Tampere University

Jere Aho

QUALITY ASSESSMENT OF MOBILE PHONE VIDEO STABILIZATION

Master of Science Thesis
Faculty of Information Technology and Communication Sciences
Examiners: Joni Kämäräinen
Veli Tapani Peltoketo
January 2023

ABSTRACT

Jere Aho: Quality assessment of mobile phone video stabilization
Master of Science Thesis
Tampere University
Computing Sciences
January 2023

Smartphone cameras are used more than ever for photography and videography. This has driven mobile phone manufacturers to develop and enhance cameras in their mobile phones. While mobile phone cameras have evolved a lot, many aspects of the mobile phone camera still have room for improvement. One is video stabilization which aims to remove unpleasant motion and artifacts from video. Many video stabilization methods for mobile phones exist. However, there is no standard video stabilization quality assessment (VSQA) framework for comparing the performance of the video stabilization methods.

Huawei wanted to improve the video stabilization quality of their mobile phones by investigating video stabilization quality assessment. As a part of that endeavor, this work studies existing VSQA frameworks found in the literature and incorporates some of their ideas into a VSQA framework established in this work. The new VSQA framework consists of a repeatable laboratory environment and objective sharpness and motion metrics.

To test the VSQA framework, videos were captured on multiple mobile phones in the laboratory environment. These videos were first subjectively evaluated to find issues that are noticeable by humans. Then the videos were objectively evaluated with the objective sharpness and motion metrics. The results show that the proposed VSQA framework can be used for comparing and ranking mobile devices. The VSQA framework successfully identifies the strengths and weaknesses of each tested device's video stabilization quality.

Keywords: Video, stabilization, mobile phone, camera, quality assessment

TIIVISTELMÄ

Jere Aho: Älypuhelimen kameran videon stabiloinnin laadun mittaaminen
Diplomityö
Tampereen yliopisto
Tietotekniikka
Tammikuu 2023

Älypuhelimien kameroita käytetään nykyään valokuvaukseen enemmän kuin koskaan. Tämä on saanut älypuhelimien valmistajia kehittämään heidän puhelimiensa kameroita. Vaikka paljon edistystä on tapahtunut, niin moni älypuhelimen kameran osa-alueista kaipaa vielä kehitystä. Yksi heikoista osa-alueista on videostabilointi. Videostabiloinnin tarkoitus on poistaa videosta epämiellyttävä liike. Monia ratkaisuja löytyy, mutta mitään standardoitua tapaa vertailla eri stabilointi ratkaisuja ei ole.

Huawei haluaa parantaa tuotteidensa videostabiloinnin laatua. Saavuttaakseen tämän tavoitteen, tässä työssä tehdään katsaus kirjallisuudesta löytyviä videostabiloinnin laadun mittausmenetelmiä ja jalostetaan näistä ideoita, joiden avulla kehitetään oma videonstabiloinnin laadun mittausmenetelmä. Menetelmä koostuu toistettavasta laboratorioympäristöstä, jossa voi kuvata heiluvia videoita eri älypuhelimilla. Näitä videoita vertaillaan objektiivisesti mittaamalla videoista terävyyttä ja liikkeen miellyttävyyttä.

Työn videostabiloinnin laadun mittausmenetelmää testattiin kuvaamalla toistettavassa laboratorioympäristössä usealla älypuhelimella videoita, joissa on simuloitua käden tärinää. Ensin kuvattuja videoita arvioitiin ja vertailtiin subjektiivisesti, jotta niistä löytyisi ongelmat, joita videostabilointi ei ole onnistunut korjaamaan. Tämän jälkeen videoita arvioitiin objektiivisilla terävyys- ja liikemittareilla. Tulokset osoittavat, että työssä esitetty videostabiloinnin laadun mittausmenetelmää voidaan käyttää eri älypuhelimien videostabilointimenetelmien vertailuun. Työn mittausmenetelmä onnistui havaitsemaan eri video stabilointimenetelmien vahvuudet ja heikkoudet.

Avainsanat: Video, stabilointi, älypuhelin, kamera, laadun mittaaminen

PREFACE

This paper was created as a part of the Tampere University signal processing master's thesis seminar. This work in video stabilization quality assessment was conducted as a part of my work at Imaging System Performance at Huawei Technologies Finland.

I would like that thank all my colleagues and supervisor for all the help with this thesis. It has been a tremendous learning opportunity which I greatly appreciate. Last but not least, a big thanks goes to family and friends for their support in this endeavor.

Tampereella, 31st January 2023

Jere Aho

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

3A	Denotion for a group 3 image algorithms: auto white balance, auto exposure and auto focus
AE	Auto exposure. An image processing algorithm
AF	Autofocus. An image processing algorithm
AWB	Auto white balance. An image processing algorithm
CFA	Color filter array
CIE-XYZ	A perceptual color space
CMOS	Complementary metal-oxide semiconductor. A part in the camera module
DIS	Digital image stabilization. A image stabilization method
DR	Dynamic range. An image and video illumination characteristic
DSLR	Denoting or relating to a camera that combines the optics and mechanisms of a single-lens reflex camera with a digital imaging sensor, rather than photographic film.
EIS	Electronic image stabilization. A image stabilization method
FOV	Field of view
FPS	Frames per second. A video characteristic
HVS	Human visual system
Hz	Hertz, Frequency unit
lux	One lumen per square metre, Illumination measurement unit
IMU	Inertial measurement unit. A motion measurement device in mobile phones
ISO	International Organization for Standardization
ISP	Image processing pipeline
JND	Just noticeable difference
K	Kelvin, Color temperature measuring unit
MP	Megapixel. The number pixels in a camera module
MPEG	Moving Picture Experts Group. A video compression standard

MTF	Modular transfer function. A function used for calculating the sharpness image quality attribute
NR-VQA	No reference video quality assessment
NR-VSQA	No reference video stabilization quality assessment
OIS	Optical image stabilization. A image stabilization method
RAW	A term for unprocessed images and videos
SFR	Spatial frequency response
SOTA	State-of-the-art
SSS	Sensor-shift stabilization. A image stabilization method
SVM	Support vector machine
VCM	Voice coil motor. A part of a camera module
VQA	Video quality assessment
VSQA	Video stabilization quality assessment

1. INTRODUCTION

In 2017, there were 4.435 billion smartphone users worldwide. In 2022, that figure has increased to 6.648 billion which is 83 % of the world's population. This shows the rapid adoption of smartphones. [1] Due to the increase in the network speed and the widespread adoption of smartphones, users share more videos online than ever before [2]. Smartphone manufacturers have recognized this trend and therefore are incentivized to improve the quality of their mobile phone cameras to gain a competitive advantage. [3]

One area of mobile phone cameras that could be improved is video stabilization. Video stabilization aims to remove unpleasant motion like jitter from video, making videos more pleasant to watch. Figure 1.1 illustrates the goal of video stabilization. Different large and expensive video stabilization equipment have been used but due to their high cost, regular consumers will not utilize them. Instead, consumers will rely on the video stabilization mechanisms available in mobile phones, which are less robust than the large and expensive equipment. Rapid motion can cause mobile phone video stabilization methods to fail. [4]

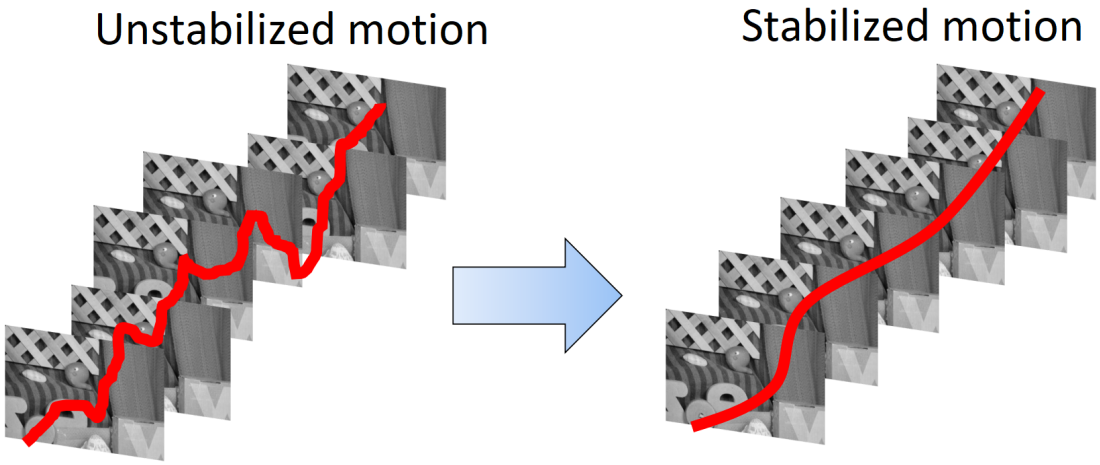


Figure 1.1. *The goal of video stabilization [5]*

Video stabilization quality assessment (VSQA) measures the performance of video stabilization. VSQA has been studied since the 1970s, and more extensively for the past 20 years. Many objective metrics measuring video stabilization quality have been proposed throughout the years. Still, no objective metrics correlate very well with the subjective

perception of the human visual system. [5] [6] The best reported correlation between an objective metric and subjective perception is 0.23, which is low [7].

This work aims to create a better framework to accurately, reliably, and repeatably measure video stabilization quality in mobile phones. First, this work surveys existing video stabilization evaluation methods found in the literature. Based on these methods, some objective metrics are proposed. Their applicability for mobile phone VSQA is tested on videos captured in a controlled indoor laboratory environment.

2. MOBILE PHONE CAMERA FUNDAMENTALS

This section covers the basics of how a mobile phone camera module works. First, some of the physical camera module components used to capture a RAW image are covered. Then some algorithms to process the RAW in the image processing pipeline are discussed. Also, some artifacts that these algorithms may generate are covered. Finally, this section will cover how camera zooming works in mobile phones.

2.1 Camera module

A camera module houses the image sensor, color filter array, lens array, voice coil motors (VCM), and other components, as shown in Figure 2.1. Modern mobile phones are very compact, which limits the camera module size that can fit inside the phone. Mobile phone cameras mostly have fixed focal lengths since the lenses in the lens array do not have space to move. For this reason, mobile phones often have multiple camera modules with different focal lengths. Common camera modules include wide, main, and telephoto, each suited for different tasks. [3] [2] For example, the Apple iPhone 12 Pro Max has a wide camera with a focal length of 26 mm and a telephoto camera with a focal length of 65 mm. Hence the phone can achieve a 2.5x magnification level with optics [8].

The camera sensor is part of the camera which converts an analog signal, photons, to a digital signal. The camera sensor has a mosaic of photodiodes, each capturing photons. The sensor's AD converter converts this analog signal to a digital signal, which can be further processed. A commonly used camera sensor in mobile phones is the CMOS sensor with rolling shutter [10] [3]. This can make images with fast-moving objects look tilted, as shown in Figure 2.2.

Before photons arrive at the camera sensor, they first travel through a color filter array (CFA), as illustrated in Figure 2.3. The CFA forms a mosaic on top of the photodiodes. One of the most commonly known mosaics is the Bayer mosaic shown in Figure 2.3. In the Bayer mosaic, each color filter filters arriving photons according to either short, medium or long wavelength bands. This is because the human eye's retina has three types of cone cells, each sensitive to short, medium and long wavelength bands, as shown in Figure 2.4. These wavelengths correspond to the blue, green and red color channels humans perceive. [12] The HVS is most sensitive to the medium wavelengths

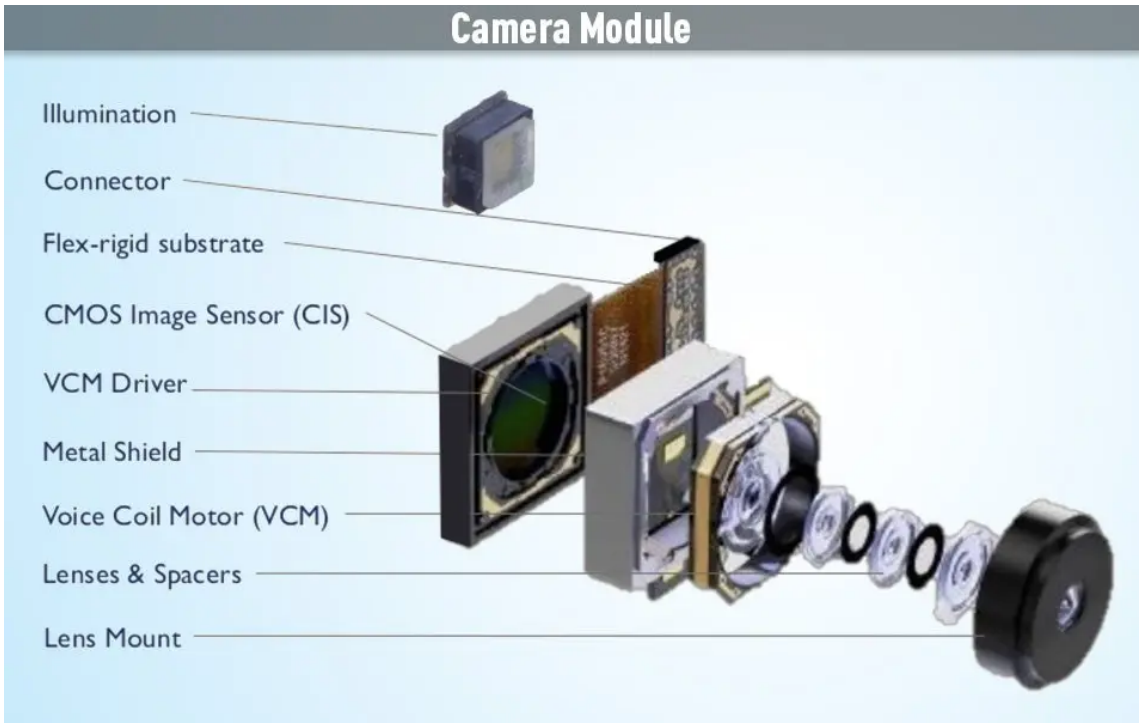


Figure 2.1. *The basic structure of the camera module [9]*

Figure 2.2. *Image of a fast-moving object taken with a rolling shutter camera [11]*

which is the green color channel [10]. This is why the Bayer mosaic has double the number of color filters for the green color channel [10]. The image obtained from the image sensor is often called a Bayer image. [3]

Before the photons arrive at either the CFA or the camera sensor, they go through a lens

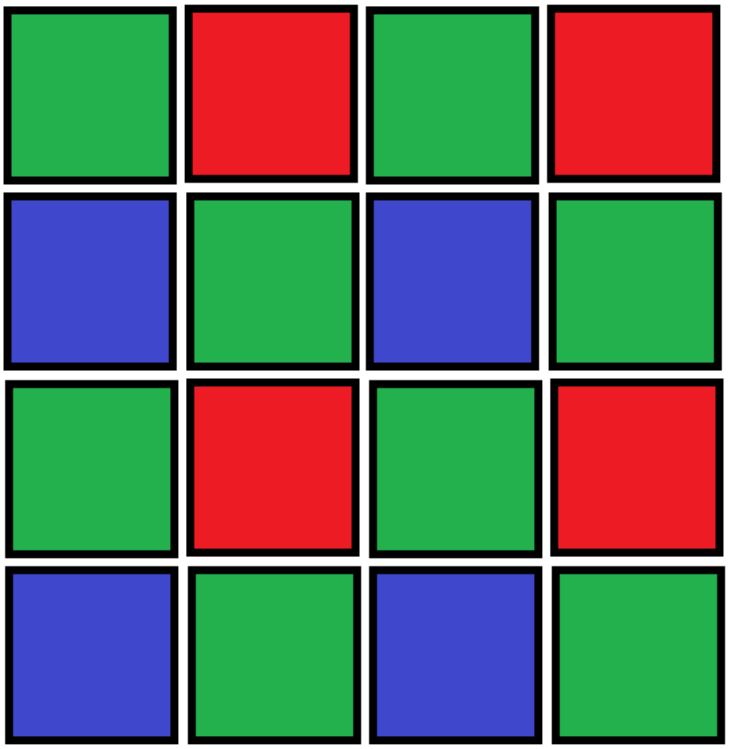


Figure 2.3. *Bayer color filter array mosaic*

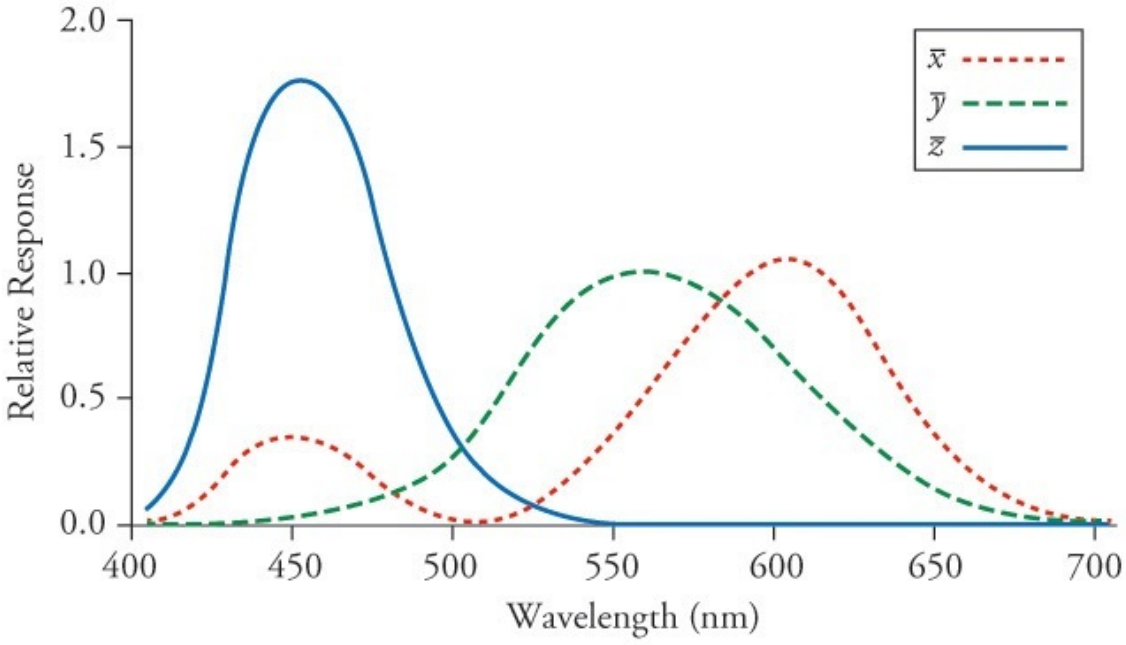


Figure 2.4. *Spectral response of the human visual system cone cells. The z, y, and x lines are the short, medium, and long wavelengths, respectively. [12]*

array. The lens array often has 5 - 6 lenses [10], each with different curvature, diameter, and focal length. By combining multiple different lenses into a lens array, photons can be focused on the image sensor accurately. In recent years, mobile phone manufacturers have been able to fit larger zoom lenses into mobile phones. This has increased the

camera’s focal length, enabling images or video capturing at larger zoom ratios with better quality.

