# A Study of High Frame Rate Video Formats

Alex Mackin, Fan Zhang, *Member, IEEE,* and David R. Bull, *Fellow, IEEE*

**Abstract**

High frame rates are acknowledged to increase the perceived quality of certain video content. However the lack of high frame rate test content has previously restricted the scope of research in this area - especially in the context of immersive video formats. This problem has been addressed through the publication of a high frame rate video database BVI-HFR, which was captured natively at 120 fps. BVI-HFR spans a variety of scenes, motions and colours, and is shown to be representative of BBC broadcast content. In this paper temporal down-sampling is utilised to enable both subjective and objective comparisons across a range frame rates. A large-scale subjective experiment has demonstrated that high frame rates lead to increases in perceived quality, and that a degree of content dependence exists - notably related to camera motion. Various image and video quality metrics have been benchmarked on these subjective evaluations, and analysis shows that those which explicitly account for temporal distortions (e.g. FRQM) provide improved correlation with subjective opinions compared to generic quality metrics such as PSNR.

**Index Terms**

High frame rates, video database, immersive video, UHDTV, HFR

## I. Introduction

As the demand for higher quality and more engaging video experiences increases, the pressure to extend the video parameter space beyond current spatial and temporal resolutions, dynamic ranges and screen sizes becomes ever greater [1]. Most of the significant activity in recent years has been focused towards the implementation of high spatial resolution (4K/8K) [2, 3], high dynamic range (HDR) [4, 5] and immersive multi-view formats [6, 7]. However the progress related to high frame rate formats has been comparatively slow, shown by the frame rates used in television and cinema having remained constant for many years - rarely exceeding 60 fps.

There are many reasons why increased frame rates have been overlooked previously, and they include:

- Video parameters such as spatial resolution (4K, 8K), bit-depth and colour gamut (HDR) being prioritised.
- The perceptual benefits associated with higher frame rates are relatively unknown to wider audiences, partly due to the lack of commercial and/or publicly available high frame rate material.
- Inadequate camera/display technology.
- Audiences being accustomed to lower frame rate material, with moves towards 'higher' frame rates being met with criticism. *The Hobbit: An Unexpected Journey* which was captured natively at 48 fps (3D) is a prime example of this assertion, as audiences complained that the 'magic was lost' [8]. However in this context it is difficult to uncouple the benefits of high frame rates from 3D.
- Integral elements of the filming process, such as: camera motion, capture parameters, computer graphics and lighting, need to be readdressed.
- Based on the contrast sensitivity function [9], there is a commonly held belief that the maximum perceptible temporal resolution of the human visual system is 50-60 fps. While this may be true for the case of flicker (which is fairly uncommon in video), when it comes to objects in motion, the frame rate determines the spatial displacement between samples. As such we may need to sample above 900 fps to recreate optical reality [10, 11].

High frame rates have recently stimulated interest in the broadcast [12, 13], online streaming (Youtube supports frame rates up to 60 fps), film (*Billy Lynn's Long Halftime Walk*, *Avatar 2*), gaming [14] and virtual reality (VR) [15] communities. Frame rates up to 120 fps are specified in the UHDTV (ultra-high-definition) video standard (Rec. 2020) [16].

Manuscript drafted October 27, 2018

The authors acknowledge funding from EPSRC (grant EP/M000885/1), and funding and support from BBC Research & Development.

All the authors are with the Bristol Vision Institute, University of Bristol, Bristol, UK. E-mail: {A.Mackin, Fan.Zhang, Dave.Bull}@bristol.ac.uk

Before future video formats begin to exploit higher frame rates, further investigation is required into the role that frame rates play in the complete video pipeline, from acquisition through compression and transmission to visual perception.

As is common across the range of emerging immersive formats (4K/8K, HDR, HFR, multi-view etc.), there are a range of artefacts and distortions that may arise due to compression, packet-loss and/or under-sampling [17]. These may have significant effects not just on visual quality, but also on user immersion (thus undermining the goal of immersive video formats).

In the context of this paper we explore the suitability of HFR formats across a range of different content types, with a view of enabling informed decisions about the suitability of higher frame rates, allowing for perceptually optimised frame rate recommendations to be made, and hopefully convincing audiences, broadcasters, content creators and manufacturers alike that higher frame rates can enhance visual quality, and provide far more immersive video experiences than is currently realised.

The lack of available high frame rate content has often been a limiting factor when conducting research, and has meant that robust conclusions about increased frame rates have been difficult to make. In order to address this problem, we have created a publicly available high frame rate video database BVI-HFR that contains a diverse set of 10 second HD video sequences captured at 120 fps.

In this paper we utilise BVI-HFR to study the impact of frame rate variation in a number of key areas for future immersive video formats. Building upon our previous work in this area [18–20], we exploit temporal down-sampling to enable comparisons across a range of frame rates.

The primary contributions of this work are as follows:

1) A publicly available high frame rate video database BVI-HFR, which allows researchers to investigate video frame rates in a content dependent manner, is presented.
2) The influence of temporal down-sampling methods on video compression is explored, and some practical considerations of increased frame rates are discussed.
3) The link between frame rates and perceived quality is characterised using a subjective experiment, and content dependence is assessed using statistical techniques.
4) Various generic and frame rate dependent quality metrics are benchmarked on content with varying frame rates.

The remainder of this paper is organised as follows: Section II summarises the state-of-the-art; Section III characterises the BVI-HFR video database; Section IV provides an analysis of temporal down-sampling; Section V presents a subjective experiment that quantifies the relationship between frame rate and visual quality; Section VI benchmarks existing quality metrics. Conclusions and suggestions for future work are then presented in Sections VII and VIII respectively.

## II. Background

### A. Benefits of Increased Frame Rates

Previous research has demonstrated that there are a number of clear benefits associated with increased frame rates, including: the visibility of temporal aliasing artefacts being diminished [10, 21–26]; a reduction in perceptible motion blur [23–25, 27–29]; increased realism, smoother motion, improved depth perception for both expert [30] and non-expert [31] viewers (especially when tracking using smooth pursuit eye movements); more realistic motion image quality (confirmed using EEG data) [32]; a reduction in viewer stress levels [33] (signified by a lower blinking frequency [34]); an improvement in speed and spatial discrimination, and reading ability [35]; and an increase in perceptual quality [18, 36], at least up to 240 fps [29]. The use of higher frame rates also enhances the ability to capture and playback slow-motion videos [37].

Viewers have been shown to prioritise frame rates over spatial resolution for computer generated (CG) content [38].

In previous work, we demonstrated using a novel experimental setup, that frame rates close to 900 fps can be required to fully eliminate perceptible temporal aliasing artefacts in certain scenarios, and that frame rates up to 30% higher may be required for future high dynamic range (HDR) displays [11].

Regardless of these benefits, high frame rate material may only be advisable when a 'hyper-realistic' representation of the scene is required (e.g. sports programming), as there may be a conflict with the 'cinematic look' at lower frame rates. Directors and content providers currently have little flexibility in this regard (as the frame rates within legacy formats have remained static for many years), and therefore the choice of frame rate - enabled through the use of temporal down-sampling methods - could be considered an artistic choice.

### *B. High Frame Rate Video Databases*

Few (if any) high frame rate (60 fps+) video databases have ever been released [39]. Related work has either considered relatively low frame rates (up to 30 fps [40–42]), or the research data has not been published in its entirety (missing subjective evaluations and/or source sequences) [25, 27–31].

