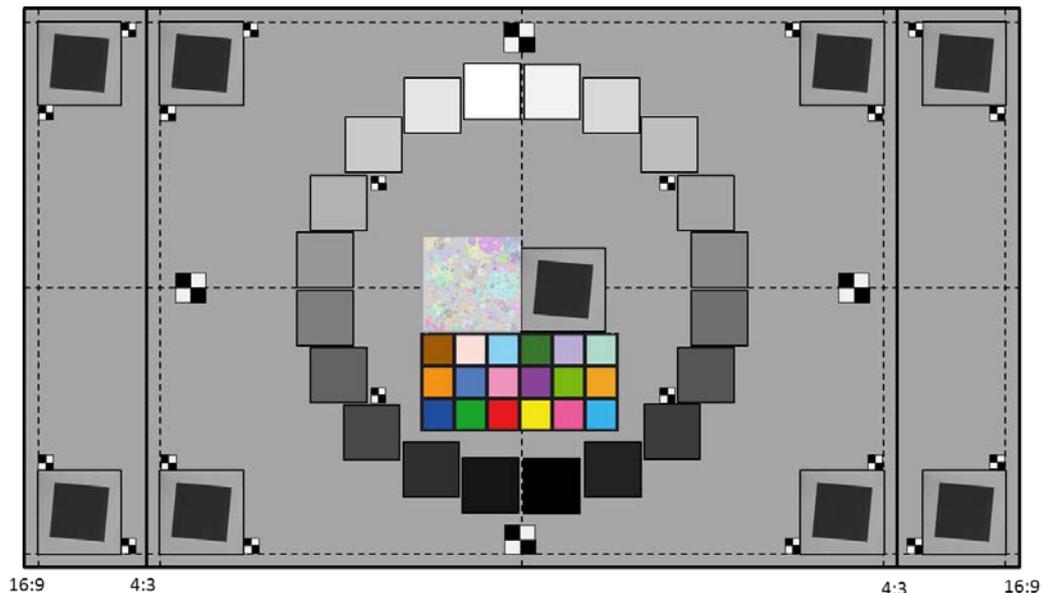


Figure 23 Test scene of benchmarking

6.3 Article III: Evaluation of mobile phone camera benchmarking using objective camera speed and image quality metrics

The article was published in Journal of Electronic Imaging 2014. The article is a continuation of the work in the previous paper and includes partially the same considerations. However, several changes were made to the previous work.

The visual noise metric was taken into use and corresponding measurements were updated. Correspondingly, the benchmarking algorithms were updated. The new algorithm did not use any weight components, but each measured metric was used as such. Moreover, the number of measured devices was increased from five to twenty five and each device was named. The manuscript contained detailed results from five devices and the other twenty were used to describe the dependency

between speed and quality features of mobile phone cameras. Finally, the very latest versions of mobile phone models were used in the measurements and the results were updated accordingly.

The work revealed an interesting fact. It seemed that mobile phone cameras with a high quality score tended to have small speed score and vice versa. The algorithm creating high quality images seemed to have some drawbacks for speed performance. Figure 24 defines the relation between quality and speed scores.

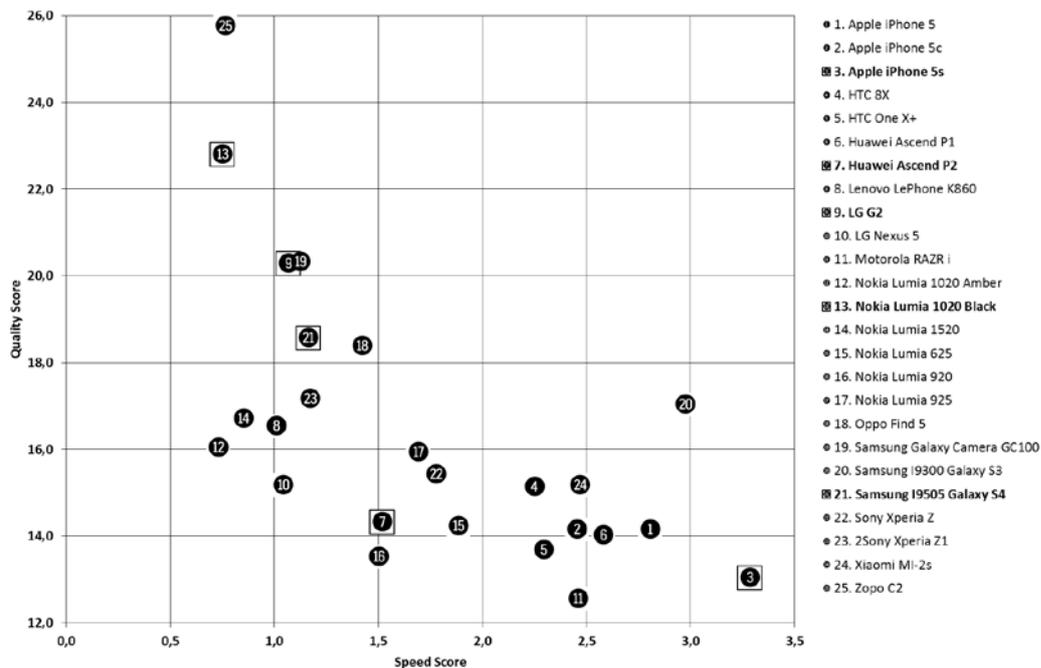


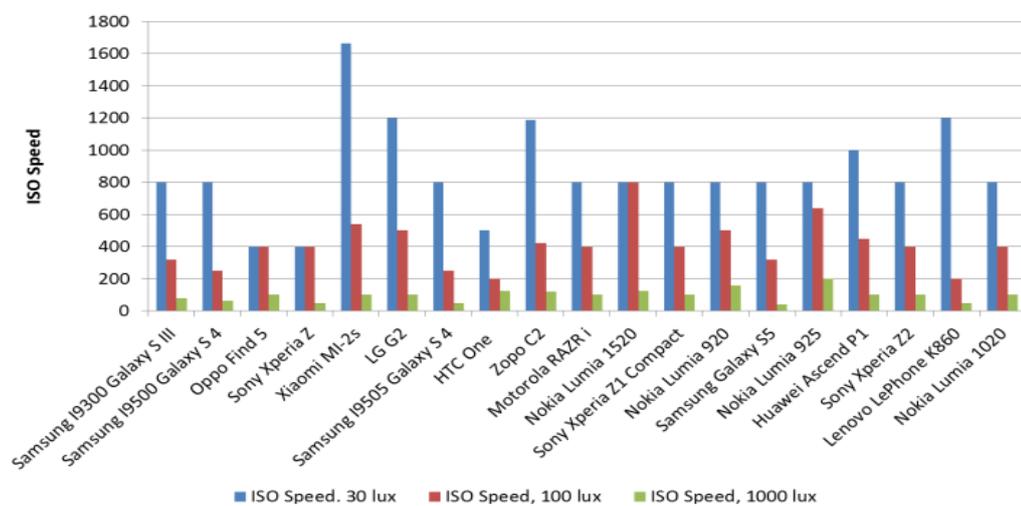
Figure 24 Measured devices in speed-quality coordinate system

6.4 Article IV: Mobile phone camera benchmarking in low light environment

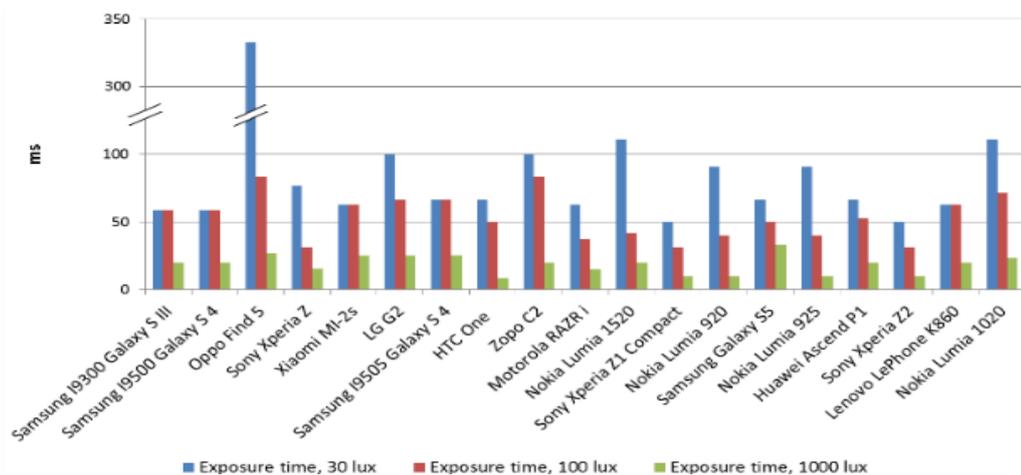
The article was published at the Image Quality and System Performance of Electronic Imaging conference in San Francisco 2015 and it is part of the proceedings of the conference. The paper includes a new approach to benchmarking of the mobile phone cameras where several illumination environments are used to validate the benchmarking system defined in previous papers. Also detailed metrics of image qualities and speed performance measurements were published for different illumination environments to investigate, how the light environment changes affect each metric.

The measurements were made in three illumination environments: 1000 lux, 100 lux, and 30 lux representing overcast day, general indoor lighting, and dim indoor lighting correspondingly. Nineteen mobile phone cameras were tested.

To inspect, how different camera systems adapt to the light environment changes, the ISO speed and exposure time were stored and published in the paper. ISO speed and exposure time are the main adjustable factors a mobile phone camera can use. Since ISO speed affects significantly noise level of images and exposure time has a strong correlation with motion blur, handshaking blur, and noise, the balancing of these two parameters is significant especially in low light imaging. The parameter values of each camera can be seen in Figure 25.



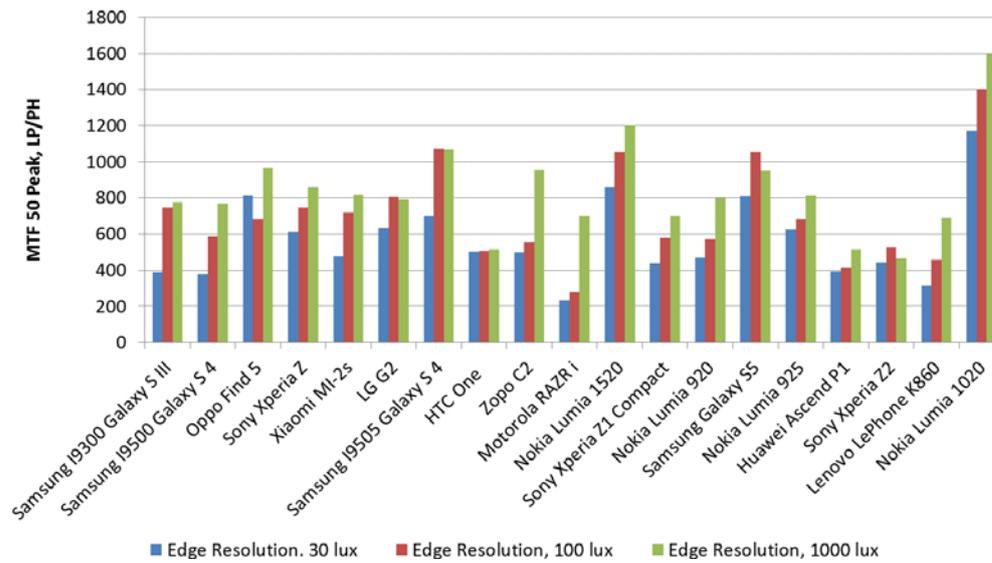
(a)



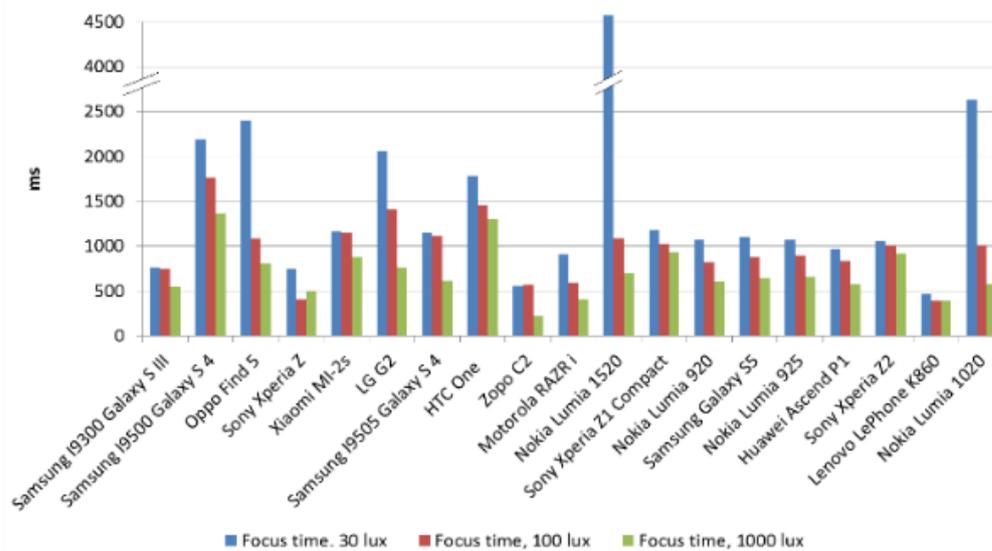
(b)

Figure 25 Mobile phone camera parametrization in different illumination environments: a) ISO speed and b) Exposure time

When each quality and speed metric was investigated in different illumination environments, it was noted that noise levels were increased in low light environment, as expected. Also the resolution metrics, spatial resolution and texture blur were decreased in low light environments. However, the low light environment did not affect color accuracy. In case of speed metrics, the low light environment increased significantly focus time, whereas other speed metrics remain quite stable. Examples of metrics are shown in Figure 26.



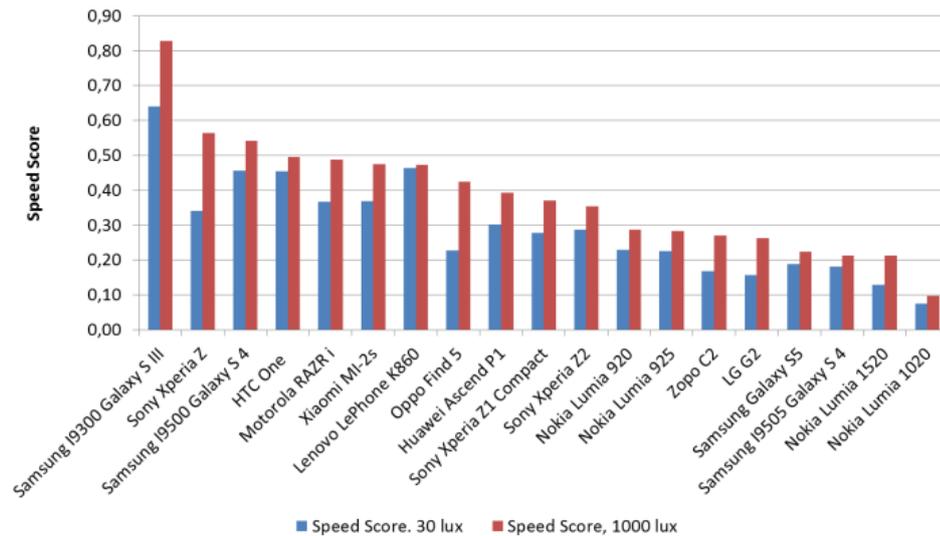
(a)



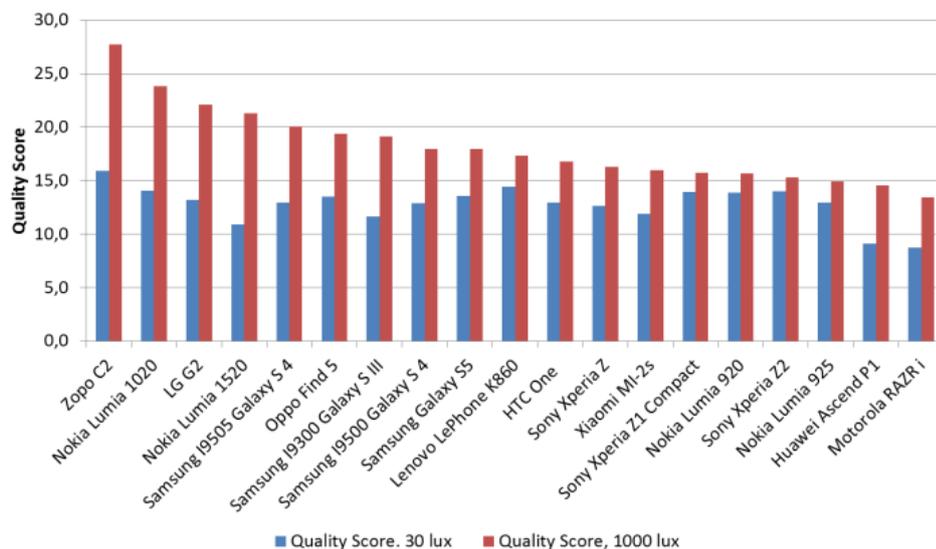
(b)

Figure 26 Quality and speed metrics in different illumination environments: a) Spatial resolution and b) Focus time

When the benchmarking scores were investigated, it was noted that the light environment does not influence significantly the speed score order of mobile phone cameras. The correlation between rank values was as high as 0.92 between 1000 lux and 30 lux illuminants. On the other hand, different illumination environments clearly influenced the quality metrics and benchmarking. The corresponding rank correlation was 0.31. The benchmark score changes between illumination environments are shown in Figure 27.



(a)



(b)

Figure 27 Benchmarking in 30 and 1000 lux illumination environments: a) Speed score and b) Quality score

The work included several conclusions. Where the illumination changes did not seem to influence the speed benchmarking, they generated significant changes in the quality benchmarking. It was also noted, that the dead leaves measurements were quite unreliable in the high noise i.e. low light environment and improvements in the measurement algorithm were required. The work concluded that creating a single value and static benchmarking score which could include different environments and which could adapt to the changes in camera systems is problematic.

6.5 Article V: SNR and visual noise of mobile phone cameras

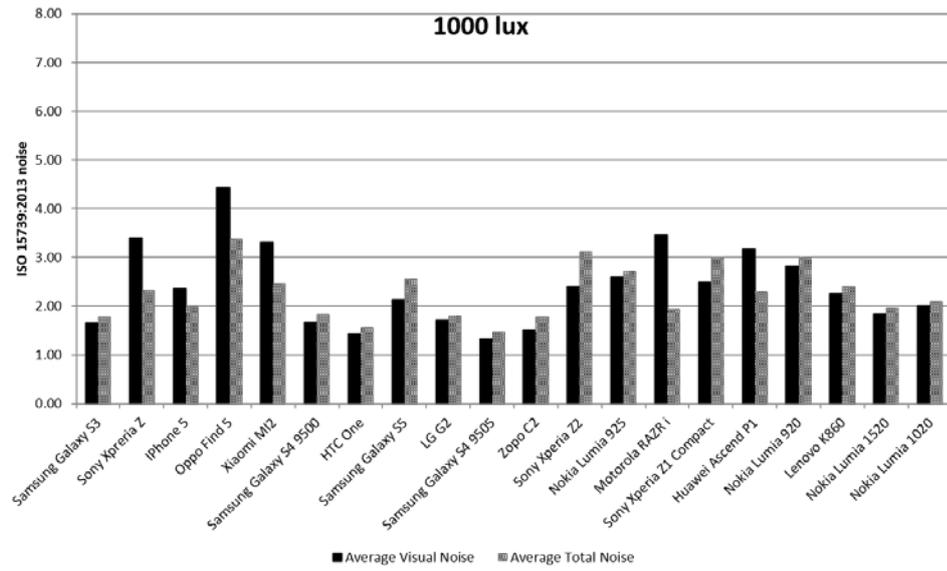
The work was initially published in International Congress of Imaging Science, in Tel Aviv 2015 and formally published in Journal of Imaging Science and Technology 2015.

The work concentrates on only one essential image quality metric, noise. The paper describes the differences between ISO 15739 standard versions and how they affect practical noise measurements. Finally, the paper defines how traditional signal to noise results and visual noise results differ in mobile phone camera measurements.

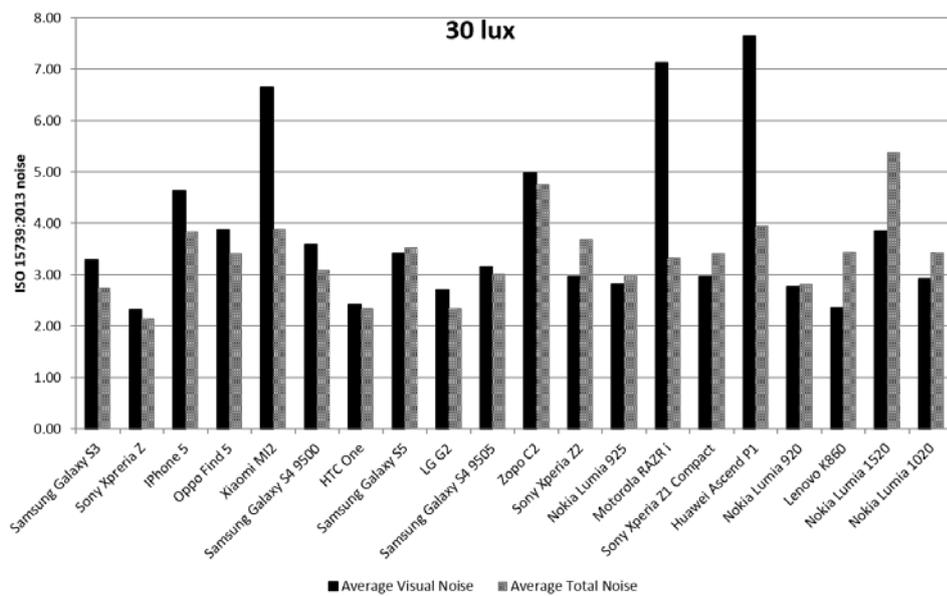
Firstly, the paper included a detailed description, of how ISO 15739 standard versions 2003 and 2013 differed and how they affected measurement in practice and end results. The paper contained descriptions of changes in the measurement environment and methods, SNR calculations, and dynamic range calculations. Finally, new visual noise measurements and calculations were explained.

Twenty mobile phone cameras were measured and the results were compared. In general, the SNR based noise and visual noise values were surprisingly equal in all light conditions. Still there were some exceptions, which were inspected more closely. It revealed that the devices, which had clearly higher visual noise values than SNR based noise values, were also perceptually noisier. This highlighted the importance of the visual noise metric. The work contained also a closer look at ISO speed values, exposure values and their relation to the noise levels.

Figure 28 shows the results of different noise results in 1000 lux and 30 lux environments.



(a)



(b)

Figure 28 SNR based noise and visual noise in different illumination environments: a) 1000 lux and b) 30 lux

7 CONCLUSIONS, DISCUSSION AND FUTURE

The objective of the thesis was to evaluate and introduce a comprehensive and novel benchmarking system which can be used to compare and rank mobile phone cameras.

Based on the five articles published in years 2012-2015 and the large introduction section, the thesis includes considerations, evaluations, trials, and a solution for a benchmarking system for mobile phone cameras. The introduced benchmarking system is the first method which does not entirely rely on image quality metrics but takes into account also performance of mobile phone cameras. It can be foreseen that the quality of mobile phone cameras will reach a level, where quality is not anymore the ultimate parameter and the performance and usability of the camera will have more significant role.

It was obvious and essential to make a detailed survey of different quality features and artefacts which should be taken into account before the benchmarking system was developed. The thesis contains a comprehensive overview of quality features, distortions, and artefacts which occur in modern digital cameras. Moreover, corresponding quality metrics are described, how they are used to evaluate the image quality and how they have evolved together with camera algorithms.

In addition to the traditional camera quality, a performance approach was used, too. The thesis includes a detailed overview of existing camera performance research and standards. The attached articles include several measurements for camera delays in different environments. The performance metrics were integrated into part of the benchmarking.

Overall, the objectives were achieved. The proposal solution for benchmarking mobile phone cameras was introduced but even more important, a generic survey of challenges and requirements of a benchmarking system was made. Certainly, the introduced benchmarking system has room for several improvements: the work towards more suitable metrics and benchmarking equations should be continued and considerations of new benchmarking environments should be started.

The first research question was formulated as: *Which requirements should a comprehensive benchmark system of mobile phone cameras fulfill?* Obviously, there are several requirements, which should be fulfilled. According to the trials and evaluations, here are the main findings.

- The system has to be a public one. Combining and averaging different metrics is always misleading and without knowledge of the original metric

values and equations, the result is very difficult to interpret. The benchmarking score has to be reproducible by third parties.

- The benchmarking result should include also detailed measurement values. Combination of a single main score and detailed results gives a comprehensive overview of the features of a camera system. A single number score can be used for straightforward comparison, whereas detailed results are used to get more information about a certain feature of a camera. A single benchmarking score alone gives a very narrow impression of a camera system.
- Several measurement environments should be considered for use, when mobile phone cameras are benchmarked. There can be significant differences in the features of cameras, when they are measured, for example, in a low light environment.
- The role of perceptual quality measurements is increasing all the time and it should be part of the benchmarking. Traditional measurements do not offer proper results, when camera algorithms adjust the final image according to the best perceptual impression and do not replicate exactly the original scene.

