Tampere University

Ilpo Viertola

AUTOMATIC OPTIMIZATION OF IMAGE QUALITY WITH RESPECT TO VCX V.2020 STANDARD

Master of Science Thesis
Faculty of Information Technology and Communication Sciences
Examiners: Prof. Esa Rahtu
Timo Gerasimow
April 2023

ABSTRACT

Ilpo Viertola: Automatic Optimization of Image Quality with Respect to VCX v.2020 Standard
Master of Science Thesis
Tampere University
Computing Sciences
April 2023

The image signal processor (ISP) takes care of handling the image signal captured by the image sensor. ISP consists of multiple blocks, each block focusing on a certain aspect of the image quality (IQ). Each block may process the signal with multiple different algorithms and these algorithms may contain multiple parameters altering the way they manipulate the signal. These parameters must be tuned to meet the expected IQ.

Tuning the ISP is a time-consuming and tedious job to do manually. Traditionally, IQ engineers have been tuning the parameters using the tuning software provided by the ISP vendor. During this process, they have used their IQ expertise and objective IQ metrics like the amount of visual noise in the resulting image to guide the IQ to a satisfactory level.

In this thesis, an automatic ISP tuning framework will be used to tune the ISP. This thesis aims to answer the following research questions: "Will optimizing ISP parameters using fully objective metrics produce subjectively good image quality?" and "Will tuning ISP in order to maximize the outcome of VCX score metrics produce better results than using alternative ways?". The set of fully objective metrics that are used in this thesis is called Valued Camera eXpression (VCX) score. By maximizing the VCX score, the best overall performance of the camera should be obtained. The alternative way used to measure IQ in this work is a deep learning-based method, which is proven to correlate well with human perception of IQ. Finally, a subjective IQ analysis is conducted, and the resulting IQ is compared between the results and a commercial reference device.

The research revealed promising results by using only objective metrics as IQ targets when tunings the ISP. The subjective IQ analysis revealed that using the deep learning-based approach produced slightly better IQ than using only objective metrics. Still, the ISP tuning produced using only objective metrics was selected more frequently over the commercial reference device in the subjective IQ analysis conducted as a part of this study.

The work done in this thesis gives a good foundation for future research on using different types of IQ metrics in automatic ISP tuning procedure. It is argued that even though the fully objective metrics did not produce the overall best outcome, they have their place in the IQ tuning. For example, a device manufacturer may demand a certain level of noise performance from the ISP. In this case, the ISP tuning has to be guided to a result that optimizes the overall IQ, still maintaining the desired level of noise.

Keywords: Image quality tuning, Optimization, Image quality metrics, Image quality

TIIVISTELMÄ

Ilpo Viertola: Automaatinen kuvanlaadun optimointi VCX v.2020 standardin suhteen
Diplomityö
Tampereen yliopisto
Tietotekniikka
Huhtikuu 2023

Kuvasignaaliprosessori (ISP) käsittelee kamerasensorin tallettaman kuvasignaalin. ISP rakentuu monista eri osioista, joista jokainen keskittyy tiettyyn kuvanlaadun aspektiin. Jokainen osio voi prosessoida digitaalista kuvasignaalia monin erilaisin algoritmein, joista jokainen voi sisältää monia erilaisia parametreja. Parametreilla algoritmin käyttäytymistä ja tapaa muokata signaalia voidaan ohjata. Näiden parametrien säätäminen on erityisen tärkeää, jotta ISP:n tuottama kuvanlaatu saadaan maksimoitua.

ISP:n parametrien säätäminen on aikaavievää ja hankalaa. Tavallisesti kuvanlaatuinsinöörit ovat säätäneet ISP:n parametreja käyttäen ISP-valmistajan siihen kehittämää ohjelmistoa. He ovat hyödyntäneet asiantuntemustaan sekä erilaisia objektiivisia kuvanlaadun metriikoita, kuten kohinan määrää, saadakseen laitteen tuottaman kuvanlaadun välttävälle tasolle.

Tässä työssä käytetään automaattista kuvasignaaliprosessorin säätämisohjelmistoa kuvanlaadun maksimoimiseksi. Tämä työ pyrkii vastaamaan seuraaviin tutkimuskysymyksiin: "Voiko kuvasignaaliprosessorin parametrien säätäminen käyttäen vain objektiivista kuvanlaatumetriikkaa tuottaa subjektiivisesti hyvän kuvanlaadun?"ja "Tuottaako kuvasignaaliprosessorin säätäminen VCX standardin mukaan paremman kuvanlaadun, kuin käyttäen vaihtoehtoisia tapoja?". Tässä työssä käytetty objektiivinen kuvanlaadun metriikka on nimeltään Valued Camera eXpression (VCX) standardi. VCX mittaa kameran suorituskykyä yhdellä numeerisella arvolla, ja sen maksimoinnin pitäisi tuottaa paras mahdollinen suorituskyky valitulle laitteelle. Toinen työssä käytetty vaihtoehtoinen tapa kuvanlaadun mittamiseen on koneoppimispohjainen neuroverkko, jonka on osoitettu korreloivan hyvin ihmisen tavan kanssa ymmärtää kuvanlaatua. Lopuksi subjektiivisen kuvanlaatuanalyysin tuloksia vertaillaan eri ISP säätöjen välillä. Tässä työssä tuotettuja säätöjä vertaillaan myös kaupallisen laitteen tuottamaan kuvanlaatuun.

Työn tutkimus paljasti lupaavia tuloksia objektiivisen metriikan käytöstä subjektiivisen kuvanlaadun säätämisessä. Subjektiivisessa kuvanlaatuanalyysissa parhaiten pärjäsivät säädöt, joiden tuottamisessa oli käytetty neuroverkkopohjaista ratkaisua kuvanlaadun mittaamiseen. Kuitenkin ISP säätö, jossa oli käytetty vain VCX standardia kuvanlaadun mittaukseen säädön aikana, valittiin useammin paremmaksi kuin kaupallinen vastine subjektiivisessa kuvanlaadun testissä joka suoritettiin osana tutkimusta.

Tässä tutkielmassa tehty työ antaa hyvän pohjan tulevalle tutkimukselle, joka koskee erityyppisien kuvanlaatumetriikoiden käyttöä automaattisessa ISP:n säädössä. Vaikka täysin objektiiviset kuvanlaatumetriikat eivät tuottaneetkaan parasta kokonaistulosta, niillä on paikkansa ISP:n säätämisessä. Laitevalmistaja voi esimerkiksi vaatia ISP:ltä tiettyä kohinatasoa. Tällöin ISP:tä on ohjattava tulokseen, joka optimoi yleisen kuvanlaadun säilyttäen silti halutun kohinatason.

Avainsanat: Kuvanlaadun optimointi, Optimointi, Kuvanlaatumetriikka, Kuvanlaatu

PREFACE

Many thanks to my former manager Timo Gerasimow and co-workers, people who attended the image quality study, and my thesis supervisor Esa Rahtu for the support.

Many thanks also to my flatmates, friends, and family for supporting me in my darkest times. Also, I would like to thank Pikilinna for such brilliant student-priced beverages.

At Tampere, 12th April 2023

Ilpo Viertola

CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

ABC	Artificial Bee Colony
AI	Artificial Intelligence
CCM	Compact Camera Module
CNN	Convolutional Neural Network
CSF	Contrast Sensitivity Function
DL	Dead Leaves
DR	Dynamic Range
DUT	Device Under Test
e-SFR	Edge-based Spatial Frequency Response
EPC	Effective Pixel Count
ESF	Edge Spread Function
HDR	High Dynamic Range
IQ	Image Quality
IQA	Image Quality Assessment
ISO	International Organization for Standardization
ISP	Image Signal Processor
LR	Limiting Resolution
MFNR	Multi-Frame Noise Reduction
MTF	Modulation Transfer Function
OECF	Opto-Electronic Conversion Function
s-SFR	Sine wave-based Spatial Frequency Response
SFR	Spatial Frequency Response
SNR	Signal-to-Noise ratio
TMR	Theoretical Maximum Resolution
TPC	Theoretical Pixel Count
VC	Viewing Condition
VCX	Valued Camera eXpression

VN	Visual Noise
WB	White Balance

1. INTRODUCTION

Modern mobile devices act as a replacement for traditional digital cameras for many people. Due to the physical limitations of these devices, they are not able to collect the same amount of light or have the same quality of optical lenses as full-size digital cameras [31]. The artifacts caused by these limitations must be overcome with digital image processing. Digital image processing happens in the image signal processor (ISP) of the device. The signal-processing algorithms contain multiple tunable parameters whose optimal value must be found in order to get the best performance out of the imaging device. In this work, the ISP is aimed to tune automatically using different image quality targets to find the optimal set of these parameters.

While manufacturers release smartphones with ever-increasing megapixel amounts and artificial intelligence (AI) based image enhancement algorithms, it might be difficult for a consumer to get an idea which smartphone has a sufficient camera for their needs [50]. Being able to define the camera's quality is important for the end user due to various social media platforms prioritizing images as the main medium, and modern mobile devices acting as a replacement for traditional digital cameras. Cameras have achieved their place as one of the most important components in today's smart devices [49].

Defining which mobile camera produces the best image quality (IQ) is challenging since IQ is subjective, and other observers may value it differently than others [36]. Quality judgment of an image is called image quality assessment (IQA). It can be divided into two different main categories: objective and subjective IQA [36]. Subjective IQA is based on surveys conducted in groups of observers who rank the images shown based on their subjective opinion. Objective IQA is based on different IQ metrics. An IQ metric measures analytically a certain aspect of the quality, for example, the amount of noise present. Combining multiple metrics can give a comprehensive interpretation of the subjective IQ. In order to produce reliable subjective or objective IQA results, assessments must follow standardized testing procedures. One of these procedures is called the Valued Camera eXpression (VCX) standard [49] which is a fully objective IQA method.

Subjective IQA is time-consuming. In order to attain reproducible, statistically significant, and reliable results, a large group of people has to be questioned using specific methodologies. Due to the amount of work it takes, objective IQ metrics are introduced.

These metrics aim to predict and model the results of subjective IQA using conventional algorithms or machine learning-based methods. In this work, these methods allow us to measure IQ instantly and use that as feedback when we tune the ISP to obtain the best possible final IQ.

The VCX standard aims to tackle these aforementioned challenges of comparing the quality of imaging devices. It aims to characterize the IQ in a way that helps consumers to select the best-fitting device for their needs. VCX score combines different aspects of camera performance and adds them up to a single score. This way comparison of different devices is easy and straightforward. In this work, the IQA during the tuning procedure is done using the IQ metrics defined in the VCX standard. Since the VCX forum claims that the higher the VCX score better the camera performance for consumers, maximizing the VCX score while tuning the ISP should produce the best possible outcome. [49], [50]

This work is done for *AAC Technologies Oy* and an automatic image tuning optimizer framework of theirs will be exploited in the tuning of the ISP.

This thesis aims to answer the following research questions:

1. Will optimizing ISP parameters using fully objective metrics produce subjectively good image quality?
2. Will tuning ISP in order to maximize the outcome of VCX score metrics produce better results than using alternative ways to measure the IQ?

The outline of the thesis is the following: chapter 2 presents the theory behind image quality, image quality assessment, and image quality tuning, and chapter 3 elaborates on the theory behind the VCX standard and score. Chapter 4 defines the hardware and laboratory environment. Also, the fundamentals of the image processing pipeline will be described. IQA and IQ tuning methodologies are covered in chapter 5. The tuning experiments are conducted in chapter 6 and finally, conclusions are drawn in chapter 7.

2. ASSESSMENT AND TUNING OF IMAGE QUALITY

This chapter will describe how image quality is defined in the scope of this work, how one can assess it using different methods and techniques, and the basic principle of IQ tuning. Image quality assessment aims to produce consistent and reproducible analysis results on the quality of an image. Chapter 3 describes an IQA standard, VCX Standard. The VCX Standard introduces a way to benchmark and compare imaging devices with each other in a consumer-friendly way by reporting the IQ of a device using a single number [49]. Before diving into the assessment and tuning of IQ any deeper, the IQ itself is described in a more elaborated manner in the following section 2.1.

2.1 Image Quality

It is hard to define what is actually good IQ. It has to be kept in mind that an image's content also contributes to the IQ perceived by a human. The emotional response might be so substantial that it makes the poor technical quality of an image diminish completely while a person is assessing its quality. One might even introduce artifacts to an image, which lowers the technical IQ, to make it look better from an artistic aspect. For example, introducing a vignetting effect degrades the technical quality of an image but might be preferred by the observer. Figure 2.1 displays this phenomenon. In the field of medical imaging, computer vision applications, or surveillance cameras, IQ is not measured as how pleasing the image is visually but as how detectable objects are. Often the division is made between artistic, human-perceived IQ and measurable, technical IQ. These categories of IQ are referred to as subjective and objective. One can also aim to measure subjective IQ using objective metrics. In the scope of this work, IQ is perceived from the subjective consumer viewpoint but aimed to measure in an objective manner to ensure consistent and reproducible results with rapid analysis time. [35], [36, p. 27-29], [53]

Image Quality Attributes

IQ can be described as a limited set of attributes that aim to describe and represent certain aspects. Breaking the IQ into smaller parts makes the whole assessment procedure more manageable. These attributes consist of four categories which are personal, aesthetic, artifactual, and preferential. Personal IQ attributes are often linked between image

Figure 2.1. *Comparison between the original image and image with added vignette for artistic purposes.*

content and a specific person. For example, a loved-one present in the image might increase the perceived quality. Aesthetic attributes include things like the composition of the image which can also be relevant not only for a specific observer but for observers in general. A good composition makes the image look more pleasing. These two categories are very subjective and are difficult to measure in an objective reproducible manner. However, artifactual and preferential attributes can be. Preferential attributes include aspects like color saturation and accuracy of color reproduction. Artifactual attributes include aspects like the sharpness of an image and the amount of visual noise present in the image. The main difference between artifactual and preferential attributes is that an optimum value exists for preferential attributes, and after that point, the IQ starts to decrease. For artifactual attributes increasing the strength of the attribute will make the IQ only more objectionable. Figure 2.2 illustrates the effect of oversharpening. Adding too much sharpening will make the IQ degrade because more sharpening artifacts are introduced, like bright halos around the black chair. [36, p. 29-31]

Figure 2.2. *Comparison between the original and oversharpened image. Note the artifacts caused by oversharpening, like bright halos around the black chair.*

IQ attributes can also be divided into global and local attributes. The scale and viewing distance of an image distinguishes between these two categories. A global attribute can be perceived regardless of the scale or viewing distance. For example, color saturation is a global attribute. On the other hand, the sharpness of an image is a local attribute since the scale of an image contributes to perception. Figure 2.3 presents the differences between local and global IQ attributes. The left image is the original, the middle image is desaturated, and the right image is blurred. The scale of the image does not affect the perceived color saturation. On the other hand, the sharpness difference is not distinguishable from the bottom row's downscaled images. The scale of the image affects the perception of sharpness, thus making it a local image quality attribute. [36, p. 29-31]

The main IQ attributes that are measured in VCX Standard, and in the scope of this work, are resolution, texture reproduction, sharpening, colors, and visual noise [50]. From these attributes, color reproduction is a global attribute. The rest can be referred to as local attributes. All of these attributes are also objectively measurable. VCX Standard is described in a more elaborated manner in chapter 3.

2.2 Image Quality Assessment

As mentioned before, image quality can be assessed objectively and subjectively. Because subjective IQ is tightly linked with human judgment and perception, it might feel counterintuitive to measure it using objective metrics. However, it is important to note that subjective assessment does not mean arbitrary or random judgment of an image. Subjective IQA is done within a large enough population of observers using statistical methods which will produce accurate and repeatable objective results. This can be a very

Figure 2.3. *Illustration of differences between global and local image quality attributes. The left image is the original, the middle image is desaturated, and the right image is blurred. In the bottom row's downscaled images sharpness difference between the original and blurred images is not visible, making sharpness a local image quality attribute. [36, p. 31].*

time-consuming task. [36, p. 31-32]

One clear advantage of objective metrics is almost instant feedback of the measured IQ which allows for predicting it fully automatically. Predicting here describes that the objective metrics try to predict subjective IQ measurement results. In the scope of this work, subjective IQ judgment is mimicked with objective IQA during the IQ optimization phase. In other words, a set of objective metrics are combined and weighted in a manner that tries to predict the results of a subjective IQA made for the specific image. After the IQ optimization is done, the final IQ is evaluated with a subjective image quality assessment. [50], [51]

The objective IQ measurement methods can be further divided into three different subcategories: full reference, no reference, and reduced reference methods. The reference is the original, perfect, image that does not contain any artifacts caused by the image-capturing process. It is distortion-free and has the perfect IQ. No reference IQA is quite an uncomplicated task for human observers since we can identify what is a good IQ relatively well, but is not so straightforward to do computationally. No reference IQ metrics are, for example, measuring a signal-to-noise ratio or color saturation. Most of the metrics used in VCX Standard are no reference metrics [50]. [36, p. 31-32], [51]

Another way to objectively assess IQ is to use deep learning-based methods. Deep learning is a field of machine learning that focuses on deep neural network models which are inspired by the behavior of the human brain. A neural network consists of multiple neurons that are connected to each other in different ways. The structure of these networks

allows them to learn complex patterns from large datasets making them able to recognize, classify, and describe objects present in the data. Essentially, the network divides a large function, for example, an imaging pipeline, into simpler more comprehensible functions. [24, p. 7-10]

When the data consists of images, often convolutional neural networks (CNN) are used which were originally designed for image recognition tasks. CNNs extract local visual features in the early layers of the network and later combine them into high-order features. For example, early layers extract information like edges, and neurons in later layers learn to combine this information into greater features like body parts. This is achieved by using the convolution operation and weight sharing between the neurons of a single layer. In this work, a deep learning method is used to assess the IQ during the automatic tuning procedure. A more elaborated explanation of the method is presented in section 5.2. [24, p. 160-168]

2.3 Image Quality Tuning

The main task of the ISP is to reconstruct and enhance the sensor's raw data to output the best possible final image which is then viewed or manipulated by the consumer. In the scope of this work, the consumer is a human observer but it could also be, for example, a computer vision application. This is a very complex task with multiple intermediate steps. Due to the very complicated nature of the task, the ISP has many parameters which modify the algorithms used to obtain the best possible final result. The functions of the ISP are further discussed in the section 4.2. With IQ tuning, camera manufacturers aim to get rid of artifacts, for example, distortions and color errors, which are introduced by the imaging system. [33], [44]

In a normal IQ tuning project, a human being has to use their prior knowledge and trial-and-error to determine which parameters to adjust to optimize the IQ. After adjusting parameters by hand, they evaluate the IQ, either subjectively or using objective metrics like signal-to-noise ratio. Since this optimization problem has multiple dimensions, it is very plausible that the best possible result is never obtained. Also, producing customer and image-sensor-specific IQ tunings is often the bottleneck in delivering different solutions to customers [33]. [44]

In this work, an automatic IQ tuning framework will be introduced. The tuning framework will take an advantage of a set of objective IQ functions defined in the VCX standard. VCX standard assumes that this set of metrics reflects subjective IQ well, and thus maximizing the VCX score will produce an image with good subjective quality. The automatic tuning framework will attempt to tune the ISP parameters so, that this score will be maximized. Thus, no subjective IQ evaluation performed by a tuning engineer is needed during the IQ tuning process.

Automatic Image Quality Tuning

Automatic IQ tuning is a relatively widely researched topic, and also commercial solutions already exist [1]. Existing solutions can be divided into two main categories: solutions that use bio-inspired algorithms and solutions that use deep learning-based optimization techniques.

Due to the nature of the optimization problem at hand traditional optimization techniques, like gradient-based methods, do not work efficiently enough. To cope with high complexity, traditional techniques are enriched with bio-inspired optimization techniques, like evolutionary algorithms, to better handle the nonlinear nature of image processing problems. Bio-inspired methods are often based on swarm intelligence. Swarm intelligence-based algorithms aim to imitate the social behavior of insect colonies. By mimicking the collective behavior of a population of unsophisticated and less intelligent agents, the complex problem is divided into smaller parts. Agents are collectively capable through direct interaction, and by modifying their surroundings to solve difficult problems. Each agent represents a potential solution to an optimization problem and moving within the problem domain aims to improve the existing solution. [38], [46]

As stated before, plenty of work has already been done around this topic. Nishimura et al. utilize swarm intelligence-based artificial bee colony (ABC) algorithm for the global exploration of the ISP parameter space [33], Portelli and Pallez take advantage of speed-constrained multi-objective particle swarm optimization algorithms [38], and Hevia et al. use differential evolution search algorithm to find the optimal set of the ISP parameters [18]. Bhat and Pavithra propose a multi-stage optimization approach utilizing a non-sorted genetic algorithm to automatically tune the ISP parameters [41].

Another popular approach to automatically optimize the ISP parameters is to use deep learning-based methods. A very brief introduction to the concept of deep learning is provided in section 2.2 above. Shen et al. utilize a parameter tuning policy network that is trained using reinforcement learning strategies to achieve good IQ of computed tomography images [44]. Kim et al. propose a deep neural network-based automatic IQ tuning framework [25]. Ethan Tseng et al. optimize the ISP parameters using a differentiable mapping between the parameters and IQ metrics [48].

3. VALUED CAMERA EXPRESSION STANDARD V. 2020

In this chapter, the Valued Camera eXpression score (VCX score) version 2020 is described. VCX score is an objective ranking method developed by VCX-Forum whose goal is to assess the IQ of mobile devices with only a single number, the VCX score [49]. The VCX score is calculated under a regulated environment using standardized testing procedures which allows reliable comparison of the results. VCX can also be seen as a standard or test procedure that helps mobile camera manufacturers develop better-quality products [3]. It is important to note that VCX-Forum is a non-profit organization, and nobody involved in developing the VCX score, or calculating the VCX score, gains any benefits or economic profits from doing so [49]. This will guarantee the objectivity and transparency of the VCX score.

3.1 The Five Tenets of VCX Score

This section will describe the five tenets of the VCX score which characterize its nature more thoroughly [49]. The later sections will take a deeper dive into the VCX score and how it is calculated but before that, it is good to get an overview of the whole scoring system. These tenets are adapted from [49].

- **1. tenet:** VCX measurements shall ensure the out-of-the-box experience, which means that the device under test (DUT) is bought from a regular store to ensure that no special software or hardware is being used to achieve a better score. Also, the device is tested using the default camera application using the default settings.
- **2. tenet:** VCX shall remain 100% objective. The purpose of this is to the repeatability of the tests and to minimize the subjective opinions of human beings which differ from one person to another.
- **3. tenet:** VCX shall remain open and transparent. The VCX score can be accessed by anyone, and the formulas how to calculate the scores are available to everyone. Also, the scores of different mobile devices are publicly available on the VCX-Forum's website www.vcx-forum.org.
- **4. tenet:** VCX shall employ and use an independent imaging lab for testing. Meaning that even though the VCX score is open source and publicly available, anyone can't publish their VCX score results on the official page. The final scores on the

VCX-Forum's website are always provided by a trusted and independent laboratory.

- **5. tenet:** VCX shall seek continuous improvement. VCX standard is constantly under development guided by feedback from customers and mobile device vendors.

These five tenets are the core values of the VCX standard and have been adopted by all the contributors that have supported the development of the standard. As an example, a few large contributors are Google, Microsoft, Huawei, and Vivo. [54]

3.2 The Prerequisites of VCX Score

In this section, the ways of measurement and calculation of the VCX score are specified. As mentioned above, the final result will be a single score. It is calculated from many different individual scores that describe the DUT's IQ and performance under different conditions [50]. The final single score, the VCX score, aims to reflect the user experience of the DUT [54]. The measurement bases itself on objective data captured with well-defined test procedures [50]. As mentioned [50], the main IQ aspects covered are:

- Spatial resolution – What level of details can I see?
- Texture loss – How does the device reproduce low contrast, and fine details?
- Sharpening – Does the device apply so much sharpening that disturbing artifacts occur?
- Noise – How much disturbing noise do I see?
- Dynamic range – What is the maximum contrast in a scene the device can reproduce?
- Color reproduction – Are there any issues in color processing?

