VCX-WebCam

Version 2023

First Release

VCX-Forum e.V. www.vcx-forum.org

10 Nov 2022

CONTENTS

EXECUTIVE SUMMARY

VCX-forum (Valued Camera eXperience) is a recognized independent industry consortium with 5 key tenets — out-of-box experience, 100% objective, open and transparent, independent lab testing, and continuous improvement. The mission of VCX is the creation and dissemination of a standard for objectively assessing the image quality (IQ) of cameras in mobile devices such as smartphones, tablets, computers, notebooks, drones, etc. The objective is to provide an independent and credible measure of quality mapped to the end-user experience for consumers and industry alike.

The VCX-WebCam benchmark focuses on video conferencing. The scope of the benchmark covers both peripheral and integrated cameras in personal computers (PCs) and laptops. Boutique solutions like whole-room conferencing cameras can certainly benefit from the same approach, but may not be in scope for the scoring system. There are 6 test areas and 24 key performance indicators (KPIs) that are driven by specific consumer pain points in areas such as auto-exposure, white balance, spatial frequency response, texture loss, sharpening, noise, color reproduction, exposure/dynamic range, and frame rate. Image quality KPIs are measured in consistent, clearly defined scenarios under a variety of controlled lighting environments that mimic the environments and conditions webcams are primarily used in.

VCX's objective scoring and weighting process is correlated with both expert and untrained end-user subjective feedback. The single score is reflective of the typical end-user experience. With a strong and reliable setup of different evaluation and scoring algorithms, the VCX system does not allow a camera to get a high score without delivering a strong performance.

This specification was developed within the webcam standards department of VCX-forum, which includes members from across the PC ecosystem, including leading OEMs, ISP developers, software and hardware vendors, and other interested parties. It is the ecosystem's participation in developing and driving this benchmark that ensures both it's credibility and reliability.

For questions about this benchmark or other work being done within VCX-forum, see the website at vcx-forum.com or send a message to info@vcx-forum.org.

1 INTRODUCTION

The first web-based video camera, or "webcam", is said to have been created by students at Cambridge University in 1991. The image was 129x129 pixels, ran at a blistering 1 frame per second (fps), and was used to keep an eye on the coffee pot just outside of the computer lab. By the summer of 2010, 23% of internet users had tried video conferencing [1]. By 2022, post-pandemic usage statistics demonstrate how video calling has become entrenched in our society and culture, with more than 300 million daily users of Zoom [2], and 77% of all video calls being made from a laptop or desktop computer (as opposed to a smartphone or other mobile device) [3]. Rapid technological advances in camera technology, network speed, and software have led to increasing but widely diverse levels of image quality across the mobile device and personal computer industries. Exaggerated marketing claims & competing technologies across a wide variety of devices have created a confusing situation for customers looking for a good performance. There is a clear need not only to characterize image quality and performance in simple terms for consumers but also to help the industry measure quality to drive improvements. A market devoid of a transparent, open, and objective scoring system has exacerbated this situation.

VCX-WebCam 2023 was developed to address the unique challenges that come with evaluating webcam image quality and performance. The result is a transparent, objective score that reflects the end-user experience of that device.

2 VCX OVERVIEW

VCX-forum is based on 5 tenets, acting as a framework to drive results mapped to real-life experience:

1. VCX measurements shall be representative of out-of-the-box experience
2. VCX shall remain 100% objective
3. VCX shall be open and transparent
4. VCX shall employ/use an independent imaging lab for testing
5. VCX shall seek continuous improvement

Tenet 1. **VCX measurements shall be representative of out-of-the-box experience**: The VCX score is designed to reflect the user experience with a camera to make it much easier for end-users to decide what they need or want from a new device. The results are obtained from a device launched onto the market to ensure that neither special samples from suppliers nor custom hardware/software are accepted. The device is tested using the default camera application and setting (except for flash/burst mode test cases, if applicable) from the device.

Tenet 2. **VCX shall remain 100% objective**: The score is calculated from measurements solely based on objective analysis of the device under test, followed by fixed and unbiased processing of the numerical results. No human interaction or subjective scoring is involved when generating the VCX score for a device.

Tenet 3. **VCX shall remain open and transparent**: The VCX score is intended to be useful to the consumer, rather than just an internal target for ecosystem participants. VCX members may publish the score in marketing campaigns or other literature, and qualified VCX scores may be published on the website www.vcx-forum.org. The specification itself, as well as the scoring methodology, is open to critique and scrutiny by the imaging community at large - within the VCX forum. High-level weighting criteria are published along with the details of high-level components that make up the final VCX score.

Tenet 4. **VCX shall employ/use an independent imaging lab for testing**: Anyone is welcome to build and utilize a VCX-certified lab for the purposes of testing and benchmarking devices. To combat bias, however, the results on vcx-forum.org are restricted to results measured or audited by trusted labs to ensure the highest quality and consistent results in order to qualify for publishing. Other independent imaging labs are welcome to join the VCX initiative; processes and procedures for the inclusion of other entities to actively contribute to VCX are defined in the VCX handbook.

Tenet 5. **VCX shall seek continuous improvement**: VCX forum members are leading ecosystem partners from the PC and Mobile industry. Members closely watch and follow industry trends and contribute positively to continuous specification development. VCX specifications have been developed over several years and will be regularly updated on an annual or as-needed basis.

3 PHILOSOPHY

Video conferencing is the primary application for webcams at a time when image quality analysis is often firmly rooted in a "still camera" approach with KPIs from years past and tuning time spent optimizing still use cases instead of video. The philosophy behind the VCX-WebCam specification is to prioritize video first and to modernize the current analytical approach used to evaluate camera systems. The basic idea is to:

- Develop the specification based on common pain points that affect video conferencing based on industry experience and user studies
- Pay special attention to dynamic behavior of 3A and image processing algorithms over multiple frames
- Modernize test methods to reflect the complexity of algorithms commonly in use today

To a consumer, the measure of "quality" for a camera is really a gestalt of their experience with all aspects of that device, often weighted by what is most distracting. For this reason, VCX checks for many different aspects of image quality in viewing conditions resembling situations commonly found in the field. Though there are many philosophies of how to tune a camera, user studies show there is an easily observable range of what is considered "good quality" across devices. With a strong and reliable lineup of different evaluation and scoring algorithms, the VCX approach does not allow a camera to attain a high score without delivering a strong performance in terms of image quality and dynamic behavior.

4 SPECIFICATION PROFILES & SCOPE

A profile is a set of test cases that are targeted at a particular type of device. VCX specifications are currently divided into two profiles: "VCX-PhoneCam", developed for smartphone cameras, and "VCX-WebCam", covering both integrated and peripheral webcams. Boutique devices such as high-end, whole-room conferencing cameras may be out of scope for the VCX score but can certainly benefit from many of the same testing approaches.

Each profile is further divided into metrics subgroups. This is a subset of test cases within a profile that covers at least one related use-case. For example, "Spatial Frequency Response" groups use multiple sharpening-related metrics together in the VCX-WebCam profile.

The weighting for metrics-groups within the profile is derived from usage data from user studies, industry experience, and testing. For example, in the case of the VCX-PhoneCam profile, usage statistics suggest that the main camera of the smartphone is used 80% of the time, hence this use case is weighted heavily.

The weighting for metrics groups is determined similarly regardless of profile but it is data from that profile's ecosystem that is used. This means that a similar metrics group may have different weighting under different profiles. For example, "Video" has a higher overall weight in the VCX-WebCam profile than the corresponding "Video" metric group under the VCX-PhoneCam profile, given that video is the primary use-case in the VCX-WebCam profile.

This document pertains to VCX-WebCam, which addresses cameras usually integrated with or connected to laptops and PCs for the purpose of video conferencing and other web-streaming usages.

5 TEST INTRODUCTION

5.1 Approach

The general approach taken in defining the VCX-WebCam profile started with a list of critical image quality issues observed in various devices in the market. At the time of writing, easily observable problems related to exposure, white balance, skin tone reproduction, fine detail reproduction and poor dynamic range are prevalent across the webcam ecosystem, whether they be discrete or built into a PC device. The image quality testing and scoring methodology built into the VCX specification is to not only provide a fair baseline evaluation but to also fill gaps in how camera devices have been evaluated that allow those pain points to persist. This approach drives improvements through measurements across the webcam ecosystem.

5.2 Test Areas

The 2022 VCX-WebCam profile specification focuses primarily on 6 key areas, measured across a broad range of lighting conditions, with 24 KPIs representative of how people use their cameras (see Appendix A).

Noise

Is luma or chroma noise obvious? How visible and/or disturbing is it to the viewer?

Visual noise - A visual noise approach as specified in ISO 15739 is used, which better reflects the way a human being experiences the camera under test. Visual noise in the lightness channel L* and the two chrominance channels a* and b* are determined for all patches of the color checker, then a contrast sensitivity function (CSF) is applied to weight this data in a similar way to how a human eye would by considering the temporal aspect of noise in video against both luma and chroma signals [4]. See Appendix D for KPI references.

Contrast Response, Dynamic Range, Exposure

Is the image bright enough? Does it look washed out? Are details preserved in high-dynamic scenes?

Contrast Response and Dynamic Range – Measurements are taken from an OECF chart across a range of light conditions. The resulting score is derived from how well these measurements fit within an envelope curve to ensure good performance while still allowing for individual preference tuning.

Exposure – Tests are focused on evaluating image brightness and exposure accuracy based on a target condition. Speed and stability of the convergence process, in the presence of faces and no faces, are tested as well L* values of the face ROI are included as a measure of face brightness, and a comparison is made to L* values in the colorchecker chart to verify overall image exposure is not significantly compromised to meet face brightness metrics..

Spatial Frequency Response

What level of detail can I see? How well does the device reproduce low contrast, fine details? Does the device apply so much sharpening that disturbing artifacts occur? Does the image look blurry?

Resolution - Describes the capability of the system under test to reproduce details in the scene. Both the level of details and the acutance, a metric that has shown good correlation with human perception of sharpness are derived using the s-SFR method (ISO12233) based on the spatial frequency response measured on the Siemens star. Unlike the e-SFR method based on slanted edges, the s-SFR method has shown its robustness against image enhancement algorithms.

Texture loss - Defined as the loss of low contrast fine details, mainly due to noise reduction or compression algorithms. This measurement is obtained using a colored dead leaves target, which consists of various circles with random color, intensity, and size stacked on top of each other, by following the method defined in ISO19567-2.

Sharpening - The act of enhancing the subjective impression of sharpness of an image by increasing the local contrast along edges using image enhancement algorithms. While a gentle amount of sharpening is beneficial for the subjective image quality, excessive sharpening will lead to strong and disturbing artifacts in the image. Sharpening is defined by the maximum response in the SFR_{DL} while the sharpening artifacts are reflected by the undershoot and overshoot measurements based on the edges spread function (ESF) method on slanted edges (see also ISO12233).

Color Reproduction

Does color and hue look natural? Are skin tones preserved?

Accuracy - The VCX-WebCam specification adopts common industry approaches utilizing a color-checker chart to measure mean and max Delta E (CIE Lab color difference), Delta C (chroma difference), & Delta H (hue difference) across a range of scene conditions.

Skin tone reproduction - The presence of a face in a scene should be a key driver of ISP and 3A behavior, as this is the reason the webcam exists. Unfortunately, many cameras fail to adequately adjust in such situations (especially for darker skin tones). Throughout the VCX-webcam specification, targets are adapted to the presence and skin tone of realistic test mannequins developed for lab-based camera testing. C* and H* values of an ROI over the face are used for color accuracy metrics.

Image Uniformity

Are obvious luminance / color shading effects or other spatial non-uniformities present in an image?

Image Uniformity - Poor optics, thermal effects, light leaks, sensor calibration, and image processing can all have a dramatic effect on spatial uniformity of an image. While e luminance non-uniformity (often called vignette) is something users have become accustomed to over the years, color non-uniformities can be extremely noticeable in many scenes. Thus, the score for image uniformity is a mix of luminance and color shading measures across a range of illuminants..

Frame Rate

Are frames captured and displayed at a frequency fast enough to keep up with the world around you?

Frame rate – Evaluated based on an LED-based measurement system. Combining ground-truth exposure data from this measurement with data from collected videos enables the derivation of true frame rate.

6 VCX SCORE

The intent of the VCX score is to create a transparent, objective benchmark that correlates to the perceptual image quality and performance experience of a webcam. This provides value to the manufacturer as a repeatable, granular measure of performance of their own products and to the consumer as an unambiguous benchmark to compare competitive devices.

Objective data is generated by well-defined and transparent test procedures following international standards where possible. The entire analysis is based on frames extracted from captured videos. Numerical analysis algorithms are applied to these frames to generate KPI results, which become the input metrics to the score system to generate a single numerical result.

