BVI-SynTex: A Synthetic Video Texture Dataset for Video Compression and Quality Assessment

Angeliki Katsenou, *Member, IEEE,* Goce Dimitrov, Di Ma, *Member, IEEE,* and David Bull, *Fellow, IEEE*

***Abstract*—Highly textured video content is challenging to compress since the bit-rate to video quality trade-off is high and complex perceptual masking influences performance. Test datasets that cover a wide range of texture types are thus important for codec evaluation, but few exist. In order to study the properties of video texture, this paper introduces a Synthetic video Texture dataset (BVI-SynTex) that was generated using a Computer-Generated Imagery (CGI) environment. It contains 196 sequences clustered in three different texture types and offers the capability of being able to generate many versions of the same scene with different video parameters. It therefore provides a flexible basis for studying the influence of texture type and parameters on video compression and perceived video quality. A thorough validation and comparison of BVI-SynTex with similarly textured natural video content is performed. The comparisons show that BVI-SynTex exhibits a comparable coverage over the spatial and temporal domain and that it produces similar encoding statistics to real video datasets. A subset of the BVI-SynTex dataset was selected to perform a subjective evaluation of compression using the MPEG HEVC codec.The results show the impact of the content parameters to both the compression efficiency and the perceived quality. The publicly available BVI-SynTex dataset contains all source sequences, the objective and subjective analysis results, providing a valuable resource for the research community.**

***Index Terms*—Video Texture, CGI Video Dataset, Video Content Analysis and Compression, HEVC.**

I. Introduction

THE technologies to support emerging video formats are rapidly advancing, creating an enormous demand for higher compression rates while preserving quality and delivering a satisfying experience to the end users [1]. Cisco reports that, by 2022, 82% of the consumer traffic will be video data [2], with Internet video traffic growing fourfold from 2017 to 2022. This implies a massive demand and an ever-increasing requirement for efficient video coding solutions.

It is well known that the trade-off between compression efficiency (or bit-rate) and quality is content related. This poses the hypothesis that, if we can better comprehend the relationship between content properties and compression, then we will be able to design more efficient video coding solutions. Highly textured video, particularly scenes that include dynamic textures, are some of the most challenging in this respect. Several researchers have presented approaches that enhance compression of textured content either via synthesis [3]–[5], by adopting different quantization strategies for textured areas, e.g. [6], or by proposing better local motion compensation techniques for those areas, e.g. [7], [8]. All of these approaches however base their studies on a limited number of video sequences that are publicly available, allowing only a partial exploration of the full video texture space.

It has been widely reported [1], [9], [10] that different textures exhibit different compression performance in terms of perceived quality. In order to fully understand and model the relationship between video texture type and compression, we need a dataset that provides within-sequence homogeneity and flexibility and diversity in terms of content (e.g. different levels of coarseness, or object motion) and acquisition (e.g. camera motion) parameters. Ideally, we would require full parametric control over the acquisition process, but this is impossible for natural video.

A. Related Work

1) Existing Video Datasets: Since the Video Quality Expert Group (VQEG) released one of the first subjective video quality databases, VQEG FR-TV Phase I [11], numerous video databases for either video compression or quality evaluation have been generated, e.g. VQEG-HD [12], IRCCyN/IVC 1080i [13], EPFL-POLIMI [14], LIVE [15], BVI-HFR [16], BVI-HD [17]. Typically, a video database for testing compression performance and evaluating quality assessment methods should comprise diverse content that is representative of everyday viewing consumption. For the latter reason, Zhang et al. [17] compare BVI-HD to BBC Redux [18].

In this context, there are few freely available video datasets that are designed to represent homogeneous textured scenes BVI-Texture [9], DynTex [19], and HomTex [20], but none that offer flexibility in texture parameters. DynTex [19] is one of the most cited datasets on texture related research. It has been generated for dynamic texture classification and recognition research purposes and contains 650 PAL resolution dynamic video sequences (spatial resolution 720×576 at 25 frames per second (fps)) with a wide range of texture types. However, the video parameters (spatial resolution and frame rate) are obsolete compared to current requirements and thus are not suitable for conducting subjective experiments. HomTex [20], [21] and BVI-Texture [9] datasets were generated with the aim of understanding and analysing the properties of video textures and their associated coding performance. The BVI-Texture dataset [9] contains 20 video sequences with Full High Definition (FHD) resolution (1920×1080) at a frame rate of 60 fps and 8 bits depth. This dataset was used to test

All the authors are with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, UK.

The work presented was supported by the Leverhulme Early Career Fellowship ECF-2017-413 (awarded to Dr. A. Katsenou) and by the Engineering and Physical Sciences Research Council (EPSRC), grants #: EP/L016656/1 and EP/M000885/1.

HEVC compression efficiency and its perceptual quality versus bit rate performance. It has been also used to develop video quality assessment metrics and future video coding algorithms, e.g. [6], [8]. Although the BVI-Texture dataset satisfies modern video specifications, the small number of available sequences is not sufficient for an extensive analysis of texture properties. It is also heterogeneous in terms of the different textured and non-textured areas it contains. Another publicly available dataset focusing on video textures derives from the former two. HomTex [20] contains 120 video textures that are manually cropped and selected from the DynTex (101 out of 120) and BVI-Texture (19 out of 120) datasets. This results in spatially homogeneous video textures at a 256×256 resolution and a frame rate of 25 fps and 60 fps (depending on whether the sequence originates from DynTex or BVI-Texture).

Although the aforementioned datasets have been valuable to the research community, they are not sufficient to fully study video textures for compression related research. The main reasons for this are:

- The number of sequences is small, thus insufficient to allow full exploration and understanding of the video parameter space that includes an enormous number of potential combinations of parameters related to video content (e.g. spatial patterns, colourfulness, complex local motion patterns).
- None of the datasets provide parametric variability of the same content (different camera motions, frame rates, spatial resolutions, etc.) to support studies of different compression performance that these parameters induce.
- DynTex and HomTex include content with obsolete spatial and temporal resolutions that are not suitable for contemporary subjective studies.
- BVI-Texture is FHD and has been subjectively evaluated at different compression levels. In spite of that, it is hard to draw conclusions for the different textural content itself as the majority of the sequences are heterogeneous. Thus, the subjects who evaluate the sequences as a whole are unconsciously performing some type of spatial and temporal pooling.

Furthermore, due to the enormous number of potential combinations of parameters related to video content (e.g. spatial patterns, colourfulness, complex local motion patterns) and acquisition (e.g. frame rate, shutter angle, camera position), the cost in person-hours of capturing multiple variants is prohibitive. Also the randomness of some textural content (e.g. foliage, falling leaves), makes capturing an identical scene with different video parameters unfeasible.

2) Synthetic Video Datasets: The use of synthetic data is a common practise in many research areas, especially in situations where real data may be difficult to acquire, due to budget, time or privacy concerns. For example, in computer vision, synthetic data are used for scene understanding [22] or object recognition [23] and have been proven reliable and useful especially for training neural networks [24], [25]. In the field of video compression, there is one synthetically-generated dataset [26] (to the best of our knowledge), designed to simulate a multi-lens stereoscopic video system with the aim to be used for multi-view compression, streaming, or other computer graphics related research.

The use of CGI content is also common practise in cinematography, not only for animated content, but also for movies. It is typically combined with real scenes, especially when special effects are to be utilised. In practice this means that a large amount of data that video technologists are compressing is actually synthetically generated. Drawing inspiration from this fact and the recent use of synthetic data in other research fields, we propose the generation of a synthetic video texture dataset for video compression purposes. Such an approach has the benefit of using parameterised models for the production of the synthetic content. This translates to datasets that are reproducible and can densely cover the video parameter space.

B. Contribution

The lack of publicly available of homogeneous textured video content, along with the prohibitive cost and weakness of capturing natural textured content with identical features, are limiting factors that prevent us from drawing robust conclusions about the perception of compressed video textures with varying features (such as motion). Furthermore, it makes it harder to develop robust content-driven algorithms for compression. To address this shortfall, we have created a publicly available synthetic texture video dataset, BVI-SynTex, that contains a diverse set of 5 sec FHD video sequences captured at 60 fps and 8 bits of depth [27]. Building upon our previous work in this area [27], we further extended it and this paper presents the following contributions:

i. the requirements established for the design of the synthetic dataset are outlined;
ii. the study of coding performance is extended by applying clustering on the encoding statistics and by associating those to texture types;
iii. the relationship between common content parameters, such as granularity and motion levels, and encoding statistics is explained;
iv. a subjective study on a subset of BVI-SynTex sequences is performed in order to evaluate the perceived quality for different types of texture and to study the effect of different content parameters, such as granularity and motion levels on video textures;
v. a statistical analysis of the collected opinion scores is performed following the recommendations of ITU [28];
vi. state-of-the-art objective quality assessment metrics on the subset of BVI-SynTex that was subjectively evaluated are calculated, the rate-quality curves for the different compression levels are plotted and their correlation to mean opinion scores is examined;
vii. the publicly available BVI-SynTex dataset is updated by including the objective and subjective results on the encoded subset of sequences.