Data and data capture

As mentioned in section 3.1 tenet number one, the DUT is kept in very default settings. This is ensured with a factory reset before capturing the data. The data capturing procedure includes four captures per test scenario, from which the image with the highest resolution at the image center is selected. The data is in JPEG format. VCX standard uses two different test charts. A test chart is a target that contains content that is hard to reproduce for cameras. Figure 3.1 illustrates the test target TE42-LL used in the VCX standard. This chart is standardized by the International Organization for Standardization (ISO) with standard number 19093 and is designed to contain test structures that enable the measurement of different aspects of IQ [12]. This test chart is needed in three different sizes to cover all the different aspect ratios, and also testing of the selfie camera. [50]

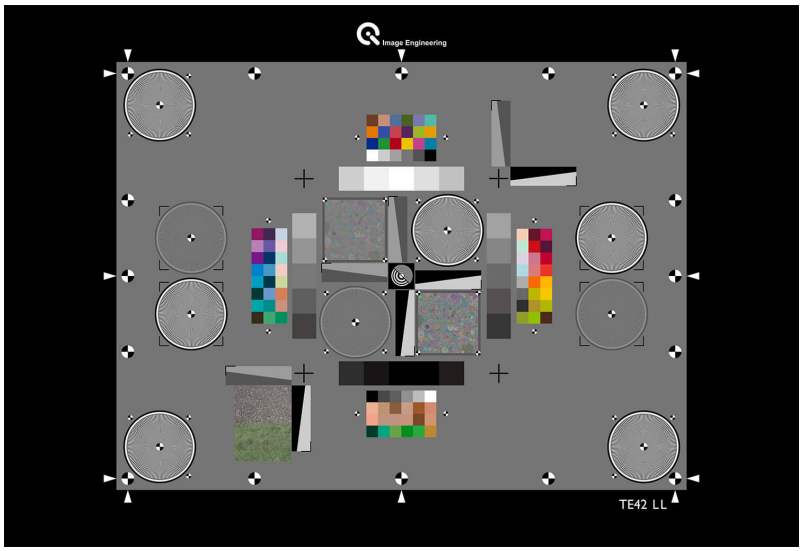

Figure 3.1. *TE42-LL multi-purpose test chart for measuring low light performance (ISO 19093) [12].*

The second test chart needed to measure the VCX score of a mobile device is called TE269B [50]. TE269B is a transparent test chart used to measure DUT's dynamic range which follows closely the standard, defined by ISO 14524 [11]. This chart is presented in figure 3.2 below. Dynamic range score is not measured in the scope of this work.

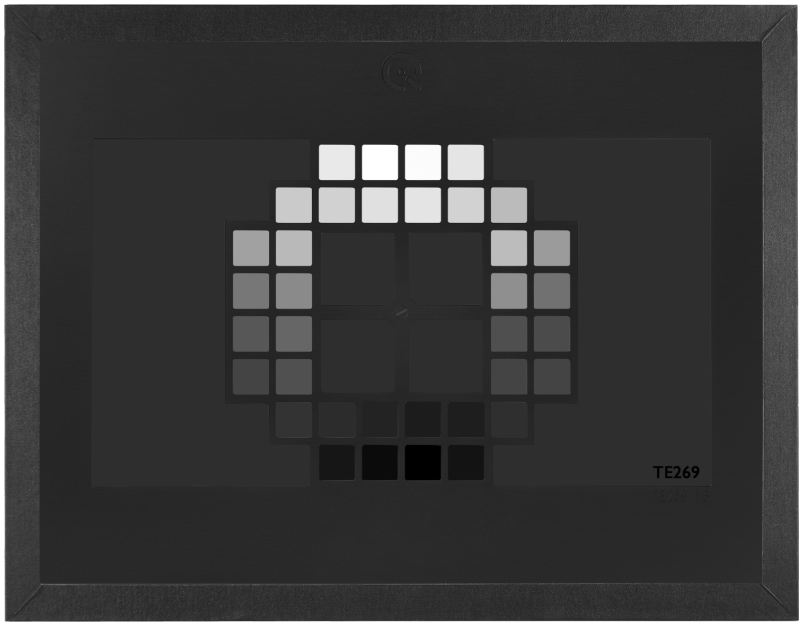

Figure 3.2. *TE269 test chart for measuring dynamic range of the imaging device (following ISO 14524/15739 and IEC 62676-5) [11].*

In order to calculate the complete VCX score for a device, also other hardware is needed. This includes a handshake simulator to see the effect of a natural handshake on the perceived IQ, and timing kits to measure the performance of the camera [50]. This equipment is not relevant to the scope of this work, and since only been noted here.

One of the main pieces of equipment, in addition to test charts, is lights. It is very important that the light sources used while capturing the data have matching properties to the ones defined in the [50]. IQ of the main camera of the DUT is measured under five different lighting conditions:

- bright light, 2000 lux, daylight (D55)
- mid-light, 250 lux, natural LED
- low light, 10 lux, warm LED
- extended low light, 10 lux, 7.5 lux, 5 lux, 3 lux, 2 lux, 1 lux
- DUT's flashlight, under mid-lighting conditions [50].

The device is also tested using 4x zoom under bright, mid, and low light described above [50]. The VCX standard also defines lighting conditions for video and selfie (front) camera testing but these test cases do not fall under the scope of this research. In this thesis, only bright, mid, and low light levels are used for the default use case.

Viewing conditions

Sometimes the IQ may vary between different ways of viewing the image. For example, one can't see the visual noise of an image that clearly from a small screen compared to a larger display. The following ways of viewing, viewing conditions (VC), of the image content are assumed in the VCX standard version 2020 whitepaper [50]:

- **VC1 100% view:** user can see most of the details, the viewing distance is 0.5 meters and a 100% view on a 96 pixels per inch (PPI) display.
- **VC2 small print/smartphone display:** the complete image is scaled from 100% to a height of 10cm, viewing distance is the same as the image's diagonal length.
- **VC3 large print/computer display:** the image is scaled to a height of 40cm, viewing distance is the same as the image's diagonal length.

The upcoming section 3.3 is going to introduce the IQ metric categories used in VCX Standard. Each category includes different IQ metrics, that characterize a certain IQ aspect. These measurements are then combined into one device-specific VCX score.

3.3 VCX Metrics

All the metrics are strictly objective, measured from the data that is captured under a controlled environment, using standardized test targets, and controlled capture procedures. The metrics can be divided into 11 different subcategories which are: resolution, texture loss, sharpening, colors, visual noise, dynamic range, shading, distortion, extended low light test, response time, and motion control [50]. The resolution, texture loss, sharpening,

colors, and visual noise metrics are used when optimizing the IQ in the later stages of this work. Next, the metrics are defined in more detail. All the explanations are adapted from the VCX standard v2020 Whitepaper [50]. For a more elaborate and detailed explanation of individual metrics and formulas, see Appendix A.

Resolution

Resolution is defined as a level of detail that the camera is able to reproduce, and not as a pixel count of the camera sensor used in the DUT. Since ISPs are able to process parts of the image differently depending on the content, the VCX standard uses many different structures to obtain reliable resolution measurements. The different metrics defining the resolution of an image are sine wave-based spatial frequency response (s-SFR) - limiting resolution (LR) calculated from the center and corner of an image, effective pixel count (EPC) calculated from the center and overall, s-SFR acutance calculated from center and corner of an image, 50% value of modulation transfer function (MTF50) calculated from edge-based spatial frequency response (e-SFR), and e-SFR acutance.

Spatial frequency response (SFR) is a metric defined by ISO12233 that describes the loss of contrast in the image as a function of spatial frequency. Contrast tends to decrease as a function of spatial frequency to a point where one can't distinguish any details [22]. This limiting frequency value is also known as the resolution of the camera [22]. The MTF is a similar metric to SFR and is often used interchangeably [19]. MTF*nn* means the spatial frequency value where the MTF is *nn*% of the low zero frequency MTF [19]. s-SFR refers to a method of how the SFR is calculated. s-SFR is deducted using Siemens star -targets because these are sine wave modulated targets and thus less influenced by image enhancement algorithms. That's why LR is calculated using it. An example of a sinusoidal Siemens star is presented below in figure 3.3. These targets can also be found in the TE42-LL test chart 3.1.

Texture Loss

The combination of compact sensors and high pixel count in smartphones today results in a high need for noise reduction in the signal processing pipeline of these devices [2]. Texture loss is a significant metric that describes how balanced the noise removal of the ISP is. As stated before, the image content dictates how the image is processed. Traditionally can be said that for flat areas of an image, like the sky, ISP applies stronger noise reduction than to areas where it detects structure. This is done because too strong noise filtering on edges will blur the resulting image, making it too soft and unpleasant to view for a human observer. When the image content has low contrast, fine details, or both, it gets harder to make a clear distinction between flat and structured areas. The

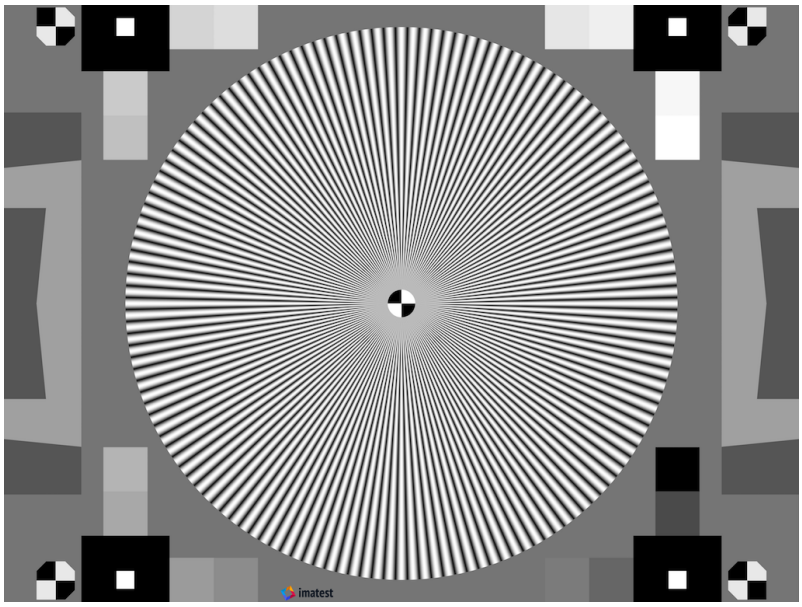

Figure 3.3. *Sinusoidal Siemens Star target used to measure resolution of the imaging device [20].*

loss of this kind of detail is also referred to as texture loss.

VCX Standard uses well-known dead leaves (DL) -target to measure this metric. The test chart used in the VCX score, presented in figure 3.1, contains two of these targets: high and low contrast versions. Both of these are used in score calculations. An example of dead leaves is presented in figure 3.4 below. The method used to analyze this target is developed by Image Engineering, and it is called $SFR_{DeadLeaves_{cross}}$ which is a fully objective reference-based method tackling the previous problems of sharpening corrupting the spatial frequency response measured from dead leaves target [2]. The different metrics composing the VCX sub-score for texture loss are texture loss MTF10/MTF50, texture loss acutance, artifacts, and chrominance loss.

Sharpening

The perceived sharpness can be increased by the ISP of the device. The way the sharpness is often increased is to enhance the local contrast around edges which makes the image look sharper. As mentioned before, sharpening increases the SFR of the camera because sharpening increases the response of the imaging system to certain spatial frequencies. To obtain an in-depth understanding of the image quality the DUT is able to produce, it is important to identify the effect of sharpening on the SFR. The sharpening subscore of the VCX score consists of two measurements which are overshoot and undershoot area, and maximum SFR.

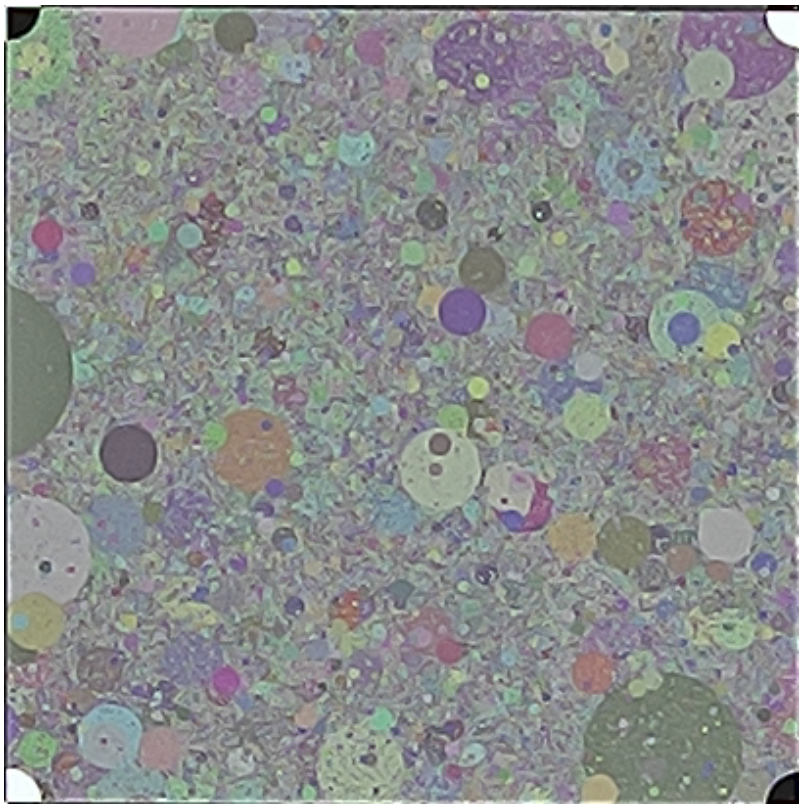

Figure 3.4. *Low contrast dead leaves target cropped from TE42-LL-target. Used to measure loss of textures.*

Colors

Even though smartphone cameras are not made for perfect and accurate color reproduction, it is important to measure the reproduction quality. Smartphone manufacturers aim to reproduce nice colors. For example, a light blue sky looks more beautiful than a gray one which it often is. Manufacturers make their devices in a way that often this gray-looking sky has a beautiful light blue color cast in the final image. The color part of the VCX score is measured using the four color patches of the TE42-LL 3.1. The actual values of the color targets are known, and this reference data is provided in the CIE-XYZ color space. Both, the captured image data and the reference data are converted to CIEL-L*a*b color space to get more accurate values of color differences. The color difference subscore is then formed by combining the color error, luminance error, chrominance error, and hue error.

Visual Noise

Noise is a basic element when measuring real-world phenomena with digital measurement devices. A camera sensor is not any different. Each pixel of a sensor can be perceived as a measurement device of the object's luminance and thus measurements contain noise. The signal-to-noise ratio (SNR) expresses the amount of noise in the im-

age signal. SNR value does not correlate well with the human perception of a good image. This is why the VCX standard uses visual noise (VN) as a unit to measure the amount of noise in the image. It correlates better with human perception and with the viewing condition. The analysis follows the standard ISO15739:2013 and is described there.

Dynamic Range

Dynamic range (DR) describes the camera's ability to reproduce details in shadowed areas and retain highlights. A low dynamic range will result in clipping in highlights which means that the pixels saturate and do not contain any relevant image data anymore. Saturation means that the maximum digital value of a sensor is reached. For most sensors, this is 255, as the sensors operate in the 8-bit domain. The DR calculations in the VCX score are based on the optoelectronic conversion function (OECF). OECF defines the relationship between exposure or reflectance, and the digital value of a pixel [14]. This measurement uses the TE269B chart displayed in figure 3.2. In TE269B different patches have varying reflectance from light to dark with different optical densities provided by the manufacturer.

Shading

In the context of the VCX standard, shading means a change of intensity or color over the target. Since TE42-LL features a uniform grey background which is illuminated with uniform illumination, its digital values in the resulting image should also be the same and not change as a factor of spatial location. Shading is an artifact introduced mainly by the optics of the DUT but it is often corrected digitally. This correction may vary under different illuminants and affect differently to noise level or color processing.

Distortion

Distortion is also an artifact caused by the lens of the camera and is often corrected by the ISP. VCX score measures the distortion with TV distortion metric using the black and white round position markers circling around the TE42-LL-chart which is presented in figure 3.1.

Extended Low Light Test

Extended low light test procedure contains capturing of test images in decreasing illumination. The test is continued until the metrics calculated from the captured image fail to reach certain thresholds. The lowest illumination level that the DUT can pass is stated as its low light performance. This test follows the procedures defined in ISO19093:2018. The

color temperature of the illumination used is D55 and the illumination levels are 2000 lux, 10 lux, 7.5 lux, 5 lux, 3 lux, 2 lux, and 1 lux. The 2000 lux image is used as a reference. The extended low-light test consists of 5 metrics:

- **Exposure** value is the luminance channel value of 50% grey patch.
- **Resolution** is calculated as s-SFR MTF10 A.5 value from the middlemost Siemens star.
- **Texture** is calculated as $SFR_{DeadLeaves_{cross}}$ MTF10 value.
- **Visual noise** is calculated as VN_{mean} A.17.
- **Color error** is calculated as $\Delta C*_{condition}$ A.10 using the top most color chart in TE42-LL.

3.4 Calculating the VCX Score

Formulas and explanations for every individual IQ metric used in VCX score are presented in appendix A. This section only describes how the scores calculated under different settings are combined into a single VCX score. The graph in figure 3.5 presents how different types of measurements are weighted when the VCX score is formed. In the scope of this work, every irrelevant part is being grayed out. This thesis will only focus on the image quality score of the main camera, and the VCX score is only optimized for default capture mode under bright, mid, and low illumination. This restriction is done to simplify the problem, and because this part of the VCX score covers a large set of different camera use cases. This is an integral part of the score and many different IQ problems can be identified with these measurements and scenarios.

The total range of the VCX score is 0-100 and the full score means that the DUT achieves the best results from every possible metric. In this work, when the VCX score is calculated the weights are adjusted so that the score of 100 is still reachable. The adjustment will not affect the relation of different weights. The rest of the score is ignored since the ISP tuning conducted in this work will not affect these aspects of IQ. Optimizing the imaging pipeline with respect to the rest of the VCX criteria is left as future work.

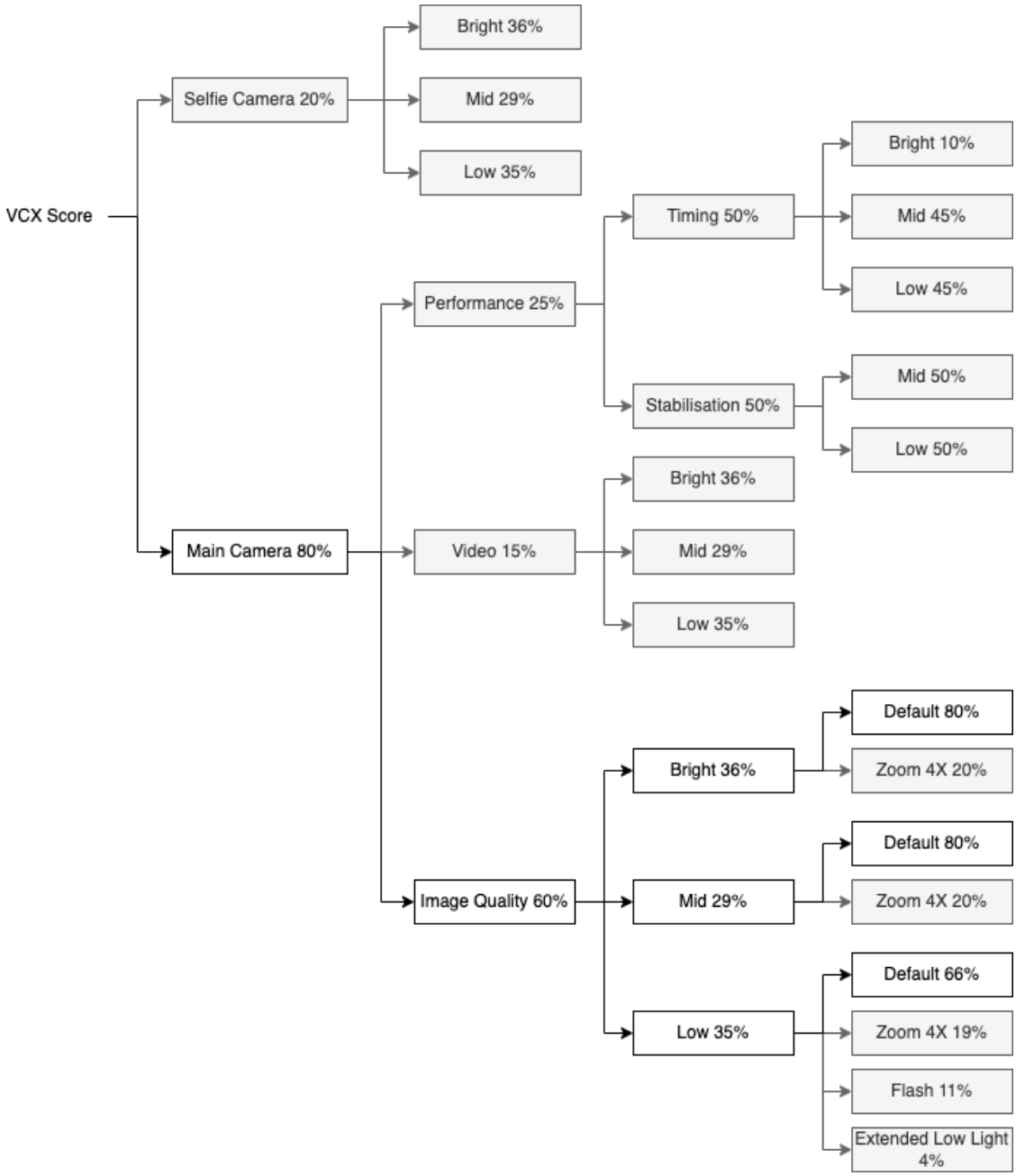

Figure 3.5. *Weight graph of total VCX score calculation [50].*

4. DESCRIPTION OF OPERATING ENVIRONMENT

In this chapter, the basic operating environment is described. Every part of the operating environment contributes to image quality and IQ tuning in their own way. These ways are discussed in the following sections. First, a brief introduction to the fundamental operations of the compact camera module (CCM) is given in 4.1 and then to the fundamentals of ISP in 4.2. Lastly, the laboratory environment is covered in section 4.3.

4.1 Compact Camera Module

Compact camera modules are small, integrated imaging devices that can be used in a wide range of applications. They are widely used in consumer electronics, like smartphones, laptops, and tablets. CCMs have become increasingly popular in recent years due to their ability to capture high-quality images and videos in a compact and convenient form. The camera module used in this work is **Sony's IMX586 CMOS image sensor**. It has 48 effective megapixels paired with a 0.8 micrometer pixel size [45]. [27]

One major advantage of compact camera modules is their size, making them ideal for use in applications where space is at a premium, like in smartphones [27]. Compact camera modules also have high resolution and can capture videos with a high frame rate [29]. They also have features such as autofocus and image stabilization, which help to capture sharp and clear images and videos even in challenging lighting conditions. Some also come with additional functionalities like infrared imaging or thermal imaging, which makes them useful for various fields such as security, medical, and industrial inspection.

