ULTRA HD: BANDWIDTH PLANNING AND VERIFICATION FOR 4K

Sean T. McCarthy

ARRIS, USA

ABSTRACT

UHD 4k TVs are quickly taking their place in people's homes. Content creators are beginning to produce Ultra HD 4k in earnest to fill the programming pipeline. It is time for broadcasters to consider how UHD 4k fits into their overall offering. The critical considerations are bandwidth and video quality; but the industry doesn't yet have the same level of experience setting bitrate & video-quality targets for 4k content as it does for high-definition (HD) and standard-definition (SD) content. The industry needs tools to validate the quality of uncompressed and compressed UHD 4k content so that broadcasters can plan bandwidth resources with confidence. This paper will describe an innovative and practical set of statistical methods that can be used to validate UHD 4k content and examine the impact of high-efficiency video coding (HEVC) compression.

INTRODUCTION

Only a decade ago, high definition HD was the big new thing. With it came new wider 16:9 aspect ratio flat screen TVs that made the living room stylish in a way that old CRTs couldn't match. Consumers delighted in the new better television experience. Studios and broadcasters delivered a new golden-age of television. HD is now table stakes most places, and where that is not yet the case, it will be soon enough.

Yet now, before we hardly got used to HD, we are talking about Ultra HD (UHD) with at least four times as many pixels as HD. Is all that extra UHD 4k resolution going to make a difference to consumers? If yes, what bandwidth will UHD 4k programming need? Those are two big questions our industry is exploring with respect to planning UHD services; yet they are not independent questions.

UHD 4k is still new enough in the studios and post-production houses that 4k-capable cameras, lenses, image sensors, and downstream processing are still being optimised. Can we be sure yet that the optics and post processing are preserving every bit of "4k" detail? On the distribution side, could video compression change the amount of visual detail to an extent that it could conceivably turn "4k" quality into something more like "HD" or even less?

If the UHD 4k content we have available today for bandwidth and video quality testing does not truly have a "4k"-level of detail, then we could go astray and plan for less bandwidth than we might need for future UHD 4k broadcasts. If the 4k content we have available today is truly "4k", then we should also want to be sure that we do not over compress and turn UHD 4k into something less impressive.

During our UHD 4k testing, we have found several test sequences that appeared normal to the eye but turned out to have unusual properties when examined mathematically. Such content could lead to wrong conclusions when planning for UHD 4k services.

In this paper, we present practical mathematical techniques to help answer the question “How 4k is it?” Our method examines UHD 4k video to see if it has a statistically-expectable distribution of spatial detail as a function of 2-dimensional spatial frequency. The benchmark for our statistical expectations is drawn from numerous studies of the statistics of natural scenes.

Our main objective in writing this paper is to describe a methodology that might be useful in helping to decide which UHD 4k content should be included in any video test library intended for use in bandwidth and video quality planning.

SPATIAL FREQUENCY

An image is normally thought of as a 2-dimensional array of pixels with each pixel being represented by red, green, and blue values (RGB) or luma and 2 chrominance channels (for example, YUV or YCbCr). An image can also be represented as a 2-dimensional array of spatial-frequency components as illustrated in Figure 1. The visual pixel-based image and the spatial-frequency representation of the visual image are interchangeable mathematically. They have identical information, just organised differently.

