Image quality evaluation using moving targets

Uwe Artmann[a]

[a]Image Engineering GmbH & Co KG, Augustinusstrasse 9d, 50226 Frechen, Germany

ABSTRACT

The basic concept of testing a digital imaging device is to reproduce a known target and to analyze the resulting image. This semi-reference approach can be used for various different aspects of image quality. Each part of the imaging chain can have an influence on the results: lens, sensor, image processing and the target itself. The results are valid only for the complete system. If we want to test a single component, we have to make sure that we change only one and keep all others constant. When testing mobile imaging devices, we run into the problem that hardly anything can be manually controlled by the tester. Manual exposure control is not available for most devices, the focus cannot be influenced and hardly any settings for the image processing are available. Due to the limitations in the hardware, the image pipeline in the digital signal processor (DSP) of mobile imaging devices is a critical part of the image quality evaluation. The processing power of the DSPs allows sharpening, tonal correction and noise reduction to be non-linear and adaptive. This makes it very hard to describe the behavior for an objective image quality evaluation. The image quality is highly influenced by the signal processing for noise and resolution and the processing is the main reason for the loss of low contrast, fine details, the so called texture blur. We present our experience to describe the image processing in more detail. All standardized test methods use a defined chart and require, that the chart and the camera are not moved in any way during test. In this paper, we present our results investigating the influence of chart movement during the test. Different structures, optimized for different aspects of image quality evaluation, are moved with a defined speed during the capturing process. The chart movement will change the input for the signal processing depending on the speed of the target during the test. The basic theoretical changes in the image will be the introduction of motion blur. With the known speed and the measured exposure time, we can calculate the theoretical motion blur. We compare the theoretical influence of the motion blur with the measured results. We use different methods to evaluate image quality parameter vs. motion speed of the chart. Slanted edges are used to obtain a SFR and to check for image sharpening. The aspect of texture blur is measured using dead leaves structures. The theoretical and measured results are plotted against the speed of the chart and allow an insight into the behavior of the DSP.

Keywords: image quality evaluation, noise reduction, spatial frequency response, SFR, Dead Leaves, MTF, video, h264, mpeg2

1. INTRODUCTION

Image Engineering is a manufacturer of a huge variety of test equipment for imaging devices and runs a test lab for image quality evaluation. One very important rule in a test lab for digital cameras is, that the camera does not move during exposure and that the target itself does not move during exposure. This is true for most tests.

One exception: A lot of cameras and some mobile phones have an image stabilization system. So in this case, it is obvious that a test of the image stabilization system needs a test setup which simulates the hand-shake of the user. So the test target is fixed, the camera is shaken in a controlled way and the quality of the image stabilization can be obtained from the images captured.[1]

For an internal project, we developed a system which allows us to move a test target in a controlled way. The idea is to move an object in a test scene and to give subjective scores if the shown object can still be recognized. This is for special interest in security applications, so for example to check if the the surveillance system is suitable to capture useful images of a passing car or person.

This paper is about our experience we made while investigating the possibilities of such test setup for objective image quality evaluation.

Further author information: (Send correspondence to Uwe Artmann)
Uwe Artmann: E-mail: u.artmann@image-engineering.de, Telephone: +49 2234 9955950 www.image-engineering.de

2. SETUP

2.1 Hardware and Chart

The setup (Figure: 1) consists of a moveable chart mount in a test scene. The test scene it illuminated with a controllable light source. The chart mount can be moved from left to right in a defined speed. The speed is controlled in high accuracy and the construction is designed that way, that the speed is constant while the chart is visible (and not in the parking position left or right from the scene).