### *C. Characterisation of Video Content*

The coverage of a video database over low-level descriptors is typically used to characterise its content [39]. While this does allow for an objective comparison with other video databases, there is little information pertaining to how representative the database is of typical video content. As unless designed with a specific purpose in mind, a video database should not simply contain superficial or novel sequences (which could be devised to exploit descriptors), and instead should be archetypal of consumer content. This should ensure that research is relevant to a wider, non-expert, audience.

An extensive analysis of recent broadcast content has shown that the distribution of five uncorrelated factors (generated using PCA) can be used to quantify the representativeness of a small-scale video database compared to the vast population of modern broadcast content (in this case BBC Redux data) [43].

### *D. Video Compression/Transmission*

In previous work [19], we investigated the impact of frame rates on the rate-quality performance of an HEVC encoder utilising frame-averaging. Results showed that high frame rates offer clear perceptual benefits at current data rates, and that for the same quality, sequences containing camera motion require higher frame rates compared to those without.

*Nasiri et al.* [36] used a subjective experiment to assess to the impact of frame rate and H.264 compression on perceived video quality. Their results show a positive, yet diminishing, relationship between quality and frame rate, and that any gains are dependent on quantisation level (QP), spatial resolution and the statistics of the video content (characterised by measures of both spatial and motion complexity).

Alongside video compression, there has also been a number of recent advancements with respect to high frame rate video transmission/streaming. *Kurdoglu et al.* [44] demonstrated the effect of video frame rates in streaming environments which exhibit bursty packet losses. They show that it can sometimes be beneficial to use lower encoding frame rates when transmitting video data over very noisy channels, and that visual quality can be increased by using temporal layering techniques at the encoder. *Wu et al.* [45, 46] have also proposed solutions (ROCHET, FRIED and JASCO) to enhance video quality when transmitting high frame rates over wireless networks.

### *E. Frame Rate Dependent Video Quality Metrics*

Video quality metrics, which estimate the relative quality of a low frame rate video compared to its higher frame rate counterpart, should account for both spatial and temporal distortions. Examples of such video quality metrics include: TCFQ [47], MNQT [42], PQD-FRS [48] and FRQM [20]. TCFQ and MNQT model degradation in perceptual quality as frame rates decrease with a parametrised exponential function, PQD-FRS uses a machine learning approach to predict the satisfied user ratio (SUR) between two distinct frame rates, while FRQM estimates the relative quality between frame rates using wavelet decomposition and spatio-temporal pooling.

## III. BVI-HFR Video Database

The high frame rate (120 fps) video database BVI-HFR is presented in this section. The content of the database is characterised using two distinct methods: one involves computing the coverage of BVI-HFR over three low-level descriptors, while the other compares BVI-HFR to typical broadcast material from BBC Redux archives.

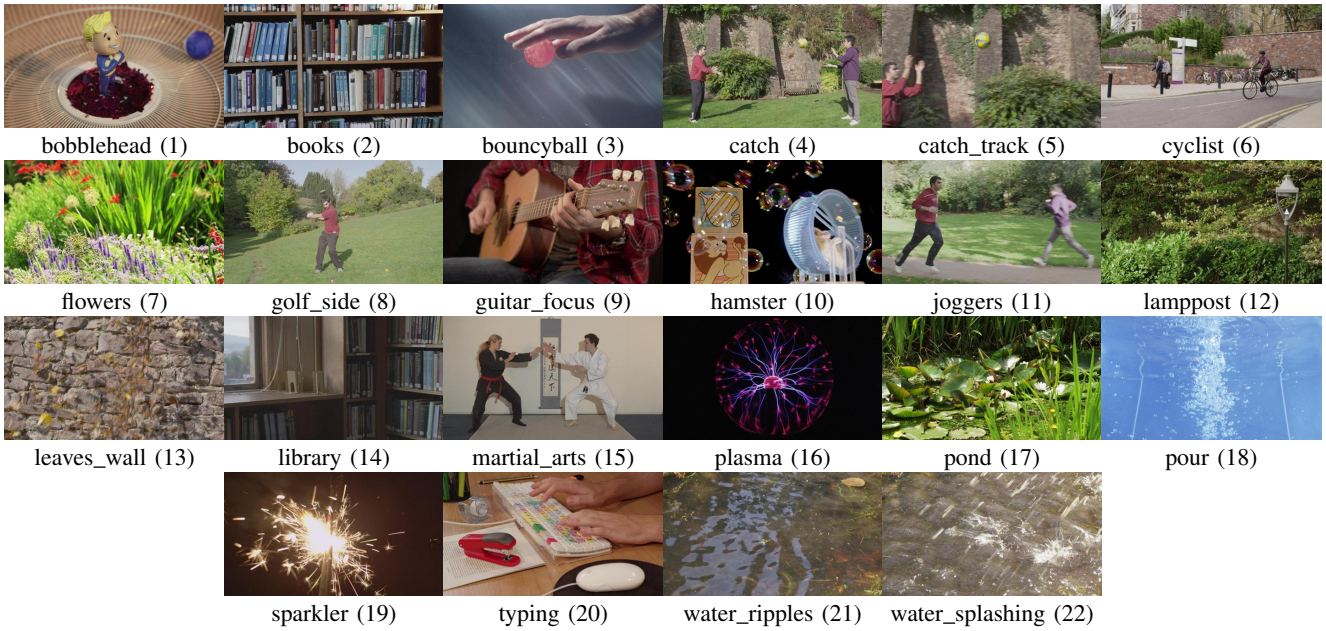



Fig. 1: A sample frame from each of the 22 video sequences in the BVI-HFR video database, along with the names and associated indices.

### *A. Source Sequences*

The Bristol Vision Institute High Frame Rate (BVI-HFR) video database [18] contains 22 unique video sequences that were captured natively using a RED Epic-X video camera with a 3840×2160p (UHD-1) spatial resolution, a frame rate of 120 fps and a 360° shutter angle. All 22 sequences were spatially down-sampled to 1920×1080p (HD) resolution using REDCINE-X software (which was also used for post-processing) into YUV 4:2:0 format (8 bit). The sequences are 10 seconds in duration, and contain no shot transitions or audio components. The name and associated index of each of sequence, alongside a sample frame, is shown in Fig. 1. The BVI-HFR video database, alongside the test sequences and subjective results from Section V, are available to download from the following link: https://vilab.blogs.ilrt.org/?p=1563.

### *B. Content Description*

Using the method proposed by Winkler [39], we characterise the content of BVI-HFR using three low-level descriptors: Spatial Information (SI), Temporal Information ($TI_{MV}$) and Colourfulness (CF). SI is an estimator of the amount of edge energy in the video sequence, and can be used to quantify the spatial complexity of a scene, $TI_{MV}$ predicts the magnitude of motion, whereas CF quantifies the variety and the intensity of colours within a scene. The coverage and distributions of the descriptors for BVI-HFR are shown in Fig. 2 and 3.

The coverage of a video database over these descriptors can be used to evaluate whether it successfully spans a variety of scenes, motions and colours, and therefore ensure fair and sufficient scrutiny when benchmarking new and existing algorithms. Winkler [39] proposes three evaluation metrics (relative range, uniformity of coverage and relative total coverage) to quantify coverage. The performance in this regard for BVI-HFR is reported in Table I. When comparing to existing video databases [39], BVI-HFR has excellent uniformity of coverage, indicating that the database provides unbiased scrutiny across the range of content types. While the relative range and relative total coverage of BVI-HFR compares favourably to existing video databases, these values have been diminished by the fact that BVI-HFR was explicitly designed to highlight temporal variations. As a consequence it contains large amounts of motion blur, and a range of dynamic textures. This culminates in reduced spatial complexity and motion prediction accuracy (which affects $TI_{MV}$) - thus modulating the range and the relative total coverage of the descriptors.