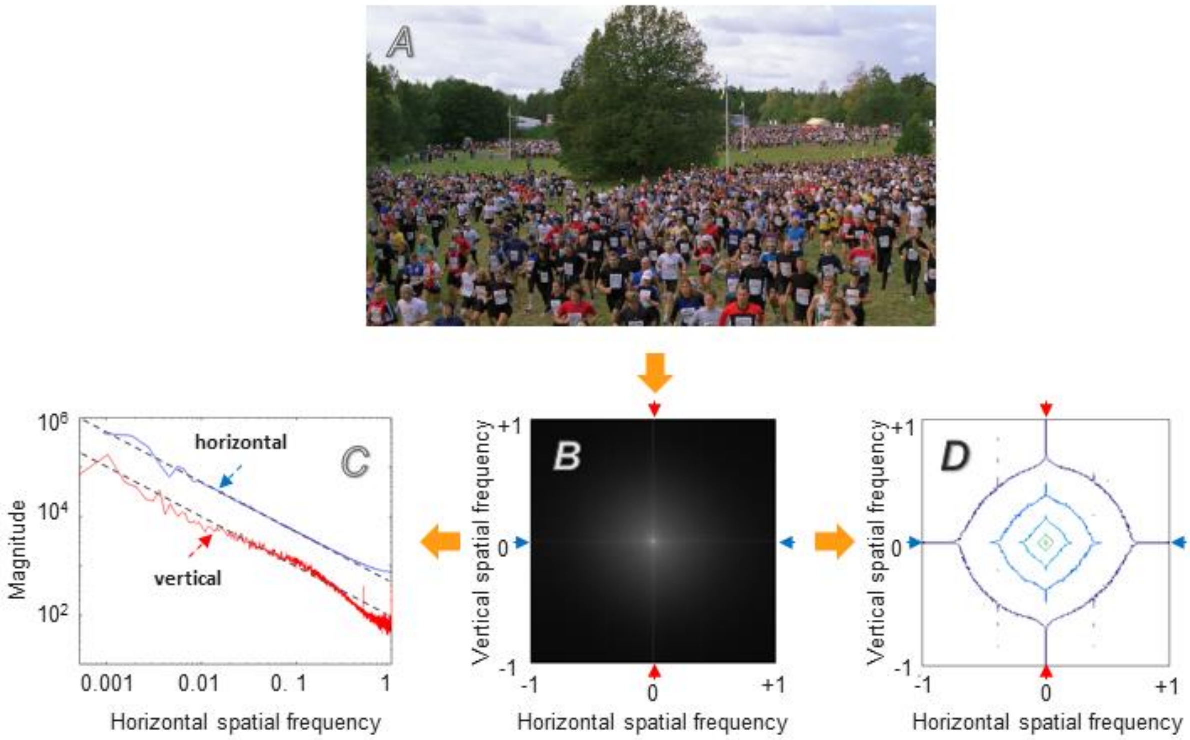

Figure 1 – Illustration of representing an image in terms of spatial frequencies.

Spatial-frequency representations of images have a magnitude component and a phase component. The magnitude component, called the magnitude spectrum, provides information on how much of the overall variation within the visual (pixel-based) image can be attributed to a particular spatial frequency. (Spatial frequency is 2-dimensional having horizontal and vertical parts.) The phase component, called the phase spectrum (not shown), provides information on how the various spatial frequencies interact to create the features and details we recognise in images.

As illustrated in Figure 1, the visual pixel-based image (***A***) can be represented as a 2-dimensional array of complex numbers using Fourier transform techniques. The absolute-value of the complex numbers is shown as a 2-dimensional magnitude spectrum (***B***) in which brighter areas correspond to larger magnitude values. (Note that the log of the magnitude spectrum is shown in ***B*** to aid visualisation. The horizontal and vertical frequency axes are shown relative to the corresponding Nyquist frequency (±1).) The magnitudes of the main horizontal and vertical spatial frequency axes are shown in ***C*.** The main horizontal spatial frequency axis corresponds to zero vertical frequency (blue arrows in ***B***), and the main vertical spatial frequency axis corresponds to zero horizontal spatial frequency (red arrows in ***B***). The dashed lines in **C** indicate the 1/spatial frequency statistical expectation for natural scenes (the 1/spatial frequency appears as a line in the log plot).

We find it useful to use contour maps of the log of the magnitude spectra to create a *gestalt* of the 2-dimensional spatial frequency composition of images. A contour map of the log of the magnitude spectrum is shown in Figure 1, ***D***.

[Note that the data shown in Figure 1 were obtained by averaging the magnitude spectrum of individual frames over 250 frames (5 seconds) of the SVT CrowdRun 2160p50 test sequence available from the European Broadcasting Union (1) . All video processing and analysis discussed in this paper were performed using MATLAB (2) and ffmpeg (3).]