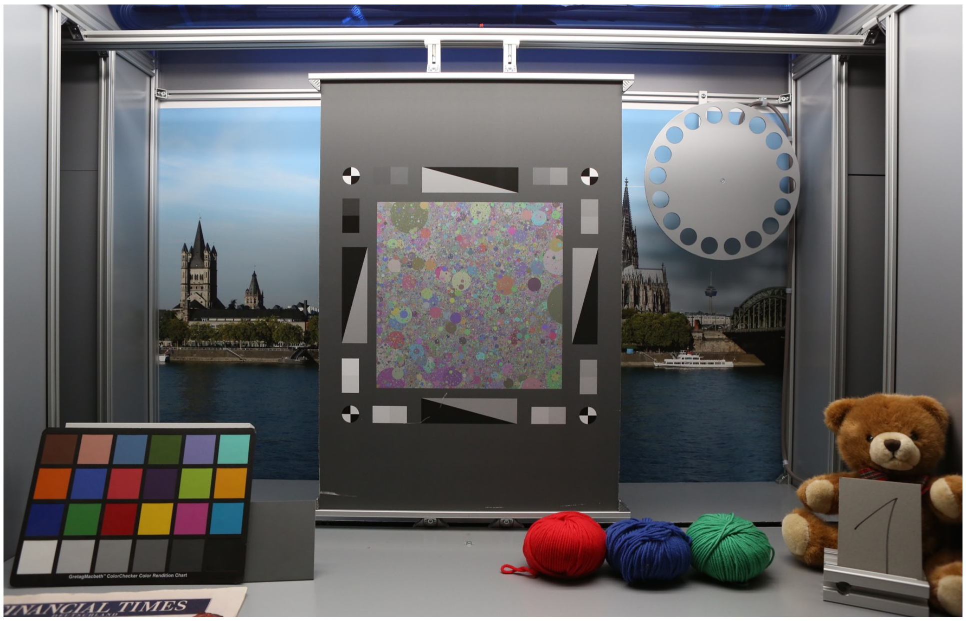


Figure 1. The hardware setup, showing the chart, mounting and surrounding

The device under test takes pictures or captures video files while the chart moves from left to right orthogonal to the optical axes. (Figure: 2) During our tests, we moved the chart with speed between 0.01 m/s up to 2m/s.

The used chart (Figure: 3) consists of different structures than can be used for analysis:

"Colored Dead Leaves " The "Dead Leaves" structure is used for some time in the area of objective image quality evaluation. First used in this context and presented by Cao et. al.,[2] it is used for the evaluation of the so called texture loss, which is the loss of low contrast, fine details due to noise reduction and adaptive filter. The first approach was adopted by other work groups and further developed. McElvain et. al.[3] presented a new approach to reduce the influence of added image noise and to reduce the contrast from the first presented chart design. In 2012 we presented the result of our research using the dead leaves structure and proposed a colored version of this chart.[5] This proposed structure is used in this paper.

"Slanted Edges " Slanted edges are a standard structure to evaluate the Spatial Frequency Response (SFR) of a camera system. The basic approach is described in ISO12233.[7] The contrast of the slanted edge can have an influence on the resulting SFR. We used a slanted edge with a 80% Modulation, which is a relatively high contrast. We decided to use this contrast as we focus to get information about image enhancement and sharpening rather than a lens resolution measurement. In modern camera systems, the signal processing is an important part of the signal chain and can highly influence the results. The higher the contrast, the higher the sharpening that is applied to the reproduction of this edge in the image[6] .

Figure 2. Sample of test procedure: images are taken with increasing chart speed. (Chart movement: *top*: 0 m/s, *center*: 0.15 m/s, *bottom*: 1 m/s)

"Neutral Area " Camera Noise that is added to the reproduction of the dead leaves structure has an influence on the measurement. To reduce this influence, the algorithm proposed by McElvain et.al.[3] uses a correction obtained from a neutral gray patch that has the same mean reflection than the used dead leaves structure. [*]

[*]The neutral area is not shown in Figure 3

"Gray Patches " Most cameras for photography or videography produce image data with a non-linear tonal response curve. As the analyzing algorithms assume a linear response, the image data has to be linearized. The gray patches are used to obtain an opto electronic conversion function (OECF), a function of the digital values against optical reflection of the patches. Using the OECF, an inverse look up table (LUT) can be created and applied to the image which results in a linear version of the original image content.

"Marker " The marker are used for registration only and are not direct part of the analysis. The distance of the marker to each other is used to calculate the speed of the chart in the image plane.

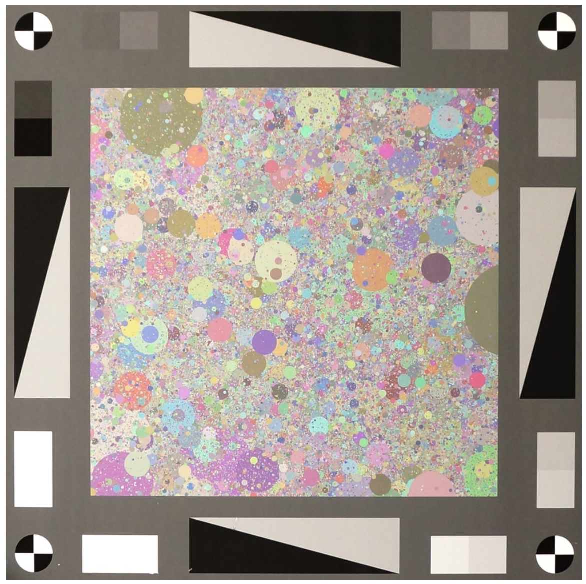


Figure 3. The used test target. *Center*: Colored Dead Leaves target, *Around*: slanted edges for SFR calculation, gray patches for linearization and markers for registration

2.2 Algorithms

2.2.1 SFR_Edge

The SFR-Edge algorithm is described in ISO12233 and is based on the reproduction of a slanted edge in the image field. The first step is to localize the edge in each row. Using this data, the offset and the slope of the edge in the image is calculated. The edge description is used to calculate an over-sampled pixel row. This is done by a binning process, placing each pixel of the image into a bin which describes a certain distance to the fitted edge. So the two-dimensional position of each pixel with column x and row y becomes a one-dimensional description with its distance to the edge.