2.2 Image processing pipeline

Modern mobile phones have limited light collection ability since they have a limited aperture due to the desire for a thin form factor. This has led manufacturers to develop computationally demanding image processing pipelines (ISP). [2] [5] The goal of an image processing pipeline is to produce a visually pleasing image or video from a RAW image or video which is the unprocessed data received from the camera sensor. An ISP is implemented by applying many cascaded algorithms on the RAW image or video. The cascaded nature entails that the algorithms at the beginning of the image processing pipeline will have a larger effect on the result. Also, changing the parameters of one algorithm may affect the performance of the following algorithms. [10] On top of this, image processing is computationally demanding. Therefore, the ISP algorithms must be computationally efficient. The cascaded nature and efficient computational demand make tuning an ISP a difficult task.

Different companies design their ISPs to their needs. This means that there is no single way of implementing an ISP. However, the algorithms used within the ISPs are often quite similar in principle. Figure 2.5 shows a simplified process diagram of the ISP. The ISP algorithms can roughly be divided into preprocessing, conversion, and manipulation algorithms, which have been color-coded in yellow, blue, and green colors, respectively, in Figure 2.5 [10]. Next, the algorithms which affect video stabilization performance are discussed.

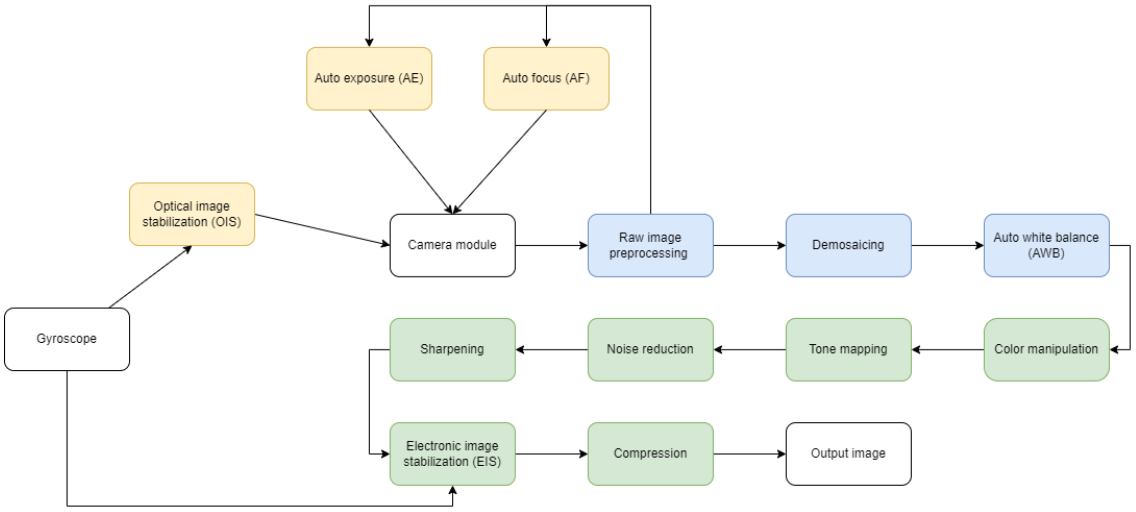


Figure 2.5. *Image processing pipeline.*

2.2.1 Preprocessing algorithms

Preprocessing algorithms are algorithms that are executed during light collection. They are time-critical algorithms and control camera functionality [10]. These algorithms can

significantly affect the performance of a camera [13]. Common and important preprocessing algorithms include autofocus (AF), auto exposure (AE), and optical image stabilization (OIS). OIS is covered in Section 3.1.2

Autofocus

Autofocus (AF) plays a crucial role in ensuring that images captured by a camera are properly focused. It adjusts the lens to place the subject of the photo or video precisely at the focal plane, which results in sharp, detailed images and videos. If the lens is out of focus, the resulting images and videos may appear blurry. Autofocus is one of the first algorithms used in the image processing pipeline making it an essential part of the camera's functionality. Two common types of AF methods exist, active and passive AF.[13] Certain phones combine active AF and passive AF, often referred to as hybrid AF [14]. The Huawei P50 Pro's wide camera uses hybrid AF [15].

Active AF systems have an active focus component, often an infrared or ultrasonic sensor that measures the distance between the scene and the camera. This information is then used to calculate the optimal focal point and adjust the lens accordingly. The downside of using active AF is that it consumes much power, is expensive to manufacture, and takes up space. Also, reflective surfaces like glass and mirrors can make distance measurements fail. [13]

Passive AF measures the sharpness in parts of the image and determines whether the image is out-of-focus. Then the lens can be adjusted accordingly. Unlike active AF, passive AF does not require extra components, making it a cheaper and less space-demanding method. [13]

Auto exposure

Controlling exposure is important for capturing images but even more important for video [16] since video illumination changes are more apparent. If the exposure is not properly adjusted according to the scene's illumination, the resulting image or video may be under or over-exposed. The sharpness and colors may also be affected [10]. There are mainly three methods for automatic exposure control: aperture control, automatic gain control (AGC), and electrical shutter control. The first method is not usually utilized in mobile phones since their camera aperture is usually constant. [3] AGC controls the gain of the analog signals. In cameras equipped with a CMOS sensor, such as mobile phones, the electrical shutter control adjusts the camera's exposure time. [16] When gain is increased, temporal noise is amplified [10].

2.2.2 Conversion algorithms

After light is captured and transformed into a digital signal, a RAW image or video is formed. Conversion algorithms such as demosaicing and auto white balance (AWB) are carried out to turn the RAW image into a RGB image. This RGB image or video is then converted into the CIE-XYZ color space, which is a perceptually uniform color space that is independent of any device or monitor.

Demosaicing

Demosaicing is the first step in converting the Bayer image into an image with red, green, and blue color channels. Demosaicing interpolates the missing colors of each pixel from its neighboring pixel colors in the CFA. For example, if the pixel color is green, then the red and blue colors are interpolated from adjacent pixels of the same colors. Many demosaicing algorithms exist. Some of the most common ones are nearest neighbors, bilinear interpolation, and cubic spline interpolation. If the interpolation is inaccurate, blur and color errors may appear. [3] [10]

Auto white balance

A camera adjusts its white balance to mimic the human visual system's (HVS) chromatic adaptation mechanism. The mechanism adjusts the perceived colors of a scene according to the scene's illumination. Auto white balance (AWB) is an automatic algorithm in mobile phone cameras to estimate the scene's illumination and adjust the white balance accordingly. AWB estimates the red, green, and blue color channels' responses to scene illumination from the captured image. [3] The gray world algorithm is a simple illumination estimation method. It assumes that an image's red, green, and blue color channel values are, on average, the same. If they are not, the deviation from the gray value can be used to estimate the color cast of the light source in the image. [17] Once the responses are obtained, each pixel's red, green, and blue values are divided by the corresponding illumination value [3]. Color errors may occur if the illumination estimation fails. [10]

2.2.3 Manipulation algorithms

After the RBG image or video has been converted into CIE-XYZ color space, manipulation algorithms are used to further process the image or video according to the manufacturer's needs. Some mobile phone manufacturers aim for visual pleasantness, while others aim for realism. Many more algorithms exist, but common manipulation algorithms include color manipulation, tone mapping, sharpening, denoising, and compression. These algorithms and the artifacts they might generate will be covered in more detail.

Color manipulation

Color manipulation is commonly the first applied manipulation algorithm. Color manipulation aims to change the image's colors to be visually appealing. 3D lookup tables (LUT) can map colors to the desired color space. The preferred 3D LUT can vary geographically. [3]. Hence, color manipulation is often tailored to a geographical region.

Tone mapping

Next, tone mapping uses a 1D lookup table to map each color channel to a more visually appealing tone via an increase in contrast. Also, tone mapping decreases each color channel's bit count's dynamic range (DR). [3] This decrease in DR can cause a contouring artifact that shows up edges in areas that should be uniform [10].

Sharpening

Sharpening is often done after noise reduction. Sharpening does not have any exact visual model, so it is hard to apply the appropriate amount of sharpening. This fact makes the development of sharpening algorithms hard. A simple method for sharpening is unsharp masking. [3]. Over-sharpening may lead to ringing artifacts in high-contrast areas like edges [10].

Denoising

The denoising algorithm aims at removing undesired noise from an image. If the noise reduction is overdone, the result will be an unpleasant blurry image. If too little noise is removed, then noise will be present. Textured areas can be mistaken for noise making denoising difficult. [10] Denoising can be done in the conversion, the manipulation, or both of these phases. [3]

Compression

The final stage of the image processing pipeline is compression. Compression aims to reduce the size of the stored file by removing spatially and temporally redundant information. Information is redundant if humans do not notice a difference when the information is removed. For example, parts of a video frame that have high spatial frequency and color components. Videos often have static and dynamic parts in frames. Static parts visible in multiple consecutive frames do not need to be encoded multiple times as the information stays the same. Thus, only dynamic parts of those frames need to be encoded. [12].

Videos are commonly compressed using the MPEG video codec [12]. Since compression removes information, artifacts might appear. Common artifacts which appear in videos

compressed with MPEG include blocking, ringing, blur, and flickering [18][10]. Blocking appears as blocks segmented by vertical and horizontal edges due to coarse quantization of frame blocks. Ringing appears along edges that have a high spatial frequency. The blur effect can occur when there is an absence of high frequencies. Flickering is a temporal problem in video where the image rapidly changes or flashes. It occurs when there is a difference between I-frames, B-frames, and P-frames used in video compression. These frames contain information needed to display a complete video and when they don't match up, the video flickers. [19]

2.3 Zoom

The zoom capability aims to magnify a chosen target by decreasing the camera's field of view [20]. Mobile phone zooming methods can be either through hardware, software, or a combination of them. These methods are also often referred to as optical zoom, digital zoom, and hybrid zoom, respectively.

Digital zoom is a software-based solution. It works by cropping a region in the center of a full-resolution image. This cropped region is then upscaled to the full resolution of the original image with interpolation [3]. A commonly used interpolation method is bicubic interpolation. [21, p. 276] The downside of using interpolation is that it decreases image quality and often causes artifacts, like jagged diagonal edges [21, p. 462].

Optical zoom is a hardware-based solution implemented through camera optics. Optical zoom is preferred since it does not degrade image quality like digital zoom. Traditional cameras, like DSLRs, implement optical zooming by moving lens elements. As discussed before, mobile phones have space constraints that give little room for moving lens elements. [20] Phone manufacturers have had to design periscope cameras to enable larger fixed zoom magnification levels. Light is guided through a prism to periscope the camera's lens array of 1 or 2 lens elements laid horizontally in the phone. [3]

Hybrid zoom is the most commonly used zooming method in modern mobile phones since neither optical zoom nor digital zoom alone can produce a good level of quality for all the zoom magnifications. Hybrid zoom is a combination of optical zoom, digital zoom, and software algorithms to produce better zoom. Hybrid zoom tries to emulate the DSLR's optical zooming capability. Mobile phones have multiple cameras [2] with varying fixed focal lengths and features. Hybrid zoom can utilize details from each phone's cameras and merge the details intelligently when forming the image [20]. [22] Software can further enhance an image through computational photography algorithms like super-resolution. [3]

3. VIDEO STABILIZATION

Mobile phone cameras often use a variety of methods for video stabilization. Video stabilization methods are either implemented through software, hardware-based, or a combination of the previous two. To evaluate a mobile phone's video stabilization performance various video stabilization quality assessment (VSQA) objective metrics found in the literature are discussed. Their correlation with subjective perception is discussed as well.

3.1 Video Stabilization Methods

Video stabilization methods attempt to remove unwanted and unpleasant motion and distortions from video while retaining intentional motion, such as panning. [6] Often, it can be considered as a task of stabilizing the original path of the video as shown in a previous section in Figure 1.1. CMOS sensors have a rolling shutter which can produce unpleasant artifacts or geometric deformations like motion blur, shear, and wobbling in the video if it contains fast movements and a short exposure time is used [23]. [5] Rolling shutter-related deformations can be referred to as the "jello effect" as they bend straight lines and make objects appear elastic [24] [25]. [10]

Motion blur can be identified as a blur that moves in the same direction but appears unpredictably [24]. Motion blur may appear when capturing an image or video of a target moving faster than the camera's shutter speed [26]. The effect of motion blur also increases with a higher resolution [26]. Figure 3.1 has a fast-moving bus that has generated motion blur in the captured image. Wobbling occurs when the frequency of the motion inflicted on the camera exceeds the camera's frame rate. Figure 4.7 highlights a wobbling frame. Wobbling is discussed more in Section 4.2.3. Shear or skew appears as a rectangular image frame warped into another shape like a rhombus, as shown in Figure 2.2. [23]

Video stabilization can be achieved through a hardware-based solution, software-based solution, or a combination of the two. The process of stabilizing a video can be divided into three parts: motion estimation, motion compensation, and motion inpainting [6]. This section will cover some of the available solutions for mobile phones and compare their advantages and disadvantages.

Figure 3.1. *Motion blur*

3.1.1 Electronic image stabilization for videos

Electronic image stabilization (EIS), also known as Digital image stabilization (DIS), is a software solution for video stabilization. The basic idea is to use movement information to align video frames in post-processing to produce a pleasing watching experience. The movement information is often obtained either from the mobile phone's inertial measurement unit (IMU) sensor measurements [27] [5] or is calculated from frame to frame differences.

Motion estimation methods from video frames can be categorized into pixel-based methods, block matching, and feature tracking. Pixel-based methods estimate pixel movement by tracking pixel illumination changes. This approach works when the illumination conditions stay constant between frames. If they do not, other factors need to be considered. Pixel-based methods include transformation estimation and optical flow. Optical flow can be divided into sparse and dense methods. Dense optical flow is more accurate than sparse optical flow since it estimates movement for every pixel, while sparse optical flow does not estimate motion for monotonous areas. However, dense optical flow is computationally more demanding. The only difference between pixel-based and block-matching methods is the use of blocks of pixels instead pixels. Block-based methods offer a good trade-off between computational complexity and accuracy but are prone to aperture and correspondence problems. Feature matching methods use a feature detector to track the movement of an object in the scene. Common feature detectors include

Kanade–Lucas–Tomasi (KLT), Harris corners, SIFT, and SURF. [5] Common features to track include human faces [28]. Feature matching offers fast calculation speed with great accuracy. However, it can fail in scenes with large solid color regions which do not give features to track.