The transformation of KPIs into individual scores is done using unique transforms based on a definition of a worst and best value for each KPI. This transformation is performed in different ways between the extreme points, depending on the metric itself. For some metrics, the correlation between “metric” and “influence in image quality” is linear, so the score is a linear function of the metric. This would be the case of a simple “the higher the better” or “the lower the better” assumption.

However, some metrics require a different approach in order to reflect the perceived quality. Sharpening is a good example - no sharpening is not beneficial for the image quality, as an image would appear flat, while an excessive sharpening can quickly result in an artificial and unpleasant look of the image. So, there is a “sweet spot”, below or above which a reduction in the score is to be applied.

Additionally, the score is driven by the responsiveness of the device to scene and/or lighting changes, evaluating how fast it can stabilize exposure and white balance, and capture quality-images in dynamic conditions.

Individual score metrics are combined into related sub-groups (auto-exposure and auto-white balance are grouped together, for example). These sub-groups are then summed through a weighting process to produce the final VCX score. The combination of transformation functions and carefully selected weighting factors associated with individual score metrics and sub-groups sets up a system where good performance is rewarded based on image traits that matter more or are more noticeable to end-users, while exceptionally poor performance even in a lesser-weighted KPI can have a strong negative impact to the final score.

The transform functions and weighting factors were determined by a group of VCX experts in conjunction with subjective user studies based on a set of reference devices that were determined to be representative of the range of webcam quality in the PC ecosystem in 2022. These functions and factors are identical for every device to ensure no bias in performance comparisons and are tied to the specific VCX standard version (VCX-webcam 2023, for example).

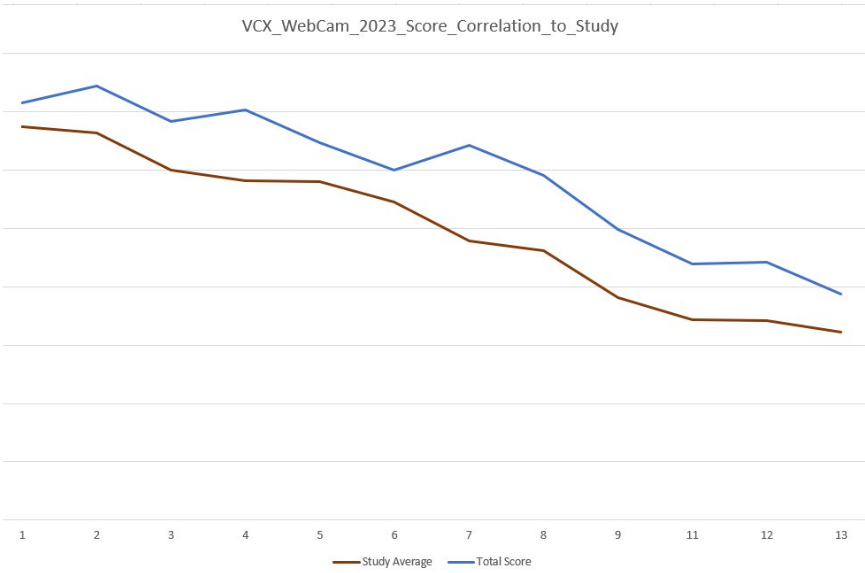


Figure 1: Correlation of webcam score to user study. One outlier device not shown due to IQ issues not covered by the study, so correlation not relevant for that device.

The total score ranges between 0 and 100, with 100 meaning that the device meets the best possible result achievable with today's camera technology across every test category as defined by the VCX-WebCam specification. As the camera industry improves ISP and camera system capabilities, VCX will update the VCX-webcam specification and the score generation process to reflect these improvements. When such a change is made, some methodology will be provided to either directly or indirectly compare scores against the previous version when changes are introduced for backwards compatibility to the most recent standard

The VCX WebCam 2023 camera score is generated entirely from captured video in a lab environment across a range of test conditions intended to reflect the environments webcams are most likely to be used in. The guiding philosophy is that, for webcams, video is the main use case.

The webcam score is driven by the 6 key focus areas. Spatial response, noise KPIs, and the relationship between them are heavily weighted in the calculation. The VCX expert study demonstrated that well performing devices appeared sharp, retained texture, and had low noise while worse performing devices often made a tradeoff for low noise at the cost of poor texture and fine detail reproduction. Exposure and white balance also significantly contribute to the overall score; 3A issues are one of the most noticeable image traits to the end user. Color, contrast, and dynamic range are weighted less as improvements in these areas tend to be less visible in systems that are poorly performing in other areas, but still impact overall score, especially in good performing devices. Frame rate is more binary in nature; it has little effect on the score unless things go wrong.

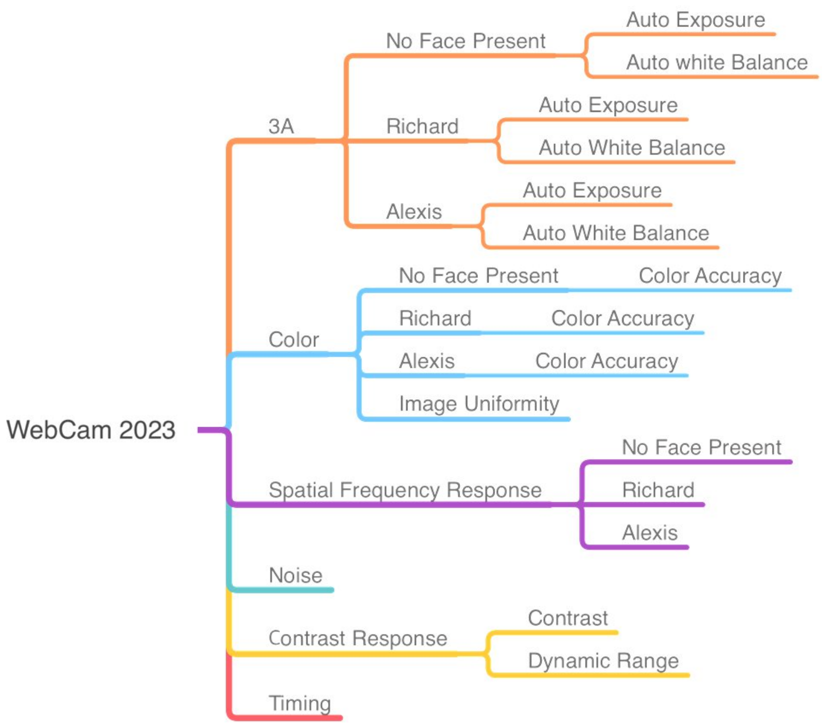


Figure 2: High level conceptual diagram of VCX score algorithm

In every case, scenes with faces are weighted more heavily than scenes without faces. In most cases light and dark faces are weighted evenly. It is worth pointing out that, across the development dataset, most cameras performed poorly with dark skin tone compared to light skin tone, often moving into the “extreme” category where scores go steeply negative, thus strongly affecting the final score. VCX recommends significant effort be put into the tuning process to ensure cameras behave well with all skin tones.

The VCX score calculation is available only to VCX members. To obtain a score, a device must be evaluated by a certified or trusted VCX laboratory. Only test data validated by a trusted lab is qualified for publishing to the VCX-forum website. For non-VCX members, see Appendix A for recommended KPI ranges. These targets were selected to represent individual KPI values that would score ≳90 and are intended to be a guide to help direct tradeoff decisions in the camera development process. It is not expected that a single camera is able to meet every one of these targets based on the current state of the art in imaging across the PC ecosystem.

7 TEST ENVIRONMENT

The test environment, designed to reproduce common conditions for video conferencing and other types of video communication, uses a variety of light sources with varying brightness levels. The webcam studio chart is centered within the field of view to support various test procedures. This studio chart includes mannequins of different skin tones and a head-turn device which mimics a human turning their face away from the camera. Each test is repeated for light skin tone, dark skin tone, and no face present. Unless otherwise stated, images will be still frames taken from a video that was recorded from the device.

7.1 Device Preparation

Each device is checked before the test procedure starts for obvious mechanical problems like scratches on the lens or other signs of mechanical shock to the device. The lens is cleaned and the whole device rebooted to avoid any interference between other processes on the device. If issues are found, the test is stopped and the test laboratory will notify the customer to resolve the issue or replace the device. The device under test is mounted on a tripod for all tests. Depending on the hardware and the position of the buttons, the best possible fixture on a tripod is used. The tripod is mounted on a rail system, so the distance between the chart and camera can be easily adjusted without impacting device alignment to the scene.

7.2 Camera Settings

Testing is aimed at the “out of the box” experience, meaning that it is performed on camera modules that are set to factory default video mode. However, various test procedures can also be used in conjunction with other camera settings and modes, including raw image data capture modes, and measurements used for other purposes, if applicable.

Devices are tested in “as-shipped” condition and in the orientation (landscape or portrait) intended for their use. Camera settings are set to default values during device preparation, with only resolution settings being changed during image capture. While some devices allow manual adjustment, of ISO, exposure, white balance, and other parameters, these settings remain unchanged during the test.

7.3 Image Capture and Analysis

There are several software options for capturing test images and videos: IQ Analyzer (preferred), the default camera application in the device’s operating system, or a capture utility provided by the customer at submission allowing single frame captures from the video pin that are equivalent to the frame(s) captured by the first two options. No special processing is allowed. If needed, common imaging software, such as FastStone or VLC player, configured to not apply any additional processing (e.g., contrast or color enhancement) should be used for frame inspection and/or video playback evaluation.

Single-frame tests: Videos to obtain static images shall be captured for at least 5 seconds to permit for 3A convergence (see appendix A for definition) and captured at the resolutions shown in Table 1 below. Frame extraction and storage shall be done in a lossless manner. VCX recommends FFmpeg [6], an open-source, cross-platform image conversion software.

Multi-frame tests: Procedure is the same as above, except frame extraction should include a time stamp or other indicator of frame progression, and both the frames and the original video should be stored.

4K	If supported
1920x1080 FHD *Strongly recommended to support!*	Test if supported
1280x720 HD	Required

Table 1: Capture resolutions

7.4 Lighting

Light sources with tunable brightness and spectral characteristics are preferred:

- Illuminant spectra need to be defined and accurately matched by the light source in the spectral range between 400 and 1000 nm, including near infrared (NIR) wavelengths for A-light.
- Intensity needs to be adjustable between 10 lux and 800 lux at the surface of the webcam studio chart, see lab layout section for details.
- Maximum of 100mS delay between stabilized illuminant changes; Flicker free
- For mixed lighting and HDR test scenes, a television is used to provide additional background light. TV specs must meet requirements listed in Section 7.5 below and use VCX-approved image set.

Lighting configurations follow industry standard illuminant definition where possible. Where not possible, related reference spectra shall be provided in this document. If standard illumination is not defined, see Appendix B for reference. Table 2 below summarizes all lighting conditions required for this specification and applies to all tests.

Name	Standard Illuminant	Lux level	Reference CCT	Spectrum Definition	Comments
Low light	A	3A: 20lux IQ:10,20,40 lux	2856K	ISO11664-2:2020 See Appendix B	https://en.wikipedia.org/wiki/Standard_illuminant#Illuminant_A
	CW (led)		4000K		No current industry standard. See appendix for reference spectra.
Mid Light	A	3A: 80, 250 lux	2856K		
	WW (led)	IQ: 80, 250 lux	3000K	See Appendix B	No current industry standard. See appendix for reference spectra.
	CW (led)		4000K	See Appendix B	No current industry standard. See appendix for reference spectra.
	D50		5003K	ISO11664-2:2020 See Appendix B	
	D65		6504K		
Bright Light	D50	3A: 500 lux	5003K		
	D65	IQ: 350, 500 lux	6504K		
Mixed Light	A foreground / D65 background	IQ: 80, 250 lux	2856K/6504 K	See above	Face-chart area illuminated as foreground, TV used for background
	D65 foreground / A background		6504K/2856 K	See above	
HDR	Television	200, 400, 700 lux*	N/A	TBD	

Table 2: Light Sources

**division by pi to get to nits for white patch (by reflection factor to get to grey levels)*

Reflective light sources, such as "iQ-Flatlights" from Image Engineering are spectrally tunable and can generate a variety of light spectra based on a mixture of various LEDs by changing not only the correlated color temperature (CCT) but the entire spectrum. The light sources are equipped with a calibrated spectrometer for high reproducibility and constancy.

Figure 3: Reflective Light Source

Integrating sphere or other uniform light source that meets the requirements in Table 2 is required for the dynamic range and shading measurements.