STATISTICS OF NATURAL SCENES

Images of natural scenes have an interesting statistical property: They have spatial-frequency magnitude spectra that tend to fall off with increasing spatial frequency in proportion to the inverse of spatial frequency (4). The magnitude spectra of individual images can vary significantly, but as an ensemble-average statistical expectation, it can be said that "the magnitude spectra of images of natural scenes fall off as one-over-spatial-frequency." This statement applies to both horizontal and vertical spatial frequencies.

As demonstrated in Figure 2, "natural-scene" images are not limited to pictures of grass and trees and the like. Any visually complex image of a 3-dimensional environment tends to have the one-over-frequency characteristic, though man-made environments tend to have stronger vertical and horizontal bias than unaltered landscape (3). The one-over-frequency characteristic can also be thought of as a signature of scale-invariance, which refers to the way in which small image details and large image details are distributed.

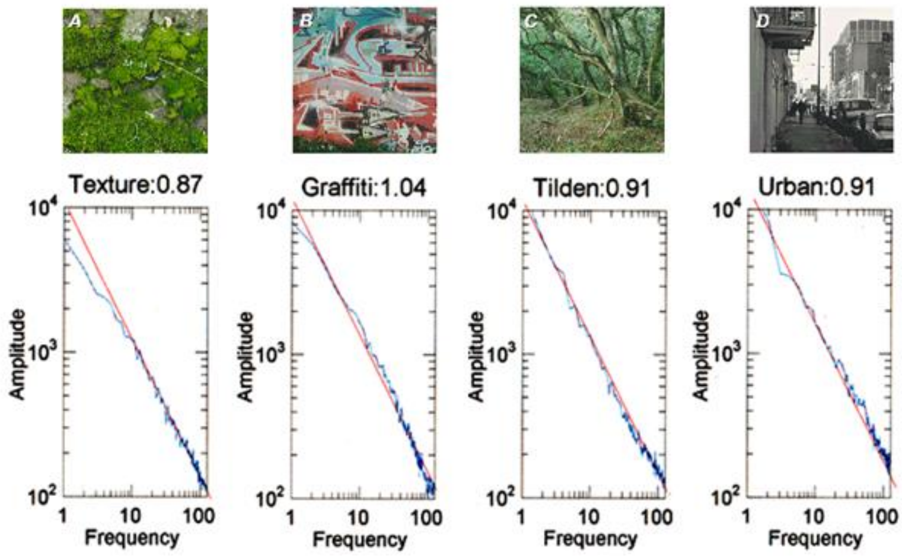

Figure 2 – Examples of adherence to the 1/f statistical expectation for images of natural and man-made environments

A BENCHMARK FOR "4K" SPATIAL DETAIL

In this paper, we leverage the one-over-frequency statistical expectation to see if it holds for UHD 4k content we have considered for use in our lab tests. Examples of 4k (2160p50) video sequences that largely adhere to the one-over-frequency statistical expectation are shown in Figure 3.