The difficulty of utilizing movement information comes with differentiating between the intentional movement to retain and unintentional movement to remove in the stabilization process. Unintentional motion in mobile phones videos can be caused by high-frequency hand shakiness when standing still or from low-frequency up-and-down movement when walking or running [29] [5]. An intentional motion could be considered as cinematic motion. Professionals often use tools like tripods, dollies, and other stabilizers to remove unintentional motion [7] and achieve cinematic effects with intentional movement. Cinematic shots can often be categorized into three types: still shots, tracking shots, and smooth transitions. Still shots have no movement. Tracking shots like panning have constant movement. Smooth transitions are ones with zero acceleration. [7] The ideal panning speed is dependent on focal length, frame rate, and total panning angle [30]. Some phones provide a cinematic video shooting mode [2].

Most EIS algorithms slightly crop the frames since they are warped slightly to smoothen motion in the video, as shown in Figure 3.2. However, cropping discards information which is not ideal. Some recent EIS methods have eliminated the need for cropping with motion inpainting while retaining good stabilization performance. One of the approaches utilized deep learning to interpolate virtual frames between captured frames, eliminating the need for cropping [31]. Another approach is to have a large image sensor since there is more area to choose a crop from and the crops can have a higher resolution [32]. For example, the GoPro10 action camera provides excellent stabilization performance with its large 23.6 megapixels (MP) image sensor and video stabilization algorithms [33].

3.1.2 Hardware-based image stabilization

Hardware-based solutions in mobile phones include optical image stabilization (OIS), sensor-shift stabilization (SSS), and micro gimbal stabilization. These hardware-based stabilization methods provide real-time video stabilization and usually use motion estimation information calculated with a gyroscope or accelerometer. They differ in which hardware components are adjusted for motion compensation. OIS rotates the camera module lens array, usually with a voice coil motor (VCM) [10] in the opposite direction of the motion inflicted on it. [34] SSS translates the camera sensor instead of the lens array.

SSS has been used in traditional DSLR cameras for a while but has recently made its way into iPhone 12 Pro Max's rear wide-angle camera [8]. The smartphone manufacturer Vivo has been the first to introduce a micro gimbal camera in its mobile phones. The micro gimbal stabilizes three rotational axes instead of two rotational axes stabilized by

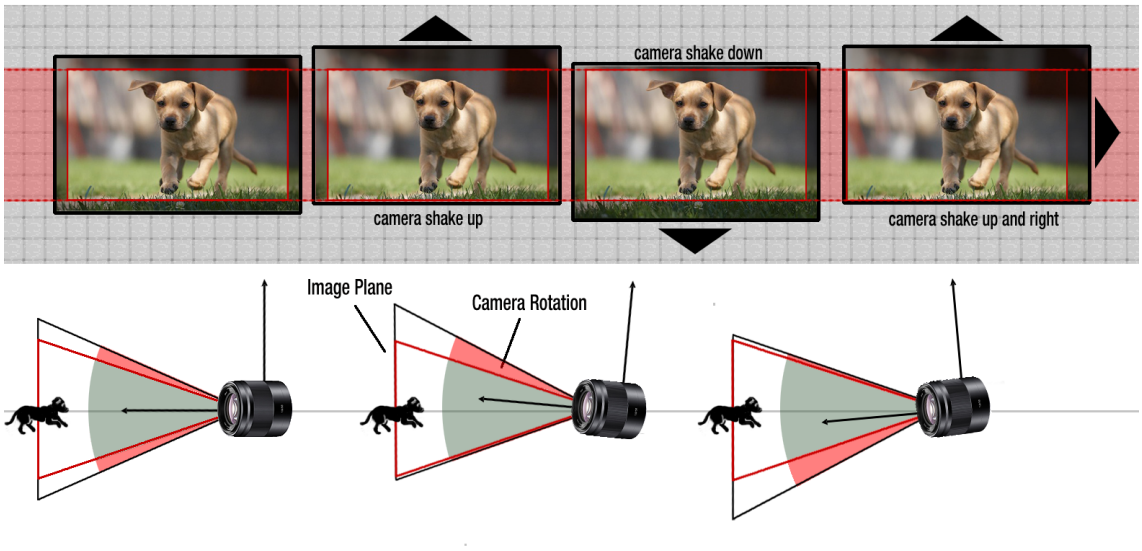


Figure 3.2. *Electronic image stabilization principle illustrated [25]*

OIS [35]. In the past, consumers had to buy an external gimbal for mobile phones to get stabilization in 3 rotational axes. The micro gimbal stabilization has yet to be widely adopted since it costs more and demands more space [36]. The OPPO Find X5 Pro wide camera uses a combination of OIS and SSS [37] which provides stabilization in three rotational axes and two translation axes. However, this combination occupies a large volume in the phone. [38] The Asus Zenfone 9 main camera has gimbal OIS, which allows stabilization in all three rotational and translation axes [39]. The gimbal OIS has excellent stabilization performance. The idea of using gimbal OIS or sensor-shift stabilization in mobile phones is intriguing and merits further investigation into their capability to enhance video stabilization.

3.1.3 Comparison of the stabilization methods

Out of all the hardware-based solutions, OIS is more widely utilized in mobile phones. Sensor-shift stabilization, micro gimbal stabilization, and gimbal OIS have only been used in a few devices. EIS is the most flexible since it can be applied later onto a video [32] and EIS algorithms can be improved without hardware changes.

The camera needs a longer exposure time to collect enough light in low light. Shaky movements during a longer exposure time can negatively impact the quality of the video. OIS can mitigate the shaky movements in real-time, but EIS cannot since it is applied afterward. Thus, OIS is essential for low-light videos. OIS is also better when using a telephoto camera since small shaky movements occurring in real time appear much more prominent in the narrow field of view [34].

OIS and micro-gimbals are more expensive to implement, meaning every smartphone

camera does not necessarily have them [34]. OIS is prone to wobbling since the springs inside the VCM may oscillate at their resonance frequency, causing distortions [10]. The VCM springs may also loosen up over time if exposed to high-frequency vibrations leading to degraded OIS performance resulting in worse image and video quality [40].

Often OIS and EIS are used together to stabilize a video. This is often referred to as hybrid image stabilization (HIS). HIS combines the best parts of EIS and OIS, providing the best video stabilization performance for mobile phone cameras.

3.2 Video stabilization quality assessment methods

This section will cover video stabilization quality assessment methods found in the literature and the challenges of measuring video stabilization. Then a number of metrics that have been used in the frameworks are discussed.

3.2.1 Challenges of measuring video stabilization quality

During the past two decades, many video stabilization quality assessment methods have been introduced in the literature. Even though much progress has been made in the field, there are still many open problems. One is the need for a good perceptual model for measuring video stabilization quality. This issue is mainly due to it being hard to quantify and model the visual discomfort of unintentional camera movement, artifacts from video stabilization algorithms, and other spatio-temporal distortions and artifacts. [41] [5]

Most state-of-the-art (SOTA) methods can only handle simple scenes with low amplitude movement, but not scenes with multiple objects. In addition, assessing the performance of video stabilization in different scenarios is difficult since each scenario may require different evaluation criteria [5].

Subjective user studies are expensive and time-consuming to conduct. Thus, many studies include an objective evaluation of video stabilization but neglect subjective evaluation even though it is of great importance. Even if subjective user studies have been conducted in some works, they often concern only certain parts of video stabilization. Due to this, it is hard to compare whether recent approaches are better than the ones from the past. [5] For these reasons, developing and establishing a general, unified, and well-acknowledged VSQA framework is important.

3.2.2 Existing video stabilization quality assessment methods

This section will go through VSQA metrics found in the literature. These metrics measure the severity of different distortions that might appear in the stabilized video. Also, some general video quality assessment (VQA) metrics will be covered. In the following section,

some of the ideas from these metrics will be used to derive a framework for measuring the video stabilization quality in mobile phones.

VSQA frameworks often measure the quality of video stabilization with the following characteristics: motion characteristics, temporal variations, spatial variations, and overall video quality. Some of the works present objective metrics which have not been validated through subjective user studies.

Some aspects of video stabilization in real-life videos have been measured with objective metrics. Multiple metrics are often used to measure each distortion to gain a complete picture of the video stabilization performance [5]. Some commonly used visual stabilization quality assessment (VSQA) metrics that evaluate the unpleasantness of unintentional motion include the Inter-Frame Transformation Fidelity (ITF), the Inter-frame Similarity Index (ISI), Average Speed (AvSpeed), and Average Acceleration (AvAcc). No-reference video quality metrics like VIDEOO can be used to evaluate the overall video quality. [7]

ITF measures the average peak-to-signal ratio (PSNR) between consecutive frames. ISI measures the average structural similarity index measure (SSIM) between consecutive frames. The idea behind these two inter-frame metrics is that if consecutive frames have a similar appearance, then these metrics should perform well. [7]

The motion-related metrics AvSpeed and AvAcc are derived from motion characteristics. These motion characteristics are often salient key points derived from video frames. AvSpeed is defined as the average difference in movement between consecutive key points. AvAcc is the average difference of difference in movement between consecutive key points. [7]

VIDEOO is a no-reference video quality assessment (NR-VQA) metric that evaluates the quality difference between the original and a stabilized version of the original video via scene statistics. This NR-VQA metric has not been trained on a dataset, making it scene-independent which suits VSQA since different scenes might need different characteristics to measure them. [7]

These objective metrics have very low correlation values with subjective perception. The correlation values are between 0 and 1, where 1 is the best. AvAcc had the highest correlation value of 0.23. The VIDEOO has the second highest correlation of 0.17. ITF, ISI, and AvSpeed had the lowest scores of 0.12, 0.13, and 0.13, respectively. This indicates that there is a need to conduct more studies on new objective metrics and measure their correlation to human's subjective perception of video and video stabilization. [7]

Many NR-VQA models have been proposed, but to our knowledge, only one no-reference video stabilization quality assessment (NR-VSQA) model [29] has been proposed. In [29], they generate a training dataset that contains mean opinion scores (MOS) for videos. They then extract video motion features, including translation, rotation, and scaling. The

motion signal is divided into low-band, mid-band, and high-band frequency compositions. Furthermore, they computed statistical quantities from each sub-band, including mean, variance, skewness, and flatness. Finally, they used the statistical features to train a support vector machine (SVM) to predict video stabilization quality. They achieved good results with this method on their dataset and showed some indication of its generalization ability. However, their method does not consider other aspects of video quality, like sharpness.

One work [24] considers both image and video quality. They measure image sharpness with texture acutance from shaky images containing a dead leaves chart. Acutance correlates well with perceptual perception in images [42], but its applicability for video has not really been studied. They also objectively measure the deforming effect of an electronic rolling shutter on video quality by evaluating the translation, rotation, and shear of 4 points. However, they did not perform any subjective studies but made an empirical observation.

An attempt to calculate a video's sharpness has been made by calculating the MTF from a natural scene by using the natural scene-derived spatial frequency (NS-SFR). It works by searching for areas within the frame that could represent edges in a natural scene via edge detection. Then the MTF is calculated from the found edges. This approach has demonstrated promising results in a laboratory setting when measuring system e-SFR from a test chart's slanted edge. The accuracy of the NS-SFR method is comparable to the ISO-12233 standard calculation method. [43]

Overall, video quality assessment frameworks have been proposed by many works, including [44] and [45]. Most of the works focus more on distortions caused by transmission errors. One recent work considers video stabilization when evaluating high-motion sports videos [46].

4. MEASUREMENT PROTOCOL

This section proposes a VSQA framework used in this work. This section covers an initially conducted prestudy, the lab environment, and the metrics used to measure the video stabilization quality.

4.1 Video stabilization quality assessment framework

Figure 4.1 shows a flow diagram of the main steps in creating our VSQA framework. The first step was identifying the most common use cases for video stabilization. We identified video capturing while walking, running, and cycling as potential consumer use cases. Based on these use cases, we captured test videos outdoors of each use case with several devices. The captured videos were then subjectively analyzed to find the main weak points like distortions present in the videos. From the observed weak points, a laboratory environment is devised in which a repeatable testing procedure could be used for testing devices. Once the laboratory was set up, videos could be captured. These laboratory videos could then be subjectively evaluated by humans and objectively with software to identify the severity of any present distortions. Finally, the objective metrics supported by subjective findings can be used to rank the devices, which is the ultimate goal of this work.

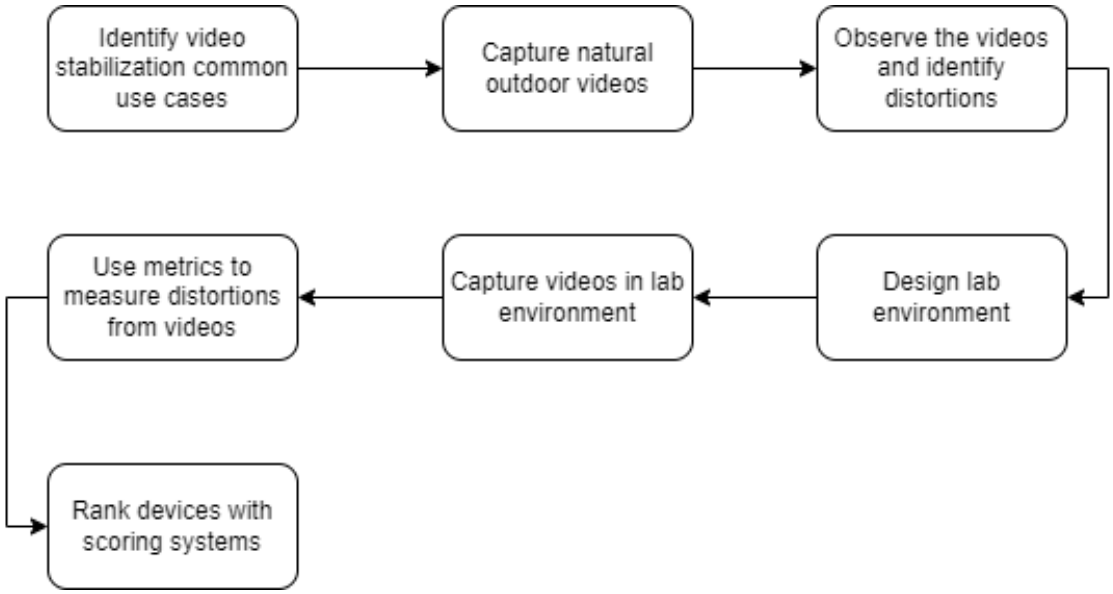


Figure 4.1. *Steps conducted to create the VSQA setup*

4.2 Subjective prestudy of outdoor videos

Initially, videos were captured outdoors in the natural environment in broad daylight while walking, running, and bicycling. The videos were captured on multiple mobile devices using 1x and 5x zoom ratios. These zoom ratios were chosen since video stabilization is relevant in a narrow FOV at 5x and the frequently used 1x zoom ratio. According to a user study presented in DxoMark's Webinar, 30% mobile phone users use their phone's zooming capability [47]. This makes evaluating the larger zoom ratio essential since many users use different zoom ratios.

In this section, the videos are analyzed to find possible distortions or artifacts. This section includes figures highlighting distortions. The temporal aspects of video cannot be assessed through a few still frames, but some examples of distortions are highlighted. The performance of the devices' videos is subjectively compared in this section.

The subjective evaluation was performed by the author on an Eizo CG319X 31.1 inch monitor which has a horizontal resolution of 4096 pixels and a vertical resolution of 2160 pixels. The monitor was chosen for accurate color accuracy and sharpness. The videos were viewed from 60cm away from the monitor

Table 4.1 shows each device's camera specifications used in this prestudy. The specifications are mentioned in the following order, resolution, aperture, focal length, and image stabilization system. The telephoto camera modules also note their optical zoom magnification factor. Device 1 has a wide camera, telephoto camera, and periscope telephoto camera. Device 2 has a wide camera and a periscope telephoto camera. Device 3 has a wide camera as well as the telephoto camera. All the cameras include OIS, except the Device 3 wide camera, which uses sensor-shift OIS. The focal length of the wide cam-

eras is nearly the same. However, the telephoto and the periscope telephoto camera focal lengths vary.

Device name	**Wide camera specifications**	**Telephoto camera specifications**	**Periscope telephoto camera specifications**
Device 1	108 MP, f/1.8, 24mm, OIS	10 MP, f/2.4, 72mm, OIS, 3x optical zoom	10 MP, f/4.9, 240mm, OIS, 10x optical zoom
Device 2	50 MP, f/1.8, 23mm, OIS	-	64 MP, f/3.5, 90mm, OIS, 3.5x optical zoom, 7x lossless zoom
Device 3	12 MP, f/1.6, 26mm, sensor-shift OIS	12 MP, f/2.2, 65mm, OIS, 2.5x optical zoom	-

Table 4.1. *Prestudy device camera specifications. The specifications are mentioned in the following order, resolution, aperture, focal length, and image stabilization method*

4.2.1 Walking videos

The walking and running videos were recorded in landscape mode with the default camera settings. The main capture target in the 5x videos was the barn shown in Figure 4.3, which contains example video frames from different parts of one of the 5x videos. The 1x videos capture the larger scene shown in Figure 4.2, with example video frames from different parts of one of the 5x videos. These videos do not contain any moving objects.