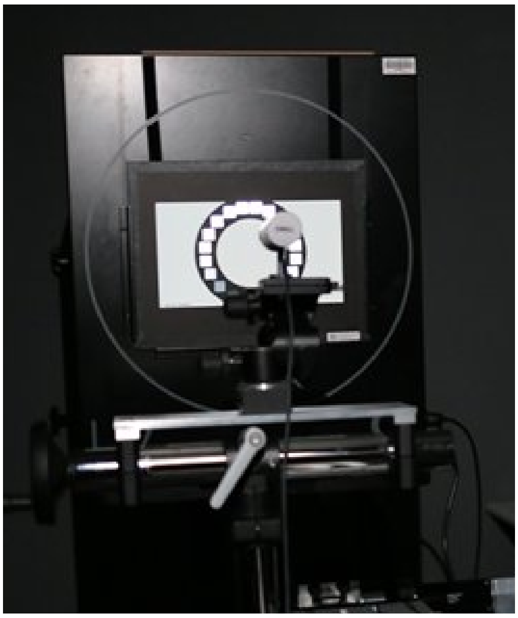


Figure 4: Webcam aligned to integrating sphere for OECF-based tests

See table in [Appendix C](#) table for lighting and other equipment recommendations.

7.5 Test charts and scenes

The *Webcam studio chart* is used for most tests in the VCX specification. As shown in Figure 5, the test chart consists of a colored dead leaves area, 24 patch color checker, and sinusoidal Siemens star of 144 cycles. In rare cases a 72 cycle Siemens star may be needed for very low resolutions cameras. (The stars on the chart are interchangeable.) The open area of the chart is used for the mannequin head or LED box. The chart mounting solution should be generally non-reflective and contribute only minimally to the scene. It should include a shelf or other mounting area to support the mannequin or LED box. The test chart should be placed orthogonally to the TV at a distance of 30 cm. Recommended mannequin mount is a high-quality ball-head mounted on an adjustable vertical post attached to the rotation device. See Appendix C for equipment recommendations.

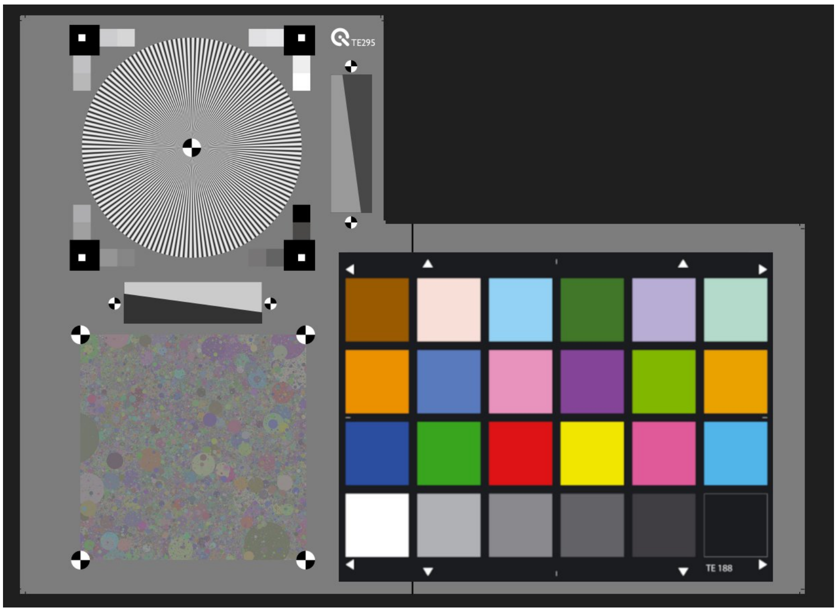


Figure 5: Webcam studio chart

20-patch OECF chart (Figure 6) refers to an OECF (Opto-Electronic Conversion Function) test target which consists of 20 neutral gray patches with different transmission and a total contrast of 100.000:1. In combination with the integrating sphere, this enables the measurement of the tone curve, gamma, and dynamic range. It follows the description of ISO15739.

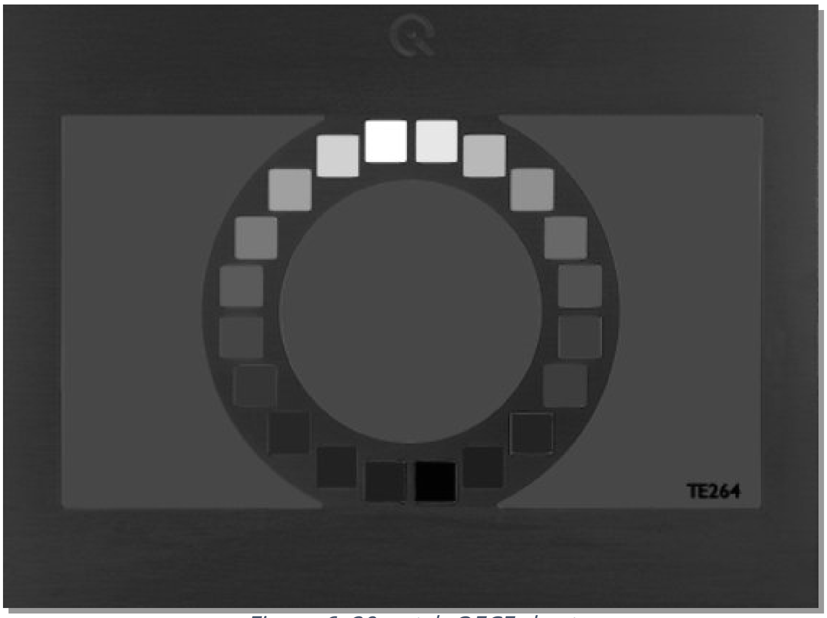


Figure 6: 20-patch OECF chart

Uniform field test chart is a spectrally-neutral diffusor plate which can be inserted into the integrating sphere, as shown in Figure 7 below. The device under test can be placed right in front of the uniform field test chart to capture flat-field images.

Figure 7: Uniform field test chart

An HDR TV with a brightness over 500 nits for 100% white shall be used for the lab scene background, permitting varying background brightness and color to meet the requirements of individual tests. A PC shall be used to drive the TV, outputting a 4K / 120 Hz signal via HDM 2.1. Table 3 summarizes the related requirements.

Size	75”
HDR brightness @ 100% White	>500 nits
Flicker	Effective
Color Accuracy, White balance	<2 dE
Refresh Rate	120Hz +
Gray Uniformity	<5
Resolution	4K+
Color Gamut	Wide (~67% Rec.2020 uv)
Response Time	<12mS

Table 3: Display device requirements

TV should be calibrated annually or as often as needed, using a commercially-available calibration device. See Appendix C for details. PowerPoint shall be used to display the required images and video on the TV, due to support for ICC profiles. An 18% grey board or curtain is required to cover the TV for non TV-based image capture.

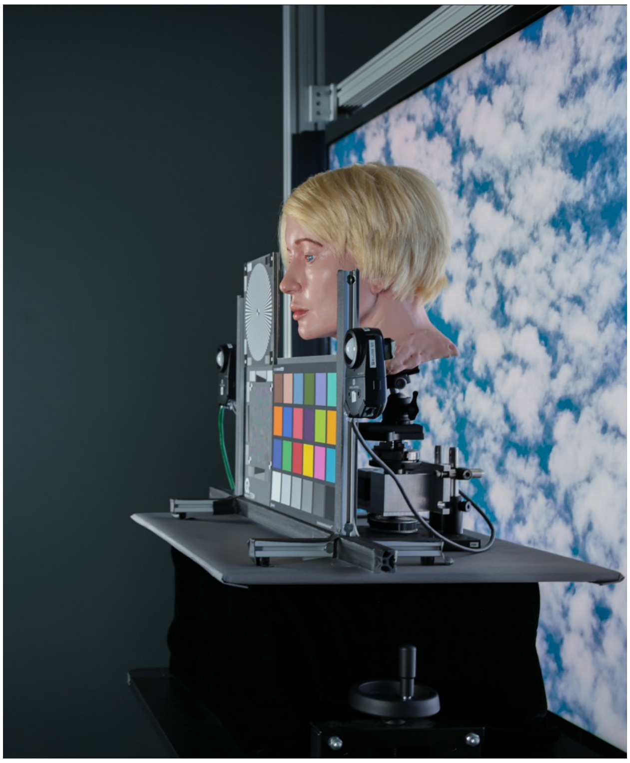


Figure 8: TV with sky image used for white balance challenge test

7.6 Mannequin

Skin tone response has been a key pain point across the PC ecosystem for some time. The test development dataset used to develop this document proves this out, with 20% of dark skin tone images being under-exposed (mostly in low light) compared to 7% for light skin tones). In addition, the range of L* values was more narrow for light than for dark skin tone. A recent study [7] also shows a general trend of lower mean opinion scores for scenes with dark skin tone present (Figures 9 & 10). To address this situation, VCX has standardized on two test mannequins that are used across nearly all VCX image quality tests. These mannequins have been designed to look as realistic as possible, representing a range of skin tones and facial features. For details on where to order these mannequins, see Appendix C.

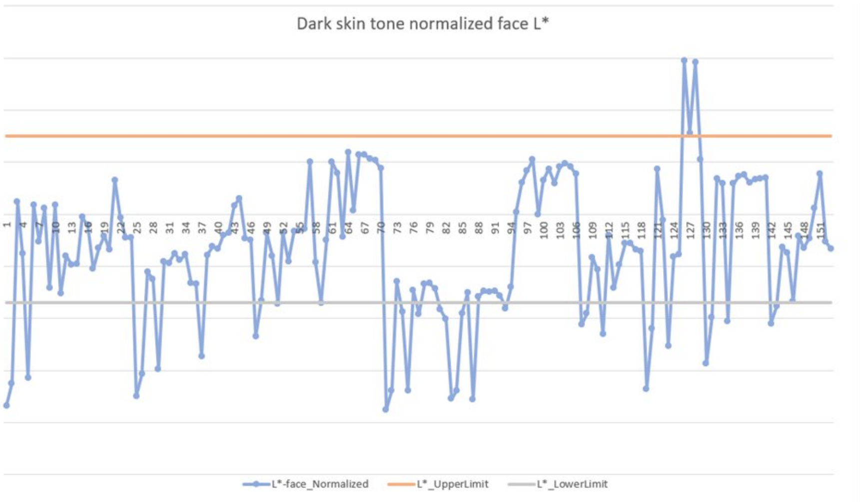


Figure 9: Normalized L of face ROI for dark skin tone*

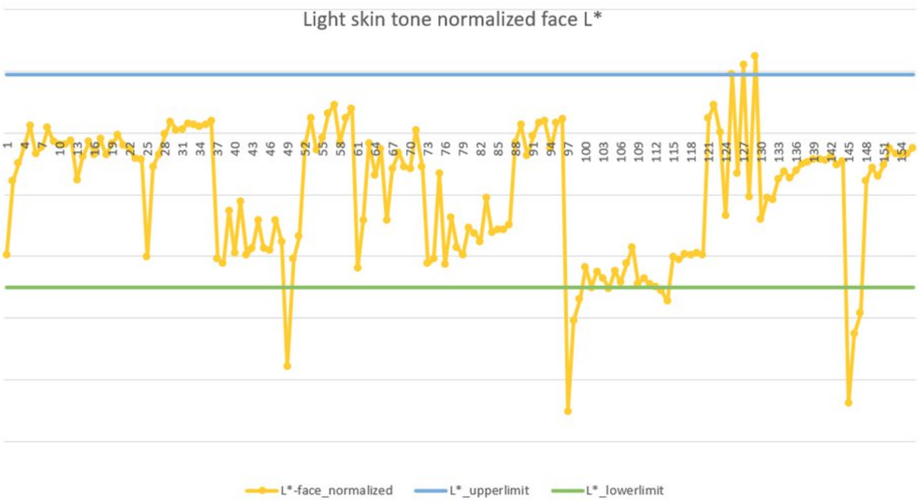


Figure 10: Normalized L of face ROI for light skin tone*

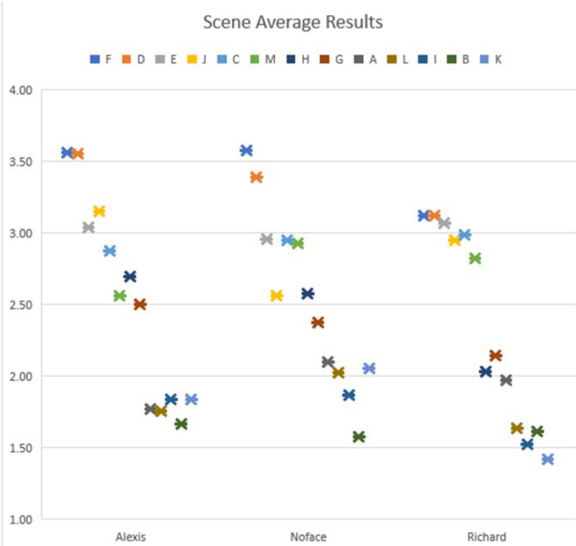


Figure 11: Results from VCX expert study showing difference of results across light & dark skin tone and with no face present.

Mannequins shall be mounted in the lab environment in such a way to allow for easily switching out the dark and light skin tone mannequins as needed at the correct position in the Webcam studio chart, leaving room for rotation. A rotation table or other device shall be used to rotate the heads from 0 to 135 degrees in less than 1.5 seconds. This device should be remote-controlled. See Appendix C for equipment recommendations.