While this analysis is important, it does not account for how representative a video database is of typical broadcast content. While a video database may exhibit excellent coverage over the descriptors, it may consist entirely of superficial or novel source sequences. Therefore to ensure that methods benchmarked on a video database are applicable to real-world applications, and that lasting conclusions are relevant to a wider audience, it is crucial that a video database contains content which is archetypical of broadcast content.

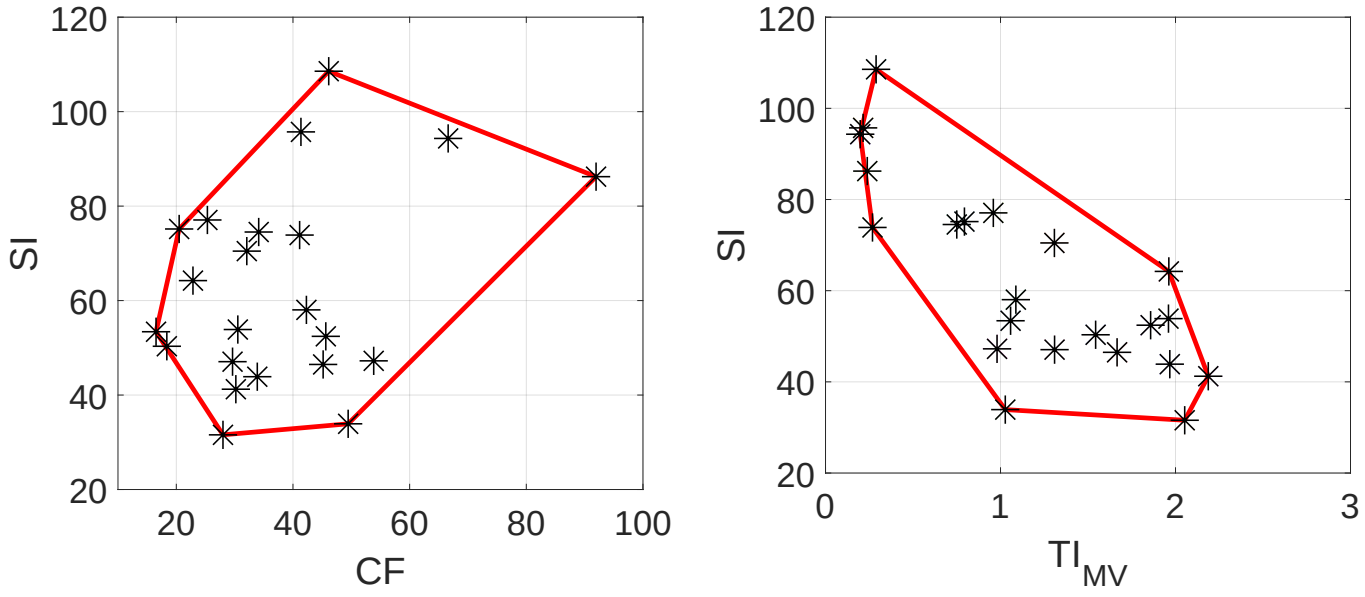



Fig. 2: Coverage of BVI-HFR where: (left) SI vs CF, (right) SI vs $TI_{MV}$.

Using the method outlined by Moss *et al.* [43], we can ascertain whether there are any similarities between BVI-HFR and BBC broadcast content (14075 sequences from BBC Redux [49]). In order to acheive this we need to calculate two further descriptors: DTP (Dynamic Texture Parameter) [50] and TP (Texture Parameter) [50] to estimate complex and irregular motion, and static textures respectively. The five descriptors (including SI, $TI_{MV}$ and CF) after normalisation are then projected using principal component analysis (PCA) into five-dimensional factor space, with the corresponding labels: *Naturalness*, *Movement*, *Brightness*, *Contrast* and *Saturation*. Fig. 4 presents the cumulative distribution of these factors, comparing BVI-HFR to BBC broadcast content.

A two-sample Kolmogorov-Smirnov (K-S) test can be used to assess whether two databases come from the same underlying cumulative distribution. The K-S statistic values when comparing BVI-HFR and BBC Redux for the five factors are 0.109, 0.113, 0.117, 0.205 and 0.154, corresponding to the $p$-values 0.946, 0.929, 0.906, 0.283 and 0.638 respectively. No significant difference between the two distributions ($p < 0.05$) is reported for all five factors. While we cannot say that BVI-HFR is a perfect representation of BBC broadcast content using this statistical test, we can conclude that the underlying distribution of BVI-HFR is similar to broadcast content.

## IV. Temporal Down-Sampling

Capturing exactly the same scene at different frame rates is prohibitively difficult for natural content. Therefore in order to enable comparisons over a range of frame rates, a method of generating lower frame rate content is required. This will be a key asset for high frame rate formats, as it will allow for conversion to legacy formats, and can be employed within future adaptive formats. Appendix A outlines a common framework for temporal down-sampling, while in this section we investigate the effect of temporal down-sampling across a range of modalities.

### A. Spectral Analysis

The video capture and display process can be expressed using a series of spatio-temporal filters and sampling operations [10]. While predominately used to model the visibility of motion artefacts using the ‘window of visibility’ [51], here this frequency representation is used to illustrate the effect that video acquisition and temporal down-sampling have on the video signal when abstracted to its simplest forms (although the same analysis could be applied to more complex scenes).

