Analyze And Predict the Perceptibility of UHD Video Contents.

Steve Göring, Julian Zebelein, Simon Wedel, Dominik Keller, Alexander Raake;
Audiovisual Technology Group; Technische Universität Ilmenau, Germany;
Email: [steve.goering, julian.zebelein, simon.wedel, dominik.keller, alexander.raake]@tu-ilmenau.de

Abstract

720p, Full-HD, 4K, 8K, ..., display resolutions are increasing heavily over the past time. However, many video streaming providers are currently streaming videos with a maximum of 4K/UHD-1 resolution. Considering that normal video viewers are enjoying their videos in typical living rooms, where viewing distances are quite large, the question arises if more resolution is even recognizable. In the following paper we will analyze the problem of UHD perceptibility in comparison with lower resolutions. As a first step, we conducted a subjective video test, that focuses on short uncompressed video sequences and compares two different testing methods for pairwise discrimination of two representations of the same source video in different resolutions. We selected an extended stripe method and a temporal switching method. We found that the temporal switching is more suitable to recognize UHD video content. Furthermore, we developed features, that can be used in a machine learning system to predict whether there is a benefit in showing a given video in UHD or not. Evaluating different models based on these features for predicting perceivable differences shows good performance on the available test data. Our implemented system can be used to verify UHD source video material or to optimize streaming applications.

Introduction

Current display technologies are increasing pixel density more and more [27]. For example, 8K resolution is currently available for consumer screens. Considering that current streaming providers such as Netflix [23] support 4K video playout, the question arises what the benefit of 4K (4096x2160) or UHD-1 (3840x2160) resolution over Full-HD (1920 x 1080) will be for specific contents. On the other hand, adaptive streaming will increase the internet traffic of the future more and more [7]. Therefore video quality estimation methods are used to ensure high quality streams, also in, e.g. rural areas in combination with adaptive streaming. Such video quality models are currently extended to support 4K resolution, e.g. VMAF [24], or our developed model *deviq* [8]. As basis for such video quality models, typical subjective quality tests are conducted.

Especially for higher resolutions, e.g. 4K, those quality tests are using a viewing distance of 1.5 or 1.6 times the height of the display, according to several recommendations [28, 14, 11, 12], e.g. of ITU-P ITU-R BT.2246-6 [14]. Considering real-world applications, for example where users are sitting in their living rooms, those viewing distances are practically never used. Viewer-Screen distances of 5.5 or even higher multiples of the display height are commonly used at home [25]. Differentiating of UHD and HD content in realistic living rooms is therefor already hard or impossible. In addition, previous studies of Berger et al. [4] showed that probably with UHD the border of visual perception is already reached. Berger et al. [4] conducted a subjective test using absolute categorical rating (ACR) for HEVC encoded material with the recommended 1.5H/1.6H viewing distance. It leads to the problem definition of UHD perceptibility, that reflects how good UHD resolution of a given video can be differentiated in comparison with lower resolutions, e.g. Full-HD. Furthermore, other studies, such as Li et al. [20]'s work focus on up-scaling algorithms for UHD. Moreover, new encoders, such as AV1 [2] are developed to reduce the bitrate of high resolution videos even more by keeping subjective video quality as high as possible [1] in e.g. adaptive video streaming applications [18].

Summarizing, there are two factors for UHD perceptibility. First, recommended viewing distances for UHD are practically not usable and second there is probably only a small perceivable difference between UHD and HD. Considering that e.g. in streaming scenarios bandwidth could be saved if a content is not streamed in UHD due to the lack of perceivable differences to HD, so that coding at HD could be used saving bitrate, a deeper analysis of UHD perceptibility is required.

We focus in this paper on perceptual differences of HD and UHD for uncompressed material including direct comparison test methods at the critical viewing distance of 1.5H. We further describe a prediction system that can already be used before encoding a video to decided if a UHD/4K version is beneficial or not for the end-viewers.

We focus in the paper to the following identified research questions. First, we check what is a suitable test method for UHD and HD comparison. The results of the tests are later used as a ground truth for the second research question, that analysis if computer vision and quality related video features are able to classify videos according their UHD perceptibility. Our classification model is based on machine learning algorithms, in our case random forest trees, however other methods like support vector machines will have similar results. We conducted two evaluation experiments. The first uses a synthetic dataset as ground-truth to verify that our features are able to distinguish between UHD and HD. We found out that our system performs well with an accuracy of approximately 82%. As second experiment we use the results from the subjective test and train our system, there we are getting an accuracy score of approximately 65%. It can be concluded that it is not only hard for a human to distinguish between UHD and HD, furthermore also for machine learning system this is a challenging task, that needs even more research.