There are many different types of compact camera modules available on the market, including complementary metal-oxide semiconductor (CMOS) and charge-coupled device (CCD) sensors. Each type has its own advantages and disadvantages, and the choice of the sensor will depend on the specific application for which the camera module is intended. For example, if an imaging device is often used in poor lighting conditions, a manufacturer may choose a CCD sensor over CMOS. This is because the CMOS image sensor noise is starting to be more visible in poor lighting conditions. [29]

As stated before, the CCM used in a mobile device contributes largely to the perceived IQ. A CCM consists generally of a lens, a lens holder, a voice coil motor (denoted as AF

Motor in the figure 4.1), an infrared cut filter, an image sensor, and a printed circuit board. The structure of one CCM is displayed in figure 4.1. [28]

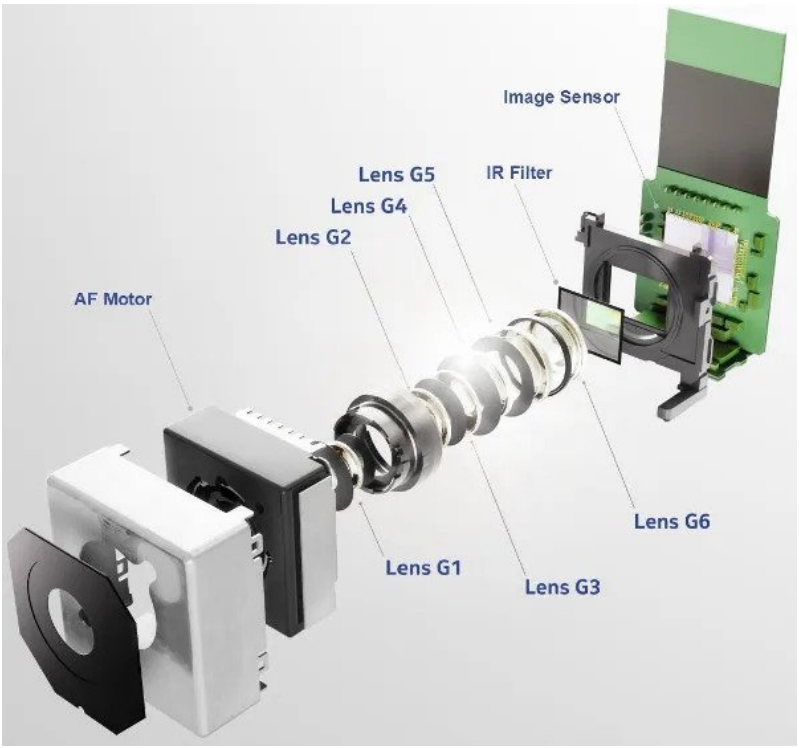

Figure 4.1. *Structure of compact camera module. [6]*

The general purpose of the lens holder is to hold all the optical components, and lenses, inside of it. A voice coil motor takes care of auto-focusing the image on the image sensor plane. The voice coil motor moves the position of the lenses to make the contents of an image appear clear and sharp. Since image sensors are capable of capturing near-infrared light that is invisible to humans, and near-infrared light also distorts the colors of an image captured under daylight conditions, these wavelengths are filtered off with an infrared cut filter. Infrared light has a wavelength ranging from 700nm to 1000nm on the electromagnet spectrum. A digital sensor consists of an array of millions of pixels that are responsible to collect photons during the exposure of an image. After exposing the relative amount of photons that fell into each pixel is converted to an intensity level of an electrical signal. [28], [30], [52]

Lens and image sensor has the largest impact on the perceived IQ. As an electronic device, the image sensor also introduces noise on top of the signal. If the sensor has a high SNR the noise comes effectively nonexistent. If SNR is low, the level of the background noise of an image sensor is high, and disturbing noise has to be removed with digital image processing. This disturbing background noise is often highly visible e.g. on smooth surfaces and when images are taken under dim lighting conditions. The lack of light is compensated by increased ISO on the image sensor. Higher ISO sensitivity will make the image sensor more sensitive to light but also increase the amount of noise in the

resulting image. The noise can be divided into luminance and chrominance noise (*luma and chroma noise*). Chroma noise means color noise which is often more disturbing than luma noise. Figure 4.2 illustrates these two components of the noise. [30], [37]

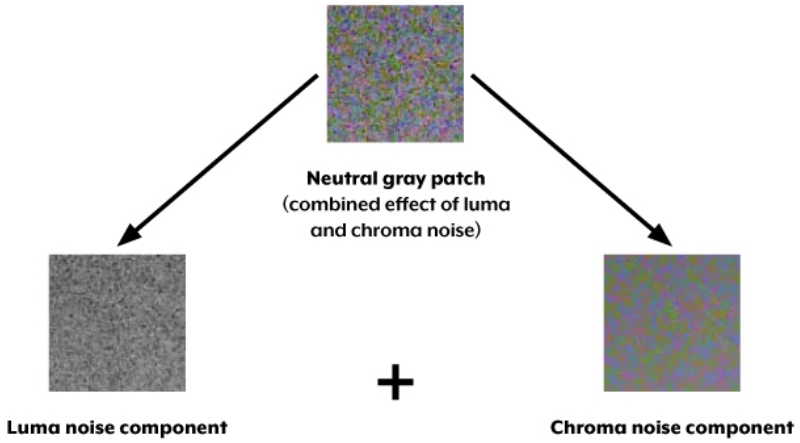

Figure 4.2. *Example of chroma and luma noise. [30]*

A smaller sensor size leads to a smaller pixel size on the sensor. Smaller pixels can not receive as much light as larger pixels on full-size digital cameras. In addition to worse noise performance, it also leads to a worse dynamic range. A low dynamic range makes mobile devices more sensitive to under and over-exposing of images. Limited space on the device also restricts the aperture size, zooming capabilities, and bokeh effect. The Bokeh effect means keeping the target in focus and sharp and blurring other areas of the image. Zooming is often implemented using different lenses for different zoom levels combined with digital zooming. The artifacts introduced by these necessary compromises are attempted to be corrected with image signal processing algorithms and methods. [31]

The optical quality of the lens means the quality of its materials, coatings, alignment of different lens components, and so on. Lens has a large impact on the perceived sharpness of an image. Resolution defines the ability of a lens to render fine details, such as hair or fur. Images will be unsharp and soft if a lens has a bad resolution. On the other hand, acutance means the lens's ability to render fine edge sharpness and sharp transitions between areas with different contrasts. Sharpness, resolution, and acutance can also be addressed with one single term: clarity. A lens with low optical quality can introduce severe optical aberrations to the image that has to be then removed digitally with image processing to improve the IQ. A few examples of such aberrations are image blurring, reduced contrast, misalignment of colors (*chromatic aberration*), distortion, and vignetting. Figure 4.3 displays the effect of sharpening on a digital image. [13], [30], [37]

In conclusion, compact camera modules are small, integrated imaging devices that have become increasingly popular in recent years due to their ability to capture high-quality images and videos in a compact and convenient form. They have many advantages including size, resolution, and versatility, and can be used in a wide range of applications.

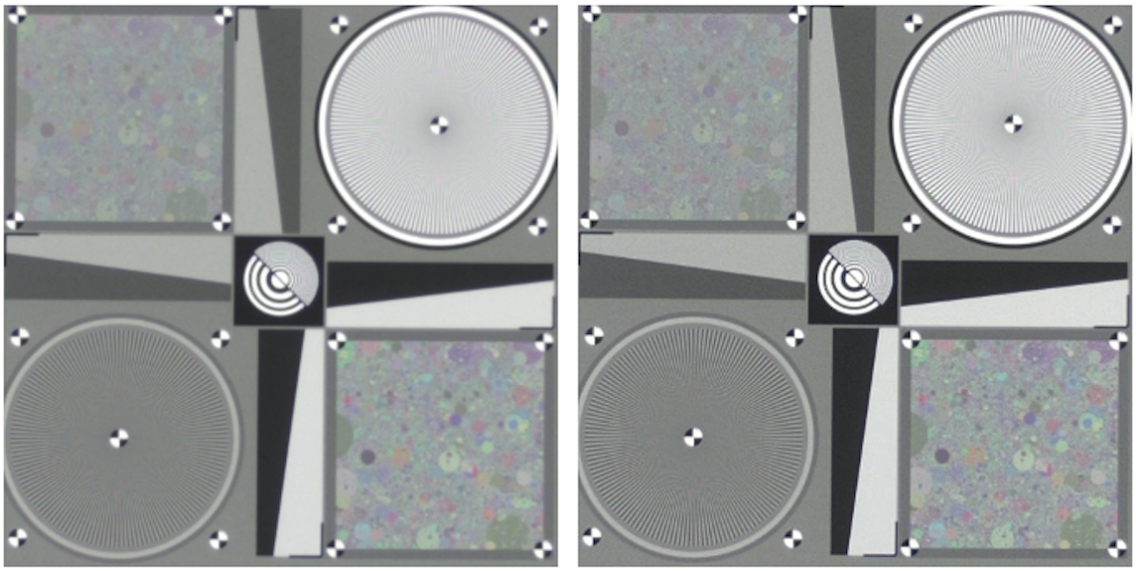

Figure 4.3. *The effect of sharpening. The left image is unsharp and the right one is the same image after sharpening.*

Many artifacts that affect the IQ of an image captured with CCM can be addressed with digital signal processing. This processing is done by the ISP and will be discussed briefly in the next section.

4.2 Image Signal Processor

One very essential component of a digital imaging device is an image signal processor. ISP is responsible for converting the RAW image captured by the CCM to a processed final image. The ISP used in this work is **Qualcomm Spectra 580** which comes with Snapdragon 888 (SM8350) -chipset. It is a modern ISP that is capable to support up to 200 megapixel image capture [40]. A common ISP pipeline is divided into separate blocks where each block takes care of one individual task. Commonly known tasks are noise reduction, adjusting white balance, demosaicing, color correction, contrast, and edge enhancement (sharpening). This image-processing pipeline has a huge effect on the final IQ. Figure 4.4 shows a simplification of a pipeline that is assumed in the scope of this work. [33], [41]

The early cameras on mobile devices produced very limited IQ compared to analog and digital cameras. The reason is the requirement for a small form factor and thus the small size of lenses and imaging sensors. This results in a relatively small amount of light available for capturing the image. To achieve better IQ, image processing techniques are required. Modern smartphones pack a lot of computing power, resulting in the ability to implement better and faster image-processing algorithms. An ISP is a dedicated piece of hardware or software that takes care of processing the digital signal acquired from the camera sensor to produce a more pleasing image for the intended use case. Next, the

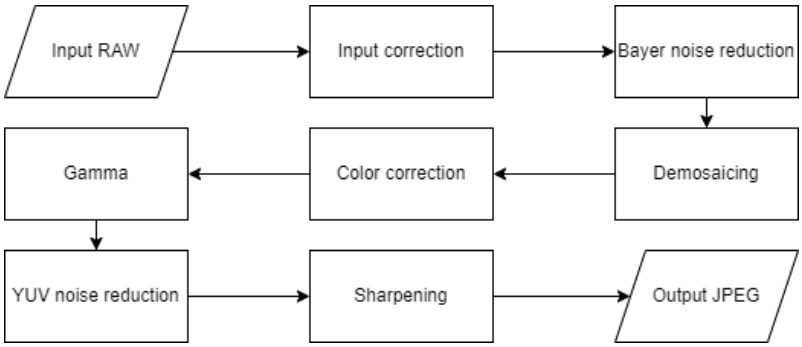

Figure 4.4. *A simplified ISP pipeline. [33]*

different ISP blocks presented in figure 4.4 are covered briefly. [31]

Input Correction

The main corrections done for the RAW data are lens shading correction, black-level correction, and bad pixel correction. Lens shading means a light falloff when moved from the center to the corner of the image. Shading originates from the lens system's radial nature to collect more light in the center [21]. This is known as luminance shading. Color shading is a different phenomenon, where a shift in color from the center to the corner of the image is visible [10]. It can be caused by the image sensor itself or by the optics. One integral part contributing to color shading in the optics is the IR-cut filter.

The output of a digital image sensor is not completely zero, even if no light reaches it. Black-level correction is needed to unify the black pixel's intensity level across the image plane. In addition to varying as a function of spatial location, the black-level may also depend on analog gain and exposure time. Black-level correction is important for accurate color and intensity level reproduction. [34]

Bad pixels can be seen as spatially varying noise caused by nonuniformly responsive pixels on the image sensor [5]. Some pixels on the sensor might be totally unresponsive or extremely responsive causing nonuniformities. These pixels must be detected and taken care of before feeding the RAW data forward in the image processing pipeline.

Bayer Noise Reduction and Demosaicing

In order to capture color images, the image sensor has a color filter array on top of it. This allows capturing the light of only one of the three primary colors, red, green, or blue, into each cavity on the sensor. The most common type of color filter array is a Bayer array which is presented in figure 4.5. Bayer array contains two times more green pixels than any other since the human eye is the most sensitive to green light. Due to this color filter array, the image RAW data consists only out of red, green, and blue pixels. [30]

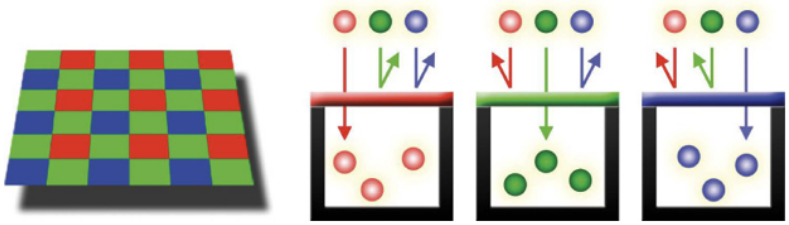

Figure 4.5. *A Bayer array that is used in a digital camera to capture color images. [30]*

Noise reduction that is done for the RAW image data consisting only of three primary color pixels is called Bayer noise reduction. This is done to prevent undesired color cast and color errors further down the image processing pipeline [8]. Demosaicing is also known as color interpolation. It is a technique used to convert the data from a color filter array domain into a multi-component-per-pixel domain [8]. This allows evening out the spatial resolution between different color channels.

Color and Gamma Correction

Color correction is an important step in order to reach a good subjective IQ. Here different aspects of the colors will be modified, like saturation. Also, faulty reproduced colors will be corrected. The CCM's color space can be arbitrary. It is often mapped to a well-known sRGB color space to introduce accurate colors. Another example of a well-known multi-component-per-pixel color domain is the luma/blue-difference/red-difference (YUV) color domain.

Gamma correction is a non-linear process where the pixel intensities of the image are modified to better match how humans perceive the changes in light intensity levels [39]. Gamma correction changes the light intensity levels in the image. By doing so, it affects the DR of the imaging device.

YUV Noise Reduction and Sharpening

Noise reduction is also done in the YUV domain, mainly to address luminance (Y) and chrominance (UV) noise with different techniques. An example of different noise components is presented in figure 4.2. One common approach in digital image processing is to use linear low-pass filtering. These averaging filters calculate the new value of a processed pixel as a weighted average of the neighboring pixel values [15, p. 119]. This will reduce sharp transitions of the pixel intensities in the processed image.

After filtering, the image may appear too blurry. Local contrast enhancement around the edges is done to make the image look crisp again. Multiple noise filtering and sharpening techniques exist, and it is ISP-vendor dependent which are used. One common method is unsharp masking. Unsharp masking is a technique where an image is sharpened by

subtracting a blurred version of itself from it [15, p. 132].

Additional Methods

One main approach to correct the noise level or the zoom quality of an image is to use multi-frame noise reduction (MFNR). MFNR means that the device captures multiple images in a short period of time, and fuses them together using different algorithms. The main problem of the MFNR is the misalignment of preceding frames. This introduces a ghosting artifact to the image where a moving object is trailed by a ghost of itself in the final image. This can be tackled with special algorithms and using a lower shutter speed. It has to be kept in mind that using too high shutter speed prevents enough light from hitting the sensor. This forces the gain of the imaging system to be set higher. The gain amplifies the noise, making images appear less pleasing. Fusing multiple images together and averaging the pixel values is the most effective and simplest way to reduce the noise. [31]

High dynamic range (HDR) -compression is a method to enlarge the dynamic range of the imaging device. DR means the range from the light intensity at the darkest part of the scene to the highlights. The DR of a scene can be too high for the sensor and thus it is not able to reproduce shadows and highlights of it. Human eyes have a very wide DR and images captured with low dynamic range CCMs can come off as unnatural to a human observer. The basic principle of HDR is to capture multiple images with different exposures and fuse them together. This allows the highlights of an image to be retained without making the image too dark overall. [31]

Algorithms and methods are very ISP-vendor dependent. This section only introduced a few common functionalities that are needed to obtain a decent IQ in the eyes of modern-day requirements. IQ tuning, described in the section 2.3, aims to find a good set of parameters for the algorithms executed in the ISP to obtain an acceptable IQ. The requirements are also use-case and end-user-dependent, and not always the produced image meant for a human being or for artistic purposes. For example, computer vision applications may care more about the distinctiveness of the objects in the image and less about the colors.

4.3 Laboratory Environment

The VCX standard expects certain pieces of laboratory equipment. These are defined in the VCX Whitepaper [50]. The whole laboratory must be temperature controlled to a standard room temperature 23 °C $\pm$ 2 °C. The device-under-test must be mounted on a tripod and also the reflective test charts have to be mounted properly. The test charts used in the VCX score measurement are described in section 3.2.

The illumination of the test chart is also very important. The illumination must be uniform across the whole chart. This is achieved by illuminating the chart from left and right using two LED light sources. This way no reflections are introduced which could affect IQ measurements. Also, multiple light levels with different color temperatures can be achieved using only one LED light source. The properties of used illuminants are described in the section 3.2. An image of the laboratory used in this work is displayed in figure 4.6 below.

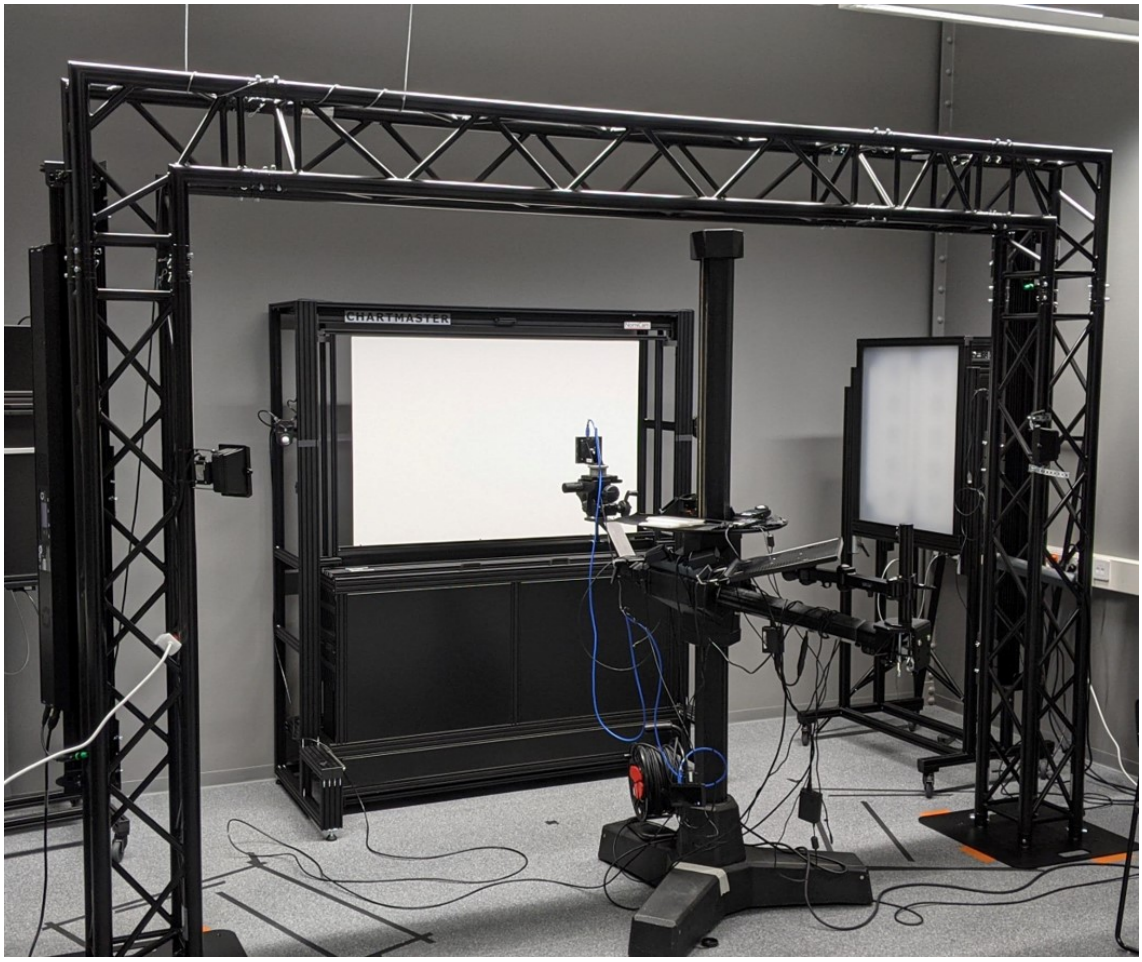

Figure 4.6. *The laboratory environment where the RAW data was captured.*

The DUT's distance to the test chart is defined by the field of view of the device. The test chart is framed so, that it fills the largest possible area of the frame so that all the content of the chart is still visible. The distance between the chart and the phone varies between different smart devices.

5. METHODS

Previous chapters have elaborated on the theory behind the work done. This chapter will focus on the methods used to automatically optimize the image quality and assess it during and after the ISP tuning efforts. During the ISP tuning procedure, the IQ is assessed objectively with a VCX score, deep learning-based method, and with their combination producing three different IQ tuning solutions. After the optimizations are done, a subjective IQ study is conducted to rank the final ISP tuning results.

5.1 Automatic Image Quality Tuning Framework

Optimization problems using images can grow up to be highly computationally expensive because the amount of data, the number of pixels, in modern imaging devices is enormous and ever-increasing. Often these kinds of optimization dilemmas are nonlinear which also contributes to the overall complexity of the problem. Optimizing the IQ produced by the ISP in a traditional manual way is a tedious and time-consuming task since ISPs generally have several parameters that must be modified by the IQ engineer by hand. Perfecting the IQ can be seen as an optimization problem that has a very high dimensionality. This work introduces an automatic image quality optimizer solution that takes care of this laborious task and optimizes the IQ according to certain targets or objective functions. In this work, the objective functions are IQ measurement methods whose outcome is attempted to be maximized for a certain ISP and CCM. Different objective functions used to measure the IQ are covered in section 5.3. [38]

The ISP used in this work is Qualcomm Snapdragon 888 (SM8350) combined with Sony's IMX586 48-megapixel CMOS image sensor. To simulate the effects of the ISP parameter changes on the perceived IQ, Qualcomm's Spectra 580 simulator is used. The simulator output matches very closely to the actual output of the ISP given the same input RAW image. The optimization framework is implemented with Python version 3.7.7. This implementation is derived from the framework introduced by Nishimura et al. [33].

The automatic ISP parameter tuning framework used in this work is utilizing a bio-inspired swarm intelligence-based ABC algorithm for global ISP parameter search to avoid getting stuck to arbitrary local minima [33]. ABC algorithm is an optimization algorithm mimicking the intelligent behavior of a honey bee swarm. It is suitable for optimizing multivariable

functions, like the function of the ISP where the input is a RAW image and the output is a processed final image. [23]

The local optimization is done using a subplex algorithm. This way a higher precision of obtained solution is achieved [33]. The combination of these two algorithms has been shown to work stably with this high-dimensional, discontinued, nonlinear, and noisy optimization problem. Regardless of the starting point of the optimization, the resulting subjective IQ converges to a similar result given the same objective function and IQ target. It is important to note that similar IQ can be achieved with multiple different combinations of ISP parameters. Thus, the ISP parameters may not be similar between different solutions but the output of the objective function (perceived IQ) is similar. This can happen due to the similarity of different ISP blocks. For example, multiple blocks in the pipeline take care of noise reduction in different ways. Many combinations of different parameters will then result in a similar IQ.

Figure 5.2 illustrates the resulting IQ of three separate optimization runs, using the same objective functions and IQ target. Only the starting point of the optimization, the initial ISP parameters, are different between the runs. The subjective IQ is very close between different results even though the final ISP parameter sets differ. The final ISP parameter sets of these results are displayed in figure 5.1.

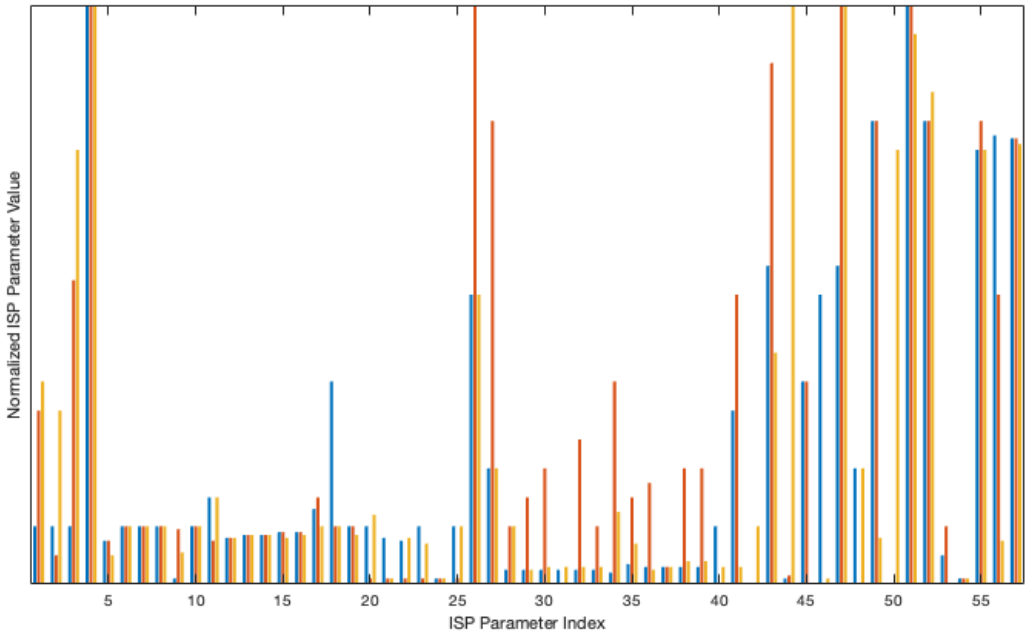

Figure 5.1. *Comparison of final ISP parameter sets after three individual optimization runs using the same final IQ target and different starting points.*

This result gives confidence in the robustness of the automatic tuning framework. Despite the different starting points of the optimizations, the framework is able to find a solution with the best subjective IQ.