The remainder of this paper is organised as follows. Section II characterises the BVI-SynTex video dataset, provides an analysis of BVI-SynTex coding statistics and its comparison to real video texture datasets and studies the content parameters

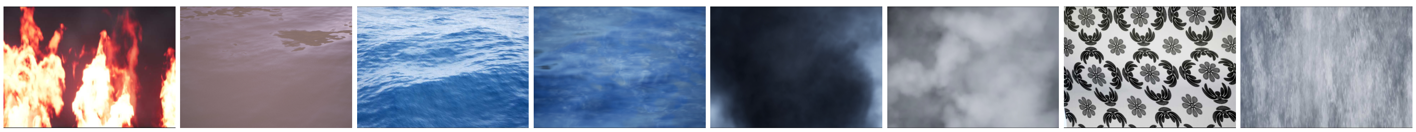
TABLE I: Sample frames from the three video texture types of the BVI-SynTex dataset and a short description of the model parameters per texture type.

Type of Texture: Static

S1. Brick Cut | S.2 Brick Hewn | S.3 Ceramic Tile | S.4 Clay Brick | S.5 Cobble Stone | S.6 Fabric | S.7 Hexagon Tile | S.8 Panel

Model Parameters: three levels of camera distance; three different levels of speed of moving camera.

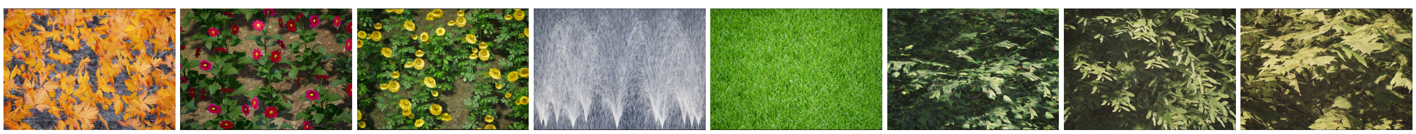
Type of Texture: Dynamic Continuous



C.1 Fire | C.2 Flowing River | C.3 Ocean Wave | C.4 Pond Water | C.5 Smoke | C.6 Steam | C.7 Swing Fabric | C.8 Waterfall

Model Parameters: spread width of fire; wave speed; intensity of smoke; initial velocity of smoke and steam; different wind direction of fabric; swing speed; density of waterfall; granularity.

Type of Texture: Dynamic Discrete



D.1 Falling Leaves | D.2 Flower I | D.3 Flower II | D.4 Fountain | D.5 Grass | D.6 Tree I | D.7 Tree II | D.8 Tree III

Model Parameters: three levels of wind speed; three levels of density of fountain; three levels of spread width of fountain; three levels of camera distance for the flowers, the grass and the trees.

and the relation to coding performance. Section III reports the results of the subjective evaluation of a subset of the dataset, while Section IV the results of the objective quality assessment. Conclusions and suggestions for future work are then presented in Section V.

II. BVI-SynTex - A Synthetic Video Texture Dataset

The synthetic video texture dataset BVI-SynTex is presented in this section. The main purpose of this dataset is to provide a publicly available CGI-based dataset of homogeneous textures, created with parameterisable models.

In order to evaluate the video texture compression performance and its perceived quality, the new dataset was designed to meet the following requirements:

- The reference video content that will be the basis for the different variations of the content should be interesting and representative of real videos.
- The synthetic content should be as realistic as possible in order not to influence viewer experience during subjective quality evaluation.
- There should be a reasonable variation of the content parameters to allow the study of their interactions.
- The acquisition parameters should be the same for all videos within a class, to prevent any resulting differences of the same content, e.g. same speed of moving camera.
- The scene capture process should be as realistic as possible, simulating camera parameters such as shutter angle.
- The video content should be homogeneous both spatially and temporally to allow the study of video textures without any bias.
- The motion patterns of the dynamic textures should be stochastic and similar to real video textures.

A. Description of the Dataset

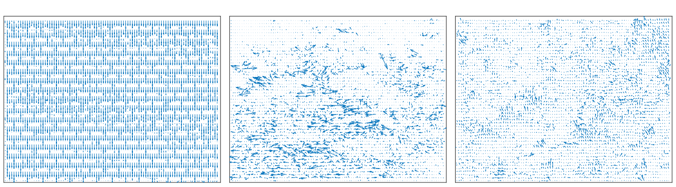
BVI-SynTex was created using a widely used CGI tool, Unreal Engine 4 (UE4) [29]. UE4 is a C++ based tool employed by the games industry and also by movie makers. Recently, it has also been used for research purposes, for example for the analysis of stereo vision [22] and studying virtual reality (VR) [30]. UE4 offers a variety of assets that include models for different scenes and objects. For each of these models there is a set of different parameters that can be adjusted, for example the amplitude or the speed of a water wave.

UE4 also includes universal parameters for capturing video that simulate a real camera, such as resolution, frame rate, shutter angle, and viewing angle. It is notable that, based on these parameters, it delivers content that has similar properties to real content, e.g. motion blur. Hence UE4 can be controlled to capture similar texture content using different parameters, e.g. frame rate, shutter angle or camera position. We carefully selected models that resemble datasets that already exist (e.g. HomTex) and typically exist as part of video scenes, for example foliage, grass or water. A further criterion was to include video textures that fitted our taxonomy developed previously [19], [20], [31]. In those works, video textures are classified into three types, static (e.g. a camera panning over a still scenery) and dynamic continuous (e.g. a scene of ocean waves) and dynamic discrete (e.g. a scene of moving foliage). In this paper, we followed the same definitions to generate synthetic video textures.

BVI-SynTex is open and publicly available on-line at [32]. It consists of 196 homogeneous video sequences and, in Table I, we illustrate sample frames of the generated video sequences grouped according to texture type. For video generation, we used typical parameters like those from other datasets: a spatial resolution of 1920×1080, a frame rate of 60 fps, a $180°$ shutter angle, 8 bit depth and YUV420 colour sampling. Each scene has a wide range of associated parameters that can be flexibly modified.

The parameters we used for the different models are also reported in Table I. These parameters are related to the motion and granularity levels of the different textures. In UE4 these parameters are defined within context, for example instead of defining the granularity level for sequence containing smoke, there is a parameter called intensity that regulates it. Also, parameters related to texture granularity and the amount of motion were designed to be uniform for the different versions of the videos, e.g. in three levels low-medium-high (wherever that was applicable). In Fig. 1 an example of the variations of the content parameters is illustrated. It is important also to mention that the motion patterns in the dynamic textures are based on stochastic models and are different for the different spatial patterns. We provide examples of the optical flow (OF) fields extracted using Farnebäck's method [33] in Fig. 2. Last, we note that the video acquisition parameters are the same for all videos.

Fig. 1: Example of the scene S.5 Cobble Stone at three different levels of granularity: low, medium and high, from left to right, respectively.



(a) Clay Brick. (b) Pond Water. (c) Tree I.

Fig. 2: Examples of OF fields for the different texture types.

B. Content Feature Analysis

In this section, the content of the dataset is characterised using a method employed by many other researchers in the field [9], [17], [34]–[37]. It involves computing the coverage of BVI-SynTex over four low-level descriptors as introduced by Winkler in [38]. This method was introduced with the aim to ensure fair and sufficient scrutiny when benchmarking new and existing algorithms. According to this method, we compared the coverage of content features extracted from BVI-SynTex with the BVI-Texture and HomTex datasets. The content features employed are: Spatial Information (SI) as an indicator of the variety of scenes (edge energy), the Motion Vectors (MV) and the Temporal Information (TI) as indicators of the variety of motion, and last the Colourfulness (CF) as indication of the variety of colours and contrast.

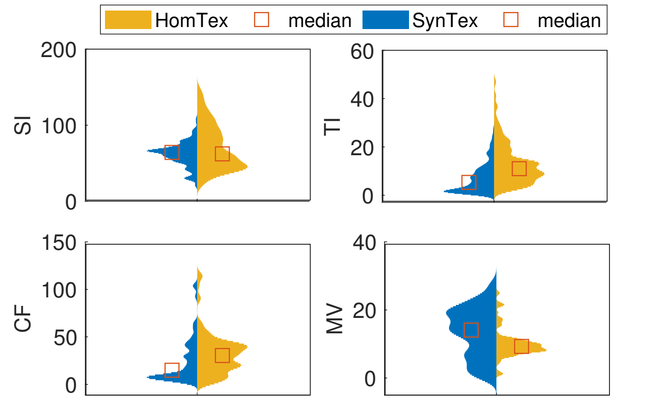


Fig. 3: Distribution plots of the low-level content descriptors of BVI-SynTex-256 and HomTex.