The over-sampled description of the edge is called the edge spread function *ESF*. The first derivative of the ESF is the line spread function *LSF*[†]. The SFR-Edge is the Fourier transform of the LSF. Before the transformation, the data is windowed to avoid leakage. (see Fig. 4)

2.2.2 SFR_Dead Leaves

The basic concept to obtain the Spatial Frequency Response (SFR) is to measure the power spectrum (PS) found in the image. The image is a reproduction the camera under test has made from the dead leaves target. The PS of the target is known, so by simply dividing these two PS, one gets the SFR. As could be shown[4] that this algorithm is influenced by camera noise, it was extended[3] by a reference measurement on a gray patch that has the same intensity as the mean value of the dead leaves structure.

[†]The LSF can be imagined as a 1-D representative of the point-spread function PSF

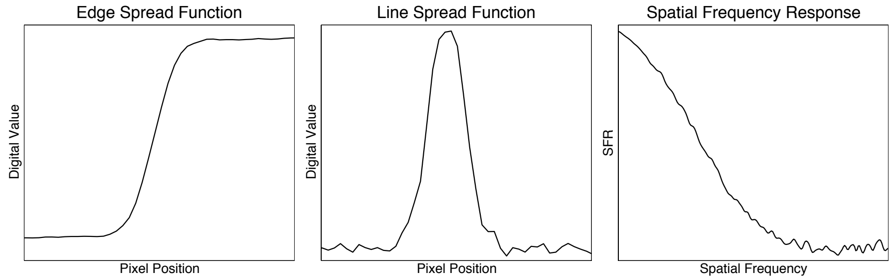


Figure 4. Basic concept of the SFR_Edge approach: The edge spread function is the over sampled intensity function of the edge. The derivative of this function is the line spread function. Transferred to the fourier spatial frequency domain, it is the spatial frequency response.

$$SFR(f) = \sqrt{\frac{PS_{image}(f) - PS_{noise}(f)}{PS_{target}(f)}} \tag{1}$$

The calculation is done in these steps, assuming the camera under test has reproduced the dead leaves target and a reference patch which is the mean value of the dead leaves structure and is homogenous.

- Calculate $PS_{target}(f)$ (PS of dead leaves target) using the meta information from chart production process.
- Read ROI of dead leaves patch, reference patch and gray patches.
- Calculate OECF with image data from gray patches and the known reflectance of these patches. The OECF here is a function of reflectance vs. Y (Y is a weighted sum of R, G, B)
- Calculate Y image from the RGB image of dead leaves patch and reference patch.
- Linearize using the inverse of the OECF.
- Calculate $PS_{image}(f)$ (from dead leaves patch) and $PS_{noise}(f)$ (from the reference patch).
- Calculate SFR(f) using Equation 1.

The calculation of the PS includes a reduction process from the 2D spectrum to 1D data and the calculation of the SFR includes a normalization process for presentation purposes.

Both algorithms provide a spatial frequency response, so a function of response versus spatial frequencies. As this is hard to compare, these functions are reduced to a single number. We use two different approaches:

"MTF50" The "MTF50" value is the highest spatial frequency that results in a modulation or spatial frequency response of $\leq 50\%$. If checking for lens performance, this value is more related to the performance in the mid frequencies the lens-camera system delivers. The higher the value, the better.

"Acutance" This value needs a more complex calculation. The idea is to take the Contrast Sensitivity Function (CSF) of the human visual system into account, so to weight the performance of a system against the importance of the spatial frequencies for the perception. The implementation we have chosen was discussed among imaging experts at working groups of ISO and I3A. The obtained SFR is filtered with the CSF and the integral of the resulting function is divided by the integral of an ideal MTF filtered with the CSF in the same spatial frequency range. The higher, the better the performance of the camera under test. As the CSF needs to be calculated for a specific viewing condition, we have chosen a 100% view on a 96ppi display in 0.5m distance.

3. VIDEO TEST

The most obvious test using a moving target is to test the image quality of a video file. This file can be created by different devices (Camera comparison) or by different video codecs (Codec Comparison).