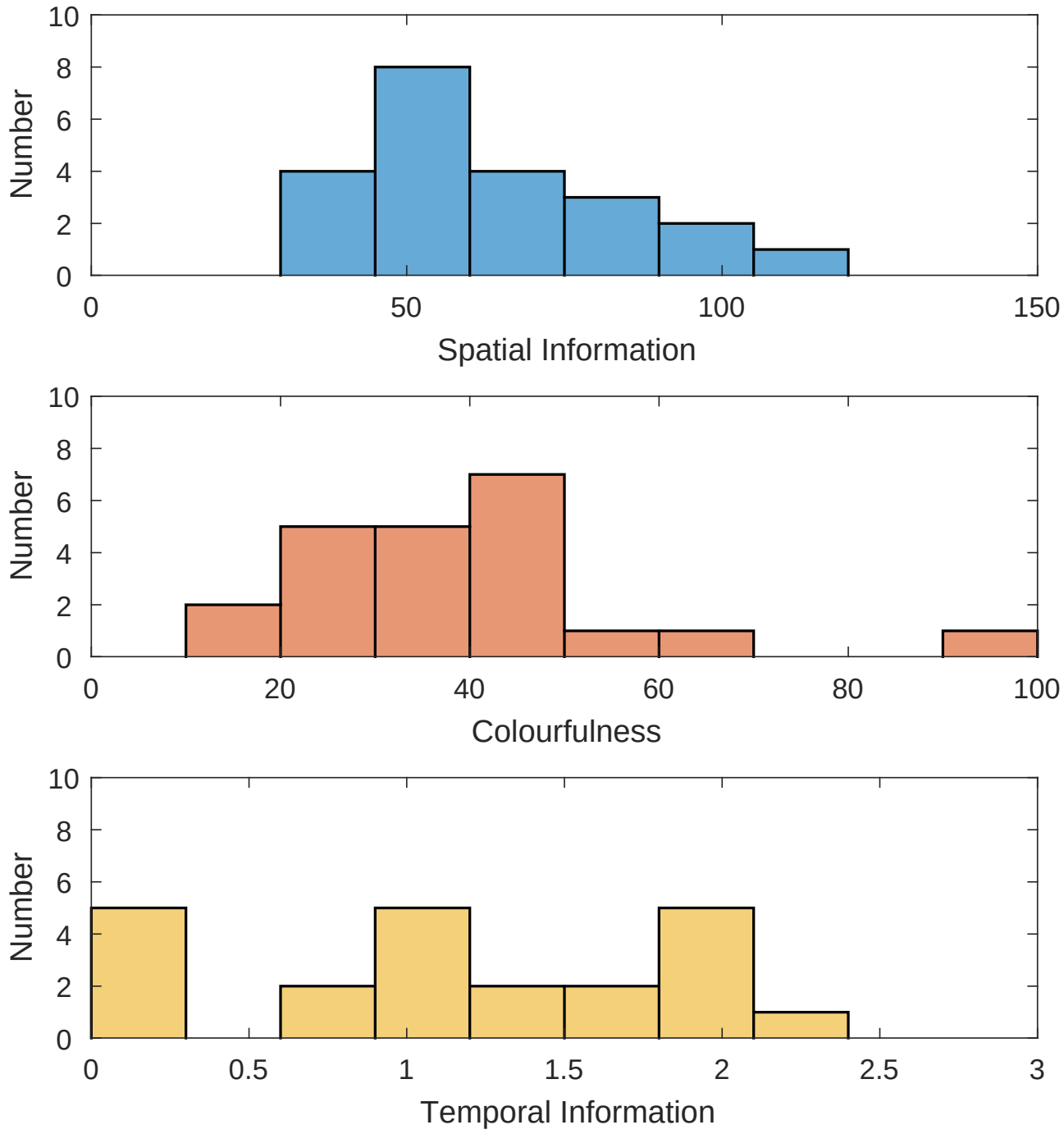



Fig. 3: The distribution of the three low-level descriptors in BVI-HFR.

TABLE I: The relative range, uniformity of coverage and relative total coverage for the source sequences (120 fps) in BVI-HFR.

| | SI | $TI_{MV}$ | CF |
|---|---|---|---|
| Relative Range | 0.51 | 0.66 | 0.75 |
| Uniformity of Coverage | 0.85 | 0.90 | 0.77 |
| Relative Total Coverage | 0.36 | | |

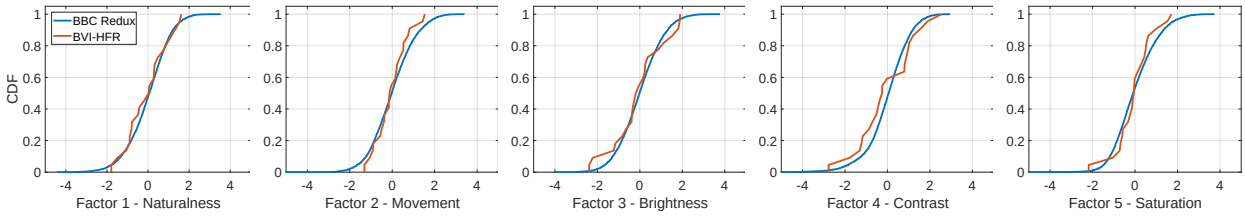



Fig. 4: The cumulative distribution function (CDF) of the five factors identified in [43]. The blue curves show the results from a high density sampling of BBC broadcast content, while the red curves are calculated from the source sequence (120 fps) in the BVI-HFR video database.

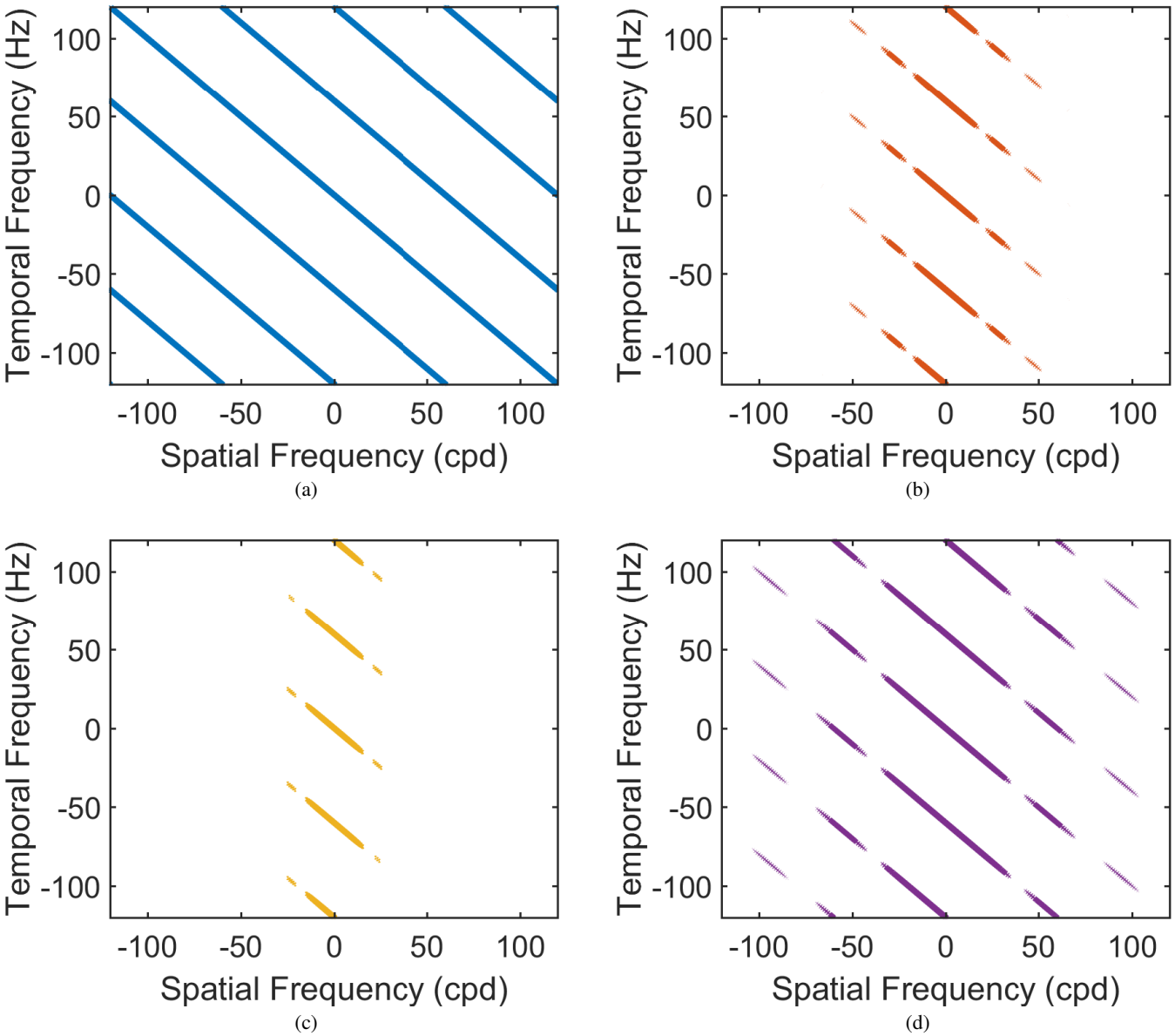



Fig. 5: (a) The spectra of a moving line sampled at 60 fps, (b) the same line but captured using a video camera (360°), and the effect of temporal down-sampling from 120 to 60 fps by (c) averaging and (d) dropping frames.