This paper is organized as follows. The following Section describes the conducted subjective test methods and test results. The results will be used in Section as ground truth for our developed prediction model, that uses computer vision features to classify videos according their UHD perceptibility. Last Section will conclude and identify some future work.

Perception Test Setup

To generate suitable ground truth data, we conducted subjective video quality tests. Considering the state of the art, there are several used methods for comparison of e.g. different codecs, resolutions or other settings [4, 16, 15, 20, 29, 13].

Based on some pre-tests we analyzed and identified the following methods that can be applied in UHD recognition scenario.

First, a side-by-side test with two screens, one screen showing UHD content and another one showing the upscaled HD content. Based on a short pretest, this approach showed to be less suitable, due to the fact that the subject needs to move the head quite often over longer viewing angles, if typical large consumer screens are used (65 inch). Furthermore, both screens need to be calibrated in the same way, which makes this approach even more complex.

Therefore, we decided to go for a test setup with only one screen. With such a one-screen setup, several methods for showing the two contents are conceivable.

Classical methods based on ITU-T Rec. P.910/913 [16, 15] would show each stimulus one after another. This method is simple to implement, however no direct comparison of the stimuli is implemented in this method.

Used Test Methods

In [20, 29] a one stripe method, where the video signal is split in the middle into two separate views, e.g. on the left the HD version and on the right side the UHD version of the video. In an informal pre-test we evaluated such an approach, revealing that this approach may work well only when critical parts for UHD are in the middle. Furthermore, the compared parts of the video are not identical, so that resolution differences are hard to identify.

Figure 1. *Example of the used STRIPES comparison method.*

We extended the one stripe method to a multi-stripe method using in total 12 stripes, we referred to this test method as *STRIPES*. In Figure 1 an example of the used stripes is shown, the colored bars (blue and green) are introduced to support the participant in the judgment which of the two versions is of higher quality (A vs. B). The *STRIPES* method can be considered as an extension of the side-by-side test method and the one-stripe test.

Other methods are possible, e.g. a sliding change from one quality setting to another one, one-stripe with same crops of the video, or a temporal change of the two resolutions. In small expert tests we evaluated these methods, and concluded that the last mentioned method – a temporal switch – was the most promising one.

In the temporal switch method *TEMP* the quality levels are switched back and forth over time, e.g. the first quality of the video is shown and later the second one, and so on. In Figure 2 an example is shown. We use the same color scheme that we use in the *STRIPES* method. Every two seconds we change the stimulus. Our *TEMP* method is a specialized version of the ITU-R BT.500-13 [13] method without the possibility to manually change the stimuli. We decided to use a forced stimulus change to have fever interactions of the participants during the test and to be more comparable with the *STRIPES* method, where only the rating is required as interaction.

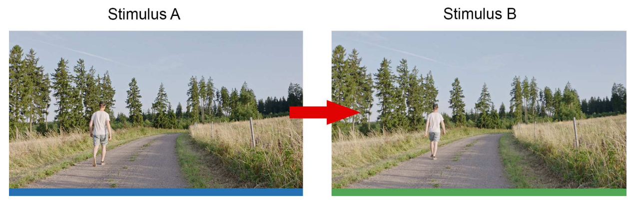

Figure 2. *Example of the used TEMP comparison method.*

In both test methods we ask which (blue or green) video part has the higher video quality.

To collect and automatically present all the test results, we use the new version of our rating software AVRate [19], referred to as AVRateNG[1]. In the employed instance of AVRateNG, the stimuli are presented in random order, providing the rating form in a web-browser. To ensure that a participant will not make decisions based on the provided color bars, we decided to repeat each stimulus at another random location in the test with the opposite color mapping.

Used videos and technical setup

We use 20 different source videos in both test cases. Each source video has a duration of 10 seconds; color subsampling 4:2:2, a resolution of 3840x2160 and a framerate of 60 frames per second. We selected the videos from publicly available uncompressed footage. 10 out of 20 videos are from harmonic.com [9], one sequence from big buck bunny (blender.org [5]), one from the BennuProRes animated video [22] and additional 8 sequence are self recorded. For the recordings we used a Sony PXW-FS-7 in combination with a Sony SELP28135G lens. All sequences are selected to satisfy a wide spatial and temporal complexity (equally low, mid, high classified SI/TI values based on [16] measurements, using our publicly available implementation[2]).