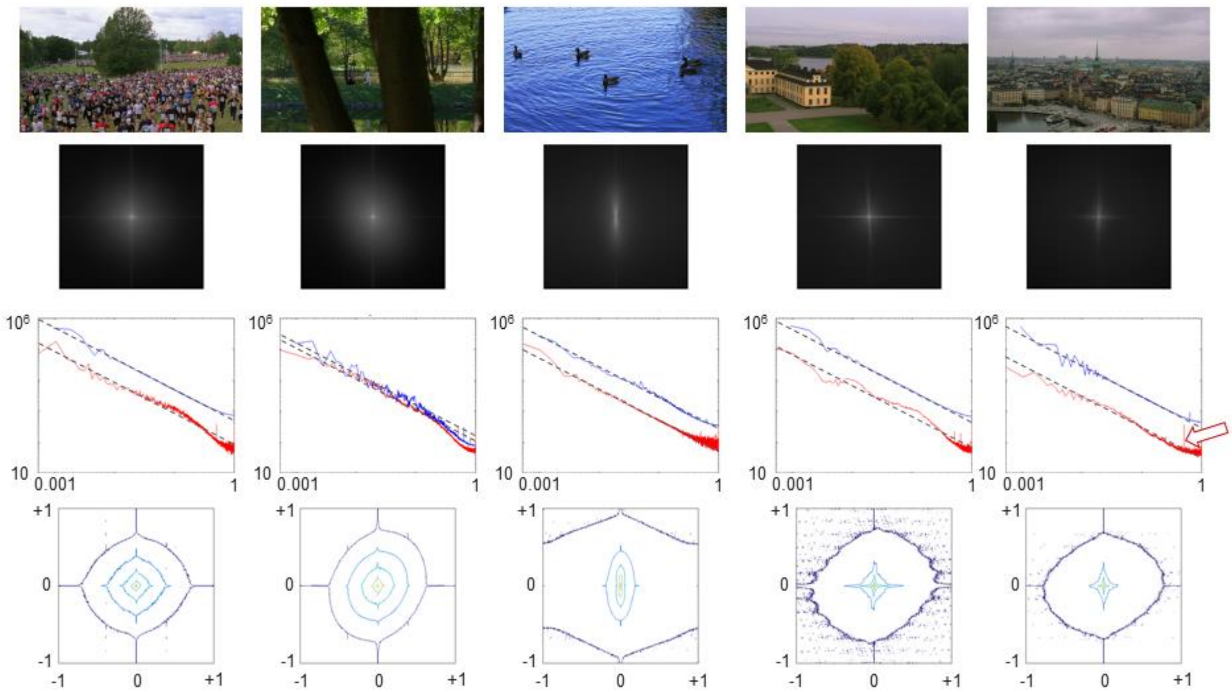

Figure 3 – UHD 4k test sequences that have statically-expectable magnitude spectra

Figure 3 illustrates analysis of the SVT UHD 4k test sequences available from the EBU (1): “CrowdRun”, “ParkJoy”, “DucksTakeOff”, “InToTrees”, and “OldTownCross” left to right in columns. All sequences are 3840x2160 at 50 frames per second. Note that each of the sequences can be well-described by the one-over-frequency statistical expectation (dashed lines in the plot on the third row from the top); though there are some subtle deviations from statistical expectations in the form of narrow-band peaks (see arrow). Note that the contour maps provide concise distinguishing information about each image sequence.

Examples of candidate UHD 4k (3840x2160) videos that violate the one-over-frequency statistical expectation are shown in Figure 4. Some UHD 4k test sequences that we have obtained from various sources that appeared normal to the eye were found to have spatial magnitude spectra that were inconsistent with statistical expectations. Typical deviations from statistical expectations included: notch-like frequency distortions; excessive or diminished high or low frequency spatial detail (non-one-over-frequency behaviour); and extraneous noise.

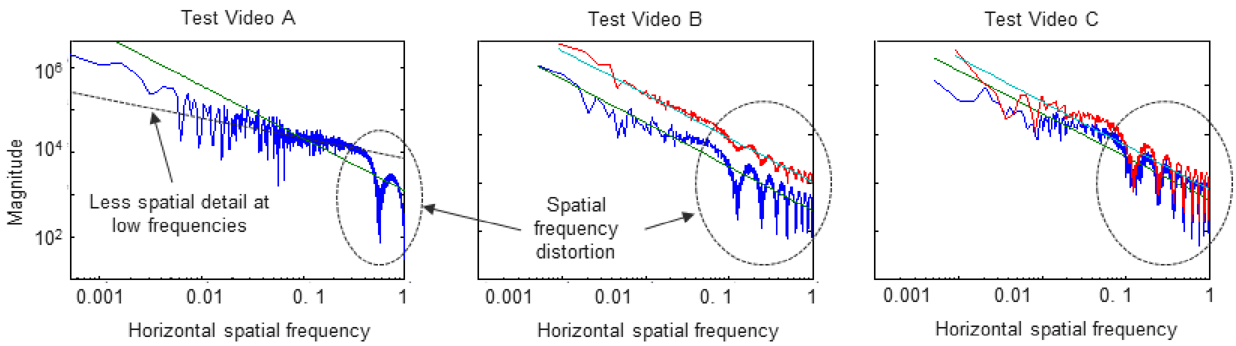

Figure 4 – Examples of candidate UHD test sequences
that do not have statically-expectable magnitude spectra

Figures 3 & 4 illustrate a method of validating UHD 4k test sequences for use in bandwidth planning. The test sequences shown in Figure 3 remain viable candidates; but we would exclude the test sequences represented in Figure 4 from our UHD 4k test library.