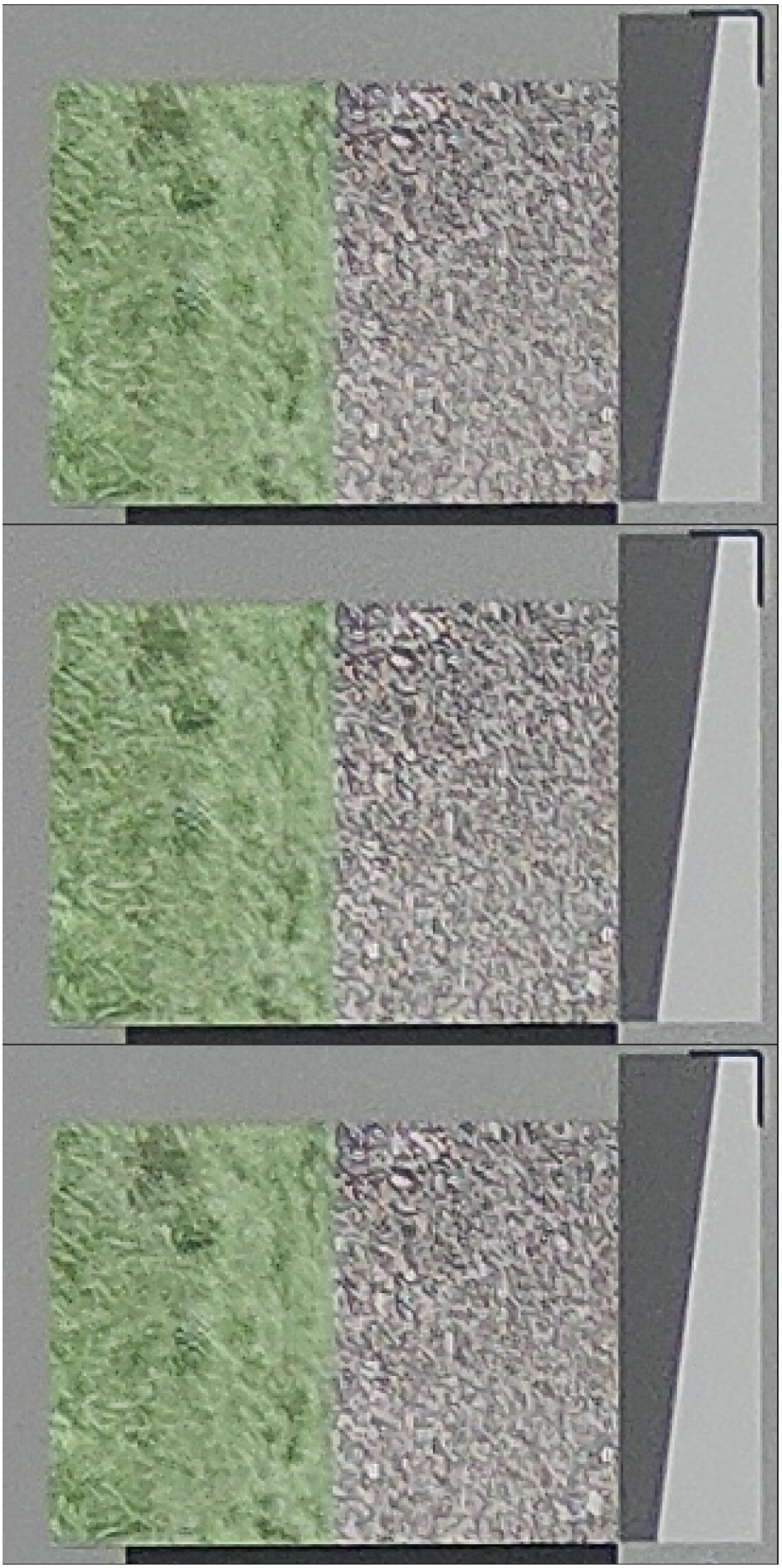

Figure 5.2. *Comparison of texture target after three individual optimization runs using the same final IQ target and different starting points.*

Overview of the Automatic ISP Parameter Tuning Process

Before the automatic tuning process can be initiated, some preliminary steps are required to perform. Firstly, RAW data needs to be captured using the camera module that is being used together with the optimized ISP. This RAW data is used as input data to the image processing pipeline. The data capture process is performed under illuminations and guidelines defined in the VCX-Standard which are summarized in section 3.2. It is important to follow these instructions during the process as closely as possible. Secondly, the camera module of the DUT must be characterized. Characterization is done to set certain parameters in the ISP. For example, the lens shading correction is dependent on the quality of the optics. Measuring the lens shading profile accurately allows near-perfect correction of it, making the problem diminish. The characterization includes, but is not limited to, calibration of the black level, lens shading, and noise profile of the camera module.

The assumed structure of the ISP pipeline is shown in figure 4.4. Figure 5.3 shows the blocks which are under the scope of optimization. These blocks of the image processing pipeline affect the produced IQ tremendously. These parts are selected by internal IQ engineers with a high amount of expertise with the ISP used in this thesis. Also, demosaicing and gamma blocks affect the IQ and the dynamic range of the DUT tremendously but they are not tuned automatically in the scope of this work. Gamma is tuned manually and demosaicing is left as default. This approach is selected since the default parameter set provided by the ISP vendor works sufficiently well.

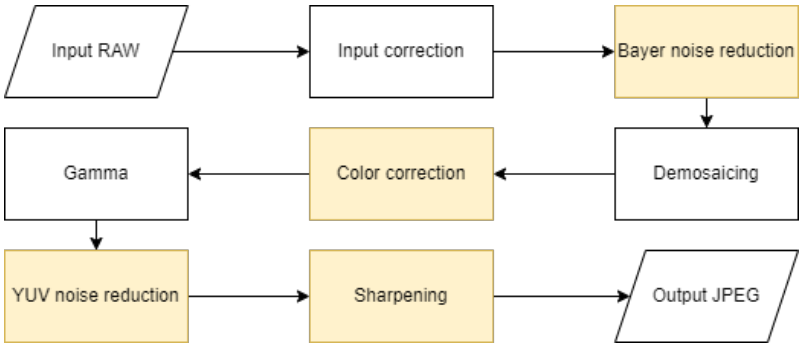

Figure 5.3. *Simplified ISP pipeline, colored blocks are being automatically tuned. [33]*

Figure 5.4 presents the overview of the automatic ISP parameter tuning process. Bolded and underlined sections are essential parts of the automatic ISP parameter tuning software.

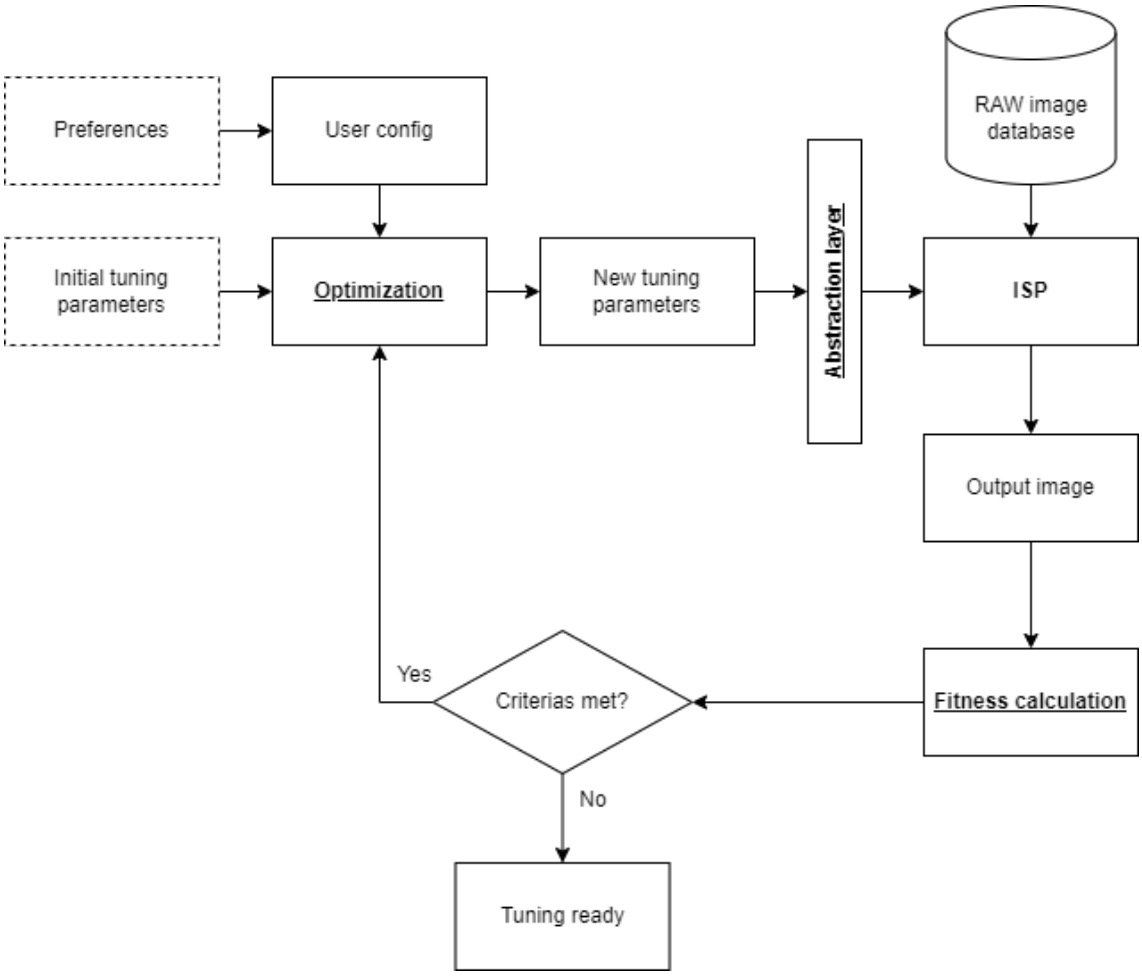

Figure 5.4. *Overview of the automatic ISP parameter tuning process used in this work.*

The optimator is initialized with an initial parameter guess and user configuration. User configuration includes, for example, algorithm-specific parameters, information about objective functions used, and other user IQ preferences. The optimator outputs a set of new tuning parameters which are then fed to the abstraction layer. The following subsection, *Reducing the Problem Space*, covers this part more thoroughly. After the abstraction layer, the parameters are in the ISP vendor-specified format and the resulting image will be generated. Using this image, the objective functions are evaluated. This happens in the *Fitness* block. In the scope of this work, the objective functions, or fitness functions, are the VCX score IQ metrics or the deep learning-based method which correlates well with human perception of IQ. These methods are covered in section 5.2. Using the fitness information, the optimator makes a new guess of optimal tuning parameters if the IQ was not satisfactory and the optimization goes on. Criteria of the termination can be reaching a certain number of iterations, reaching the best possible fitness, or reaching a minima where the fitness is not changing anymore.

Reducing the Problem Space

Since an ISP has several tunable parameters, selecting those that have the most effect on the perceived image quality is important. Also, the parameter ranges on the ISP side can be arbitrary. It is also possible that one set of parameters could be represented with a single parameter. For example, one can define blur kernel coefficients by only defining a cut-off frequency without defining every individual value used in the kernel. In general, the parameters are represented as lookup tables on the ISP side. The blue line in figure 5.5 illustrates a lookup table with 128 individual values. One IQ algorithm on the ISP might consist of multiple parameters defined as different lookup tables containing multiple individual values. This results in a huge dimensionality of the optimization problem. However, lookup tables often follow a geometric shape, for example, a piecewise linear function.

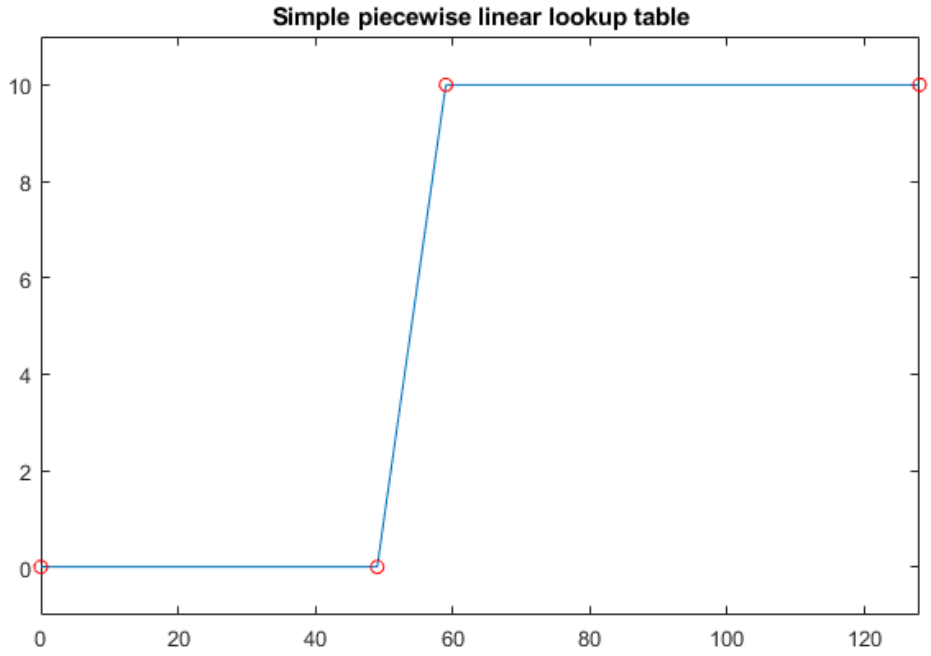

Figure 5.5. *An example of ISP parameter shape. ISP vendor expresses this parameter as a lookup table of 128 values. The automatic ISP tuning framework operates only using 4 parameters represented in the image with red circles.*

To overcome the issues with optimization stated above, the automatic tuning framework introduces an abstraction layer between the ISP and the parameter optimization algorithm. By configuring this abstraction layer, the user can select which ISP parameters to tune from which ISP blocks and how to abstract these parameters. The abstraction implies the selection of the geometrical shape that the lookup table can be represented with. This requires a high level of knowledge about the ISP that is being optimized since the user has to know the shapes of the optimized lookup tables defined by the ISP vendor. This information is usually provided in an ISP tuning guide. The tuning guide explains the parameters, their shapes, and how to tune them. This helps the IQ tuning engineer

in the optimization of the IQ. The tuning guide also explains how to use the ISP tuning software which presents the parameters in a human-readable fashion. Tuning software is an interface between an IQ engineer and an ISP. The software takes care of turning the parameters from complicated ISP parameter sets to tunable and understandable sets. All this is being mimicked by the automatic tuning framework. This way the number of parameters being optimized can be reduced significantly compared to the original.

The piecewise linear lookup table presented in figure 5.5 can be defined with four parameters marked with red circles in the figure. These four parameters are optimized by the automatic ISP tuning framework. Before exporting the parameters to the ISP, the abstraction layer reconstructs the lookup table to the ISP vendor specified format. The ISP vendor has defined the lookup table shown with a set of 128 individual parameters, and optimizing every parameter individually is not desirable. The shape of the lookup table and its length are configured in the abstraction layer but aren't fed to the algorithm.

Another approach to make parameter search more robust and optimal is to limit the total search space. This requires expertise in reasonable parameter ranges. This could also be done with an automatic search algorithm that is not a part of this thesis. The minimum or maximum limit of a certain parameter defined by the ISP vendor might be unfitting to real-world applications. This leads to too extensive search space. With knowledge of the ISP and its behavior, the user is able to reduce the search space by configuring tighter limits for the optimized parameters.

Global Optimization Using Artificial Bee Colony Algorithm

The artificial bee colony (ABC) algorithm is used for global parameter optimization. The target of the ABC algorithm is to find the rough location from the parameter space where the global minima exist [33].

The ABC algorithm [23] divides the bee colony into three different groups: employed, onlooker, and scout bees. A bee waiting to decide which solution (food source) to select from the parameter space is called an onlooker. A bee going to a solution that it has previously visited is called an employed bee, and a bee doing a random search for food sources is called a scout.

For every solution, only one employed bee exists. The number of employed bees is half of the colony and the second half consists of the onlookers. A bee whose solution is discarded as a bad one becomes a scout. The main steps of the ABC algorithm are:

The employed bee goes to a solution, and if the onlooker bee finds a better neighboring solution, the employed bee will become a scout who looks for a randomly new solution in the parameter space. Thus, the ABC algorithm can be divided into four different selection processes:

Algorithm 1 The main steps of ABC algorithm

```
Initialize
while Requirements are not met do
    Place the employed bees on the solutions
    Place the onlooker bees on the solutions
    Send the scouts to search for new solutions
end while
```

1. A global search where artificial onlooker bees discover promising new regions from the whole parameter space.
2. A local search where artificial employed bees and onlooker bees determine the neighboring solutions.
3. A local greedy selection where the candidate solution is selected over the present one if the reward of that solution is better. Otherwise, the candidate solution is discarded and the present one is kept as the best solution.
4. A random selection process carried out by the artificial scout bees.

The three control parameters in the ABC algorithm are the number of solutions which is equal to the number of employed and onlooker bees, the cycle limit for abandoning a solution whose value can not be further improved, and the maximum cycle limit.

Local Optimization Using Subplex Algorithm

The result of global optimization is used as the starting point for the local parameter optimization. The local phase of the ISP parameter optimization utilizes a subplex algorithm. Subplex algorithm suits well to noisy and discontinuous optimization problems. It is a variant of the Nelder-Mead simplex method [32]. In mathematics, simplex is a generalization of a triangle or a tetrahedron to higher dimensions. A simplex in n-dimensional space is defined as a set of $n + 1$ points that are not all collinear. The simplest example of a simplex is a triangle in two-dimensional space.

This algorithm is well suited for derivate-free optimization in high-dimensional problem space because typically the number of function evaluations needed for convergence increases only linearly when the problem grows. The subplex is more efficient for most optimization problems than the traditional simplex method. The general idea is to create a sequence of nested subsets in the problem domain and search for a minimum in each subset. At each optimization iteration, a subset of the simplex's vertices is used to define the search direction. After the direction is defined, a line search along that direction is performed. [26], [32], [33]

5.2 Image Quality Assessment Methods

This section elaborates on the IQA methods used in this work. VCX is used as the objective IQA method during the ISP tuning procedure. Also, a deep learning method called Deep Image Structure and Texture Similarity (DISTS) Metric is introduced to evaluate the IQ in an alternative way [9]. Forced-choice method is the subjective IQA method used after the IQ tuning to determine how different tuning solutions compare to one another.

VCX as Objective Image Quality Assessment Method

During the IQ optimization, the ISP parameters are tuned with the framework defined above in section 5.1. During the optimization, VCX metrics are used to assess the resulting IQ. This section explains the method of how these objective IQ metrics are exploited in the prediction of subjective IQ.

In objective IQA it is fundamental that the results correlate with subjective analysis. VCX standard scores the camera's overall performance from 0 to 100, where 0 stands for the worst possible subjective stimulus and 100 for the best. As stated before, one part of this score is the IQ of the primary camera which this thesis focuses on. The primary camera's IQ score is combined from measurements under multiple illumination levels, every illumination level including at least default and zoomed captures. From each capture session, for example, using 4X zoom under mid-illumination, multiple different IQ metrics are calculated. These separate IQ metrics are fused into one score using custom weighing functions. With these functions every separate IQ metric, for example, the amount of artifacts present in the dead leave target, illustrated in figure A.9, is scaled between [-1.0, 1.0].

The scaling functions provided by the VCX standard imitate how a human being would react to an image with a specific IQ metric value. For example, if an image has a very low maximum texture SFR value, it means that the image is blurry and the textures are not reproduced in a good manner. Subjectively IQ is not treated as a good one; thus, the VCX metric defining the texture SFR is scaled to be near the minimum -1.0. Figure 5.6 illustrates an example of an aforementioned weighing function. Note that it is not a real weighing function used by the VCX standard.

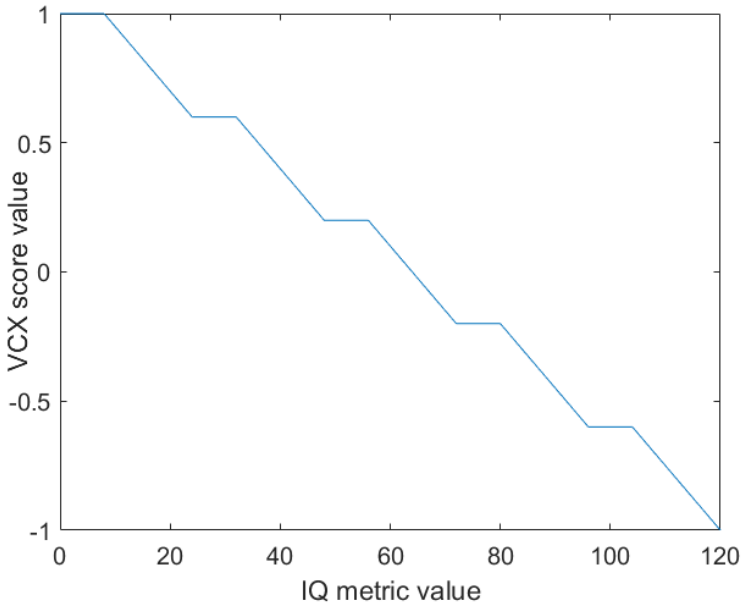

Figure 5.6. *An arbitrary example of a VCX score component weighing function which maps the result of an image quality metric between -1.0 and 1.0.*

Furthermore, every individual part of the score, for example, texture loss or colors which consist of multiple IQ metrics, is scaled between [-1.0, 1.0]. This is done by weighing the separate VCX metrics after scaling them. Combining every individual part of the scores calculated in different shooting conditions, defined in figure 3.5, will result in one single numerical value describing the overall quality of the camera. The score is often multiplied by 100 and rounded to the closest integer value before reporting.

DISTS as Image Quality Assessment Method

Deep Image Structure and Texture Similarity is a deep learning-based method that aims to mimic the human IQ evaluation. It is also full-reference-based, meaning that the quality of an image is defined by comparing the analyzed image to a perfect, artifact-free, reference image. Traditionally, pixel values of degraded images have been compared to the reference by using traditional quality metrics like a mean squared error. These kinds of metrics correlate poorly with human image quality perception and they are also dependent on the perfect alignment of the compared images. [9]

The proposed approach constructs a function that transforms images into multi-scale representations using a convolutional neural network. By doing so, some degree of texture resampling is achieved. The network is able to capture the texture appearance and form an IQ score between analyzed and reference images representing the similarity of the textures by comparing the texture properties and the structural details. Also, the analyzed and reference images do not have to be perfectly aligned since it is shown that the model

is robust for mild geometric transformations. [9]

DISTS is used with the automatic tuning framework to emphasize the importance of texture reproduction. Extensive texture loss degrades the subjective IQ [50]. Past experiments have shown that using DISTS with the automatic tuning framework exploited in this thesis has given good results considering the final IQ.

Subjective Image Quality Assessment with Forced-Choice Method

Because the goal of this thesis is to produce an image quality tuning solution that produces subjectively good images, it is important the final IQ is also judged subjectively and not only using the objective VCX score. It is possible to evaluate subjective IQ as predictable and repeatable when following the methodologies of psychophysics. Psychophysics studies the relationships between the perceptual and physical realms. In this section, a psychophysical technique that is used in this work will be explained. [36, p. 117]

The subjective IQA study will be conducted in a controlled environment, where the ambient lighting stays the same, the screen used to observe the images is calibrated, and the viewing distance to the screen is regulated. This way the error in the results caused by the environment is minimized.

The psychophysical technique selected is called forced-choice or pairwise comparison. In forced-choice comparison, the observer is presented with image pairs and they have to choose which one they prefer more. The drawback of the method is its time-consuming nature. Every image needs to be paired with every other test sample in order for the study to be fully experimental. For n images the amount of comparisons is $\frac{n(n-1)}{2}$. This experiment suits the needs of the thesis since every compared image has the same content, and the only differences will be in, for example, the sharpness. [36, p. 125]

The forced-choice comparison is based on the Thurstonian scaling -method [47]. During the test, the observer will be displayed two different images at a time. They have to choose, which one they prefer more and their preferences are recorded by the software used to conduct the test. After displaying every image paired with every other, the assessment is done. Thurstonian scaling -method assumes that the repeated judgment of the images produces a response following a normal distribution. Also, the variation in assessments between two images is assumed to follow a normal distribution. [36, p. 130]

A sample size of at least 30 respondents is selected for the test. Selecting a sample size of 30 has been shown to produce statistically significant results and increasing it would not affect drastically to mean preference score value (MPS) [4, p. 45]. An MPS is aggregated preference score divided by the number of respondents. An aggregated preference score states the number of times an item was chosen across all paired comparisons made by the respondents [4, p. 10]. It tells how many times an image was selected over all other

images in the same scene. Here a scene means the illumination level where the image was captured. There are three scenes: bright, mid, and low. It has to be noted that the results of this pairwise comparison test are not scaled and thus not comparable to any other subjective IQ test.

Table 5.1 presents an arbitrary response matrix of 33 respondents. A matrix of this kind will be produced as a result of the pairwise comparison for each scene and also overall. Sum-row represents the aggregated preference score for a certain item. The matrix is read column-wise and it tells how many times, for example, item 1 was selected over items 2, 3, 4, and 5. The proportion percentage states how many times an item was selected relative to the maximum possible number of times [4, p. 11]. The maximum number of selections per item here is $33 \times 4 = 132$ since the sample size is 33.

Table 5.1. *Arbitrary example of a response matrix. The matrix is read column-wise and it tells how many times an image was selected over others. The mean preference score (MPS) tells how many times on average a respondent selected an image over other options. Proportion stages the percentage of an image selected relative to the maximum selection amount.*

	1	**2**	**3**	**4**	**5**
1	-	21	20	24	28
2	12	-	8	18	17
3	13	25	-	22	24
4	19	15	11	-	25
5	5	16	9	8	-
Sum	39	77	48	72	94
MPS	1.18	2.33	1.45	2.18	2.85
Proportion	0.30	0.58	0.36	0.55	0.71

Images will only differ by the tunings of the ISP used to process them and the images with different illumination levels are not compared. After the subjective IQA, a single ISP tuning is selected as the best one from bright, mid, and low illumination level images. Also, the ISP tuning with the best overall performance will be stated. This way, a subjective understanding of the best ISP tuning will be formed.