The second research question was probably the most challenging and the answer will vary between interested parties such as end users, phone operators, and mobile phone vendors. The answer given today will also vary with an answer given in the future. The second question was: *Which metrics should be included in a benchmarking system?*

- No doubt, the fundamental image quality metrics should be included in the benchmarking, color fidelity including exposure accuracy, spatial resolution and noise level have to be investigated.
- Capturing the moment requires a quick camera response. Performance metrics are a valuable addition to the quality dominated measurement. The selection of the best performance metrics requires more comprehensive use case studies.
- There is a demand for inclusion of video specific metrics, too. Compression based artefacts in particular are highlighted in the literature. For example temporal resolution metrics would tackle the most severe compression artefacts.
- A single, dominant artefact may ruin the image quality. To some extent the artefacts are measured by fundamental measurements but signs of artefacts can be difficult to isolate from generic measurement results. Usual artefacts which may not affect other measurements are vignetting, and to some extent color shading, geometric distortion, blooming and banding. It should be

considered, whether geometric distortion, vignetting, and color shading should be added to the benchmarking.

The problematic issue of different imaging environments were discussed in the third question: *How should different environmental factors be taken into account in a benchmarking system?* An imaging environment undoubtedly affects the quality of the final image and the result varies between devices and models. In the case of lighting, it seems that the quality features vary more than performance ones. This work included only measurements of different illumination environments. Probably different distances between the captured scene and movement of the object would reveal more interesting results and variations between camera devices and models. However, the combination of different environments is challenging. Including all measurements in one benchmarking score would be misleading.

Finally, the fourth research question was described as: *How would the evolution of digital cameras, algorithms and testing methods affect the benchmarking system?* As seen in the evolution of the resolution measurements, new image processing algorithms generate continuous changes and additions to the measurement methods and therefore to benchmarking, too. Even where a static benchmarking system ensures compatibility between new and old camera models, without continuous evaluation and improvements, a benchmarking system will expire and the usefulness of the score will decrease.

All in all, this work and the answers show that a benchmarking system is always a trade-off between numerous demands. Moreover, the development of the camera industry forces the measurements and benchmarking systems to evolve continuously.

7.1 Future

As defined in section 2.3, the future of digital imaging is very exciting. A significant number of new inventions and methods are being developed to create new features for digital cameras and to improve the quality of images and videos. New inventions require new ways of testing, new standards and new benchmarking methods. This research can be used as the first public benchmarking method for mobile phone camera, but it requires continuous development towards new camera features.

Moreover, image quality has been studied and developed more than video related quality. The huge use of video related features and applications forces to reduce this disparity. Video is also being used in new ways: presence capturing is one of

the features which may revolutionize totally the way recorded video and live stream material is consumed. Related to presence capture quality measurements, the author has published a new paper in Photonics Europe 2016 conference. The topic of the paper was ‘Presence capture cameras – a new challenge to the image quality’. Together with new image quality metrics, the presence capture camera systems will require also methods to compare and benchmark these cameras.

Consumer cameras are only a small part of the growth in digital imaging. Digital cameras are used more and more in automotive, industry, security, military, and medical applications. Each specific area of digital imaging requires its own features and thus quality measurement methods. It can be foreseen, that there will be more industry specific image quality requirements and standards.

This research investigates mainly objective quality and performance metrics of mobile phone cameras. A clear continuation for this work is to use perceptual metrics or convert objective metrics to perceptual ones. In case of image quality area and perceptual benchmarking, it will be very interesting to follow the success of the CPIQ standard, which will be released during year 2016.

In case of performance metrics, the perceptual approach is not properly studied. There are no research which are validating how different delays of a camera system will affect to the usability. Especially new performance features like negative shutter lag have certainly some acceptance limits which should be investigated.

It should be highlighted that the novel benchmarking metric is not the only outcome of this research. The findings, challenges and requirements found during this study can help following research to avoid same difficulties and ease to develop better systems.

All in all, the traditional image quality features of mobile phone cameras will reach at some point a maturity level where cameras have a good quality in general. This evolution gives mobile phone manufacturers opportunity to build more sophisticated camera features and eventually forces to improve better metrics and benchmarking systems.

REFERENCES

- Adimec Blooming [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://info.adimec.com/blogposts/ccd-versus-cmos-blooming-and-smear-performance>.
- Adimec noise [Homepage]. [Cited 23rd Oct. 2015]. Available at: <http://info.adimec.com/blogposts/read-noise-versus-shot-noise-%E2%80%93-what-is-the-difference-and-when-does-it-matter>.
- Agranov, G., Berezin, V. and Tsai, R. H. (2003). Crosstalk and microlens study in a color CMOS image sensor. *IEEE Transactions on Electron Devices*, 50:1, 4-11. DOI: 10.1109/TED.2002.806473.
- Altek [Homepage]. [Cited 15th May 2016]. Available at: https://3gltesummit.qualcomm.com/sites/default/files/pdf/3GLTE2015_Altek-JLin.pdf
- ANSI/CEA-639 (2010). Consumer Camcorder or Video Camera Low Light.
- Artmann, U and Wueller, D. (2009) Differences of digital camera resolution metrology to describe noise reduction artifacts. In *Image Quality and System Performance VII*. San Francisco, USA: SPIE. 7529. DOI: 10.1117/12.838743.
- Artmann, U. (2015) Image quality assessment using the dead leaves target: experience with the latest approach and further investigations. In *Digital Photography XI*. San Francisco, USA: SPIE. 9404. DOI: 10.1117/12.2079609.
- Artmann, U. and Wueller, D. (2012). Improving texture loss measurement: spatial frequency response based on a colored target. In *Image Quality and System Performance IX*. San Francisco, USA: SPIE. 8293, DOI: 10.1117/12.907303.
- Baker, S., Bennett, E., Sing Bing Kang, and Szeliski, R. (2010). Removing rolling shutter wobble. In *IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, USA: IEEE. 2392-2399. DOI: 10.1109/CVPR.2010.5539932.
- Bhagavathy, S., Llach J., and fu Zhai, J. (2007). Multi-scale probabilistic dithering for suppressing banding artifacts in digital images. In *IEEE International Conference on Image Processing*. San Antonio, USA: IEEE. 397-400. DOI: 10.1109/ICIP.2007.4380038, 397-400.
- Bucher, F-X., Cao, F., Viard, C. and Guichard, F. (2014). Electronic trigger for capacitive standard time lag measurements for smartphones. In *Digital Photography X*. San Francisco, USA:SPIE 9023. DOI: 10.1117/12.2042162.

Business Insider [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://uk.businessinsider.com/iphone-6-plus-rear-camera-ois-bug-2014-11?r=US&IR=T>.

Business Wire [Homepage]. [Cited 15th May 2016]. Available at: <http://www.businesswire.com/news/home/20130717005332/en/Aptina-Leaps-13-Megapixel-Smartphone-Image-Sensor>.

Cambridge in colour [Homepage]. [Cited 26th Oct. 2015]. Available at: <http://www.cambridgeincolour.com/tutorials/cameras-vs-human-eye.htm>.

Cao, F., Guichard, F. and Hornung, H. (2009). Measuring texture sharpness of a digital camera. In *Digital Photography V*. San Francisco, USA: SPIE. 7250. DOI: 10.1117/12.805853.

Caponigro [Homepage]. [Cited 27th Oct. 2015]. Available at: <http://www.johnpaulcaponigro.com/blog/7747/7-sharpening-artifacts-to-avoid/>.

CIPA DC-003 (2003). Resolution Measurement Methods for Digital Cameras.

CIPA DC-004 (2004). Sensitivity of digital cameras.

CIPA DCG-001-Translation (2005). Guideline for Noting Digital Camera Specifications in Catalogs.

CIPA DCG-002 (2012) Specification Guideline for Digital Camera.

CIPA DC-X011 (2012). Measurement and description method for image stabilization performance of digital cameras (Optical method)

CMOSIS (2012). Application note for CMV4000 v2 and CMV4000 v3.

CMOSIS (2012). CMV4000 v3 Datasheet.

CMOSIS Global Shutter [Homepage]. [Cited 15th May 2016]. Available at: http://www.cmosis.com/technology/technology_overview/global_shutter_cmos_image_sensor_pixels_with_excellent_shutter_efficiency.

CPIQ (2016). Draft Standard for Camera Phone Image Quality.

CPIQ Phase 1 (2007). CPIQ Initiative Phase 1 White Paper – Fundamentals and review of considered test methods.

CPIQ Phase 2 (2009). CPIQ Phase 2 - Acutance – Spatial Frequency Response.

CPIQ Phase 2 (2009). CPIQ Phase 2 – Color Uniformity.

CPIQ Phase 2 (2009). CPIQ Phase 2 – Lens Geometric distortion (LGD).

CPIQ Phase 2 (2009). Introduction.

CPIQ Phase 2 (2009). CPIQ Phase 2 - Initial Work on Texture Metrics.

da Silva, R. D., Minetto, R., Schwartz, W. R. and Pedrini, H. (2013). Adaptive edge-preserving image denoising using wavelet transforms. *Pattern Analysis and Applications*. 16:4, 567-580.

Dabov, K., Foi, A., Katkovnik, V. and Egiazarian K. (2006). Image denoising with block-matching and 3D filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*. San Jose, USA: SPIE. 6064. DOI: doi:10.1117/12.643267.

Danes Picta [Homepage]. [Cited 24th Oct. 2015]. Available at: <http://www.danes-picta.com/>.

DigitalOptics [Homepage]. [Cited 15th May 2016]. Available at: <http://www.doc.com/Company/Documents/Smartphone-Camera-Advances-Made-Possible-by-mems-cam-Technologies-February-20131.pdf>

Digital Trends [Homepage]. [Cited 4th Nov. 2015]. Available at: <http://www.digitaltrends.com/mobile/camera-phone-history/>.

DxO Color sensitivity [Homepage]. [Cited 28th Oct. 2015]. Available at: <http://www.dxomark.com/About/In-depth-measurements/Measurements/Color-sensitivity>.

DxO Color shading [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://www.dxo.com/us/more-information-about-color-shading>.

DxO Mark [Homepage]. [Cited 26th Oct. 2015]. Available at: <http://www.dxomark.com/>.

DxO Perceptual MPix [Homepage]. [Cited 3rd Nov. 2015]. Available at: <http://www.dxomark.com/Reviews/Looking-for-new-photo-gear-DxOMark-s-Perceptual-Megapixel-can-help-you>.

DxO Sharpness [Homepage]. [Cited 22nd Oct. 2015]. Available at: <http://www.dxomark.com/About/In-depth-easurements/Measurements/Sharpness>.

DxOMark Mobile [Homepage]. [Cited 6th Nov. 2015]. Available at: <http://www.dxomark.com/Mobiles>.

EBU R37 (2007). The relative timing of the sound and vision components of a television signal.

EBU Tech. 3249–E (1995). Measurement and analysis of the performance of film and television camera lenses.

EEtimes [Homepage]. [Cited 4th Nov. 2015]. Available at: http://www.eetimes.com/document.asp?doc_id=1328183.

Euro NCAP [Homepage]. [Cited 9th Nov. 2015]. Available at: <http://www.euroncap.com/en>.

Fenimore, C. P. and Nikolaev A. I. (2003). Assessment of resolution and dynamic range for digital cinema. In *Proceedings of Image and Video Communications and Processing 2003*. San Francisco, USA: SPIE. 5022. DOI: 10.1117/12.484325.

Fleming P. J. and Wallace J. J., (1986). How not to lie with statistics: the correct way to summarize benchmarking results. *Communications of the ACM*. 29:3, 218-221.

Foveon [Homepage]. [Cited 15th May 2016]. Available at: <http://www.foveon.com/article.php?a=67>.

GlobeNewswire [Homepage]. [Cited 15th May 2016]. Available at: <http://globenewswire.com/news-release/2015/10/26/780079/0/en/Parc-Scientists-Develop-Tiny-Low-Cost-Hyperspectral-Imaging-Cameras.html>.

Guarnera, M., Messina, G. and Tomaselli, V. (2010). Adaptive Color Demosaicing and False Colors Removal. *Journal of Electronic Imaging*. 19:2, DOI: 10.1117/1.3432486.

Habekost, M. (2013). Which color differencing equation should be used? *International Circular of Graphic Education and Research* 6, 20-33.

Hecht, E. (2002). *Optics*. San Francisco, US: Addison Wesley. ISBN 0-321-18878-0, 253-273.

Heptagon [Homepage]. [Cited 15th May 2016]. Available at: <http://hptg.com/technology/time-of-flight/>.

Hirakawa, K. and Parks, T. W. (2005) Adaptive homogeneity-directed demosaicing algorithm. In *IEEE Transactions on Image Processing* 14:3, 360 – 369. DOI: 10.1109/TIP.2004.838691.

Hoefflinger, B. (2007). *High-Dynamic-Range (HDR) Vision*. Berlin, Germany: Springer. ISBN: 978-3-540-44432-9.

Hsu, T. H., Fang, Y. K., Yaung, D. N., Wu, S. G., Chien, H. C., Wang, C. S., Lin, J. S., Tseng, C. H., Chen, S. F., Lin, C. S. and Lin, C.Y. Color mixing improvement of CMOS image sensor with air-gap-guard ring in deep-submicrometer CMOS technology. In *IEEE Electron Device Letters* 26:5, 301-303. DOI: 10.1109/LED.2005.846574.

Hunter, R. S. (1958). Photoelectric Color-Difference Meter. In *Journal of the Optical Society of America*, 48:12, 985.

IEC 61146-2 (1997). Video cameras (PAL/SECAM/NTSC) - Methods of measurement - Part 2: Two- and three-sensor professional cameras.

Image Sensors World AF [Homepage]. [Cited 23rd Oct. 2015]. Available at: <http://image-sensors-world.blogspot.fi/2015/09/st-tof-assisted-af-adopted-in-many.html>

Imaging resource [Homepage]. [Cited 4th Nov. 2015]. Available at: <http://www.imaging-resource.com/news/2014/06/25/happy-20th-birthday-apple-quicktake-100-the-first-consumer-digital-camera>

Imatest Chromatic [Homepage]. [Cited 14th Oct. 2015]. Available at: http://www.imatest.com/docs/sfr_chromatic/.

Imatest Image quality factors [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://www.imatest.com/docs/iqfactors/>.

Imatest Moiré [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://www.imatest.com/docs/iqfactors/#moire>.

Imatest Sharpening [Homepage]. [Cited 30th Oct. 2015]. Available at: <http://www.imatest.com/docs/sharpening/>.

Imatest Sharpness [Homepage]. [Cited 22nd Oct. 2015]. Available at: <http://www.imatest.com/docs/sharpness/>.

Imatest Slanted-Edge versus Siemens Star [Homepage]. [Cited 30th Oct. 2015]. Available at: http://www.imatest.com/docs/slant_edge_star_comparison/.

Imatest Tilt [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://www.imatest.com/2012/04/measuring-the-effects-of-tilt/>.

IMEC [Homepage]. [Cited 15th May 2016]. Available at: http://www2.imec.be/be_en/research/image-sensors-and-vision-systems/hyperspectral-imaging.html.

Invisage [Homepage]. [Cited 15th May 2016]. Available at: <http://www.invisage.com/quantumcinema/>.

ISO 12232 (2006). Photography - Digital still cameras - Determination of exposure index, ISO speed ratings, standard output sensitivity, and recommended exposure index.

ISO 12233 (2000). Photography - Electronic still-picture cameras - Resolution measurements.

ISO 12233 (2014). Photography – Electronic still picture imaging – Resolution and spatial frequency responses.

ISO 12233 (2014). Photography - Electronic still-picture cameras - Resolution measurements.

ISO 13406-2 (2001). Ergonomic requirements for work with visual displays based on flat panels -- Part 2: Ergonomic requirements for flat panel displays.

ISO 14524 (2009). Photography -- Electronic still-picture cameras -- Methods for measuring opto-electronic conversion functions (OECFs).

ISO 15739 (2003). Photography - Electronic still-picture imaging - Noise measurements.

ISO 15739 (2013). Photography -- Electronic still-picture imaging -- Noise measurements.

ISO 15781 (2013). Photography - Digital still cameras - Measuring shooting time lag, shutter release time lag, shooting rate, and start-up time.

ISO 17321-1 (2012). Graphic technology and photography - Colour characterisation of digital still cameras (DSCs) - Part 1: Stimuli, metrology and test procedures.

ISO 17957 (2015). Photography - Digital cameras - Shading measurements.

ISO 15780 (2015). Photography - Digital cameras - Geometric distortion (GD) measurements

ISO 19084 (2015). Photography - Digital cameras - Chromatic displacement measurements

ISO 20462-3 (2012). Photography - Psychophysical experimental methods for estimating image quality - Part 3: Quality ruler method.

ISO 9039 (2008). Optics and photonics -- Quality evaluation of optical systems -- Determination of distortion.

ISO 9358 (1994). Optics and optical instruments -- Veiling glare of image forming systems -- Definitions and methods of measurement.

ISO/CIE 11664-6 (2014). Colorimetry – Part 6: CIEDE2000 Colour-Difference Formula.

ITU-R BT.500-11 (2002). Methodology for the subjective assessment of the quality of television pictures.

ITU-T P.800 (1996). Methods for subjective determination of transmission quality.

ITU-T P.910 (2008). Subjective video quality assessment methods for multimedia applications.

ITU-T P.913 (2014). Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment.

ITU-T Rec. P.910 (2008). Subjective video quality assessment methods for multimedia applications.

ITU-T T.81 (1992). Information technology - digital compression and coding - of continuous-tone still images - requirements and guidelines.

Johnson, G. M. and Fairchild, M. D. (2003). A top down description of S-CIELAB and CIEDE2000. In *Color Research & Application*. 28:6, 425–435. DOI: 10.1002/col.10195.

Keelan, B. W. (2002). *Handbook of Image Quality*. Boca Raton, US: CRC Press. ISBN 978-0-8247-0770-5.

Keelan, B. W., Jenkin, R. B. And Jin, E.W. (2012). Quality versus color saturation and noise. In *Proceedings of Digital Photography VIII*. San Francisco, USA: SPIE 8299. DOI: 10.1117/12.905377.

Kingslake, R. (1992). *Optics in Photography*. Bellingham, US: SPIE Press. ISBN 0-8194-0763-1.

Kirk, L., Herzer, P., Artmann, U. and Kunz, D., (2014). Description of texture loss using the dead leaves target: Current issues and a new intrinsic approach. In *Proceedings of Digital Photography X*. San Francisco, USA: SPIE 9023. DOI: 10.1117/12.2039689.

Kleinmann, J. and Wueller, D. (2007). Investigation of two Methods to quantify Noise in digital Images based. In *Proceedings of Image Quality and System Performance IV*. San Francisco, USA: SPIE 6494. DOI: 10.1117/12.701899.

LensVector [Homepage]. [Cited 15th May 2016]. Available at: <http://lensvector.com/technology/how/>.

Lepistö, L., Nikkanen, J and Suksi, M. (2009). Blemish detection in camera production testing using fast difference filtering. *Journal of Electronic Imaging*. 18:2. DOI: 10.1117/1.3132004.

Light [Homepage]. [Cited 15th May 2016]. Available at: <https://light.co/camera>.

Lilja, D. J., (2005). *Measuring Computer Performance*. Cambridge, UK: University Press. ISBN: 9780511036279.

Lukac, R. (2013). *Perceptual Digital Imaging, Methods and Applications*. Boca Raton, US: CRC press. ISBN 978-1-4398-6856-0.

Lytro [Homepage]. [Cited 15th May 2016]. Available at: <https://www.lytro.com/imaging>.

Mantiuk, R. and Seidel, H-P. (2008). Modeling a Generic Tone-mapping Operator. In *Journal compilation of Eurographics Association 27:2*.

Masson, L., Cao, F., Viard, C. and Guichard, F. (2014). Device and algorithms for camera timing evaluation. In *Image Quality and System Performance XI*. San Francisco, USA: SPIE 9016. DOI: 10.1117/12.2042161.

Menon, D., Andriani, S. and Calvagno, G. (2006). A novel technique for reducing demosaicing artifacts. In *14th European Signal Processing Conference*. Florence, Italy: IEEE.

Microsoft [Homepage]. [Cited 4th Nov. 2015]. Available at: <https://blogs.windows.com/devices/2013/09/20/the-magic-camera-ingredients-inside-the-nokia-lumia-1020/>.

Mokrzycki, W. S. and Tatol, M. (2012). Color difference Delta E – A survey. In: *Machine graphics and vision* 20:4, 383-411.

Nixon, R.H., Kemeny, S. E., Staller, C. O. and Fossum, E. R. (1995). 128x128 CMOS Photodiode-Type Active Pixel Sensor with On-Chip Timing, Control and Signal Chain Electronics. In *Charge-Coupled Devices and Solid State Optical Sensors V*. San Jose, USA: SPIE 2415. DOI: 10.1117/12.206529.

Nobel Prize [Homepage]. [Cited 4th Nov. 2015]. Available at: http://www.nobelprize.org/nobel_prizes/physics/laureates/2009/smith-facts.html.

Nokia (2013). A white paper - Pushing the boundaries of digital imaging.

Nuutinen, M., Valkonen, V., Oittinen, P. and Virtanen, T. (2013). Automatic Exposure and White Balance Control in Video Cameras: Time Course Characterization and Preference. In *International Symposium on Image and Signal Processing and Analysis*. Trieste, Italy: IEEE, 25-29. DOI: 10.1109/ISPA.2013.6703709.

P1858 (2015). IEEE P1858 CPIQ Overview.

Panasonic [Homepage]. [Cited 15th May 2015]. Available at: <http://news.panasonic.com/global/press/data/2016/02/en160203-5/en160203-5.html>

Patent US 20130153748 A1 Solid-state image sensor and electronic apparatus.

Pelican Imaging [Homepage]. [Cited 15th May 2015]. Available at: <http://www.pelicanimaging.com/technology/mobile.html>.

Peltoketo, V-T. (2014). Evaluation of mobile phone camera benchmarking using objective camera speed and image quality metrics. In *Journal of Electronic Imaging*. 23:6. DOI: 10.1117/1.JEI.23.6.061102.

Peltoketo, V-T. (2015). Mobile phone camera benchmarking in low light environment. In *Image Quality and System Performance XII*. San Francisco, USA: SPIE 9396. DOI: 10.1117/12.2075630.

Peltoketo, V-T. (2015). SNR and Visual Noise of Mobile Phone Cameras. In *Journal of Imaging Science and Technology*. 59:1. DOI: 10.2352/J.ImagingSci.Technol.2015.59.1.010401.

Peres, M. R. (2007) *Focal Encyclopedia of Photography*. Burlington, USA: Focal Press. ISBN: 978-0-240-80740-9.

PetaPixel [Homepage]. [Cited 4th Nov. 2015]. Available at: <http://petapixel.com/2010/11/04/first-digital-photograph-ever-made/>.

Rambus [Homepage]. [Cited 15th May 2016]. Available at: <https://www.rambus.com/emerging-solutions/lensless-smart-sensors/>

Recon Analytics (2011). International comparisons: the handset replacement cycle.

Säämänen, T., Virtanen, T. and Nyman, G. (2010). Videospace: classification of video through shooting context information. In *Image Quality and System Performance*. San Francisco, USA: SPIE 7529. DOI: 10.1117/12.839414.

Samsung ISOCELL [Homepage]. [Cited 12th Nov. 2015]. Available at: <http://global.samsungtomorrow.com/get-the-big-picture-cmos-image-sensors-and-isocell/>.

Samsung Tomorrow [Homepage]. [Cited 23rd Oct. 2015]. Available at: <http://global.samsungtomorrow.com/samsung-announces-mass-production-of-industrys-first-mobile-image-sensor-with-1-0%CE%BCm-pixels/>.

Samsung [Homepage]. [Cited 4th Nov. 2015]. Available at: <http://www.samsung.com/us/news/531>.

Sharp [Homepage]. [Cited 4th Nov. 2015]. Available at: http://www.sharp-world.com/corporate/info/his/only_one/item/t34.html.

Skype (2013). Skype Hardware Certification Specification Video requirements for Computer Accessories and Computers, Version 6.3.2_ACC.

Smith, J. E. (1988). Characterizing computer performance with a single number. In *Communications of the ACM*. 31:10, 1202-1206.

Sony [Homepage]. [Cited 15th May 2016]. Available at: http://www.sony.net/Products/SC-HP/IS/sensor1/img/products/ProductBrief_IMX278_20150715.pdf.

Sony IMX174LLJ [Homepage]. [Cited 15th May 2016]. Available at: http://www.sony.net/Products/SC-HP/new_pro/december_2013/imx174_e.html.

Sony Optics [Homepage]. [Cited 15th May 2016]. Available at: https://www.sony.com/en_us/SCA/company-news/press-releases/sony-electronics/2015/sony-introduces-new-palm-sized-rx1r-ii-camera-with.html.

Stapenhurst, T. (2009). *The Benchmarking Book: A How-to-Guide to Best Practice for Managers and Practitioners*. Oxford, UK: Elsevier Ltd. ISBN-978-0-7506-8905-2.

Sun, Y. and Liu, G. (2012). Rolling shutter distortion removal based on curve interpolation. In *IEEE Transactions on Consumer Electronics*. 58:3, 1045-1050. DOI: 10.1109/TCE.2012.6311354.

Theuwissen Blooming [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://harvestimaging.com/blog/?p=1471>.

Theuwissen PRNU [Homepage]. [Cited 14th Oct. 2015]. Available at: <http://harvestimaging.com/blog/?p=916>.

Tian, H. (2000). *Noise analysis in CMOS image sensors*. A PHD dissertation submitted to the department of applied physics and the committee on graduate studies of Stanford University.

Tom, A. (2014) *Photography: The Definitive Visual History*. London, UK: Dorling Kindersley Limited. ISBN: 9781465422880.