Figure 4.2. *Example 1x video frames showing walking and running outdoor environment. The top frame is the first frame and the bottom frame is the last frame.*

The 1x walking was the most stable use case. Device 1 was the most stable because it used a higher frame rate. In the device 3 1x video, the trees seemed to wobble. The device 2 1x video had illumination changes which appeared as flickering in the greenery. Flickering may be caused by the 3A algorithms [10].

The 5x videos showed results from good to poor quality. Device 3 5x video was the most

Figure 4.3. *Example 5x video frames showing walking and running outdoor environment. The top frame is the first frame and the bottom frame is the last frame.*

stable of all the devices but had some illumination changes. The shakiness of the device 2 5x video made it unwatchable. Device 1 5x video was shaky, but the shakiness was smoother and not as large Device 2 5x video shakiness made it more watchable.

4.2.2 Running videos

In 1x videos, flickering occurs in different parts of the frame when the handheld running motion occurs in all devices except Device 1. Figure 4.4 illustrates the flickering that occurred with a sequence of 18 consecutive cropped frames. Each of the cropped frame's average red, green, and blue (RGB) values are mentioned to highlight the flickering. The cropped frames were cropped from the red box region shown on top of figure 4.4. Flickering only occurs in the sky in device 2 1x video, while the whole scene flickers in device 3 1x video. Device 1 by default, uses a higher frame rate of 60 frames per second (fps), while the other devices use 30 fps, so flickering does not occur in Device 1. The flickering probably occurs due to insufficient stabilization leading to problems with auto exposure. The motion in the device 2 and 3 1x videos is more smooth compared to the jerky motion in the device 1 1x video.

The 5x videos are extremely shaky due to large handheld movements amplified by the narrow field of view. This phenomenon made the device 2 and 3 5x videos entirely unwatchable due to the large magnitude shakiness. Device 1 was slightly more watchable, since the shakiness magnitude was smaller.

Figure 4.4. *Illumination flickering in the device 3 5x video.*

4.2.3 Cycling videos

The cycling videos were recorded in portrait mode. The mobile device was mounted onto the bicycle's handlebar with a mounting attachment shown in Figure 4.5. A portion of the videos was recorded on a rocky road and the other portion on the pavement. The former uneven road surface produces higher frequency impacts on the bike than the latter flatter surface. Figure 4.6 highlights the environment via video frames. These videos contain some moving objects on the pavement. Their contribution is assumed marginal.

All of the devices produced less severe artifacts and distortions on the pavement. When on the rocky road, the artifact severity amplified. In contrast, the artifacts were less severe on the pavement. The ground surface difference is especially evident in the 5x, where

Figure 4.5. *Phone mounted to the bicycle*

Figure 4.6. *Example video frames showing bicycling outdoor environment*

videos are extremely blurry on the rocky road. Figure 4.7 illustrates an extremely blurry frame with other wave-like geometric deformations. The field-of-view (FOV) is very narrow in 5x videos. Hence all distortions are amplified compared to 1x videos.

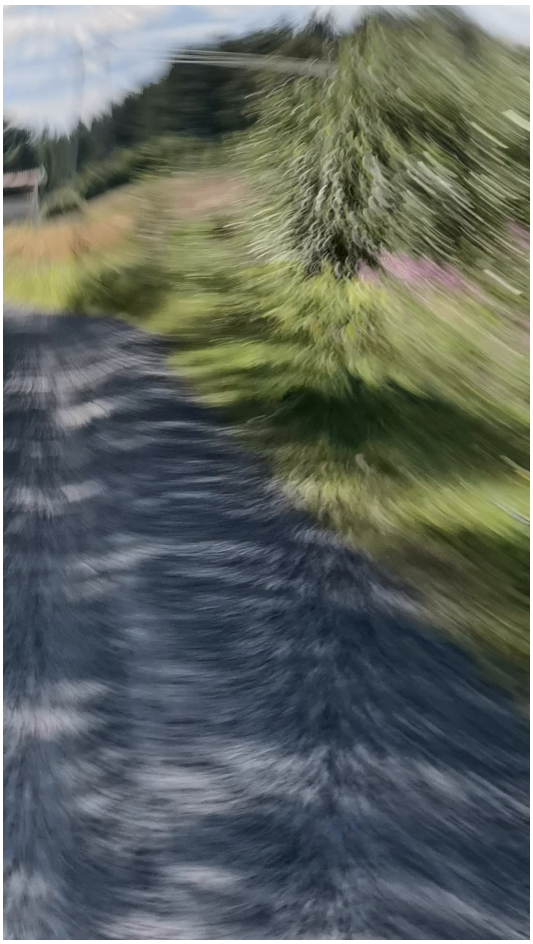


Figure 4.7. *Blurry frame from device 2 5x cycling video*

The device 3 1x video was overall the best performer, but it still has room for improvement. High-frequency jitter is present in the device 1x video, especially on the rocky road. The device 3 5x video suffers from severe rolling shutter geometric deformations throughout the video. This deformation combined with rough rocky road first portion of the video unwwatchable. On the other hand, the pavement portion of the video was more watchable since the ground surface was not as rough.

The device 2 1x and 5x videos suffer from severe rolling shutter geometric deformation. In the 5x video, the deformation is amplified. The deformation of the 5x video is shown in Figure 4.7.

The device 1 1x and 5x videos were surprisingly on par with each other compared to the other devices 1x and 5x differences. The narrow FOV on the 5x video did not have as of an affect compared to the other devices. The device 1 1x video seems to be highly out of focus on the rocky road, as shown in the consecutive frame sequence in Figure 4.8. The

device 1 5x video was more visually pleasing compared to the device 1 1x video on the pavement.

Figure 4.8. *Blurry frame sequence from device 3 1x cycling video*

4.2.4 Prestudy conclusions

The main observations for each use case in this prestudy are included in Table 4.2. Overall, the cycling videos had the worst geometric deformations due to the rolling shutter and insufficient stabilization in both 1x and 5x videos. The 1x videos running were quite watchable. The only noticeable artifact is illumination changes in the scene's greenery or sky and some wobbliness. On the other hand, the 5x running videos were dominated by shaky handheld motion, which appears as significant movement in the narrow FOV. The cycling videos were naturally the worst since there was high-frequency shakiness from the bike and the rough riding surfaces.

Use case	Observations
Walking	Wobbling trees in the background. Illumination flickering was present in the greenery. Unpleasant large rapid movements in the 5x videos.
Running	Illumination changes in the sky, flickering. Significantly large handheld motion made 5x videos unwatchable
Cycling	Geometric deformations due to the rolling shutter made the rocky road portion of the videos quite unpleasant to watch. High-frequency jitter can be seen in the pavement portion of the videos. Most of the videos were unpleasing to watch.

Table 4.2. *Observed distortions in each use case*

4.3 Lab environment

To measure video stabilization quality repeatably and reliably, a laboratory environment was setup to capture videos. The lab environment is a dark room with gray and black walls, carpet, and ceiling to minimize all reflected light [48]. The lab equipment consists of two Image Engineering iQ-Flatlight illumination devices, an Image Engineering STEVE-6D motion platform, and an Imatest SFRPlus test chart. Figure 4.10 shows the laboratory environment with the iQ-flatlights, the STEVE-6D, and the SFRPlus which are marked with the numbers 1, 2, and 3, respectively.

The STEVE-6D motion platform in Figure 4.11 is used to evaluate the performance of image stabilizer systems. STEVE-6D was chosen since it allows movement in 6 degrees of freedom and simulates human handshaking. [50] Image engineering provides a handshaking profile compatible with the STEVE-6D [51]. The profile only includes yaw, pitch, and roll rotations shown in Figure 4.12. The profile is based on the ISO 20952-2 standard [52].

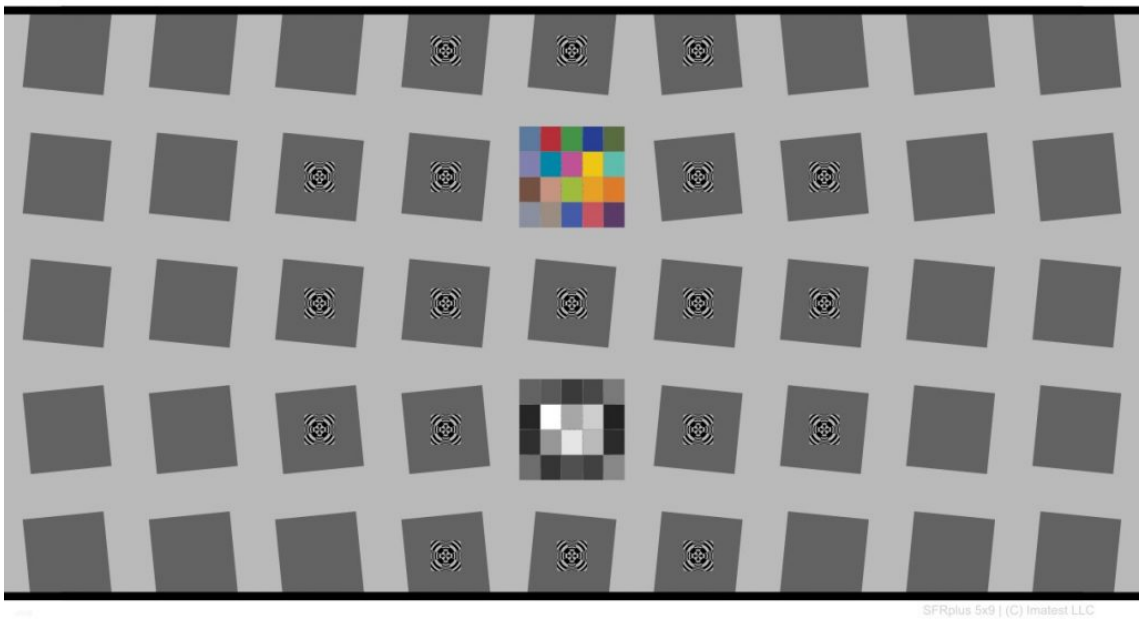


Figure 4.9. *Imatest's SFRPlus chart [49]*

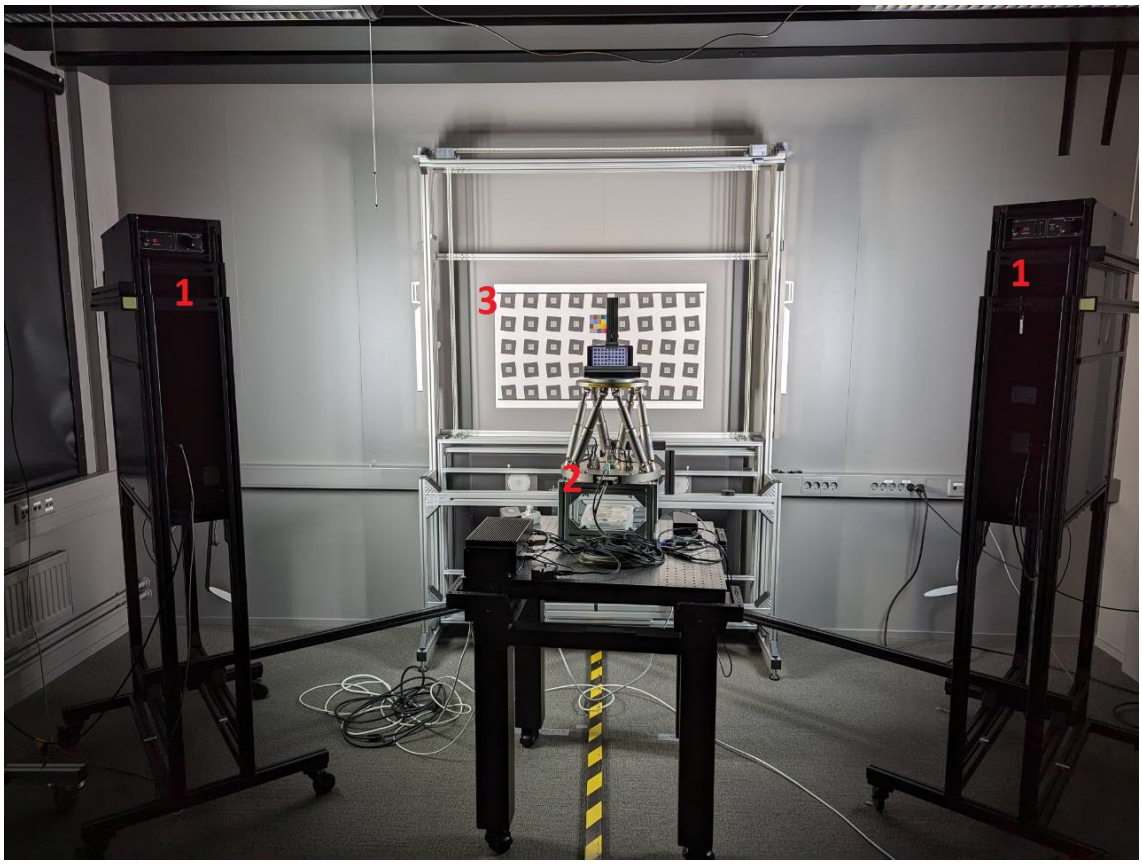


Figure 4.10. *Laboratory environment.*

The iQ-Flatlight illuminants were chosen since mobile phone manufacturers commonly use them for their camera testing. The lighting conditions were set to D65 illuminant (Artificial Daylight, color temperature at 6500 K) at an intensity of 2000 lux. Both lights were

Figure 4.11. *Image engineering's STEVE-6D motion platform [50]*

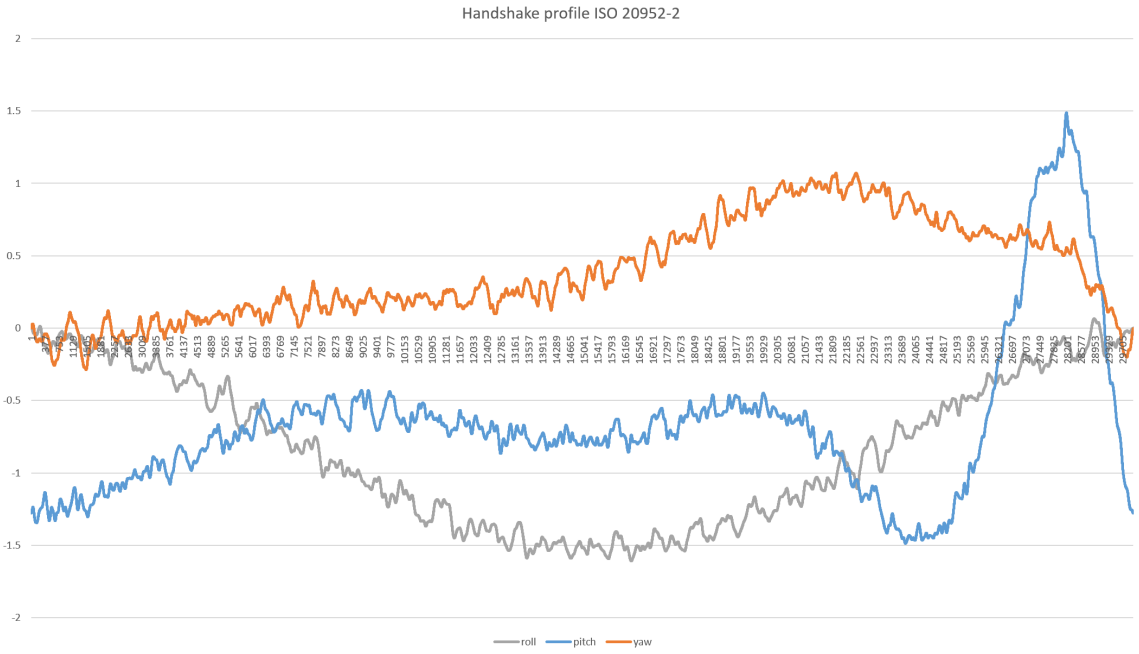


Figure 4.12. *The yaw, pitch, and roll rotations in the handshake profile in ISO 20952-2 standard [52]*

positioned at a 30 - 45 degree angle relative to the chart's normal to avoid glare as advised in Imatest's lab setup guide [48]. The SFRPlus chart was chosen since sharpness can be measured from its slanted edge features using Imatest's software. Other image quality metrics such as color accuracy, noise, lateral chromatic aberrations, and delta E can be measured from it. However, this work will not analyze those since they are not the primary focus [49]. The lighting uniformity of the test chart was ensured and measured according to Imatest's lighting uniformity requirements [48].