5.3 Image Quality Tuning Approaches

Three different IQ tunings will be completed using different sets of objective functions. The alternative IQ tunings, the second and third approaches, will be generated based on

the expertise of the automatic IQ tuning framework and commonly meaningful subjective IQ attributes.

The first approach will be using only VCX metrics with weighing factors defined by the VCX. The weights will be scaled so, that a score of 1.0 is achievable but the relations of the contents, for example, the texture metrics weight relation to the weight of sharpness metrics, will stay the same. Not all VCX metrics will be evaluated during the optimization run because the tuning framework will not affect every aspect that the VCX score is measuring. Evaluated metric categories are texture loss, color, sharpness, resolution, and visual noise. This tuning approach is noted with an identifier 002 in the list of different tuning approaches presented in table 6.1.

The second tuning approach fuses the VCX score and Deep Image Structure and Texture Similarity (DISTS) Metric together [9]. Since texture reproduction is important, texture content and slanted edges will be assessed using the DISTS network to reach a better level of textures. Color, visual noise, and resolution metrics from the VCX metrics will be used to assess other aspects of the IQ. This tuning approach is noted with an identifier 003 in the list of different tuning approaches presented in table 6.1.

The third and final tuning approach, the VCX score will be discarded. Only DISTS will be used to analyze the IQ. Compared to the second tuning approach, also the luma and chroma noise levels and resolution will be analyzed using DISTS. This tuning approach is noted with an identifier 004 in the list of different tuning approaches presented in table 6.1.

This thesis will also introduce two different tunings as the baseline. One of the baselines will be the default IQ tuning solution provided by Qualcomm with the ISP. The default tuning solution is noted with an identifier 000 in the list of different tuning approaches presented in table 6.1. Another baseline will be a commercial smartphone, OnePlus 7T. Oneplus 7T uses the same CCM, **Sony's IMX586 CMOS image sensor**, as is used in the tuned device [16]. The ISP of the OnePlus 7T is **Spectra 380**, which is older and less powerful than the **Spectra 580** used in the device [16]. OnePlus baseline is marked with an identifier 001 in table 6.1.

6. EXPERIMENTS AND DISCUSSION

This chapter will go through different image quality tuning experiments and the subjective image quality evaluation done for the tuning efforts. Section 6.1 with section 6.2 aims to answer the research questions: Will tuning the ISP in order to maximize the outcome of VCX score metrics produce better results than using alternative ways? The section 5.3 explains and covers alternative ways. A subjective IQ test will be conducted, described in section 2.2, in a controlled environment to determine which tuning result reaches the best overall quality. Section 6.2 presents the results to answer the second research question: Will optimizing ISP parameters using fully objective metrics produce subjectively good image quality?

6.1 Image Quality Tuning Experiments

In this thesis, three different IQ tuning solutions were generated fully automatically using an automatic IQ tuning framework described in section 5.1. The primary target of these tunings was to determine which IQA methods worked the best as the objective function of the automatic IQ tuning framework. As a baseline, the default IQ tuning supplied by Qualcomm will be used. Also, a commercial IQ tuning will be used as a reference. This commercial baseline is OnePlus 7T smartphone. Oneplus 7T was selected since it uses the same CCM, **Sony's IMX586 CMOS image sensor**, as is used in the other device [16]. The ISP of the OnePlus 7T is **Spectra 380**, which is older and less powerful than the **Spectra 580** used in the tuned device [16]. Table 6.1 presents a summary of different IQ tuning approaches and gives every IQ tuning an identifier. When IQ tunings are compared these identifiers will be used. These tuning approaches are explained more extensively in section 5.3.

Table 6.1. *Identifier and a brief summary of different image quality tuning approaches.*

IQ tuning solution identifier	Description
000	Qualcomm's default tunings
001	OnePlus 7T commercial reference
002	Tuned using only VCX metrics as target
003	Tuned using VCX and DISTS as target
004	Tuned using only DISTS as target

IQ will be verified with a special TE42-LL test chart. The verification chart has more complex content, like portrait images and flowers, than the regular TE42-LL. This aims to test the IQ more thoroughly. Figure 6.1 presents the test chart used. The chart is captured under the same illuminations defined in the VCX score.

Figure 6.1. *The verification test chart used to assess the final image quality.*

All the different metrics used in the figures 6.2–6.11 are defined in a more elaborated manner in appendix A. Only the results of different IQ metrics are covered here.

Visual Noise

Figures B.1–B.5 present the final IQ of every individual tuning. The tunings 002, 003, and 004 generated in this work have a very low level of visual noise compared to the VN levels of 000 and 001 tunings. Preserving some of the VN allows the ISP to better reproduce textured content. Too strong noise filtering will destroy textures and it is challenging to recover them in the later stages of the image processing pipeline. This phenomenon is

visible from the results of 002 and 003 solutions. Figure 6.2 presents the level of texture acutance calculated under three illuminations from all of the results.

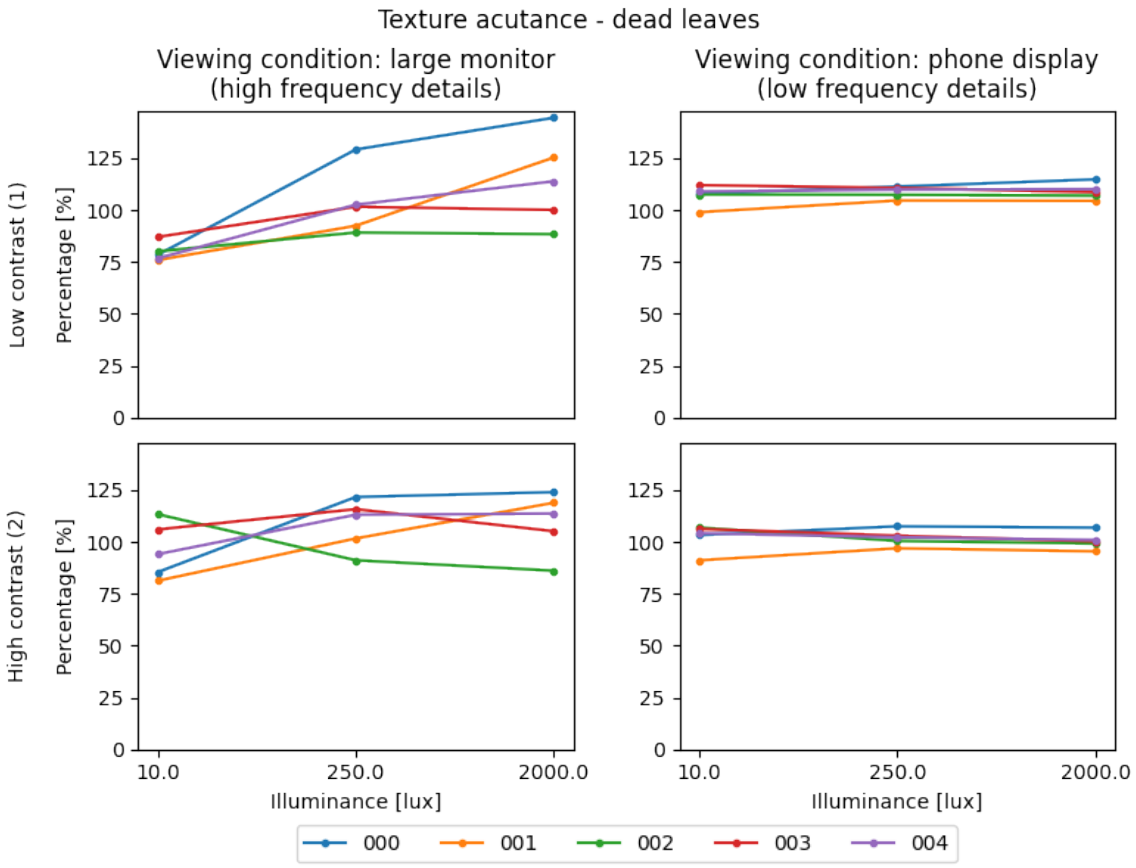

Figure 6.2. *Texture acutance comparison between tuning solutions.*

Figure 6.3 presents the level of VN across all the tunings, calculated from the same images. Tuning solutions 002 and 003 have the best noise performance. The noise filters of these specific tuning solutions were tuned using the VCX score's VN-metric as the target. The lower the VN level, the better the score [50]. This is one factor why the automatic tuning framework converged to such strong noise filters.

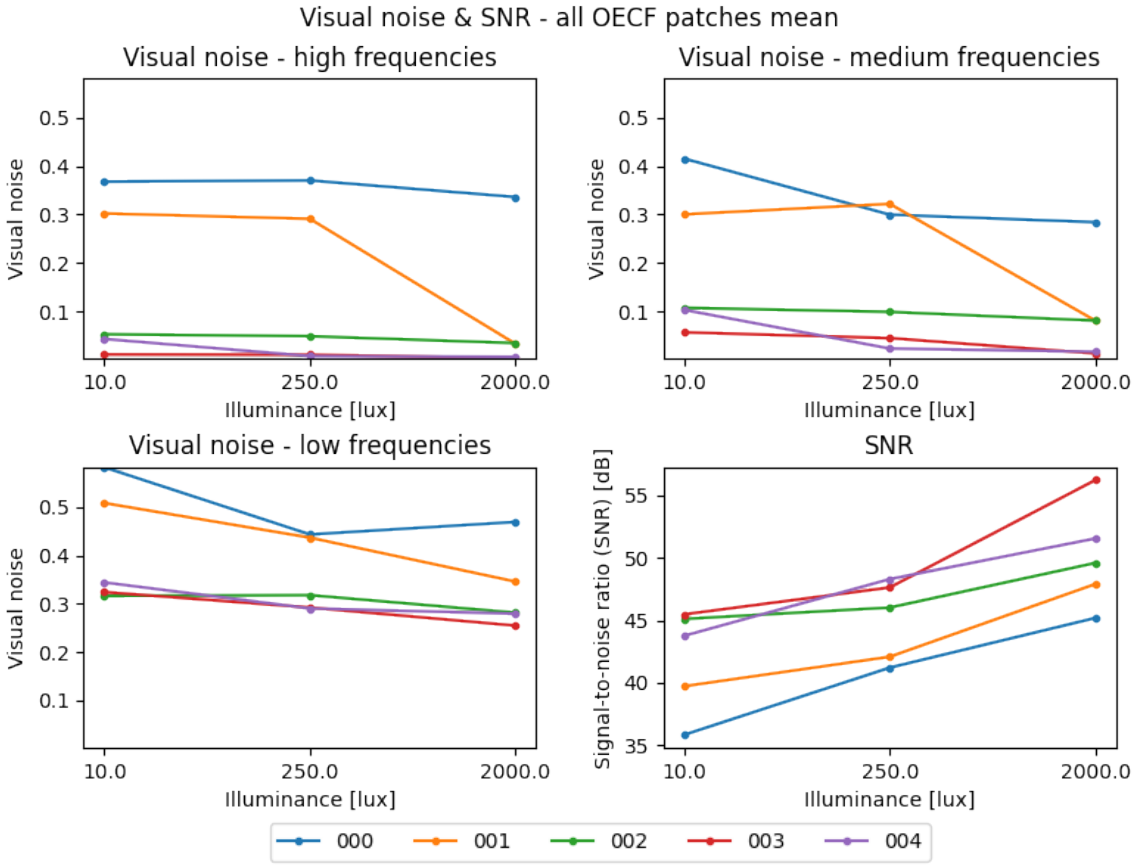

Figure 6.3. *Visual noise comparison between tuning solutions.*

Figure 6.4 presents the VN differences between the tuning solution 000 and 002 under bright illumination. The image generated using solution 000 is on the left. The better noise performance of the tuning solution 002 is clearly visible. If high noise performance is not crucial for the use case of the camera, it might be a good idea to leave some noise on the flats. This might ensure a slightly better texture reproduction. Also, uniform noise on the flat areas of the image is rarely too disturbing for the end user. This is a trade-off that the tuning engineer must do during the IQ tuning process.

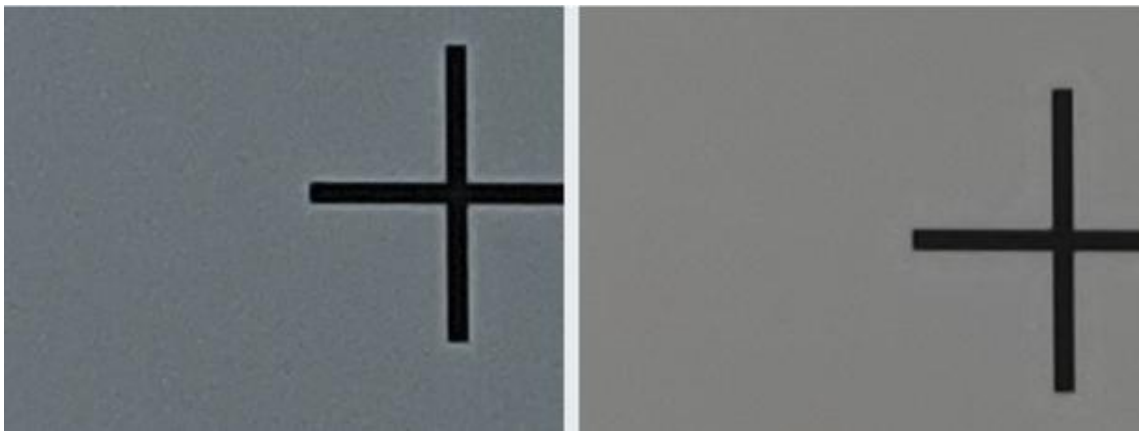

Figure 6.4. *Visual noise comparison under bright illumination between Qualcomm default tunings (000) on the left and tunings generated using only VCX score as target metric (002) on the right. Note the substantially smaller amount of visual noise in the flat areas of the image generated with tuning solution 002.*

Sharpening

Texture reproduction is also affected by the amount of sharpening. The VCX score determines the level of sharpness mainly by using two metrics: overshoot/undershoot area and maximum SFR [50]. The overshoot and undershoot area does not tolerate contrast enhancement around the edges well. Contrast enhancement means that the edge is amplified by making the area around the edge lighter and another side darker. This will introduce an artifact known as a halo or ringing. A more elaborated explanation can be found in appendix A. This metric controls the shape of the edge and the amount of sharpening heavily. Figure 6.5 presents the edge profile curves for tuning solutions 001 and 002 under bright conditions. Solution 002 has a significantly smaller amount of over- and undershoot, and thus the level of sharpening is smaller. This also affects the texture reproduction, as can be seen from figure 6.2. On the other hand, this metric is not uniform across the illumination levels. Under low illuminant, a much greater amount of overshoot and undershoot is tolerated. This can be seen as an abrupt change in figure 6.2 bottom left plot, where the graph of solution 002 abruptly changes. This discontinuity in the behavior of the sharpening is also visible in the images of solution 002, presented in figure 6.9. The way the VCX score weighs sharpening in higher illumination levels is not in line with the lowest illumination level.

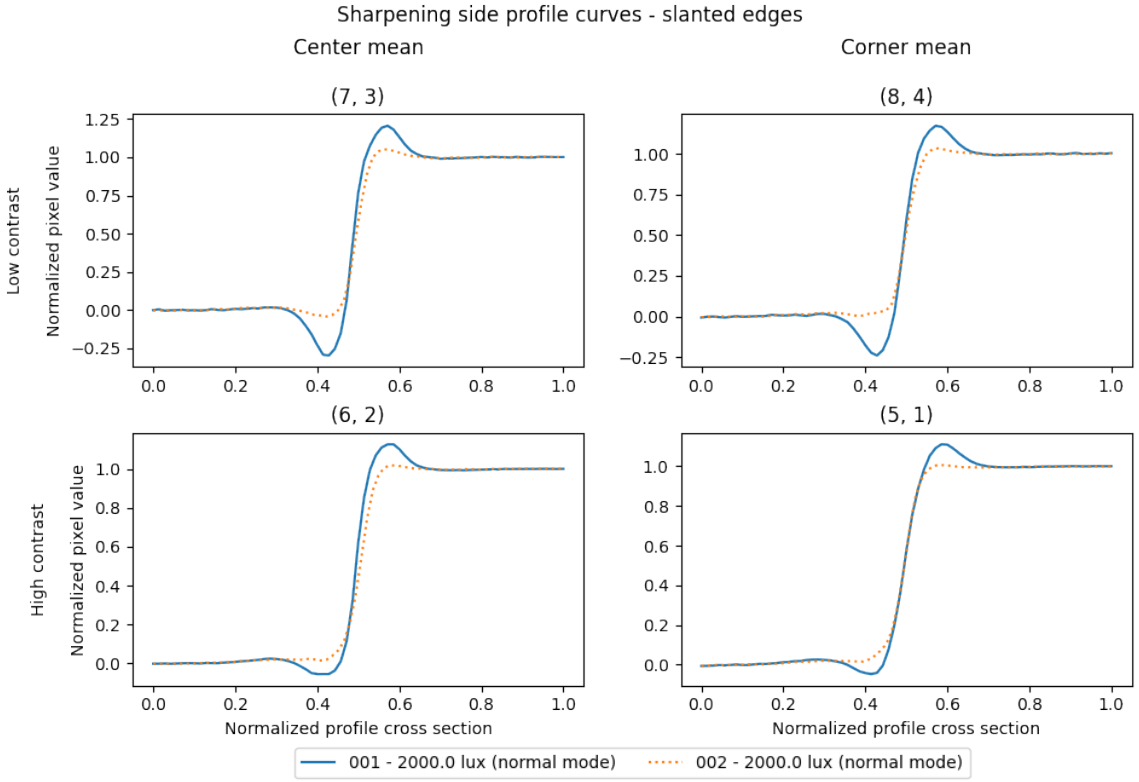

Figure 6.5. *Sharpening profile curve comparison between solutions OnePlus 7T (001) and tuning solution generated using only VCX metrics as IQ target (002) under bright illumination.*

Using the VCX score's overshoot and undershoot –metric as a tuning target is practical if the target is to produce a certain shape of an edge. Figure 6.6 presents an example of how using only VCX metrics as the IQ target controls the edge shape compared to the edge shape of solution 001 in the bright illumination. To be more precise, the metric should include information about the height of the edge. Two edges may match the area, but the other might be wider, and the other has more overshoot. That is very relevant information when tuning the IQ.

Figure 6.6. *Left image is the commercial OnePlus 7T (001) baseline image, right image is generated with the tuning solution achieved using only VCX score as target metric (002). Images are taken under bright illumination. Note the differences in the edge shape.*

Solutions 003 and 004 provided better texture acutance although their VN levels are in line with solution 002. These solutions used DISTS IQ metric to assess texture content and edge profiles. This highlights the importance of sharpening in texture reproduction, and the automatic tuning framework's ability to match the VN levels very closely since the solutions 002 and 003 used the VCX score VN-metrics as targets. Solution 004 targeted also a low level of VN, only using DISTS to do so. Figure 6.7 presents an example comparison of texture targets between solutions 004 and 002 under bright illumination. The image generated with the tuning solution 004 is on the left. It can be seen that a higher amount of sharpening results to better texture reproduction.

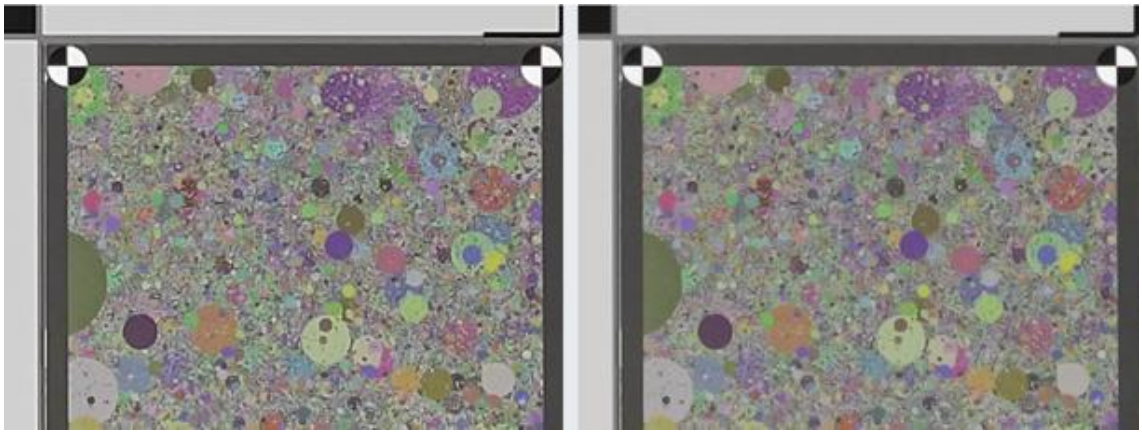

Figure 6.7. *Comparison of the texture targets between tuning solution generated using only DISTS (004) on the left and tuning generated using only VCX score as the target metric (002) on the right. Images were taken under bright illumination. Note how higher sharpening on the left produces better textures.*

The sharpening metrics of the VCX score favor too low sharpness under higher illuminants if the target was to reach similar texture reproduction as 001. Figure 6.8 presents the comparison between sharpening amounts of different tuning solutions. Here the

aforementioned inconsistency of sharpening in solution 002 is very visible. Also, the increase of sharpness in solutions 003 and 004 is noticeable when compared to solution 002. This is because DISTS was used to target better texture reproduction and a higher level of sharpening in general.

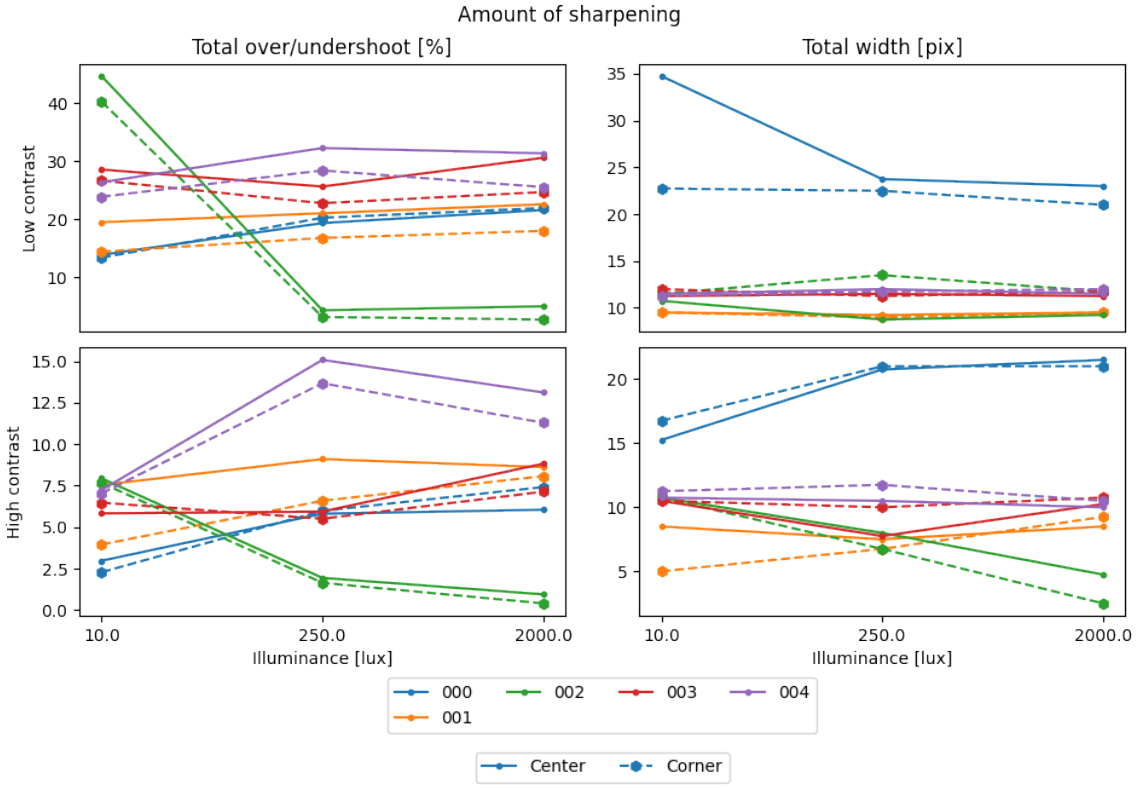

Figure 6.8. *Sharpening comparison between tuning solutions.*

A high level of sharpening can easily cause disturbing artifacts when edges are over-sharpened. This might be one reason why the VCX score's take on the sharpness is very conservative. This is not optimal when tuning the ISP. It might leave the total amount of sharpening to be too low and the image may appear blurry. Although, the allowed level of sharpness is not in line with different illumination levels in the VCX Standard. Figure 6.9 shows an example of the inconsistency of the sharpening with tuning solution 002. Over-sharpening the low light image causes sharpening artifacts like amplified noise around the edges to appear, making the edges jaggy. The right-hand image taken under a bright illuminant does not have the same level of sharpness.