For up- and down-scaling we use the Lanczos-3-algorithm, because this algorithm ensures a better quality than simpler (e.g. bi-linear) scaling algorithms [20]. The Lanczos scaling algorithm is already included in FFmpeg[3]. For both methods, *STRIPES* and *TEMP*, we use FFmpeg to perform the split, add the stimulus-indication bars and/or temporal changes.

Our test setup uses the Panasonic VIERA TX-65CXW804 65 inch screen that was calibrated based on ITU-R BT.500-13 [13]. To ensure a smooth play-out, the test PC uses a Nvidia Ge-Force GTX 970 graphics card connected to the screen of 3840x2160 resolution (UHD-1) via HDMI 2.0a. Furthermore, the

[1] see `https://bit.ly/2QlCGft`
[2] `https://bit.ly/2oXxQIN`
[3] `http://ffmpeg.org/`

screen has a native framerate of 60 frames per second. The viewing distance of our subjects was $1.5 \cdot$ screen height. We did not include other viewing distances due to the fact that the pre-tests already showed that a higher distance will lead to nearly no visual difference of UHD and HD videos.

In addition, based on pre-tests we decided to extend our HD vs UHD test to UHD vs 900p, UHD vs 720p. We decided to include lower resolutions to ensure that participants are not annoyed during the test. For example, we observed in the pre-tests that it is really hard to distinguish between UHD and HD for mostly all the contents.

Results of both conducted test methods

In total 60 participants conducted the tests. In the first test we used only the *STRIPES* method, and in the second test only the *TEMP* method. Before the test started, the participant passed a Snellen-Charts visual acuity test.

For each video we calculated the probability for all users if they were able to recognize UHD correctly or not, referred to as UHD recognition rate. For all trials, we calculate for all users how often UHD was correctly identified for a given source video, and divide the final count by the number of users.

In both tests we found out that participants were able to recognize UHD well in comparison with the 900p and 720p resolution. In general in these cases the *TEMP* method was especially well suited for UHD recognition compared to more recognition errors in the 720p/900p case for *STRIPES*. In general for the comparison of UHD with 720p/900p resolutions the users mostly recognized the UHD video correct with both methods.

However, our main goal of the conducted test was to compare UHD with HD.

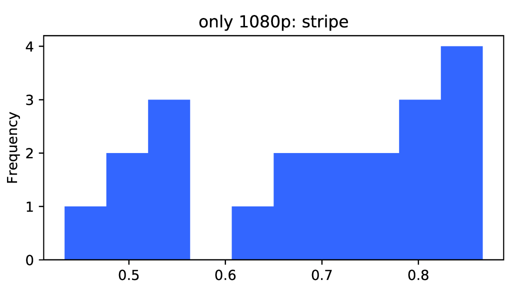

Figure 3. *Distribution of UHD recognition rate for 1080p case – STRIPES setup.*

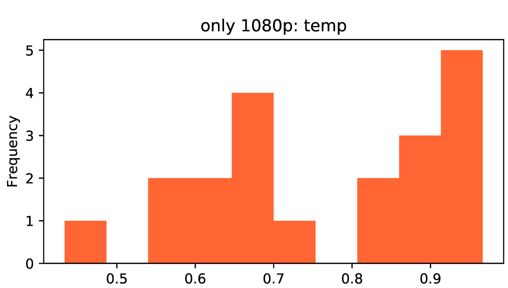

Figure 4. *Distribution of UHD recognition rate for 1080p case– TEMP setup.*

In Figures 4 and 3 the distributions of the final UHD recognition rate of both used test methods are shown. It can be observed, that in case of the *TEMP* method, more videos were correctly recognized as UHD. Also, for the wrongly classified videos the *TEMP* method seems better. In Figure 4 and 3 a possible separation of video contents can be observed, e.g. in case of the videos shown with the *TEMP* method with more than 0.8 recognition rate are "easy" to recognize as UHD and the remaining videos are "hard" to identify. The situation is similar for the *STRIPES* method, where videos with a UHD recognition rate lower than 0.6 are "hard" to classify. It can be concluded that whether UHD is recognized is mostly content depended. This confirms our assumptions that there are some videos where nearly no benefit of watching a UHD video in comparison to the HD version.