A video camera captures a scene using a shutter, which remains open for a proportion of the frame duration (called the shutter angle). The camera shutter integrates the incoming luminance signal during each exposure, and therefore acts as a temporal filter. The frequency response of this filter is [10]:

$$S(f) = \text{sinc}\left(\frac{d\pi f}{F}\right) \tag{1}$$

where $d$ is the normalised shutter angle (divided by 360), $f$ is temporal frequency in Hz and $F$ is frame rate in Hz (fps).

Temporal down-sampling also filters the video signal temporally, where the resulting frequency response is dependent on the weights in Eq. 8. For the case of averaging frames, the frequency response of the filter can be modelled as [10]:

$$A(f) = \text{sinc}\left(\frac{k\pi f}{F}\right) \tag{2}$$

Dropping frames is equivalent to capturing at the lower frame rate - albeit with the reduced shutter angle $d/k$.

Using the method outlined by Watson [10], we can analyse the effect of these two down-sampling methods on the spectra of a simple line (used for simplicity and to aid visualisation).

Fig. 5 (a) shows the spatio-temporal frequency spectra of a line moving at constant speed when sampled at 60 fps. The original spectrum passes through (0,0), while the other lines in the figure are aliases (spectral replicas). Fig. 5 (b) shows the same moving line, but captured natively at 60 fps with a fully open shutter. Fig. 5 also show the case when the moving line is captured at 120 fps, and then temporally down-sampled by (c) averaging frames and (d) dropping frames to 60 fps.

For the case of averaging frames, there is a greater attenuation of high spatio-temporal frequencies compared to capturing natively, and the higher the down-sample factor ($k$), the greater this attenuation (Eq. 2). Attenuation of the original spectrum will result in the visual sensation of motion blur, while attenuation of the aliases will diminish perceptible temporal aliasing artefacts. Therefore, ideally a rectangular pre-filter would be applied to the video signal before capture, as to only preserve perceptible frequencies below the Nyquist frequency ($F/2$). However as such a filter would be non-causal and have an infinite delay, it would not be feasible.

(a) (b) (c) (d) (e) (f)

Fig. 6: The *cyclist* sequence at: (a) 120, (b) 60, (c) 30 and (d) 15 fps; and the *catch* sequence at (e) 120 and (f) 15 fps (all generated by averaging frames).

Fig. 6 shows a sample frame from the *cyclist* and *catch* sequence at a range of frame rates after averaging frames. There is a clear increase in motion blur at lower frame rates when averaging frames, and it can also be observed that the faster the speed of an object, the larger the extent of any motion blur. As the faster the speed, the higher the corresponding temporal frequency [52].

There is less attenuation across the whole frequency spectrum when dropping frames. This will as a consequence reduce the visibility of motion blur, but at the expense of increased levels of temporal aliasing. In terms of the visibility of these two motion artefacts, averaging frames and dropping frames represent the extreme cases given the constraints in Eq. 8.

A filter could be designed such that no perceptual difference existed between capturing natively at the lower frame rate, and down-sampling from the high frame rate. However such a filter would be content dependent, and would require a three-dimensional Discrete Fourier Transform (DFT) to be computed for each sequence to generate the weights in Eq. 8.

### B. Source Statistics

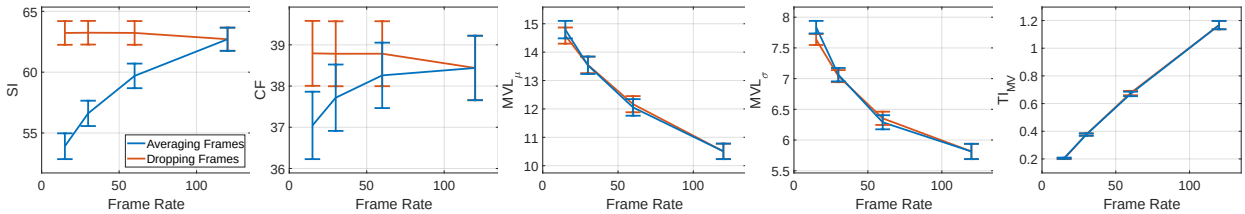



Fig. 7: The relationship between descriptors and frame rate for all the sequences in BVI-HFR. Error bars represent standard error of the mean.

While a simple moving line is a useful tool when exploring the fundamental effects of temporal down-sampling in signal processing terms, it does not account for the fact that video signals typically exhibit more complex spectral signatures [53]. Therefore in this context we can study the impact of temporal down-sampling (and to a greater extent frame rate) on the source statistics of video content, by using some the content descriptors outlined that were in Section III.

The results from Fig. 7 show that SI and CF are invariant with frame rate when dropping frames (the sequence is unchanged spatially), whereas the increased motion blur when averaging frames (Fig. 6) results in higher spatial correlation between adjacent pixels. This reduces the 'sharpness' of edges (lower SI) and the dispersion of pixel values (lower CF).

At higher frame rates, motion vectors are smaller and less varied, shown by the average ($\text{MVL}_\mu$) and standard deviation ($\text{MVL}_\sigma$) of motion vector length decreasing with frame rate. Interestingly the choice of down-sampling has a negligible effect on motion prediction - even though there is a clear spatial difference between averaging and dropping frames.

$\text{TI}_{\text{MV}}$ should be invariant with frame rate as it has been normalised by it [39]. Instead $\text{TI}_{\text{MV}}$ increases with frame rate (due to the non-linear relationship between $\text{MVL}_\mu$ and frame rate). Therefore when comparing video sequences across a range of frame rates, a different normalisation factor for average motion vector length should be used.

### C. Video Compression

An unavoidable consequence of higher frame rates is increased data rates, due to the fact that more frames need to be coded within the same time interval [19]. However in our previous work we only considered averaging frames. Therefore it is important that we also explore dropping frames, as the choice of temporal down-sampling method affects the source statistics of the video content (see Fig. 7).

The 22 source sequences (120 fps) in BVI-HFR were temporally down-sampled by both averaging and dropping frames to 60, 30 and 15 fps. The resulting 154 video sequences were then encoded using the HEVC [54] reference codec (HM 16.4) using five Quantisation Parameters (QP): 22, 27, 32, 37, 42; and three compression configurations: All Intra (AI), Low Delay (LD) and Random Access (RA) [55] (2310 in total). The coding structure of these three

TABLE II: The coding structure of the three HEVC compression configurations: All Intra (AI), Low Delay (LD) and Random Access (RA).

| | IntraPeriod | IntraQPoffset | GOPSize | GOP |
|---|---|---|---|---|
| AI | 1 | - | 1 | I |
| LD | Only First | -1 | 4 | PPPP |
| RA | 32 | -3 | 16 | Hierarchical |

configurations can be viewed in Table II. The temporal overheard of compression was reduced by encoding only the middle three-seconds of each sequence [19].

In order for a fair comparison across the tested frame rates, we need to ensure that the intra period for Random Access mode spans a certain length of time, rather than a fixed number of frames. In order to achieve this, we normalise the reported bitrates to an equivalent intra period of 32 at 30 fps (≈1s).