In order to perform a fair comparison of the content features with those of other (real) datasets, we must match at least their spatial resolution and frame rate. Since the spatial resolution of BVI-SynTex is 1920×1080 at a frame rate of 60 fps, it was directly compared with BVI-Texture [9] which has the same characteristics. However, BVI-Texture does not exclusively comprise of homogeneous textured scenes. Therefore, in order to compare BVI-SynTex with the only available homogeneous video texture dataset, HomTex [20], each sequence in BVI-SynTex was downscaled from 1920×1080 to 256×256 using Lanczos-3 filter [39], as implemented by FFmpeg [40]. This downscaled version will be referred to as BVI-SynTex-256.

In Fig. 3 the distributions of the content features of BVI-SynTex-256 and HomTex are depicted using violin plots in a side-by-side comparison. As can be seen, the distributions of the two datasets overlap in all four content descriptors with the median values. The distribution of SI in BVI-SynTex-256 is narrower but quite symmetric (median value is in the centre of the distribution) compared to HomTex that is skewed towards the lower SI values. This is expected as HomTex contains a wider variety of scenes compared to BVI-SynTex-256, which has only 24 scenes but many different versions of the same scene. For the same reason, CF is shifted towards the lower range and has a narrower range. TI is expected to have a lower range of values as BVI-SynTex-256 has a higher frame rate, thus the temporal consistency is higher. Additionally, we note that the motion patterns in BVI-SynTex-256 are generated using models and this ensures a uniform motion across all frames. Finally, it should be noted that the distribution of MV

TABLE II: The relative total coverage of BVI-SynTex, BVI-Texture, BVI-SynTex-256 and HomTex.

	BVI-SynTex	BVI-Texture	BVI-SynTex-256	HomTex
Rel. Total Coverage	0.34	0.35	0.40	0.33

for BVI-SynTex-256 is significantly wider than in HomTex. This is explained by one of the BVI-SynTex requirements that required for the same content three different levels of motion: low, medium and high, thus, the wider distribution of the MV descriptor.

To further compare the video datasets, we computed the relative total coverage based on the cube root of the convex hull of all source sequences in the normalized MV×SI×CF space, as in [38]. A higher relative coverage implies a better representation of the feature space with more variability across the different feature dimensions. The relative total coverage for BVI-SynTex is reported in Table II. BVI-SynTex-256 achieves a better relative total coverage compared to HomTex. On the 1920×1080 resolution comparison of datasets, BVI-SynTex has almost the same coverage as BVI-Texture. Also, compared to other video datasets that are reported in [38], BVI-SynTex's total relative coverage is around the mean value of the video datasets reported.

C. Coding Statistics for BVI-SynTex-256 and HomTex

To further validate the effectiveness of BVI-SynTex, we compare the coding statistics of BVI-SynTex-256 to those of HomTex [20], encoding both with HEVC HM16.2 at five quantization levels, QP=$\{22, 25, 27, 32, 37\}$ using the Random Access configuration [41], which is one of the most commonly used mode in practice in broadcasting and streaming applications [42].

First, despite the fact that the two data sets contain different textural patterns, we are showing in Fig. 4, examples of the resulting quality-rate curves of video textures from the compared datasets that appear similar both in spatial and temporal characteristics. It is clear that for all three video texture types, the real and synthetic example sequences have similar curves. Moreover, between the three texture types there is a noticeable difference in the compression efficiency. The static textures exhibit very high quality even at low bit rates, while the dynamic discrete sequences require a much higher number of bits per pixel (bpp) for the same quality. For example, to achieve a PSNR value of 36 dB, the static textures of Fig. 4 require less than 0.012 bpp, the dynamic continuous 0.025-0.04 bpp and the dynamic discrete 0.4-1.3 bpp.

Furthermore, in Table III a subset of the extracted coding statistics are presented (for the full list refer to our previous work [20]). We have selected the specific statistics as they are adequately representative for the comparison and validation purposes[1]. The employed statistics are reported in four general categories, two of which are related to the encoding decisions (i.e the prediction modes and the partitioning) and the other two categories are related to the impact of certain encoding decisions (i.e. the residual statistics and the motion vectors). In Fig. 5, a comparison of the distributions of the statistics mentioned in Table III of BVI-SynTex-256 and HomTex is presented.

[1]For more plots and comparisons on the encoding statistics of BVI-SynTex-256 and HomTex, please, refer to our previous work [27].

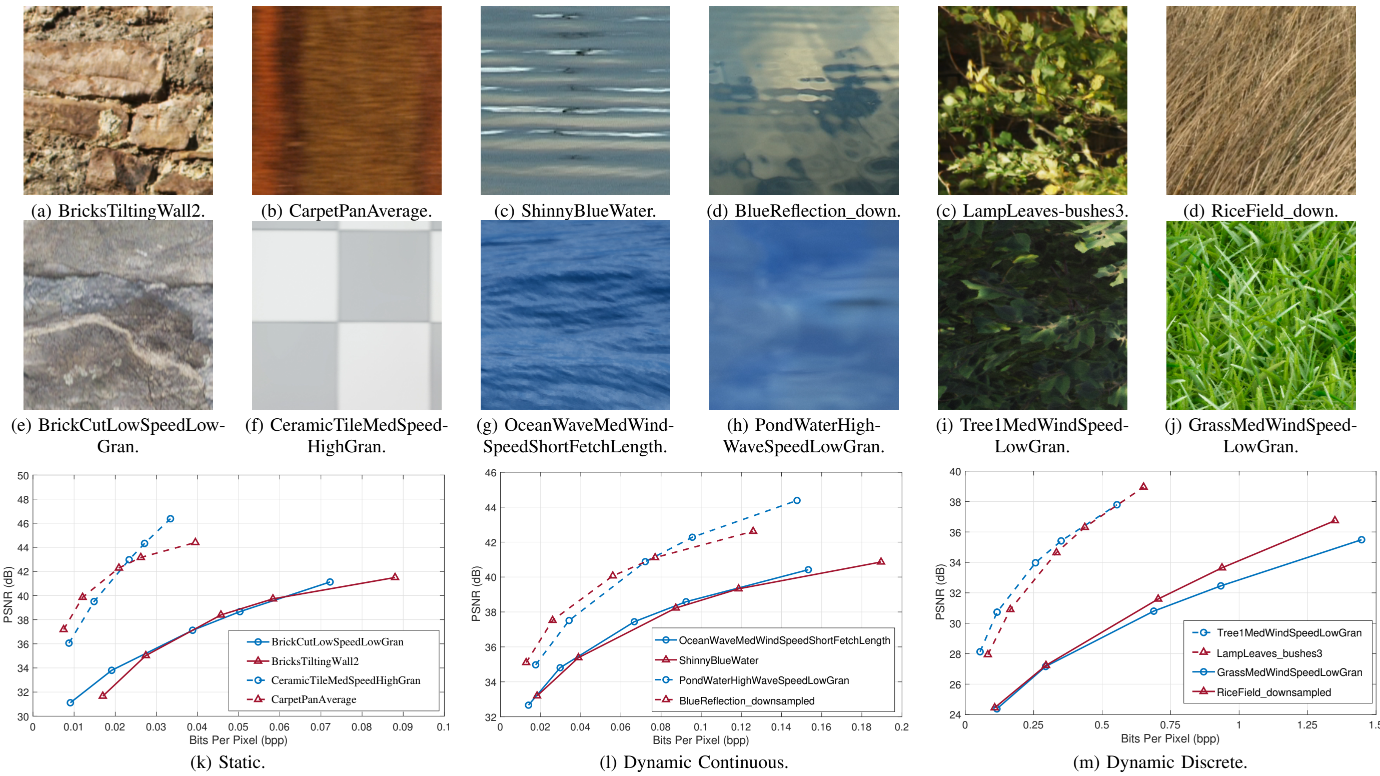
TABLE III: The subset of the presented extracted encoding statistics [20].

Category	Statistics
Prediction Modes	Average percentage of the prediction mode - Intra, Skip, Merge, Inter - selected for the Coding Unit (CU).
Partitioning	Average of the number of partitions per Coding Transform Unit (CTU).
Residual Statistics	Percentage of bits used to encode the residual signal. Average MSE of the reconstruction error. Average of the correlation between original and residual frames.
Motion Vectors	Average length of motion vectors.

1) Prediction Modes: A first observation from Fig. 5 (a)-(d) is that, for both BVI-SynTex-256 and HomTex, there are significant differences between the different types of video textures. Static textures exhibit a high percentage of Skip modes and a low percentage of Intra modes, mainly because they only have simple camera motion with a fixed movement direction [43]. Intra mode is also mainly used to encode dynamic textures. A noticeable difference between BVI-SynTex-256 and HomTex is in the Intra and Skip modes, particularly for the dynamic continuous textures. This is explained by the different acquisition parameters of BVI-SynTex, the impact of the downsampling process on BVI-SynTex-256 (greater than a 4 downsampling factor) and by the fact that many of sequences from HomTex suffer from noise, as also explained in [31]. Another reason that explains the differences in Skip statistics for the dynamic continuous textures is the frame rate. BVI-SynTex is captured at 60 fps which is higher than that of HomTex (the frame rate for most continuous textures in HomTex is only 25 fps). A higher frame rate results in a higher redundancy, thus more CUs can be skipped.