Both methods show similar results, however comparing with the low resolution cases (720p and 900p) the *TEMP* is a bit more suitable than the *STRIPES* method. This can be explained by the fact that in the *STRIPES* method not the full video in both resolutions is shown and that users still focus on the borders between two quality levels.

Based on the conclusions of our test, we decided to use the *TEMP* method and the resulting UHD recognition rates as ground truth for our prediction system. Furthermore, we define that a video is "UHD recognizable" ($class = 1$) if it has a UHD recognition rate of more than 80%.

Prediction of UHD perceptibility

Considering that it is already for a human hard to distinguish if a video is in UHD or HD resolution, we focus in the following Section on a prediction system that can estimate automatically if a video is perceivable as UHD or not. First, we describe the used video features that are motivated based on perceptual assumptions or features that are used in state of the art for video and image analysis.

Based on the calculated features a machine learning system can be trained. In general our formulated problem can be defined as a binary classification problem, where $class = 1$ means that humans can perceive a difference between UHD and HD and $class = 0$ is the case where no differences are perceivable.

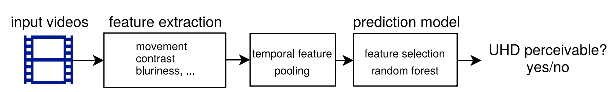

Figure 5. *General structure of our prediction approach.*

Our general structure of the approach is summarized in Figure 5. Staring from a set of videos we calculate different features, see Table 1. These features are considering temporal and image aspects, e.g. movement and spatial information. For each frame of one video we calculate different feature values. All collected values are finally aggregated – so called pooling. For our pooling method we use different statistic measurements. Assume that f is such a per frame estimate feature vector. For f we calculate: mean value, standard deviation, skewness, kurtosis, inter quartile range, the last and first value of f. Furthermore, we split the values of f in 5 temporal groups and for each group we calculate mean and standard deviation. In total, we have a fixed number of 17 values for each of the input videos. After temporal pooling of all videos, we train a feature selection and random forest pipeline with our ground truth labels. For machine learning we use the scikit-learn python library [26], features are based on scikit-video[4] or open-cv[5]. Our software is written in Python 3[6].

Most important for our system are features, that are able to describe and measure the perceptual differences between UHD and HD. One advantage of UHD over HD is the increasing of details, however in videos with high motion these details cannot be perceived.

Features

We therefore grouped our developed features into two main categories. First category considers only pure image features – *img* – without knowledge of the surrounding frames catching the detail improvements of UHD. The second category focuses on movement and changes over time –*mov*– to describe temporal changes.

Table 1: Features that are used for prediction with marked sources

feature name	*img*/*mov*	source
contrast	img	own
blur	img	own
fft	img	[17]*
si	img	[16]
niqe	img	[21]
colorfulness	img	[10]*
tone	img	[3]*
saturation	img	[3]*
uhdhdsim	img	own
ti	mov	[16]
temporal	mov	own
blockmotion	mov	own
movement	mov	own
staticness	mov	own

In Table 1 all used features with the corresponding sources are summarized, not for all features are open source implementations available, those features that are re-implemented are marked with *. We will further describe our own implemented features in detail.

image aspects – img

Especially the spatial information is higher in UHD videos, therefore we use our implementation of the SI measure *si* that is based ITU-T Recommendation P.910 [16].

Furthermore, luminance differences are important factors for the human visual system to detect objects in scenes [6]. They can be evaluated using color histograms to measure contrast. For the *contrast* feature we use histogram equalization. We can estimate histograms of the uncorrected (before) and corrected image. Using the cumulative distribution function (CDF) of each histogram the average difference before and after correction can be estimated, we use this value normalized as feature value.

Another important factor for high quality images or videos is sharpness or blurriness. Generally it can be said the lack of sharpness in an image or video is to be equated with low quality.

As a blurriness measure *blur* we implemented the following procedure based on Laplacian variance. Each frame is converted to grayscale, then a bilateral filter is applied which functions as an instrument to prevent unwanted noise or blocking artifacts from being wrongly detected as edges. Lastly the frame is convolved with the 2D Laplacian filter kernel. For each frame we are calculation a blurriness score. Furthermore also the *fft* feature is a blurriness estimation feature [17].