The SFRPlus test chart should stay in the camera frame when the device shakes on the STEVE6D motion platform. It was easy to keep the chart in the frame on the 1x zoom ratio as the movement was small. On the other hand, framing the camera at a 5x zoom ratio was slightly more challenging since even slight movements greatly moved the narrow field of view. The solution to overcome this was to keep the chart in a smaller region of the frame. This solution is not ideal since geometric distortions may occur at the outer parts of the frame. Another issue was that Imatest's software is not able to process a frame if the region of the chart occupies a region smaller than 25% of the frame's size [49]. This limitation was resolved by cropping the chart from each frame with template matching and inputting the cropped frame into Imatest's software.

4.4 Metrics

In this work, two metrics are used to characterize the quality of video stabilization: movement characteristics and sharpness. The following sections cover how these metrics were measured from the laboratory videos containing the SFRPlus chart.

4.4.1 Motion characteristics from optical flow

This work uses optical flow to characterize the movement between consecutive video frames. Optical flow computes the displacement between each pixel in two consecutive frames. All the pixels' displacements produce a displacement field, also known as a flow field, for each consecutive frame pair, as shown in Figure 4.13.

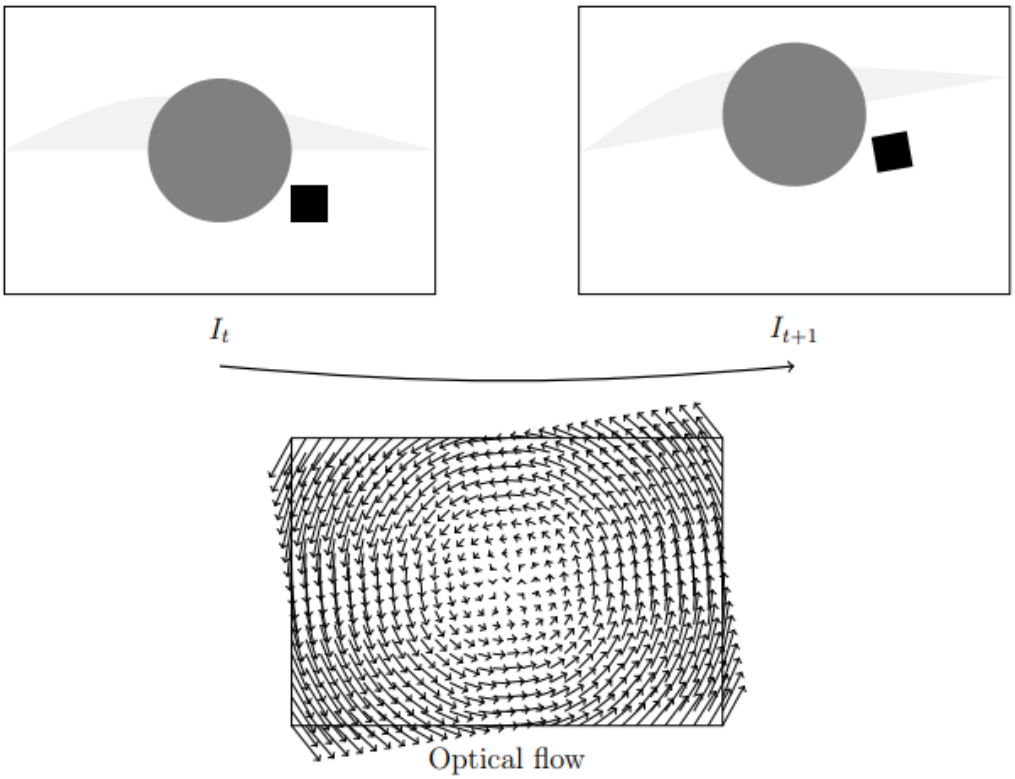


Figure 4.13. *Optical flow illustration [7]*

Many approaches have been proposed to solve optical flow, which is an actively investigated topic. Many traditional optical flow approaches define handcrafted optimization objectives. The problem with handcrafted optimization objectives is that it is hard to design a robust solution that handles many edge cases. This has slowed the advancement of handcrafted optimization-based solutions. Recently many deep learning-based approaches have been proposed to tackle solving optical flow. These approaches can directly train a network to predict the optical flow and avoid the need to formulate an optimization problem. [53]

In this work, the Recurrent All-Pairs Field Transforms (RAFT) [53] approach is used to compute optical flow. It was chosen since it had state-of-the-art performance in optical flow benchmarks, KITTI and Sintel [53]. Computational speed was not a major consideration since accuracy was the most crucial criterion. Some preprocessing image filters were applied to images before computing optical flow, as shown in Figure 4.14. First, the input images are resized to a third of their original to make computations faster in the following steps. The resizing operation did not affect the accuracy of the results. Next, a ridge filter is applied to the resized image to extract meaningful features like edges. Finally, the chart's coordinates within the frame are extracted via template matching since optical flow can struggle with detecting motion in monotonous gray areas, which are present in the background of the image [53].

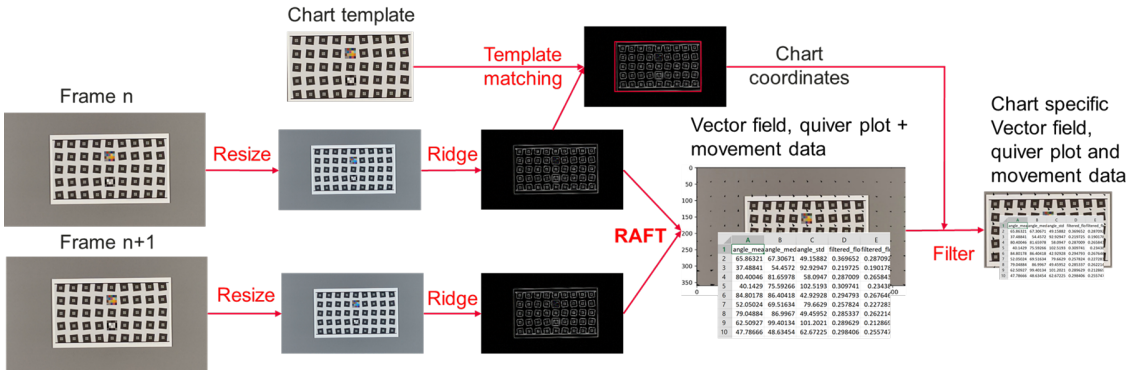


Figure 4.14. *Diagram of RAFT with preprocessing steps*

In this work, the movement of the laboratory videos was characterized by the optical flow movement data of the chart since it provides the most accurate movement data. Some motion characteristics are derived from the movement data. One simple characteristic is the average euclidean distance between two consecutive frames. The euclidean distance is described by the following equation $\sqrt{u^2 + v^2}$, where *u* is the average x-axis movement of all the pixels between consecutive frames and *v* is the average y-axis movement of all the pixels between consecutive frames.

Confidence ellipses and frequency responses can be calculated from the euclidean distances as illustrated in Figures 4.15, 4.16 and 4.17. The 95% confidence ellipse plot represents the global movement amplitude and direction of 95% of the euclidean distances. Only 95% of the data is used to remove outliers. In Figure 4.15, some data points are outside the single 95% confidence ellipse. A more thorough explanation of the confidence ellipses can be found here [54]. In Figure 4.16 Device 1 has the smallest movement amplitude while Device 2 has the largest movement amplitude. Device 1 and 2 have a similarly shaped confidence ellipse, indicating similar directional movement. On the other hand, Device 3 has a vastly different movement direction.

Figure 4.17 is a frequency response plot of the euclidean distance movement. It highlights

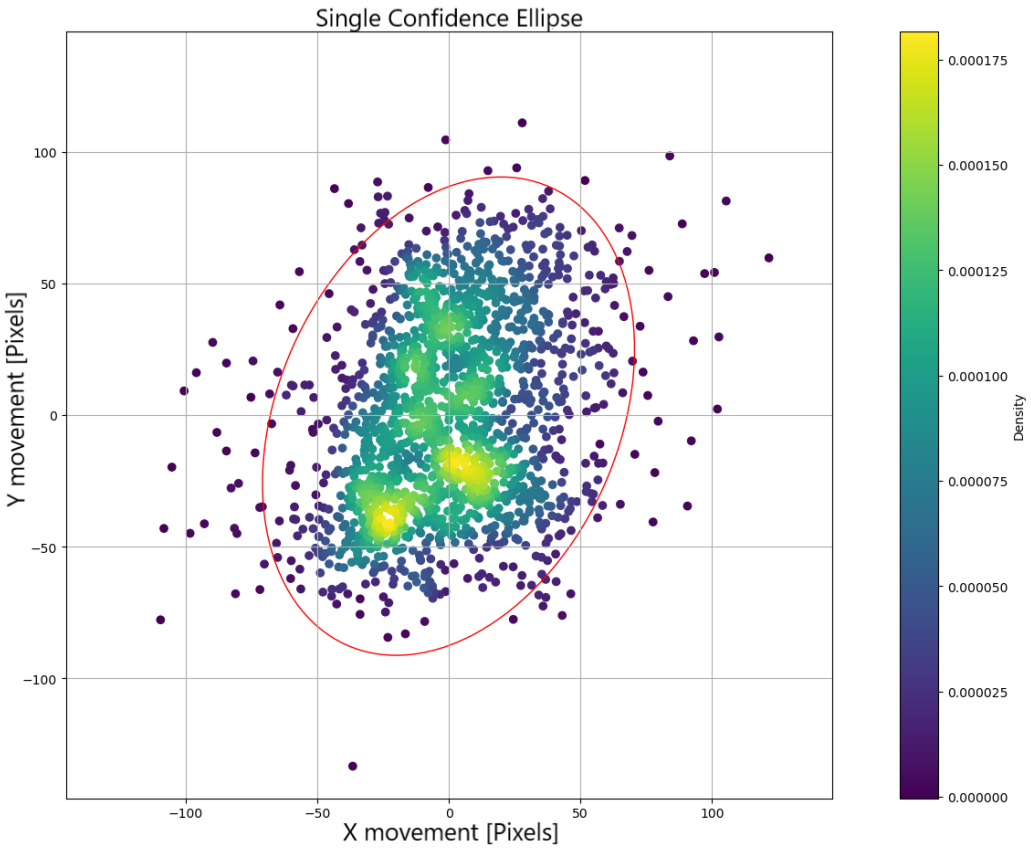


Figure 4.15. *Example of a single confidence ellipse which encompasses 95% of the plotted data points. The density color coding is used to highlight where the majority of data points are.*

the most prominent frequencies of movement with blue dots. Device 1 has frequencies up to 30 Hz since its video was captured at 60 fps. The prominent frequencies indicate whether the video has low or high-frequency shakiness movement. In this example, Device 1 and Device 2 have significantly higher frequency prominent responses than Device 3.

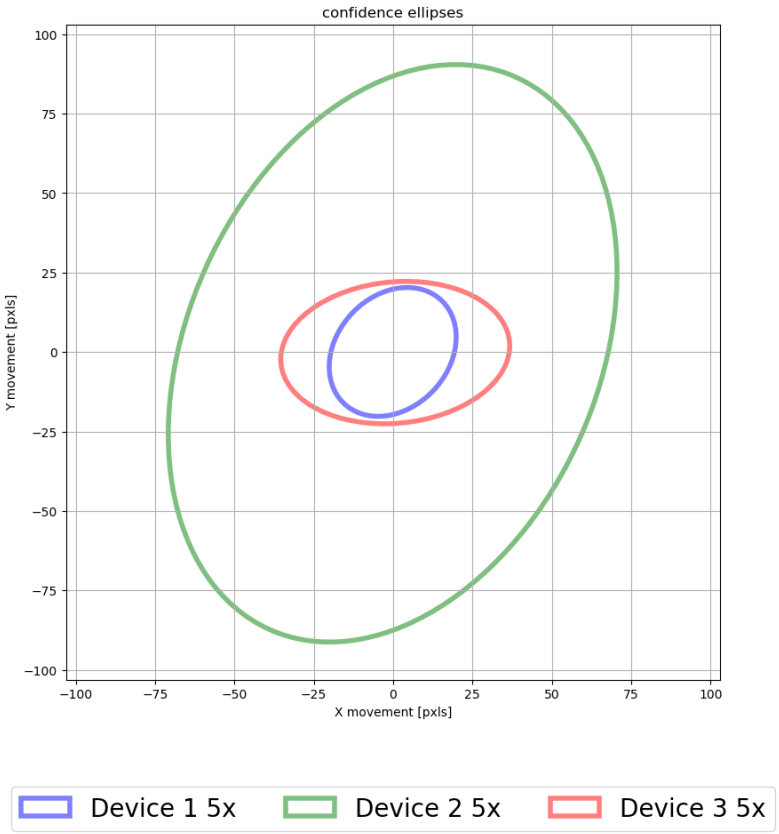


Figure 4.16. *Example of multiple confidence ellipses*

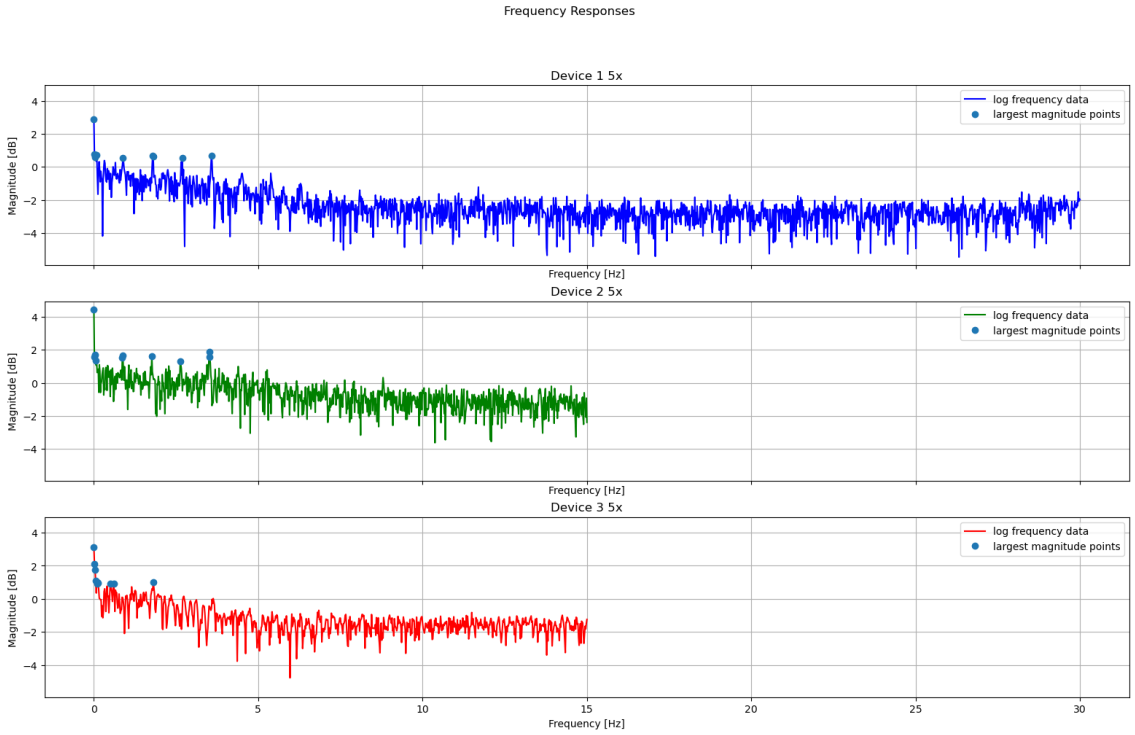


Figure 4.17. *Example of frequency analysis*

4.4.2 Sharpness metrics

Sharpness is commonly measured with modulation transfer function (MTF), closely related to the spatial frequency response (SFR). The SFR is usually the combination of all the camera component MTFs. However, SFR and MTF are quite often used interchangeably. [55] The ISO-12233 standard specifies three different methods for measuring the SFR: CIPA resolution, low contrast edge SFR (e-SFR), and sine-based SFR (s-SFR). These can be measured from the following chart elements shown in Figure 4.3: wedges, slanted edges, and siemens stars, respectively. [56]

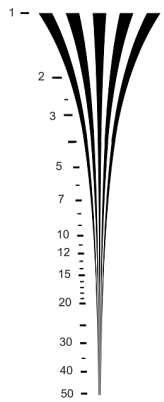
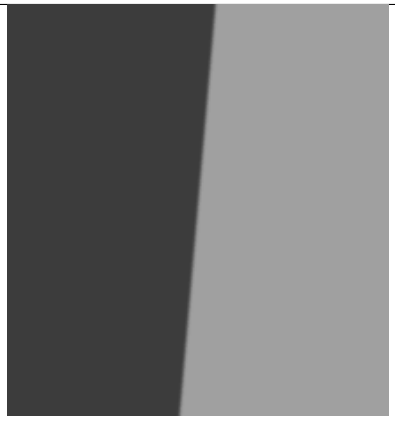
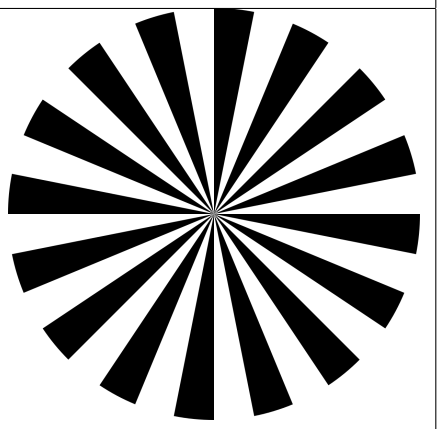
Wedge	Slanted edge	Siemens star
		

Table 4.3. *This table contains chart elements which can be used to measure sharpness*

Sharpness was measured from the laboratory videos with the e-SFR measurement method. Figure 4.18 illustrates how the sharpness was computed from each image. First, Imatest's software automatically detects the slanted edges from the input image. Then, Imatest's software computes an MTF curve for each slanted edge within the SFRPlus test chart. From each MTF curve, Imatest's software extracts a value that is 50% of the MTF curve's peak value (MTF50P). The MTF50P correlates well with perceived sharpness in still images [55]. However, the MTF50P's subjective correlation with temporal aspects has not been studied and therefore the MTF50p results for video frames can only be objectively evaluated currently. The MTF50P was chosen over MTF50 since it is not as susceptible to over-sharpening done by the ISP [55].