The mannequin shall be placed in the open quadrant of the webcam studio chart; the face must be fully visible and not shadowed by the chart, with the “chin” placed just above the edge of the color chart. Spatially, the mannequin should be positioned such that the eyes and mouth are roughly parallel with the surface of the color chart. The test area shall be uniformly illuminated with less than 10% variation across the camera FOV. The test chart and mannequin must fill 50-70% of the camera FOV. When needed for frame rate measurements, the LED box shall be placed in the open quadrant of the webcam studio chart in place of the mannequin or below the Webcam studio chart.

Figure 12: Richard and Alexis mannequin heads

7.7 Timing-Based Measurements

Frame rate measurements require the presence of a ground-truth measurement of exposure. The main component in this setup is the LED-panel (Figure 13). This device features 100 LEDs, which can be illuminated one after another at a pre-selectable speed, thus acting as a high precision clock. Device is usually placed within the scene in place of the mannequin head.

Figure 13: LED panel

7.8 Device Handling Equipment

Camera device shall be mounted on a system that can reproduce different object distances in a repeatable fashion. The mounted device should be placed on a dolly or rail system that allows the operator to vary the distance between the camera and the chart. Automation is not required but is strongly recommended.

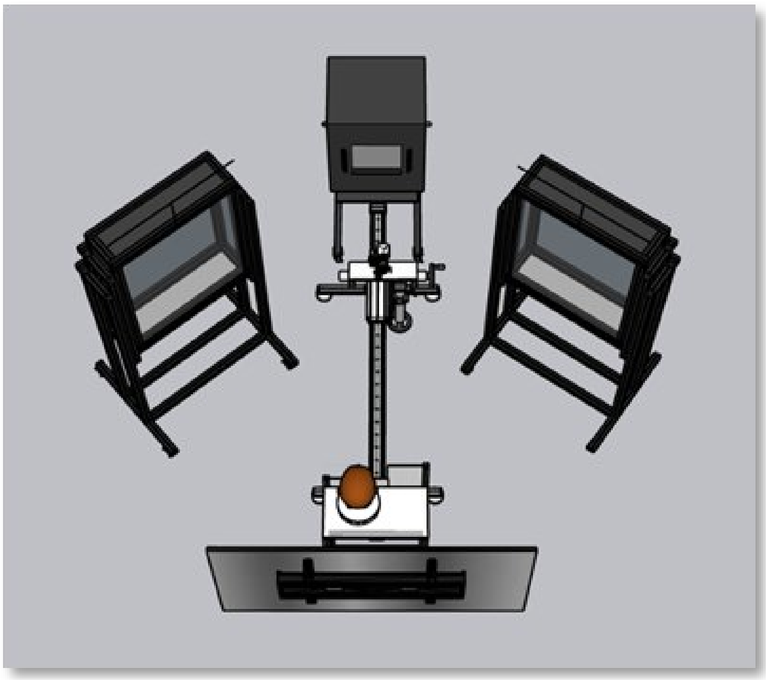


Figure 14: Lab layout illustration

Camera handling equipment details:

Current VCX-webcam tests occur at 75cm. Future revisions likely to add additional tests that require different distances. VCX recommends these ranges for future-proofing lab design:

* Minimum Distance: 30 cm
* Maximum Distance: ≥100 cm

Device mount should be adjustable to provide for alignment of a variety of devices to the target scene. VCX recommends the following adjustment ranges:

- X: +/- 25 cm
- Y: +/- 25 cm
- Z: +/- 25 cm
- Pitch/Yaw/Roll in 1-degree increments

All equipment used should be able to safely and reliably hold a device with a weight of up to 20 lbs in a stable state in multiple orientations.

It is preferred that the device and scene motion is automated. If this is not possible, test operator can implement manual movement of the device and mannequin, but engineering controls must be in place to meet the timing and position requirements as called out in this document. See timing requirements for head turn in AE and AWB sequence tests in Section 8. For details of the head-tuning mechanism see the proof-of-concept device descibed in the spec release package. Commercial head-turn solutions are just coming to market.

7.9 Lab Layout

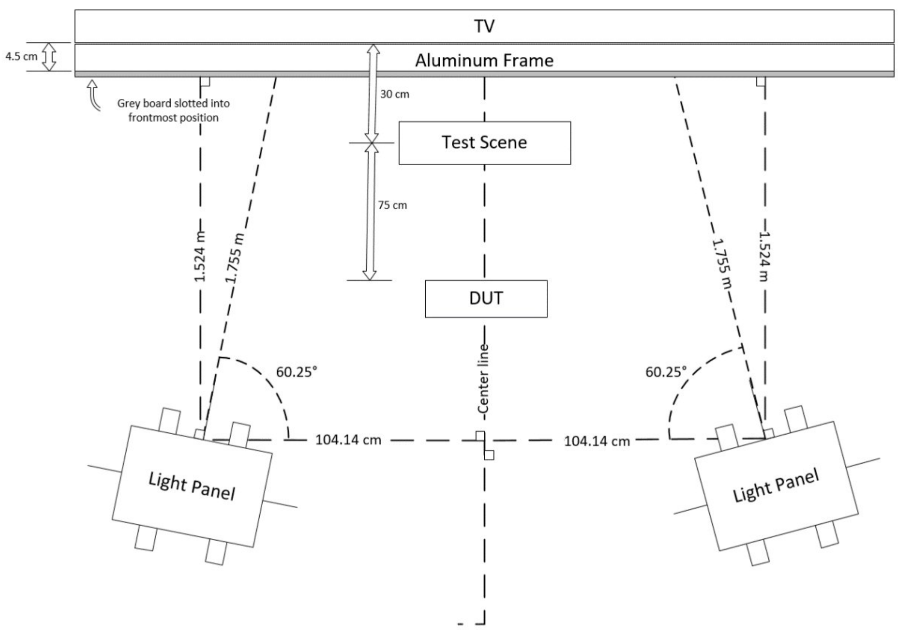


Figure 15: Lab layout

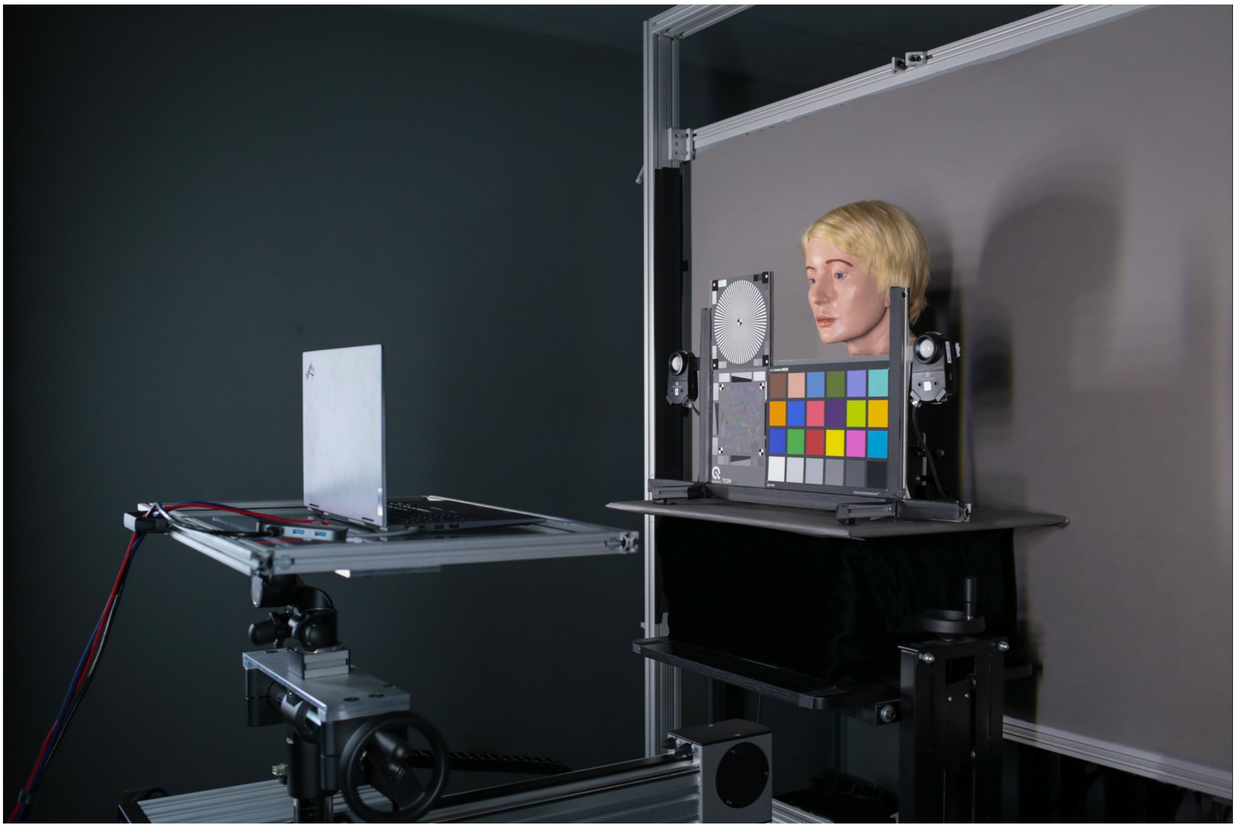


Figure 16: Laptop device aligned to webcam studio chart for color, spatial frequency response and face-based tests

8 TEST DESCRIPTION

8.1 Visual Noise

Denoising algorithms in modern ISPs vary widely in performance. In addition, cost-optimized hardware choices may result in higher noise. Often, textures and fine details are flattened out in the name of a "clean" image. In other cases, noise may express itself as low frequency blotches in both luma and chroma or even unpleasant graininess throughout as if it was a part of the image.

Unlike the traditional signal-to-noise ratio (SNR) [6], a visual noise approach as specified in ISO 15739 better represents the way a human being perceives noise in the camera under test. Visual noise in the lightness channel L* and the two chrominance channels a* and b* are determined for all patches of the color checker and then weighted through a contrast sensitivity function (CSF) to mimic human perception for the 100% viewing on a 140 ppi monitor at 50 cm distance.

One change from ISO 15739 is to adopt a CSF that is specific to video, rather than still images, since the temporal aspect of noise in video leads to a difference in noise perception compared to single images [4]. The aspect that changes is the shape of the contrast sensitivity function (Figures 17 and 18).

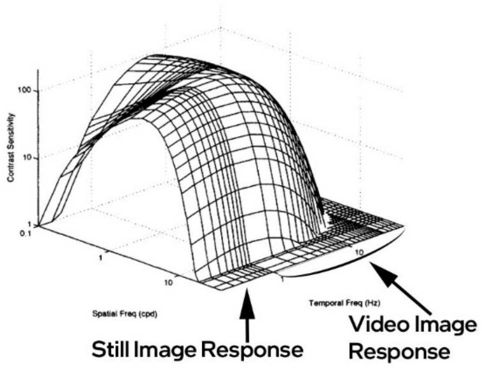


Figure 17: Luminance Contrast Sensitivity Function

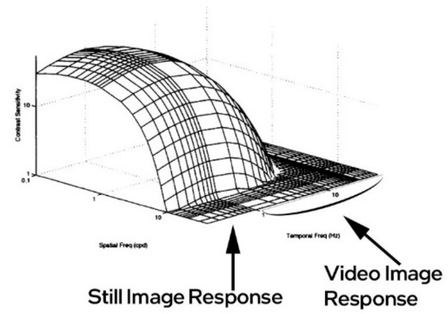


Figure 18: Chroma Contrast Sensitivity Function

The test procedure requires setting the distance between the camera and the OECF chart such that the field of view (FOV) is aligned to the OECF chart flags for the intended resolution. A five second video shall be captured following the conditions in Table 4. A frame captured from the last second of the video shall be used for noise evaluation. The resulting KPI corresponds to the "set 3" visual noise value as reported in the OECF result file from the analysis software described in Appendix C.

Test	Name	Standard Illuminant	Lux level	Comment
8.1.1, 8.1.2	Low light	A	20 lux	
8.1.1, 8.1.2	Mid light	CW (led)	80 lux	
8.1.1, 8.1.2	Mid light	CW (led)	250 lux	
8.1.1, 8.1.2	Bright light	D65	500 lux	

Table 4: Visual Noise test conditions

8.2 Contrast, Dynamic Range, Exposure

Exposure, gamma correction, and tone mapping have a dramatic effect on brightness, dynamic range, and an overall usefulness of an image.

8.2.1 Contrast Response

Test shall be conducted utilizing a transmissive 20-patch OECF chart, with transmissive light source adjusted to the conditions in Table 5. See Appendix C for specific equipment recommendations.

The device is set such that the FOV is aligned orthogonally to the OECF chart flags for the intended resolution. Exposure should be left in auto mode and a five second video is captured for each lighting condition in table 5.