We re-implemented features for *colorfulness*, *tone* and *saturation* from the image aesthetics research area, to include liking/aesthetics aspects of video frames in our model.

Also we include quality features, e.g. *niqe* using the scikit-video implementation [21].

As last feature *uhdhdsim* we measure the psnr-similarity of a video frame compared to a down-scaled version of the frame, to calculate psnr values a backward up-scaling to UHD resolution is required. The *uhdhdsim* feature is an indication of information loss in downscaling, assuming that e.g. a HD frame that is rescaled to UHD would get a high score.

movement aspects – mov

To handle motion perception in our video sequences, we implemented several features. The simplest feature is the *ti* calculation that is based on ITU-T Recommendation P.910 [16].

Another simple method for movement characterization is our *temporal* feature, that is just the RMSE of the current and previous played frame. Furthermore, similar to codecs, we estimate *blockmotion*, based on the blockmotion implementation of scikit-video with the SE3SS method. We use 10% of the video height as blocksize for blockmotion estimation, after extraction the similar blocks, we count how often a block is moved left, right, top, down, or not. These counted values are our feature values.

As an addition we implemented a *movement* feature that uses open-cv background subtraction based on [30, 31]. We subtract the background mask from the frame and then calculating the sum of the remaining foreground pixels. As measure, we implemented the ratio to the total number of pixels in one image. The background re-movement ensures that, e.g., foreground objects are more considered, this is comparable to a typical viewing of a user.

As last feature, we observed that some videos have a huge part of *staticness*. For this feature we simply calculate a mean frame based on all currently played frames, if the video is mostly static the estimated mean frame includes a lot of spatial information. As final feature value we calculate the SI measure of the current mean frame.

Prediction results

Considering that for each video we calculate and pool a lot of features it is clear that we also need to validate our used features. For a simple validation we use a synthetic dataset and train our system. After verification of our introduced features, we use our UHD-recognition classification based on the UHD recognition rates of our *TEMP* test. We will conclude with a short discussion of our experiments.

[4] http://www.scikit-video.org
[5] https://opencv.org/
[6] https://www.python.org/

Synthetic dataset

For a better feature validation we use a synthetic dataset. It consists of 36 different video segments (with the same conditions that we used in our comparison test: UHD-1 resolution, 60 fps, 4:2:2 chroma subsampling). For each of the UHD videos we created a down-scaled HD version using the Lanczos scaling algorithm. Furthermore, for feature calculation we up-scaling the HD videos again to UHD, so called fake UHD videos. Using this approach, we know already possible classes, all down-scaled – fake UHD– videos are not UHD recognizable. The original videos are all UHD recognizable. However, such a trained system cannot be used to validate human perception, because we also include probably videos that are per definition not UHD recognizable. We selected this setup only for feature validation.

As parameters for our machine learning model, we use 10 decision tress, a feature selection threshold of "0.01*mean" and we perform a 10 fold cross validation. Other parameters are default parameters of scikit-learn.

Table 2: Classification results for the synthetic experiment.

class	precision	recall	f1-score	support
0	0.77	0.92	0.84	36
1	0.90	0.72	0.80	36
avg / total	0.83	0.82	0.82	72

In Table 2 our classification results are presented. For both classes precision, recall, f1-score are quite high, larger than 0.8. We further get an accuracy score of rounded 0.82 for our system. Meaning that our system is able to predict approximately 80% of all cases correctly. This result is similar to our results from the subjective test.

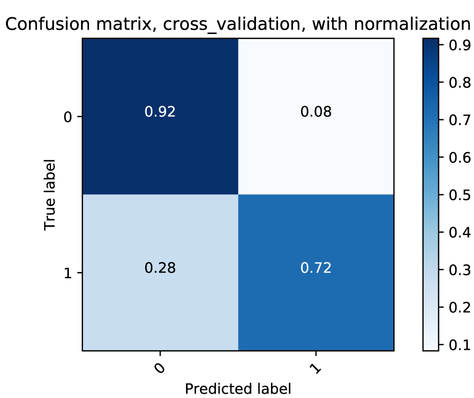

Figure 6. *Confusion matrix of the synthetic prediction system.*

The confusion matrix is shown in Figure 6, only a few cases are miss-classified, a more detailed analysis why there is this miss-classification will be done in future work. We verify that our features are able to be used in a prediction system for UHD-recognition classification.