Figure 6.9. *Inconsistency of the sharpening between illumination levels of the tuning solution generated by using only VCX metrics as target (002). The left image is taken under low illuminant and the right is taken under bright illuminant. Note the amount of sharpening artifacts in the left image.*

Resolution

Figure 6.10 presents the Siemens star acutance calculations between all tuning solutions and under all illuminations. It is clear that an increased level of sharpness also increases the Siemens star acutance. The sharpening discontinuity of tuning solution 002 is also distinguishable here.

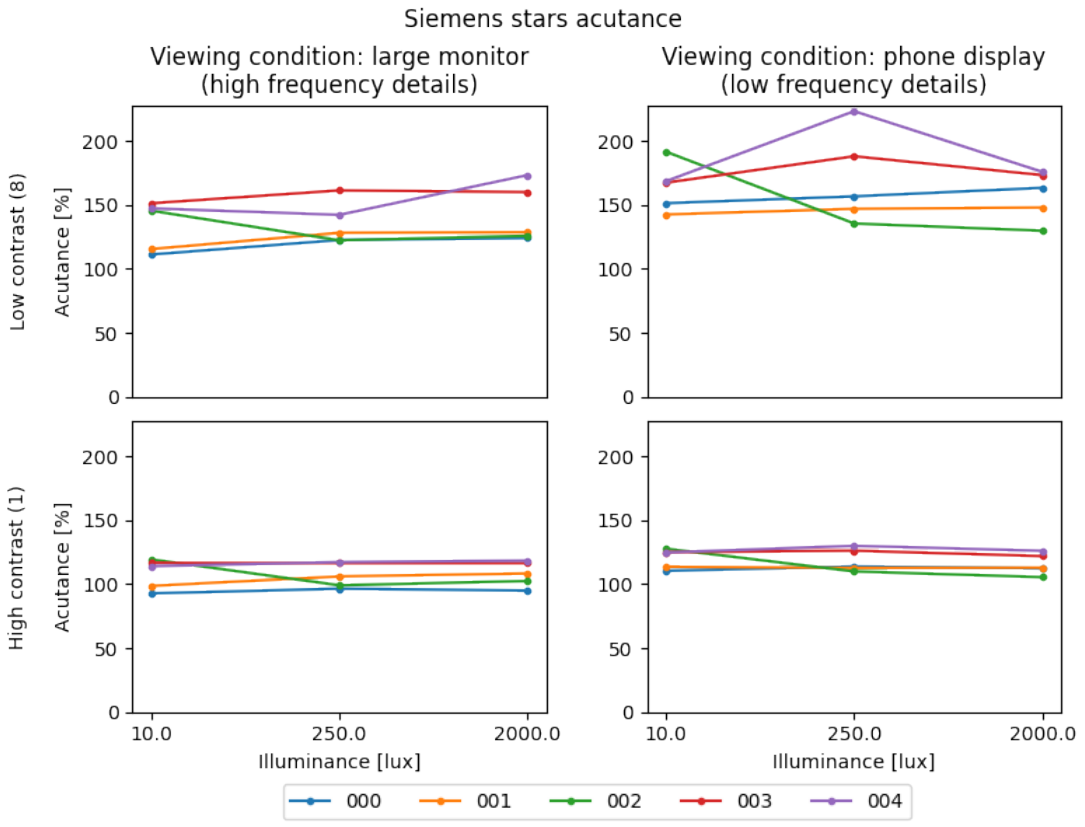

Figure 6.10. *Siemens star acutance comparison between tuning solutions.*

Even though tunings 003 and 004 have the highest values almost throughout every viewing condition on siemens star acutance, the texture acutance does not behave the same way. Especially with viewing condition 1, high-frequency details, the baseline solutions 000 and 001 produce the best results on average with mid and bright illumination. This might imply that the noise filtering in solutions 002, 003, and 004 is too strong and those high-frequency details are being washed away. This might also imply that the sharpening algorithm used with baseline solutions differs slightly from the one used with tuning solutions generated in this thesis.

Colors

Figure 6.11 presents the color errors between the tuning solutions. On average the tuning solutions generated in this work provide the most accurate results. Color reproduction is most frequently producing worse results under 10 lux illumination. This is due to the white balance shift which produces a slightly warmer image. This phenomenon can be seen from the resulting images presented in figures B.3–B.5. This color temperature change is done to make images look more natural to the human eye since the low-illuminant images are shot under a warm LED light.

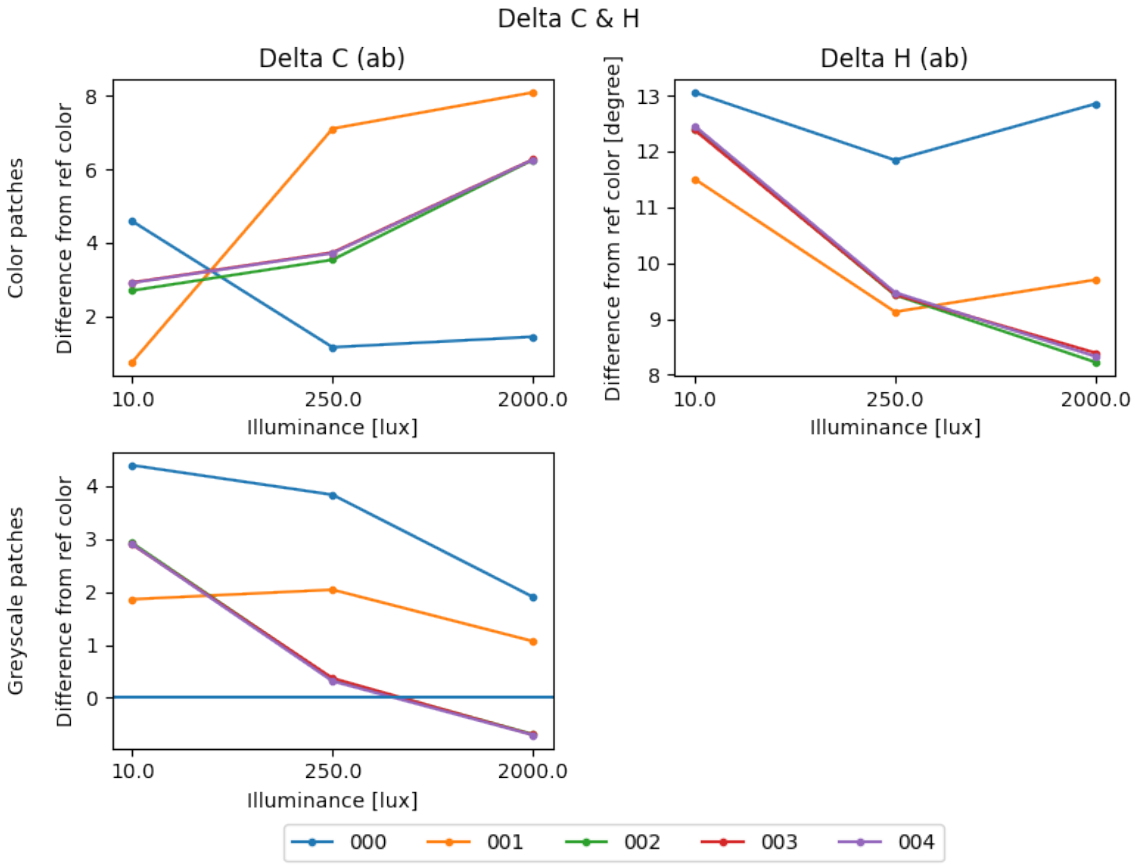

Figure 6.11. *Chrominance and hue difference comparison from reference colors between tuning solutions.*

Providing very accurate color tunings might not be desired. Often camera manufacturers want to modify certain colors to make images look more pleasing. For example, camera manufacturers may produce color tunings where the color of the sky renders a beautiful light blue. Overall manufacturers might prefer more saturated colors for artistic purposes. These are all tuning solutions that the IQ engineer must perform when tuning the ISP.

Summary of the Image Quality Tuning Results

Tuning the ISP using purely VCX metrics as a target, as was done with tuning solution 002, did produce a bit blurry final IQ in bright and mid illumination. The main reason for this is the overshoot and undershoot –metric of the score. The metric decreases too abruptly when a small amount of sharpening is applied. This leads to a blurry solution. This combined with the VN metric, where a smaller amount of VN provides always a better result, produces a filtered and blurry result compared to other tuning solutions. On the other hand, the low illuminant image was not in-line with other illuminant levels. The result seemed oversharpened and it caused some discontinuity in solution 002. The following section will test the IQ subjectively to test does this match with a common perception of good IQ.

Applying DISTS to analyze textured content and edge shape during the optimization procedure allowed the tuning framework to produce sharper images. Even though the sharpness and textures were assessed using different methods compared to solution 002, the framework was still able to match the VN levels which were assessed using VCX metrics. This can be seen from the VN levels of tuning solutions 002, 003, and 004 presented in figure 6.3. The tuning solutions produced in this work reached the lowest overall VN values compared to baseline solutions 000 and 001.

The resolution did improve after applying different IQ evaluation methods for sharpening. The Siemens star acutance of tuning solution 002 is very in-line with baseline solutions, even though solution 002 produced worse texture acutance results on average. This implies different sharpening algorithms and approaches between solutions 000, 001, and 002.

Color tunings produced accurate colors on average. White balance caused some color errors in low-illuminant images as expected. This is not critical since making the color temperature of the image match the color temperature of the prevalent light source is crucial in order to make the image look natural to a human observer.

In general, targeting a single objective number might be preferred. With objective metrics, it is easy to target certain performances on selected IQ aspects. For example, a customer might want to have a certain noise level or edge shape. It is difficult from a set of objective measurements to determine which tuning solution is the best overall. Some tuning so-

lutions have better performance in resolution and noise but others are able to reproduce textures on a better level. The weighing between these different metrics is difficult but it is done in the VCX score. The following section 6.2 will define does the solution tuned by maximizing the VCX score actually correlate with the subjective stimulus of the IQ.

6.2 Subjective Image Quality Assessment

A subjective image quality assessment is conducted to determine which tuning solution reached the best subjective image quality under the measuring conditions. The used subjective IQA method is a forced-choice or pairwise comparison method, described in section 5.2. The actualized sample size was **33 respondents**, which is above the suggested sample size. Roughly two-thirds of the respondents were not working in the imaging industry. The performance of each tuning solution was measured under three illumination levels: bright, mid, and low. These are defined in the VCX score [50]. Each illumination level represents one scene. Also, the overall performance of every tuning solution will be presented to define the solution with the best overall performance.

Table 6.2 presents the response matrix of the bright scene. The tuning solution 004 was selected most frequently over others. This might be due to the increased sharpness.

Table 6.2. *The response matrix of the bright illuminant scene. The matrix is read column-wise and it tells how many times an image was selected over others. The mean preference score (MPS) tells how many times on average a respondent selected an image over other options. Proportion stages the percentage of an image selected relative to the maximum selection amount.*

Scene of bright illuminant					
Tuning solution	**000**	**001**	**002**	**003**	**004**
000	-	21	20	24	28
001	12	-	8	18	17
002	13	25	-	22	24
003	9	15	11	-	25
004	5	16	9	8	-
Sum	38	77	48	72	94
MPS	1.18	2.33	1.45	2.18	2.85
Proportion	0.30	0.58	0.36	0.55	0.71

Table 6.3 presents the response matrix of the mid illuminant scene. The tuning solutions 003 and 004 were selected most frequently over other solutions.

Table 6.3. *The response matrix of the mid illuminant scene.*

Scene of mid illuminant					
Tuning solution	**000**	**001**	**002**	**003**	**004**
000	-	26	23	28	28
001	7	-	20	26	25
002	10	13	-	26	22
003	5	7	7	-	19
004	5	8	11	14	-
Sum	27	54	61	94	94
MPS	0.82	1.64	1.85	2.85	2.85
Proportion	0.20	0.41	0.46	0.71	0.71

Table 6.4 presents the response matrix of the low illuminant scene. The tuning solution 004 was selected most frequently over other solutions.

Table 6.4. *The response matrix of the low illuminant scene*

Scene of low illuminant					
Tuning solution	**000**	**001**	**002**	**003**	**004**
000	-	17	29	29	31
001	16	-	21	30	29
002	4	12	-	18	16
003	4	3	15	-	28
004	2	4	17	15	-
Sum	26	36	82	92	94
MPS	0.79	1.09	2.48	2.79	2.84
Proportion	0.20	0.27	0.62	0.70	0.71

Table 6.5 presents the overall response matrix. Tuning solutions produced as part of this thesis were preferred over the baseline tunings. The most preferred IQ was produced by the tuning solution 004.

Table 6.5. *The overall response matrix.*

Overall					
Tuning solution	**000**	**001**	**002**	**003**	**004**
000	-	64	72	81	87
001	35	-	49	74	71
002	27	50	-	66	62
003	18	25	33	-	62
004	12	28	37	37	-
Sum	92	167	191	258	282
MPS	2.79	5.06	5.79	7.82	8.55
Proportion	0.23	0.42	0.48	0.65	0.71

These results reflect the quality of the tuning solutions proposed as a result of this thesis. Table 6.5 shows that every solution generated in this work produced better results in the subjective IQ analysis compared to the baseline solutions. This gives confidence to the first research question. It shows that optimizing the ISP parameters using fully objective metrics can produce subjectively good IQ. The tuning solution 002, which aimed to maximize the VCX score, produced better results in a subjective IQA than the baseline devices. Note that the baseline solution 001 is a commercial reference device OnePlus 7T using the same CCM as the device that was used to produce other solutions.

Even though 002 produces better IQ than baseline solutions, the alternative ways produced even better results than only optimizing the IQ against the VCX score. Solutions 003 and 004 used DISTS full-reference IQA method [9] to determine the IQ of the resulting image. Solution 003 used DISTS to achieve better texture reproduction, still aiming to maximize the VCX score of resolution, color, and visual noise. Solution 004 used only DISTS to determine the IQ of the result, to achieve an even better level of texture reproduction allowing a bit more VN on the final image.

Subjective IQA results show that solutions 003 and 004 produced better results than solution 002. Solution 002 used only the VCX score to measure the IQ during the optimization procedure. Utilizing DISTS as an objective function resulted in a better outcome. The usage of objective VCX metrics as the target function of the automatic IQ tuning framework is intentional when, for example, a certain level of VN is desired. The overall problem of optimizing the IQ using only VCX metrics as the target is the blurriness of bright and mid illumination scenes. Also, another problem is the inconsistency of the IQ when the low illumination scene result is compared to the bright and the mid illuminant scene results. The abrupt large increase in the level of sharpness is a clear discontinuity in the final IQ.

The tuning solutions 003 and 004 do not have this kind of the discontinuity. The behavior of the VCX sharpness metric is not consistent between the illumination levels.

When comparing the objective results and the final images of tuning solutions 003 and 004, it is clear that they are very close to one another. Close call choices introduce the most inconsistencies to the comparison results [4, p. 83]. Due to this, the results might experience distortion, and making clear conclusions about which tuning solution is actually subjectively better might not be feasible. Also, the IQ was not verified in a thorough manner. Since the objective of this thesis was to only optimize the IQ under the scenarios defined by the VCX score, no field testing of the tuning solutions was conducted. Field testing means thorough testing of the IQ solution by exposing it to varying illuminations and conditions of the real world. These field tests might bring up more inconsistencies in the IQ of each solution that was missed in a controlled laboratory environment.

7. CONCLUSIONS

This thesis investigated the different ways to optimize the image quality (IQ) of a mobile device by automatically optimizing the image signal processor (ISP) parameters using different objective functions as targets of an automatic IQ tuning framework. The motivation for the research was to determine that does measuring IQ using fully objective metrics and maximizing their output reflect the subjective stimuli of IQ. The set of objective metrics that were used are part of Valuated Camera eXperience (VCX) standard.

VCX standard defines a VCX score, which is conducted as a result of a fully objective IQ analysis. The VCX score is a single number that aims to represent the overall performance of the imaging system by measuring it under defined conditions using standardized methods. Thus, the VCX score results are highly reproducible. The score gives an easy reference to consumers to compare different devices and find a device for their needs. The VCX score consists of multiple different metrics which aim to comprehensively evaluate camera performance. In this thesis, only metrics measuring visual noise, sharpness, texture loss, colors, and resolution were used.

This research aimed to answer two research questions. The first research question concerned will ISP which was optimized using fully objective IQ metrics produce subjectively good results. This was investigated by optimizing an ISP using parts of the VCX score as a target and then comparing these results to a commercial device that used similar hardware. The IQ was compared by conducting a subjective IQ assessment (IQA). The subjective IQA method that was used is known as a pairwise comparison or a forced-choice comparison method. After reviewing the answers of 33 respondents, it was clear that under selected scenarios the ISP optimized using only fully objective VCX metrics produced better IQ. It has to be noted that the verification of the IQ was not comprehensive, since it was only tested under the scenarios defined by the VCX score and inconsistencies might exist.

The second research question considered alternative ways to define the objective functions of the automatic IQ tuning framework to achieve a better final IQ. Even though VCX claims that a larger score reflects the better performance of the imaging system, this thesis presents alternative ways to measure the IQ during the optimization procedure to gain even better subjective results. The alternative IQ target was a full-reference IQA model

which aims to correlate with human perception of the IQ called Deep Image Structure and Texture Similarity (DISTS) Metric [9]. This model was used to get a more consistent and higher level of texture reproduction throughout the different measuring conditions. Two different IQ tuning solutions were generated. The first solution used DISTS to determine the IQ of textured areas and edge shapes, and the VCX score to measure visual noise, color errors, and resolution of the result. The second solution used only DISTS to measure every aspect of IQ. In the subjective IQA, these two tuning solutions outperformed the one using only VCX metrics as IQ targets. The main difference between the solutions using DISTS and the solution using only VCX metrics was the sharpness of the image and the consistency of the tunings under different conditions. The solution using only VCX metrics produced a slightly blurrier image under bright and mid illumination and an over-sharpened result under low illuminant. This inconsistency was not present in the solutions using DISTS to measure texture loss and sharpness.

Based on these results, it is clear that there is a place for fully objective IQ metrics in the tuning of the ISP parameters. Although, using only objective metrics might not be appropriate. Using only objective metrics might introduce artifacts to the image which are not visible from the metrics. For example, a too high level of sharpening might emphasize noise around the edges making them seem jagged. At the same time, if the noise level on the flat parts of the image is low, this phenomenon might not be visible from the visual noise metric. Using objective metrics is very justifiable when a certain level of, for example, visual noise is aimed to achieve. This way, making the automatic tuning framework converge towards a low-enough noise value is easy by penalizing the high amount of noise by a great factor. Also, measuring the edge shape allows the tuning framework to achieve an output with a desired edge profile.

Modern deep learning-based solutions, like DISTS, allow measuring IQ in a subjective manner without conducting time-consuming subjective IQA. Using a full-reference method allows guiding the automatic tuning framework to converge to a certain IQ by using a reference image which is then used by the network to determine how close the IQ of a given input image is to the reference. Defining the IQ with a reference image is also more logical and easier to understand than defining subjective IQ as a weighted sum of different objective metrics. This might prevent discontinuity in the IQ between, for example, different illuminations. Nevertheless, both of these methods have their place as the objective functions of the automatic tuning framework.

REFERENCES

[1] Algolux. *Automatic Image Quality; Computer Vision Optimization: ISP tuning*. Jan. 2023. URL: https://algolux.com/atlas/.

[2] Uwe Artmann. “Image quality assessment using the dead leaves target: experience with the latest approach and further investigations”. eng. In: *DIGITAL PHOTOGRAPHY XI*. Vol. 9404. Proceedings of SPIE. BELLINGHAM: SPIE, 2015, 94040J-94040J–15. ISBN: 9781628414943.

[3] Uwe Artmann. “VCX Version 2020 - Further development of a transparent and objective evaluation scheme for mobile phone cameras”. eng. In: *Electronic Imaging* 2021.9 (2021), pp. 204-1-204–6. ISSN: 2470-1173.

[4] Thomas C. Brown and George L. Peterson. “An enquiry into the method of paired comparison: Reliability; scaling; and thurstone’s law of comparative judgment”. eng. In: *USDA Forest Service - General Technical Report RMRS-GTR* 216 (2009), pp. 1–98. ISSN: 0277-5786.

[5] Brandon J Brys. *Image Restoration in the Presence of Bad Pixels*. eng. 2010.

[6] *Camera Module: Definition and Types*. Jan. 3, 2019. URL: https://insightsolutionsglobal.com/camera-module-definition-and-types/ (visited on 10/30/2022).

[7] Internation Color Consortium. *Introduction to the ICC profile format*. URL: https://www.color.org/iccprofile.xalter. (accessed: 12.11.2022).

[8] GIACALONE DAVIDE, BOSCO ANGELO, and BRUNA ARCANGELO RANIERI. *Image chroma noise reduction in the bayer domain*. eng. 2015.

[9] Keyan Ding et al. “Image Quality Assessment: Unifying Structure and Texture Similarity”. eng. In: *IEEE transactions on pattern analysis and machine intelligence* 44.5 (2022), pp. 2567–2581. ISSN: 0162-8828.

[10] Image Engineering. *Shading*. URL: https://www.image-engineering.de/library/image-quality/factors/1073-shading/. (accessed: 09.03.2023).

[11] Image Engineering. *TE269*. URL: https://www.image-engineering.de/products/charts/all/406-te269. (accessed: 05.11.2022).

[12] Image Engineering. *TE42-LL*. URL: https://www.image-engineering.de/products/charts/all/1034-te42-ll. (accessed: 04.11.2022).

[13] Angela. Faris-Belt. *The elements of photography understanding and creating sophisticated images*. eng. 1st edition. Amsterdam ; Focal Press, 2008. ISBN: 1-281-09623-7.

[14] G. Geleijnse et al. “Measuring Image Quality of ENT Chip-on-tip Endoscopes”. eng. In: *The Journal of imaging science and technology* 65.2 (2021), pp. 20503-1-20503–7. ISSN: 1062-3701.

[15] Rafael C. Gonzalez. *Digital image processing*. eng. 2. ed. Upper Saddle River (NJ): Prentice Hall, 2002. ISBN: 0-201-18075-8.

[16] GSM-Arena. *OnePlus 7t - Full Phone Specifications*. URL: https://www.gsmarena.com/oneplus_7t-9816.php. (accessed: 13.03.2023).

[17] R.M. Harold. “An introduction to appearance analysis”. eng. In: *GATFWORLD (Pittsburgh, Pa.)* 13.3 (2001), pp. 5–. ISSN: 1048-0293.

[18] Luis V Hevia et al. “Optimization of the ISP Parameters of a Camera Through Differential Evolution”. eng. In: *IEEE access* 8 (2020), pp. 143479–143493. ISSN: 2169-3536.

[19] Imatest. *Sharpness: What is it and How it is Measured*. URL: https://www.imatest.com/docs/sharpness/. (accessed: 06.11.2022).

[20] Imatest. *Sinusoidal Siemens Star*. URL: https://www.imatest.com/product/sinusoidal-siemens-star/. (accessed: 06.11.2022).

[21] Imatest. *Uniformity*. URL: https://www.imatest.com/solutions/uniformity/. (accessed: 09.03.2023).

[22] *Photography – Electronic still picture imaging – Resolution and spatial frequency responses*. Standard. International Organization for Standardization, Jan. 2017.

[23] Dervis Karaboga and Bahriye Basturk. “A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm”. eng. In: *Journal of global optimization* 39.3 (2007), pp. 459–471. ISSN: 0925-5001.

[24] J.D. Kelleher. *Deep Learning*. The MIT Press Essential Knowledge series. MIT Press, 2019. ISBN: 9780262354905. URL: https://books.google.fi/books?id=ZU6qDwAAQBAJ.

[25] Younghoon Kim et al. “DNN-based ISP parameter inference algorithm for automatic image quality optimization”. eng. In: *Electronic Imaging* 2020.9 (2020), pp. 315-315–6. ISSN: 2470-1173.

[26] Aaron A. King and Tom Rowan. *Subplex: Unconstrained optimization using the SUBPLEX algorithm*. Apr. 2022. URL: https://cran.r-project.org/web/packages/subplex/subplex.pdf.

[27] Kuk Won Ko et al. “An automatic optical inspection system for inspection of CMOS compact camera module assembly”. eng. In: *International Journal of Precision Engineering and Manufacturing* 10.5 (2009), pp. 67–72. ISSN: 1229-8557.

[28] Yanni Liu. *Camera System Characterization with Uniform Illuminate*. eng. Informaatioteknologian ja viestinnän tiedekunta - Faculty of Information Technology and Communication Sciences, 2020.

[29] A. Lustica. “CCD and CMOS image sensors in new HD cameras”. eng. In: *Proceedings ELMAR-2011*. IEEE, 2011, pp. 133–136. ISBN: 9781612849492.