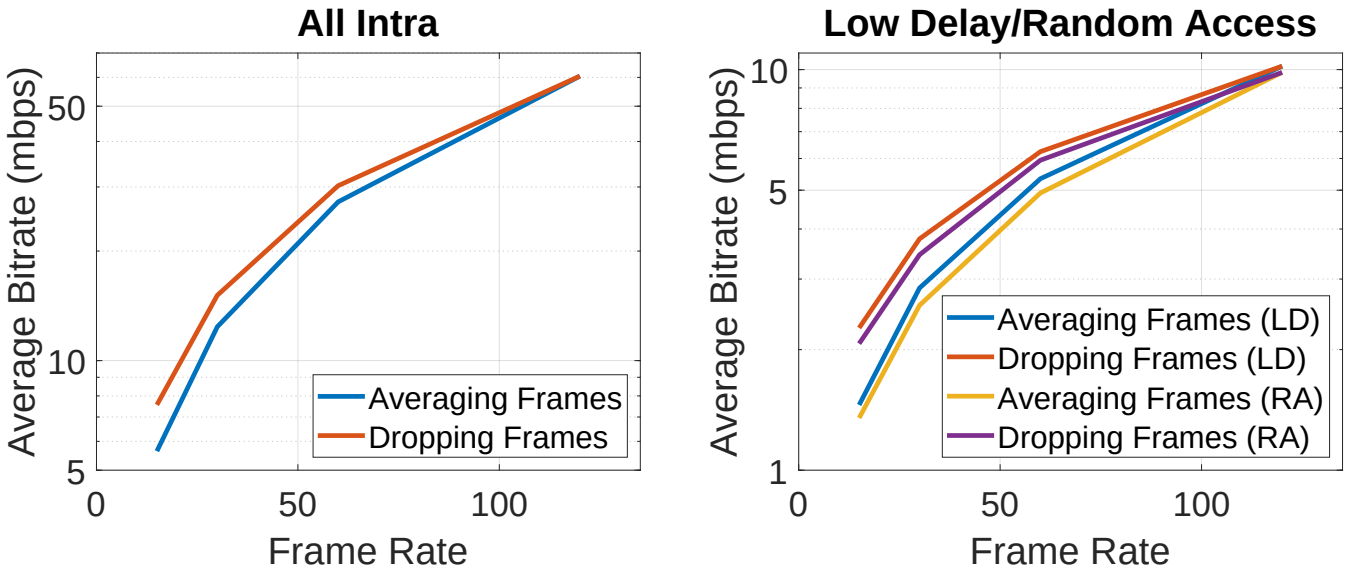



Fig. 8: The relationship between frame rate and average bitrate (mbps) for (left) All Intra and (right) Low Delay/Random Access.

The relationship between frame rate and average bitrate for All Intra, Low Delay and Random Access configurations is shown in Fig. 8, while Table III summarises the distribution of bits by the HM encoder per frame. The reported values are the average of all tested QPs and video sequences with the same frame rate.

TABLE III: Average kb consumed by the HM encoder per frame. FR = Frame Rate, R = Residual Coding, MP = Motion Prediction, I = Intra Direction, MI = Merge Index, MS = Mode Signalling, P = Partitioning and O = Other.

| | | **Averaging Frames** | | | | | | | **Dropping Frames** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FR | R | MP | I | MI | MS | P | O | R | MP | I | MI | MS | P | O |
| **AI** | 15 | 345 | - | 24 | - | - | 6 | 1 | 459 | - | 34 | - | - | 9 | 2 |
| | 30 | 378 | - | 27 | - | - | 7 | 1 | 459 | - | 34 | - | - | 9 | 2 |
| | 60 | 415 | - | 30 | - | - | 8 | 2 | 459 | - | 34 | - | - | 9 | 2 |
| | 120 | 459 | - | 34 | - | - | 9 | 2 | 459 | - | 34 | - | - | 9 | 2 |
| **LD** | 15 | 81 | 5 | 3 | 2 | 3 | 3 | 0 | 122 | 9 | 6 | 3 | 4 | 6 | 1 |
| | 30 | 76 | 5 | 3 | 3 | 3 | 4 | 1 | 100 | 8 | 5 | 3 | 4 | 5 | 1 |
| | 60 | 70 | 5 | 3 | 3 | 3 | 4 | 1 | 82 | 6 | 3 | 3 | 4 | 5 | 1 |
| | 120 | 67 | 4 | 3 | 3 | 3 | 4 | 1 | 67 | 4 | 3 | 3 | 3 | 4 | 1 |
| **RA** | 15 | 75 | 4 | 4 | 2 | 2 | 3 | 0 | 111 | 8 | 7 | 2 | 4 | 5 | 1 |
| | 30 | 70 | 5 | 4 | 2 | 2 | 3 | 0 | 92 | 7 | 6 | 2 | 3 | 4 | 1 |
| | 60 | 67 | 4 | 4 | 2 | 2 | 3 | 0 | 78 | 6 | 5 | 2 | 3 | 4 | 1 |
| | 120 | 66 | 4 | 4 | 2 | 2 | 3 | 1 | 66 | 4 | 4 | 2 | 2 | 3 | 1 |

*1) Averaging Frames:* For All Intra mode, the average number of bits increases in all areas with frame rate, suggesting that the increased spatial complexity (SI) associated with higher frame rates (see Fig. 7) is harder to encode. The use of motion prediction for Low Delay and Random Access modes dramatically decreases the number of bits, and results in the number of bits per frame decreasing with frame rate.

*2) Dropping Frames:* Compared to averaging frames, more bits are consumed when dropping frames. The increased spatial complexity for the same frame rate (see Fig. 7) results in finer partitioning, more intra-prediction modes and higher valued high frequency DCT coefficients. The number of bits consumed by HM is integer invariant with frame rate for All Intra mode, as the same frames are effectively encoded.

A Wilcoxon signed-rank test[1] shows that averaging frames has a significantly ($p < 0.05$) smaller bitrate compared to dropping frames (ignoring 120 fps) for all compression modes:

- All Intra: $Z$ = -4.1, $p$ = 0
- Low Delay: $Z$ = -4, $p$ = 0
- Random Access: $Z$ = -4.1, $p$ = 0

This indicates that the choice of down-sampling method, and as a consequence the shutter angle (as dropping frames decreases the shutter angle), has a significant impact on bitrate.

## V. Subjective Evaluations

An experiment which quantifies the relationship between frame rate and perceptual quality is described in this section.

### A. Test Sequences

As discussed in Section IV the choice of temporal down-sampling method will effect the visibility of both temporal aliasing and motion blur, and subsequently may impact perceived quality. Given the fact that we are investigating 22 sequences across 4 frame rates (15, 30, 60 and 120 fps), in order to manage experimental complexity, we exclusively employed frame averaging to generate the lower frame rate sequences in BVI-HFR. This is due to a number of considerations: i) The source sequences were captured with a fully open shutter angle - thus mitigating any ghosting artefacts. ii) While averaging frames may introduce more motion blur into the scene, the use of dropping frames may result in judder/strobing artefacts. During informal experimentation we discovered that participants broadly preferred averaging frames to dropping frames - especially at the lowest tested frame rate of 15 fps (postulated to be due to a very short shutter angle being simulated). (iii) Averaging frames has been adopted by a number of broadcasting companies including BBC [12] for their high frame rate video processing.

All the sequences used in the experiment were uncompressed in order to reduce the number of independent variables tested.