Toshiba PDAF [Homepage]. [Cited 23rd Oct. 2015]. Available at: <http://toshiba.semicon-storage.com/ap-en/product/sensor/cmos-sensor/tech-pdaf.html>.

Umbaugh, S. E. (2005). *Computer Imaging, Digital Image Analysis and Processing*. Boca Raton, US: CRC press. ISBN 0-8493-2919-1.

University of Texas [Homepage]. [Cited 4th Nov. 2015]. Available at: <http://www.hrc.utexas.edu/exhibitions/permanent/firstphotograph/>

VQEG (2008). Multimedia Group Test Plan Version 1.21. Available at: <http://www.its.bldrdoc.gov/vqeg/projects/multimedia-phase-i/multimedia-phase-i.aspx>.

Walker, B. H. (1998). *Optical Engineering Fundamentals*. Bellingham, US: SPIE Press. ISBN 0-8194-2764-0.

Wang, X. (2008). *Noise in Sub-Micron CMOS Image Sensors*. Enschede, Netherlands. ISBN: 9789081331647

Wang, Z. and Bovik, A. C. (2006). *Modern Image Quality Assessment*. Lexington, US: Morgan & Claypool. ISBN 1-59829-022-3.

Welbourne, L. E., Morland, A. B. and Wade, A. (2015). Human colour perception changes between seasons. *Current Biology* 25:15, 646-647. DOI: 10.1016/j.cub.2015.06.030.

Winkler, S. (2005). *Digital Video Quality: Vision Models and Metrics*. Chichester, UK: Wiley. ISBN 0-470-02404-6.

Wu, H. R. and Rao, K. R. (2006). *Digital Video Image Quality and Perceptual Coding*. Boca Raton, US: CRC Press. ISBN: ISBN 0-8247-2777-0.

Wueller, D. (2013). Low light performance of digital still cameras. In *Proceedings of Multimedia Content and Mobile Devices*. San Francisco, USA: SPIE 8667. DOI: 10.1117/12.2003080.

Wyszecki, G. and Stiles, W. S. (2000). *Color Science – Concepts and Methods, Quantitative Data and Formulae*. Danvers, US: Wiley. ISBN: 978-0-471-39918-6

Yole (2015). Status of the CMOS image sensor industry.

Youtube [Homepage]. [Cited 4th Dec. 2015]. Available at: <https://www.youtube.com/watch?v=gWHajkbAsYY>

Zhang, X. and Wandell, B. A. (1997). A spatial extension of CIELAB for digital color-image reproduction. *Journal of the Society for Information Display* 5:1, 61–63. DOI: 10.1889/1.1985127.

Zhou, W. and Bovik, A.C. (2009). Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *Signal Processing Magazine* 26:1, 98-117. DOI: 10.1109/MSP.2008.930649.

Objective Verification of Audio-Video Synchronization

Veli-Tapani Peltoketo

*Sofica Ltd, Tiedekatu 2, 60320 Seinäjoki, Finland
veli-tapani.peltoketo@sofica.fi*

Abstract — The digital multimedia and especially video presentations are very widely used nowadays. Quality issues, like audio-video synchronization problems, are very common due to varying hardware and software environment. Several digital video encoding standards and different hardware and software implementations creates very complicated combinations and the verification of each combination is demanding but essential. There are several real time algorithms that detect the synchronization problems and correct them in real time. Still, the real time solutions have some weaknesses and they concentrate more to fix the possible problems than to verify and measure the quality of the whole video pipeline. There is a demand on objective and general testing system, which could offer comparable test results of audio-video synchronization between digital video systems. In this research, the most promising algorithms are compared and the most suitable algorithm combination is selected and tested. According to the results, a general method can be found even if more work and comprehensive testing is required before robust solution can be released.

Index Terms — audio codecs, video codecs, synchronization, automatic testing

I. INTRODUCTION

THE modern world is full of digitalized multimedia and the role of this presentation type is increasing all the time. There are a lot of different multimedia components, for example, in the mobile phones, computer applications, home pages and advertising. The definition of the multimedia is quite wide, it may include texture, still images, audio, video, animation, interactive features and combinations of all these. However, the role of the digital video presentation is more and more popular. This trend has highlighted the quality of the digital video broadcasting.

The quality of the video can be measured using several different parameters like frame count, noise, resolution and color accuracy. An essential quality parameter is also the audio-video synchronization which defines the timing difference of the sound and vision components in the video representation. Humans are very sensitive to detect the difference between visual representation and corresponding audio. Less than 100 ms difference can be detected especially when it is question of a lip synchronization problem, i.e., the voice is not synchronized to the lip movements.

There are several reasons why the audio-video synchronization is a quite common quality problem. The fundamental reason of the problem is that the digital video pipeline is very long and contains several different components, like recording devices, encoding, encoder buffering, multiplexing, transmission, demultiplexing, decoder

buffering decoding and presentation devices [1]. Moreover, in most of these components the audio and video data are handled separately and the adjusting between audio and visual data is very dependent on the implementation and hardware and software environment.

Onwards, there are several encoding and decoding methods, for example, MPEG2 (Moving Picture Experts Group), AVI (Audio Video Interleave) and Quicktime. Especially the newest codecs are quite heavy to execute. Also the modern audio systems with several audio channels, like Dolby surround voice with six audio channels, increase the complexity of the video handling. Moreover, there are several hardware and software techniques to implement encoding and decoding; the functionality can be done using DSPs (Digital Signal Processor), FPGA-chips (Field Programmable Gate Array), unique ASICs (Application-Specific Integrated Circuit) or by software. Especially, when the codecs are made using software, the performance is dependent on the used processor, the load of the processor and, the most importantly, the software implementation skills.

Why the performance is so crucial in case of audio-video synchronization? The amount of the visual data is much bigger than audio data and the delays which are generated to the audio and video signals are typically unequal [1]. The adjustment of the audio and video data inside the heavy encoding and decoding process is very demanding and the requirements of the adjusting operation are strict. Moreover, the current market trend forces to raise the frame count of the video which also increases the load of the codecs. The result can be noticed quite often when there is a timing difference between audio and video signals.

There are several real-time algorithms which are verifying and adjusting the audio-video synchronization during video recording and playback. Still, the real-time solutions have some weaknesses and they concentrate to fix the possible problems rather than to verify and evaluate the quality of the whole stream pipeline. There is a strong demand on objective and general testing system, which could offer comparable test results of the audio-video synchronization between digital video systems. Fortunately, different mathematical algorithms and statistical methods give powerful tools, which can be used to measure and validate this phenomenon.

This paper describes the first steps towards the general audio-video synchronization measurement algorithm. The paper concentrates mainly on evaluating and comparing the most promising methods of the audio-video synchronization detection. The final optimization and implementation as well as the comprehensive testing with large reference data will be done in the next phase.

II. CURRENT SITUATION

There are two main facts which are affecting to the evaluation of the audio-video synchronization. Firstly, the audio-video synchronization requirements are defined in several standards, which give a good base to verify and judge the quality of this feature. Furthermore, the current real time algorithms offer several solution proposals even though they are not suitable as such.

A. Standardization

The audio-video synchronization requirements are standardized very accurately because the same problems are valid both in the television broadcasting and the modern digital video broadcasting, because the acceptable delay is dependent on the end user experience, not on the used technology.

The standardized limits help significantly the evaluation of the video signal and give strict limits to the synchronization. Three different organizations have given recommendations of the audio-video synchronization as described in Table 1. The delayed audio represents situation, where the audio data of the video stream is played too late and correspondingly the advanced audio is the artifact where the audio data is played too early.

TABLE I
AUDIO-VIDEO SYNCHRONIZATION LIMITS OF DIFFERENT STANDARDS

Standard	Year	Maximum difference (milliseconds)	
		Delayed audio	Advanced audio
ITU-R BT.1359, [4]	1998	30	22.5
ATSC IS-191, [1]	2003	45	15
EBU R37, [2]	2007	60	40
Subjective evaluation by ITU-R, undetectable limits	1998	100	20

Even if the standards give certain limits of the acceptable difference between the sound and visual component in the video handling, the verification of the video quality has to be much more accurate. The evaluation of the video has to detect also such kind of problems which are acceptable, but near to the limits. Also, a jitter, a small variance of the audio-video synchronization, has to be detected.

B. Available Methods

In general, modern codecs have already functionalities, which are ensuring the audio-video synchronization. For example, the MPEG standard contains timestamp based approach, where the MPEG encoding calculates the difference between audio and video signals and stores the difference to the stream. Onwards, when the video stream is decoded, the difference data is used to adjust the audio and visual signal correctly. Still, the synchronization result depends on the

implementation of the standard and the quality of the final video stream is not always the best one.

As mentioned before, there are several real time algorithms which are handling audio-video synchronization. For example, the synchronization can be recovered using correlation analysis [7]. The study proposes a method which is based on face detection and adjusts the synchronization when human faces are recorded. Also a bimodal linear prediction model is used to handle a speech synchronization detection [6].

Furthermore, there are methods, which are detecting the audio-video synchronization using hidden watermarks of the visual data and also algorithms which are inserting the audio data to the corresponding video frames [5], [9].

There are obvious advantages when real time methods are used. The audio-video synchronization issues are not only evaluated but also fixed as well as possible. The real time methods measure the video stream and have possibility to adjust to the current situation.

However, there are also many restrictions in the real time methods or, actually, the real time methods are not suitable to the strict evaluation and verification. They do not test the video quality by giving results to the tester but try to ensure as good synchronization as possible. Furthermore, the characteristics of the real time methods bind the method implementation tightly to the used codec implementation and the outcome is not universal but suitable only to corresponding codec or even suitable to the specific video content when face detection based algorithms are used.

Maybe the most obvious restriction is that the real time methods are executed in the target environment, which is normally an embedded system like a mobile phone. The high quality evaluation of the audio-video synchronization requires heavy mathematical algorithms, a lot of time and a powerful processor. In case of an embedded real time system, such resources are unusual. One solution is to evaluate the existing video and audio data in a separate environment which have enough capacity. This solution offers also possibility to create a codec independent evaluation algorithm.

In the beginning of this study, the most promising research was done by Radhakrishnan, Bauer, Cheng and Terry [8]. Based on the signature calculations of the audio and video data, the correlations were calculated between the reference and processed data. According to the correlation and Hamming distance calculations, the quality of the audio-video synchronization can be measured [8]. This research was selected to be the first algorithm, which was implemented and tested. However, the hash results of the audio stream were not as unique as expected but clear trends were noticed which caused faulty hits during the stream comparison. Therefore, another study by Haitsma and Kalker [3] was also evaluated.

III. METHODS

The used algorithms [8] and [3] are based on the usage of two streams, the reference and processed ones. Furthermore, the [3] is intended only to audio fingerprinting. Figure 1 describes the main calculations steps of both algorithms. The figure does not contain the signature and robust hash calculation blocks of the reference data because the reference data is static and the calculations can be done beforehand. The

feedback loop from the best hit selection block to the processed hash selection is an addition to the original structure and done to both optimize and extend the current algorithms.

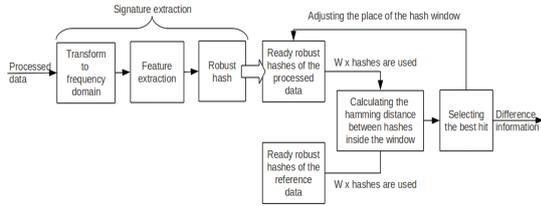


Fig. 1. Steps of the difference calculation

The calculation of the video signatures is very similar than the audio one, only the spectrogram calculation differs; in case of the video, spectrograms are not used, but the differences of the consecutive frames are used. The coarse presentation, hash calculations and hamming distance are same than in the audio measurements when method [8] is used.

The feature extraction and robust hash parts include the differences between algorithms [8] and [3]. Where the former method uses coarse representation, the latter calculates energies of certain frequency ranges. Also the hash calculation differs; [8] uses random matrices and the [3] detects the changes between consecutive frequency values. Following chapters defines the details and differences of the methods.

A. Signature Extraction

The goal of the signature extraction is to create a signature which is robust against different changes that are done to the stream, such as compression and time scale modifications [8]. Despite this kind of changes, the signature of a certain clip has to be unique and identifiable against the reference data. The same calculations are made to the reference and processed streams.

During the signature extraction, the audio data is divided to parts and one signature is calculated from each part. Moreover, the parts are overlapped so that the consecutive parts are almost equal [8], [3]. In this study, the T_0 , which defines the difference between consecutive parts, is 50 audio samples and defines the accuracy of the synchronization verification. The size of the part T_p , is 5120 which ensures the uniqueness of the signatures. Furthermore, the audio samples are transformed to the frequency domain, which is defined as S in the (1) and (2).

Equation (1) defines the coarse representation Q_a of [8], where W_f and W_t represent the time and frequency blocks of frequency-time representation S. The coarse representation averages the magnitude of frequency coefficients in time-frequency blocks. This study uses the same constant values $F=20$ and $T=10$ as the reference study.

$$Q_a(k, l) = \frac{1}{W_f * W_t} \sum_{i=(k-1)W_f}^{kW_f} \sum_{j=(l-1)W_t}^{lW_t} S(i, j) \quad (1)$$

$k = 1, 2, \dots, F; l = 1, 2, \dots, T$

Equation (2) defines the frequency energy $E(n, m)$ calculation of [3]. The frequency area of the certain audio samples W_i are divided to the frequency blocks k and sum of those frequency amplitudes are calculated. The number of the frequency blocks F is 33 which defines later the size of the hash number. However, the size of the frequency block W_i is not a constant but the frequency has a logarithmic spacing.

$$E(n, m) = \sum_{i=(k-1)W_f}^{kW_f} \sum_{j=0}^{W_t} S(i, j) \quad (2)$$

$k = 1, 2, \dots, F;$

In the first phase, when method [3] is used, the bandwidth of the audio signal is filtered so that only the frequencies between 300-2000 Hz were used. This bandwidth is the the most relevant spectral range for the human auditory system [3]. Despite this filtering, the [3] generates more unique values than [8]. However, the full range of audio frequencies has to be also evaluated and study, if the filtering causes inaccuracy to the measurements.

B. Robust Hash Extraction

The purpose of the robust hash calculation is two-fold: it filters the small changes by signal processing and it reduces the size of the signature.

The method [8] uses random matrices P_k and the size of the Q_a can be reduced to the k bits, where the k is the number of the random matrices. Before the projection H_k is calculated in (3), the matrix P_k is changed by removing the mean of the matrix from the components of the matrix.

Each bit of the robust hash is gotten by calculating H_k using corresponding P_k . If the value of the H_k is greater than the median of all projections H_k , the bit gets value '1', otherwise '0'. Finally, the robust hash is the combination of these bits, and the size of the robust hash is k bits.

$$H_k = \sum_{i=1}^F \sum_{j=1}^T Q_a(i, j) * P_k(i, j) \quad (3)$$

The principle of method [3] is different. It is based on the difference of consecutive audio samples as well as the difference between consecutive frequency blocks. If previous values are greater, the hash bit gets value '1', otherwise '0' (4). The number of frequency blocks defines the size of the hash number which is 32 bits.

$$F(n, m) = 1, \text{ if } E(n, m) - E(n, m+1) - (E(n-1, m) - E(n-1, m+1)) > 0 \quad (4)$$

$F(n, m) = 0, \text{ otherwise}$

C. Hamming Distance

The hamming distance is calculated between multiple consecutive signatures (signature block W) of reference and processed data. The main procedure contains a loop, which compares a signature block of the reference data R against the same size of signature block of the processed data A . The same reference block is compared against several blocks in the processed data using a certain window size $2xL$. The

corresponding hamming distances of each signature pairs are stored. Onwards, the next signature block is selected from the reference data and again compared against processed data inside the correspondingly shifted window. Using this loop, the whole data has been measured and the best hit of each signature block can be selected. [8]

$$D(m, i) = \sum_{j=0}^{j=W} \text{HammingDistance}(R(i+j), A(m+j)) \quad (5)$$

$$m = i-L, \dots, i+L; i = 1, 2, \dots, I$$

Figure 2 illustrates the hamming distance calculation of the (3). The signature block 1-W is compared between reference and processed data. When the whole window 2xL is measured, the signature block of the reference data and the window of the processed data are stepped forward.

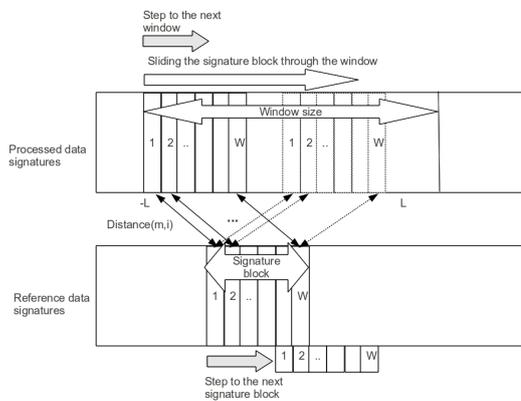


Fig. 2. Hamming distance calculations from signatures

There are three crucial parameters, which affect to the hamming distance calculation: the size of the signature block W , the size of the searching window against the processed data $2xL$ and the step size i.e. how much the signature block is moved between the comparisons. The step size affects significantly to the performance of the measurement whereas the number of the signatures affects to the reliability of the measurement and also to the filtering characteristics of the comparison - big signature block filters quick changes. The larger size of the searching window reduces the accuracy requirements of the feedback loop in the Figure 1, but also reduces the performance of the calculation.

D. Selecting the Best Hit

The obvious solution to calculate the best hit is defined in (4), where the $D(m,i)$ is the result of the Hamming distance calculation [8].

$$\text{best hit} = \arg \min_m D(m, i) \quad (6)$$

However, there are cases in which the best hit (i.e., the smallest hamming distance), is found from incorrect place and the second or third best hit is the correct one. Therefore, the

minimum distance should not be the only rule when selecting the right hit. The best hit can be validated, for example, by calculating the probability of the placement of the next hit according to the previous ones. In case of stereo and surround audio, the data of the other audio channels can be also used to filter incorrect hits. Nevertheless, this logic is not yet used in this study.

E. Adjusting Window of Processed Data

Normally, the window of the processed data is stepped on the same rate than the signature block of the reference data. However, there are situations in which this kind of logic does not work. For example, if the processed audio contains cumulative delay, sooner or later the corresponding signatures of the processed data will locate outside the measured window. This issue can be avoided by dynamically adjusting the window location depending on the latest measurements of the hamming distance.

IV. SOLUTION

The implementation of the methods is based on a layered architecture, which isolates the evaluation algorithm from the codec specific implementation as described in Figure 3. This kind of implementation is essential, when a generally usable and codec independent algorithm implementation is required.

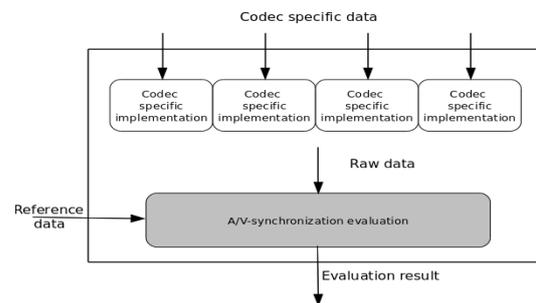


Fig. 3. Layered architecture

The role of the codec specific part is to transform the processed data of the codec to raw audio and video format. The raw audio contains the digitalized samples of the audio data and raw video contains the frame based images.

The evaluation part makes the real verification work. It receives the raw data of the tested codec and the corresponding reference data. Using the signature based algorithms, it generates the evaluation results which contain the difference values between reference and processed streams. Different statistical values, like mean difference, maximum difference and variance can be measured from these values.

There is two ways to use the reference data. It can be even raw stream or it can contain ready signatures which are calculated beforehand. In this case, the random matrices have to be same when the signatures of the processed stream are calculated, if method [8] is used.

V. THE PROGRESS OF THE STUDY AND RESULTS

The first evaluation of the signature based algorithm contains hash distribution investigation and verification of three different audio-video synchronization problems: clipped audio data with decoding delay, clipped video data and cumulative audio delay.

A. Hash Distribution Investigation

Even if the method [8] seemed very usable algorithm, it was noticed that it generated several wrong hits when reference and processed audio streams were compared. When the corresponding hash numbers were evaluated, clear trends was noticed as Figure 4 shows. If there are hash numbers which are very near to each other's or even same, the probability of wrong hits increases significantly.

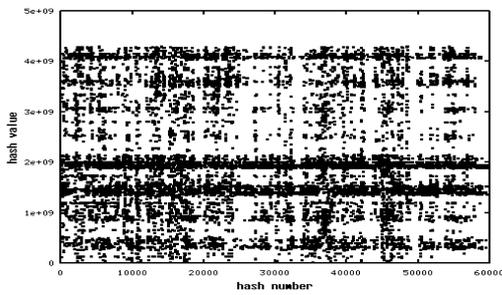


Fig. 4. Hash distribution of the method [8]

As Figure 5 defines, the hash distribution of the method [3] is significantly more uniformly distributed than previous method. (The number of hash values is equal in both figures.) This was also noticed in the best hit selection; the number of wrong hits was reduced notably.

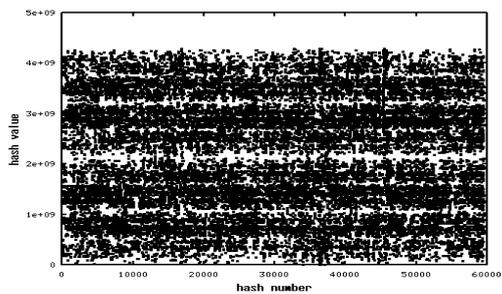


Fig. 5. Hash distribution of the method [3]

The corresponding problem was not noticed when video data was evaluated. It is noteworthy to notice that method [8] uses the difference between consecutive frames as an input data when the signatures of video part were calculated. Onwards, the method [3] uses the difference between consecutive audio frequency energies when signatures were calculated.

This study uses the combination of these methods; method

[8] is used to calculate video signatures and method [3] is used to calculate audio signatures.

B. Clipped Audio Data With Decoding Delay

Figure 6 describes a problem, when 500 ms of audio data is removed from the processed video. Moreover, during the audio removing process, the audio part of the video was converted from AAC to MP3 and back to AAC -format. Lossy codecs, like AAC and MP3 have characteristic which add delay to the beginning of the audio stream. The constant delay of MP3 decoding is 528 samples and in case of AAC decoding it is 2112 samples. However, the decoding delay of AAC may vary depending on the implementation. Due to decoding delay, the result has two opposite asynchronous artifacts; in the beginning there is decoding delay and in the middle the removed audio data can be seen. The amplitude of both delays follow the decoding and clipped audio data delays.

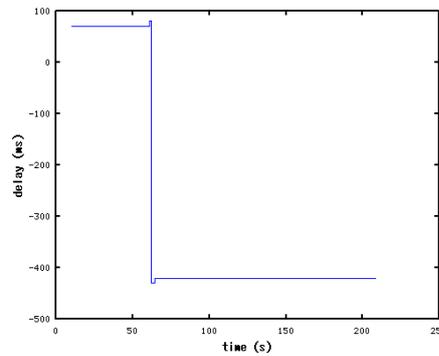


Fig. 6. Delay result of clipped audio data with decoding delay

C. Clipped Video Data

Figure 7 describes a problem, when 10 frames of video data is removed from the processed video. Ten frames represents 400 ms when fps is 25. The asynchronous can be clearly seen from the figure and it is opposite to the previous measurement. The amount of the clipped data follow the generated delay exactly.

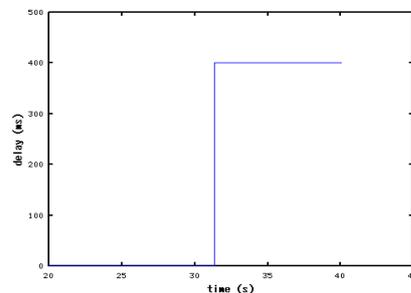


Fig. 7. Delay result of clipped video data

D. Cumulative Audio Delay

The processed audio data is cumulative and uniformly delayed 1%, meaning 2 seconds during the video. Figure 8 defines the result where the asynchronous can be noticed and the result follows the generated artifact.

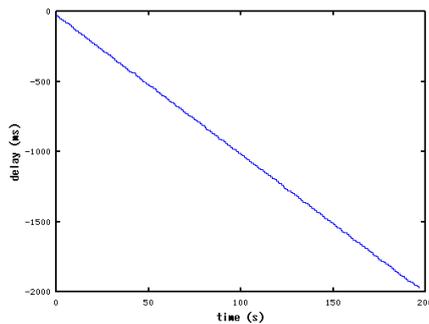


Fig. 8. Delay result of cumulative audio delay

VI. CONCLUSIONS

The combination of two signature based algorithms is a promising method for the measurement of the audio-video synchronization. The tested problems like cumulative delay and clipped audio and video data were detected correctly.

Also the result of the algorithms reveal more than audio-video synchronization evaluation. It also points out, how the timings of the audio and video parts are changed separately. This evaluation method can detect such problems that cannot be verified from the basic audio-video synchronization results.

There are quite many steps in the signature based algorithm. However, the implementation of the method is quite straightforward and can be done, for example, using C++ language. Also the performance facts encourage to use efficient language, like C++. It was clearly noticed, that the implementation is a tradeoff between performance and accuracy. It is possible to achieve a great exactness using, for example, big has numbers, but the verification time of videos will increase correspondingly.

Even if the parametrization of the methods requires more study as well as the optimization of the implementation, the current results give very positive signals that signature based evaluation can be used to measure the delays in audio and video streams. The next steps will reveal the how general implementation can be done using signature based methods.