The Merge and Inter modes are very similar for both datasets, as can be seen in Fig. 5 (a) and (c). For the Merge mode a slightly higher median value is observed for the dynamic textures. This can be explained by the higher variety of granularity levels in BVI-SynTex-256 compared to HomTex. On the other hand, the small differences of the median values for the Inter mode are related to the variety of motion levels of BVI-SynTex-256 (low, medium, high). The effect of the different levels of motion on the Inter mode is also illustrated in Section II-E.

2) Partitioning: It can be seen from Fig. 5 (e) that the distribution of BVI-SynTex-256 is within the value range of HomTex with similar median values. It is also noticeable that for both datasets, different types of video textures also have different ranges of values. Static textures have the lowest number of partitions (medians of 2 and 4 partitions per CTU for BVI-SynTex-256 and HomTex, respectively). The highest number of partitions is observed for dynamic discrete textures. This is attributed to the higher density of high frequencies

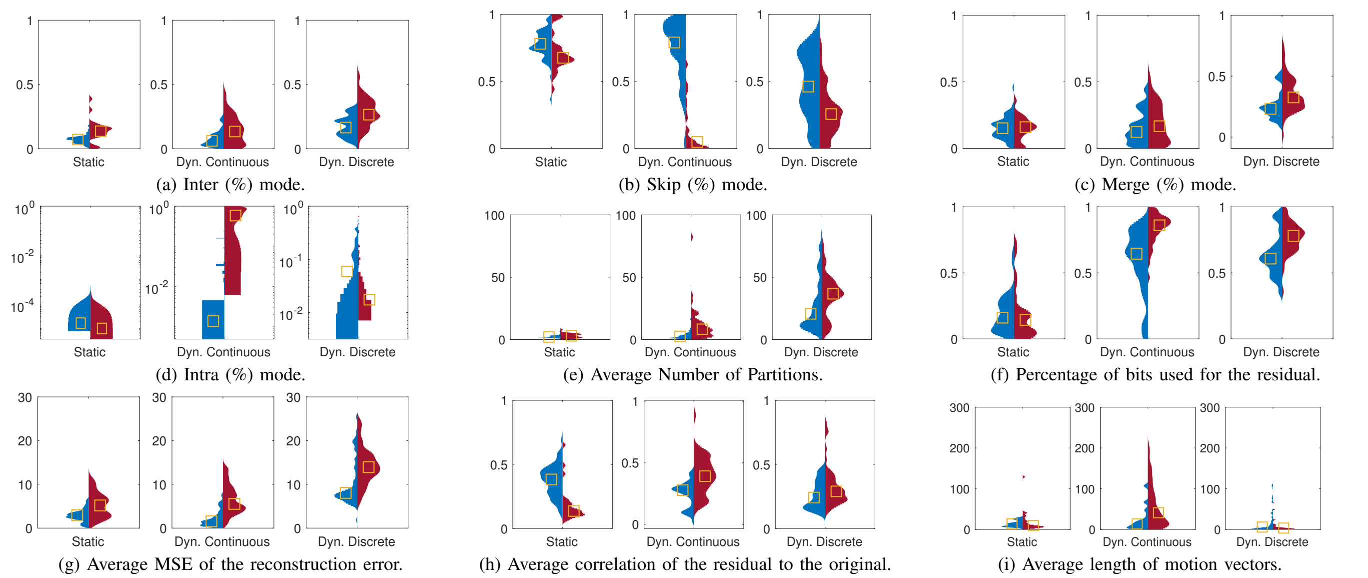


(a) BricksTiltingWall2. (b) CarpetPanAverage. (c) ShinnyBlueWater. (d) BlueReflection_down. (c) LampLeaves-bushes3. (d) RiceField_down.

(e) BrickCutLowSpeedLow-Gran. (f) CeramicTileMedSpeed-HighGran. (g) OceanWaveMedWind-SpeedShortFetchLength. (h) PondWaterHigh-WaveSpeedLowGran. (i) Tree1MedWindSpeed-LowGran. (j) GrassMedWindSpeed-LowGran.

(k) Static. (l) Dynamic Continuous. (m) Dynamic Discrete.

Fig. 4: Examples of Rate-Quality (RQ) curves for sequences with similar textural content from BVI-SynTex-256 (blue lines) and HomTex (red lines) [27]. Example frames from the sequences of the displayed RQ curves are displayed on top of each plot.



(a) Inter (%) mode. (b) Skip (%) mode. (c) Merge (%) mode.

(d) Intra (%) mode. (e) Average Number of Partitions. (f) Percentage of bits used for the residual.

(g) Average MSE of the reconstruction error. (h) Average correlation of the residual to the original. (i) Average length of motion vectors.

Fig. 5: Distributions of coding statistics of BVI-SynTex-256 (in blue) and HomTex (in red). The orange squares on the distributions mark the median values.

in discrete textures [44]. Also, BVI-SynTex-256 has a lower mean value of partitions for the dynamic textures. This is related to the wider range of granularity of the structures, as the same content has been represented in three different levels. On the other hand, HomTex consists mainly of fine structures which explains the higher number of partitions.

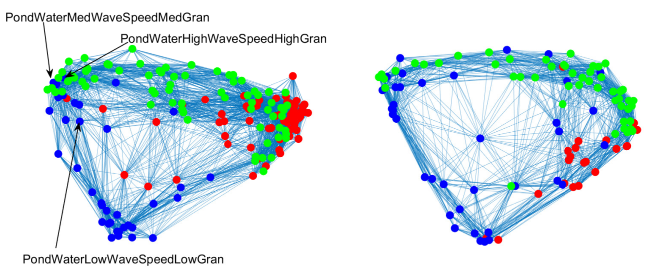
3) Residual Signal: Figures 5 (f)-(h) show that the residual statistics are well aligned for BVI-SynTex-256 and HomTex. Also, both datasets show a variability for the different texture types. It should be noted that the average correlation of the residual to the original frames[2] for static textures of BVI-SynTex-256 is slightly higher than HomTex. This is because the camera motion was simulated in UE4 with models that ensure a uniform motion trajectory across all frames, while for HomTex this would only be feasible if a motorised dolly was

[2]The Pearson correlation coefficient was used only on the luminance component of the frames.

used for the acquisition of the content[3]. Also, higher levels of motion were selected which as expected impacts to the residual energy, as also explained in [45], [46]. Finally, the percentage of bits used for the residual and the average MSE of the reconstructed signal are slightly higher for HomTex. This is expected as many of these sequences suffer from acquisition noise, as mentioned above.

4) Motion Vectors: For both BVI-SynTex-256 and HomTex, static and dynamic discrete textures produce short motion vectors with a similar median value. This is expected for static textures due to simple global translational (camera) motion. It is also expected for the case of dynamic textures due to the restricted range of local random motions of small structures [20].

Different distributions of the average length of motion vectors is observed for dynamic continuous textures. These have the widest distribution, especially for HomTex. Both datasets however have very similar mean and median values.



(a) BVI-SynTex-256. (b) HomTex.

Fig. 6: Undirected graph representation of the (a) BVI-SynTex-256 and (b) HomTex sequences using $k = 33$ neighbors (equal to the number of extracted coding statistics).

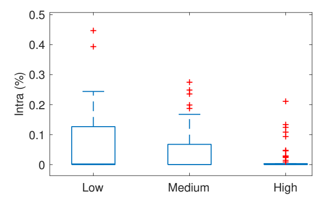
D. Relation of Texture Types with Coding Statistics

Given the large number of features, visual inspection of the data challenging. We therefore represent them as an undirected graph where each node is a sequence. The edges that connect two nodes of the graph consist of the k nearest neighbors (k-NN) of each sequence, based on the distance expressed as the correlation of the extracted statistics. The resulting graph is depicted in Fig. 6 (a), where static, dynamic continuous and dynamic discrete textures are distinguished by the colors red, blue and green, respectively. Inspecting the graph, it is possible to identify three distinct clusters in the data in addition to a number of centrally located nodes. These latter nodes correspond to textures that do not fit into only one class due to their characteristics. Also, it is evident that the clustering due the application of k-NN on the coding statistics is not fully aligned with the expert annotations, which were purely based on the context. Similar clustering patterns appear in the HomTex dataset, as also shown in Fig. 6 (b).

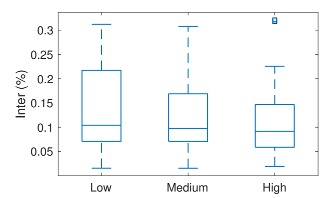
An example of a sequence whose encoding statistics bias them towards another texture group is the C.4 Pond Water sequence. This sequence, according to its statistics, is closely related to dynamic discrete textures and it appears that, as the motion and the granularity increase, the more its encoding performance resembles the dynamic discrete class. Similarly, other sequences that are static appear closer to dynamic continuous textures. It is also interesting that dynamic discrete sequences demonstrate encoding statistics that are similar to static textures. This is attributed to sequences with low granularity and low motion.