UPCONVERSION & THE EFFECTIVE RESOLUTION OF “4K” VIDEO

UHD 4k displays have such high resolution, and upconversion algorithms have become so good, that it is sometimes difficult to see by eye if a particular video is pristine full resolution or if some upconversion has occurred in the preparation of the content.

In Figure 5, we illustrate a method of analysing the effective resolution of “4k” (3184x2160) resolution test content more quantitatively than can be done by eye. It is well-known that a reduced effective resolution correlates to a loss of high-frequency spatial detail. This loss could, in principle, be evident by inspecting the main horizontal and vertical axes of the magnitude spectrum. We find that is not always the case.

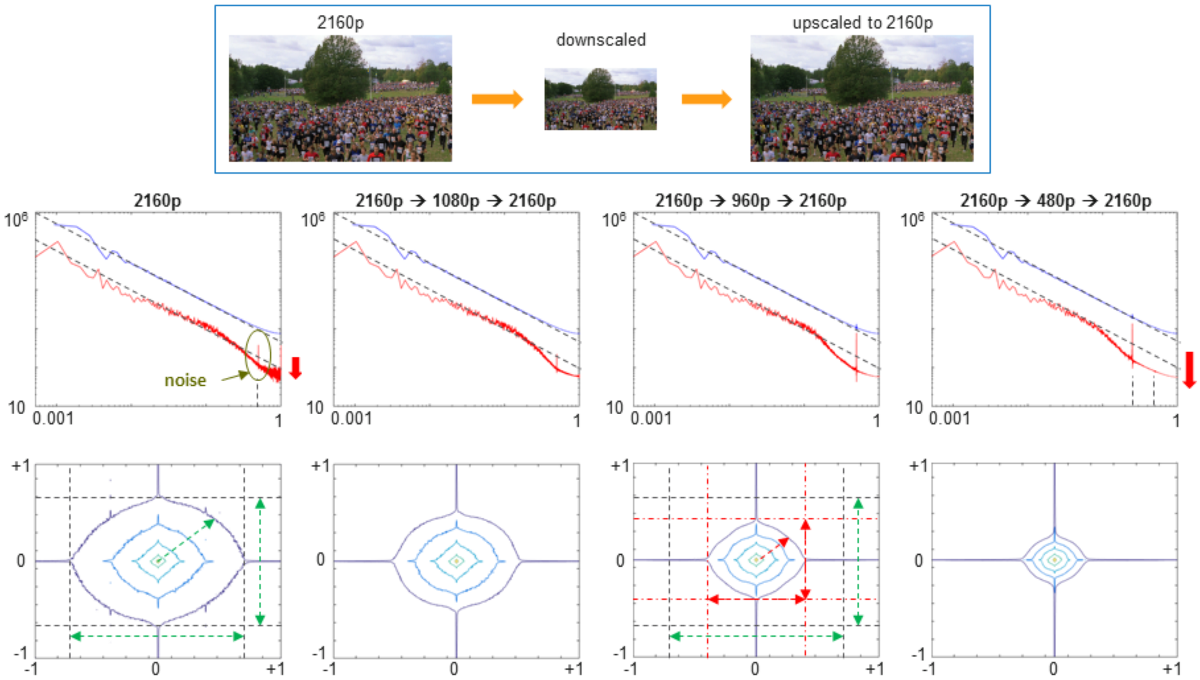

Figure 5 – An example of using contour maps of magnitude spectra to examine effective spatial resolution

As illustrated in Figure 5, we simulated the impact of upconversion by downscaling and then upscaling back to 3840x2160 resolution using ffmpeg. From left to right in Figure 5, the downscaled resolution is: unaltered 3840x2160; 1920x1080; 960x540; and 480x270 as an extremum. Note that examination of only the main horizontal and vertical axes of the magnitude spectrum (middle row) reveals some differences, most notably some reduction in high-frequency spatial detail and shift in the narrow-band noise; but these details are too subtle to make confident decisions, particularly in situations in which the original full resolution content would be unavailable for comparison. The contour maps of the log of the average magnitude spectrum provide more clear-cut evidence. The contour levels are that same for all columns. Thus the constriction of the contours towards the centre indicates that the magnitude spectrum narrows (loses high-frequency spatial detail) thus quantifying the reduced effective resolution.