The available image data is transferred to the CIE-Lab color space and the mean L* value per patch is calculated. A target L* value is calculated based on the reference transmission value of each patch of the OECF target. L* difference is calculated for all adjacent patches within the test image then compared to the target L* difference values. The resulting KPI is derived from the mean ratio of measured and target ΔL values in the three image intensity regions, namely low ((10 < L* < 30,), mid (40 < L*< 70) and high (70 < L* < 90).
The higher the metric, the more the local contrast can be perceived in the specified region. While some contrast enhancement is beneficial for the subjective contrast impression, a very high value will lead to an unpleasant effect for the observer. KPIs are compared to upper and lower limits to allow room for preferential tuning, while ensuring extreme contrast that is unpleasant to the viewer will negatively impact the VCX score.

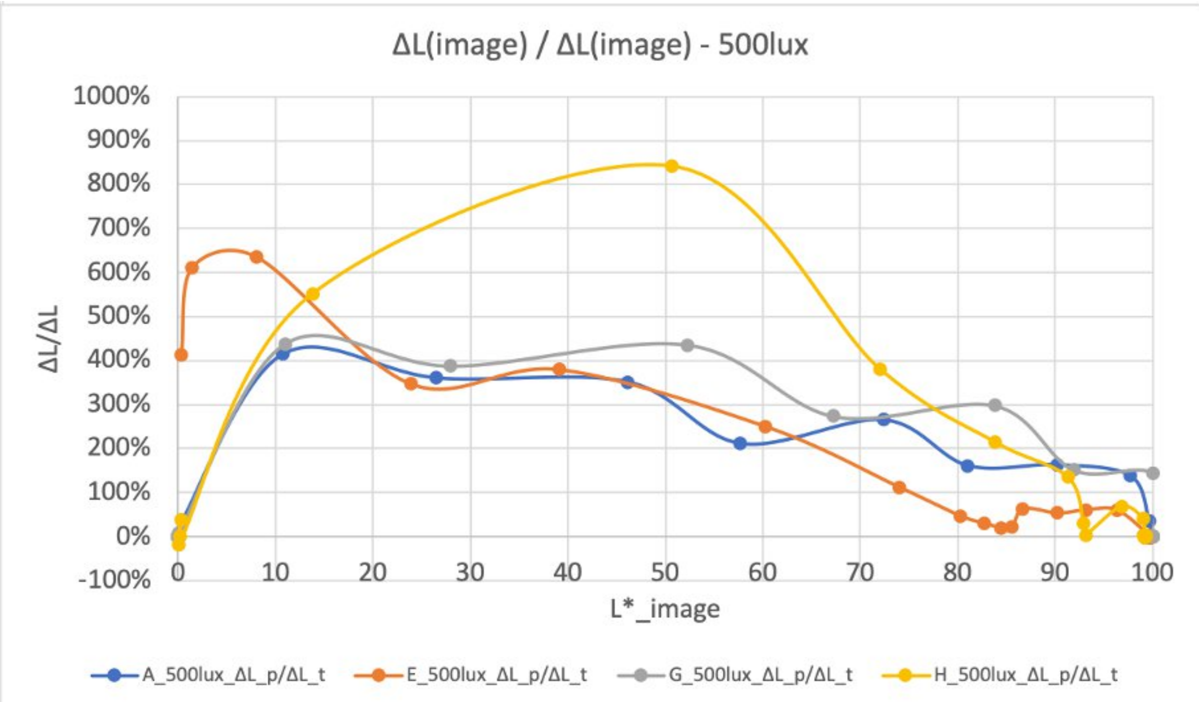


Figure 19: Sample data from different devices - device E showing high contrast in low region, device H showing very high contrast in med region. Overly aggressive contrast for device H would have a strong impact on the VCX score.

Test	Brightness	Light Source
8.2.1, 8.2.2	20 lux	CW
8.2.1, 8.2.2	80 lux	
8.2.1, 8.2.2	250 lux	
8.2.1, 8.2.2	500 lux	

Table 5: Contrast response test conditions

8.2.2 Dynamic Range

Test uses the same image(s) as in the Contrast Response test. The dynamic range shall be calculated according to ISO15739:2013 as the ratio of L_max and L_min. The value of L_max shall be defined as saturation point while L_min refers to the luminance that leads to an SNR of 3. Reporting shall be done in f-stop units.

8.2.3 Exposure Accuracy

Exposure accuracy is evaluated based on the lightness value L* of patch 21 (18% gray) - in the classic color checker (24 patches, included in webcam studio chart) according to ISO 11664-2:2020. Test shall be conducted for the lighting conditions in Table 6 unless otherwise specified using the following KPIs:

- Convergence time – the time in ms from scene change to L* convergence of patch 21 or mannequin ROI.
- Convergence target – L* value of the 18% gray patch falls within spec limits.
- Convergence stability – the standard deviation of the L* value of the patch 21 after convergence.
- Static L* of patch 21 ColorChecker
- Head-turn - face L* target and patch 21 standard deviation across short and long head-turn sequences

Each test condition is evaluated for these three targets:

Webcam studio chart with light skin tone mannequin (“Alexis”)
Webcam studio chart with dark skin tone mannequin (“Richard”)
Webcam studio chart with no face present (mannequin removed)

Test	Name	Standard Illuminant	Lux level
8.2.3.1, 8.2.3.3	Low light	A	10 lux
8.2.3.1, 8.2.3.3	Mid Light	A	80 lux
8.2.3.1, 8.2.3.3		CW (led)	250 lux
8.2.3.1, 8.2.3.3	Bright Light	D65	500 lux
8.2.3.1, 8.2.3.3	HDR	Television: Front light (D65):	700 lux effective (X nits) 80 lux on face
8.2.3.2	AE Sequence	CW (led)	500, 10, 80,20, 250, 500, 0 lux; 10 seconds each

Table 6: Exposure test conditions - See Table 2 for full details

8.2.3.1 Exposure Behaviour in Static Scenes

The device shall be tested at a distance of 75 cm from the target. For every condition in Table 6, five-second videos shall be captured for every condition in Table 6 to allow for convergence of 3A and ISP algorithms. Frames shall be extracted between the 4th and 5th second of the video for analysis; these frames shall be re-used in later tests.

8.2.3.2 Sequence Auto Exposure Convergence Test Sequence

The device shall be tested at 75 cm from the target by following these steps:

- Set scene illumination to 500 lux and CW. Allow to stabilize for 30 seconds to warm up the lighting system, then turn off.
- Set scene illumination to 0 lux (dark room).
- Start and stop the camera after several seconds to reset any kind of "history" data that might be saved.
- Start camera and begin video capture.
- With as close to 2 seconds delay as possible, start the following sequence. This is to introduce ~ 2 sec 0 lux condition at the start of the recorded video.
    - 500 lux, 10 seconds
    - 10 lux, 10 seconds
    - 80 lux, 10 seconds
    - 20 lux, 10 seconds
    - 250 lux, 10 seconds
    - 500 lux, 10 seconds
    - 0 lux, 2+ seconds
- Stop video capture.

L* component according to ISO CIE 11664-1 for patch 21 (18% gray) in the color checker is extracted for each frame in the test video to calculate the KPIs described in Section 8.2.3. Color encoding of the video is assumed to follow ITU-R Recommendation BT.709 unless otherwise indicated and converted into CIELAB from there.

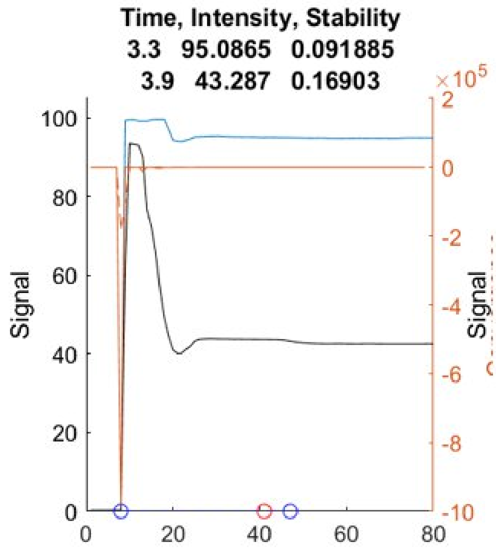


Figure 20: Snapshot of video showing ROI being evaluated, example of AE convergence plot

8.2.3.3 Exposure Behaviour in Moving Scenes

The objective of these tests is to ensure exposure stability and reliability with repeated convergence and moving subjects. Exposure shall be tested in the following moving scenarios using KPIs from Section 8.2.3, the lighting conditions summarized in Table 6 above, and at the distance of 75 cm from the target

- Face turn away, return in less than 5 seconds, AE should stay stable (converged to face-present values). Test shall be conducted for both LDR and HDR background.
- Face turn away, return after 5 or more seconds, AE should start converging or completely converge to non-face detected scene. When face returns AE should reconverge to face value within 5% of original face convergence value (mean L* of the face region). Test shall be conducted for both LDR and HDR background.
- With face stable in foreground at object distance 75 cm, introduce peripheral motion via video on the TV behind face. Motion should fill outer 25% of FOV and include people walking back and forth plus random motion.

8.3 Spatial Frequency Response

Spatial frequency response (SFR) includes all measurements (resolution, texture loss, and sharpening) related to the capability of the camera to reproduce details in the scene. It is important to note that the term SFR is more general than the modulation transfer function (MTF), which is a special version of SFR. In this context, only SFR is used here even though the s-SFR (SFR measured on the sinusoidal Siemens star) could also be called MTF. Due to the established terminology, the derived KPI from the SFR is referred to as MTF.

Resolution describes the capability of the system under testing to reproduce details in the scene. The level of details is measured as well as the acutance - a metric that has shown good correlation with human perception of sharpness of a displayed or printed image.

Texture loss reflects how well contrast and fine details are reproduced. A strong texture loss, mainly due to strong noise reduction and/or data compression, results in aquarelle like images with reduced or completely lost natural variations in texture regions, such as skin, hair, fabric, etc.

Sharpness impression of an image can be improved by applying edge sharpening algorithms. While such algorithms can substantially enhance images, too aggressive processing can result in various artifacts and reduced image quality.

Due to the adaptive and nonlinear behavior of image enhancement algorithms, different measurement configurations will lead to different SFRs, each defining unique KPIs which describe certain aspects of image quality.

The test procedure requires aligning the device such that it is orthogonal to the scene, centered on patch 2 of the color checker chart, and positioned 75 cm away from the chart. A five-second video shall be captured for each lighting condition in Table 7 below and relevant metrics calculated using the last captured frame.

Test	Name	Standard Illuminant	Lux Level	Reference CCT
8.3.1, 8.3.2, 8.3.3	Low light	A	20 lux	2856K
8.3.1, 8.3.2, 8.3.3	Mid Light	CW (led)	250 lux	4000K
8.3.1, 8.3.2, 8.3.3	Bright Light	D65	500 lux	6504K

Table 7 – Overview of light conditions for spatial frequency response measurement

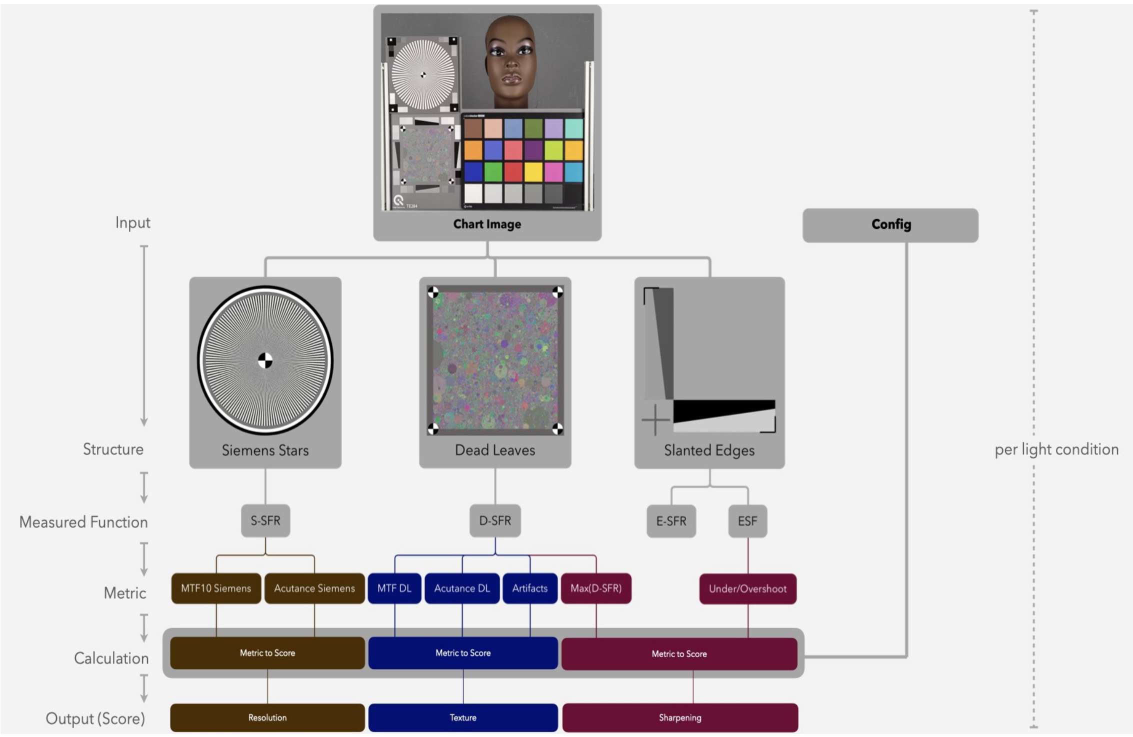


Figure 21: Representation of spatial frequency response measurement

8.3.1 Resolution

The resolution score is defined by the $MTF25_{siemens}$ describing the level of details that can be captured by the device and $Acutance_{siemens}$ reflecting the subjective impression of sharpness, both derived from the s-SFR (spatial frequency response measured on the Siemens star) according to ISO12233. The s-SFR method is known for its robustness against image enhancement algorithms, as opposed to the e-SFR method based on slanted edges which can be easily modified by such algorithms.