3.1 Camera Comparison

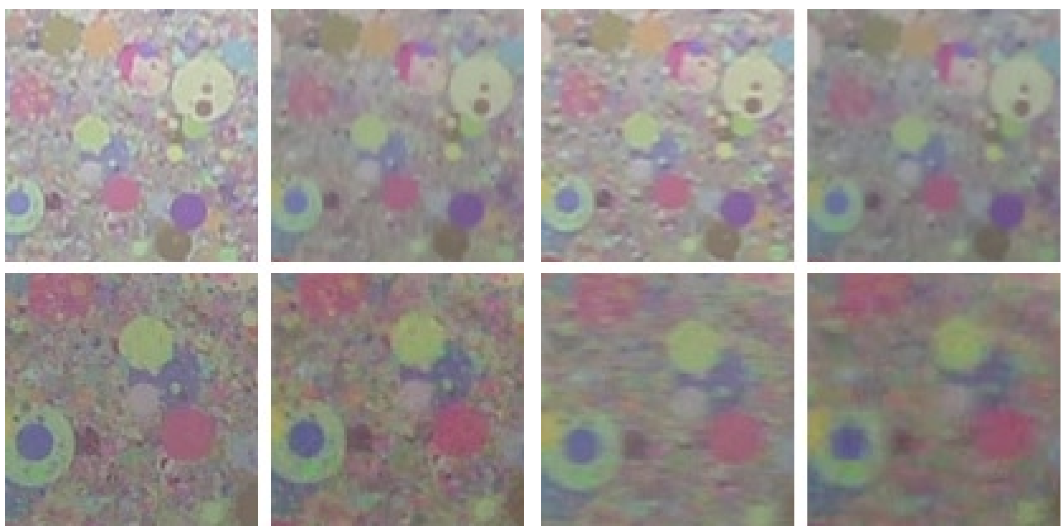
3.1.1 Procedure

The device under test has to capture the moving target under two different light conditions. "Bright" is a daylight illumination (D65) with a illuminance of 1310lx on the ground of the test scene. The "dark" light situation is tungsten light with 107 lux on the ground of the test scene. The captured files have been transferred to a PC and frames have been extracted using ffmpeg[‡] These files have been analyzed.

3.1.2 Results

The speed of the moving target is known, the speed of the object in the image plane is depending on the exposure time and therefore most of the time unknown. For a lot of devices, we do not know the exposure time, especially in video mode.

In this paper, we show the results of the Canon 5D MkIII and the Apple iPhone 5. The Canon camera is a Fullframe D-SLR, the Apple iPhone 5 a smartphone with a unknown sensor size in its camera module.



(a) No object movement (b) Object moved with 0.2 m/s

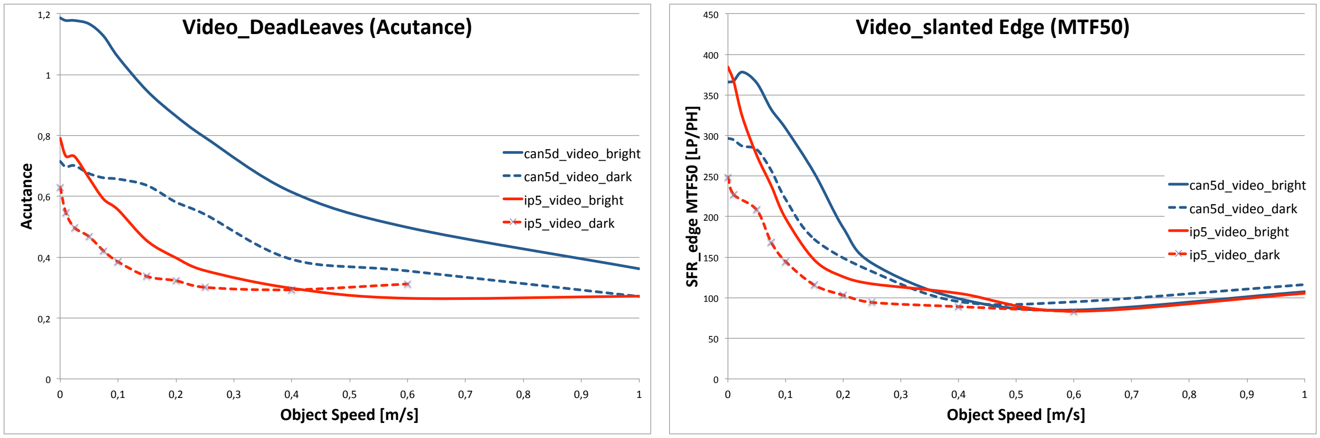
Figure 5. Enlarged crop of the video frames, showing the Dead Leaves structure. Top: Canon 5D MkIII, Bottom: Apple iPhone 5, Left: Bright, Right: Dark

The images in Figure 5 show the colored dead leaves structures as a detail of the entire frame. Figure 5(a) shows the details for a non-moving object, Figure 5(b) show the details of the frames while the object is moving with 0.2 m/s. All images are available for the two different cameras and for the bright and dark scene.

The images show that the D-SLR shows a much better performance in terms of reproduction of low contrast, fine details in bright conditions compared to the dark scene. So the texture loss increases significantly with the loss of illumination. The iPhone 5 does not perform that well compared to the D-SLR, but the difference between the dark and bright scene is much lower.