E. Relation of Content Parameters with Coding Statistics

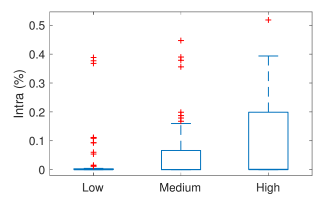
In the previous subsections, the varying levels of granularity and motion and the lack of homogeneity have been suggested as factors that might cause misalignment of the distributions of the two compared datasets. Therefore, in Fig. 7, we show examples of the effect of granularity and motion on encoding statistics on the percentages of Intra and Inter mode. As can be seen, for Intra mode, the lower the granularity the higher the selection of Intra mode and the vice versa for the motion levels.



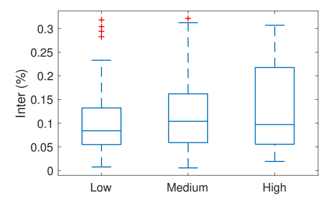
(a) Percentage of Intra mode versus Granularity levels.



(b) Percentage of Inter mode versus Granularity levels.



(c) Percentage of Intra mode versus Motion levels.



(d) Percentage of Inter mode versus Motion levels.

Fig. 7: Examples of the effect of granularity and the motion level on the encoding statistics on the whole dataset.

Interesting cases relating to effect of granularity and motion level are given in Section III where both an objective and subjective evaluation is performed on a subset of BVI-SynTex that takes into account the different versions of the same source content.

F. Discussion

The detailed comparison of BVI-SynTex-256 and HomTex in the previous subsections has shown that the two datasets are aligned and exhibit similar content characteristics and coding statistics. For the observed differences, we have to account for the following fundamental differences of the two datasets:

i. The datasets were acquired with different parameters, i.e. camera, resolution, frame rate and shutter angle, that influence the spatio-temporal features of the source sequences.
ii. BVI-SynTex-256 was heavily downsampled from its native resolution, 1920×1080, to 256×256 and this has an important impact on the spatial characteristics.

[3]The detailed shooting parameters of the HomTex sequences that originate from DynTex dataset are unknown.

iii. BVI-SynTex was captured at 60 fps frame rate, whilst most HomTex sequences (101/120) were captured at 25 fps and the remaining at 60 fps.
iv. Moreover, the HomTex sequences that originate from the DynTex dataset [19] (they are cropped patches from DynTex sequences) suffer from inherent noise caused by the de-interlacing.
v. BVI-SynTex comprises only of 24 different types of content but with a combination of varying levels of motion and granularity. On the other hand, HomTex contains more variation in spatial texture patterns, however not many variations of the same texture pattern.
vi. HomTex contains some sequences with characteristics that are a mixture of texture types as some of the sequences are not purely homogeneous. Although this could affect the comparison to BVI-SynTex, HomTex was selected as a benchmark because it is the only real video dataset with homogeneous textural content.
vii. Finally, HomTex has a low number of static textures, only 25 out of 120 sequences, thus less variability compared to a more balanced composition of BVI-SynTex.

Despite these differences, we believe that the presented comparison is “fair”, as these dataset differences are encountered in the real world, where there is no one-to-one alignment of equipment and specifications in video production.

III. Subjective Evaluation of BVI-SynTex

Besides the study of the content characteristics and encoding performance related to BVI-SynTex, we also performed a subjective study to evaluate compression performance. Since the number of video sequences is large for such a study, we have selected only two scenes per category (six scenes in nine different combinations of model parameters, i.e. 54 source sequences). The motivation behind the selection of the specific representative sequences was to study the effects of granularity and the amount of motion in the same content. The selected sequences are: S.4 Clay Brick, S.5 Cobble Stone, C4. Pond Water, C.8 Waterfall, D.5 Grass, and D.6 TreeI.

All sequences are encoded with the latest version of the HEVC reference software, HM16.20, with a configuration as defined in the JVET Common Testing Conditions [47]. The intra-period is set to 64, GoP length is 16 frames and the quantization range is $QP = \{22, 27, 32, 37, 42\}$.

A. Experimental setup

The viewing environment conformed to home environment conditions, as outlined in BT.500-13 [28]. A consumer display, SONY KD65Z9D, was used with a peak luminance of 250 cd/m^2 (measured using a Konica Minolta CS-2000 spectroradiometer), BT.2020 colour space (full range), set to a 1920×1080 spatial and 60 fps temporal resolution. The viewing distance was chosen as $3 \times H$, where H denotes the height of the screen, as indicated in [48]. The display was connected to a Windows PC with a high performance GPU and an open source software, BVI-SVQA [49], developed within our group, was employed to run the test. The software is lightweight, user friendly and only requires the user to use either a wireless mouse or a wireless keyboard to provide their scores. The software includes other functionalities useful to the experiment operator, such as the option to perform the statistical analysis of the opinion scores and to directly plot the results.

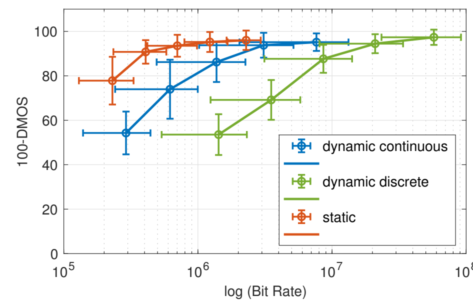


Fig. 8: Average subjective quality-rate curves per different texture type.

B. Experimental Procedure

Because of the large number of selected sequences and the fact that we cover a wide range of quantization levels, we decided to employ a single stimulus methodology with a continuous scale from $0-100$ and with the original sequences as hidden references (a slight variation of the Absolute Category Rating with Hidden Reference (ACR-HR), as defined in BT.500 [28]).

After visual acuity and colour blindness were tested using a Snellen chart and a Ishihara chart respectively, participants were given instructions and presented with a training trial containing videos that were not featured in the main experiment and that covered similar quality levels as the main experiment. The subjective data in the training session were not collected for subsequent analysis. Eighteen observers (16 non-experts and 2 expert) participated in the experiment, aged from 20-40, and with a proportion of 4 females over 14 males.

C. Subjective Rate-Quality (RQ) Curves

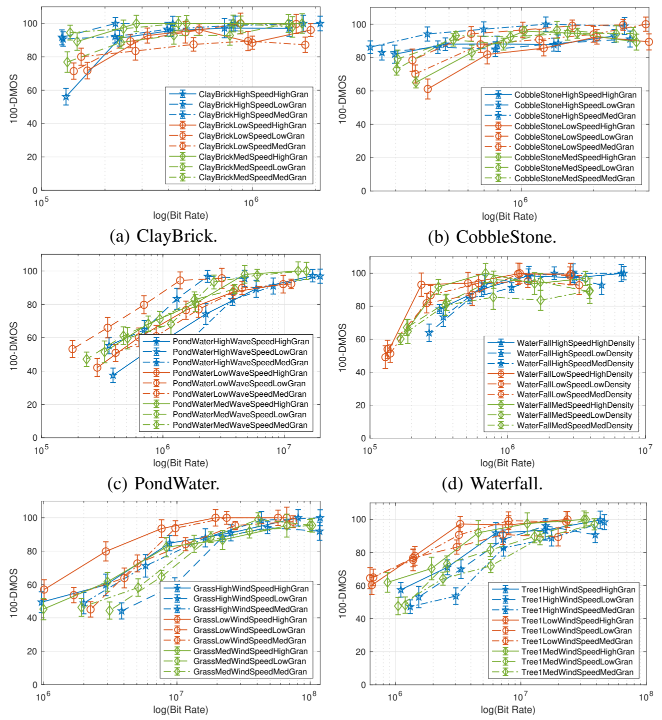
Prior to the processing of the scores, subject screening was performed on the participants' scores using the outlier rejection method recommended by BT.500 [28], the β_2 test. None of the participants were rejected.

The Differential Mean Opinion Scores (DMOS) is used for the analysis of the subjective results, and is calculated as follows:

$$DMOS = \frac{1}{K}\sum_{i=1}^{K}(OS_{\text{org},i} - OS_{\text{cmp},i}) \tag{1}$$

where OS_{org} represent the opinion score of participant $i \in \{1, \ldots, K\}$ for the original reference sequence and $OS_{\text{cmp},i}$ the opinion score of participant i for the compressed version of the same video.

In Fig. 8, we plot the average subjective quality-rate curves per different groups. The vertical ranges denote the standard deviation of the DMOS, while the horizontal ranges denote the bit rate standard deviation. A first strong observation is that there is clear separation of the average perceived quality-rate



(a) ClayBrick. (b) CobbleStone.

(c) PondWater. (d) Waterfall.

(e) Grass. (f) TreeI.