MTF25 is the spatial frequency that leads to an SFR of 25%. Acutance is calculated by employing the contrast sensitivity function (CSF) to model relevant aspects of human perception by weighting different spatial frequencies according to their contribution to the human perception. The required definition of viewing condition is the same as used for visual noise in Section 8.1.2.

8.3.2 Texture Loss

A colored dead leaves chart is used to measure texture loss. This target consists of many circles with random color, intensity, and size stacked on top of each other and is much harder to reproduce for image processing algorithms compared to some other test charts with edge targets.

The score for texture loss is calculated based on $MTF25_{DL}$ using the $DeadLeaves_{cross}$ method defined in ISO19567-2. The final texture value is a sum of these parameters and a ratio of the resolution and noise values, which reflects the complex relationship between noise reduction strength and texture preservation. See Appendix A for calculation details.

8.3.3 Sharpening

Sharpening is defined by the undershoot and overshoot criteria (Figure 19) based on the edge spread function (ESF) on slanted edges (ISO12233), as well as the maximum response in the SFR_D. and the measured nyquist value from the siemens star. To increase the subjective impression sharpness of an image, image enhancement algorithms will increase the local contrast along edges. While a gentle amount of sharpening is beneficial for the subjective image quality, excessive sharpening will lead to strong and disturbing artifacts in the image.

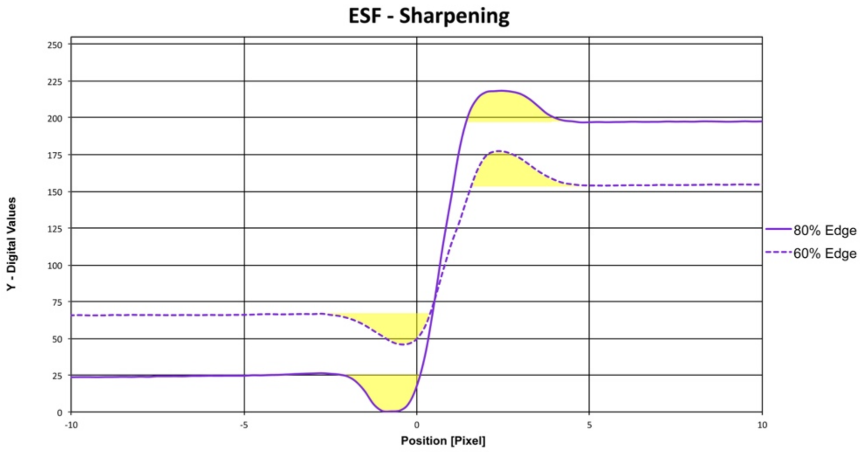


Figure 22: Undershoot and Overshoot in the ESF (yellow area).

8.4 Color Reproduction

Color reproduction quality is expressed as a difference between the target values and the corresponding color measurements collected from the captured video.

8.4.1 Color Accuracy

Both mean and maximum Delta E (CIE Lab color difference), Delta C (chroma difference), and Delta H (hue difference) values according to ISO17321 shall be determined for the color checker chart. Separate measurements shall be collected for the face ROI of the mannequin for both skin tones. The device shall be set at 75 cm from the target and a five-second video shall be captured for all lighting conditions specified in Table 2 and three tests targets listed below:

- Webcam studio chart with light skin-tone mannequin(“Alexis”)
- Webcam studio chart with dark skin tone mannequin (“Richard”)
- Webcam studio chart with no face present (mannequin removed)

8.4.2 White Balance

White balance quality is determined by the Delta C_{2000} for patch 21 of the color checker. For face-present scenes, Delta C_{2000} of the whole-face ROI shall be evaluated. While this test shall reuse the images from the color accuracy test, new captures are collected to evaluate the impact of background variation on white balance for each test combination listed in Table 8 below, using a TV setup. The color difference values for patch 21 in the ColorChecker are calculated and the mean and maximum values are reported for each of the background conditions.

Lighting Condition	**TV setting**
CW, 20, 80, 250lux	Sky
A, 20, 80, 250lux	Grass
D65, 250lux	Stones
	Red
	Yellow
	Green
	Blue
	18% Gray
	Coffee Shop
	Moving Heads (includes both AE and AWB test)

Table 8: Background variation test settings

8.4.3 Auto White Balance Convergence

C* component, according to ISO CIE 11664-1 for patch 21 (18% gray) in the color checker, is extracted for each frame in the test video to calculate the following KPIs (in each case, the maximum value shall be reported):

- AWB convergence time – the time in ms from the scene change to convergence of the C* signal in patch 21 or mannequin ROI.
- Convergence target – C* value of the 18% gray patch should fall within spec limits.
- AWB convergence stability – defined as the standard deviation of the C* value of the 18% gray patch for the remaining number of frames after convergence
- Static DeltaC - patch 21 of Color Checker

The device shall be tested at 75 cm from the target by following these steps:

- Ensure lighting system is warmed up for all CCTs.
- Set scene illumination to 0 lux (dark room).
- Start and stop the camera after several seconds to reset any kind of "history" data that might be saved.
- Begin video capture at 0 lux.
- Within less than 2 seconds, start the following sequence
    - 250 lux A, 10 seconds
    - 250 lux D65, 10 seconds
    - 250 lux WW, 10 seconds
    - 250 lux CW, 10 seconds
    - 250 lux D50, 10 seconds
    - 20 lux A, 10 seconds
    - 80 lux CW, 10 seconds
    - 250 lux D65, 10 seconds
- Stop video capture

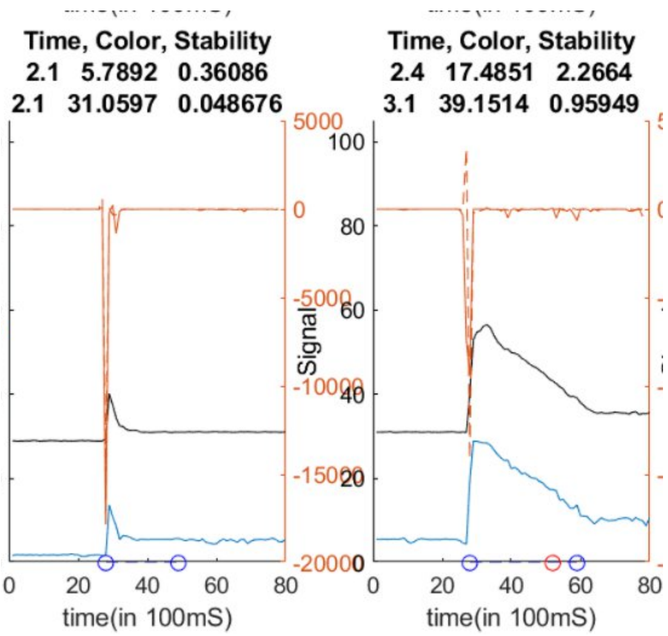


Figure 23: Example of a good and bad AWB convergence plot

8.5 Image Uniformity

There are many factors, such as poor optics, thermal effects, light leaks, sensor calibration, and image processing, which can have a dramatic effect on spatial uniformity of an image. This test evaluates spatial uniformity in both luminance and color across a range of common illuminants. Results from the worst performing illuminant will be reported as the test KPI.

Test shall be conducted utilizing a uniform field test chart with transmissive light source adjusted to the conditions in Table 9. The device shall be aligned orthogonally to the chart, such that the illuminated area fills the entire field of view of the camera. The device display and any LEDs should be turned off or covered to avoid any reflected light in the image. A five-second video shall be captured for each lighting condition; 5 frames from the last second of the video will be averaged for analysis. Luminance uniformity is calculated as 1-max/min using the maximum and minimum CIE-L* found in the image. Color uniformity is expressed as maximum ΔEab over the image field. See Appendix C for specific equipment recommendations for this test.

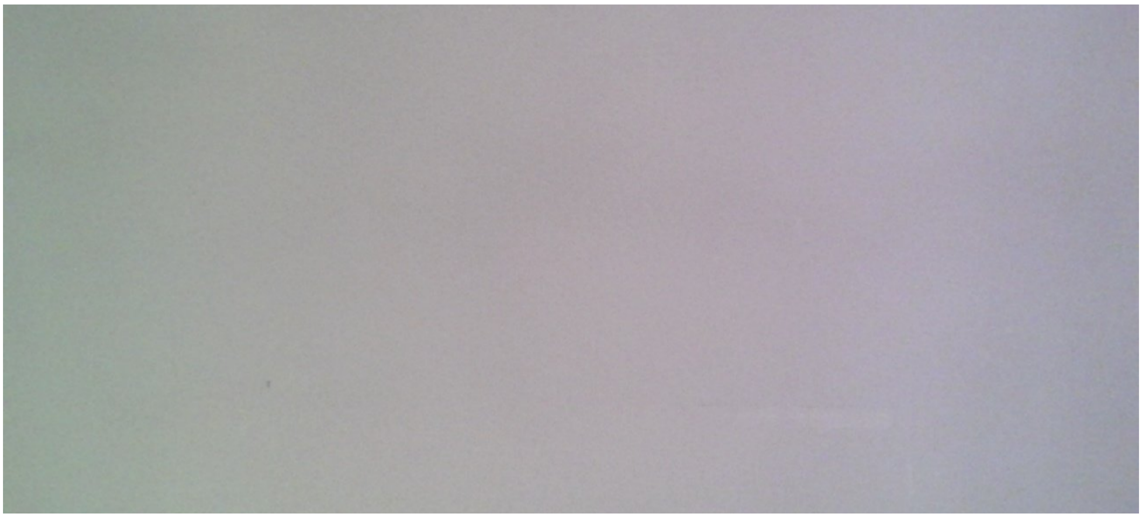


Figure 24: Example of strong spatial nonuniformity in both luminance and color

A	IQ: 250 lux
WW	
CW	
D50	
D65	

Table 9: Conditions for image uniformity test

8.6 Frame Rate

Device shall be tested 75 cm from the target. Frame rate is calculated by measuring true exposure with an LED panel positioned in the VCX PC test scene mannequin slot. This approach allows to check for differences between the number of frames shown and the number of frames captured. Theoretically, a device could capture 15 frames per second at low light and use the same frame twice to still provide 30 frames per second in the video stream.

The LED panel shall be configured for at least 1 ms interval. In the captured 5 second video, the last 30 frames will be analyzed to determine the time between the exposure start of each frame and the mean value shall be reported as the resulting frame rate. This test shall be performed at 250 lux, 80 lux, 20 lux - all D50. See Appendix C for equipment recommendations.

Note: A good LED panel frequency setting can be determined as frame rate times the number of LEDs (e.g., 3 kHz for 30 fps). This will in most cases result in the same LEDs to be illuminated in each frame, making analysis easier.

Figure 25: Led panel within test scene

____

9 JOB AIDE

A Job Aide [8] is available in the spec package showing reference tables for conditions for each test outlined in this document. VCX labs will continually update this document and the latest document will be available on the VCX-forum website.