Modern rescaling algorithms are very sophisticated and the visual difference between lowered effective resolution and full resolution can be subtle. We find that the contour maps of the log of the magnitude spectrum are a much more sensitive indicator of effective resolution. A test for accepting candidate UHD 4k test content into a master library could be along the lines of determining the average radius of the outermost contour, accepting test content only when a certain radius threshold is exceeded.

EFFECT OF HEVC COMPRESSION

Video compression changes the amount of spatial detail in video, but the extent to which spatial detail is lost depends of the content itself and the aggressiveness of compression; i.e. the target bitrate.

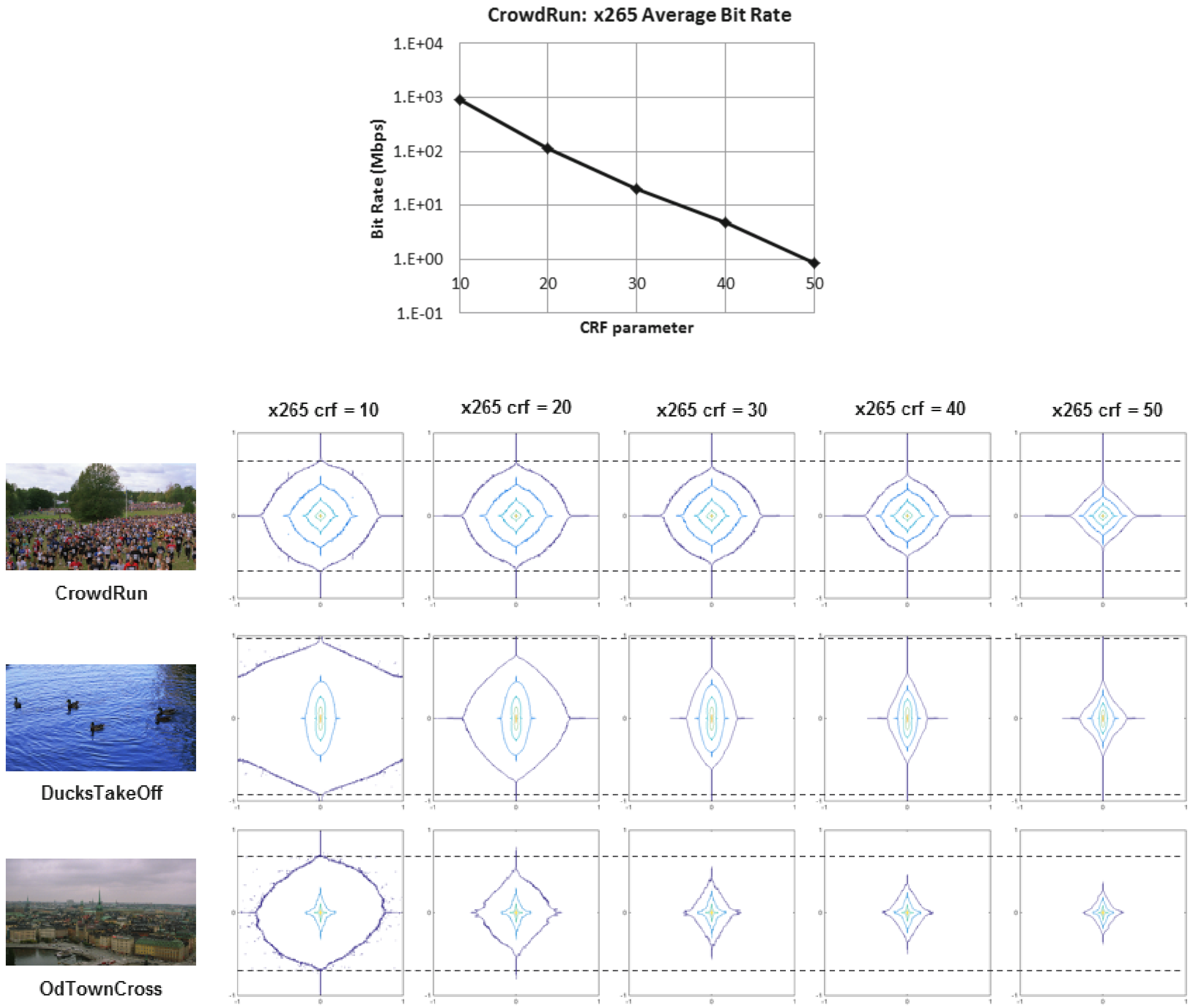

Figure 7 – An example of using contour maps of magnitude spectra to examine the effect of video compression

In Figure 7, we demonstrate that our method of evaluating test content provides a way of testing the effective resolution of HEVC compressed content. Shown are the contour maps of the average (250 frames, 5 seconds) of the log magnitude spectrum of each of the SVT UHD 4k test sequences compressed with HEVC to various extents. We used the libx265 (6) library with ffmpeg to perform the HEVC compression. The ***crf*** value noted at the top of each column indicates the value of the constant rate factor (***crf***) parameter used in the ffmpeg libx265 command line. Smaller values of ***crf*** created more lightly compressed video. Video compressed with a ***crf*** value of 50 is typically very heavily artefacted. Video compressed with a ***crf*** value of 10 produces contour maps that are very similar to those for uncompressed video (see Figure 3).

Note that the impact of the ***crf*** value is content dependent. For “CrowdRun, the ***crf*** values below ~30 do not a have a major impact on effective resolution. On the other hand, a noticeable change in effective resolution is evident for a ***crf*** value of 20 for “DucksTakeOff” and “OldTownCross”. (The dashed lines provide a reference for the radial extent of the outer contour of the lightly compressed and uncompressed versions of the video.)