Fig. 9: 100-DMOS-Rate curves for all tested sequences.

performance of the different texture types. Previous work [9], compared static textures to dynamic textures (aggregating both dynamic continuous and discrete) and mixed textures. The new findings from our tests, show the different performance of dynamic texture types, namely dynamic continuous and dynamic discrete. We notice that there is a significant difference in the performance of the two different dynamic texture types. On average, dynamic discrete textures require the highest bit rate compared to the other two texture types. The curves also indicate that for the tested dynamic synthetic textures there is a saturation of the scored quality in most cases at the third tested quantization QP point (32). For the static textures, we notice this at even higher QP.

In Fig. 9, we plot the 100-DMOS-Rate curves for all tested sequences. It can be observed that sequences with low motion have subjective results that show a wider range of perceived quality. This is expected since lower motion improves the perception of the content and thus enables better judgment of quality level. Similarly, for lower granularity, where the density of the high spatial frequencies is low, the effect of the spatial masking is weaker and it is easier to spot the compression artifacts.

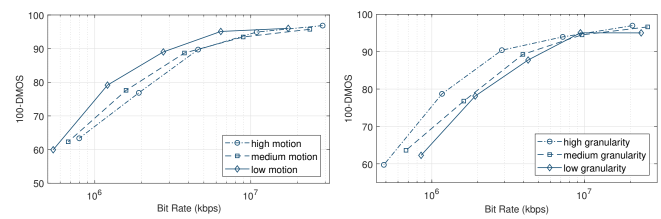
D. Statistical Analysis of the Subjective Quality Assessment

We performed a significance test using one-way Analysis of Variance (ANOVA) between the different QPs. Whenever the ANOVA shows significant difference, i.e. $p < 0.05$, we signal this with "1", otherwise with "0". In Table IV, we report the results of the aggregation of the significance point system. A first observation is that the highest compression levels ($QP = 42$) that correspond to the lowest rate points, were indicated as

TABLE IV: Aggregated significant difference of perceived quality between QPs. The maximum points that can be reached is 54 (aka the number of tested compressed sequences).

QP	22	27	32	37	42
22	-	30/54	44/54	45/54	45/54
27	30/54	-	17/54	27/54	27/54
32	44/54	17/54	-	0/54	0/54
37	45/54	27/54	0/54	-	0/54
42	45/54	27/54	0/54	0/54	-

significantly different from the other four compression levels in almost all tested sequences.



(a) Motion Levels. (b) Granularity Levels.

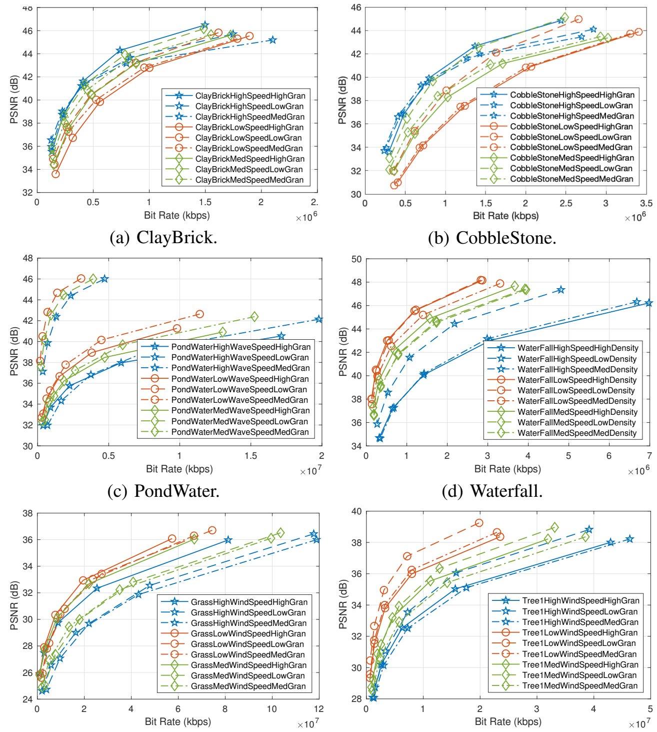
Fig. 10: Average 100-DMOS-Rate curves over all tested sequences grouped according to their (a) motion and (b) granularity level.

E. Relation of Content Parameters with RQ Curves

In order to explore the content parameters effect on the perceived quality at various compression levels, we are illustrating in Fig. 10 the average RQ curves grouped according to their (a) motion and (b) granularity level. We observe that the sequences with low motion have been scored with higher quality by the observers. Then for the high QPs (low bit rates), the medium motion sequences are scored with higher than the high motion sequences. This is naturally expected as the compression artifacts are more intense in the high motion sequences where a lot of blockiness is observed. In the low QP tested range (high bit rates), the curves converge which is expected as the quality range is very high (> 90). In Fig. 10 (b), it is clear that the high granularity average RQ curve is on top of the medium and low granularity curves for almost the whole range of tested QPs. This is attributed to the stronger spatial masking in content that is dense of high frequencies, making distortion less visible.

IV. Objective Quality Assessment of the Subjectively Tested SynTex

To complete the quality assessment, we computed nine popular objective Image and Video Quality Assessment (IQA/VQA) metrics, including PSNR, PSNR that takes into account the Contrast Sensitivity Function (CSF) and the between-coefficient contrast masking of Discrete Cosine Transform (DCT) basis functions (PSNR-HVSM) [50], Structural Similarity Index (SSIM) [51], multi-scale SSIM (MS-SSIM) [52], Visual Information Fidelity measure (VIF) [53], Video Quality Metric (VQM) [54], Spatio-Temporal Most Apparent Distortion model (ST-MAD) [55] and Video Multi-method Assessment Fusion (VMAF) metric (using model vmaf_v0.6.1.pkl) [56]. PSNR, PSNR-HVSM, SSIM, MS-SSIM, and VIF are commonly used IQA metrics, while VQM, ST-MAD and VMAF are VQA methods.

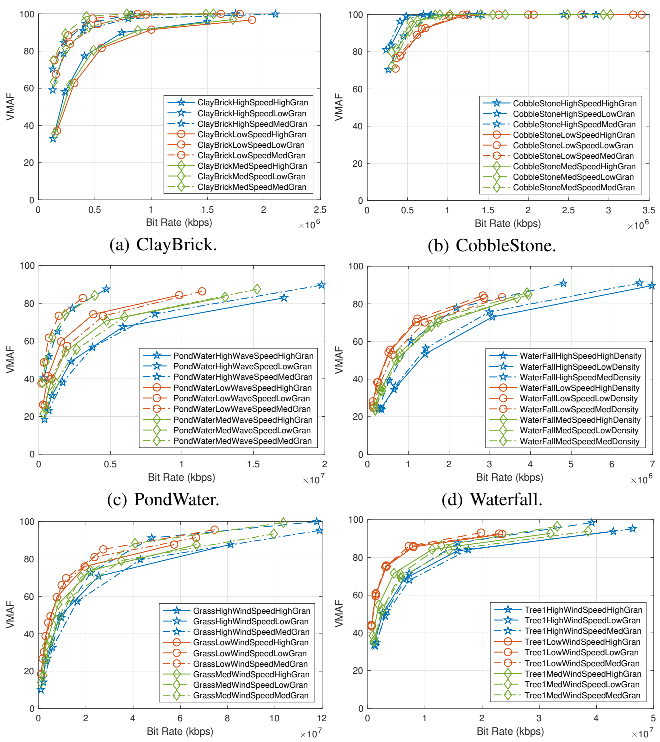


(a) ClayBrick. (b) CobbleStone.

(c) PondWater. (d) Waterfall.

(e) Grass. (f)TreeI.

Fig. 11: PSNR-Rate curves for all tested sequences.



(a) ClayBrick. (b) CobbleStone.

(c) PondWater. (d) Waterfall.

(e) Grass. (f)TreeI.

Fig. 12: VMAF-Rate curves for all tested sequences.

A. Objective RQ Curves

As PSNR and VMAF are commonly used by the video compression research community, we plotted the RQ curves of the selected sequences in Figs. 11 and 12. Each subplot aggregates the curves of the different versions of the same content. As mentioned earlier, the static textures are more effectively compressed in all their versions compared to dynamic textures. Also it can be seen here in more detail that both the spatial and temporal variation of features highly influences compression performance. Particularly, for the dynamic textures, the "HighSpeed" versions (blue curves) are harder to compress as expressed by both quality metrics and the wider bit rate range. Also, dynamic discrete textures (Grass and Tree sequences) are harder to compress compared to dynamic continuous and static, as shown by the objective metrics RQ curves of Figs. 11 and 12 and also explained earlier in Section II-C.

It is also important to note, that although the objective metrics indicate significant quality difference, in many cases the subjective results overthrow these figures as many of the 100-DMOS-Rate curves have overlapping confidence intervals.

B. Correlation of Objective Metrics to DMOS in synthetic video textures

It is well known [1] that traditional objective quality metrics do not correlate well with subjective video quality assessments. This of course is content-dependent and is more noticeable for textured content. We tested the performance of the aforementioned objective quality metrics using: the Spearman Rank Order Correlation Coefficient (SROCC), the Pearson Linear Correlation Coefficient (LCC), the Outlier Ratio (OR) and the Root Mean Squared Error (RMSE) after performing a nonlinear regression with a four-parameter logistic function. SROCC assesses the monotonicity of the objective metric prediction with respect to opinion scores, while LCC measures the prediction accuracy [57], [58].