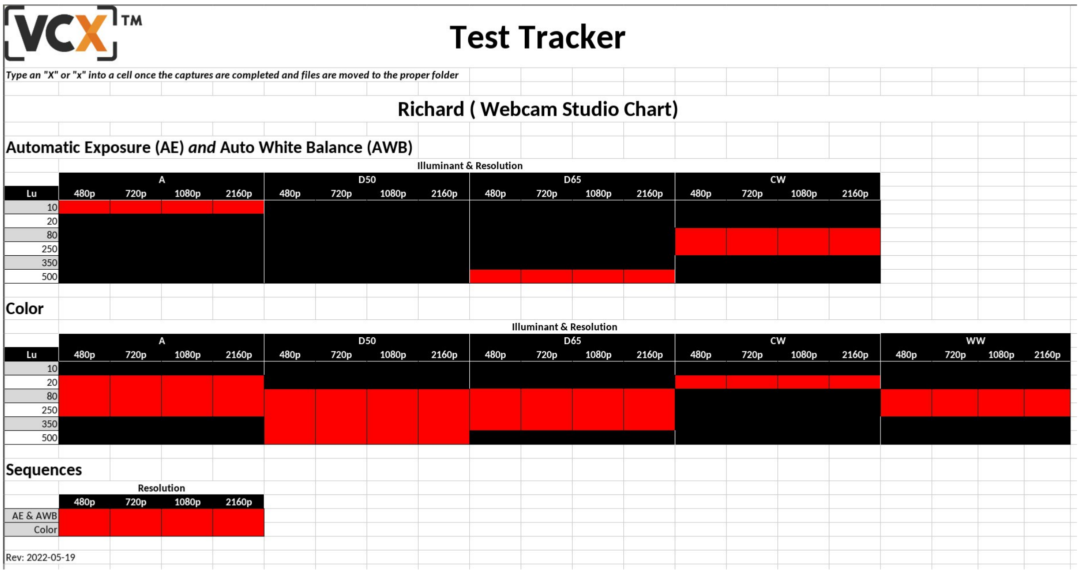


Figure 26: Screenshot from most recent Job Aide

10 TABLE OF DEFINITIONS

Below are definitions of some core terms used in this document.

Acutance	A measure of the sharpness of an edge in a displayed or printed image, often used as a measure of how well detail is preserved.
Contrast Curve	A graphical depiction of the relationship between the amount of exposure given and its density after processing an image.
Color Checker	A color calibration target consisting of a cardboard-framed arrangement of 24 colored squares for the "classic" chart.
Correlated color temperature (CCT)	A specification of the color appearance of the light emitted by a light source, relating its color to that of the reference light source when heated to a particular temperature measured in Kelvin (K) units.
DUT	Device under test.
Frame rate	The frequency at which consecutive images are displayed.
Jitter	A measurement of video frame rate consistency.
Orthogonal	A direction to align the camera centered at a right angle to the test chart.
Resolution	Often referred to as the number of distinct pixels that can be displayed in each dimension of an image. The VCX webcam specification differentiates between pixel count and resolution, described here as the level of details the camera can reproduce after considering the optical system and any processing that is applied to the image.
Sharpening	Sharpening enhances edges in the image by increasing the local contrast along the edge.
Spatial Frequency	The number of repeating elements in a pattern per unit distance.
Texture loss	The loss of low-contrast fine details in an image, often due to noise reduction.
White Balance	Color processing applied to prevent unnatural color cast of an image and establish the true color of white.

11 APPENDIX A: KPI Summary andDefinitions

NOTE: The target values in the table are intended to assist in delivering image quality that will perform well against the VCX score algorithm. In most cases, the provided target value coincides with a KPI score >90. It is not expected that camera systems will be able to meet all target values. Where applicable, calculations are provided showing weighting of particular KPIs and how they are related to one another.

Category	KPI
Auto-Exposure	Convergence Time – the time in ms from initial detected scene change to L* convergence of patch 21or mannequin ROI. Recommended target: ≤3 seconds *Convergence is defined as the point at which the measured signal of the patch or ROI stops changing after major scene change is detected. This shall be measured by the VCX convergence test algorithm, integrated with VCX approved test analysis software (refer to Appendix C)*
	Convergence Target – Lightness value (L*) of the 18% gray patch falls within the targets defined in the VCX score. Recommended target ranges as follows: Alexis Mannequin: 65-75 across the mannequin face Richard Mannequin: 30-45 across the mannequin face Patch 21 (with no face present): 80-85
	Convergence Stability – expressed as the standard deviation of the L* value of the test ROI after convergence is achieved. Recommended target: ≤0.1
	Patch 21 L* - refers to the lightness value of patch 21 in the color checker chart of the webcam studio scene Alexis Mannequin: 80-85 across the mannequin face Richard Mannequin: 78-82 across the mannequin face. The target range for this KPI is narrower than the other two scenes, recommended to target as close to 80 as possible for the reference mannequin. Patch 21 (with no face present): 80-85
	Head-turn - scene consistency KPI measured by the L* target and standard deviation value of patch 21 of the color checker,and face exposure KPI defined as the mean L* value of the face ROI of the mannequin. Short head-turn: L* targets for Patch 21 L* above. L* stability same as convergence stability. Long head-turn: Ratio of L* targets per Patch 21 L* above before and after head turn. Ideal target ratio of 1.
Contrast Response Curve and Dynamic Range	Contrast Response - L* target low, medium and high target windows, measured by contrast patches in the OECF chart as defined by (deltaL_image/deltaL_target)/L*_Image

	Dynamic Range: Ratio of L_max and Lin_min per ISO 15739:2013 in unit f-stop.
Frame Rate	Frame rate: True exposure measured via an LED panel in the scene set to 1mS sample time. The result is 1/((mean LED count -1) of 5 sequential frames, measured individually). The 5 frames come from the last second of a 5 second video sample. Recommended targets: 20lux, ≥15fps 80lux, ≥30fps
Visual Noise	Visual Noise: Measure of the "visibility" of noise following methodologies from ISO 15739, with a human vision contrast sensitivity function that takes into account temporal change. Reported as visual noise set 3 in the analysis software XML. Reported KPI is the weighted sum of the mean and max visual noise values. Recommended ideal targets: Mean: <2 Max: <2.4 Calculation: $((VN3_{mean} * 10) * 0.4) + ((VN3_{max} * 10) * 0.6) < 20$.
Resolution, Sharpness, & Texture	Resolution: Measure of how well a camera system reproduces both low and high frequency edges. Measured on the siemens star of the webcam studio chart. $(MTF10_{siemens3} * 0.5)+(10*(Acutance_{siemens3} * 0.5)) \geq 500$
	Sharpness: Measure of the likelihood of sharpening artifacts in the image. Defects such as overshoot/undershoot and jagged edges will show up in this KPI. Undershoot <5 Overshoot < 5 Max(D)SFR < 1 Nyquist >700 Weighting: (US * 0.2) + (OS * 0.5) + Max(D)SFR * 0.1) + (Nyquist * 0.2)
	Texture: Measure of how well a camera system reproduces fine detail and texture. Measured on the dead leaves portion of the webcam studio chart per ISO 19567_2. $(MTF10_{DL3} * 0.5)+(10*(Acutance_{DL3} * 0.5)) \geq 600$
Color Accuracy	Mean and max Delta E (CIE LAB color difference) of the colorchecker chart. dE target: 30
	Mean Delta C (chroma difference) of the colorchecker chart dC target: <10

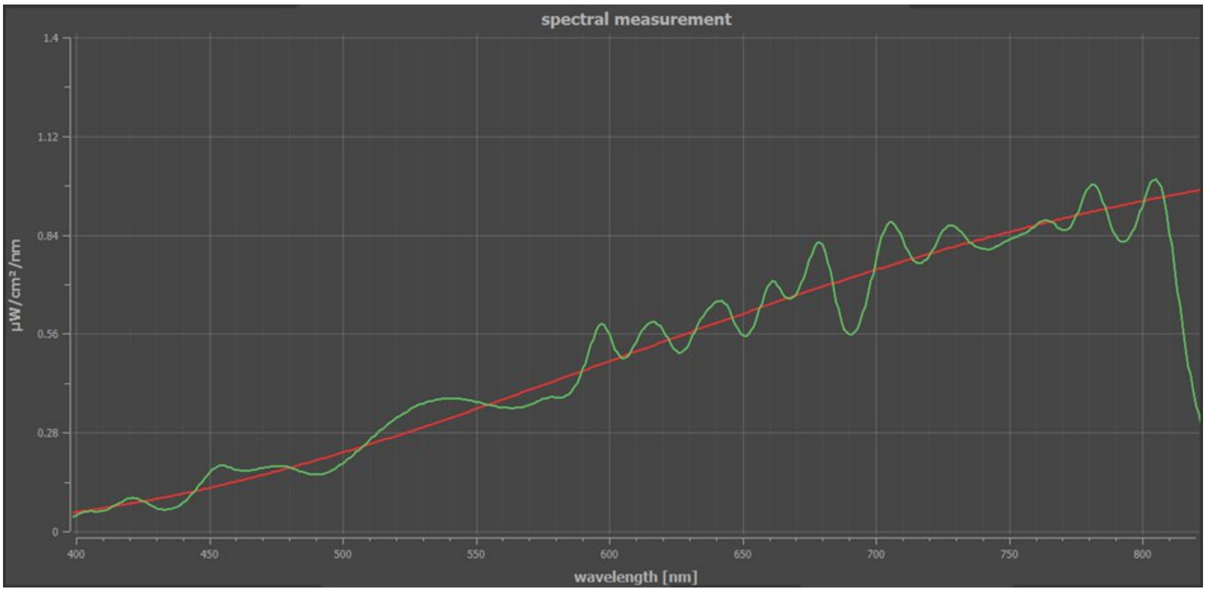
	Mean Delta H (hue difference) of the colorchecker chart dH target:
	C* value of the face ROI Alexis: >-1, < 6 Richard: +/-15
	H* value of the face ROI Alexis: +/- 10 Richard: >8, <45
Auto White Balance	C* for patch 21 of the color checker for static scenes. C* Target < 10
Auto White Balance	AWB Convergence Time –The time in ms from initial detected scene change to C* convergence of patch 21or mannequin ROI. Recommended target: ≤3 seconds *Convergence is defined as the point at which the measured signal of the patch or ROI stops changing after major scene change is detected. This shall be measured by the VCX convergence test algorithm, integrated with VCX approved test analysis software*
Auto White Balance	Convergence Target – Lightness value (C*) of the 18% gray patch falls within the targets defined in the VCX score. Recommended target ranges as follows: Alexis Mannequin: 28-32 across the mannequin face Richard Mannequin: 22-28 across the mannequin face Patch 21 (with no face present): ≤4
Auto White Balance	Convergence Stability – the standard deviation of the C* value of the test ROI after convergence. Face C* stability target range: 0-0.5. Patch 21 (no face present) stability target range: 0-0.3
Image Uniformity	Luminance uniformity reported as worst case (min/max) Target: ≤28% non-uniformity
Image Uniformity	Color uniformity measured as max delta E_{ab} Target: ≤**0.24**

12 APPENDIX B: Light Spectra References

Reference spectrum measured at 1M with an Avantes AvaSpec Mini2048CL-IMG11 spectrometer. The red line is the target spectrum (see spectrum folder for files), the green line is the output of the light source used for this measurement.