A. Next Steps

The goal of the research is demanding; to find a general and objective method which is suitable to all video formats. This study is definitely the very first step towards the goal and much more work is required.

Even if the first evaluated methods gave good results, more methods have to be evaluated and tested. As the study has revealed so far, the combination of several methods may be the best way to implement comprehensive audio-video

synchronization detection.

Another improvement is to optimize the signature algorithm and to find the optimal parameter values to the audio and video part. Also different video formats require parameter optimization as well as the usage of several audio channels.

Finally, the implementation requires more testing. A comprehensive testing using numerous audio and video formats is needed as well as validation using known audio-video synchronization problems.

REFERENCES

- [1] ATSC - Advanced Television Systems Committee, "Relative timing of sound and vision for broadcast operations". in standard IS-191, 2003, pp. 3-4.
- [2] EBU - European Broadcasting Union, "The relative timing of the sound and vision components of a television signal". in standard R37-2007, 2007, p. 3.
- [3] J. Haitsma, T. Kalker, "A highly robust audio fingerprinting system". in proceedings of ISMIR, 2002, pp. 1-4.
- [4] ITU-R - International Telecommunication Union-Radiocommunication, "Relative timing of sound and vision for broadcasting". in standard BT.1359, 1998, pp. 2.
- [5] L. Kezheng, Y. Wei, L. Pie, "Video watermarking temporal synchronization on motion vector". in Intelligent System and Knowledge, 2008, p. 1105.
- [6] K. Kumar, J. Navratil, E. Marcheret, V. Libal, G. Ramaswamy, G. Potamianos, "Audio-Visual speech synchronization detection using a bimodal linear prediction model". in Computer Vision and Pattern Recognition Workshops, 2009, p. 54.
- [7] Y. Liu, Y. Sato, "Recovering audio-to-video synchronization by audiovisual correlation analysis". in Pattern Recognition, 2008, p. 2.
- [8] R. Radhakrishnan, K. Terry, C. Bauer, "Audio and video signatures for synchronization". in Multimedia and Expo, 2008, pp.1549-1552.
- [9] M. Yang, N. Bourbakis, Z. Chen Zishong, M. Trifas, "An efficient audio-video synchronization methodology". in Multimedia and Expo, 2007, p. 768.

Mobile phone camera benchmarking – Combination of camera speed and image quality

Veli-Tapani Peltoketo*

Sofica Ltd., Kampusranta 9C, 60320 Seinäjoki, Finland
Vaasa University, Wolffintie 34, 65200 Vaasa, Finland

ABSTRACT

When a mobile phone camera is tested and benchmarked, the significance of quality metrics is widely acknowledged. There are also existing methods to evaluate the camera speed. For example, ISO 15781 defines several measurements to evaluate various camera system delays. However, the speed or rapidity metrics of the mobile phone's camera system have not been used with the quality metrics even if the camera speed has become more and more important camera performance feature.

There are several tasks in this work. Firstly, the most important image quality metrics are collected from the standards and papers. Secondly, the speed related metrics of a mobile phone's camera system are collected from the standards and papers and also novel speed metrics are identified. Thirdly, combinations of the quality and speed metrics are validated using mobile phones in the market. The measurements are done towards application programming interface of different operating system. Finally, the results are evaluated and conclusions are made.

The result of this work gives detailed benchmarking results of mobile phone camera systems in the market. The paper defines also a proposal of combined benchmarking metrics, which includes both quality and speed parameters.

Keywords: Camera speed, camera benchmarking, image quality, camera testing, automated testing

1. INTRODUCTION

When a mobile phone camera is tested and benchmarked, the significance of quality metrics is widely acknowledged. Generally, sharpness, color reproduction and noise metrics are usually defined as the most significant quality parameters of the camera¹. Also several other quality metrics are available. For example International Imaging Industry Association (I3A), currently Institute of Electrical and Electronics Engineers (IEEE) P1858 working group, has defined the lens distortion and lateral chromatic aberration as quality components^{2, 3}. However, the speed or rapidity metrics of the mobile phone's camera system have not been used with the quality metrics even if the camera speed has become more and more important camera performance feature.

Capturing the moment is an essential requirement in modern mobile phone cameras and it requires a great performance and speed from camera systems and pushes camera developers to find new innovative breakthroughs to fulfill users' needs. However, features like great pixel count, image stabilization, auto focus, auto exposure and auto white balance together with different mixtures of hardware and software based image pipelines and very complex image defect correction generates different and unpredictable delays to the image capturing. An excellent image quality may be ignored, if the camera is always too slow to capture the needed action. It is also notable that according to statistical analysis of large image databases, over 70% of digitally captured images contain human faces⁴. Therefore, different face, smile and blink detection algorithms may cause extra delays to the image capturing pipeline.

As well as quality benchmarking metrics, there are existing methods to evaluate the camera speed. For example, very recently accepted standard ISO 15781 defines several methods to evaluate the shooting time lag, shutter release time lag, shooting rate and start-up time⁵. However, the standard is more suitable to compact and DSLR cameras and the methods can be difficult to use in mobile phone environment. It is also notable that the standard is focused only to the still image capturing measurements. The role of video recording is increasing even faster than still image capturing and therefore video performance measurements are also needed.

*veli-tapani.peltoketo@sofica.fi; phone +358 44 517 8552; sofica.fi

It is quite obvious that individual camera standards focus to specific quality or speed metrics. Therefore, combinations of different quality and speed metrics are missing. If mobile phone cameras are supposed to compare in a comprehensive way, this kind of combined benchmarking metrics is needed.

There are several tasks in this work. Firstly, the most important image quality metrics are collected from the standards and papers. Secondly, the speed related metrics of a mobile phone's camera system are collected from the standards and papers and also novel speed metrics are identified. Thirdly, combinations of the quality and speed metrics are validated using mobile phones in the market. Mobile phones are selected from three different operating systems and they represent the flagship models of five mobile phone manufacturers. The measurements are done using software based and automatic test system which is executed towards application programming interface of different operating systems. This approach gives comparable speed measurement values between different operating systems and removes the influence of mobile phone specific camera applications. Finally, the results are evaluated and conclusions are made.

The result of this work gives detailed benchmarking results of mobile phone camera systems in the market. The paper defines also a proposal of combined benchmarking metrics, which includes both quality and speed parameters

2. SUITABLE CAMERA QUALITY BENCHMARKING METRICS

2.1 Generally

When a mobile phone camera is tested and benchmarked, the significance of quality metrics is widely acknowledged. There are numerous standards from the analog camera era which are partially valid and partially updated to the digital camera testing. I3A started to define camera phone image quality metrics in CPIQ (Camera Phone Image Quality) group in 2007. Currently, work group P1858 of IEEE continues the work of the CPIQ group.

Imaging quality area contains a notable amount of different metrics and using all the metrics a camera can be validated in a very comprehensive way. However, the target of the benchmarking is to give an easily understandable score of the target devices which can be then used to sort the devices. Usually a single score value is used to make the comparison. If the score is derived from too many components, the influence of a quality parameter is difficult to weight to a single score. Also the interpretation of the score will become difficult.

Sharpness, color reproduction and noise metrics are usually defined as the most significant quality parameters of a digital camera. Even if several other quality metrics are available like lens specific quality metrics as lens distortion and lateral chromatic aberration, the quality metrics of this work are selected from sharpness, color accuracy and noise areas.

2.2 Image Sharpness and Resolution

Modulation transfer function curves (MTFs) are commonly used in the resolution measurements as ISO standard 12233 describes⁶. The standard defines a high contrast slanted edge type reference image which is captured by measured device. The sharp edge acts like an impulse type signal to the imaging device and the impulse response of the edge can be defined as a MTF curve. The MTF method is used, for example, to measure the quality of the lens systems. Because the camera system MTF is a product of the camera component's MTFs, the evaluation of the result is more complicated.

In modern mobile phones the captured image is processed heavily before it is accepted as a final image. Quite often the image processing adds artificial sharpness and denoising to the image. Especially combination of sharpening and denoising may corrupt the MTF curve to show too good results for resolution⁷. Artificial image processing algorithms may also decrease the texture resolution dramatically which cannot be evaluated using slanted edge type testing charts.

One option to minimize the sharpness effect to the MTF curve is use a low contrast version of the slanted edge because sharpness algorithms do not modify low contrast edges as much as high contrast ones. Another way to avoid sharpening is to use sinusoidal Siemens star reference images⁸, where the high contrast edge is replaced using an edge which density varies according to sinusoidal curve.

Onwards, so called dead leaves method has been defined in several papers^{7, 10, 11} and shortly by CPIQ group⁹. The method uses a reference image which contains random circles which are following common statistics of natural images. The power spectrum (PS) can be calculated from the reference image as well as captured image and MTF is calculated by dividing the spectrums. The dead leaves method is a powerful way to measure the texture resolution of the camera but is not accurate when the image is very noisy. However, the effect of noise can be decreased by calculating the noise level and decreasing the noise from captured image¹² (1).

$$\text{SFR}(f) = \sqrt{\frac{PS_{\text{image}}(f) - PS_{\text{noise}}(f)}{PS_{\text{reference}}(f)}} \quad (1)$$

Different resolution measurement targets are shown in Figure 1.

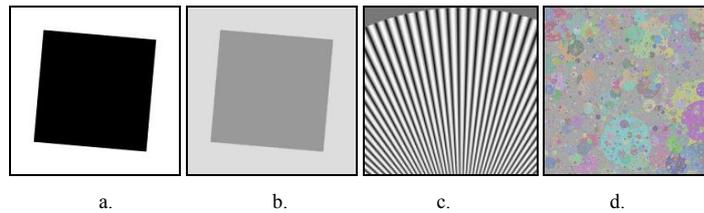


Figure 1. a. High contrast slanted edge, b. low contrast slanted edge, c. detail of sinusoidal Siemens star and d. colored dead leaves

Due to sharpness and denoise functionalities of modern mobile phone cameras, both low contrast slanted edge and colored dead leaves MTFs were selected to the benchmarking metrics. Onwards, MTF50Peak value was selected because it decreases the effect of the sharpening. Both vertical and horizontal MTF values are calculated from the slanted edge areas.

It is widely known that camera optics may cause more distortion to the edge areas of the image than to the center. Due to small optics and large field of view, this phenomenon can be very difficult in mobile phone cameras. To measure the resolution comprehensively, there are five slanted edge measurement points in the image: one in the center and one in each corner. The MTF50Peak value is the mean of these measurements but the center resolution has four times bigger weight than the corner resolution.

The dead leaves measurement point is located to the center area of the image because the sharpening and denoising affects equally over the image area.

2.3 Color Representation

There has been a lot of discussion between color difference formulas. Evaluations between CIE color difference formulas¹³ and ISO formulas¹⁴ have been done in several papers^{15, 16}. The results are quite contradictory and very dependent on the coefficients used in formulas.

In this work, CIEDE2000 formulas are used. The standard replaces earlier CIE76 and CIE94 formulas and correlates better to the visual differences of the colors than the previous formulas. However, the CIEDE2000 is very complicated formula compared to CIE76 and CIE94 formulas and the influence of single changes on L^* , a^* and b^* components is more difficult to predict. CIEDE2000 has some issues with the discontinuity, but they are evaluated not to be severe¹⁷. CIEDE2000 defines also precisely the measurement environment, which helps to build up the imaging laboratory.

In this work, CIEDE2000 formulas are used with 1:1:1 coefficient as the standard recommends when reference conditions are used. Onwards, ΔE_{00} value and mean saturation error are selected to represent the color correctness in the benchmarking. The mean saturation error is calculated from the gray patches of the ISO 15739 charts. This error defines the white balance correctness of the captured image.

2.4 Noise

The selection of noise metrics is quite straightforward. ISO standard 15739¹⁸ is widely used as a noise measurement reference. Signal to Noise (SNR) value is used in the benchmarking and according to standard, camera gamma, fixed pattern noise and temporal noise values are used to calculate the SNR. Fixed pattern and temporal noise metrics are calculated from eight pictures where the maximum location deviation between images is not more than 0.25 pixels.

However, the SNR does not represent fully the visual experience. Especially when the noise components are large, the visual experience is much worse than the SNR value defines. ISO 15739 has defines a measurement called Visual Noise which weights the noise components to correlate better to the visual experience. Visual noise measurements are not part of this work but they will be included to the next mobile phone camera benchmarking work.

3. SUITABLE CAMERA SPEED BENCHMARKING METRICS

3.1 Generally

Capturing the moment is an essential requirement in modern mobile phone cameras and it requires a great performance and speed from camera. However, features like great pixel count, image stabilization, auto focus, auto exposure and auto white balance together with different mixtures of hardware and software based image pipelines and very complex image defect correction generates different and unpredictable delays to the image capturing. An excellent image quality may be ignored, if the camera is always too slow to capture the needed action. It is also notable that according to statistical analysis of large image databases, over 70% of digitally captured images contain human faces. Therefore, different face, smile and blink detection algorithms may cause extra delays to the image capturing pipeline.

Very recently accepted standard ISO 15781 defines several methods to evaluate the shooting time lag, shutter release time lag, shooting rate and start-up time⁵. However, the standard is more suitable to compact and DLSR cameras and the methods can be difficult to use in mobile phone environment. It is also notable that the standard is focused only to the still image capturing measurements. The role of video recording is increasing even faster than still image capturing and therefore video performance measurements are also needed.

The main criterion of speed metrics is to measure the camera functionalities which represent the main use cases of the camera user. Currently, there are no standards or papers which define this kind of use cases.

3.2 First Image Capture with Camera startup

Obviously, the image capture time is one of the most important speed metrics. Adding camera startup and focus times to the measurement, the metric represents the situation where user sees an interesting object, takes the mobile phone, starts the camera and captures an image. Partially the measurement result depends on the environment. Especially light environment affects to the exposure time and also may affect to the focus time. Also the image content may affect to the image processing time. On the other hand, camera start up time and shutter lag should not be varied in different environments. A stable environment is used in this work and it offers comparable measurement result between devices.

Total image capturing time with camera startup and focus is selected one of the speed metrics of the benchmarking. However, the software based measurement methods enable to isolate different delay components from the image capture time. Camera startup, focus, shutter lag and image processing time can be measured one by one and their affect to the total image capture time can be evaluated.

3.3 Consecutive Image Capturing

Consecutive image capturing is an interesting metric because it may reveal different delays in the camera implementation than single image capturing. Features like memory handling, bus speed between image sensor and processors and even multi thread execution may affect to the consecutive image capturing speed when several images are processed in a row. Certainly, this feature is important also to the user. Especially when images are captured from moving object, the speed of consecutive image capturing acts a very important role of the camera performance.

The consecutive image capturing measurement does not include camera startup time nor focus time. It measures only the shutter lag, exposure and image processing times. In this work, time of five consecutive images is selected to the benchmarking metrics.

3.4 Video Metrics

This work contains one video related metric; the audio/video synchronization delay of the recorded video. Different implementation solutions of the camera hardware and software may generate a lot of extra delay between video and audio component of the recorded video. Especially, when audio leads clearly the video component, the user experience is unpleasant. Several television broadcasting standards have specified the limits of the delay. The latest recommendation is from year 2007, which defines that the audio delay should be less than 60 ms and audio lead should be less than 40 ms¹⁹.

Among the measured devices, the video functionality is part of the camera software stack and separate video recording startup time cannot be measured. However, the swapping delay from still image mode to video mode can be measured as well as the delay between recording start and receiving the first video frame and they will be part of the future benchmarking work.

4. MEASUREMENT METHODOLOGY

4.1 Benchmarking by Software

The measurements are done using software based and automatic measurement system which is executed towards application programming interface (API) of different operating system as shown in Figure 2. This approach gives comparable speed measurement values between different operating systems and removes the influence of mobile phone specific camera applications. In case of quality measurements, the camera control is done using the same API.

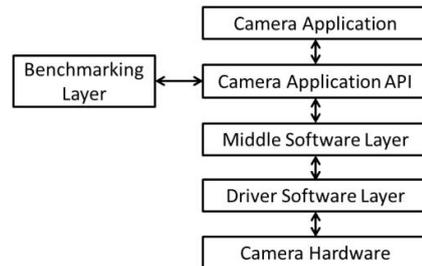


Figure 2. The measurement point of the camera software stack

The public APIs of different operating systems offers measurement points which can be used to calculate the speed of the camera. For example, Android operating system offers `Camera.takePicture` and `Camera.PictureCallback` –methods by Java camera API, which can be used to measure the image capturing time of the device. In case of quality measurements, separate measurement points are not required but the result image is the measured entity.

However, the usage of the camera application API causes some restrictions to the benchmarking. Firstly, some of camera features may be implemented only to the camera application layer and therefore they cannot be measured in this work. For example, part of measured devices includes face-, smile- and blink detection algorithms inside camera application and they cannot be activated and measured through camera application API layer. For this reason, the extra delay of these algorithms cannot be measured. Secondly, the true user experience of the camera speed is dependent on the camera application implementation and the delays of the camera application cannot be measured in this work.

It is notable, that all quality measurements are based on jpeg-compressed images because only few mobile phone camera supports raw-format images. However, the measurements are comparable because jpeg images are used in every measured device.

4.2 Measurement Environment

A separate imaging laboratory is used for benchmarking measurements. The speed and quality measurements are done in the same environment. Quality measurements are based on standardized testing charts which are located to the testing scene and the scene is illuminated using high quality lights. The illumination uniformity is measured before measurements and it has to be less than $\pm 5\%$. In this work, the measurements are done using one illumination environment; 1000 lux. One of the future tasks will be the benchmarking using different illumination levels.

The background of the scene is 18% neutral matt grey and following test charts are mounted to the scene:

- 20 grey patches to calculate ISO 14524 OECF curve and ISO 15739 noise. Charts are located circularly around the middle area of the scene.
- Macbeth color chart to color accuracy measurements. Chart is located to the middle area of the scene.
- Low contrast slanted edge charts in the middle and each corner to sharpness measurements. 5% angle and 4:1 contrast are used.
- Colored dead leaves chart to texture sharpness measurement and to detect denoising and sharpening defects. Chart is located to the middle area of the scene.

The audio/video synchronization measurements are done in an isolated environment where an accurately synchronized light and voice source is used as a reference to the audio/video synchronization measurement.

4.3 Different Measurements to Single Score

Following metrics are selected to the speed and quality camera benchmarking:

- MTF50P_{edge} value of slanted edge charts, weighted mean value of center and corners
- MTF50P_{deadLeaves} value of dead leaves chart. The value is noise corrected.
- ISO SNR value calculated from the gray charts
- ΔE_{00} and saturation error (satErr)
- Total image capture time with camera startup and focus time (t_{single})
- Time of five consecutive image capturing without camera startup or focus time (t_{five})
- Audio/Video delay of recorded video (t_{AV})

All values are mean values of several images to get reliable test results. All speed related measurements are done at least from five images and quality measurements at least from eight images. If the measurements contain clear outliers, they are extracted from the benchmarking calculation and separately mentioned with the benchmarking score.

Obviously, calculations are needed before the metrics can be combined to one score value. Part of the measurements defines a better value when they are small ones (for example, speed metrics) and vice versa (for example, SNR). Also the scale of the metrics varies a lot. Combining such metrics reveals another problem: How to define a single benchmark number which characterizes each metric fairly.

The problem has been evaluated in several papers^{20, 21, 22}. Arithmetic, harmonic and geometric mean have been evaluated to combine different metrics. There is no unambiguous solution to the problem because usage of mean values is always misleading. Thus it is necessary to reveal all the measurement values which are used to the calculations.

After evaluation, the geometric mean was observed to be suitable to combine different benchmarking metrics. According to the evaluation and using a large measurement database, each of five quality metric influenced 12-15% to the median score which was calculated from the measurement database. Onwards, each of three speed metric influenced 10-25% to the median score. Three different scores were calculated: speed score (2), quality score (3) and total benchmark score (4), which combines all metrics used in this work.

$$\text{Speed Score} = \sqrt[3]{\frac{1}{t_{single}} * \frac{1}{t_{five}} * \frac{1}{t_{AV}}} \quad (2)$$

$$\text{Quality Score} = \sqrt[5]{MTF50P_{edge} * MTF50P_{deadLeaves} * SNR * \frac{1}{\Delta E_{00}} * \frac{1}{(0.1 + satErr)}} \quad (3)$$

$$\text{Total Score} = \sqrt[8]{MTF50P_{edge} * MTF50P_{deadLeaves} * SNR * \frac{1}{\Delta E_{00}} * \frac{1}{(0.1 + satErr)} * \frac{1}{t_{single}} * \frac{1}{t_{five}} * \frac{1}{t_{AV}}} \quad (4)$$

As mentioned, a single score can be very misleading and using appropriate equations and weights the result score can be manipulated very efficiently. It is notable, that equations 2-4 do not use any weight components and each measured speed and quality metrics are used as such. The only exception is saturation error which value is summarized with 0.1 to decrease its too big influence to the total score.

5. RESULTS

5.1 Measured Devices

Five different devices are measured. The devices are selected from three different operating systems and they represent the flagship models of five mobile phone manufacturers. The sensor sizes of the cameras are equal or more than eight mega pixels. Obviously, each camera has auto focus, auto exposure and auto white balance functionalities. All measurements are done using default settings of the camera.

5.2 Benchmarking Results

Benchmarking results are summarized to three tables and one coordinate system. Table 1 includes the speed measurement of the devices and speed score. Table 2 shows corresponding quality values and scores. Finally, Table 3 defines the single benchmarking score and Figure 3 the relation between speed and quality scores.

Camera start-up and focus times are shown to Table 3 to declare the delay components of the image capturing, even if they are already included to the total image capture time. For example camera start-up time varies a lot between cameras and influences to the user experience. Onwards, Device #5 has significant long times both total image capture and five image capture times which clearly decreases the speed score.

Device #1 has extremely good five image capture time. Partially this can be explained by burst imaging mode which cannot be switched off in Device #1. All of the measured devices do not support burst mode or the mode cannot be activate without device's camera application.

Table 1. Speed measurement values and scores of the devices, sorted by the speed score

Device	Camera start-up, seconds	Focus, seconds	Total image capture, seconds	Five image capture, seconds	A/V sync delay, seconds	Speed score (3)
Device #1	0.692	0.738	1.519	0.326	0.057	3.28
Device #2	0.398	0.740	2.066	4.513	0.032	1.50
Device #3	1.040	0.611	2.706	8.114	0.029	1.16
Device #4	0.787	0.757	2.530	5.690	0.057	1.07
Device #5	0.984	0.507	4.983	21.339	0.024	0.73

The differences between quality scores are smaller than speed ones. The sharpness is measured by using unit LP/PH (Line Pairs/Picture Height). Using this unit, number of pixels of the sensor is also taken into account and different sensor sizes are compared fairly. The most interesting measurement is very good SNR value of Device #3. According to the visual inspection and MTF curves, Device #3 uses heavy image post processing algorithms as sharpness and denoising. Obviously, the denoising algorithm causes the very good SNR value.

Table 2. Quality measurement values and scores of the devices, sorted by the quality score

Device	MTF50_Peak, edge (LP/PH)	MTF50_Peak, dead leaves (LP/PH)	ISO SNR	ΔE_{00}	saturation error	Quality score (2)
Device #4	805	960	36.4	5.45	0.039	32.7
Device #3	1084	960	42.9	7.32	0.110	31.1
Device #5	1163	992	19.9	5.34	0.080	29.9
Device #2	935	935	21.8	5.81	0.110	27.5
Device #1	857	588	29.5	8.58	0.156	23.2

Clearly, the benchmarking score correlates fully with speed score and the reason is quite obvious: Among the measured devices the variance of the quality score is significantly smaller than variance of the speed score. To get more reliable single score value, some weights have to be used.

Table 3. Total benchmark score of the devices, sorted by the benchmark score

Device	Benchmarking score (4)
Device #1	11.2
Device #2	9.22
Device #3	9.07
Device #4	9.06
Device #5	7.44

A coordinate system is more informative way to present the speed and quality scores as Figure 3 shows. Even if number of samples is small, trend can be seen: High image quality means slower camera functionality.

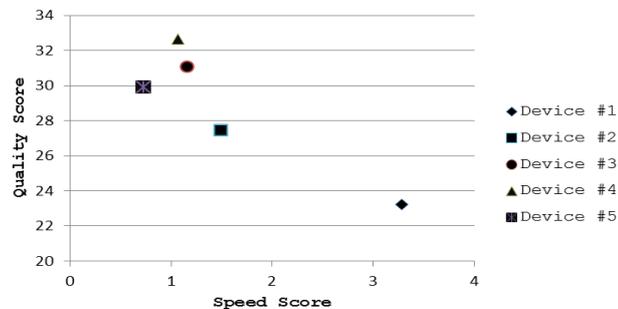


Figure 3. Measured devices in a speed-quality coordinate system

6. CONCLUSIONS

When mobile phone cameras are benchmarked using speed and quality metrics the results are significantly different. According to devices tested in this work, mobile phone cameras with good quality scores are slower than devices which do not get so high quality values.

Nowadays, the mobile phones have very powerful processors and camera specific signal processors. However, it seems that the high processor power cannot fully compensate the growing pixel amount and comprehensive image post processing. Some of the good quality cameras were so slow that they affect obviously to the user experience. Clearly, both speed and quality measurements are required to benchmark a mobile phone camera comprehensively.

Making a user friendly, single benchmark score is challenging. A comprehensive measurement requires several speed and quality metrics. Onwards, the conversion from several different units to one score is problematic. Geometric mean might be a solution, but due to variance differences between speed and quality scores, the speed score is too dominant. If weighting factors are not used with the geometric mean calculations, the speed-quality coordinate system is the best way to specify the mobile phone camera benchmark.