Our contour-map method indicates that effective resolution is more sensitive to compression for some kinds of content compared to other kinds of content. As such, our contour-map method could be used to optimise the selection of compressed UHD 4k content for testing purposes in terms of both the intrinsic image characteristics of content and the impact of bit rate. In this way, our contour-map method could serve as a content-independent method of measuring effective resolution and thus selection of useable UHD 4k test content.

DISCUSSION & CONCLUSIONS

The objective of this paper was to present techniques to evaluate UHD 4k video sequences intended for use in video quality and bandwidth planning. Out motivation was that inclusion of test content that is not representative of anticipated UHD 4k programming – including future UHD 4k programming that will be available when the end-to-end UHD 4k ecosystem has been optimised – could lead to wrong conclusions.

We show in this paper that ensemble-average statistical expectations related to spatial frequency magnitude spectra of images of natural scenes can be used as a benchmark of comparison to address the question: “How 4k is it?” We find that major deviations from statistical expectations can be considered grounds for excluding content from a test library.

We also find that examination of contour maps of the log of the magnitude spectra are sensitive indicators of effective resolution and the impact of video compression on spatial detail. Contour maps that indicate a significantly less than 4k-level effective resolution can be considered grounds for excluding content from a test library. Simple inspection of the main horizontal and vertical components of the magnitude spectrum is a good means of detecting added noise and gross distortions; but it is not as sensitive as our contour-map method for assessing effective resolution.

A key finding is that our contour-map method is a sensitive content-independent to scrutinise both uncompressed and compressed candidate UHD 4k test content to an extent not possible with the eye alone. We believe the methods we presented could aid broadcasters in making video quality and bandwidth planning decisions for UHD 4k services with confidence.

REFERENCES

1. EBU UHD-1 Test Sequences. https://tech.ebu.ch/testsequences/uhd-1
2. The Mathworks. www.mathworks.com
3. ffmpeg. www.ffmpeg.org
4. Field, D.J. 1987. Relationship between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A. Vol. 4, No. 12
5. McCarthy, S.T. and Owen, W.G.. Personal communication. University of California, Berkeley.
6. x265. www.x265.org