The SFRPlus test chart has 154 vertical and horizontal slanted edges. The MTF50P values of these slanted edges were averaged for each frame. These average values were then used to compute the total average MTF50P value for the video. The total average MTF50P metric quantifies the overall sharpness of the video and the higher, the better. A standard deviation of the frame average MTF50P values is used to measure the fluctuation in sharpness. The spatial frequency MTF unit used in the average MTF50P plots is Line Widths/Picture Height (LW/PH) which is the best unit for comparing cameras with different sensor sizes and pixel counts [55].

When 5x video frames were processed, the chart needed to be cropped from the frame since Imatest's software cannot process the regular frames if the chart is too small. The chart was cropped with template matching. Then the cropped 5x video frames could be processed by Imatest's software.

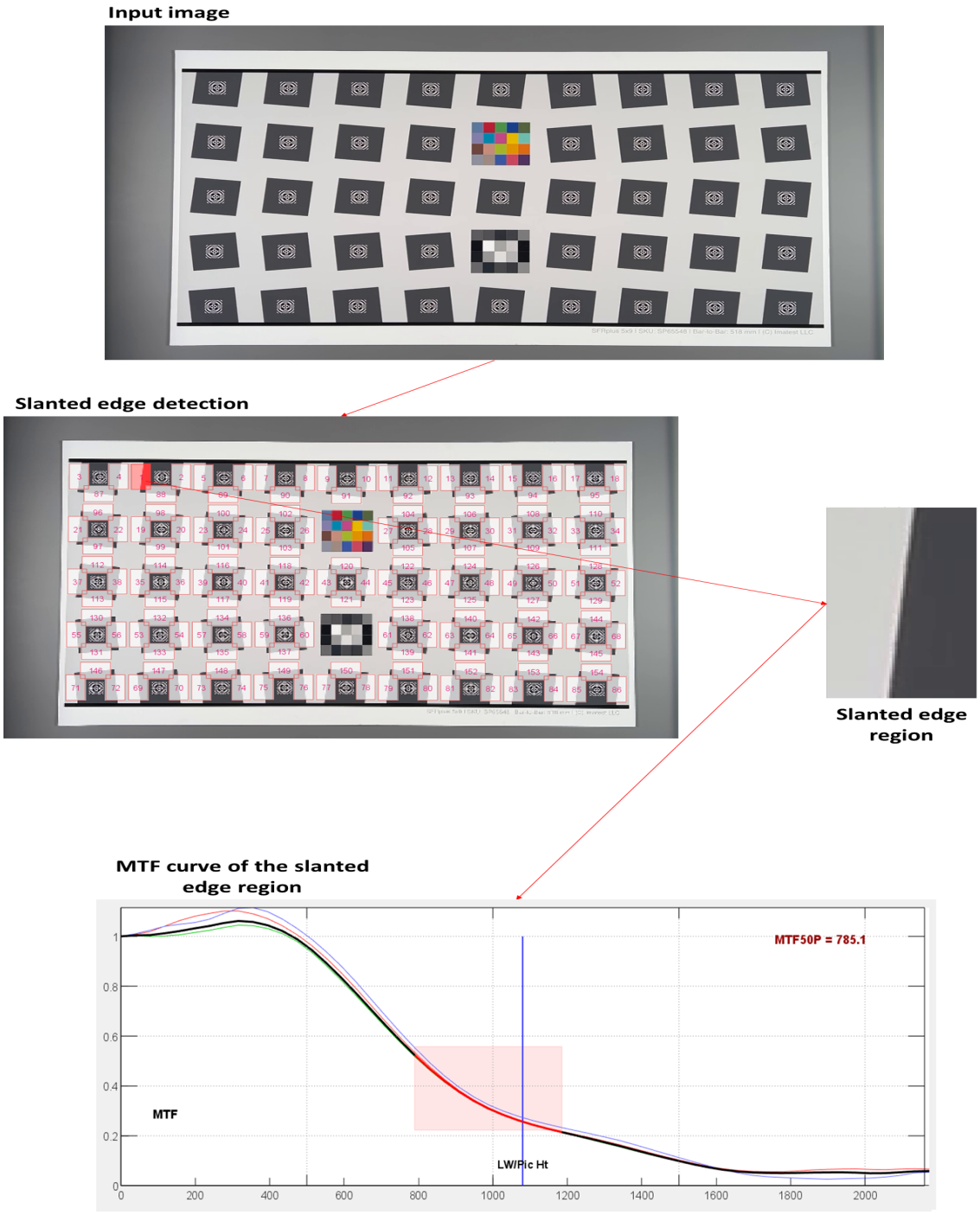


Figure 4.18. *Sharpness calculation method. First, slanted edges are automatically detected from the input image. Then a MTF Curve is computed for each slanted edge. Finally, the sharpness value is the value at 50% of the MTF curve's peak value also known as the MTF50P*

5. RESULTS

This section includes subjective and objective quality assessment of 8 videos captured on four devices with zoom ratios of 1x and 5x in the controlled laboratory environment using simulated handshaking. The videos were recorded at a vertical resolution of 1080 pixels and a horizontal resolution of 1920 pixels at 30 frames per second. The other camera settings were not changed from the default settings. The devices' performances are compared subjectively and objectively with motion characteristics and sharpness metrics. Finally, the devices were ranked according to the subjective and objective quality assessment findings.

Table 5.1 has the specifications for the mobile device cameras which were objectively and subjectively evaluated. The devices used are not the same devices that were used in the previous section. All of the camera modules have OIS, while device B's wide camera module has sensor-shift OIS. Devices A and B do not have a periscope telephoto camera and device C does not have a telephoto camera. Device D's periscope telephoto camera has a much larger focal length than Device C's periscope telephoto camera. All the focal lengths of the wide camera modules are quite the same, but the resolution varies between the devices.

Device name	Wide camera specifications	Telephoto camera specifications	Periscope telephoto camera specifications
Device A	50 MP, f/1.9, 25mm, OIS	48 MP, f/3.5, 104mm, OIS, 4x optical zoom	-
Device B	12 MP, f/1.5, 26mm, sensor-shift OIS	12 MP, f/2.8, 77mm, OIS, 3x optical zoom	-
Device C	50 MP, f/1.8, 23mm, OIS	-	64 MP, f/3.5, 90mm, OIS, 3.5x optical zoom, 7x lossless zoom
Device D	108 MP, f/1.8, 23mm, OIS	10 MP, f/2.4, 70mm, OIS, 3x optical zoom	10 MP, f/4.9, 230mm, OIS, 10x optical zoom

Table 5.1. *Device camera specifications used in the final assessment. The specifications are in the following order resolution, aperture, focal length, and image stabilization system*

5.1 Subjective comparison

The goal of this section is to gain an understanding of what are the most subjectively noticeable and annoying artifacts in indoor laboratory videos. The primary focus was evaluating the temporal sharpness and motion of the videos since those were the most annoying observations found in the subjective prestudy in Section 4.2. Also, the same monitor setup used for subjective evaluation in Section 4.2 was used here.

Figure 5.1 shows the first frames of the 1x and 5x videos. They represent the overall sharpness present in each video. In the 1x videos, the first frame's sharpness looks similar in all the devices. However, in the 5x videos, the first frame's overall sharpness is worse than in the 1x videos. Larger zoom ratios often require digital zoom, which degrades image quality. Devices with optical zoom that supports large zoom ratios often perform better. The overall sharpness also differs more between 5x videos. The Device A 5x video's first frame appears the blurriest, while the Device D 5x video's first frame looks the sharpest. The first frame's 5x sharpness in devices B and C is worse than the sharpness in device A, but both are better than the sharpness of device D. The overall sharpness of the device C 5x video looks worse than the device B 5x video's sharpness.

The motion appeared similar in all 1x videos since motion was small. However, in the 5x videos, the narrow FOV makes motion larger, making some of the 5x videos shaky and amplifying rolling shutter artifacts like motion blur. Device B 5x video showed a small amount of motion blur shown in Figure 5.3, but it was not as severe as the motion blur in device C 5x video as shown in Figure 5.2. The SFRPlus chart's size is larger in the device C 5x video's first frame than in the Device B 5x video's first frame. Hence, the motion blur in device C appears larger than in Device B.

Overall, in the 5x videos, the poor sharpness and the shaky motion were the most annoying artifacts. The 1x videos did not contain artifacts or other annoyances since the videos were all sharp and did not contain much motion. The flickering illumination observed in some of the outdoor videos, as discussed in section 4.2, did not occur in any of the laboratory videos.

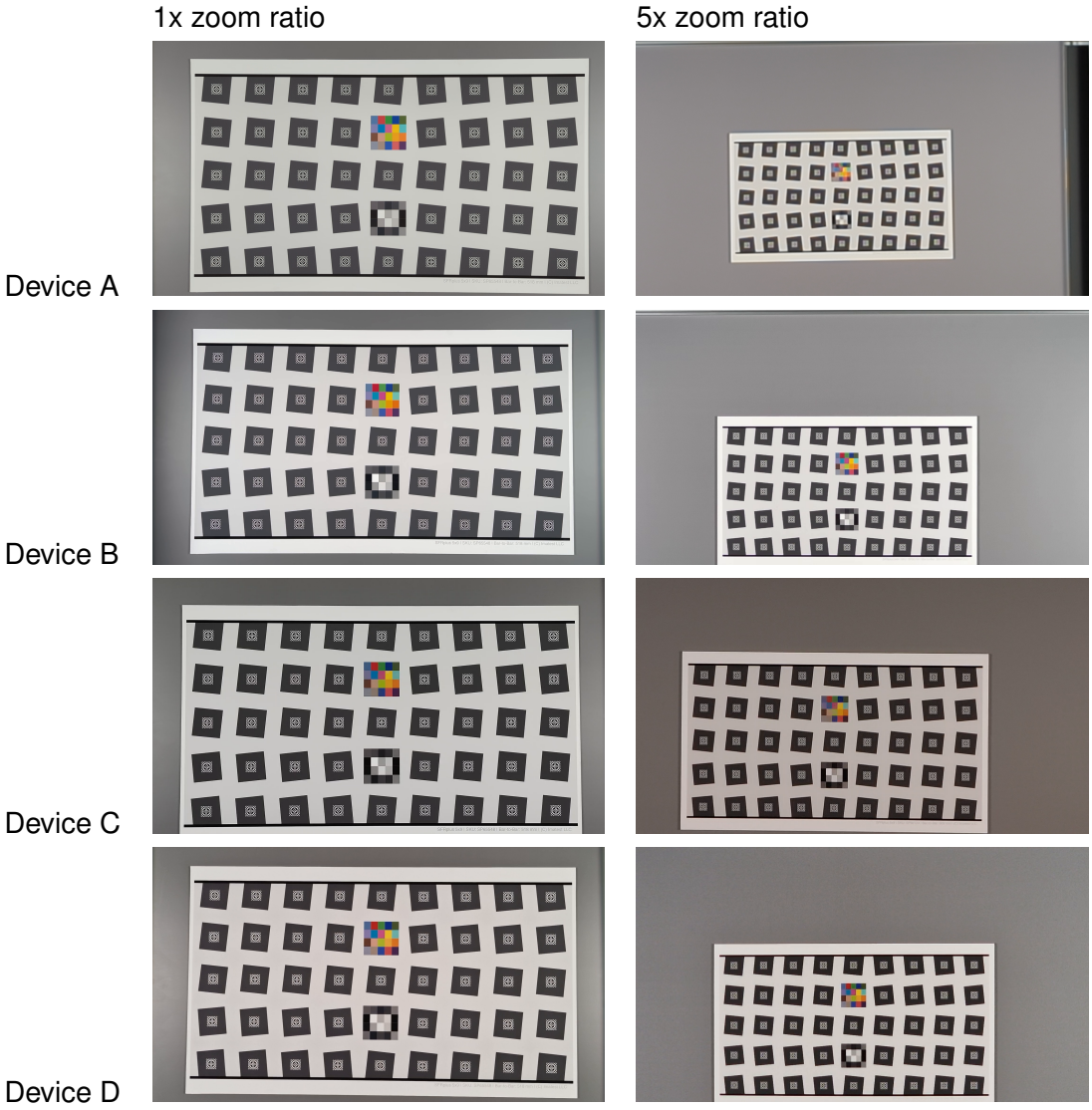


Figure 5.1. *Each video's first frames used to compare sharpness subjectively. On the are the 1x video frames and on the right are the 5x video frames*

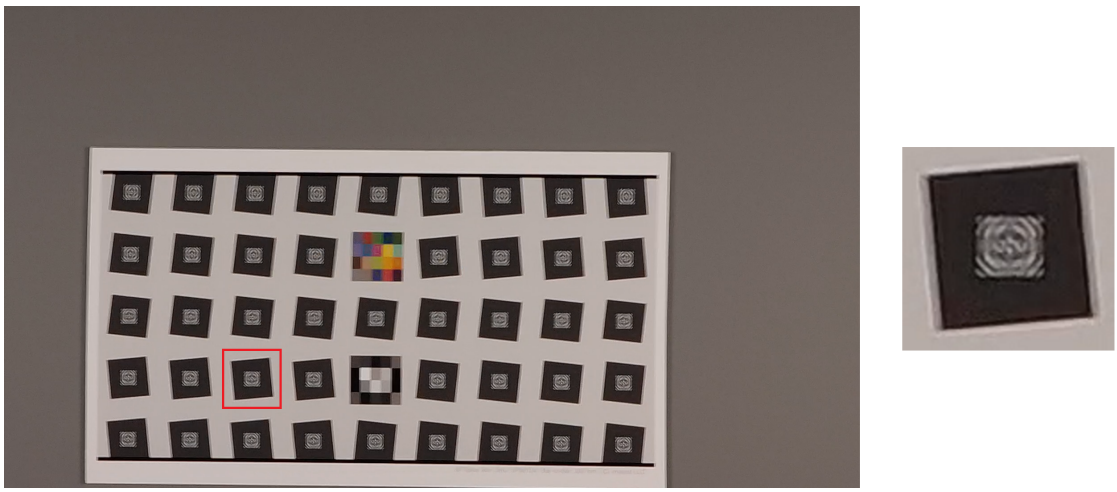


Figure 5.2. *Shaky frame which has motion blur from the device C 5x laboratory video. The cropped image on the right has been cropped from the red square region from the shaky frame on the left*

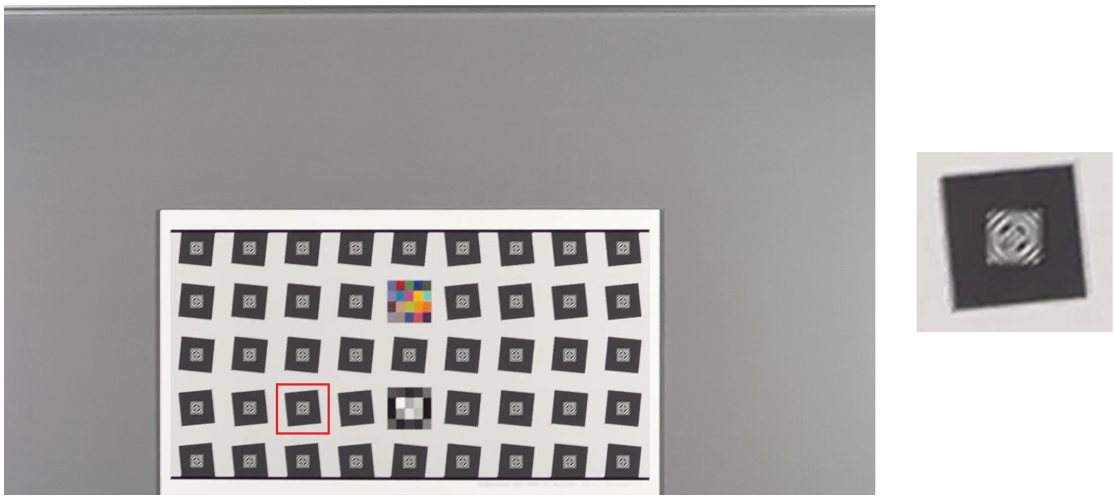


Figure 5.3. *Shaky frame which has motion blur from the device B 5x laboratory video. The cropped image on the right has been cropped from the red square region from the shaky frame on the left*

5.2 Objective comparison

This section compares each device's objective sharpness performance. The subjective sharpness evaluation is also validated with these metrics.