Some future tasks were found during this work. The most essential enhancements are be the usage of visual noise, evaluation of different weighting of score calculations and probability to measure the speed metrics from the graphical end user interface.

REFERENCES

- [1] International Imaging Industry Association, "Camera Phone Image Quality – Phase 1 White Paper, Fundamentals and review of considered test methods" (2007)
- [2] International Imaging Industry Association, "Camera Phone Image Quality – Phase 2, Lens Geometric Distortion" (2009)
- [3] International Imaging Industry Association, "Camera Phone Image Quality – Phase 2, Lateral Chromatic Aberration" (2009)
- [4] Wueller D., Fageth R., "Statistic Analysis of Millions of Digital Photos" Proc. SPIE 6817, (2008)
- [5] International Organization of Standardization, "ISO 15781 Photography – Digital still cameras – Measuring shooting time lag, shutter release time lag, shooting rate, and start-up time" (2013)
- [6] International Organization of Standardization, "ISO 12233 Photography — Electronic still-picture cameras — Resolution measurements" (2000)
- [7] Cao, F., Guichard, F. and Hornung H., "Dead leaves model for measuring texture quality on a digital camera," Proc. SPIE 7537, (2010)
- [8] Loebich, C., Wueller, D., Klingen, B. and Jaeger, A., "Digital Camera Resolution Measurement Using Sinusoidal Siemens Star," Proc SPIE 6502, (2007)
- [9] International Imaging Industry Association, "Camera Phone Image Quality – Phase 2, Initial work of Texture Metric" (2009)
- [10] McElvain, J., Campbell, S. P., Miller, J. and Jin, E. W., "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems," Proc. SPIE 7537, (2010)
- [11] Burns, P. D. and Williams, D. "Measurement of Texture Loss for JPEG 2000 Compression," Proc. SPIE 8293, (2012)
- [12] Artmann, U. and Wueller, D. "Improving texture loss measurement: spatial frequency response based on a colored target," Proc. SPIE 8293, (2012)
- [13] International Commission of Illumination, "CIE S 014-6/E:2013 Colorimetry – Part6: CIEDE2000 Colour-Difference Formula" (2013)
- [14] International Organization of Standardization "ISO 105-J03:2009 Textiles -- Tests for colour fastness -- Part J03: Calculation of colour differences" (2009)
- [15] Huang, M., Liu, H., Xiao Y. and Cai Q. "Research on Digital Images' Color-Difference by altering Lightness and Chroma: Analysis and Evaluation of Color-Difference Formulae," 3rd International Congress on Image and Signal Processing, 2347-2350 (2010)
- [16] Hao-xue, L., Meng, X. and Min, H. "Image Color-Difference Evaluation Based on Color-Difference Formula," 4th International Congress on Image and Signal Processing, 1771-1774 (2011)
- [17] Sharma, G. Wu, W. and Dalal E.N. "The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations", Color Research & Application 30(1), 21–30 (2005)
- [18] International Organization of Standardization, "ISO 15739 Photography – Electronic still-picture imaging – Noise measurements" (2003)
- [19] EBU - European Broadcasting Union, "The relative timing of the sound and vision components of a television signal," EBU Recommendation R37-2007 (2007)
- [20] Fleming, P. J. and Wallace, J. J. "How not to lie with statistics: The correct way to summarize benchmarking results," Communications of the ACM 29(3), 218-221 (1986)
- [21] Smith, J.E., "Characterizing computer performance with a single number," Communications of the ACM 31(10), 1202-1206 (1988)
- [22] Lilja, D. J., [Measuring computer performance], Cambridge University Press, Cambridge, 24-41 (2005)

Evaluation of mobile phone camera benchmarking using objective camera speed and image quality metrics

Veli-Tapani Peltoketo^{a,b,*}

^aUniversity of Vaasa, Faculty of Technology, Wolffintie 34, Vaasa, 65200 Finland

^bSofica Ltd., Kampusranta 9C, Seinäjoki, 60320 Finland

Abstract. When a mobile phone camera is tested and benchmarked, the significance of image quality metrics is widely acknowledged. There are also existing methods to evaluate the camera speed. However, the speed or rapidity metrics of the mobile phone's camera system has not been used with the quality metrics even if the camera speed has become a more and more important camera performance feature. There are several tasks in this work. First, the most important image quality and speed-related metrics of a mobile phone's camera system are collected from the standards and papers and, also, novel speed metrics are identified. Second, combinations of the quality and speed metrics are validated using mobile phones on the market. The measurements are done toward application programming interface of different operating systems. Finally, the results are evaluated and conclusions are made. The paper defines a solution to combine different image quality and speed metrics to a single benchmarking score. A proposal of the combined benchmarking metric is evaluated using measurements of 25 mobile phone cameras on the market. The paper is a continuation of a previous benchmarking work expanded with visual noise measurement and updates of the latest mobile phone versions. © 2014 SPIE and IS&T [DOI: 10.1117/1.JEI.23.6.061102]

Keywords: digital imaging; image quality; image evaluation; resolution; noise; color.

Paper 14146SSP received Mar. 24, 2014; revised manuscript received May 5, 2014; accepted for publication May 21, 2014; published online Jul. 21, 2014.

1 Introduction

When a mobile phone camera is tested and benchmarked, the significance of image quality metrics is widely acknowledged. Generally, sharpness, color reproduction, and noise metrics are usually defined as the most significant quality parameters of the camera.¹ Several other quality metrics are also available. For example, International Imaging Industry Association (I3A), currently Institute of Electrical and Electronics Engineers (IEEE) P1858 working group, has defined the lens distortion and lateral chromatic aberration as quality components.^{2,3} However, the speed or rapidity metrics of the mobile phone's camera system has not been used with the quality metrics even if the camera speed has become a more and more important camera performance feature.

Capturing the moment is an essential requirement in modern mobile phone cameras, and it requires a great performance and speed from camera systems and pushes camera developers to find new innovative breakthroughs to fulfill users' needs. However, features like great pixel count, image stabilization, auto focus, auto exposure, and auto white balance together with different mixtures of hardware- and software-based image pipelines and very complex image defect corrections generate different and unpredictable delays in the image capturing. An excellent image quality may be ignored if the camera is always too slow to capture the needed action. It is also notable that according to statistical analysis of large image databases, >70% of digitally captured images contain human faces.⁴ Therefore, different

face, smile, and blink detection algorithms may cause extra delays to the image capturing pipeline.

Apart from the quality benchmarking metrics, there are existing methods to evaluate the camera speed. For example, recently accepted standard ISO 15781 defines several methods to evaluate the shooting time lag, shutter release time lag, shooting rate, and start-up time.⁵ However, the measurement methods are difficult to use in the mobile phone environment. It is also notable that the standard is focused only on still image capturing measurements. The role of video recording is increasing even faster than still image capturing and, therefore, video performance measurements are also needed.

It is quite obvious that individual camera standards focus on specific quality or speed metrics. Therefore, combinations of different quality and speed metrics are missing. If mobile phone cameras are supposed to compare in a comprehensive way, this kind of combined benchmarking metrics is needed.

There are several tasks in this work. First, the most important image quality metrics are collected from the standards and papers. Second, the speed-related metrics of a mobile phone's camera system are collected from the standards and papers and also novel speed metrics are identified. Third, combinations of the quality and speed metrics are validated using mobile phones in the market. In total, 25 mobile phone models are measured and used as a reference market data. Five of the devices are selected and detailed measurement data are provided in this paper. The selected phones represent three different operating systems, and they are flagship models of five different mobile phone manufacturers. The measurements are done using a software-based and

*Address all correspondence to: Veli-Tapani Peltoketo, E-mail: veli-tapani.peltoketo@sofica.fi

automatic test system, which is executed toward application programming interface of different operating systems. This approach gives comparable speed measurement values between different operating systems and removes the influence of mobile phone specific camera applications. Finally, the results are evaluated and conclusions are made.

The paper defines a proposal of combined benchmarking metric, which includes objective image quality and camera speed parameters. The result of this work also gives detailed benchmarking results of mobile phone camera systems in the market.

The paper is a continuation of a previous benchmarking work.⁶ The previous paper is expanded with visual noise measurements of the latest ISO 15739 standard.⁷ Moreover, the number of measured devices has been increased from 5 to 25 and the model names of the devices are revealed. Finally, the very latest versions of mobile phone models are used in the measurements and the results are updated accordingly.

2 Suitable Image Quality Benchmarking Metrics

There are numerous image quality standards from the analog camera era, which are partially valid and partially updated to the digital camera testing. I3A started to define camera phone image quality metrics in the Camera Phone Image Quality (CPIQ) group in 2007. Currently, work group P1858 of IEEE continues the work of the CPIQ group.

The image quality measurement contains a notable amount of different metrics, and using all the metrics, a camera can be validated in a very comprehensive way. However, the target of the benchmarking is to give an easily understandable score of the measured devices, which can be used to sort the devices. Usually, a single score value is used to make the comparison. If the score is derived from too many components, the influence of a quality parameter is difficult to weight to a single score. Also, the interpretation of the score will become difficult.

It is notable that current standards are based on objective quality metrics and measurements as Keelan defines the objective quantities,⁸ which means that the corresponding benchmarking scores are objective scores. The usage of perceptual benchmarking scores would require a different approach to the score algorithms and it is not part of this work.

Sharpness, color reproduction, and noise metrics are defined as the most significant image quality parameters of a digital camera; therefore, the quality metrics of this work are selected from these three areas.

2.1 Image Sharpness and Resolution

Modulation transfer function (MTF) curves are commonly used in the resolution measurements as ISO standard 12233 describes.⁹ The standard defines a high-contrast slanted edge type reference image, which is captured by measured device. The sharp edge acts like an impulse type signal to the imaging device, and the impulse response of the edge can be defined as an MTF curve. The MTF method is used, for example, to measure the quality of the lens systems. Because the camera system MTF is a product of the camera component's MTFs, the evaluation of the result is more complicated.

In modern mobile phones, the captured image is processed heavily before it is accepted as a final image. In particular, a combination of sharpening and denoising may corrupt the MTF curve to show too-good results for resolution.¹⁰ Artificial image processing algorithms may also decrease the texture resolution dramatically, which cannot be evaluated using slanted edge type testing charts.

One option to minimize the sharpness effect to the MTF curve is to use a low-contrast version of the slanted edge because sharpness algorithms do not modify low-contrast edges as much as high-contrast ones. Another way to avoid sharpening is to use sinusoidal Siemens star reference images,¹¹ where the high-contrast edge is replaced using an edge whose density varies according to sinusoidal curve.

Moreover, the so-called dead leaves method has been defined in several papers^{10,12,13} and briefly by CPIQ group.¹⁴ The method uses a reference image that contains random circles, which are following common statistics of natural images. The power spectrum (PS) can be calculated from the reference image as well as captured image and the MTF is calculated by dividing the spectra. The dead leaves method is a powerful way to measure the texture resolution of the camera but is not accurate when the image is very noisy. However, the effect of noise can be decreased by calculating the noise level and decreasing the noise from captured image.¹⁵

$$\text{MTF}(f) = \sqrt{\frac{\text{PS}_{\text{image}}(f) - \text{PS}_{\text{noise}}(f)}{\text{PS}_{\text{reference}}(f)}} \quad (1)$$

Examples of resolution measurement targets are shown in Fig. 1.

Due to sharpness and denoise functionalities of modern mobile phone cameras, both low-contrast slanted edge and colored dead leaves MTFs were selected to the benchmarking metrics. Moreover, MTF50Peak value was selected to decrease the effect of the sharpening.

Camera optics may cause more distortion to the edge areas of the image than to the center. Due to small optics and large field of view, this phenomenon can be very difficult in mobile phone cameras. To measure the resolution comprehensively, there are five slanted edge measurement points in the image: one in the center and one in each corner. Both vertical and horizontal MTF values are calculated from the five slanted edge areas. The MTF50Peak value of each area is the mean of horizontal and vertical values. Moreover, the final MTF50Peak value is the mean of each area, but the center resolution has four times bigger weight than the corner resolution.

The dead leaves measurement point is located in the center area of the image because the sharpening and denoising affect equally the whole image area.

The sharpness is measured by using the unit line pairs/picture height (LP/PH). Using this unit, number of pixels of the sensor is also taken into account. Another option would be to use the unit cycles/pixel. In this case, pixel specific sharpness can be measured. Obviously, units are not identical because the number of pixels affects the LP/PH unit. Even if the pixel amount does not tell the true sharpness of the sensor, it may significantly affect digital zoom, which is one of the key features of the sensors with big pixel amounts. Using LP/PH

Peltoketo: Evaluation of mobile phone camera benchmarking using objective camera speed...

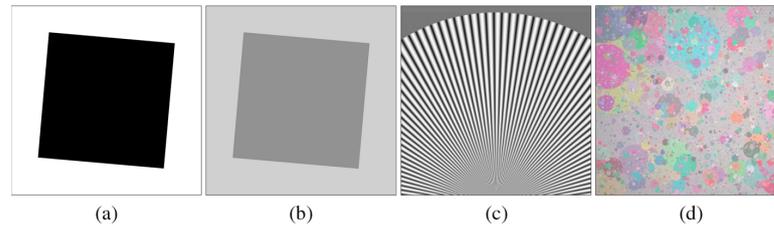


Fig. 1 Examples of the resolution test charts: (a) high-contrast slanted edge, (b) low-contrast slanted edge, (c) detail of sinusoidal Siemens star, and (d) colored dead leaves.

units, the benchmarking metric also considers the digital zoom functionality.

2.2 Color Reproduction

There has been a lot of discussion about color difference formulas. Evaluations between the International Commission on Illumination (CIE) color difference formulas¹⁶ and ISO formulas¹⁷ have been done in several papers.^{18,19} The results are quite contradictory and very dependent on the coefficients used in the formulas.

In this work, CIEDE2000 formulas are used. The standard replaces earlier CIE76 and CIE94 formulas and correlates better to the visual differences of the colors than the previous formulas. However, the CIEDE2000 is a very complicated formula compared to CIE76 and CIE94 formulas, and the influence of single changes on L^* , a^* , and b^* components is more difficult to predict. CIEDE2000 has some issues with discontinuity, but they are evaluated not to be severe.²⁰ CIEDE2000 also defines precisely the measurement environment, which helps to build up the imaging laboratory.

In this work, CIEDE2000 formulas are used with 1:1:1 coefficient as the standard recommends when the reference conditions are used. Moreover, color difference (ΔE_{00}) and mean saturation error are selected to represent the color correctness in the benchmarking. The mean saturation error is calculated from the gray patches of the ISO 15739 charts. This error defines the white balance correctness of the captured image.

2.3 Noise

The selection of noise metrics is quite straightforward. ISO standard 15739 (Ref. 7) is widely used as a noise measurement reference. Signal-to-noise ratio (SNR) value is used in the benchmarking, and according to standard, camera gamma, fixed pattern noise, and temporal noise values are used to calculate the SNR. Fixed pattern and temporal noise metrics are calculated from at least eight pictures.

However, the SNR does not fully represent the visual experience. Especially when the noise components are large, the visual experience is much worse than the SNR value defines. The newest ISO 15739 standard from 2013 defines a measurement called visual noise, which weights the noise components to correlate better to the visual experience. According to the standard, visual noise is calculated from each gray patch of the test chart, which means 20 values. In case of benchmarking, the amount of values is not usable; therefore, a root mean square value is calculated and used in the benchmarking measurements.

3 Suitable Camera Speed Benchmarking Metrics

The main criterion of speed metrics is to measure the camera functionalities, which represent the main use cases of the camera user. Currently, there are no standards or papers that define this kind of use case. Two still image speed metrics and one video recording metric is selected for the benchmarking measurements.

3.1 First Image Capture with Camera Start-Up

Obviously, the image capture time is one of the most important speed metrics. Adding camera startup and focus times to the measurement, the metric represents the situation where the user sees an interesting object, starts the camera of the mobile phone, and captures an image. Partially, the measurement result depends on the environment. Especially, light environment affects the exposure time, and it may also affect the focus time. Also the image content may affect the image processing time. On the other hand, camera start-up time and shutter lag should not be varied in different environments. A stable environment is used in this work and it offers comparable measurement result between devices.

Total image capturing time which includes camera start-up time, focus time, and image processing time is the first speed metric of the benchmarking. However, the software-based measurement methods enable to isolate different delay components from the image capture time. Camera startup, focus, shutter lag, and image processing time can be measured one by one, and their effect on the total image capture time can be evaluated.

3.2 Consecutive Image Capturing

Consecutive image capturing is an interesting metric because it may reveal different delays in the camera implementation than single image capturing. Features like memory handling, bus speed between image sensor and processors, and even multithread execution may affect the consecutive image capturing speed when several images are processed in a row. Certainly, this feature is also important to the user. Especially, when several images are captured from a moving object, the speed of consecutive image capturing plays a very important role in the camera performance.

The consecutive image capturing measurement does not include camera start-up time or focus time. It measures only the shutter lag, exposure, and image processing times. In this work, time of five consecutive images is selected to the benchmarking metrics.

3.3 Video Metrics

This work contains one video-related metric: the audio/video synchronization delay of the recorded video. Different

implementation solutions of the camera hardware and software may generate a lot of extra delay between video and audio component of the recorded video. Especially, when audio clearly leads the video component, the user experience is unpleasant. Several television broadcasting standards have specified the limits of the delay. The latest recommendation is from 2007, which defines that the audio delay should be <60 ms and audio lead should be <40 ms.²¹

Among the measured devices, the video functionality is part of the camera software stack and separate video recording start-up time cannot be measured. However, the swapping delay from still image mode to video mode can be measured as well as the delay between recording start and receiving the first video frame, and they will be part of the future benchmarking work.

4 Measurement Methodology

The measurements are done using a software-based and automatic measurement system, which is executed toward application programming interface (API) of different operating systems as shown in Fig. 2. This approach gives comparable speed measurement values between different operating systems and removes the influence of mobile phone specific camera applications. In case of quality measurements, the camera control is done using the same API.

The public APIs of different operating systems offer measurement points that can be used to calculate the speed of the camera. For example, the Android operating system offers camera class by Java camera API, which can be used to measure the image capturing times of the device. In case of quality measurements, separate measurement points are not required but the result image is the measured entity.

However, the usage of the camera application API causes some restrictions in the benchmarking. First, some of the camera features may be implemented only in the camera application layer and, therefore, they cannot be measured in this work. For example, part of measured devices include face, smile, and blink detection algorithms inside the camera application, and they cannot be activated and measured through camera application API layer. For this reason, the extra delay of these algorithms cannot be measured. Second, the true user experience of the camera speed is dependent on the camera application implementation, and the delays of the camera application cannot be measured in this work.

It is notable that all quality measurements are based on Joint Photographic Experts Group (JPEG) compressed images because only few mobile phone cameras support

raw format images. However, the measurements are comparable because JPEG images are used in every measured device.

4.1 Measurement Environment

A separate imaging laboratory is used for benchmarking measurements. The speed and quality measurements are done in the same environment. Quality measurements are based on standardized testing charts, which are located on the testing scene, and the scene is illuminated using high-quality lights. The illumination uniformity is measured before benchmarking measurements and the uniformity error is $< \pm 5\%$. In this work, the measurements are done using one illumination environment: 1000 lux. One of the future tasks will be benchmarking using different illumination levels.

The background of the scene is 18% neutral matte gray, and following test charts are mounted on the scene:

- Twenty gray patches to calculate ISO 14524 opto-electronic conversion function (OECF) curve²² and ISO 15739 noise. The charts are located circularly around the middle area of the scene.
- A Macbeth color chart for color accuracy measurements. The chart is located in the middle area of the scene.
- Low-contrast slanted edge charts in the middle and each corner for sharpness measurements. 5% angle and 4:1 contrast are used.
- A colored dead leaves chart to texture sharpness measurement and to detect denoising and sharpening defects. The chart is located in the middle area of the scene.

Figure 3 shows the entire testing scene and locations of testing charts.

The audio/video synchronization measurements are done in an isolated environment where an accurately synchronized light and voice source is used as a reference to the audio/video synchronization measurement. Moreover, the recorded target contains a changing entity, which stresses the video encoders.

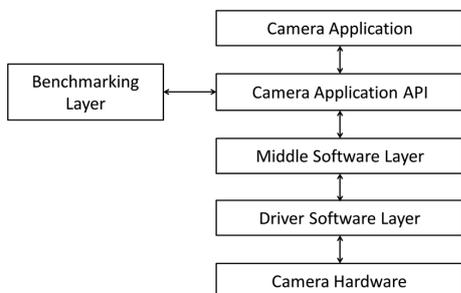


Fig. 2 The measurement point of the camera software stack.



Fig. 3 The testing scene.

4.2 Different Measurements to Single Score

Following metrics are selected for the speed and quality camera benchmarking:

- MTF50Peak value of the slanted edge charts. The value is weighted mean of center and corners (MTF50P_{edge}).
- MTF50Peak value of the dead leaves chart. The value is noise corrected (MTF50P_{deadLeaves}).
- ISO SNR value calculated from the gray charts (SNR).
- Visual noise value calculated from the gray charts (VN).
- CIEDE2000 color difference from the Macbeth color chart and saturation error from the gray charts (ΔE_{00} , satErr).
- Total image capture time with camera start-up and focus time (t_{single}).
- Time of capturing five consecutive images without camera start-up or focus time (t_{five}).
- Audio/video delay of recorded video (t_{AV}).

All values are average values of several images to get reliable test results. All speed-related measurements are done from at least five images and quality measurements from at least eight images. If the measurements contain clear outliers, they are extracted from the benchmarking calculation and separately mentioned with the benchmarking score.

Obviously, calculations are needed before the metrics can be combined to one score value. Part of the measurements define a better value when they are small ones (for example,

speed metrics) and vice versa (for example, SNR). Also, the scale of the metrics varies a lot. Combining such metrics reveals another problem: how to define a single benchmark number that characterizes each metric fairly.

The problem has been evaluated in several papers.²³⁻²⁵ Arithmetic, harmonic, and geometric mean have been evaluated to combine different metrics. There is no unambiguous solution to the problem because the usage of mean values is always misleading. Thus, it is necessary to reveal all the measurement values that are used in the calculations.

The selected metrics are based on objective measurements and, thus, the benchmarking scores are also objective ones. A perceptual approach would require calibrating each metric to the perceptual space as Keelan defines and calculating the overall quality using multivariate formalism.⁸ However, in case of objective metrics and benchmarking, a more straightforward approach can be reasonable to calculate scores.

After evaluation, the geometric mean was observed to be suitable to combine different benchmarking metrics. According to the evaluation and using a large measurement database, it was noticed that both quality and speed component influenced the final score equally. Three different scores were calculated: speed score [Eq. (2)], quality score [Eq. (3)], and total benchmark score [Eq. (4)], which combines all metrics used in this work.

$$\text{Speed score} = \sqrt[3]{\frac{1}{t_{\text{single}}} * \frac{1}{t_{\text{five}}} * \frac{1}{t_{AV}}} \quad (2)$$

$$\text{Quality score} = \sqrt[6]{\text{MTF50P}_{\text{edge}} * \text{MTF50P}_{\text{deadLeaves}} * \text{SNR} * \frac{1}{\text{VN}} * \frac{1}{\Delta E_{00}} * \frac{1}{\text{satErr}}} \quad (3)$$

$$\text{Total score} = \sqrt[9]{\text{MTF50P}_{\text{edge}} * \text{MTF50P}_{\text{deadLeaves}} * \text{SNR} * \frac{1}{\text{VN}} * \frac{1}{\Delta E_{00}} * \frac{1}{\text{satErr}} * \frac{1}{t_{\text{single}}} * \frac{1}{t_{\text{five}}} * \frac{1}{t_{AV}}} \quad (4)$$

As mentioned, a single score can be very misleading, and using appropriate equations and weights the result score can be manipulated very efficiently. It is notable that Eqs. (2) to (4) do not use any weight components and each measured speed and quality metric is used as such. The addition of a visual noise metric enables to remove the satErr related weight component of previous work.⁶ Both quality and speed components have now equal influence on the total score without any weight components.

One obvious weakness can be seen when this approach is used: the measured values in denominator cannot be zero.

5 Results

Twenty-five different mobile phone devices are measured and five of them are selected and detailed result data are provided in this paper. There are two exceptions on the device list: Samsung Galaxy Camera GC100 is an example of a compact camera and two software versions of Nokia Lumia 1020 are represented to give an example of software influence on the camera benchmark.

The selected five devices represent three different operating systems, and they are flagship models of five different mobile phone manufacturers. Every device has a

Table 1 Speed measurement values and scores of the devices, sorted by the speed score.

Device	Camera start-up (s)	Focus (s)	Total image capture (s)	Five image capture (s)	A/V sync delay (s)	Speed score (2)
Apple iPhone 5s	0.692	0.738	1.519	0.326	0.057	3.3
Huawei Ascend P2	0.398	0.740	1.977	4.513	0.032	1.5
Samsung Galaxy S4	1.040	0.611	2.705	8.114	0.029	1.2
LG G2	0.787	0.757	2.529	5.690	0.057	1.1
Nokia Lumia 1020	0.979	0.500	5.070	21.271	0.022	0.8

Peltoketo: Evaluation of mobile phone camera benchmarking using objective camera speed...

Table 2 Quality measurement values and scores of the devices, sorted by the quality score.