The observation from the images can also be seen in the numerical results in Figure 6. The graph in Fig. 6(a) plot the acutance vs. the object speed. One can observe the significant loss for the Canon 5D MkIII when

[‡]http://ffmpeg.org



(a) Object speed vs. Acutance (Dead Leaves)

(b) Object speed vs. MTF50 (slanted edges)

Figure 6. Two cameras compared: Canon 5D Mark III, a D-SLR with high quality video functionality (can5d) and the Apple iPhone 5 (ip5)

changing from bright to low illumination. For slow object speeds, the acutance does not change for the Canon camera, but it already does for the iPhone 5. This behavior can also be observed in Figure 6(b), a plot of SFR_edge (MTF50) vs. the object speed. While for the iPhone5 we see nearly the same behavior as we have in the dead leaves analysis, we have a slightly different behavior for the Canon. The slope of the decreasing acutance with increasing object speed is lower compared to the slope of the SFR_edge (MTF50) value. We explain this with the sharpening that is applied to the image frames of the Canon. The applied sharpening applied by the iPhone 5 to the images is much lower.

The loss of acutance with increasing object speeds is obviously depending on the actually used exposure time of the devices under test. As this time is hard to measure in high precision and may change during exposure, using a moving target is a good way to see differences between cameras.

3.2 Codec Comparision

While we have just a few image file formats that are regular used, the video world has a huge variety of video codecs in use. One important difference between the codecs is how to deal with moving objects. While simple codes just compress every captured frame and store the full information for each frame, more complex codecs store key frames and the information what happens to these key frames over the time. So to check for the resulting image quality, it makes sense to check using a moving object.

3.2.1 Procedure

For this evaluation, a video file recorded by the Canon 5D MkIII was transcoded using different codecs. The original file was generated at bright conditions and had a file size of 4.02GB (Duration of 06:27 mm:ss). This file (originally H.264) was transcoded using H.264 and MPEG2 to much smaller file sizes. The file sizes were kept constant for both codecs. The files were transcoded using an average bitrate of 1500kBps / 1000kBps / 500 kBps / 250 kBps with resulting files sizes of 85MB / 59 MB / 34 MB and 21MB.

3.2.2 Results

The numerical results are shown in Figure 8. The graphs show plots of acutance vs object speed for different bitrates including the original file. Figure 8(a) shows the results for the H.264 codec, while Figure 8(b) show the results for the MPEG-2 codec. Samples of the frames are shown in Figure 7

The H.264 transcoded files with high bitrate follow the original very closely, so we can see in the numerical results what we also see in visual inspection: The texture loss is very low and the difference to the original very low. At very low speed, the difference also for lower bitrates is not significant. The difference in the quality of the codecs gets visible (in images and numerical results) in the area of slow moving objects.

While the moving object get blurry in H.264, we see block artifacts in the MPEG-2 codec. This block artifacts are considered as the reason why the acutance gets better for high object speeds compared to the original.

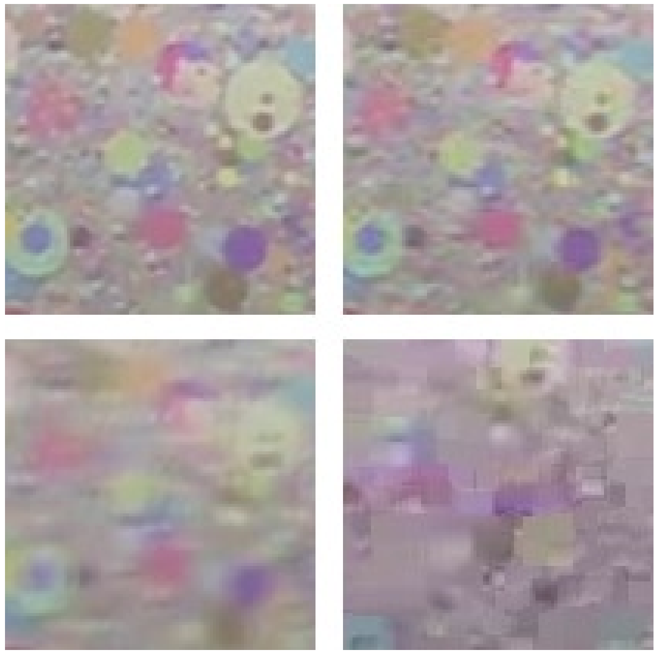
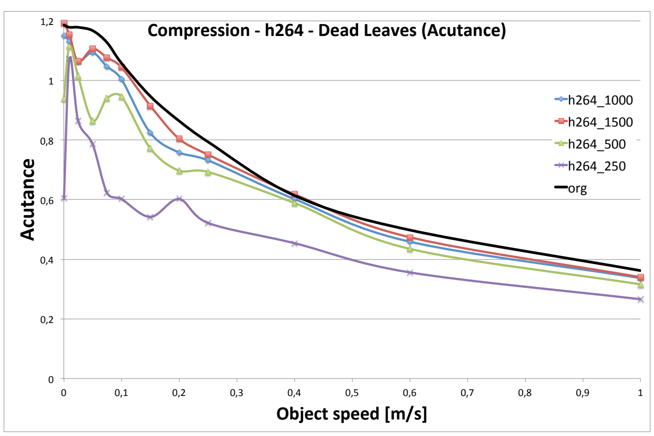
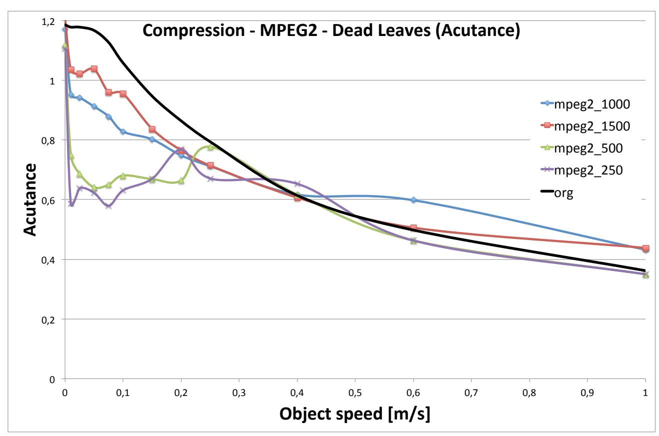