[30] Sean (Sean T.) McHugh. *Understanding photography : master your digital camera and capture that perfect photo*. eng. 1st edition. San Francisco: No Starch Press, 2019. ISBN: 1-4920-6946-9.

[31] Chamin Morikawa et al. “Image and video processing on mobile devices: a survey”. eng. In: *The Visual computer* 37.12 (2021), pp. 2931–2949. ISSN: 0178-2789.

[32] JA NELDER and R MEAD. “A SIMPLEX-METHOD FOR FUNCTION MINIMIZATION”. eng. In: *Computer journal* 7.4 (1965), pp. 308–313. ISSN: 0010-4620.

[33] Jun Nishimura et al. “Automatic ISP Image Quality Tuning Using Nonlinear Optimization”. eng. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 2471–2475. ISBN: 9781479970612.

[34] KALEVO OSSI, ALAKARHU JUHA, and NIKKANEN JARNO. *Method and system for black-level correction on digital image data*. eng. 2010.

[35] Marius Pedersen and Jon Yngve Hardeberg. “Full-reference image quality metrics: classification and evaluation”. eng. In: *Foundations and trends in computer graphics and vision*. Foundations and trends in computer graphics and vision 7.1 (2012), pp. 1–80. ISSN: 1572-2740.

[36] Jonathan B Phillips and Henrik Eliasson. *Camera Image Quality Benchmarking*. eng. The Wiley-IS&T Series in Imaging Science and Technology. Newark: John Wiley & Sons, Incorporated, 2018. ISBN: 9781119054498.

[37] Jonathan B Phillips and Henrik Eliasson. *Camera Image Quality Benchmarking*. eng. The Wiley-IS&T Series in Imaging Science and Technology. Newark: John Wiley & Sons, Incorporated, 2018. ISBN: 9781119054498.

[38] Geoffrey Portelli and Denis Pallez. “Image Signal Processor Parameter Tuning with Surrogate-Assisted Particle Swarm Optimization”. eng. In: *Artificial Evolution*. Vol. 12052. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 28–41. ISBN: 3030457141.

[39] Charles A. Poynton. *Digital video and HD algorithms and interfaces*. eng. 2nd ed. The Morgan Kaufmann Series in Computer Graphics. Amsterdam: Morgan Kaufmann, an imprint of Elsevier, 2012. ISBN: 1-280-58146-8.

[40] *Qualcomm Snapdragon 888 5G Mobile Platform: Qualcomm*. Feb. 2020. URL: https://www.qualcomm.com/products/application/smartphones/snapdragon-8-series-mobile-platforms/snapdragon-888-5g-mobile-platform.

[41] Pavithra G. Radhesh Bhat. “Automatic image quality tuning framework for optimization of ISP parameters based on multi-stage optimization approach”. eng. In: *Electronic Imaging* 2021.9 (2021), pp. 197-197–7. ISSN: 2470-1173.

[42] R. Ruslan and N. Roslan. “Assessment on the skin color changes of Carica papaya L. cv. Sekaki based on CIE Lab and CIE LCh color space”. eng. In: *International food research journal* 23 (2016), S173–S178. ISSN: 1985-4668.

[43] O.H. SCHADE Sr. “Optical and photoelectric analog of the eye”. eng. In: *Journal of the Optical Society of America (1930)* 46.9 (1956), pp. 721–739. ISSN: 0030-3941.

[44] Chenyang Shen et al. “Intelligent Parameter Tuning in Optimization-Based Iterative CT Reconstruction via Deep Reinforcement Learning”. eng. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1430–1439. ISSN: 0278-0062.

[45] *Sony releases stacked CMOS image sensor for smartphones with industry's highest 48 effective megapixels*. July 2018. URL: https://www.sony.com/en/SonyInfo/News/Press/201807/18-060E/.

[46] *Swarm intelligence and bio-inspired computation theory and applications*. eng. 1st ed. Elsevier insights Swarm intelligence and bio-inspired computation. Boston, Mass: Elsevier, 2013. ISBN: 0-12-405177-4.

[47] L. L Thurstone. “A Law of Comparative Judgment”. eng. In: *Psychological review*. The Centennial Issue of the Psychological Review 101.2 (1994), pp. 266–270. ISSN: 0033-295X.

[48] Ethan Tseng et al. “Hyperparameter optimization in black-box image processing using differentiable proxies”. eng. In: *ACM transactions on graphics* 38.4 (2019), pp. 1–14. ISSN: 0730-0301.

[49] VCX-Forum. *What is VCX*. URL: https://vcx-forum.org/what-is-vcx/who-we-are. (accessed: 04.11.2022).

[50] VCX-Forum. *White Paper*. URL: https://vcx-forum.org/standard/white-paper. (accessed: 04.11.2022).

[51] Zhou Wang and Alan C. Bovik. “Modern image quality assessment”. eng. In: *Synthesis lectures on image, video, and multimedia processing* 3 (2005), pp. 1–156. ISSN: 1559-8136.

[52] *What is an IR-cut filter – and why do embedded vision applications need it?* Feb. 24, 2022. URL: https://www.e-consystems.com/blog/camera/technology/what-is-an-ir-cut-filter-and-why-do-embedded-vision-applications-need-it/ (visited on 10/30/2022).

[53] Chyuan-Tyng Wu et al. “VisionISP: Repurposing the Image Signal Processor for Computer Vision Applications”. eng. In: *2019 IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP)*. Vol. 2019-. IEEE International Conference on Image Processing ICIP. IEEE. NEW YORK: IEEE, 2019, pp. 4624–4628. ISBN: 9781538662496.

[54] Dietmar Wueller et al. “VCX: An industry initiative to create an objective camera module evaluation for mobile devices”. eng. In: *Electronic Imaging* 30.5 (2018), pp. 172-1-172–5. ISSN: 2470-1173.

APPENDIX A: VCX STANDARD V2020 SCORE CALCULATION FORMULAS

s-SFR - LR center and corner are defined as spatial frequency line pairs (LP) per picture height in megapixels (PH) that will result in an MTF10. Figure A.1 illustrates how limiting resolution is obtained. The LR is the frequency at which the SFR crosses the dashed red line. When calculating the LR for the center, the Siemens star is divided into eight equal segments. Similarly, the Siemens star is divided into three segments when calculating the limiting resolution for image corners. For each segment, a LR value is calculated and then an average of these values is determined. If the limiting resolution is higher than the theoretical maximum resolution (TMR), the limiting resolution is reduced to the theoretical limit. Theoretical maximum resolution can be determined as $TMR = \frac{PH}{2}$. If LR is low, it will lower the overall VCX score.

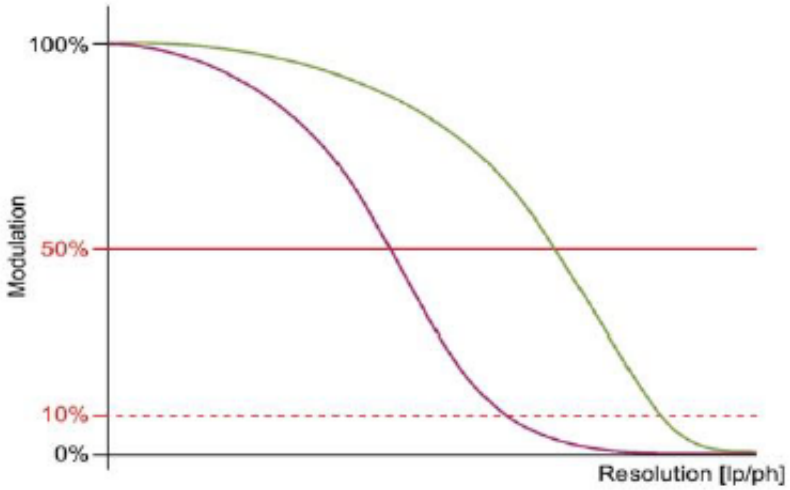

Figure A.1. *Modulation transfer as a function of spatial frequency [50].*

Segments are further divided into two different groups: diagonal and horizontal/vertical. This is done because horizontal and vertical segments are limited to TMR, as the diagonal line pair can have a limiting resolution of $TMR * \sqrt{2}$. The equation A.1 expresses how limiting resolution is calculated

$$LR_{all} = \begin{cases} \frac{\sum_{segment=1}^{segment=m} LR_{segment}}{m}, & \text{if } LR_{all} \leq TMR \\ \frac{\sum_{segment=1}^{segment=m} LR_{corr}}{m}, & \text{if } LR_{all} > TMR \end{cases} \quad \text{(A.1)}$$

where m is amount of segments and LR_{corr} is

$$LR_{corr} = \begin{cases} \min(LR_{segment}, TMR), & \text{for horizontal and vertical segments} \\ \min(LR_{segment}, TMR * \sqrt{2}), & \text{for diagonal segments.} \end{cases} \quad \text{(A.2)}$$

EPC center and overall is describing the level of detail the system can reproduce. EPC is based on a ratio of LR defined above and TMR multiplied by the theoretical pixel count TPC. TPC can be calculated as the product of PH and picture width (PW) in megapixels divided by 10^6, $TPC = \frac{PH*PW}{10^6}$, to obtain the number of pixels DUT has. Equation A.3 presents the EPC of the image center

$$EPC_{center} = \left(\frac{LR_{center}}{TMR}\right)^2 * TPC. \quad \text{(A.3)}$$

The overall EPC formula is presented in the equation A.4 below

$$EPC_{overall} = \left(\frac{LR_{center} + LR_{corner}}{2 * TMR}\right)^2 * TPC. \quad \text{(A.4)}$$

s-SFR acutance for center and corner aims to describe the actual sharpness that the DUT produces. However, sharpness is a subjective experience of an observer which depends on SFR and VC. This is why the calculation of acutance which means the perceived sharpness of an image [13]. In the VCX standard, the calculation of the acutance is based on the contrast sensitivity function (CSF). The CSF aims to model the subjective experience of SFR, meaning that is an edge visible to a human observer, and thus used when calculating the acutance [43]. CSF decomposes edges into a series of sinusoidal gratings of different frequencies which allows the discrimination of edge from non-edge content [43]. The CSF used with this metric is described in the ISO15739:2013-standard. Here, the acutance is a ratio between two integrals computed from the minimum spatial frequency range (f_{min}) to the Nyquist frequency (f_{nyq}). f_{min} is the smallest frequency that was analyzed in the SFR when it is based on the Siemens star 3.3. The CSF is scaled according to the VC. This metric can be then defined as

$$Acutance(SFR, VC) = \frac{A}{A_r}, \quad \text{(A.5)}$$

where A and A_r can be expressed as

$$A = \int_{f_{min}}^{f_{nyq}} SFR(f) * CSF_{VC}(f)\, df \tag{A.6}$$

$$A_r = \int_{f_{min}}^{f_{nyq}} CSF_{VC}(f)\, df. \tag{A.7}$$

e-SFR MTF50 metric is implemented to measure the sharpening of the DUT. With s-SFR methods, the spatial frequency response was calculated using the harmonic Siemens star 3.3 as a target. Now with the e-SFR method, it is calculated using the slanted edge target. An example of a slanted edge target cropped from 3.1is in figure A.2 below. This method is very useful when determining the amount of sharpening device applied to the image. e-SFR is not used to obtain the limiting resolution, since the sharpening affects the result heavily. A detailed description of how sharpening measurements are obtained is in section 3.3. Recording of the edge spread function (ESF) is the core of the analysis. ESF states how the edge is reproduced. Then a Fourier transform of the derivative of the ESF is taken to form the SFR function. Then MTF50 number is obtained from the point where the spatial frequency drops to half of its peak value. TE42-LL 3.1 contains 4 high contrast slanted edge targets, also visible in the figure A.2 and 4 low contrast slanted edges. Two spatial frequency (LP/PH) values are reported. MTF50 values calculated from the slanted edges having the same contrast are averaged together.

Figure A.2. *High contrast slanted edge target cropped from TE42-LL test target.*

e-SFR acutance is calculated the same way as s-SFR acutance described above. The only difference is that SFR functions are now calculated using slanted edges. Two different e-SFR acutance values are obtained, one for low-contrast slanted edges and one for high-contrast slanted edges.

SFR-dead-leaves-cross ($SFR_{DeadLeaves_{cross}}$) is a full reference method, meaning that it also needs vector-based data of the printed target which can be considered as the ideal image. This is done to distinguish between the spatial information that has been added to the image by the imaging pipeline of the device and spatial information that is in the target itself. This enables a more complex transfer function of the camera system to be obtained which takes into account also the phase information. The full flow chart of this

algorithm is presented in figure A.3. [2]

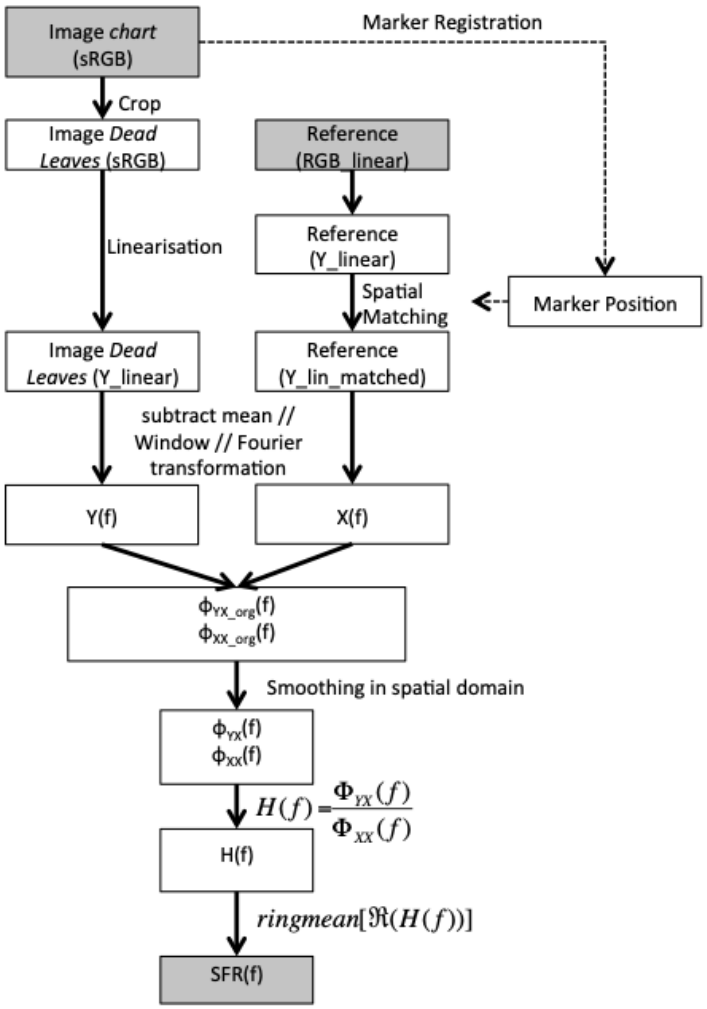

Figure A.3. *The full flow chart of the $SFR_{DL_{cross}}$ algorithm. [2]*

The most critical parts are to obtain the cross power density $\phi_{YX}(f)$ of the reference target and the device and to obtain the auto power density of the target $\phi_{XX}(f)$. With these two power spectrums, we can calculate the transfer function of the imaging pipeline $H(f)$

$$H(f) = \frac{\phi_{YX}(f)}{\phi_{XX}(f)}. \tag{A.8}$$

The key of the algorithm is to match the vectorized image of the dead leaves target to the one taken with the DUT, and the cross-power density of the target and image can be then calculated leading to the transfer function shown in the equation A.8. The SFR is finally calculated after smoothing the power density distributions as a 1D representation of the transfer function's real part. The transformation from 2D to 1D is done with a ring mean, which uses the average of all coefficients belonging to the same spatial frequency regardless of their orientation. [2]

Texture loss MTF10/MTF50 reflects the level of detail that the DUT can reproduce. The MTF10 value, also known as limiting resolution, reflects the smallest detail visible to human beings in low-contrast fine detailed structures. The MTF50 value correlates to the sharpness of an image. Both of these are derived from spatial frequency response obtained with $SFR_{DeadLeaves_{cross}}$ method.

Texture loss acutance describes how a human observer perceives the low contrast fine details under different VCs. The calculation of this metric is based on $SFR_{DeadLeaves_{cross}}$ method but otherwise identical to the s-SFR acutance metric introduced in section 3.3.

Artifacts is a metric reflecting the spatial information added by the DUT. This added information is referred to as artifacts. This is done with an algorithm introduced in [2]. Because $SFR_{DeadLeaves_{cross}}$ is a full reference method and since insensitive to the content in the measured image, it can be used as a baseline. An older method, $SFR_{DeadLeaves_{direct}}$ is not insensitive to the image content [2]. The full algorithm is described in Image Engineering's article [2]. Thus the difference between these spatial frequency responses is caused by artifacts introduced by the imaging pipeline of the DUT. The equation A.9 describes how this metric is calculated. The algorithm does not include weighting by VC, so CSF is assumed as a perfect perception for all frequencies.

$$Artifacts = 100 - \frac{Acutance(SFR_{DeadLeaves_{cross}}, all) * 100}{Acutance(SFR_{DeadLeaves_{direct}}, all)} \tag{A.9}$$

Chrominance loss describes how the saturation level drops as a function of illumination. In low-lit scenes, the saturation of an image is dropped in order to get rid of color noise in flat and neutral areas. The baseline for this metric is calculated under bright illumination described in subsection 3.3 under the data and capture part. This means that no chrominance loss is assumed in bright conditions.

In order to calculate the chrominance loss score, the red, green, and blue (RGB) values of each pixel are first converted into a color space defined by the Commission Internationale de l'éclairage (eng. International Commission on Illumination) (CIE) called CIE-XYZ. This color space reflects how the human visual system responds to different wavelengths of light on average [17]. The conversion is done using the appropriate International Color Consortium (ICC) color profile. This profile allows making an accurate color conversion exploiting the color profile used by the DUT [7]. Then, the data is converted into CIE-L*a*b* color space and finally, the data is transformed to CIE-L*C*h* format. CIE-L*a*b* color space is a uniform opponent color scale which allows a more accurate estimate of perceived color differences [17]. Opponent-colors theory bases itself on the fact that the human visual system compares red responses are compared to green to generate a red-to-green color dimension (a), and the green response is compared to blue to generate a yellow-to-blue color dimension [17] (b). The conversion from CIE-XYZ to CIE-L*a*b*

requires a definition of a white reference. This reference is measured from the white patch in the image of the TE42-LL chart 3.1. In the final color space, CIE-L*C*h*, the C* stands for chrominance and is thus used in this metric [42]. The chrominance loss is derived for each lighting condition with an equation A.10

$$\Delta C*_{condition} = \frac{C*_{condition}}{C*_{bright}} * 100, \tag{A.10}$$

where $C*_{bright}$ is a chrominance value under bright conditions measured from CIE-L*C*h* color space and $C*_{condition}$ is the same value measured under some different, dimmer, lighting condition.

Overshoot and undershoot score describes how the DUT reproduces edges. One of the most noticeable artifacts produced by sharpening is light and dark areas around the edges. This makes the edge look sharper. This artifact is also known as halos or ringing. Figure A.4 is an example of an edge with a lot of overshoot and undershoot. Also, the figure below the image shows the amount of ringing as a function of position in the image.

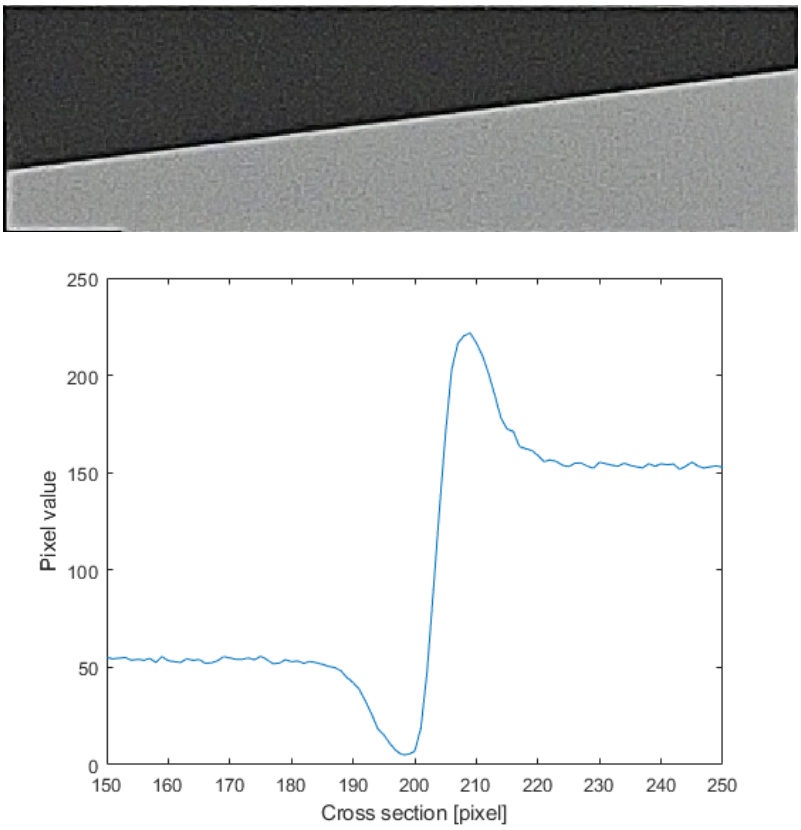

Figure A.4. *Slanted edge with a ringing effect and corresponding edge spread function.*

The score is derived from the ESF. It is calculated as the integral of the area beneath (dark region, undershoot) or above (bright region, overshoot) the edge spread function. The amount of sharpening is averaged between the edges with the same modulation. The measurement is also dependent on the pixel count of the DUT since devices with

high pixel counts can introduce more sharpening without it being distractive for a human observer. Thus, not only the total area of over- or undershoot is reported but it's also scaled with assumed pixel density for the given VC. The scores OS_{scaled} and US_{scaled} are derived from equation A.11

$$\begin{aligned} OS_{scaled} &= \frac{OH}{PH} * OS_{abs} \\ US_{scaled} &= \frac{OH}{PH} * US_{abs}, \end{aligned} \tag{A.11}$$

where OH is object height, PH is picture height, OS_{abs} is the absolute value of overshoot, and US_{abs} is the absolute value of the undershoot. For example, OH would be 100mm for the VC2.

Maximum SFR aims to measure if the ISP has increased the SFR. If the maximum amount of SFR is above 100%, it is clear that the image processor has increased it since that can't be achieved with optics. The maximum SFR value is measured for all the structures where SFR can be calculated from in the TE42-LL-target. These are Siemens stars, slanted edges, and dead leaves-targets.

Euclidean distance (∆E) color error for all patches is calculated for all the 96 color patches of the TE42-LL chart, and then averaged. The calculation is presented in the equation A.12

$$\Delta E_{all} = \frac{\sum_{i=1}^{96} \sqrt{(L_{i,ref} - L_{i,sample})^2 + (a_{i,ref} - a_{i,sample})^2 + (b_{i,ref} - b_{i,sample})^2}}{96}, \tag{A.12}$$

where $L_{i,ref}$ is the value of the L-coordinate of CIE-L*a*b* color space for the ith color patch in the reference data, $L_{i,sample}$ is the value of the L-coordinate of CIE-L*a*b* color space for the ith color patch in the sample data, and so on.

∆E color error for red channel is calculated using the equation A.12, this time excluding all the other patches than red ones.

∆E color error for green channel is calculated using the equation A.12, this time excluding all the other patches than green ones.

∆E color error for blue channel is calculated using the equation A.12, this time excluding all the other patches than blue ones.

∆E color error for skin tones is calculated using the equation A.12, this time excluding all the other patches than those containing skin tones.

Luminance error (ΔL) is calculated for all patches from the L*-component of the CIE-

L*a*b* color space. The subscore is formulated as described in the equation A.13

$$\Delta L = \frac{\sum_{i=1}^{96}(L_{i,ref} - L_{i,sample})}{96}, \tag{A.13}$$

where $L_{i,ref}$ is the amount of luminance for the ith patch in the reference data and $L_{i,sample}$ is the amount of luminance for the ith patch in the sample data.

Chrominance error (Δ C) is calculated from a* and b* components of CIE-L*a*b* color space as follows

$$\Delta C = \frac{\sum_{i=1}^{96}\left(\sqrt{a_{i,sample}^2 + b_{i,sample}^2} - \sqrt{a_{i,ref}^2 + b_{i,ref}^2}\right)}{96}, \tag{A.14}$$

where the components of the same data type are squared, summed together, and then the square root is obtained. This means that the ΔC can be positive or negative. Thus the chrominance error of zero does not mean that the DUT reproduces the chrominance of each patch perfectly. Negative chrominance error means that on average the saturation is lower than the reference. The positive chrominance error means that the saturation is higher on average.