Device	MTF50_Peak, edge (LP/PH)	MTF50_Peak, dead leaves (LP/PH)	ISO SNR	Visual noise	ΔE_{00}	Saturation error	Quality score (3)
Nokia Lumia 1020	1320	1176	29.2	2.07	4.34	0.036	22.8
LG G2	797	956	36.5	1.79	5.77	0.039	20.3
Samsung Galaxy S4	1057	1026	44.5	1.42	6.92	0.120	18.6
Huawei Ascend P2	875	892	21.7	3.08	5.79	0.110	14.3
iPhone 5s	853	593	29.3	2.24	8.50	0.159	13.0

Note: LP/PH, line pairs/picture height

Table 3 Total benchmarking score of the devices, sorted by the benchmarking score.

Device	Benchmarking score (4)
iPhone 5s	8.2
LG G2	7.6
Samsung Galaxy S4	7.4
Nokia Lumia 1020	7.3
Huawei Ascend P2	6.8

complementary metal oxide semiconductor) camera sensor. The sensor sizes of the cameras are more than or equal to 8 megapixels. Obviously, each camera has autofocus, autoexposure, and auto white balance functionalities. All measurements are done using default settings and maximum resolution of the camera.

5.1 Benchmarking Results

Benchmarking results are summarized to three tables and a coordinate system. Table 1 includes the speed measurement of the devices and speed score. Table 2 shows corresponding quality values and scores. Finally, Table 3 defines the single benchmarking score and Fig. 4 the relation between speed and quality scores of all 25 devices.

Camera start-up and focus times are shown in Table 1 to declare the delay components of the image capturing, even if they are already included in the total image capture time. For example, camera start-up time varies a lot between cameras and influences the user experience. Moreover, Lumia 1020

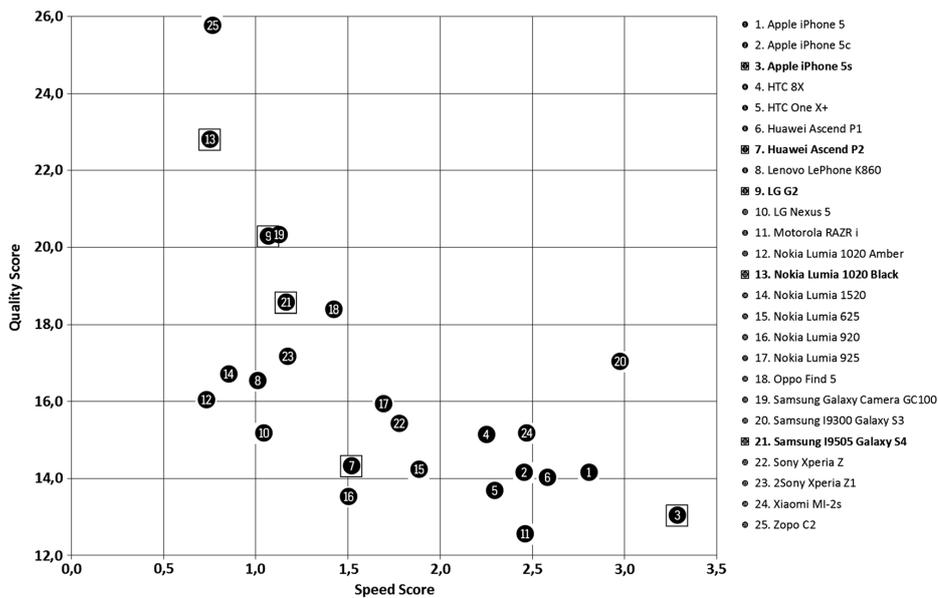


Fig. 4 Measured devices in speed-quality coordinate system.

Peltoketo: Evaluation of mobile phone camera benchmarking using objective camera speed...

has significant long times, both total image capture and five-image capture times, which clearly decreases the speed score.

iPhone 5s has extremely good five-image capture times. Partially, this can be explained by burst imaging mode, which cannot be switched off in iPhone 5s. All of the measured devices do not support burst mode or the mode cannot be activated without the device's camera application.

The ratio between the smallest and biggest quality scores is smaller than in the speed case. The most interesting measurement is very good SNR value of Galaxy S4. According to the visual inspection and MTF curves, Galaxy S4 uses heavy image postprocessing algorithms such as sharpness and denoising. Obviously, the denoising algorithm causes the very good SNR value.

It can be seen that the speed and quality scores rank the devices in exact opposite order and the total benchmarking score averages the order quite well. The result differs from the previous paper⁶ mainly for two reasons. First, the quality score of Lumia 1020 is improved significantly due to new software version of the model. This is a good example of the importance of software when an image is processed in a mobile phone. Second, poor visual noise decreased the benchmark rating of Ascend P2.

However, a coordinate system is probably a more informative way to present the speed and quality scores than a single benchmarking score as Fig. 4 shows. Scores of all 25 devices are shown in the figure and the selected devices are highlighted with a square.

Even if the operating system, hardware performance, and software implementation affects the camera speed, it can be noted that devices with high quality score have low speed scores and vice versa. High image quality seems to denote slower camera functionality.

6 Conclusions

When mobile phone cameras are benchmarked using objective speed and image quality metrics, the results are significantly different. According to devices tested in this work, mobile phone cameras with high image quality scores are slower than devices that do not get such high quality values.

Nowadays, mobile phones have very powerful processors and camera-specific signal processors. However, it seems that the high processor power cannot fully compensate for the growing pixel amount and heavy image postprocessing. Some of the cameras with high image quality were so slow that they obviously affect the user experience. Clearly, both speed and quality measurements are required to benchmark a mobile phone camera comprehensively.

Making a user-friendly, single benchmark score is challenging. A comprehensive measurement requires several speed and quality metrics. Moreover, the conversion from several different units to one score is problematic. Geometric mean could be a solution, but a single score value gives a very narrow impression of the different features of the camera. Probably, the speed-quality coordinate system is a better way to specify the mobile phone camera benchmark.

Some future tasks were found during this work. The most essential enhancements are the benchmarking using different illumination levels, probability to measure the speed metrics

from the graphical end user interface, and comprehensive video-related measurements.

References

1. International Imaging Industry Association, *Camera Phone Image Quality—Phase 1 White Paper, Fundamentals and Review of Considered Test Methods*, International Imaging Industry Association, White Plains, New York (2007).
2. International Imaging Industry Association, *Camera Phone Image Quality—Phase 2, Lens Geometric Distortion*, International Imaging Industry Association, Boston, Massachusetts (2009).
3. International Imaging Industry Association, *Camera Phone Image Quality—Phase 2, Lateral Chromatic Aberration*, International Imaging Industry Association, Boston, Massachusetts (2009).
4. D. Wueller and R. Fageth, "Statistic analysis of millions of digital photos," *Proc. SPIE* **6817**, 68170L (2008).
5. International Organization of Standardization, "Photography—digital still cameras—measuring shooting time lag, shutter release time lag, shooting rate, and start-up time," ISO 15781, ISO Copyright Office, Geneva (2013).
6. V.-T. Peltoketo, "Mobile phone camera benchmarking: combination of camera speed and image quality," *Proc. SPIE* **9016**, 90160F (2014).
7. International Organization of Standardization, "Photography—electronic still-picture imaging—noise measurements," ISO 15739, ISO Copyright Office, Geneva (2013).
8. B. W. Keelan, *Handbook of Image Quality: Characterization and Prediction*, pp. 149–180, CRC Press, New York (2002).
9. International Organization of Standardization, "Photography—electronic still-picture cameras—resolution measurements," ISO 12233 (2000).
10. F. Cao, F. Guichard, and H. Hornung, "Dead leaves model for measuring texture quality on a digital camera," *Proc. SPIE* **7537**, 75370E (2010).
11. C. Loebich et al., "Digital camera resolution measurement using sinusoidal Siemens star," *Proc. SPIE* **6502**, 65020N (2007).
12. J. McElvain et al., "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems," *Proc. SPIE* **7537**, 75370D (2010).
13. P. D. Burns and D. Williams, "Measurement of texture loss for JPEG 2000 compression," *Proc. SPIE* **8293**, 82930C (2012).
14. International Imaging Industry Association, *Camera Phone Image Quality—Phase 2, Initial Work of Texture Metric*, International Imaging Industry Association, Boston, Massachusetts (2009).
15. U. Artmann and D. Wueller, "Improving texture loss measurement: spatial frequency response based on a colored target," *Proc. SPIE* **8293**, 829305 (2012).
16. International Commission of Illumination, *CIE S 014-6/E:2013 Colorimetry—Part 6: CIEDE2000 Colour-Difference Formula*, CIE Central Bureau, Vienna, Austria (2013).
17. International Organization of Standardization, "Textiles—tests for colour fastness—Part J03: calculation of colour differences," ISO 105-J03:2009 (2009).
18. M. Huang et al., "Research on digital images' color-difference by altering lightness and chroma: analysis and evaluation of color-difference formulae," in *3rd Int. Congress on Image and Signal Processing*, pp. 2347–2350, IEEE Operations Center, Piscataway, NJ (2010).
19. L. Hao-Xue, X. Meng, and H. Min, "Image color-difference evaluation based on color-difference formula," in *4th Int. Congress on Image and Signal Processing*, pp. 1771–1774, IEEE Operations Center, Piscataway, NJ (2011).
20. G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations," *Color Res. Appl.* **30**(1), 21–30 (2005).
21. European Broadcasting Union, "The relative timing of the sound and vision components of a television signal," EBU Recommendation R37-2007 (2007).
22. International Organization of Standardization, "Photography—electronic still-picture cameras—methods for measuring OECFs," ISO 14524 (2009).
23. P. J. Fleming and J. J. Wallace, "How not to lie with statistics: the correct way to summarize benchmarking results," *Commun. ACM* **29**(3), 218–221 (1986).
24. J. E. Smith, "Characterizing computer performance with a single number," *Commun. ACM* **31**(10), 1202–1206 (1988).
25. D. J. Lilja, *Measuring Computer Performance*, pp. 24–41, Cambridge University Press, Cambridge (2005).

Veli-Tapani Peltoketo leads the R&D at Sofica Ltd. For 20 years, he has been working in several companies in the telecommunication and imaging industry. He has acted in several positions and projects producing new products and innovations. In recent years, he has led the development of automated testing systems to the imaging industry. He is a member of IEEE and IS&T.

Mobile phone camera benchmarking in low light environment

Veli-Tapani Peltoketo*

Sofica Ltd., Kampusranta 9C, 60320 Seinajoki, Finland
Vaasa University, Wolffintie 34, 65200 Vaasa, Finland

ABSTRACT

High noise values and poor signal to noise ratio are traditionally associated to the low light imaging. Still, there are several other camera quality features which may suffer from low light environment. For example, what happens to the color accuracy and resolution or how the camera speed behaves in low light? Furthermore, how low light environments affect to the camera benchmarking and which metrics are the critical ones?

The work contains standard based image quality measurements including noise, color, and resolution measurements in three different light environments: 1000, 100, and 30 lux. Moreover, camera speed measurements are done. Detailed measurement results of each quality and speed category are revealed and compared. Also a suitable benchmark algorithm is evaluated and corresponding score is calculated to find an appropriate metric which characterize the camera performance in different environments.

The result of this work introduces detailed image quality and camera speed measurements of mobile phone camera systems in three different light environments. The paper concludes how different light environments influence to the metrics and which metrics should be measured in low light environment. Finally, a benchmarking score is calculated using measurement data of each environment and mobile phone cameras are compared correspondingly.

Keywords: Low light imaging, image quality, camera speed, benchmarking

1. INTRODUCTION

High noise values and poor signal to noise ratio are traditionally associated to the low light imaging. Still, there are several other camera quality features which may suffer from low light environment. For example, what happens to the color accuracy and resolution in low light imaging and how the camera speed behaves in low light? Furthermore, how low light environments should be taken into account in the camera benchmarking and which metrics are the critical ones in low luminance?

Definitely high noise values may decrease color accuracy metrics since the noise causes wrong colored pixels to the image. Also sharpness may decay when the noise will dim sharp edges of the image. On the other hand, texture resolution measurements will be challenging in a high noise environment because the measurement algorithms may interpret noise as a high frequency texture.

The low light environment will affect especially to the focus speed of the image capturing. Obviously, exposure time will increase also. However, it is interesting to see, if the image pipeline algorithms like denoising and sharpening will increase the image capture time in low light environment and how the high noise will influence to the image compression time. In this work, low light images are captured without flash to concentrate to the low light characteristics of the sensor and camera module.

To compare mobile phone cameras in different environments easily, a single benchmarking is commonly used. The score is a combination value of measured camera metrics. However, quality and speed metrics will be affected differently in different luminance environment and thus it should be considered which ones are useful in each environment.

The work contains standard based, objective image quality measurements including noise, color and resolution measurements in three different light environments: 1000, 100, and 30 lux which represent overcast day, general indoor lighting, and dim indoor lighting correspondingly. Moreover, camera speed measurements are done and important image capture factors like exposure time and ISO speed are recorded in each environment. Detailed measurement results of

*veli-tapani.peltoketo@netikka.fi; phone +358 44 517 8552;

each quality and speed category are revealed and compared between light environments. Also a suitable benchmark algorithm is evaluated and corresponding score is calculated to find an appropriate metric which characterizes the camera performance in different environments.

Measured mobile phones are selected from different price categories and operating systems. The measurements are done using software based and automated test system which is executed towards an application programming interface of different operating systems. This approach gives comparable measurement values between different operating systems and removes the influence of mobile phone specific camera applications.

The work is a continuation to the previous papers of the author. The previous researches have been concentrated to the generic mobile phone benchmarking metrics and noise characteristics of mobile phone cameras in different light environments.

The result of this work introduces detailed image quality and camera speed measurements of mobile phone camera systems in three different light environments. The paper concludes how different light environments influence to the metrics and which metrics should be measured in low light environment. Finally, a benchmarking score is calculated using measurement data of each environment and mobile phone cameras are compared correspondingly.

2. LOW LIGHT TESTING CHARACTERISTICS

2.1 Generally

The first mobile phone cameras were published in 2000. Since that time, the quality of the camera systems has improved tremendously and quality differences between camera systems are decreased. Even if there are still differences between the mobile phone cameras, it seems that new and more demanding environments are required to solve true disparities between camera models. A low light environment is very suitable one because it is commonly used when still images are captured¹ and especially in case of mobile phone cameras^{2,3}. Moreover, the low light environment is very stressful environment to the mobile phone cameras due to small sensor size.

There are some good research and standards of low light camera performance. ANSI/CEA-639⁴ defines very straight forward metrics for video camera in low light environment and a novel research by Wueller⁵ defines additions for the still imaging performance. Both papers are based on method, where low light metrics are compared to the reference values, i.e. measurements, which are measured using, for example, 1000 lux luminance. When a low light metric reaches a certain threshold, corresponding luminance value is recorded as a low light performance of the camera system. However, the corresponding standard is currently considered in the working group of ISO technical committee 42 and there are not any official metrics ready.

2.2 Noise

The small sensor size of modern mobile phone cameras causes several different challenges to the low light imaging. The small sensor size with a great pixel amount means small pixel size which requires a lot of photons to get enough information from the imaging scene. This leads to long exposure times and forces to increase analog and digital gains which increase the noise in the image significantly. Moreover, the long exposure time increases the possibility of blur in the image due to hand shake issues and the probability to have a motion blur in the image.

Even if the noise is only one of the quality artifacts of the digital imaging, it may cause several different problems to the imaging and therefore it should be considered as one of the dominant issues especially in low light imaging.

Noise may cause direct or indirect artifacts to the image. Noise influences straightly to the quality of the image by blurring the image and decreasing the contrast of the image. When the image has a low contrast, it may cause problems for example to the auto-focus algorithms and generate extra delays to the image capturing. On the other hand, too aggressive noise removal may corrupt the texture of the image and thus generate indirect anomalies to the image.

The noise metrics of this paper follows the previous research of the author⁶. Signal to Noise Ratio (SNR) and visual noise are calculated according to the latest ISO 15739 standard⁷. Moreover, the important factors of the noise; ISO speed and exposure time are recorded in every light environment and presented in the paper.

It should be also noted that there might be a fundamental noise measurement issue with the modern digital cameras which are using the latest denoising algorithms. ISO 15739 standard defines noise measurements which are based on the noise calculation from twenty uniformly gray test patches. The latest denoising algorithms may remove noise more

efficient from homogenous areas than other areas of the image which contains, for example, different textures⁸. If these kinds of algorithms will be commonly used in the digital cameras, the ISO 15739 based noise measurements may not give real noise values of the whole image.

2.3 Image Resolution

The low light imaging with increased noise may cause several resolution and sharpness related issues to the images. Obviously, high noise will disturb the edges of the image and thus decrease the objective and perceptual sharpness of the image. Moreover, high noise may flatter the contrast and dynamic range of the image and in that way, decrease the perceptual sharpness.

Image resolution metrics of this paper are based on ISO 12233 standard⁹ and the latest texture resolution researches^{8,10-12}. It is noteworthy to mention, that the latest version of ISO 12233 standard from year 2014 defines resolution as an objective metric of the camera and sharpness as a perceptual metric of the camera. The metrics which are used in the low light benchmarking are MTF (Modulation Transfer Function) calculations of the low contrast and slanted edge charts and texture resolution from the dead leaves test chart. The detailed measurement metrics can be found from the previous research of the author⁶. However, very noise images and aggressive denoising may cause problems especially to the texture resolution measurements using the dead leaves chart.

Without noise compensation, the noise particles in the captured dead leaves area may be interpret as a high frequency texture and thus they can be measured as too good texture resolution. The effect of noise can be decreased by calculating the noise level and decreasing the noise from captured image¹² (1). The approach has been used in this paper.

$$\text{SFR}(f) = \sqrt{\frac{PS_{image}(f) - PS_{noise}(f)}{PS_{reference}(f)}} \quad (1)$$

However, the problems of this approach have been pointed by Kirk et al.⁸. The power spectrum of the noise is not calculated from the dead leaves chart but from the uniformly gray patch which has same mean brightness than the dead leaves chart. As defined in the previous chapter, some denoising algorithms may handle uniform areas differently as other areas of the image and remove noise more efficiently from the uniform areas. In this case, the effect of the noise power spectrum is too small when the texture resolution is calculated. The problem was noticed in this research.

Kirk et al. proposes to calculate the full transfer function $H(f)$ as follows (2).

$$H(f) = \frac{\phi_{xy}(f)}{\phi_{xx}(f)} \quad (2)$$

Where ϕ_{xy} is the cross correlation power density between reference and captured dead leaves area and ϕ_{xx} corresponding auto correlation power density. The algorithm seems to remove the noise component very efficiently and the results of the Kirk et al. paper are convincing ones. However, the algorithm requires full reference approach when the texture resolution is calculated. Moreover, the paper does not describe the accuracy requirement between the placements of reference dead leaves chart and dead leaves chart which is cropped from the captured image. Even if the approach is sensible, it is not yet used in the benchmarking metrics of this paper.

Finally, if the camera system uses artificial sharpening, the sharpening algorithms may skip very noisy and low contrast edges and therefore cannot highlight edges and improve perceptual sharpness of the image. On the other hand, very aggressive artificial sharpening may increase noise by detecting noise components as edges and highlight them.

Improved denoising and sharpening algorithms are very good examples how new innovations force to change the measurement metrics all the time. The algorithm development make difficult to maintain static benchmarking metrics which would also work when new digital camera generations are developed.

2.4 Color Fidelity

The most obvious color fidelity issue in low light imaging is too dark images with low contrast and therefore faulty colors. Also high noise values may influence to the color fidelity since noise causes wrong colored pixels to the image. Even if the color characteristics are not the main issues in the low light imaging, it should be investigated carefully. It

has been interesting to notice that there are very few researches which have concentrated to this area. A very recent research by Rezagholizahed et al. describes an issue which may declare this shortage¹³. The role of the light source is extremely important in low light measurement when the color fidelity is tested. The fluctuations of the photon stream may be unpredictable in low light environment and these characteristics of the light source may cause faulty illumination to the testing scene since the standardized color fidelity measurement defines the color temperature of the light sources very precisely. The research by Rezagholizahed et al. does not specify the illumination level where the fluctuation problems will start. However, it could be presumed that these kinds of anomalies will be essential much lower light environment than the lowest one in this research (30 lux).

The most obvious way to avoid low light anomalies is to use a flash. However, when the flash is used, several new measurements should be added to verify the functionality. At least the color temperature, uniformity and magnitude of the flash should be measured and also reflections from the imaging scene should be noted. The flash was not used in the research because it would require several environment specific metrics which should be taken into account.

Since the light source could be adjusted so that the light temperature does not vary too much, standardized ΔE_{00} value was measured according to the CIEDE2000 standard¹⁴. Also the mean saturation error was calculated from the gray patches of the ISO 15739 chart.

2.5 Camera Speed

The camera speed measurements in different light environments are very interesting ones. There are several camera parameters and algorithms which are influencing to the cameras speed. Firstly, a camera adjusts the exposure time and ISO speed to ensure the lightness of the captured image. Camera manufacturers have clearly different approaches to search the balance between exposure time and ISO speed. Secondly, lack of ambient light complicates the functionality of the auto-focus algorithm: A flat contrast characteristics and noise of the captured image together with long exposure time affect to the speed of the focus algorithms¹⁵. Thirdly, the low light environment may require more time to the image processing pipeline to make denoising, sharpening and compression functionalities.

To investigate closely the characteristics of the camera in different light environments, the image capturing time was measured step by step and each component of the time was revealed. The image capturing time includes following components: startup time, focus time, exposure time and image processing time. The image processing time contains all functionalities which are done after the auto-focus acceptance and exposure of the image. It has to be noted that the ISO speed and exposure time are not measured but they are fetched from the metadata of the captured images and are based on the information of device manufacturers.

The total time of a single image capturing and time of five consecutive images were used to camera speed related benchmarking metrics. The five image time does not contain startup and focus time.

3. MEASUREMENT METHODOLOGY

The low light measurements are done in the same environment as the previous paper of the author⁶. The measurements are based on automatic measurement system which is executed towards application programming interface (API) of the devices. All speed measurements, camera configuration and image capturing are done using the test automation. It is notable, that all quality measurements are based on JPEG (Joint Photographic Experts Group) compressed images because only few mobile phone camera supports raw-format images.

A separate imaging laboratory is used for speed and image quality measurements. The quality measurements are based on standardized testing charts which are located to the testing scene and the scene is illuminated using high quality lights. Different light conditions are built up by decreasing active components of the lights and using diffusors. The illumination uniformity over the whole scene is less than $\pm 5\%$ in each light environment and the light temperature differences between different light environments is below 200 K.

The background of the measurement scene is 18% neutral matt grey and following test charts are mounted to the scene:

- 20 grey patches to calculate ISO 14524 OECF curve¹⁶ and ISO 15739 noise.
- Macbeth color chart for CIEDE2000 color accuracy measurements.
- Low contrast slanted edge charts in the middle and each corner for ISO 12233 resolution measurements. 5% angle and 4:1 contrast are used.

- Colored dead leaves chart to texture resolution measurement and to detect denoising and sharpening defects.

Figure 1 shows the measurement scene and details of the scene in each light environment.

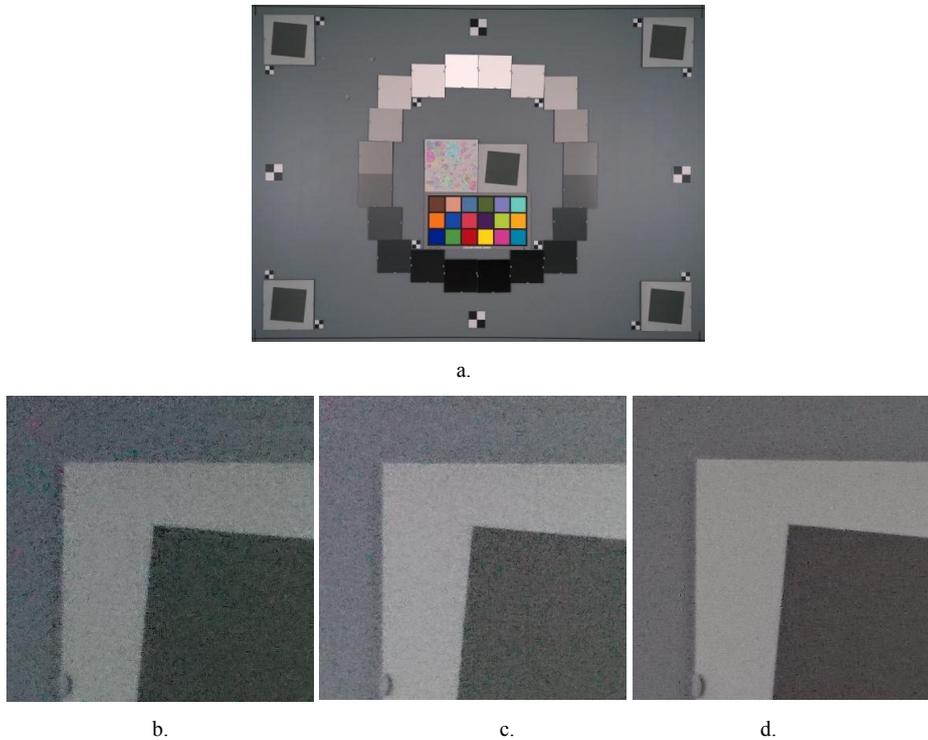


Figure 1. a. Measurement scene, upper left corner using b. 30 lux, c. 100 lux and d. 1000 lux

3.1 Different Light Environments and Single Score

As mentioned before, a proper light environment does not necessarily solve all differences between devices and more stressful environments are needed. This requirement applies another issue; which low light metrics should be added to the benchmarking score without making the benchmarking too complicated and difficult to interpret? Or is it more practical to calculate own low light benchmarking score?