### B. Experimental Setup and Methodology

A calibrated BenQ XL2720Z LCD monitor (hold-type) with a peak luminance of 200 cd/m$^2$, a spatial resolution of 1920×1080 (measuring 59.8×33.6 cm), a static contrast ratio of 1000:1, and a refresh rate of 120 Hz was used in this experiment. This display was connected to a PC with Matlab R2013a and PsychToolbox 3.0 [56]. The viewing distance was chosen as 168 cm (5 H), as to ensure that the spacing between adjacent pixels is less than the spatial acuity of the human visual system [1]. The viewing environment conformed to the laboratory conditions outlined in BT.500-13 [57].

Prior to the experiment, each participant took part in a training session to acclimatise them with the testing process. This consisted of explaining the testing methodology, while answering any queries or concerns. When ready, participants viewed four sequences not contained within BVI-HFR (but captured using the same parameters) across the range of tested frame rates, and recorded their subjective opinion using the methodology described below. These were subsequently discarded.

A complete session lasted no longer than 30 minutes, and involved viewing all 88 test sequences (see Section V-A) using a single stimulus methodology. Each trial consisted of the participant viewing a 3 s mid-level grey screen

[1]A paired t-test was not used as the data violated the normality assumption.

before viewing a randomly selected sequence. Participants' then recorded their opinion on a continuous quality scale from 0 (bad) to 5 (excellent) [57]. A single-stimulus, rather than a double-stimulus methodology, was chosen as it considerably reduces the length of the experiment (88 versus 132 presentations). A continuous scale was used to allow for sufficient granularity between responses to investigate content dependence.

Fifty-one[2] undergraduate and postgraduate students (33 male and 18 female) from the University of Bristol were paid to participate in the experiment. The average age ($\pm\sigma$) of the participants was 25.6 $\pm$ 3.8 years. They all had normal or corrected-to-normal colour vision (verified with a Snellen chart). No participants were removed during screening [57].

### C. Analysis of Opinion Scores

The opinion scores collected in the experiment were scaled to the range 0-100 (from worst to best). Fig. 9 presents the distribution of these scaled opinion scores. Higher frame rates tend to lead to increases in perceived quality (albeit with a few outliers). Clear content dependence can also be observed.

The cumulative distribution of opinion scores is shown in Fig. 10 (left). Across the range of opinion scores, there is no overlap between these distributions - although the differences between them diminish as frame rates increase. The mean (MOS) and standard deviation of opinion scores for each of the 88 test video sequences is shown in Fig. 10 (right). There is a greater consensus in opinion at the higher frame rates, shown by the smaller and more compact standard deviations.

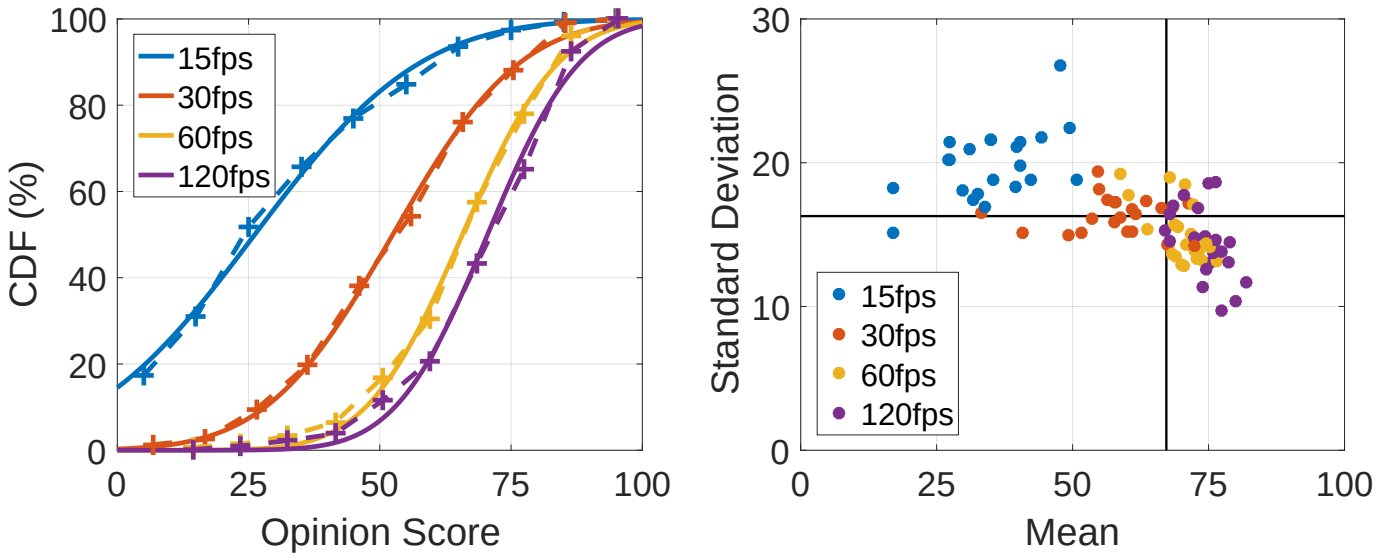



Fig. 10: (left) Cumulative distribution of opinion scores across the tested frame rates, and (right) the relationship between the mean and standard deviation of the opinion scores. The black lines represent the respective medians.

### D. The Influence of Frame Rate on Opinion Scores

Fig. 11 (left) presents the measured relationship between MOS and frame rate. The increase in perceived quality with frame rate is postulated to be due to a reduction in the visibility of both motion blur and temporal aliasing artefacts [11]. Improvement in quality beyond 120 fps is predicted for some content [29], as the faster the speed of a stimulus relative to the viewer, the higher the frame rate required to eliminate perceptible motion artefacts [10]. The effect of diminishing returns with increased frame rates can be observed. A one-way repeated measures ANOVA with Greenhouse-Geisser correction (as the assumption of sphericity was violated) shows that the effect of frame rate is statistically significant ($p<0.05$) with respect to MOS: $F(1.2, 58.5) = 222$, $p \approx 0$.

[2]The subjective evaluations that were originally provided with the BVI-HFR video database was based on 29 participants [18]. We have conducted further experimentation here to increase the number of participants to 51.