Figure 7. Enlarged crop of the video frames, showing the Dead Leaves structure at a object speed of 0.2 m/s. Left: H.264, Right: MPEG-2; Top: 1500kBps, Bottom: 250kBps



(a) h264 codec - Object speed vs. Acutance



(b) MPEG2 codec - Object speed vs. Acutance

Figure 8. The original (org) file is the standard output of a Canon 5D MkIII, capturing the moving target. This file has been transcoded using the h264 and the mpeg2 codec with decreasing bitrate. Filesize: org: 4.02GB, 1500: 85MB, 1000: 59MB, 500; 34MB, 250: 21MB

4. PHOTO TEST

While the test of video functionality is the most obvious test that can be done using a moving target, we were investigating the possibilities for the analysis of the photo functionality.

As we can extract the exposure time from the meta data of the image files, we can calculate a comparable unit of the movement by using the unit "pixel per exposure"[px/exp]. This unit represent the distance (measured in pixel) the chart has moved during the exposure and is a good indicator for the potential motion-blur that can be observed in the image.

4.1 Sharpening and Moving targets

The signal processor of modern digital cameras allows an intensive use of image enhancement algorithms to optimize the image appearance and image quality (or to hide shortcomings of the hardware). Sharpening is a process to boost the high frequency content of an image, to improve the edge appearance. If the input signal to the sharpening algorithm does not contain much high frequency content, the sharpening is reduced. A moving object can be considered as a low-pass filter to the image content, so with increasing speed the high frequency content is reduced and we can learn about the behavior of the sharpening algorithm.

Figure 9 shows the graph SFR_DL (the spatial frequency response based on the dead leaves pattern) for different sharpening level of the same camera (Sony NEX-7) and for non-moving object and at a movement of 5px/exp.

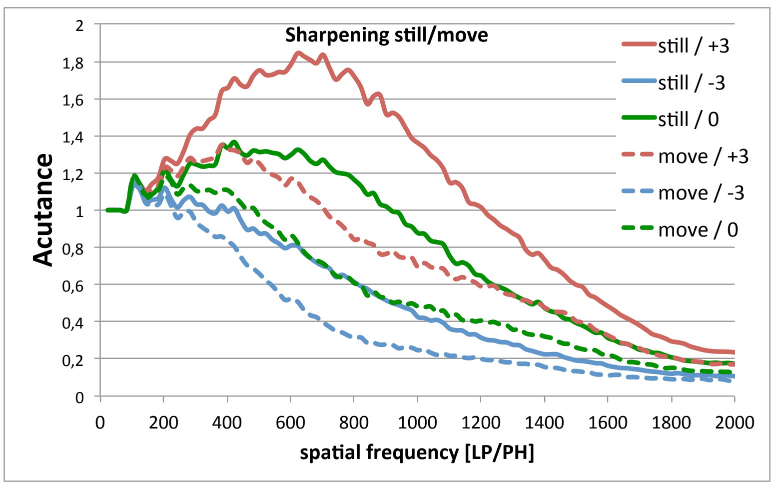
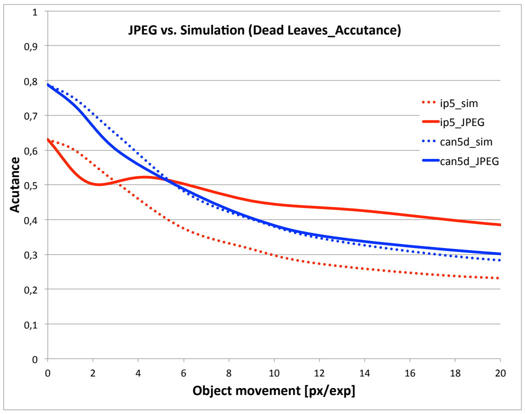


Figure 9. The SFR calculated from the dead leaves structures; Sharpening was applied in the camera from +3 to -3; Results for non moving object and moving object at 5px/exp