It would be tempting to select only one or two low light measurement metrics to represent the low light performance of the devices and add them to the original benchmarking score. The method maintains the user friendly, single score benchmarking.

The low light speed measurements are supporting this approach. The rank correlation using speed score between 30 lux and 1000 lux environment is 0.92 and corresponding value between 100 lux and 1000 lux as high as 0.95. This means that the speed performance in different light environments does not change the benchmarking order of the devices significantly.

However, when the low light image quality metrics are investigated, it is noted that any subset of the quality metrics does not individually represent the low light quality performance of the devices. Even if the noise is the dominant factor which influences to the other quality features, different image quality pipelines reacts in various ways to the noise and for example the resolution results do not follow the noise amount in the image. To keep the benchmarking score as straightforward as possible, the same score equation is used in every light environments (3).

$$\text{Score} = \left(\prod_{i=1}^n a_i \right)^{1/n} \quad (3)$$

Three different scores are calculated; speed score including single image capture time and time of five consecutive images, quality score including edge resolution, texture resolution, color error, saturation error, SNR, and visual noise, and finally benchmarking score including all above mentioned metrics. Edge resolution, texture resolution, and SNR values are used as such and rest of the values are used as reciprocal values. Differ from the earlier paper, audio/video synchronization measurements are not done in the research because the audio/video testing environment is not suitable for low light measurements.

4. RESULTS

Nineteen different mobile phone cameras are measured and detailed results are revealed. The devices are using even Android or Windows Phone operating system. Every device has a CMOS (complementary metal oxide semiconductor) camera sensor.

The sensor sizes of the cameras are more than or equal to 8 megapixels with one exception: HTC One with UltraPixel sensor which contains 4 megapixels. Obviously, each camera has autofocus, autoexposure, and auto white balance functionalities. All measurements are done using default settings and maximum resolution of the camera. Majority of the devices has 4:3 aspect ratio when maximum resolution is used, only Samsung Galaxy S5 and HTC One use 16:9 ratio.

Results are divided to three categories, image quality results and camera speed results in three different light environments and corresponding benchmarking result.

4.1 Image Quality in Different Light Environments

The image quality results include two metrics of each main quality areas: noise, resolution, and color fidelity.

Figures 2a and 2b show very expected values of SNR and visual noise. The SNR values decrease when the light environment is getting worse. Also visual noise follows quite well the light environment changes. It has to be noted that the total noise value is a complex mixture of several noise sources of lens system, sensor and image processing pipeline. The noise amount can be partially controlled by camera parameters like ISO speed and exposure time and denoising algorithms. However, the separation of different noise sources from the total noise value is a difficult task and requires much detailed research.

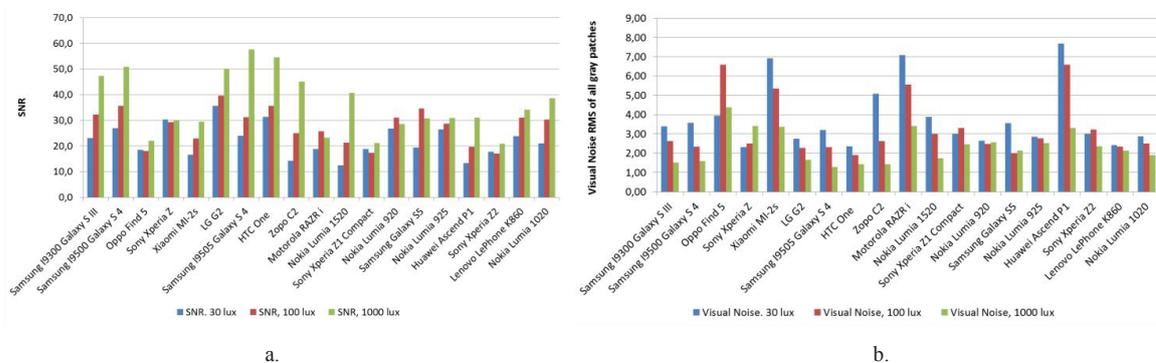


Figure 3 defines the results of the resolution measurement. As figure 3a shows, the slanted edge based resolution results are quite expectable, the resolution increases when the light environment is improved. HTC One is the most immune device to the light changes due to big pixel size. On the other hand, the lowest pixel count decreases the resolution towards other devices.

The high noise problem of dead leaves measurement⁶ can be seen in some devices in figure 3b. Especially Zopo C2 and Huawei Ascend P1 have clearly better texture resolution values in low light than in proper light conditions. The corresponding SNR values of the devices also declare very high noise values in the low light. It seems that the current dead leaves algorithm is not valid with high noise values.

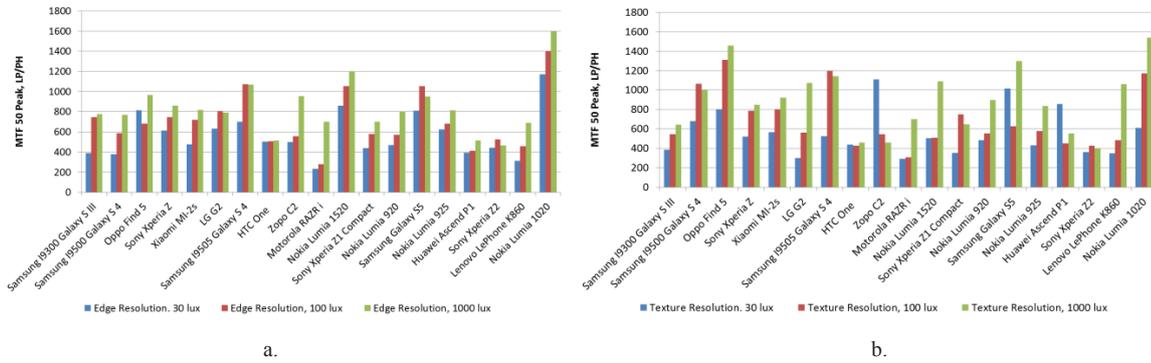


Figure 3. a. Slated edge based resolution and b. texture resolution in different light environments

The color fidelity results are quite controversial; there does not seem to be a clear correlation between color error, saturation error and light environments.

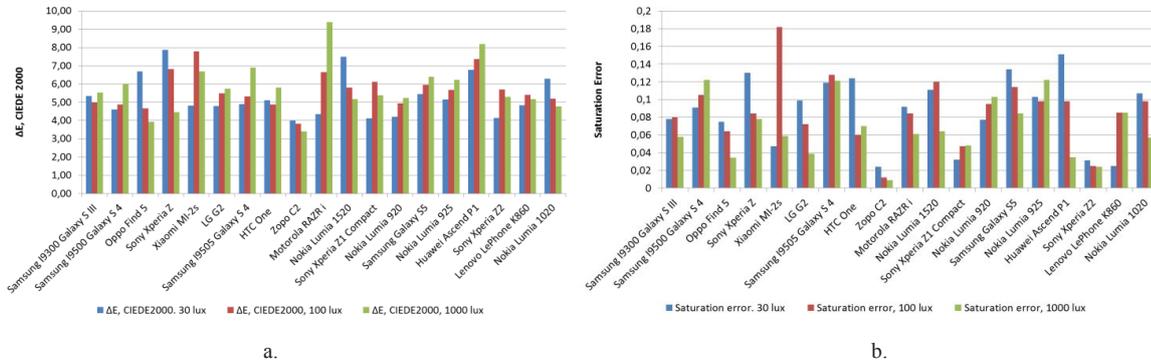


Figure 4. a. Color error and b. saturation error in different light environments

4.2 Camera Speed in Different Light Environments

The main adjustable factors of different light conditions, ISO speed and exposure time are shown in figure 5a and 5b correspondingly. The combinations between factors vary significantly between devices. Even if handshaking issues and motion blur are not measured in the research, they are significant problems when a long exposure time is used. The old thumb rule defines the relation between focal length and exposure time: the exposure time should not be longer than the reciprocal of the focal length, otherwise the handshaking will affect to the sharpness of the image. In case of 30 and 100 lux measurements almost all devices have too long exposure time. On the other hand, motion detectors which informs that camera is well mounted and different image stabilization algorithms may increase the acceptable exposure time.

Generally, the exposure time is not a significant factor when the total capture time is measured. Only Oppo Find 5 has such a long exposure time which affects to the total image capture time.

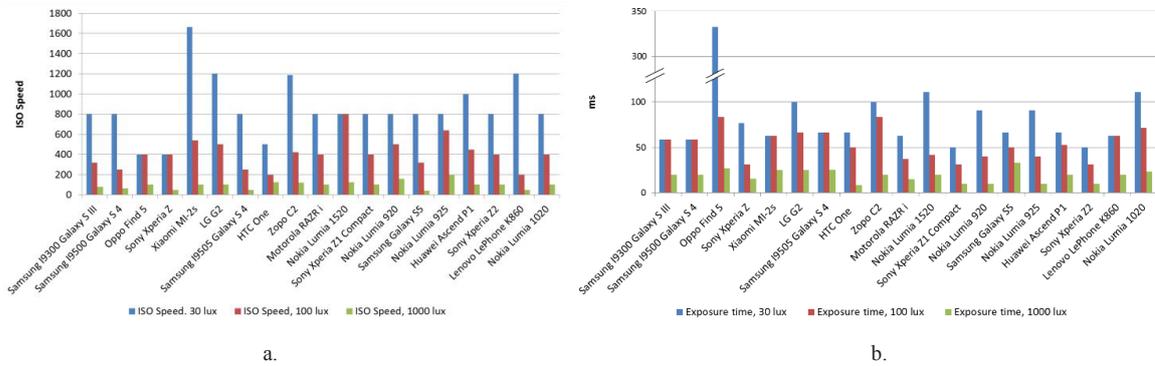


Figure 5. a. ISO speed and b. exposure time in different light environments

Image capture times are separated to startup, focus, and image processing time. It is quite expectable that light environment do not affect to the startup time as figure 6a defines. However, it is quite surprising that the light changes affect only slightly to the image processing time in figure 6c. Figure 6b shows that the focus time is the dominant metric when differences between light environments are investigated. Finally, figure 6d shows the total image time which is the summation of figures 6a-6c and 5b.

It can be seen from figure 6d that generally the light environment does not affect to the order between devices and one light environment would be enough to solve the speed benchmarking. However, there are exceptions like Oppo Find 5 which low light speed performance is poor and decreases the corresponding speed benchmarking rank significantly.

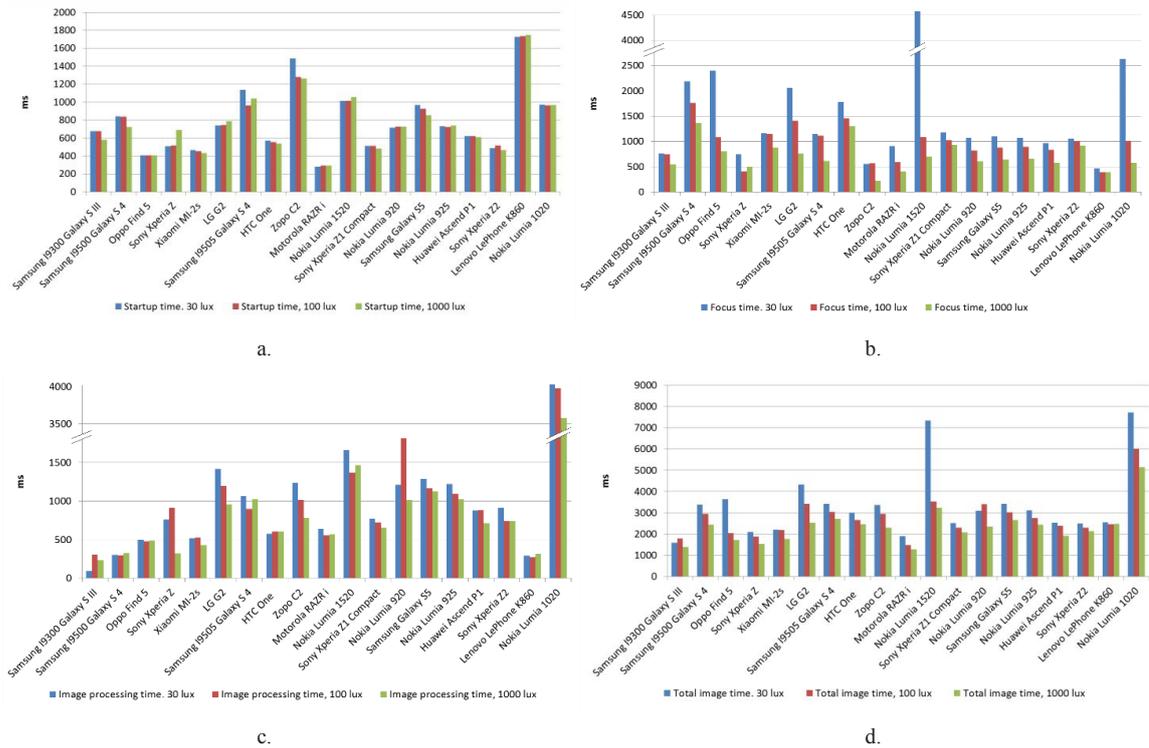


Figure 6. a. Camera startup time, b. focus time, c. image processing time, and d. total time in different light environments

4.3 Camera Benchmarking in Different Light Environments

Speed, quality and total benchmarking score were calculated from the speed and quality metrics. Benchmarking scores of 30 and 1000 lux are used to highlight the trends between the light environments. Score values of the 100 lux environment fit quite linearly between 30 and 1000 lux values. Each benchmarking figure is sorted by the 1000 lux values of the corresponding score which declares the different ranking between scores.

Figure 7a shows the speed metric of 30 and 1000 lux environments. Without few exceptions, the order of 1000 lux scores and 30 lux scores are very similar. The correlation between the ranks of the devices is as high as 0.92. Clearly, the light environment do not influence significantly to the benchmarking order.

On the other hand, the quality score in figure 7b behaves differently. The trend between 1000 lux and 30 lux environment are not similar and corresponding rank correlation is 0.31. Different lens systems, sensors and image processing pipelines react differently to the low light environment challenges.

Finally, figure 7c shows the total benchmarking score which combines speed and quality metrics. According to the used benchmarking algorithm, the 30 lux rank differs significantly from the 1000 lux values.

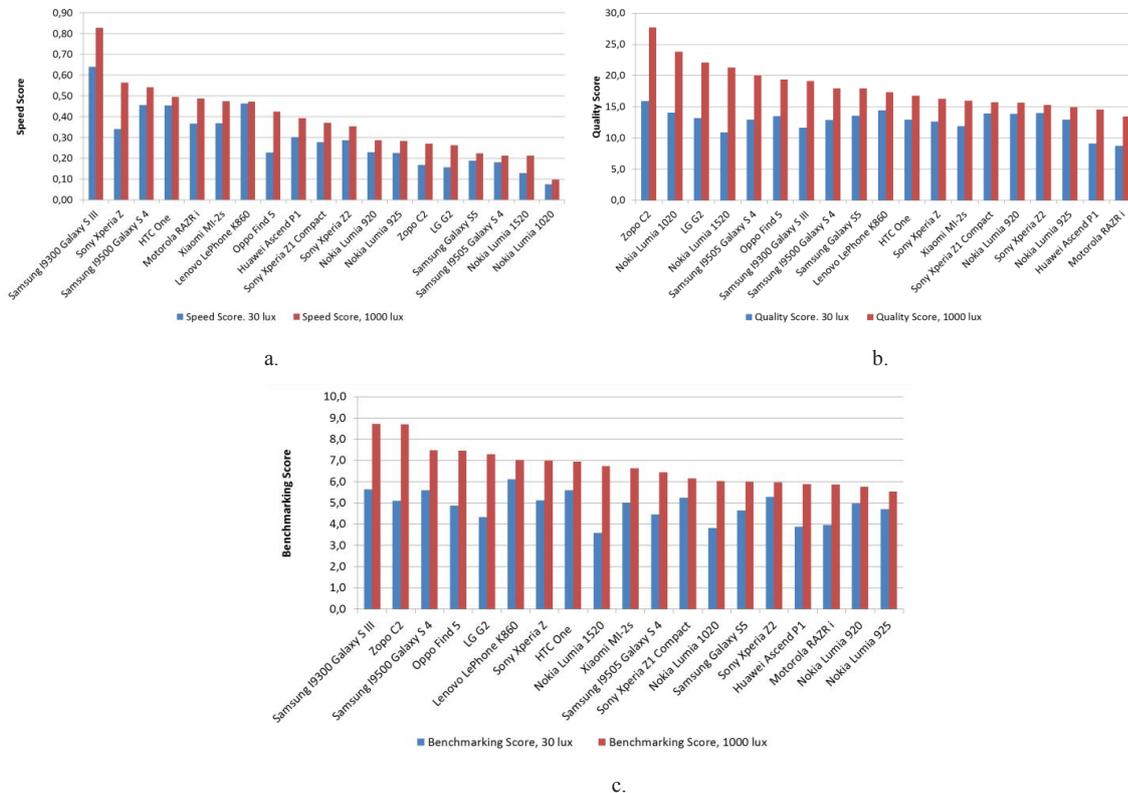


Figure 7. a. Speed score, b. quality score and c. benchmarking score in different light environments

5. CONCLUSIONS

The research revealed several factors from the low light measurements and benchmarking. Firstly, the light environment does not influence significantly to the speed based benchmarking order of the devices. The speed benchmarking in 1000 lux environment correlates very well to the 100 lux and 30 lux environments. Moreover, the focus time seems to be the dominant metric when camera speed differences between light environments are investigated. Secondly, the texture measurement issue reported by Kirk et al.⁸ is a real problem and the content based denoising algorithms may decrease the reliability of the ISO 15739 noise measurement, too.

Thirdly, even if the speed benchmarking seems to be quite static in different light environments, the quality benchmarking reacts differently. The quality benchmarking does not correlate between light environments. To keep the benchmarking score as simple and straightforward as possible, a separate benchmarking score was measured to the low light imaging.

Finally, it seems to be quite problematic to create a static benchmarking score with static metrics because new innovations and algorithms in camera industry force to develop new image quality measurements all the time. The sharpening and denoising algorithms are very good examples of this kind of evolution.

REFERENCES

- [1] Keelan, B. W., [Handbook of Image Quality: Characterization and Prediction], CRC Press, New York, 398-399 (2002).
- [2] International Imaging Industry Association, "Camera Phone Image Quality – Phase 1 White Paper, Fundamentals and review of considered test methods," (2007).
- [3] Hultgren, O. and Hertel, D.W., "Megapixel mythology and photospace: estimating photospace for camera phones from large image sets," Proc. SPIE 6808, (2008).
- [4] Consumer Electronics Association, "ANSI/CEA-639 Consumer Camcorder or Video Camera Low Light Performance," (2010).
- [5] Wueller, D., "Low Light Performance of Digital Still Cameras," Proc. SPIE 8667, (2013).
- [6] Peltoketo, V-T., "Evaluation of mobile phone camera benchmarking using objective camera speed and image quality metrics," J. Electron. Imaging, 061102, (2014).
- [7] International Organization of Standardization, "ISO 15739 Photography – Electronic still-picture imaging – Noise measurements," (2013).
- [8] Kirk, L., Herzer, P., Artmann, U. and Kunz, D., "Description of texture loss using the dead leaves target: Current issues and a new intrinsic approach," Proc. SPIE 9023, (2014).
- [9] International Organization of Standardization, "ISO 12233 Photography — Electronic still-picture cameras — Resolution measurements," (2014)
- [10] McElvain, J., Campbell, S. P., Miller, J. and Jin, E. W., "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems," Proc. SPIE 7537, (2010).
- [11] Burns, P. D. and Williams, D. "Measurement of Texture Loss for JPEG 2000 Compression," Proc. SPIE 8293, (2012).
- [12] Artmann, U. and Wueller, D. "Improving texture loss measurement: spatial frequency response based on a colored target," Proc. SPIE 8293, (2012).
- [13] Rezagholizadeh, M. "Photon Detection and Color Perception at Low Light Levels," Canadian Conference on Computer and Robot Vision (CRV), 283–290 (2014).
- [14] International Commission of Illumination, "CIE S 014-6/E:2013 Colorimetry – Part6: CIEDE2000 Colour-Difference Formula" (2013).
- [15] Gamaida, M., Kehtarnavaz, N. and Roberts-Hoffman, K., "Low-Light Auto-Focus Enhancement for Digital and Cell-Phone Camera Imaging Pipelines," IEEE Transactions on Consumer Electronics, Vol.53, 249-257, (2007).
- [16] International Organization of Standardization, "ISO 14524 Photography—electronic still-picture cameras— methods for measuring OECFs," (2009).

Signal to Noise Ratio and Visual Noise of Mobile Phone Cameras

Veli-Tapani Peltoketo[▲]

University of Vaasa, Faculty of Technology, Wolffintie 34, Vaasa 65200, Finland

Sofica Ltd, Kampusranta 9C, Seinäjoki 60320, Finland

E-mail: veli-tapani.peltoketo@netikka.fi

Abstract. Recently, ISO standard 15739:2013 revealed the first official visual noise metrics. Until now, signal to noise ratio (SNR) has been used the most as the noise measurement metric for digital cameras, but according to several research studies it does not represent the visual perception of noise included in images. This article investigates the differences between SNR based noise measurements and visual noise measurements when real mobile phone cameras are measured. The work contains the following tasks. Firstly, the improvements between old and new standards are detailed. Secondly, the noise measurements are executed, testing 20 mobile phones using three different light environments. Finally, the results are compared between different noise measurement algorithms and conclusions are drawn. The result of this work gives detailed noise measurement results for mobile phone camera systems on the market. Total and visual noise metrics are measured, also corresponding ISO speed and exposure parameters are stored and correlations between noise levels and parameters are calculated. The differences between noise metrics are summarized and perceptual inspection is made of the images which have clear differences between total and visual noise. © 2015 Society for Imaging Science and Technology.

[DOI: 10.2352/J.ImagingSci.Technol.2015.59.1.010401]

INTRODUCTION

Nowadays, mobile phone cameras modify captured images in a very comprehensive way. Just to mention a few, the image processing pipeline may contain denoising, sharpening, antialiasing and lens aberration correction algorithms. These algorithms and their combinations may cause very unpredictable artifacts in image quality. Image quality standardization tries to follow the latest trends in digital image technology, and thus the standardization of digital imaging has been recently updated very intensively. Standards like ISO 12233¹ (resolution), ISO 12232² (ISO speed), ISO 14524³ (Opto-Electronic Conversion Function) and ISO 11664-6⁴ (color difference) have been updated or were in the review stage during the year 2014. Noise standard ISO 15739 is no exception here. The latest version is from the year 2013⁵ which replaces the older one from the year 2003.⁶

Until now, signal to noise ratio (SNR) has been used most as the noise measurement metric for digital cameras.

However, SNR measurement does not necessarily represent the visual perception of the noise present in the images; images with exactly the same SNR values can be visually very different.⁷ Moreover, when the noise frequency is low, the SNR measurement may give values that are too good. Also luminance and chromatic noise differ when they are detected by human eyes and this phenomenon cannot be measured using pure SNR values.

There are several research approaches to measure the visual noise. The ISO 15739 and the S-CIELAB⁸ visual noise methods use quite similar logic. Both use opponent color space, frequency domain and specific filter which weights different noise frequencies. However, the final visual noise calculation differs; whereas ISO 15739 calculates the standard deviation from $L^*u^*v^*$ color space, the S-CIELAB version calculates color difference (ΔE) from $L^*a^*b^*$ color space.

Furthermore, perceptually calibrated and just noticeable difference (JND) based visual noise measurements were first published by Kuang et al.⁹ and based on the research by Keelan et al.¹⁰ The papers are based on variances calculated from $L^*a^*b^*$ color space. Measured noise values are validated by observers and noise equations calibrated so that they follow JND based quality loss metrics.

This article concentrates on the noise algorithms of the ISO 15739 standard, the changes between versions 2003 and 2013 and how they affect noise measurements. Moreover, the relationship between SNR based noise and visual noise is evaluated by measuring mobile phone cameras currently on the market.

Even though the visual noise measurement method is the most notable update to ISO 15739:2013, there are several small changes which affect the measurement, calculation and results of the noise. The influences of the changes are also described in this article.

There are several tasks in this work. Firstly, the characteristics of the old and new standard and visual noise measurements are defined and the most significant updates are detailed. Secondly, the corresponding noise measurements are executed, testing mobile phones which represent models from flagships down to mid priced devices. There are 20 devices which are measured in three different light conditions: 1000, 100 and 30 lx. Finally, the results are

[▲] IS&T Member.

Received June 22, 2014; accepted for publication Mar. 6, 2015; published online Apr. 30, 2015. Associate Editor: Zeev Zalevsky.

1062-3701/2015/59(1)/010401/7/\$25.00

compared between different noise measurement algorithms and conclusions are drawn.

The result for this work gives detailed noise measurement results for mobile phone camera systems on the market. Total and visual noise metrics are measured, also corresponding ISO speed and exposure parameters are stored and correlations between noise levels and parameters are calculated. The differences between noise metrics are summarized and perceptual inspection is made of the images which have clear differences between total and visual noise.

THE EVOLUTION OF THE ISO 15739 STANDARD

The most notable update of ISO 15739:2013 is the official visual noise measurement method. There has been a lot of discussion and research relating to perceptual noise measurement and the latest ISO 15739 standard is the first official version of visual noise measurement. However, the corresponding research is ongoing and there is a great possibility that the visual noise part of the standard will change in the near future.

Nevertheless, the visual noise is not the only update to the ISO 15739 standard. There are several relatively small changes which affect the environment, measurement, calculation and results of the noise. This chapter defines the changes between the old and the new versions.