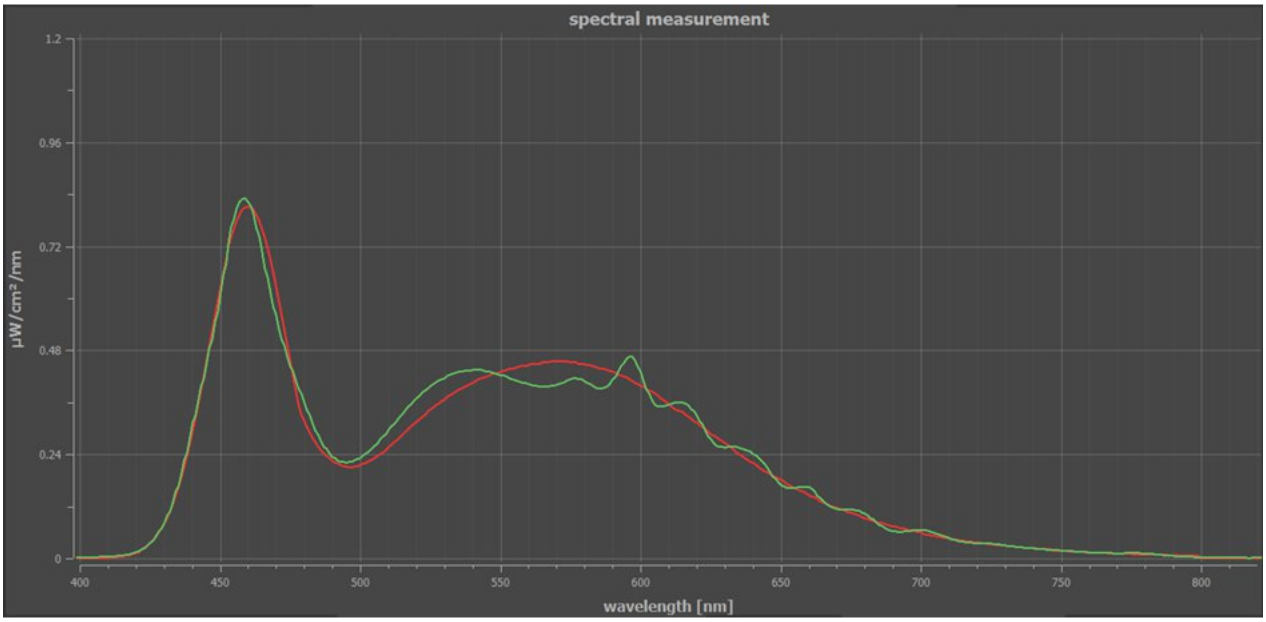
12.1 Illuminant A

Defined per ISO 11664-2:2020



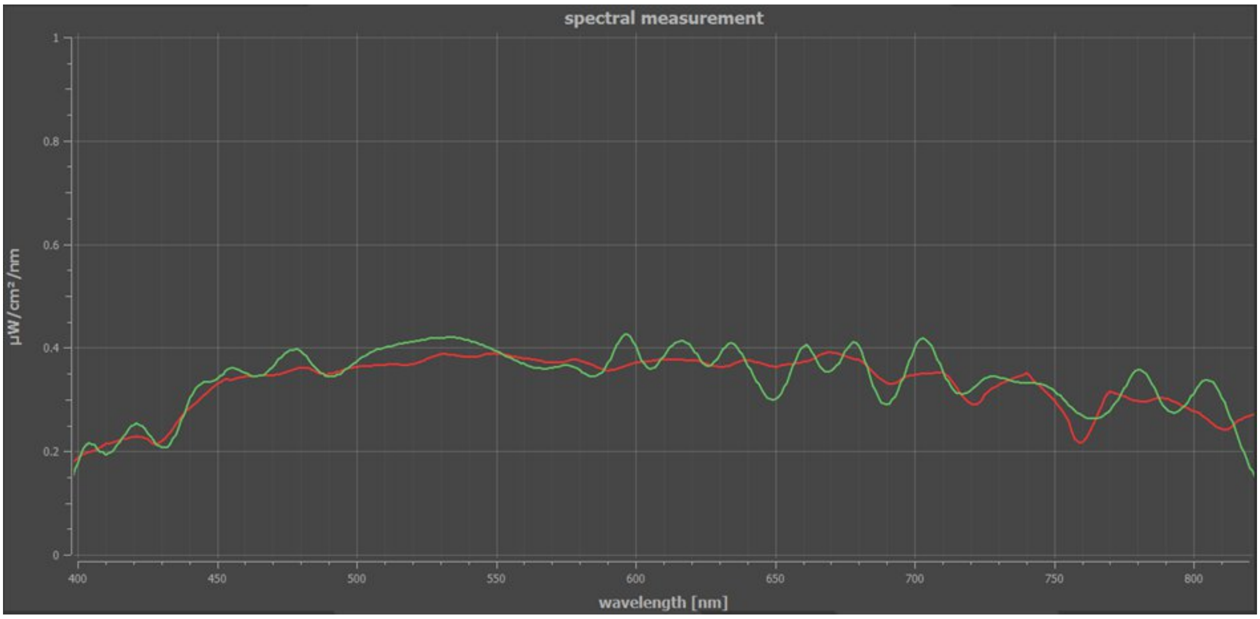
12.2 Cool White – LED at 500 lux

VCX-provided reference. No accepted international standard at this time.



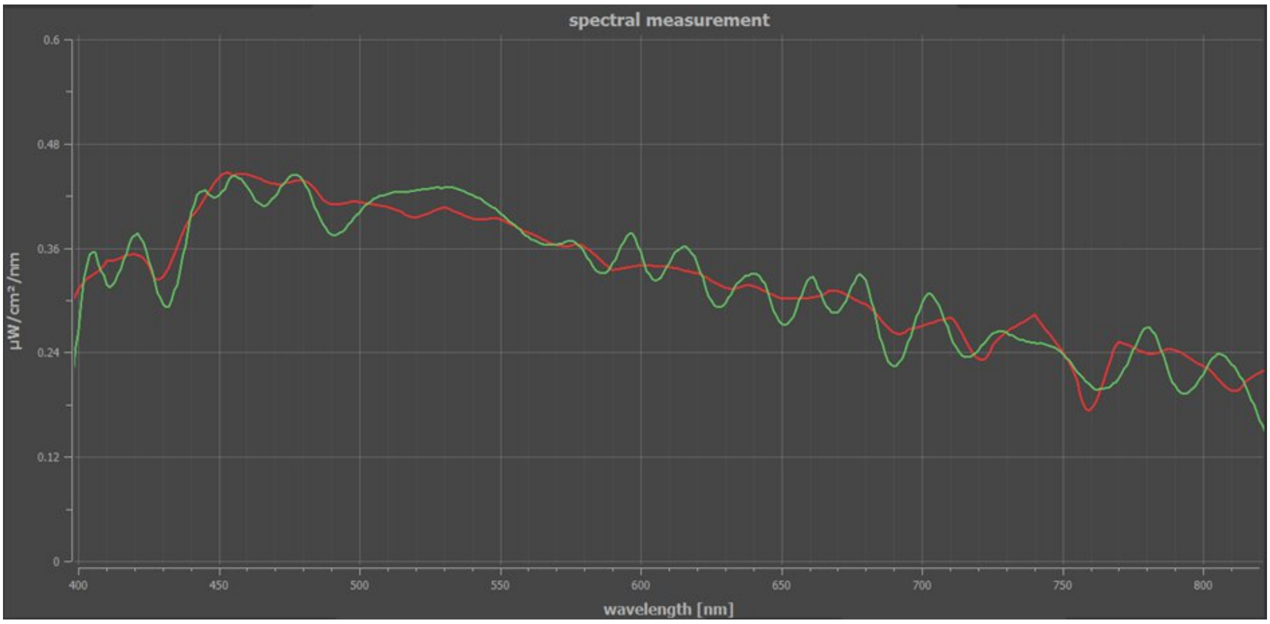
12.3 D50 at 500 lux

Defined per ISO 11664-2:2020



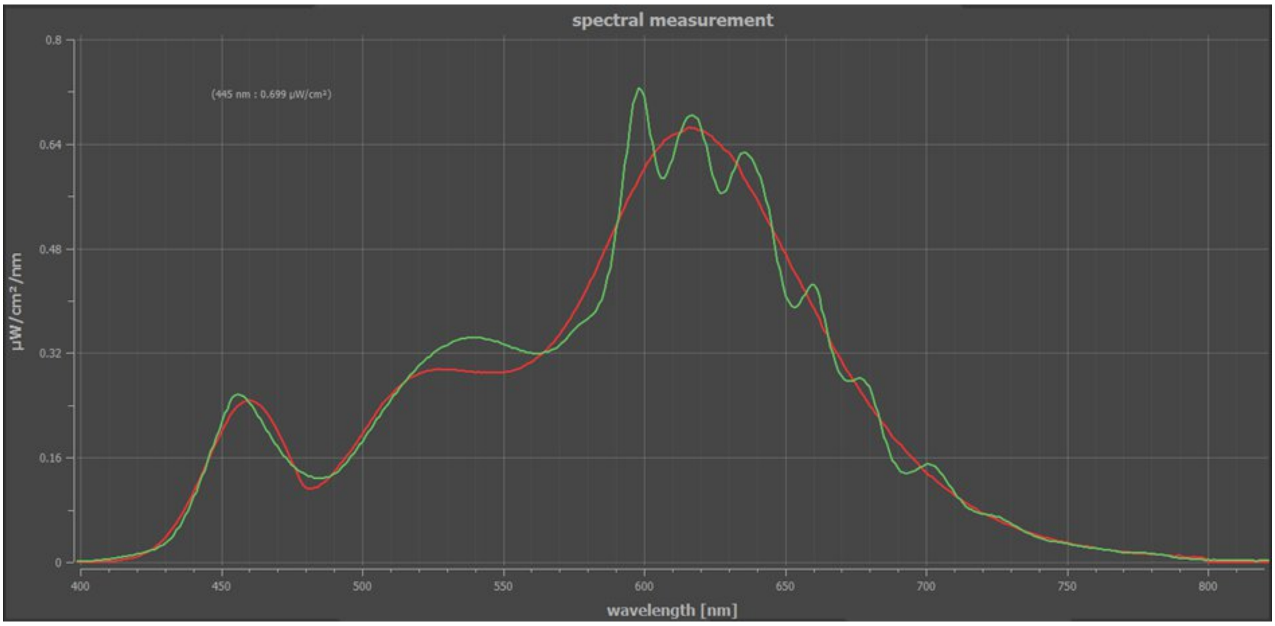
12.4 D65 at 500 lux

Defined per ISO 11664-2:2020



12.5 Warm White – LED at 500 lux

VCX-provided reference. No accepted international standard at this time.


spectral measurement
0.8
0.64
0.48
0.32
0.16
0
(445 nm : 0.699 µW/cm²)
µW/cm²/nm
400
450
500
550
600
650
700
750
800
wavelength [nm]

13 APPENDIX C: HW / SW Requirements and Recommendations

VCX will update this table with additional equipment and software recommendations as they become available and qualify to meet the VCX spec requirements.

Type	Recommended/ Required	Item	Q'ty
Test Charts	Required	Webcam studio chart	1
	Recommended	Image Engineering TE295 test scene	1
	Required	Uniform field diffuser for integrating sphere	1
	Recommended	Image Engineering TE255-D280 diffusor for LE-7	1
	Required	VCX-compliant 20 patch OECF chart	1
	Recommended	Image Engineering TE264X-V2-D280, Contrast 100,000	1
	Required	VCX-compliant LED panel	1
	Recommended	Image Engineering LED Panel	1
	Required	Mannequin Heads Required: Richard head – Ebony with mounting rod, *(Dapper Cadaver model 8007-EB*) Alexis head - Ivory with Blonde Human Hair Wig, open eyes with mounting rod, (*Dapper Cadaver model 8008-IV*) Optional: David head in Olive with Human Hair Beard and mounting rod, (*Dapper Cadaver model 8009-OV*) ¶ (David head not required, but recommended for mid-tone testing)	1
Illumination and Measurement Devices	Required	Light source that meets illuminant and lux requirements in table 1 as well as uniformity requirements in Section 3.3.1 (Lab setup)	2
	Recommended	Image Engineering iQ-Flatlights-v2 with 400-1000nm support for illuminat A NIR spectrum	2
	Required	Uniform transmissive light source capable of up to 2000 lux capable of requirements in section 4.1.5 (Contrast Response) and 4.6 (Image Uniformity)	1
	Recommended	Image Engineering LE7 VIS-IR	1
	Required	TV capable of meeting requirements in Section 3.1	1

		(Television Background) with supporting hardware	
	Recommended	SONY X90J 75" television	1
	Required	Calibration device for TV	1
	Recommended	Calibrite ColorChecker Display	1
Camera Handling Devices	Required	Device handling solution capable of meeting requirements in Section 3.3.3 (Camera Handling Equipment)	1
	Recommended	iQ-Bench-M + 1x extension (total length: 2 meters)	1
		+ iQ-Cameramount with Manfrotto 405b tripod head	1
	Required	Mannequin mounting solution capable of meeting requirements in Section 3.3.5.	1
Image Review and Analysis Software	Required	VCX compliant evaluation software	1
	Recommended	iQ-Analyzer- All modules - used for generating KPI data from images	1
	Recommended	Faststone Image Viewer or VLC Player (used for viewing images and videos)	1

14 APPENDIX D: References

14.1 Document References

[1] Video calling and video chat | Pew Research Center

[2] https://backlinko.com/zoom-users

[3] https://blog.tmcnet.com/blog/rich-tehrani/wp-content/uploads/2019/09/2019-Impact-of-Video-Conferencing-Report-Lifesize_FINAL.pdf

[4] Xin Tong et al. (1999) in Video Quality Evaluation Using ST-CIELAB (Part of the lS&T/SPIE Conference on Human Vision and Electronic Imaging IV • San Jose, California • January 1999 SPIE Vol. 3644

[5] Artmann, U. (2015) Image quality assessment using the dead leaves target: experience with the latest approach and further investigations. *Electronic Imaging Conference, Digital Photography XI. 9404.* San Francisco: SPIE. Details: EIC2015_9404-18.pdf (image-engineering.de)

[6] FFmpeg - a complete, cross-platform solution to record, convert and stream audio and video (ffmpeg.org)

[7] Orchard, A. (2022) VCX-Webcam 2022 Expert Subjective Study. Report available to VCX members.

[8] Job Aide: VCX Job Aide 2022.xlsx - Included in VCX-WebCam 2023 release package

14.2 References for KPIs used in spec

FPN, Visual noise, various ISO, CIE specs

ISO 11664-2:2020
ISO15739:2013
ISO12233:2017
ISO CIE 11664-1
ISO19567-2
Video Quality Evaluation Using ST-CIELAB (Part of the lS&T/SPIE Conference on Human Vision and Electronic Imaging IV • San Jose, California • January 1999 SPIE Vol. 3644).
Placeholder for visual noise white paper or other reference, need to move data below to another doc. Currently “Visual and FPN noise reference_0.1.doc”

15 APPENDIX E: Document History

Version	Date (DD.MM.YYYY)	Editor	Comment
1.0	11.10.2022	Anthony Orchard	VCX WebCam 2023 first release.