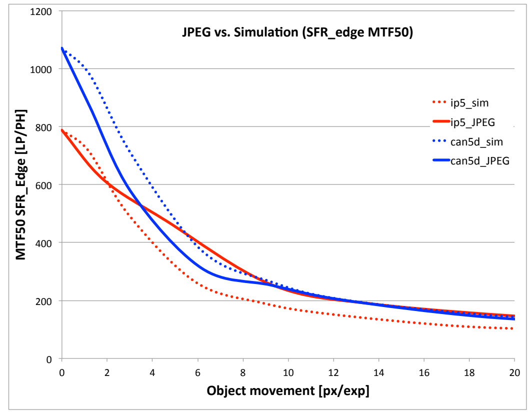
We can see, that the SFR_DL is very much influenced by the image sharpening. While checking the images, one can see that it is o.k. that the SFR_DL is higher for higher sharpening, as it shows more details, but the difference seems to be too high. This phenomena should be checked in other psychophysical studies. Comparing the two plots per sharpening level, one can see that the higher the sharpening level, the more it looses due to object movement. So this might be a way to investigate in more detail, how the image sharpening works for a particular camera and how much of the results are depending on image enhancement.

4.2 Real Image versus Simulation

If we know the movement of the object (in pixel per exposure), we can calculate the theoretical motion blur that is introduced due to this movement. We simulated the motion blur based on the speed and exposure time we measured for different images. after that, we can compare the result for real image with the simulated counterpart. In theory, the results should be at least close to each other.



(a) Acutance



(b) MTF50

Figure 10. Based on the still image, the motion blur is simulated and the analyses results compared with the real images.

For this experiment, we used the Canon 5D MkIII and the Apple iPhone 5. In Figure 10 we show the results of this experiment as a plot of acutance vs. movement (10(a)) and MTF50 vs. movement (10(b)).

We could see, that the results of the D-SLR (Canon 5DMkIII) are close between simulation and real images. For the iPhone 5, we can see that there are quite some difference, especially in the acutance, based on the dead leaves.

We have two ideas that can explain this observation:

Rolling Shutter Effect In the simulation, we assume a global shutter. The motion blur is based on the assumption, that the exposure time is equal in the entire image file and that the exposure time provided in the EXIF data is true. It is very likely, that these assumptions are not true for the mobile phone. The Rolling Shutter effect results in distortion rather than motion blur and a reduced local exposure time.

Noise and Artifacts The motion of an object is simulated by applying a filter kernel to the original image. This process is basically a low-pass filter operation which is applied to the entire image content. In the original image, the motion blur is created during the exposure process. Potential image noise, which contains all frequencies, is not influenced by the motion. So by comparing the power spectrum of the image (including the image noise) and the power spectrum of the target (low pass filtered by the motion blur) we get an over estimation of the SFR_DeadLeaves. (see Eq. 1)

We tried to check the two ideas against the observations we made. The rolling shutter effect seems to be a reasonable explanation, even though we would have to see it comparing the two images. Figure 11 show a detail of the two images. We see strong motion blur in both images, so a potential rolling shutter effect seems to be low.

For further investigation what causes the differences, we included the knowledge about the motion blur into the analyzing process. We calculated the theoretical motion blur (the same that was already used to create the simulation images) and applied the resulting transfer function to $PS_{target}(f)$ (see Eq. 1). This "motion corrected Dead Leaves" approach should give the same result for all different object motion. Figure 12 shows the result for the simulation data. We can see, that this method provides the data as we would expect.

Even if it looks promising on the simulation data, the "motion corrected Dead Leaves" approach did not work on real image data. The motion blur reduces the higher frequencies in $PS_{target}(f)$ that much, that any kind of noise or artifact that was added to the image has a huge impact on the resulting SFR_DL. Figure 13 shows this effect. With increasing motion, the response at the higher frequencies increases. So basically we increased the assumed effect in this approach.

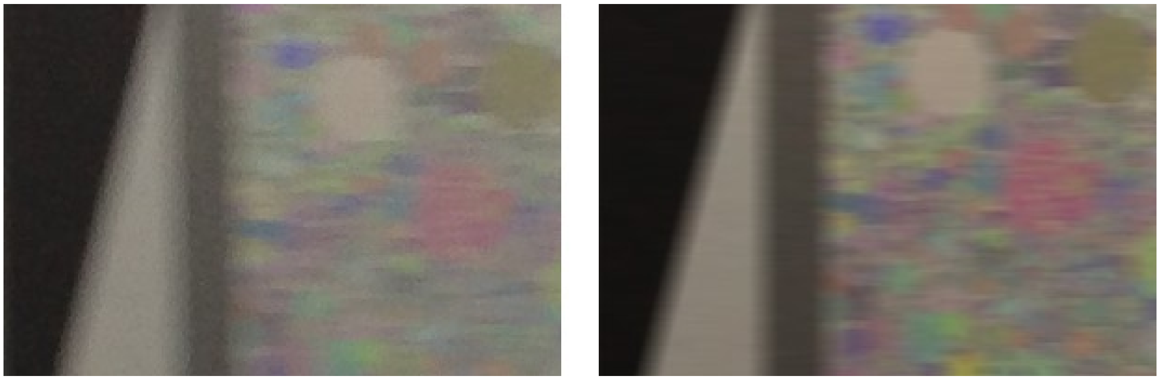


Figure 11. Enlarged detail of the original image (right) and the simulated counterpart (left). Both images show a movement of 15 px/exp