As next step, we focus on the estimated classification based on the results of the perception test.

Perception test

We defined a video as UHD-recognizable ($class = 1$) if a UHD-recognition rate of more than 80% was achieved in the *TEMP* test. Similar to the synthetic prediction experiment, we train a random forest model with feature selection using the same parameters in a 10-fold cross validation approach. The model parameters are chosen to ensure a comparison of both methods.

Table 2: Classification results for the experiment with the data from the perception test.

class	precision	recall	f1-score	support
0	1.00	0.30	0.46	10
1	0.59	1.00	0.74	10
avg / total	0.79	0.65	0.60	20

In Table 3 the results are summarized. It can be seen that our prediction system is not as good as in the synthetic case. We also calculated the accuracy of our system, and we get a value of round 0.65. This value is not outstanding, we also checked the confusion matrix, and the false positive rate is quite high. In approximately 30% of all cases a not UHD recognizable video is classified correctly.

We showed that our features are able to predict differences of UHD and HD. Using the subjective data, we are able to show that such a trained system can be used in real world applications. Even if the classification performance is not amazing, assuming that 30% of all not as UHD recognizable videos are correctly classified. Considering, e.g. streaming applications where such a system can be used to ensure that some videos are not streamed via 4K/UHD, due to the missing benefit of UHD. Furthermore, we were not considering the real world viewing distances, that are larger than 1.5 times display height.

Another application of our prediction system in case of using the synthetic dataset, is that 4K contents or HD up-scaled – fake UHD– content can be distinguished. This is for example important for content providers that buy 4K/UHD videos (similar to [17]).

Conclusion and Future Work

Even if display technologies are increasing more and more the possible resolution, the problem of UHD recognition is still a challenging task. We started with the observation that UHD content will be recognized in real world applications only in a few cases. We analyzed how hard it is for humans to recognized UHD content in comparison with HD. Therefore, we conducted a subjective video test, where we evaluated two different comparison methods. The *STRIPES* and *TEMP* methods were compared. We can conclude that the *TEMP* method, that uses a temporal switch between UHD and a lower resolution is more suitable for our UHD recognition problem. We found out that it is a challenging task for humans to distinguish between UHD and HD for uncompressed video sequences, even with the recommended 1.5 times height of the display viewing distance. Furthermore, we introduced computer vision based features, that can be used in a prediction system to automatically predict if there is a benefit of UHD. In one experiment we evaluated the meaningfulness of our features using a synthetic dataset. We found out that our system is quite well able to classify the synthetic videos. As another evaluation experiment we used the results from our *TEMP* method. It can be shown that our system also works for the subjective classification, however the results can be improved, e.g. considering more features or conducting a second comparison test. Even if the results are not outstanding, such a prediction system can be used in various applications, e.g. in source video to check if it is a true UHD video or not. Or to reduce streaming of videos that have no benefits of UHD. Future work will focus on fine tuning our prediction system.

Acknowledgments

This research work was partially funded by Telekom Innovation Laboratories, Deutsche Telekom AG, Germany.

References

[1] P. Akyazi and T. Ebrahimi. "Comparison of Compression Efficiency between HEVC/H. 265, VP9 and AV1 based on Subjective Quality Assessments". In: *10th International Conference on Quality of Multimedia Experience (QoMEX)*. CONF. 2018.

[2] AOM. *AV1 Codec*. [Online; 07.09.2018]. 2018.

[3] T. O. Aydın, A. Smolic, and M. Gross. "Automated aesthetic analysis of photographic images". In: *IEEE transactions on visualization and computer graphics* 21.1 (2015), pp. 31–42.

[4] K. Berger et al. *Subjective quality assessment comparing uhd and hd resolution in hevc transmission chains*. LUNAM Univerité, Université de Nantes, IRCCyN UMR CNRS 6597, Polytech Nantes rue Christian Pauc BP 50609, 44306 Nantes Cedex 3, France, 2015.

[5] Blender Foundation. *Bick Buck Bunny Distribution*. URL: `http://distribution.bbb3d.renderfarming.net/video/png` (Accessed: 07/07/2017).

[6] S. T. Chung and G. E. Legge. "Comparing the Shape of Contrast Sensitivity Functions for Normal and Low VisionComparing CSF for Normal and Low Vision". In: *Investigative ophthalmology & visual science* 57.1 (2016), pp. 198–207.