For the sake of conciseness, the motion metrics are not analyzed in this section because the motion metric findings correlate well with sharpness metrics and would rank devices in the same order. The motion metric findings are included in Appendices A and B.

Figures 5.4 and 5.5 contain the average MTF50P values for each video frame captured in the laboratory environment. Statistics presented in Table 5.2 are derived from the average MTF50P in Figures 5.4 and 5.5. The mean and standard deviation represent the overall sharpness and the fluctuation in the sharpness of the video.

All the 1x video mean MTF50P values are quite high. Subjectively the sharpness did not temporally change in the 1x videos. However, in Table 5.2, the device C 1x video standard deviation is high compared to the other devices. Subjectively this fluctuation in sharpness was not visible probably because the mean MTF value is quite high. This large standard deviation is caused by a slow change in the trend of the sharpness values throughout the whole video, as is evident from Figure 5.4.

The mean MTF50P values for the 5x video are much lower than the mean MTF50P values of the 1x videos. The subjective findings of Device A's lowest sharpness and Device D's highest sharpness in the 5x videos correlate well with the mean MTF50P values for the 5x video. The device B and C 5x videos had similar mean MTF50P values but their standard deviation of MTF50P is high compared to device A and D 5x videos. The high standard deviation values are most likely due to the motion blur observed in these videos.

Overall, device A was objectively the worst performer due to its low mean sharpness values in both zoom ratios. Device D had the best mean sharpness value in the 5x zoom ratio and the second best sharpness in the 1x zoom ratio. Device D also had the second lowest sharpness standard deviation. These facts make device D the best performer in terms of sharpness.

It was difficult to decide on the second and third placements. On the one hand, objectively device C has the top mean sharpness in the 1x zoom and second in the 5x zoom, whereas device B ranked third in mean sharpness in both zoom ratios. On the other hand, objectively device C has the highest standard deviation in sharpness in the 5x zoom while device B had a lower standard deviation in the 5x sharpness. It was difficult to determine subjectively which was more problematic - the general quality or the intensity of motion blur. Overall, device C places third since its 5x video's motion blur is subjectively more annoying to watch. Objectively, the motion blur's severity is reflected in the higher sharpness standard deviation and the findings in Appendices A and B.

MTF50p mean and standard deviation value summary table				
Zoom ratio	Device A	Device B	Device C	Device D
1x	724.9 ± 2.6	860.4 ± 3.7	1046.4 ± 31.0	943.1 ± 7.5
5x	119.0 ± 2.5	334.5 ± 40.6	382.8 ± 50.1	474.5 ± 8.4

Table 5.2. *MTF50p mean and standard deviation value summary table. The best scores are highlighted in green and the worst in red.*

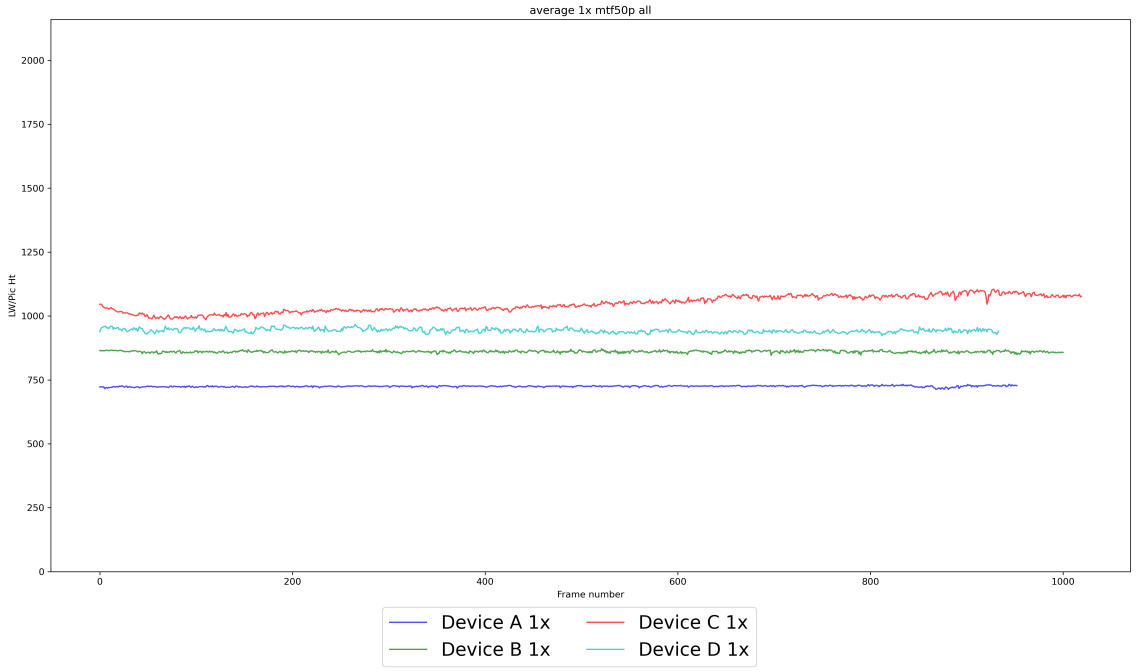


Figure 5.4. *Mean MTF50p values of each 1x laboratory video frames*

Figure 5.5. *Mean MTF50p values of each 5x laboratory video frames*

5.3 Conclusions

This work establishes a VSQA framework for testing and ranking mobile phone videos. A repeatable laboratory environment was created and objective metrics were implemented for measuring sharpness and motion characteristics.

The handshaking motion was relatively small, so the objective movement metrics differences were quite small. Consequently, in this scenario, the movement metrics discussed in Appendices A and B mainly offer valuable information to support the conclusions drawn from the sharpness metrics. For videos with greater movements, such as those captured outdoors in Section 4.2, the motion metrics would be more beneficial.

Overall, the best device, in this case, is the device with the best sharpness performance. Since the sharpness of all the 1x zoom videos appeared subjectively equally sharp, the best device is decided by each device's 5x zoom overall sharpness performance. Device D is considered the best since its 5x zoom has the best mean sharpness performance and the second-best standard deviation in sharpness. Device A had the worst performance in both 1x and 5x zoom ratios. Devices B and C performed similarly, but device B placed second since it had less motion blur, and device C placed third.

6. FUTURE WORK

This VSQA framework can still be improved in many ways. This work considered one use case where slow handshaking motion was used in bright lighting conditions. More use cases must be added to make the test suite more comprehensive. For example, testing handshaking and walking motion in more illumination conditions, such as in low light, could be the next logical use case to test. Simulating walking motion would require the creation of a more comprehensive motion simulation profile which could be used on the STEVE-6D motion model.

The objective metrics proposed in this work need to be mapped to just noticeable differences (JND) through subjective user studies to better understand how well the metrics correlate with subjective perception. Currently, the sharpness metric can be mapped to a still image JND. However, this does not consider the temporal sharpness changes present in the video and humans perceive still image sharpness differently than video sharpness.

Other objective metrics to measure overall temporal video quality artifacts like wobbling should also be implemented. The effects of using different video capturing frame rates should also be further studied. Also, capturing the videos with the same focal length instead of the same zoom ratio magnifications would increase the comparability of the results.

The ultimate goal would be to measure the video stabilization quality with subjectively correlated metrics from real-life videos in a repeatable manner. A possibility for measuring sharpness via MTF50P from real videos was discussed in Section 3.2.2. The assessment of the real-life video movement's pleasantness could be more precise with the use of a motion model that incorporates 6 degrees of freedom and utilizes optical flow movement data.

REFERENCES

[1] Turner, A. *How many smartphones are in the world?* Mar. 2022. URL: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world.

[2] Counterpoint Research. *Whitepaper: Smartphone Imaging Trends: New Directions Capturing Magic Moments*. Tech. rep. [Online; accessed 18-July-2022]. June 2022, p. 15. URL: https://www.counterpointresearch.com/wp-content/uploads/2022/06/WP-Smartphone-Imaging-Trends-New-Direction-Capturing-Magic-Moments.pdf.

[3] Delbracio, M., Kelly, D., Brown, M. S. and Milanfar, P. Mobile Computational Photography: A Tour. (Feb. 2021). URL: https://arxiv.org/abs/2102.09000v2.

[4] Ito, M. S. and Izquierdo, E. A Dataset and Evaluation Framework for Deep Learning Based Video Stabilization Systems. *IEEE Visual Communications and Image Processing (VCIP) Conference*. 2019, pp. 1–4. DOI: 10.1109/VCIP47243.2019.8966057.

[5] Guilluy, W., Oudre, L. and Beghdadi, A. Video stabilization: Overview, challenges and perspectives. *Signal Processing: Image Communication* 90 (2021). ISSN: 09235965. DOI: 10.1016/j.image.2020.116015.

[6] Wang, Y., Huang, Q., Sun, S., Ye, F. and Wang, Y. An objective assessment method for video stabilization performance. *Eleventh International Conference on Digital Image Processing (ICDIP 2019)*. Vol. 11179. International Society for Optics and Photonics. 2019, p. 111792M.

[7] Guilluy, W. Video stabilization : A synopsis of current challenges, methods and performance evaluation. PhD Theses. Université Sorbonne Paris Cité, Dec. 2018. URL: https://tel.archives-ouvertes.fr/tel-03006243.

[8] GSMArena. *iPhone 12 Pro Max phone specifications*. https://www.gsmarena.com/apple_iphone_12_pro_max-10237.php. [Online; accessed 20-May-2022]. 2020.

[9] Wannatek. *Image of camera module*. https://www.wannatek.com/wp-content/uploads/2020/07/COB-CSP-2.jpg.webp. [Online; accessed 06-June-2022]. 2020.

[10] Peltoketo, V.-T. Benchmarking of mobile phone cameras. PhD thesis. 2016. ISBN: 978-952-476-685-2.

[11] Stemmar imaging. *Rolling shutter image*. https://www.stemmer-imaging.com/en-dk/knowledge-base/rolling-shutter/. [Online; accessed 06-June-2022].

[12] Tekalp, A. M. *Digital video processing*. eng. Second edition. New York: Prentice Hall, 2015. ISBN: 0-13-399111-3.

[13] Xu, X., Zhang, X., Fu, H., Chen, L., Zhang, H. and Fu, X. Robust passive autofocus system for mobile phone camera applications. *Computers Electrical Engineering* 40.4 (2014), pp. 1353–1362. ISSN: 0045-7906. DOI: https://doi.org/10.1016/j.compeleceng.2013.11.019. URL: https://www.sciencedirect.com/science/article/pii/S0045790613003017.

[14] The Smart Phone Photographer. *Fully Explained: How Smartphone Cameras Focus*. https://thesmartphonephotographer.com/how-smartphone-cameras-focus/#:~:text=In%20terms%20of%20how%20the,to%20focus%20in%20the%20scene.. [Online; accessed 15-July-2022]. 2022.

[15] GSMArena. *Huawei P50 Pro Specifications*. https://www.gsmarena.com/huawei_p50_pro-11029.php. [Online; accessed 15-July-2022]. 2022.

[16] Su, Y., Lin, J. Y. and Kuo, C.-C. J. A model-based approach to camera's auto exposure control. *Journal of Visual Communication and Image Representation* 36 (2016), pp. 122–129. ISSN: 1047-3203. DOI: https://doi.org/10.1016/j.jvcir.2016.01.011. URL: https://www.sciencedirect.com/science/article/pii/S1047320316000201.

[17] Huo, J.-y., Chang, Y.-l., Wang, J. and Wei, X.-x. Robust automatic white balance algorithm using gray color points in images. *IEEE Transactions on Consumer Electronics* 52.2 (2006), pp. 541–546. DOI: 10.1109/TCE.2006.1649677.

[18] Winkler, S. Issues in vision modeling for perceptual video quality assessment. *Signal Processing* 78.2 (1999), pp. 231–252. ISSN: 0165-1684. DOI: https://doi.org/10.1016/S0165-1684(99)00062-6. URL: https://www.sciencedirect.com/science/article/pii/S0165168499000626.

[19] Fedak, V. and Nakonechny, A. Spatio-temporal algorithm for coding artifacts reduction in highly compressed video. (2015). URL: http://psjd.icm.edu.pl/psjd/element/bwmeta1.element.ojs-nameId-6e3e8ea9-5a94-37a7-828b-e6cd5da23db6-year-2015-article-1761.

[20] Photographer, T. S. *Optical Zoom On Smartphones: What Is It Really?* https://thesmartphonephotographer.com/what-is-optical-zoom/. [Online; accessed 27-May-2022]. 2021.

[21] Allen, E. and Triantaphillidou, S. *The Manual of Photography, 10th Edition*. eng. Routledge, 2012. ISBN: 0240520378.

[22] Android Authority. *Camera zoom explained: How optical, digital, and hybrid zoom work*. https://www.androidauthority.com/camera-zoom-optical-digital-hybrid-1021264/. [Online; accessed 27-May-2022]. 2022.

[23] Baker, S., Bennett, E., Kang, S. B. and Szeliski, R. Removing rolling shutter wobble. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 2392–2399. DOI: 10.1109/CVPR.2010.5539932.

[24] Cormier, E., Cao, F., Guichard, F. and Viard, C. Measurement and protocol for evaluating video and still stabilization systems. *Image Quality and System Performance X*. Ed. by P. D. Burns and S. Triantaphillidou. Vol. 8653. International Society for Optics and Photonics. SPIE, 2013, pp. 9–18. DOI: 10.1117/12.2003583. URL: https://doi.org/10.1117/12.2003583.

[25] InvenSense. *Electronic image stabilization*. https://invensense.tdk.com/solutions/electronic-image-stabilization/. [Online; accessed 21-July-2022]. 2022.

[26] Seibold, C., Hilsmann, A. and Eisert, P. Model-based motion blur estimation for the improvement of motion tracking. *Computer Vision and Image Understanding* 160 (2017), pp. 45–56. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2017.03.005. URL: https://www.sciencedirect.com/science/article/pii/S1077314217300590.

[27] Hanning, G., Forslöw, N., Forssén, P.-E., Ringaby, E., Törnqvist, D. and Callmer, J. Stabilizing cell phone video using inertial measurement sensors. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 2011, pp. 1–8. DOI: 10.1109/ICCVW.2011.6130215.

[28] Jin, Y., Guo, X., Li, Y., Xing, J. and Tian, H. Towards stabilizing facial landmark detection and tracking via hierarchical filtering: A new method. *Journal of the Franklin Institute* 357.5 (2020), pp. 3019–3037. ISSN: 0016-0032. DOI: https://doi.org/10.1016/j.jfranklin.2019.12.043. URL: https://www.sciencedirect.com/science/article/pii/S0016003219309561.

[29] Cui, Z. and Jiang, T. No-Reference Video Shakiness Quality Assessment. *Computer Vision – ACCV 2016*. Ed. by S.-H. Lai, V. Lepetit, K. Nishino and Y. Sato. Cham: Springer International Publishing, 2017, pp. 396–411. ISBN: 978-3-319-54193-8.

[30] RED. *PANNING SPEED BEST PRACTICES*. https://www.red.com/red-101/camera-panning-speed. [Online; accessed 29-Jul-2022].

[31] Choi, J. and Kweon, I. S. Deep Iterative Frame Interpolation for Full-Frame Video Stabilization. *ACM Trans. Graph.* 39.1 (Jan. 2020). ISSN: 0730-0301. DOI: 10.1145/3363550. URL: https://doi.org/10.1145/3363550.

[32] Shaw, M. R.-T. and Triggs, R. *EIS Is Actually More Important Than OIS for Videos*. https://medium.com/@GadgetHax/eis-is-actually-more-important-than-ois-for-videos-ccd4805b3245. [Online; accessed 13-May-2022]. 2019.

[33] GoPro. *GoPro Hero10 Black specifications*. https://gopro.com/en/us/shop/cameras/hero10-black/CHDHX-101-master.html. [Online; accessed 06-June-2022].