We report the quality metric assessment figures in Table V. It is evident that for the selected BVI-SynTex sequences, MS-SSIM shows the highest linear correlation, while SSIM the lowest. The same ranking of the IQA/VQA metrics results from the rank correlation values. SSIM, PSNR-HSVM and VMAF have the lowest OR, while the lowest RMSE is recorded for MS-SSIM. Generally, the higher the correlation statistics (LCC, SROCC), the better the perceptual alignment of the tested objective quality metric to the textured content.

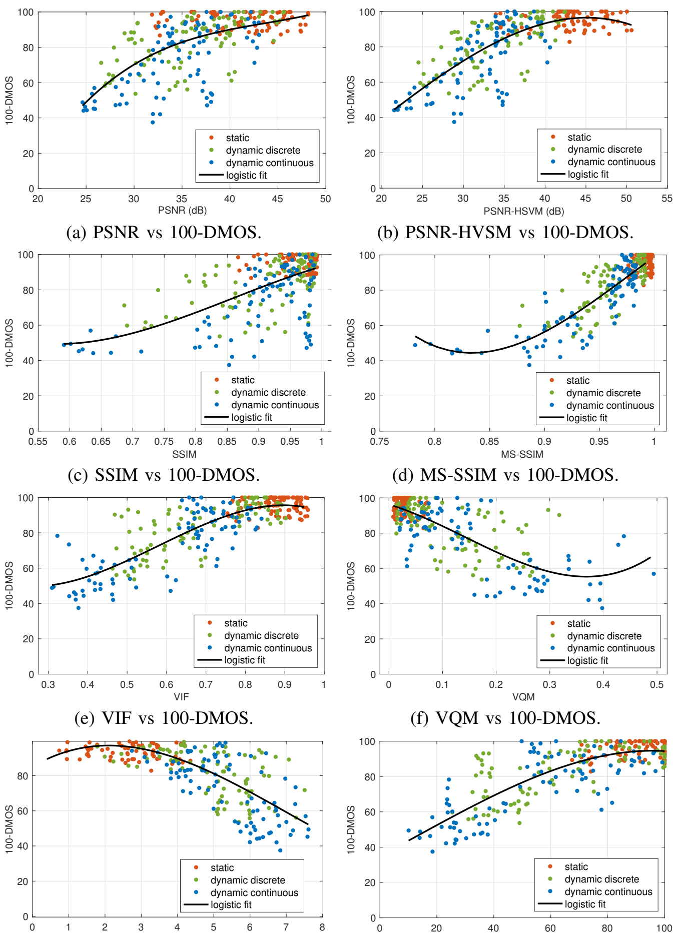
Additionally, it is worth mentioning that the non-linear behaviour of the tested objective quality metrics in relation to subjective quality is visually verified by the plots in Fig. 13, where the 100-DMOS scores are scattered against them and the fitted logistic function is plotted. As can be seen from the figure, in the case of PSNR and SSIM, the logistic function does not fit well resulting in low linear and rank correlation values. On the other hand, the logistic function fits much better on the MS-SSIM, VIF and VMAF values resulting in better linear and rank correlation values. This means that MS-SSIM, VIF and VMAF are better perceptually aligned than the rest of the IQA/VQA metrics on the tested textured content.

Fig. 13 also indicates that most objective quality metrics are performing better on static textures (red markers) as those points are closely scattered to the fitted logistic curve. This is expected as the static textures are easier to compress while maintaining a high quality. For the dynamic textures (green and blue markers), the tested IQA/VQA metrics do not

TABLE V: Comparison of Correlation Statistics for the Tested Image/Video Quality Metrics on the BVI-SynTex Selected Subset.

Metric	LCC	SROCC	OR	RMSE
PSNR	0.6695	0.6126	0.0630	11.9701
PSNR-HVSM	0.7794	0.7358	0.0593	10.0971
SSIM	0.6222	0.5289	0.0593	12.6152
MS-SSIM	**0.8714**	**0.7730**	0.0778	**7.9058**
VIF	0.8372	0.7447	0.0667	8.8129
VQM	0.7653	0.6479	0.0593	10.3735
ST-MAD	0.7417	0.6690	**0.0481**	10.8091
VMAF	0.8162	0.6977	0.0593	9.3111

correlate well and most fitted curves are not monotonic either at the high quality (PSNR-HSVM, VIF, ST-MAD) or at the low quality (MS-SSIM, VQM) range.



(a) PSNR vs 100-DMOS.

(b) PSNR-HVSM vs 100-DMOS.

(c) SSIM vs 100-DMOS.

(d) MS-SSIM vs 100-DMOS.

(e) VIF vs 100-DMOS.

(f) VQM vs 100-DMOS.

(g) ST-MAD vs 100-DMOS.

(h) VMAF vs 100-DMOS.

Fig. 13: Scatterplot of 100-DMOS versus the tested objective quality metrics. The black solid line shows the fitted logistic function.

V. Conclusions and Future Work

In this paper, we introduced a CGI generated video texture dataset, BVI-SynTex. BVI-SynTex benefits from model-driven content generation and covers a wide range of video textures with varying spatial and temporal patterns. It provides the benefit of parameterizable video content (e.g. granularity, wind speed, camera speed). The relative coverage of traditionally used content descriptors and coding performance of BVI-SynTex are shown to be comparable to real video textured datasets. Thus, we can conclude that BVI-SynTex has similar properties compared to real videos.

To the best of our knowledge, BVI-SynTex is the first video texture dataset created for video compression purposes. As all the original videos and the computed metrics are publicly available, it can be used by researchers for analysis and understanding of the video parameter space and its relation to visual perception and video compression. Its parameterisable nature, has the benefit of being extensible to include more video variations (different scene, heterogeneous content, etc.) to cover the needs of training, testing and validating of any video acquisition, analysis and compression research method.

It is part of our future work plans to extend BVI-SynTex with more realistic scenes, including diverse content, changing lighting conditions, varying camera motion (zooming in and out, rotation). Also, emphasis will be given in realistic motion representation by including motion fields that vary within sequences and by imposing extracted motion fields from real sequences on synthesized content.

References

[1] D. R. Bull, *Communicating pictures: A course in Image and Video Coding*, Academic Press, 2014.

[2] White Paper, "Cisco Visual Networking Index: Forecast and Methodology, 2016 – 2021," Tech. Rep. 1465272001663118, CISCO, Sep. 2017.

[3] D. Chen, Q. Chen, and F. Zhu, "Pixel-level Texture Segmentation Based AV1 Video Compression," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 1622–1626.

[4] F. Zhang and D. R. Bull, "A parametric framework for video compression using region-based texture models," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1378–1392, Nov 2011.

[5] B. Wandt, T. Laude, B. Rosenhahn, and J. Ostermann, "Extending hevc with a texture synthesis framework using detail-aware image decomposition," in *2018 Picture Coding Symposium (PCS)*, June 2018, pp. 144–148.

[6] M. A. Papadopoulos, Y. Rai, A. V. Katsenou, D. Agrafiotis, P. Le Callet, and D. R. Bull, "Video quality enhancement via qp adaptation based on perceptual coding maps," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 2741–2745.

[7] Z. Wang, S. Wang, X. Zhang, S. Wang, and S. Ma, "Locally refined motion compensation for future video coding," in *2018 Data Compression Conference*, March 2018, pp. 431–431.

[8] O. Chubach, P. Garus, M. Wien, and J. Ohm, "Motion-distribution based dynamic texture synthesis for video coding," in *2018 Picture Coding Symposium (PCS)*, June 2018, pp. 218–222.

[9] M. A. Papadopoulos, F. Zhang, D. Agrafiotis, and D. Bull, "A video texture database for perceptual compression and quality assessment," in *IEEE International Conference on Image Processing (ICIP)*, Sept 2015, pp. 2781–2785.

[10] K. Naser, V. Ricordel, and P. Le Callet, "Modeling the perceptual distortion of dynamic textures and its application in hevc," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 3787–3791.

[11] V. Q. E. Group, "Final report from the video quality experts group on the validation of objective quailty metrics for video quality assessment," Tech. Rep., 2000.

[12] V. Q. E. Group, "Report on the validation of video quality models for high definition video content," Tech. Rep., 2010.

[13] S. Pechard, R. Pepion, and P. L. Callet, "Suitable methodology in subjective video quality assessment: a resolution dependent paradigm," in *Int. Workshop Image Media Quality Appl.*, 2008.

[14] F. D. Simone, "EPFL-PoliMI video quality assessment database," http://vqa.como.polimi.it, 2009.

[15] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, “Study of Subjective and Objective Quality Assessment of Video,” *IEEE Transactions on Image Processing*, vol. 19, no. 6, 2010.