Environment and Methods

There are several changes in the environment definitions between the standard versions. Some of them seem to be minor but their influence on the measurement practices can be radical.

The most visible changes are to the testing chart itself. The new standard recommends using a 20 patch version of the OECF test chart whereas the older version defines a test chart with 12 different patches. Moreover, the new standard defines the background value of the test chart between 110 and 130 in the case of 8 bit standard red green blue (sRGB)¹¹ encoded signals. Two threshold values give more freedom to adjust the exposure time than the previous version which strictly defines the background value as 118. The new version also defines that the luminance uniformity of the whole test chart should not vary more than $\pm 2\%$, which is a significantly strict requirement.

Furthermore, the new version allows slight defocusing of the captured image, if the accuracy of the test chart is too low. Weak test chart accuracy can be a real problem nowadays because it may reveal the printing raster to cameras which have high resolution sensors. Defocusing may be needed, particularly when the standard requires the test chart spatial frequency to be at least ten times higher than the limiting resolution of the camera.

Another technology driven change is the movement requirement change between captured images. The noise measurement algorithm requires capturing at least eight different images to ensure correct noise measurements. The older version of the standard gives a very strict requirement for the movement between images: the difference between

eight images should be less than a quarter of a pixel. This requirement was technically feasible before optical image stabilization (OIS) technology was on the market. However, the latest mobile phone cameras may contain OIS, and even if the camera is well mounted the OIS may cause a variance between consecutive exposures. With OIS, the quarter pixel requirement can be fulfilled only by capturing a huge quantity of images and selecting those which are within the threshold, which is not a very practical technique. The change is reasonable particularly because the noise calculation does not contain this strict requirement. The requirement of the old version of the standard was intended to detect the fixed pattern noise of the sensor, but camera movement does not affect the fixed pattern noise. It is interesting to note that the new version of the standard has not reduced the movement requirement but completely removed it.

Finally, the new standard defines informative recommendations for practical viewing conditions for different output media like photo print, computer display, mobile phone display and HDTV display. These were missing in the older version of the standard. On the other hand, an informative annex of a method for measuring edge noise has been removed from the latest version.

SNR

Obviously, SNR contains two elements: signal and noise. The noise equations are very similar; fixed pattern noise, temporal noise and total noise calculations are equivalent to the former version of the standard. However, there are changes in the signal element, the SNR measurement itself and also in the environmental requirements section.

Firstly, the reference luminance value is decreased from 255 to 245. Moreover, the constant which defines the target luminance at which the SNR measurement is made is decreased from 18% to 13%. The change compensates for the removal of the 140% underexposure rule which exists in the previous standard and, in practice, the percentage change does not affect the target luminance value.

The most significant change is the SNR measurement method update. In the older version of the standard the SNR was measured only from the three density patches of the test chart where the middlemost patch has density 0.9. This required only one incremental gamma value measurement using the three signal values of the three patches. Furthermore, only one SNR value is obtained. The newest standard version does not define the exact density of patches but the reference luminance at which the SNR measurement is made. In practice, this means that several SNR values have to be calculated and the SNR value which matches 13% of the reference luminance is selected. If there is no patch which equals the required luminance, an interpolation can be made between SNR values of surrounding patches.

The new SNR method is less sensitive to the differences between how cameras reveal image highlights. Cameras use exposure time and tone mapping to display or hide image details, especially in low light, and highlight parts

of the image. Nowadays, also more aggressive highlight compression is used by HDR or WDR (high dynamic range/wide dynamic range) algorithms where details of dark and bright areas are exposed. The complexities of WDR camera SNR measurements are defined in the research by Hertel.¹²

Finally, Annex C of the new version of the standard defines a high pass filter which eliminates low frequency variations of the tested camera. Usually, this means lens shading artifacts. If the low frequency variations cause concern, the high pass filter is only recommended for the fixed, temporal and total noise measurements and should not be used for the visual noise measurements.

Dynamic Range

Even if the equations of the dynamic range chapter are changed, a closer look reveals that the principles of the dynamic range algorithm are the same as in the previous version of the standard. The latest version is more straightforward, and reflectance constants of the previous version of the standard are removed.

It is notable that all mention of the 140% underexposure rule of digital cameras is removed from the latest version of the standard. This simplifies the understanding of the standard.

Visual Noise

The most recent version of ISO standard 15739 defines the first normative version of visual noise measurements. There was a visual noise annex in the older version, but it was labeled as an informative one and a number of critical factors were missing.

Visual noise measurements mimic the human visual system. They are based on three main components. Firstly, usage of opponent color space which includes luminance, green-red and blue-yellow color channels. Secondly, evaluation of the opponent color channels in the frequency domain and filtering the channels by contrast sensitivity function (CSF). Finally, conversion back to the spatial domain and color conversion to the $L^*u^*v^*$ color space. The visual noise is calculated using the standard deviation of each gray patch of the test chart.

The latest version of the standard collects all required color transformation matrices and corresponding constants from different standards, which simplifies the visual noise calculations. Also the most critical entity, the CSF, is specified in the new version while it was not detailed at all in the 2003 version. The CSF is mainly based on the work of Jonson and Fairchild,¹³ whereas Kelly and Keelan¹⁴ proposed weights for measuring the chrominance channel noise levels of the opponent color space. However, there is still a lot of discussion about the CSF curve and particularly how the clipping problems of luminance noise should be handled.^{7,15}

The visual noise chapters of ISO 15739:2013 include a very informative algorithm definition and specific step-by-step instructions, which give good guidelines on how to calculate the visual noise values. Also exact requirements

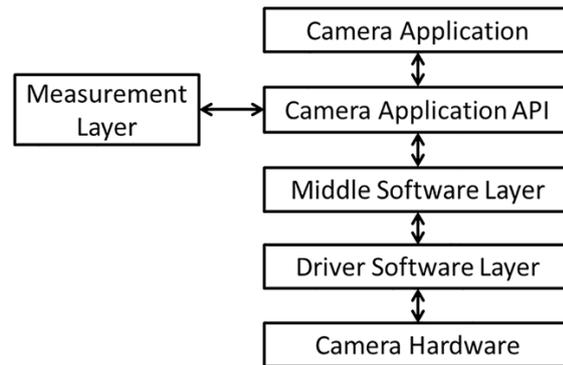


Figure 1. The measurement point of the camera software stack.

of the visual noise results streamline the output of the measurements.

MEASUREMENT METHODOLOGY

The noise measurements are made using a software based and automated measurement system which is executed toward an application programming interface (API) on different operating systems as shown in Figure 1. The camera control and image capturing is made via the API. Obviously, the captured images are used to measure the noise values. The method has been previously used to benchmark mobile device cameras.¹⁶

It is notable that all noise measurements are based on Joint Photographic Experts Group (JPEG) compressed images because only a few mobile phone cameras support raw format images. However, the measurements are comparable because JPEG images are used in every measured device.

Measurement Environment

A separate imaging laboratory is used for the noise measurements. The same environment has been previously used to benchmark the mobile device cameras.¹⁶ The noise measurements are based on ISO 15739:2013 testing charts which are located at the testing scene and the scene is illuminated using high quality lights. The illumination uniformity between gray patches is measured before measurements and the uniformity error is less than $\pm 3\%$. This value slightly exceeds the standard requirement. In this work, the measurements are made using three different illumination environments, 1000, 100, and 30 lx, which correspondingly represent overcast day, general indoor lighting, and dim indoor lighting.

The background of the scene is 18% neutral matte gray and the following test charts are mounted onto the scene.

- 20 gray patches to calculate ISO 14524 OECF curve and ISO 15739 noise. The charts are located circularly around the middle area of the scene.
- The scene contains also a Macbeth color chart for color accuracy measurements, low contrast slanted edge

Peltoketo: Signal to noise ratio and visual noise of mobile phone cameras



Figure 2. The testing scene.

charts for sharpness measurements and a so called dead leaves chart for texture sharpness measurement, but these are not used in the research.

Figure 2 shows the entire testing scene and locations of the testing charts.

RESULTS

Twenty different devices are measured. The devices are selected from three different operating systems and they represent models from flagships down to mid priced devices from twelve mobile phone manufacturers. The sensor sizes of the cameras are four mega pixels or higher. Obviously, each camera has auto focus, auto exposure and auto white balance functionalities. All measurements are made using default settings and using the maximum resolution of the camera.

Measured Values

SNR based noise metrics i.e., fixed noise, temporal noise and total noise values are measured where the total noise is a combination of the fixed and the temporal noise. The noise values are measured from at least eight images and every patch (20) has its own values. Moreover, average noise values of all patches are calculated.

In the case of visual noise, only one image is needed for measurements. To use the same image material as SNR based noise calculations, a mean visual noise value of each patch is calculated from the same images used in the SNR based noise measurements. Average sRGB pixel values and lightness values of patches are calculated as well as output pixel size. Also, the average visual noise value of all patches is calculated.

The main parameters controlling the low light imaging are ISO speed and exposure time. Both significantly affect the noise because longer exposure time increases shot noise and higher ISO speed affects analog and digital gain of the imaging and increases noise. To highlight the dependence of the parameters on noise values, exposure time used and ISO

speed of devices are also detected in all light environments and correlations are calculated between the parameters and noise levels. It has to be noted that the ISO speed and the exposure time are not measured but they are obtained from the metadata of the captured images and are based on the information of device manufacturers.

Noise Results

Obviously, both SNR based noise and visual noise values increase when light conditions are getting poorer, as Table I shows, because devices increase exposure time and ISO speed in low light conditions which increases the noise levels. However, the noise removal algorithms may affect this rule, and there is a clear example of this phenomenon. For example, all Sony devices have lower noise values in 30 lx than 100 lx and the Sony Xperia Z has the highest noise values in the best light. All in all, the Sony Xperia Z2 and the Sony Xperia Z1 Compact have very similar image quality characteristics, which may reveal equivalent or identical camera module and image signal pipeline construction.

Generally, SNR based noise has dominant values in good light conditions, whereas visual noise values are higher in low light and, especially in good light conditions, the majority of the devices have almost equal SNR based noise and visual noise values. This is quite obvious because the noise and noise removal characteristics of different devices should differ most in a low light environment. When the corresponding values are compared in low light conditions, more deviation can be observed as well as some larger differences, but still the SNR based noise and the visual noise follow each other surprisingly well. According to the pure numeric values, it might be considered that even the visual noise does not sufficiently reveal the perceptual noise quality, or that the original SNR based noise is already good enough to mimic the noise detection of human eyes.

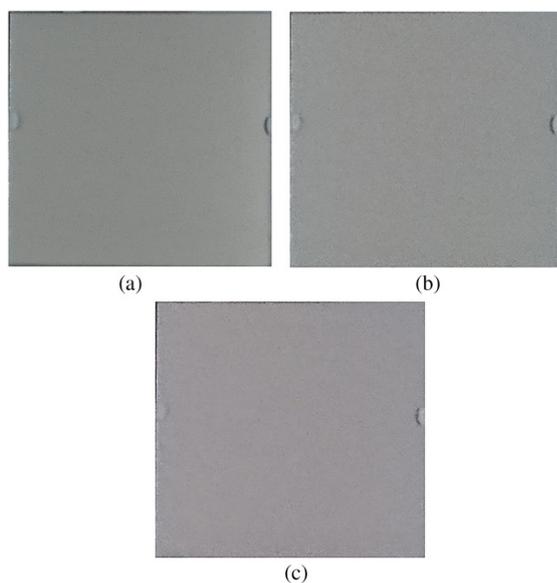
Even if the noise levels seem to be almost equal, there are still some clear exceptions in the 100 and 30 lx environments where the visual noise and total noise differ significantly. The Oppo Find 5, Xiaomi MI2, Motorola RAZR i and Huawei Ascend P1 have distinctly higher visual noise than total noise.

Figures 3 and 4 show a detailed view of tested images where the rightmost gray patch of the testing scene is represented in all three illumination environments. The example comparison is made between the Xiaomi MI2 and the Sony Xperia Z2 since the Xiaomi has better total noise values in 1000 and 100 lx and almost equal value in 30 lx, but in the case of visual noise the Sony Xperia Z2 has clearly better values. Even if the image cropping may decrease the visual differences, it is clear that the Xiaomi MI2 has significantly more chromatic noise than the Sony Xperia Z2 and in particular the structure of the noise is larger, which affects the visual noise value because of the CSF. Perceptually, the quality of the Xiaomi MI2 image is clearly poorer than the Sony Xperia Z2 one, which points out the importance of the visual noise. Also other devices with high visual noise values have the same kind of perceptual noise behavior.

Peltoketo: Signal to noise ratio and visual noise of mobile phone cameras

Table I. Noise values in different light environments.

Device	Avg. visual noise 1000 lx	Avg. total noise 1000 lx	Avg. visual noise 100 lx	Avg. total noise 100 lx	Avg. visual noise 30 lx	Avg. total noise 30 lx
Samsung Galaxy S3	1.65	1.79	2.56	2.25	3.29	2.75
Sony Xperia Z	3.40	2.33	2.51	2.25	2.31	2.14
iPhone 5	2.36	2.00	3.52	3.05	4.63	3.84
Oppo Find 5	4.43	3.38	6.52	3.74	3.88	3.42
Xiaomi MI2	3.31	2.47	5.18	2.81	6.65	3.90
Samsung Galaxy S4 9500	1.68	1.83	2.36	2.33	3.59	3.10
HTC One	1.44	1.56	2.00	2.17	2.41	2.34
Samsung Galaxy S5	2.13	2.57	2.00	2.37	3.41	3.54
LG G2	1.72	1.80	2.27	2.18	2.70	2.35
Samsung Galaxy S4 9505	1.33	1.47	2.25	2.27	3.16	3.02
Zopo C2	1.51	1.79	2.60	2.89	4.99	4.76
Sony Xperia Z2	2.40	3.12	3.17	4.00	2.97	3.70
Nokia Lumia 925	2.59	2.71	2.75	2.94	2.82	2.98
Motorola RAZR i	3.47	1.94	5.59	2.60	7.12	3.32
Sony Xperia Z1 Compact	2.48	3.00	3.26	3.90	2.97	3.42
Huawei Ascend P1	3.17	2.30	6.49	3.43	7.65	3.95
Nokia Lumia 920	2.82	3.01	2.57	2.65	2.76	2.82
Lenovo K860	2.25	2.40	2.41	2.76	2.36	3.45
Nokia Lumia 1520	1.83	1.97	3.04	3.62	3.86	5.39
Nokia Lumia 1020	2.02	2.10	2.52	2.63	2.92	3.43

**Figure 3.** Detail of the testing scene, Sony Xperia Z2: image in (a) 1000 lx, (b) 100 lx and (c) 30 lx.

The blue tone of the Xiaomi MI2 100 lx image reveals a white balance problem in the illumination environment but does not affect noise results.

Exposure time and ISO speed influence the noise significantly. Devices control the luminance of the captured

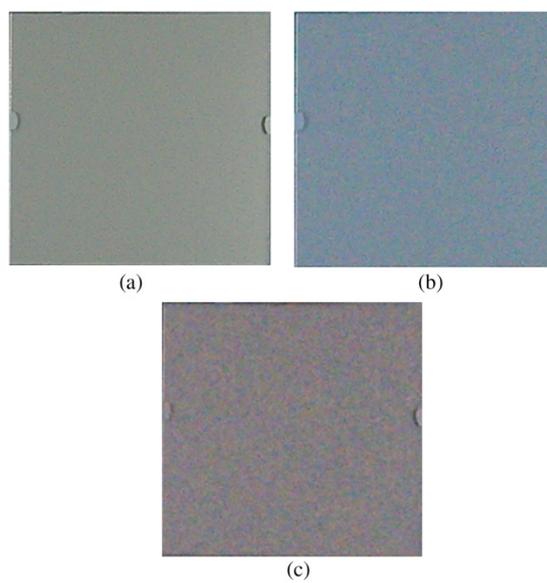
**Figure 4.** Detail of the testing scene, Xiaomi MI2: image in (a) 1000 lx, (b) 100 lx and (c) 30 lx.

image by adjusting the exposure time and the ISO speed. However, the balance between the exposure time and the ISO speed is significantly different between devices, as Table II shows.

Accurate correlations between exposure time and noise levels and correspondingly between ISO speed and noise

Peltoketo: Signal to noise ratio and visual noise of mobile phone cameras

Table II. Exposure and ISO speed in different light environments.

Device	Exp. time (ms), 1000 lx	ISO speed 1000 lx	Exp. time (ms), 100 lx	ISO speed 100 lx	Exp. time (ms), 30 lx	ISO speed 30 lx
Samsung Galaxy S3	20	80	59	320	59	800
Sony Xperia Z	16	50	31	400	77	400
iPhone 5	17	64	50	250	67	640
Oppo Find 5	27	100	83	400	333	400
Xiaomi MI2	25	100	63	541	63	1665
Samsung Galaxy S4 9500	20	64	59	250	59	800
HTC One	8	125	50	200	67	500
Samsung Galaxy S5	33	40	50	320	67	800
LG G2	25	100	67	500	100	1200
Samsung Galaxy S4 9505	26	50	67	250	67	800
Zopo C2	20	120	83	422	100	1187
Sony Xperia Z2	10	100	31	400	50	800
Nokia Lumia 925	10	200	40	640	91	800
Motorola RAZR i	15	100	37	400	63	800
Sony Xperia Z1 Compact	10	100	31	400	50	800
Huawei Ascend P1	20	100	53	450	67	1000
Nokia Lumia 920	10	160	40	500	91	800
Lenovo K860	20	50	63	200	63	1200
Nokia Lumia 1520	20	125	42	800	111	800
Nokia Lumia 1020	23	100	71	400	111	800

levels cannot be calculated in this research because both parameters affect the noise values and accurate correlations can be separated only if another parameter is kept constant. Nonetheless, some observations can still be made. The average correlations between exposure time and visual noise and total noise are 0.69 and 0.65 respectively, whereas the average correlations between ISO speed and visual noise and total noise are 0.71 and 0.72 correspondingly. However, there are some clear exceptions. Firstly, the Oppo Find 5 keeps the ISO speed low in the 30 lx environment but the exposure time is exceptionally long. Because exposure time affects the noise less than ISO speed and probably more aggressive denoising in low light, the total noise does not increase between 100 and 30 lx and visual noise even decreases. Secondly, Sony cameras overall, and the Sony Xperia Z in particular, keep both exposure time and ISO speed at a good level in low light environments and the noise levels decrease in a low light environment. Also the Nokia Lumia 1520 has the same kind of behavior. This is probably caused by aggressive low light denoising algorithms. Illumination based denoising may cause very nonlinear noise behavior and therefore mix the correlations. Without these exceptions the correlations between the parameters and noise levels are all above 0.90 which denotes, as expected, very strong dependence.

It has to be noted that denoising algorithms may decrease the other image quality factors even if the noise levels are decreasing. Overly aggressive denoising may cause blurriness and decreased texture sharpness in the images. On the other hand, excessive exposure time causes motion blur

and hand shaking issues whereas high ISO values increase noise in general.

Finally, Figure 5 visualizes the noise differences between devices. Even if the amplitude of the noise increases in low light, the order of the devices is quite stable. Moreover, there is no device which handles the low light environment exceptionally well but certain devices clearly have problems removing noise in difficult light conditions.

CONCLUSIONS

Obviously, the first official version of the visual noise method is the most significant update of the ISO 15739 standard. Clear step by step instructions, explicit color space conversions, CSF curve definition and result examples give good guidelines on how to measure, calculate and express visual noise results. However, there is a great possibility that the visual noise part of the standard will change in the near future due to new research results.

The main principles of SNR algorithm, SNR based noise, and dynamic range measurements have not been changed in the new version of standard. However, there are several small changes which affect both the measurement environment, such as the test chart recommendation and the requirement of illumination uniformity, and the measurement methods, such as the possibility of using defocus, the removal of the quarter pixel requirement of sequential images and the requirement of background luminance. These changes seem relatively minor but they may cause notable changes to the test methods, particularly in the case of the test automation

Peltoketo: Signal to noise ratio and visual noise of mobile phone cameras

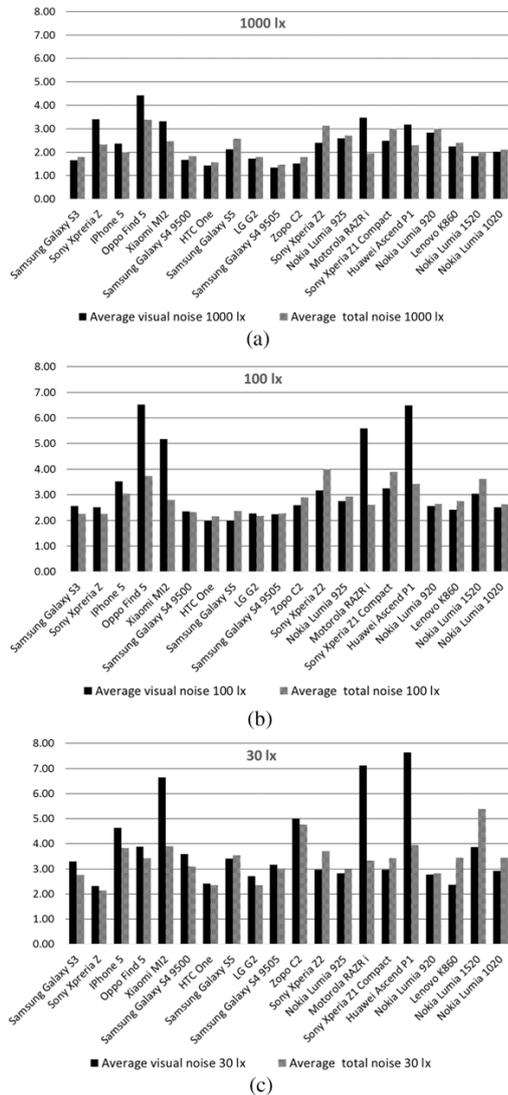


Figure 5. Noise values in different light conditions: (a) 1000 lx, (b) 100 lx and (c) 30 lx.

system. Overall, the new version of ISO 15739 is much more straightforward than the previous standard and several unclear requirements are removed or streamlined.

SNR based noise and visual noise values were measured from twenty mobile phone cameras using a test chart with twenty gray patches. Corresponding exposure and ISO speed values were detected and correlations between noise levels calculated to clarify the reasons for the noise. As expected,

the correlation is quite clear even if different denoising algorithms seem to mix the dependence.

The comparisons between SNR based noise and visual noise values were made using average noise values. Also perceptual comparisons were made between devices where visual noise and total noise values differ significantly. Even if the visual noise values were dominant in low light, generally the visual noise values and the SNR based noise values were surprisingly equal in all light conditions. However, there were some clear exceptions in the low light environments, where some of the devices had clearly higher visual noise values than total noise values. According to the visual inspection of the captured images, exceptionally high visual noise values correlate with the perceptual noise. Even if most of the devices have almost equal total noise and visual noise values, the exceptions reveal a true need for visual noise measurements.

REFERENCES

- 1 ISO 12233:2000 Photography—Electronic still-picture cameras—Resolution measurements (ISO Geneva), www.iso.org.
- 2 ISO 12232:2006 Photography—Digital still cameras—Determination of exposure index, ISO speed ratings, standard output sensitivity, and recommended exposure index (ISO Geneva), www.iso.org.
- 3 ISO 14524:2009 Photography—Electronic still-picture cameras—Methods for measuring opto-electronic conversion functions (OECFs) (ISO Geneva), www.iso.org.
- 4 ISO/CIE 11664-6:2014 Colorimetry—Part 6: CIEDE2000 Colour-difference formula (ISO Geneva), www.iso.org.
- 5 ISO 15739:2013 Photography—Electronic still-picture imaging—Noise measurements (ISO Geneva), www.iso.org.
- 6 ISO 15739:2003 Photography—Electronic still-picture imaging—Noise measurements (ISO Geneva), www.iso.org.
- 7 D. J. Baxter and A. Murray, "Calibration and adaptation of ISO visual noise for I3A's camera phone image quality initiative," Proc. SPIE **8293**, (2012).
- 8 G. M. Johnson and M. D. Fairchild, "A top down description of S-CIELAB and CIEDE2000," *Color Res. Appl.* **28**, 425 (2003).
- 9 J. Kuang, X. Jiang, S. Quan, and A. Chiu, "Perceptual color noise formulation," Proc. SPIE **5668**, (2005).
- 10 B. W. Keelan, E. W. Jin, and S. Prokushkin, "Development of a perceptually calibrated objective metric of noise," Proc. SPIE **7867**, (2011).
- 11 IEC 61966-2-1 Multimedia systems and equipment—Colour measurement and management—Part 2-1: Colour management—Default RGB colour space—sRGB (1999).
- 12 D. Hertel, "Extended use of ISO 15739 incremental signal-to-noise ratio as reliability criterion for multiple-slope wide dynamic range image capture," Proc. SPIE **7242**, (2009).
- 13 G. M. Johnson and M. D. Fairchild, "The effect of opponent noise on image quality," Proc. SPIE **5668**, (2004).
- 14 S. C. Kelly and B. W. Keelan, "ISO 12232 revision: determination of chrominance noise weights for noise-based ISO calculation," Proc. SPIE **5668**, (2005).
- 15 B. Baxter, F. Cao, H. Eliasson, and J. Philips, "Development of the I3A CPIQ spatial metrics," Proc. SPIE **8293**, (2012).
- 16 V.-T. Peltoketo, "Mobile phone camera benchmarking: combination of camera speed and image quality," Proc. SPIE **9016**, (2014).