Hue error (Δ H) is the difference in the color tone. This is expressed as follows

$$\Delta H = \frac{\sum_{i=1}^{96}\left(\sqrt{(a_{i,ref} - a_{i,sample})^2 + (b_{i,ref} - b_{i,sample})^2 - \Delta C^2}\right)}{96}. \tag{A.15}$$

White balance (WB) is an important part of color reproduction in imaging devices. A human observer can adapt to different light sources in such a way that a certain color looks similar even though the color stimulus is different due to the light. The camera has to estimate this so-called white point of the illumination and perform white balancing accordingly. The purpose is to change the color processing so that neutral grey patches appear still neutral. The white balance subscore is calculated from TE42-LL's 20 grey patches. It is formed as the average of the CIE-C* value over all of the patches, calculated as follows

$$WB = \frac{\sum_{i=1}^{16} C*_i}{16}. \tag{A.16}$$

CIE-C* value of zero means that the color is completely neutral. In low light conditions, cameras might introduce some color cast to the image with the intention to make the image look more pleasing to the observer. This is considered in the score calculations.

Mean visual noise is calculated using the procedure defined in ISO15739:2013 over 16

grey patches in the TE42-LL test chart. Two of the brightest and darkest are excluded. The subscore is calculated using the A.17

$$VN_{mean} = \frac{\sum_{i=1}^{16} VN_i}{16}, \tag{A.17}$$

where VN_i is the visual noise value of the ith grey patch.

Max visual noise is the maximum visual noise value over all the 20 grey patches. It is calculated with the equation A.18

$$VN_{max} = \max(VN_1, VN_2, ..., VN_{20}), \tag{A.18}$$

where VN_i is the visual noise value of the ith grey patch.

Dynamic range DR describes the camera's ability to reproduce details in shadowed areas and retain highlights. The chart is transparent and lit from behind. This allows calculating the OECF from the mean value of a grey patch luminance channel. The analysis procedure is described in ISO15739:2013. The luminance for each grey patch is then reported using the equation A.19

$$L_i = OECF(Y_i), \tag{A.19}$$

where L_i is the luminance of ith grey patch and Y_i is the mean value of the luminance channel of the patch. The dynamic range is then defined as the ratio of the minimum luminance and smallest the luminance value that resulted in the saturation of the pixel. Luminance is the scene which produces the SNR value of 3 is referred to as minimum luminance. This is the threshold where users can still differentiate image content from the noise, and information isn't lost. If the SNR value of 3 is not reached due to noise filtering of the ISP the L_{min} is extrapolated. The dynamic range subscore is then defined as follows

$$DR_{[f-stop]} = \frac{\log_{10}\left(\frac{L_{sat}}{L_{min}}\right)}{\log_{10}(2)}, \tag{A.20}$$

where $f - stop$ denotes that the DR, and luminance values, are reported in the unit f-stop. The test target is also captured using three different exposures, forced by tapping a certain region in the camera application of the DUT. These regions are the background, bright region, and dark region. Scene lighting conditions are the same as described in the section 3.2.

Intensity shading measures a loss in the luminance component of a specific region of

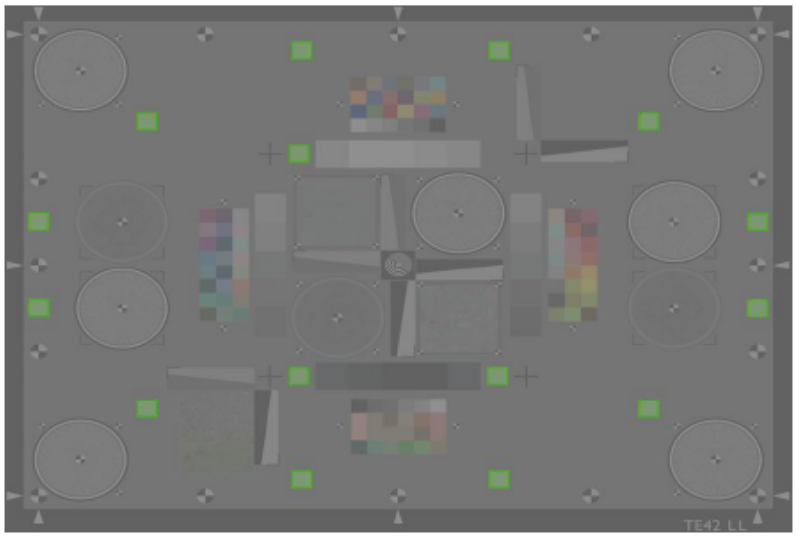

Figure A.5. *The regions of interest used in the shading analysis [50].*

interest (ROI). These ROIs are presented in figure A.5. For $Shading_{intensity}$-subscore, the mean value of the luminance channel of the 23 ROIs is calculated, and the final result is reported as stated in the equation

$$Shading_{intensity} = \frac{\log_{10}\left(\frac{L_{sat}}{L_{min}}\right)}{\log_{10}(2)}, \tag{A.21}$$

where L_{min} is the lowest luminance value and L_{max} is the maximum luminance value.

Color shading is measured using the same patch as for intensity shading, presented in figure A.5. The $Shading_{color}$-subscore is the maximum color difference computed for each patch in the CIE-L*a*b* color space. The equation

$$Shading_{color} = \max\left(\sqrt{(a_1 - a_{ref})^2 + (b_1 - b_{ref})^2}, ..., \sqrt{(a_{23} - a_{ref})^2 + (b_{23} - b_{ref})^2}\right), \tag{A.22}$$

where a_{ref} and b_{ref} are reference values calculated as the mean values of CIE-a* and CIE-b* components calculated over all patches.

TV distortion is defined as the relation between the height of the same object in the image center and in the corner. The equation is defined below A.23

$$TV Distortion = \frac{\Delta H}{H} * 100, \tag{A.23}$$

where H is the distance between the top center and the bottom center position markers and ΔH is the difference between H and the average distance between the left and right

top and bottom position markers.

APPENDIX B: IMAGE COMPARISONS OF DIFFERENT IMAGE QUALITY TUNING SOLUTIONS

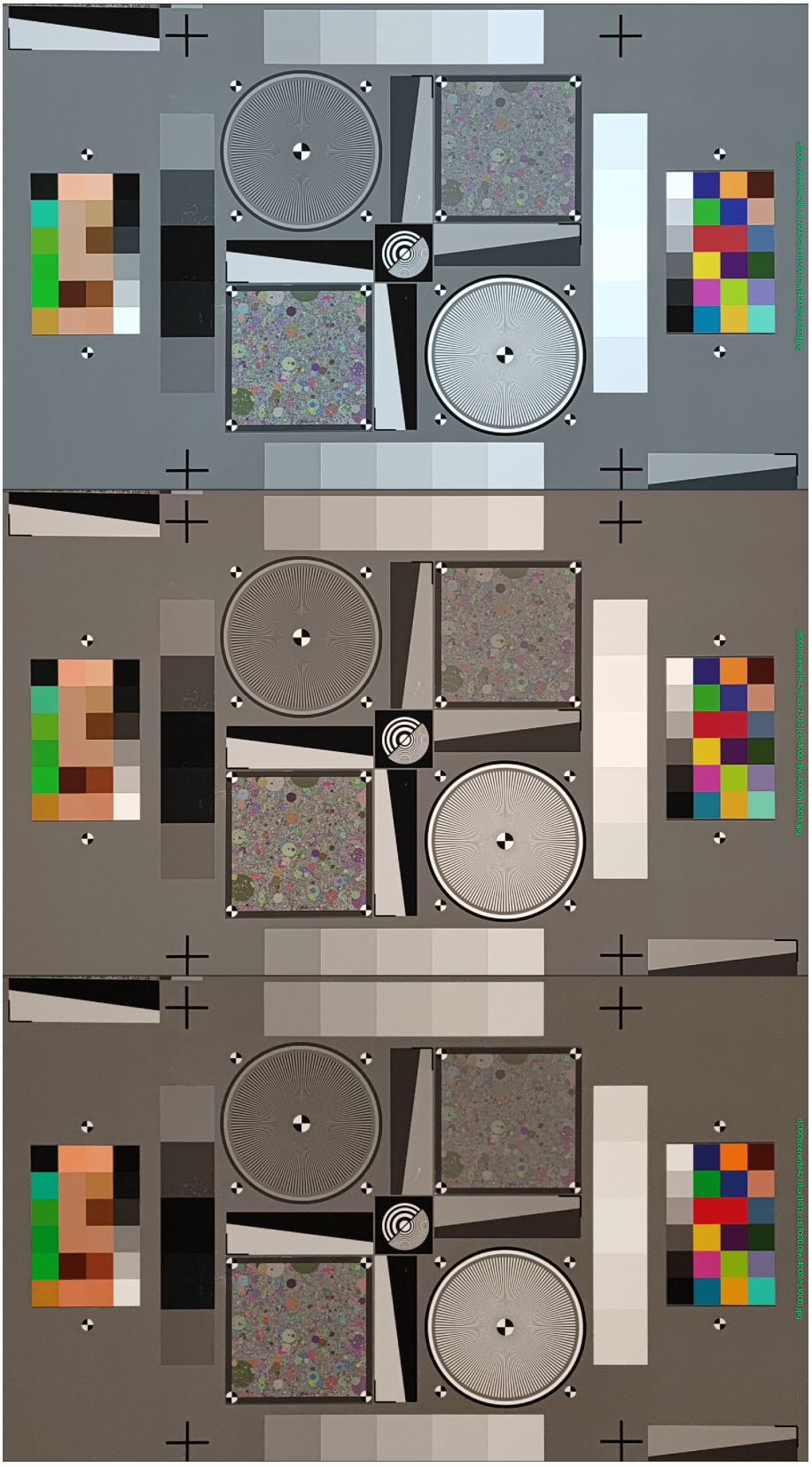

Figure B.1. *TE42-LL images of tuning solution 000.*

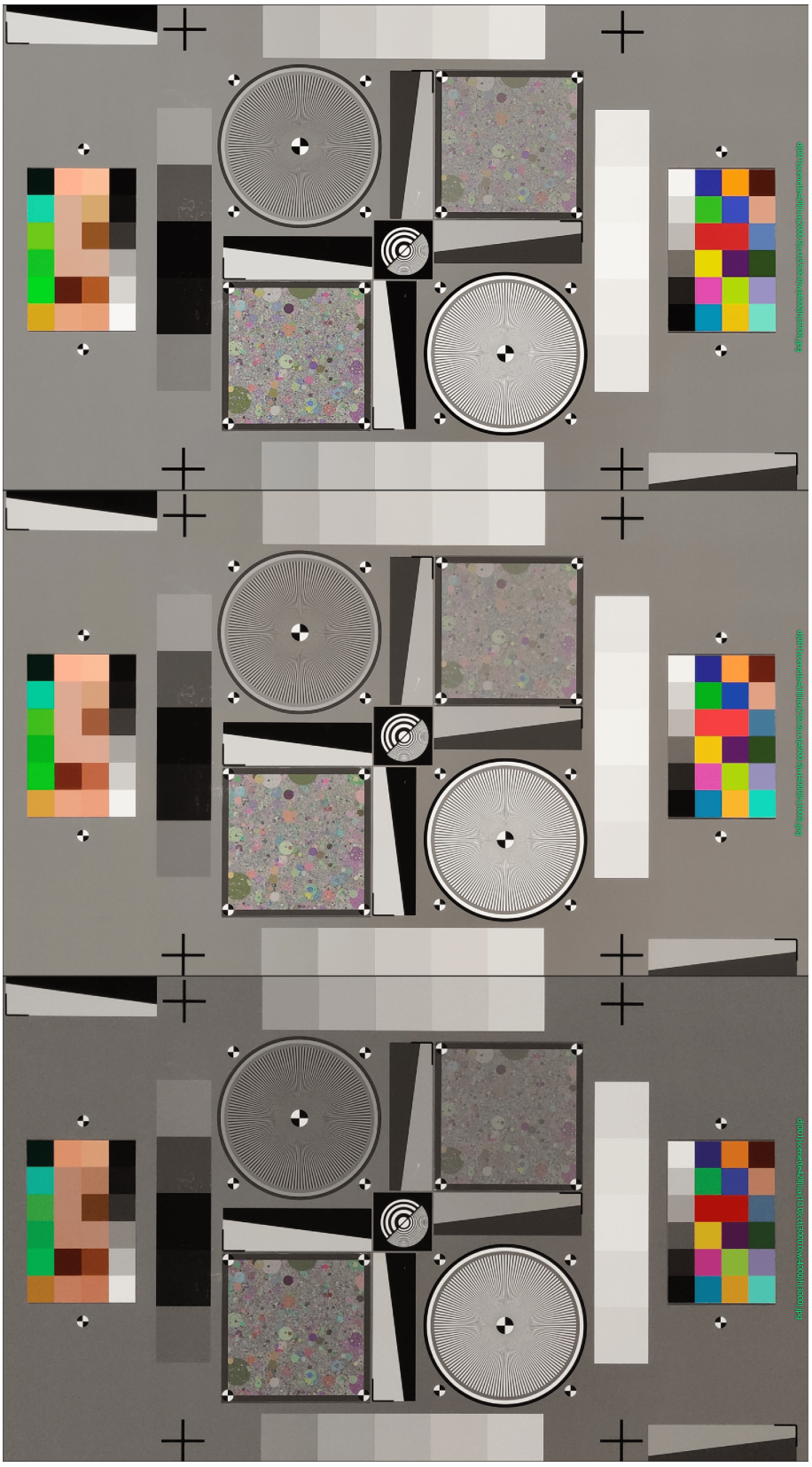

Figure B.2. *TE42-LL images of tuning solution 001.*

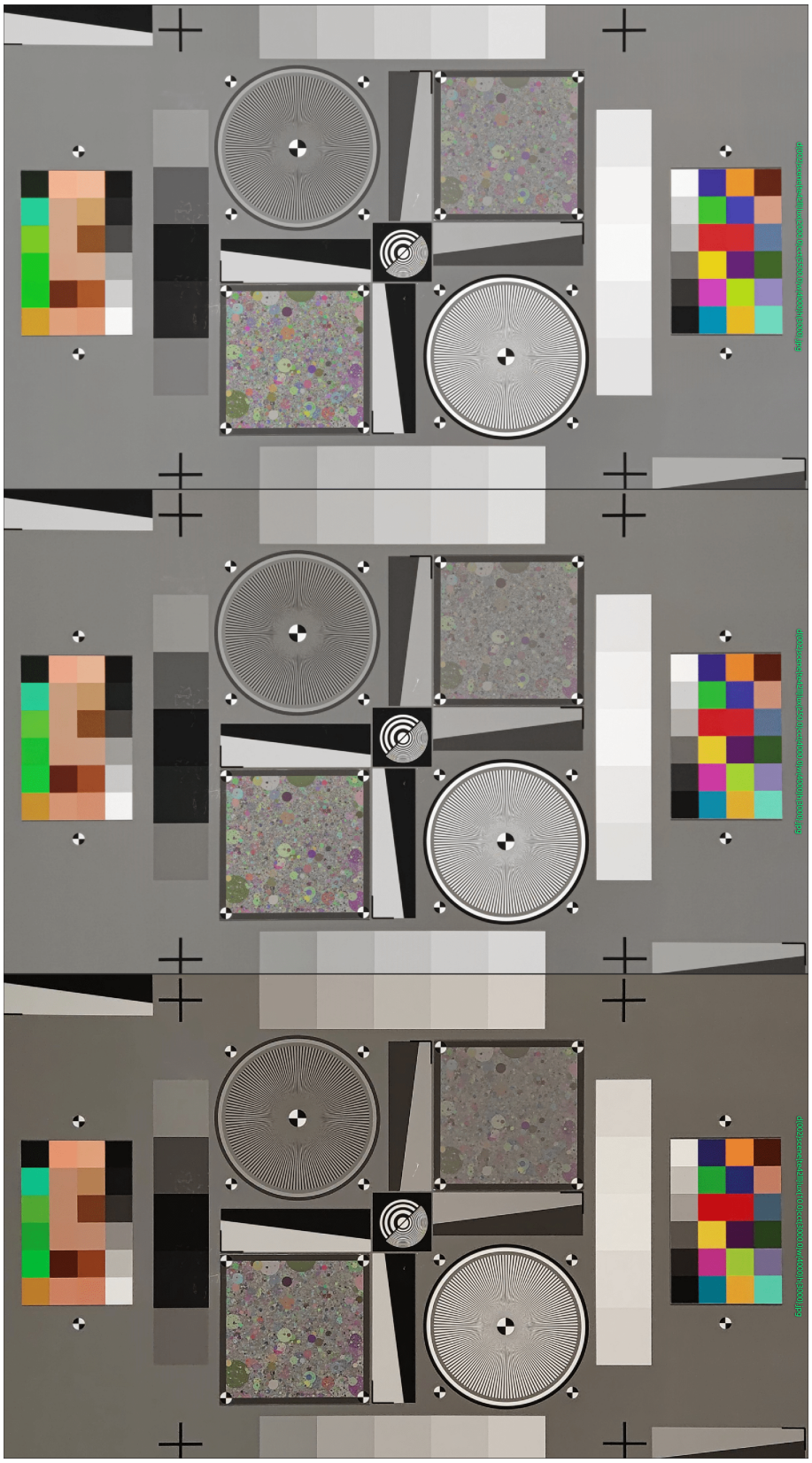

Figure B.3. *TE42-LL images of tuning solution 002.*

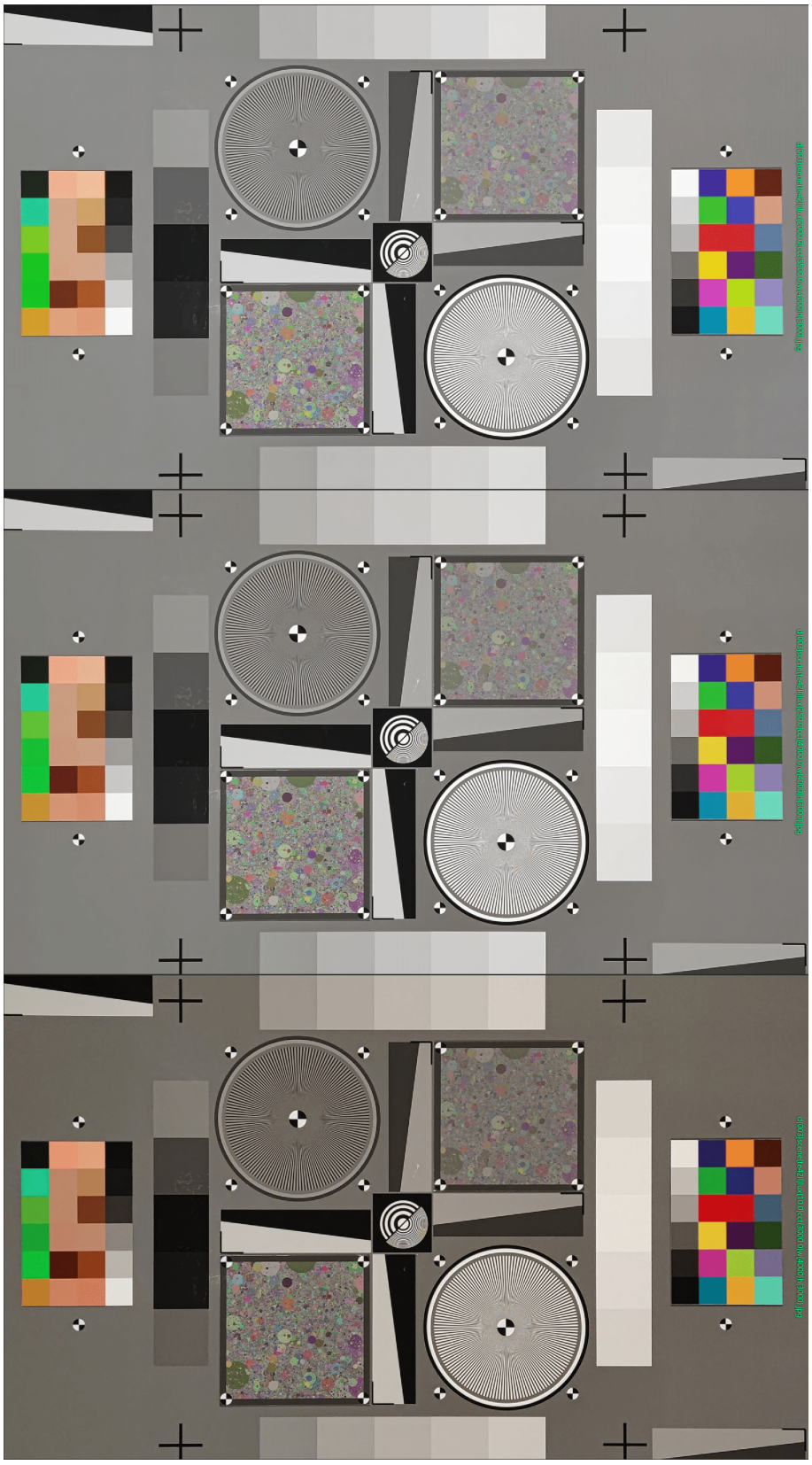

Figure B.4. *TE42-LL images of tuning solution 003.*

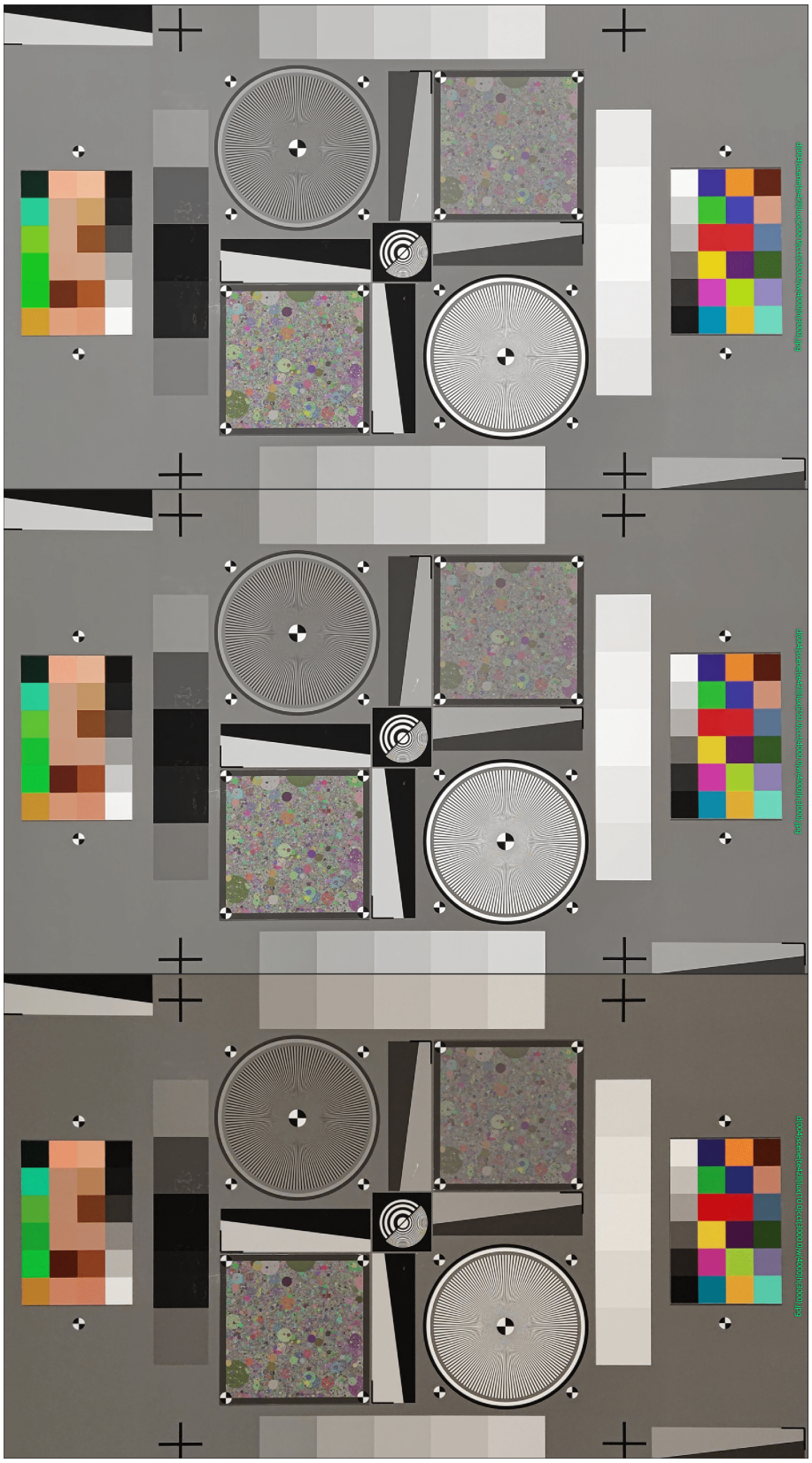

Figure B.5. *TE42-LL images of tuning solution 004.*

Figure B.6. *Superchart images of tuning solution 000.*

Figure B.7. *Superchart images of tuning solution 001.*

Figure B.8. *Superchart images of tuning solution 002.*

Figure B.9. *Superchart images of tuning solution 003.*

Figure B.10. *Superchart images of tuning solution 004.*