[34] Shaw, A. R.-T. and Triggs, R. *What is image stabilization? OIS, EIS, and HIS explained*. https://www.androidauthority.com/image-stabilization-1087083/. [Online; accessed 12-May-2022]. 2021.

[35] Forbes, B. S. *Vivo X50 Pro Review: Gimbal Camera Works, But There's More To Like Elsewhere*. https://www.forbes.com/sites/bensin/2020/07/16/vivo-x50-pro-review-gimbal-camera-works-but-theres-more-to-like-elsewhere/?sh=9e873efc93ff. [Online; accessed 06-June-2022]. 2020.

[36] Engadget, R. L. *Vivo explains the X50 Pro's gimbal-like camera stabilization*. https://www.engadget.com/vivo-x50-pro-gimbal-camera-182524376.html. [Online; accessed 15-July-2022]. 2021.

[37] GSMArena. *OPPO Find X5 Pro Specs*. https://www.gsmarena.com/oppo_find_x5_pro-11236.php. [Online; accessed 21-July-2022]. 2022.

[38] iPhoneWired. *OIS optical image stabilization, which is very common on mobile phones: it turns out to be divided into three, six, nine, etc.* https://iphonewired.com/industry/357192/. [Online; accessed 21-July-2022]. 2022.

[39] GSMArena. *Asus Zenfone 9 specifications*. https://www.gsmarena.com/asus_zenfone_9-11656.php. [Online; accessed 17-August-2022]. 2022.

[40] Apple. *Exposure to vibrations, like those generated by high-powered motorcycle engines, might impact iPhone cameras*. https://support.apple.com/en-us/HT212803. [Online; accessed 10-June-2022]. 2021.

[41] Guilluy, W., Beghdadi, A. and Oudre, L. A performance evaluation framework for video stabilization methods. *2018 7th European Workshop on Visual Information Processing (EUVIP)*. 2018, pp. 1–6. DOI: 10.1109/EUVIP.2018.8611729.

[42] *ISO 20462-3:2012 Photography — Psychophysical experimental methods for estimating image quality — Part 3: Quality ruler method*. Standard. May 2012.

[43] Zwanenberg, O. van, Triantaphillidou, S., Jenkin, R. B. and Psarrou, A. Estimation of ISO12233 Edge Spatial Frequency Response from Natural Scene Derived Step-Edge Data. *Journal of Imaging Science and Technology* 65.6 (2021), pp. 60402–1.

[44] Saad, M., Bovik, A. C. and Charrier, C. Blind prediction of natural video quality and h. 264 applications. *Seventh International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VQPM)*. 2013, pp. 47–51.

[45] Choi, L. K. and Bovik, A. C. Flicker sensitive motion tuned video quality assessment. *2016 IEEE southwest symposium on image analysis and interpretation (SSIAI)*. IEEE. 2016, pp. 29–32.

[46] Shang, Z., Ebenezer, J. P., Wu, Y., Wei, H., Sethuraman, S. and Bovik, A. C. Study of the Subjective and Objective Quality of High Motion Live Streaming Videos. *IEEE Transactions on Image Processing* 31 (2021), pp. 1027–1041.

[47] DxoMark. *Smartphone Imaging Trends: New Directions Capturing Magic Moments*. https://www.airmeet.com/e/ca661610-d756-11ec-9df2-01f3ac62b311. [Online; accessed 08-June-2022]. 2022.

[48] Imatest. *Test Lab Setup*. https://www.imatest.com/solutions/test-lab-setup/. [Online; accessed 12-July-2022]. 2022.

[49] Imatest. *Imatest test chart catalogue*. https://www.imatest.com/docs/sfrplus_instructions/. [Online; accessed 12-July-2022]. 2022.

[50] Image Engineering. *STEVE6D motion platform*. https://www.image-engineering.de/products/equipment/measurement-devices/825-steve-6d. [Online; accessed 12-July-2022]. 2022.

[51] Image Engineering. *ISO 20954-2 Handshake Profile for STEVE-6D*. https://www.image-engineering.de/news/product-news/1035-iso-20954-2-handshake-profile-for-steve-6d. [Online; accessed 12-July-2022]. 2022.

[52] Bucher, F.-X., Park, J. Y., Partinen, A. and Hubel, P. Issues reproducing handshake on mobile phone cameras. *Electronic Imaging* 2019.4 (2019), pp. 586–1.

[53] Teed, Z. and Deng, J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. Vol. 12347 LNCS. 2020. DOI: 10.1007/978-3-030-58536-5_24.

[54] Schelp, C. *An Alternative Way to Plot the Covariance Ellipse*. https://carstenschelp.github.io/2018/09/14/Plot_Confidence_Ellipse_001.html. [Online; accessed 19-Octorber-2022]. 2018.

[55] Imatest. *Sharpness: What is it and How it is Measured*. https://www.imatest.com/docs/sharpness/. [Online; accessed 26-July-2022]. 2022.

[56] *ISO 12233:2017 Photography — Electronic still picture imaging — Resolution and spatial frequency responses*. Standard. 2017.

APPENDIX A: FREQUENCY RESPONSES

The euclidean distance frequency responses for the 1x and 5x laboratory videos are shown in Figures A.1 and A.2. The goal of the frequency responses is to show what type of motion a video contains. A video can have rapidly changing high frequency motion which might cause motion blur.

The frequency responses of Device A and C in Figure A.1 do not contain much energy as they taper off quite quickly. Device B has more energy between 0 - 1 Hz while the rest of the frequencies do not contain much energy. Device D has more energy compared to the other devices and does not taper as quickly as the other devices. This indicates that larger high-frequency motion is present in the device D 1x video.

In Figure A.2, Device C has a lot of energy around 5 Hz which is likely the frequency of the motion that caused the motion blur in the video. Devices A and B have most of the energy between 0 and 1. Device D contains most of the energy between the frequencies 0 and 2.

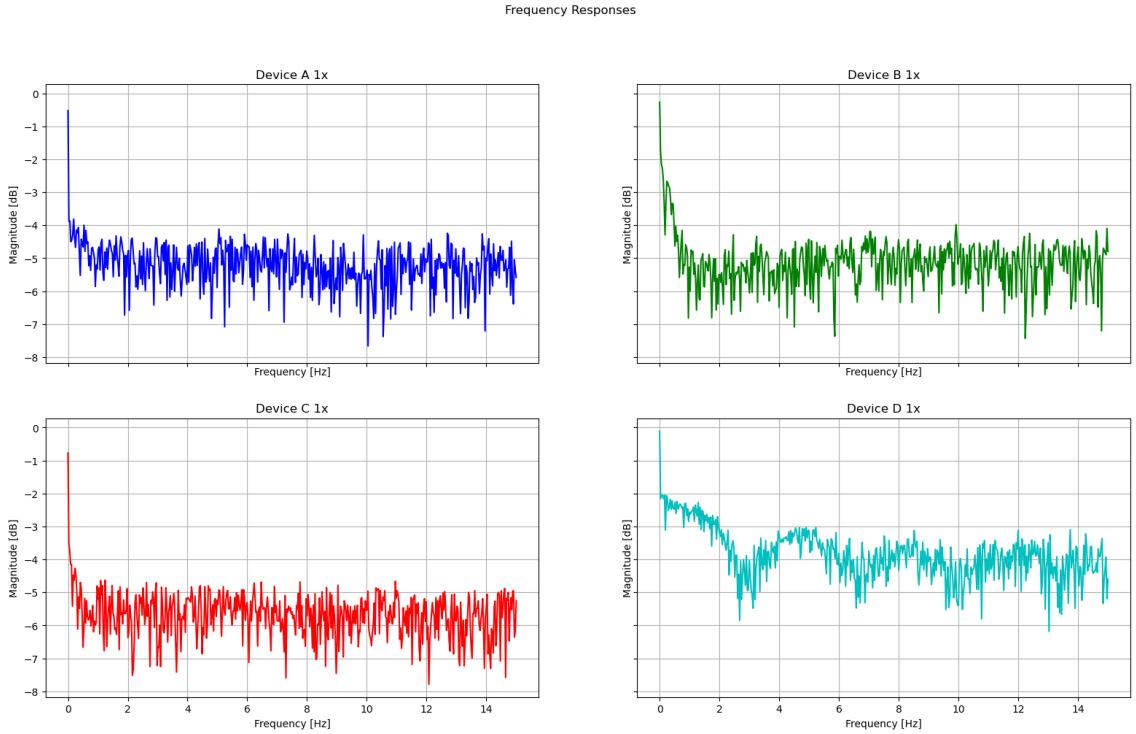


Figure A.1. *Frequency responses of 1x labratory videos.*

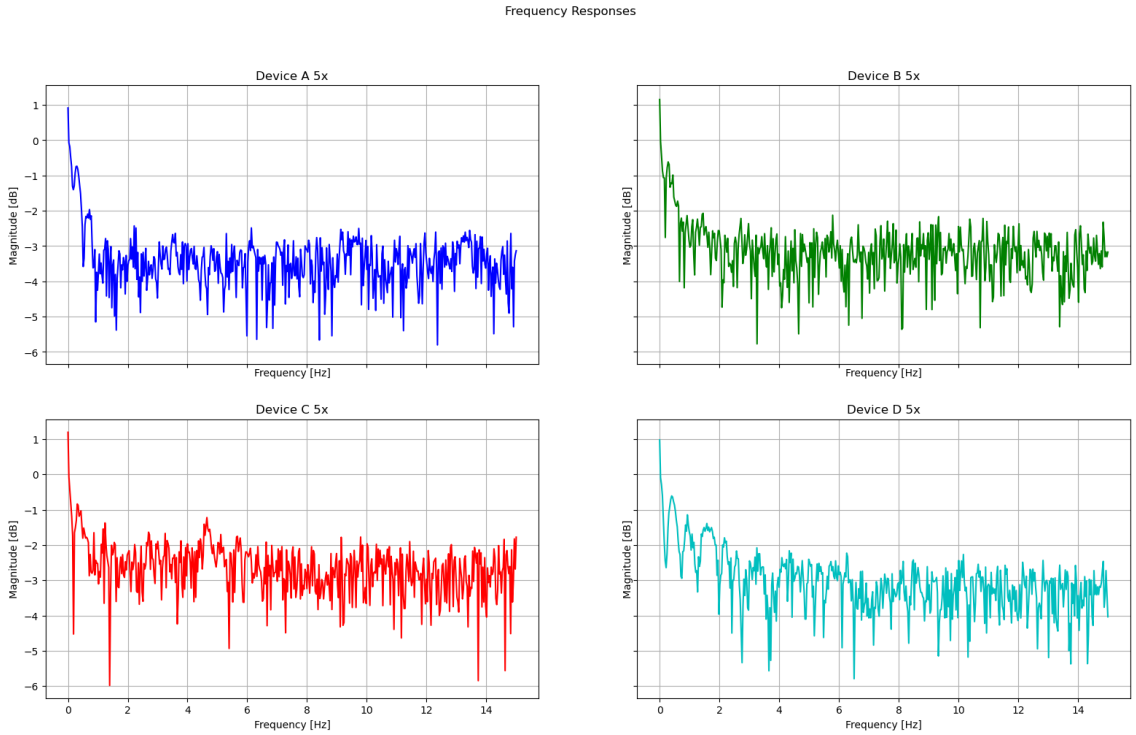


Figure A.2. *Frequency responses of 5x labratory videos.*

APPENDIX B: CONFIDENCE ELLIPSES

The euclidean distance confidence ellipses for the 1x and 5x laboratory videos are in Figures B.1 and B.2. They describe the global motion of each video. The 1x confidence ellipses show that devices A, C, and D have more of a horizontal moving pattern, while device B has a more vertical moving pattern while device B. The 5x confidence ellipses have a similar magnitude in y-axis movement while varying more in their x-axis movement magnitude. Table B.1 shows the mean euclidean distances for each device. In the 1x case, Device C is the best performer with the smallest mean magnitude in movement and device D is the worst performer with its largest mean magnitude in movement. These observations can also be seen in the confidence ellipse sizes in Figure B.1. In the 5x videos, device A is the best performer and device C is the worst performer in terms of movement's mean magnitude.

Euclidean distance movement mean value summary table				
Zoom ratio	Device A	Device B	Device C	Device D
1x	0.30	0.38	0.23	0.45
5x	1.25	1.59	1.65	1.33

Table B.1. *Euclidean distance movement mean value summary table. The best scores are highlighted in green and the worst in red.*

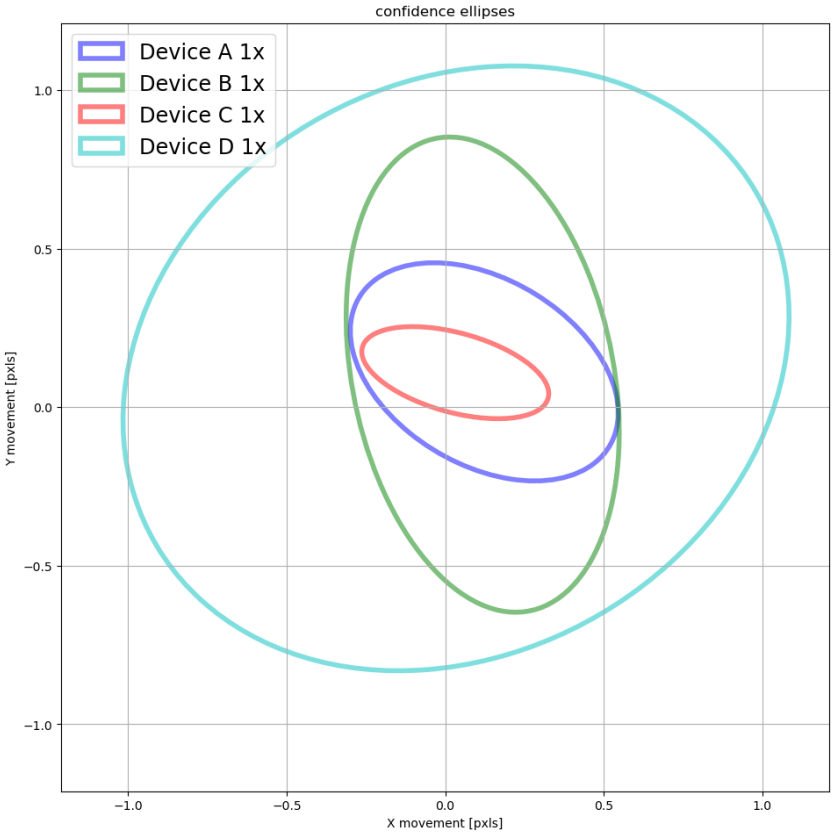


Figure B.1. *Confidence ellipses of the 1x labratory videos*

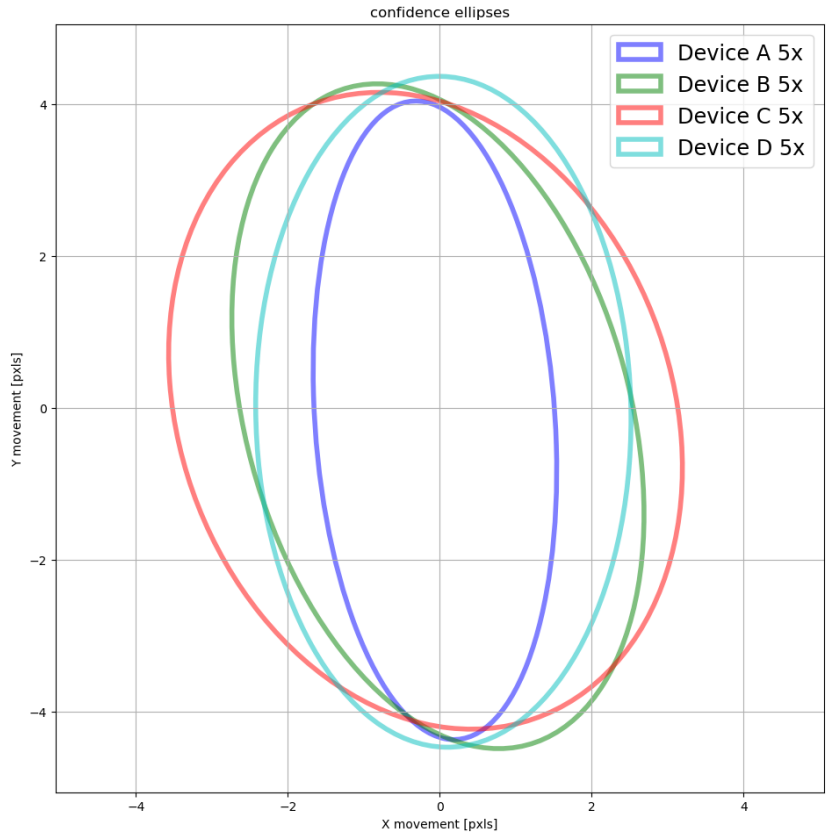


Figure B.2. *Confidence ellipses of the 5x labratory videos*