[7] Cisco. *Whitepaper: Cisco Visual Networking Index:Forecast and Methodology, 2015-2020*. 2015.

[8] S. Göring, J. Skowronek, and A. Raake. "DeViQ – A deep no reference video quality model". In: *Electronic Imaging, Human Vision Electronic Imaging* (2018).

[9] Harmonic. *Free 4K Demo Footage - Ultra HD Demo Footage*. URL: `https://www.harmonicinc.com/4k-demo-footage-download/` (Accessed: 07/07/2017).

[10] D. Hasler and S. E. Suesstrunk. "Measuring colorfulness in natural images". In: *Human vision and electronic imaging VIII*. Vol. 5007. International Society for Optics and Photonics. 2003, pp. 87–96.

[11] ITU. *Comparison of quality assessment of UHD and HD videos at two different viewing distances*. International Telecommunication Union. Geneva, 2017.

[12] ITU-R. *HDTV and UHDTV including HDR-TV test materials for assessment of picture quality*. Vol. Report ITU-R BT.2245-3. International Telecommunication Union. Geneva, 2017.

[13] ITU-R. *Methodology for the subjective assessment of the quality of television pictures*. Vol. Recommendation ITU-R BT.500-13. International Telecommunication Union. Geneva, 2012.

[14] ITU-R. *The present state of ultra-high definition television*. Vol. Report ITU-R BT.2246-6. International Telecommunication Union. Geneva, 2017.

[15] ITU-T. *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*. Serie P: Terminals and subjective and objective assesment methods. Vol. Recommendation ITU-T P.913. International Telecommunication Union. Geneva, 2016.

[16] ITU-T. *Subjective video quality assessment methods for multimedia applications*. Serie P: Telephone Transmission Quality, Telephone Installations, Local Line Networks. Vol. Recommendation ITU-T P.910. International Telecommunication Union. Geneva, 2008.

[17] I. Katsavounidis, A. Aaron, and D. Ronca. "Native resolution detection of video sequences". In: *Annual Technical Conference and Exhibition, SMPTE 2015*. SMPTE. 2015, pp. 1–20.

[18] D. K. Krishnappa, D. Bhat, and M. Zink. "DASHing YouTube: An analysis of using DASH in YouTube video service". In: *Local Computer Networks (LCN), 2013 IEEE 38th Conference on*. IEEE. 2013, pp. 407–415.

[19] P. Lebreton, M.-N. Garcia, and A. Raake. "AVRate: An Open Source Modular Audio/Visual Subjective Evaluation Test Interface". In: *8th International Workshop on Quality of Multimedia Experience (QoMEX)* (2016).

[20] J. Li et al. *Comparing upscaling algorithms from HD to Ultra HD by evaluating preference of experience*. LUNAM Université, Université de Nantes, 2014.

[21] A. Mittal, R. Soundararajan, and A. C. Bovik. "Making a "completely blind" image quality analyzer". In: *IEEE Signal Processing Letters* 20.3 (2013), pp. 209–212.

[22] NASA. *Ultra high definition video gallery*. 2017. URL: `https://www.nasa.gov/content/ultra-high-definition-video-gallery` (Accessed: 12/04/2017).

[23] Netflix. *4K Support*. [Online; 07.09.2018]. 2018.

[24] Netflix. *VMAF 4K included*. [Online; 07.09.2018]. 2018.

[25] K. Noland and L. Truong. "A survey of UK television viewing conditions". In: *BBC Research & Development White Paper* 287 (2015).

[26] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *JMLR* 12 (2011), pp. 2825–2830.

[27] samsung. *Future of display*. [Online; 07.09.2018]. 2018.

[28] R. Sotelo et al. *Subjective video quality assessments for 4K UHDTV*. Universidad de Montevideo and University of Cagliari (UNICA), 2017.

[29] G. Van Wallendael et al. *Perceptual Quality of 4K-resolution video content compared to HD*. Ghent University - iMinds - Data Science Lab, Belgium and Free University Brussels - iMinds - SMIT, Belgium, 2016.

[30] Z. Zivkovic. "Improved adaptive Gaussian mixture model for background subtraction". In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 2. IEEE. 2004, pp. 28–31.

[31] Z. Zivkovic and F. Van Der Heijden. "Efficient adaptive density estimation per image pixel for the task of background subtraction". In: *Pattern recognition letters* 27.7 (2006), pp. 773–780.