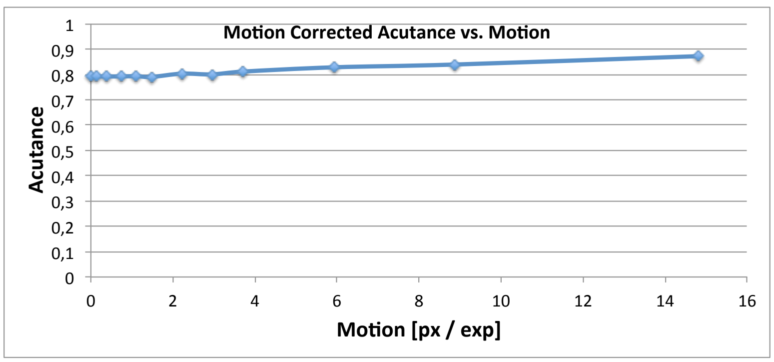


Figure 12. Acutance vs. motion - motion corrected analysis

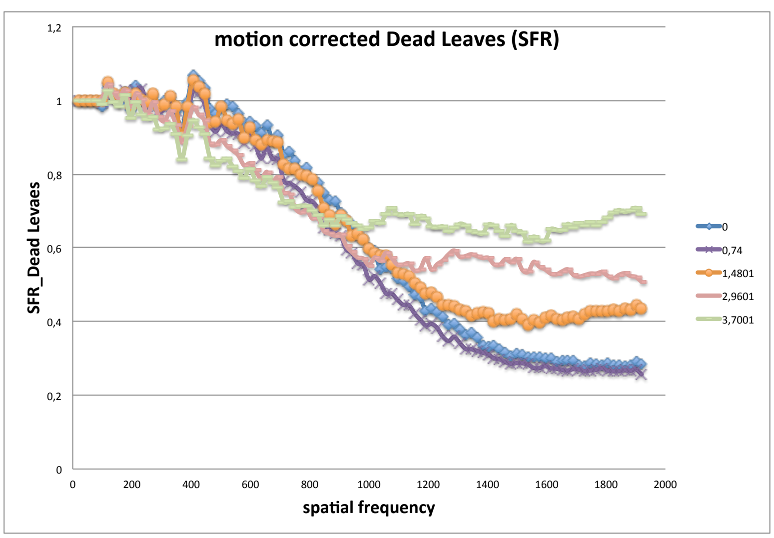


Figure 13. motion corrected Dead Leaves for different object motion

5. CONCLUSION

- While we could see different interesting aspects, it is hard to derive a new approach of image quality analysis from our observations.
- Using a moving chart is a benefit when investigating video image quality comparing different devices.
- When comparing codecs or other quality parameter, the acutance vs. known speed of the object is a interesting parameter.
- The dead leaves structure and the corresponding algorithms can describe the effects video compression has on the image content, but can be fooled by image noise and artifacts. Especially for higher motion speed, the influence of noise and artifacts on the results is increased.
- Sharpening has a significant influence on the dead leaves approach. It seems to be too high and should be part of further investigation.

REFERENCES

1. Golik, Wueller, "Measurement Method for Image Stabilizing Systems", SPIE-IST Electronic Imaging, SPIE Vol. 6502, 65020O (2007)
2. Cao, Guichard, Hornung "Measuring texture sharpness",Electronic Imaging Conference, Proc. SPIE, Vol. 7250, 72500H (2009)
3. Jon McElvain, Scott P. Campbell, Jonathan Miller and Elaine W. Jin, "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems", Proc. SPIE 7537, 75370D (2010); doi:10.1117/12.838698
4. Artmann, Wueller, "Differences of digital camera resolution metrology to describe noise reduction artifacts", Electronic Imaging Conference, Proc. SPIE, Vol. 7529, 75290L (2010)
5. Artmann, Wueller, "Improving texture loss measurement: spatial frequency response based on a colored target", Electronic Imaging Conference, Proc. SPIE, Vol. 8293-4, (2012), DOI: 10.1117/12.907303
6. Artmann, Wueller, "Noise Reduction vs. Spatial Resolution",Electronic Imaging Conference, Proc. SPIE, Vol. 6817, 68170A (2008)
7. International Organization of Standardization,"ISO12233 Photography - Electronic still picture imaging - Resolution measurements"