[16] A. Mackin, F. Zhang, and D. R. Bull, “A study of subjective video quality at various frame rates,” in *IEEE International Conference on Image Processing (ICIP)*, Sept 2015, pp. 3407–3411.

[17] F. Zhang, F. M. Moss, R. Baddeley, and D. R. Bull, “BVI-HD: A video quality database for HEVC compressed and texture synthesized content,” *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2620–2630, 2018.

[18] B. Butterworth, “History of the ’BBC Redux’ project,” BBC Internet Blog, 2013.

[19] R. Péteri, S. Fazekas, and M. J. Huiskes, “DynTex: A comprehensive database of dynamic textures,” *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1627–1632, 2010.

[20] M. Afonso, A. Katsenou, F. Zhang, D. Agrafiotis, and D. Bull, “Video texture analysis based on HEVC encoding statistics,” in *Picture Coding Symposium (PCS)*. IEEE, 2016, pp. 1–5.

[21] A. V. Katsenou, M. Afonso, and D. R. Bull, “VIL: Homogeneous Video Texture Dataset (HomTex),” https://data.bris.ac.uk/data/dataset/1h2kpxmxdhccf1gbi2pmvga6qp, Jul. 2016.

[22] Y. Zhang, W. Weichao Qiu, H. X. Chen, Q., and A. Yuille, “UnrealStereo: Controlling hazardous factors to analyze stereo vision,” in *International Conference on 3D Vision (3DV)*. IEEE, 2018.

[23] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, “Virtual worlds as proxy for multi-object tracking analysis,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4340–4349.

[24] A. Dai, A. X. Chang, M. Savva, M. Halber, T. A. Funkhouser, and M. Nießner, “ScanNet: Richly-annotated 3D reconstructions of indoor scenes.,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 2, pp. 5828–5839.

[25] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, “The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3234–3243.

[26] F. Zhang, W.-C. Feng, and F. Liu, “A multi-lens stereoscopic synthetic video dataset,” in *ACM Mutlimedia Systems*, 2015.

[27] D. Ma, A. V. Katsenou, and D. Bull, “A Synthetic Video Dataset for Video Compression Evaluation,” in *International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1094–1098.

[28] ITU-T, “Recommendation BT.500-13: Methodology for the subjective assessment of the quality of television pictures,” Tech. Rep., 2012.

[29] “The Unreal Engine (UE4),” https://www.unrealengine.com, 2018.

[30] U. C. Pendit, M. B. Mahzan, M. D. F. B. M. Basir, M. B. Mahadzir, and S. N. binti Musa, “Virtual reality escape room: The last breakout,” in *Information Technology (INCIT), 2nd International Conference on*. IEEE, 2017, pp. 1–4.

[31] A. V. Katsenou, T. Ntasios, M. Afonso, D. Agrafiotis, and D. R. Bull, “Understanding video texture - a basis for video compression,” in *Multimedia Signal Processing (MMSP), IEEE 19th International Workshop on*. IEEE, 2017, pp. 1–6.

[32] A. V. Katsenou, D. Ma, G. Dimitrov, and D. R. Bull, “BVI Synthetic Video Texture Dataset - BVI-SynTex,” https://data.bris.ac.uk/data/dataset/320ua72sjkefj2axcjwz7u7yy9, Aug. 2019.

[33] G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Scandinavian Conference on Image analysis*. Springer, 2003, pp. 363–370.

[34] Li Song, Xun Tang, Wei Zhang, Xiaokang Yang, and Pingjian Xia, “The sjtu 4k video sequence dataset,” in *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2013, pp. 34–35.

[35] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini, and S. MÃűller, “GamingVideoSET: A Dataset for Gaming Video Streaming Applications,” in *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, June 2018, pp. 1–6.

[36] A. Mackin, F. Zhang, and D. R. Bull, “A Study of High Frame Rate Video Formats,” *IEEE Transactions on Multimedia*, 2019.

[37] D. Ghadiyaram, J. Pan, and A. C. Bovik, “A Subjective and Objective Study of Stalling Events in Mobile Streaming Videos,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 183–197, Jan 2019.

[38] S. Winkler, “Analysis of Public Image and Video Databases for Quality Assessment,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, Oct 2012.

[39] C. E. Duchon, “Lanczos filtering in one and two dimensions,” *Journal of Applied Meteorology*, vol. 18, no. 8, pp. 1016–1022, 1979.

[40] “FFMPEG,” https://www.ffmpeg.org.

[41] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, “Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC),” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, 2012.

[42] M. Wien, *High Efficiency Video Coding: Coding Tools and Specification*, Springer, 2015.

[43] J. Kim, J. Yang, K. Won, and B. Jeon, “Early determination of mode decision for HEVC,” in *Picture Coding Symposium (PCS)*. IEEE, 2012, pp. 449–452.

[44] S. Cho and M. Kim, “Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 9, pp. 1555–1564, 2013.

[45] G. Tian and S. Goto, “Content adaptive prediction unit size decision algorithm for HEVC intra coding,” in *Picture Coding Symposium (PCS)*. IEEE, 2012, pp. 405–408.

[46] Y. Wang, X. Fan, D. Zhao, and W. Gao, “Mode dependent intra smoothing filter for HEVC,” in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 539–543.

[47] K. Sharman and K. Sühring, “Common Test Conditions for HM video coding experiments,” Tech. Rep., document JCTVC-AC1100 of JCT-VC, Oct. 2017.

[48] ITU-T, “Recommendation P.910: Subjective video quality assessment methods for multimedia applications,” Tech. Rep., 2008.

[49] G. Dimirov, A. V. Katsenou, and D. R. Bull, “SVQA: Subjective Video Quality Assessment software,” https://github.com/goceee/SVQA, Aug. 2019.

[50] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, “On between-coefficient contrast masking of DCT basis functions,” in *Proc. of the 3rd Intern. workshop on video processing and quality metrics*, 2007, vol. 4.

[51] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[52] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-scale structural similarity for image quality assessment,” in *Asilomar Conf. Signals Syst. Comput.*, 2003, pp. 1398–1402.

[53] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, Dec 2005.

[54] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[55] P. V. Vu, C. T. Vu, and D. M. Chandler, “A spatiotemporal most-apparent-distortion model for video quality assessment,” in *18th IEEE International Conference on Image Processing*, Sep. 2011, pp. 2505–2508.

[56] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, “The NETFLIX tech blog: Toward a practical perceptual video quality metric,” http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html, note = [Online; accessed 2018-08-04],.

[57] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, “Study of Subjective and Objective Quality Assessment of Video,” *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, June 2010.

[58] ITU-T, “Recommendation P.1401: Statistical analysis, evaluation and reporting guidelines of quality measurements Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models,” Tech. Rep., 2012.

Angeliki V. Katsenou (M'11) is a Leverhulme Early Career Fellow and is with the Visual Information Lab at the University of Bristol since 2015. She obtained her Ph.D. degree from the Department of Computer Science and Engineering, University of Ioannina, Greece (2014). She received her B.S./M.Eng. in Electrical and Computer Engineering and an M.Sc. degree in Signal and Image Processing from the University of Patras, Greece. She has experience in several FP7 EC-funded and EPSRC projects, such as MSCA-ITN PROVISION and EPSRC Platform Grant EP/M000885/1. Her research interests include perceptual video analysis, video compression, and quality.

Goce Dimitrov received his B.Sc./M.Eng. in Computer Science and Electronics at the University of Bristol, U.K., in 2019. During his final year project he developed a software for subjective video quality assessment and he did an internship on video codec comparison with the Visual Information Lab, University of Bristol, under the supervision of Prof. D. Bull and Dr. A. Katsenou. His research interests include software engineering, video quality assessment, cloud computing and artificial intelligence.

Di Ma received the MEng degree (First Class Honours) in Information and Communications Engineering at the Tianjin Key Lab for Advanced Signal Processing, Tianjin, P.R.China, in 2016. He is currently working toward the PhD degree in Communications with the EPSRC Centre for Doctoral Training in Communications, University of Bristol. His research interests include perceptual video compression, computer vision, machine learning, and radar signal processing.

David R. Bull (M'94-SM'07-F'12) received the B.Sc. degree from the University of Exeter, Exeter, U.K., in 1980; the M.Sc. degree from University of Manchester, Manchester, U.K., in 1983; and the Ph.D. degree from the University of Cardiff, Cardiff, U.K., in 1988.

Dr Bull has previously been a Systems Engineer with Rolls Royce, Bristol, U.K. and a Lecturer at the University of Wales, Cardiff, U.K. He joined the University of Bristol in 1993 and is currently its Chair of Signal Processing and Director of its Bristol Vision Institute. In 2001, he co-founded a university spin-off company, ProVision Communication Technologies Ltd., specializing in wireless video technology. He has authored over 450 papers on the topics of image and video communications and analysis for wireless, Internet and broadcast applications, together with numerous patents, several of which have been exploited commercially. He has received two IEEE Premium awards for his work. He is the author of three books, and has delivered numerous invited/keynote lectures and tutorials. Dr. Bull is a fellow of the Institution of Engineering and Technology.