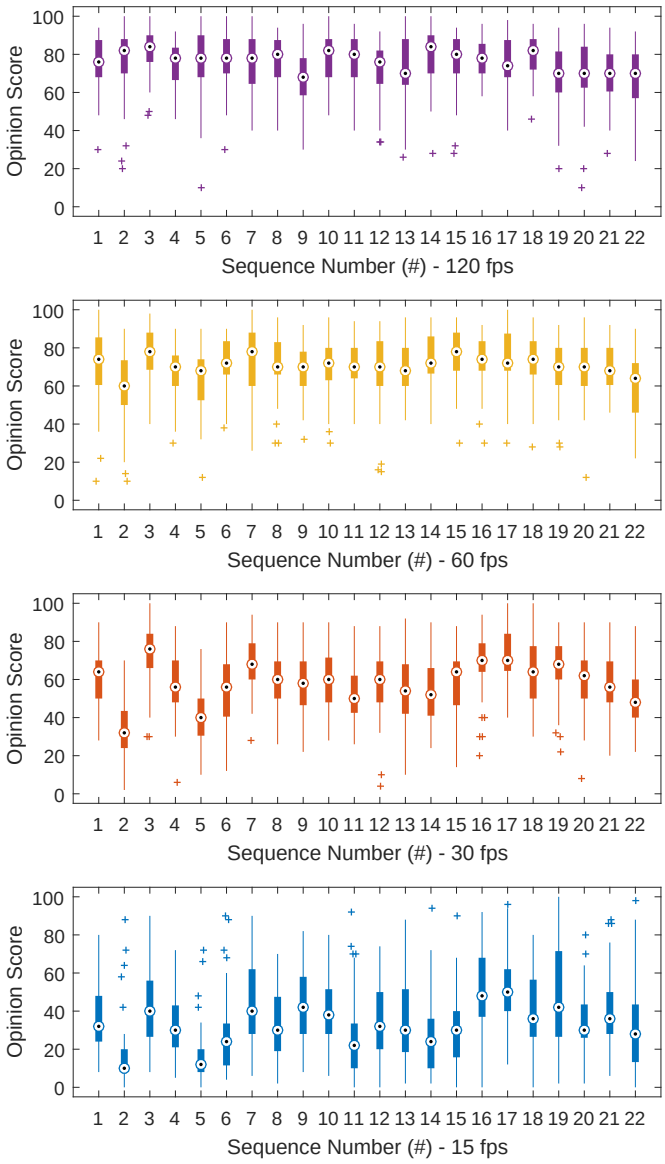



Fig. 9: Boxplots showing the distribution of opinion scores for all the sequences in BVI-HFR over the range of tested frame rates. The boxes representing the interquartile range (IQR) of the data, the whiskers are $\pm 2.7\sigma$, the target is the median and outliers are denoted as '+'.

### *E. Content Dependence*

Fig. 11 (right) shows the influence of camera motion on MOS. At 120 fps, video sequences with camera motion have higher a MOS compared to if no camera motion is present, whereas at lower frame rates, this ordering is reversed. The disparity between the two cases is predicted to be due to camera motion creating global rather than just local temporal distortions i.e. there will typically be more motion artefacts apparent in the visual field. Because of this, increasing frame rates will have a greater impact on visual quality when camera motion and/or large global motions are present.

By using the content descriptors from Section III, we can attempt to estimate the factors which most influence visual quality. However given that $\text{TI}_{\text{MV}}$ has been normalised by the dependent variable (frame rate), it would be unfair to include this descriptor in this analysis (due to the dependence of MOS on frame rate). Therefore instead we will use the average ($\text{MVL}_{\mu}$) and standard deviation ($\text{MVL}_{\sigma}$) of motion vector length to characterise motion in the sequence.

A multiple linear regression was performed to predict MOS using SI, CF, $\text{MVL}_{\mu}$, $\text{MVL}_{\sigma}$, TP and DTP, and the results show that these descriptors statistically significantly predicted MOS: $F(6, 81) = 5.69$, $p \approx 0$, $R = 0.544$.

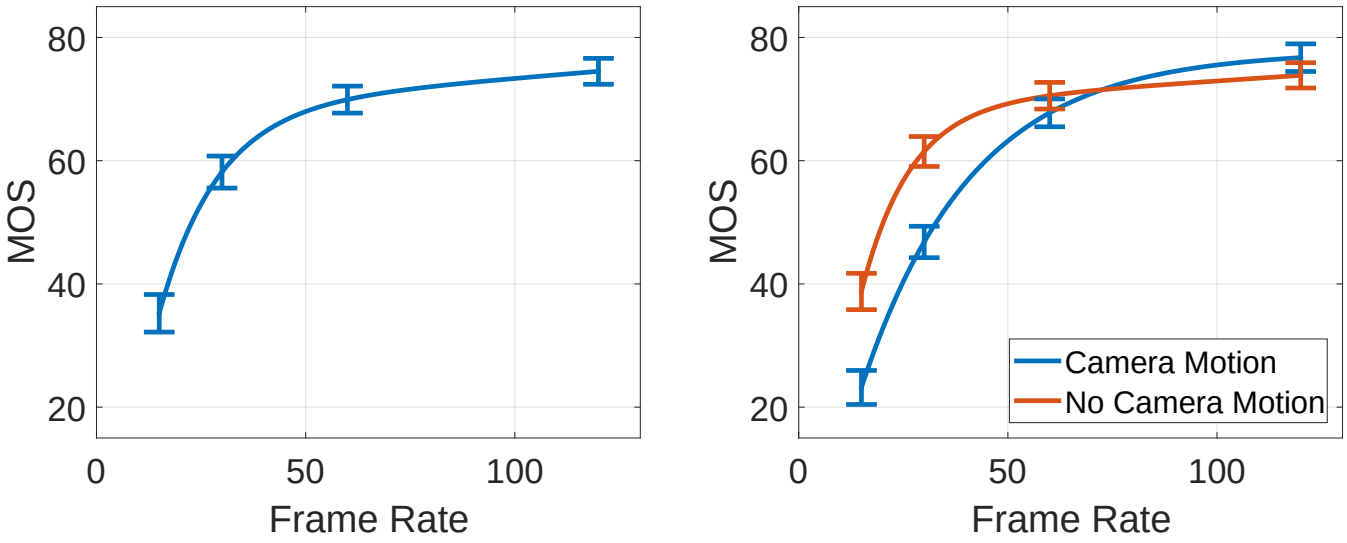



Fig. 11: (left) The measured relationship between MOS and frame rate, and (right) the influence of camera motion. A two-term exponential fitting curve is used to model the data. Error bars represent standard error of the mean.

TABLE IV: The coefficients ($\beta$) and standardised coefficients ($\hat{\beta}$) from the multiple linear regression, alongside the results from a t-test showing whether the descriptor has a significant effect on the output of the prediction model.

| | Descriptors | | | | | |
|---|---|---|---|---|---|---|
| | SI | CF | $MVL_\mu$ | $MVL_\sigma$ | TP | DTP |
| $\beta$ | -0.17 | -0.06 | -0.26 | -2.35 | -1.39 | -0.01 |
| $\hat{\beta}$ | -0.22 | -0.07 | -0.10 | -0.36 | -0.10 | -0.42 |
| $t$ | -1.05 | -0.54 | -0.49 | -1.85 | -0.53 | -3.67 |
| $p$ | 0.28 | 0.59 | 0.62 | 0.07 | 0.60 | 0 |

The constant value (intercept) for the prediction model is 96.43.

The magnitudes of the standardised coefficients $|\hat{\beta}|$ can be used to predict the influence that the descriptors have on MOS. From Table IV we can see that DTP has the greatest impact on the subjective evaluations, and was the only descriptor to significantly affect the prediction of MOS ($p < 0.05$). The negative coefficient for DTP indicates that MOS is reduced as the amount of complex and irregular motion in the scene in increased. While this indicates that the presence of dynamic textures has an effect on visual quality, this relationship may also be attributed to the greater temporal correlation between frames at high frame rates. This is because the DTP descriptor is based on the second derivative of the motion vector field [50], and when the interval between frames reduces, motion vectors become smaller (Fig. 7), and consequently reduces the magnitude of the second derivative.

The prediction model after multiple linear regression exhibits poor linear correlation (LCC) with visual quality ($R = 0.544$). As a result we will now investigate whether more advanced models can accurately and robustly predict MOS.

## VI. Quality Metrics

An accurate video quality metric, which can successfully characterise the relationship between perceived quality and frame rate, will be an asset for future adaptive and immersive video formats, as it will allow optimal frame rates to be selected in a content dependent manner. In this section a selection of generic and frame rate dependent image and video quality metrics will be benchmarked on the subjective evaluations that were collected in Section V.

### A. Methodology

In order to facilitate a fair comparison between the metrics and to reduce non-linearities [58], predictions are fitted with a four-parameter logistic function. Four evaluation metrics: Spearman Rank Correlation Coefficient (SROCC),

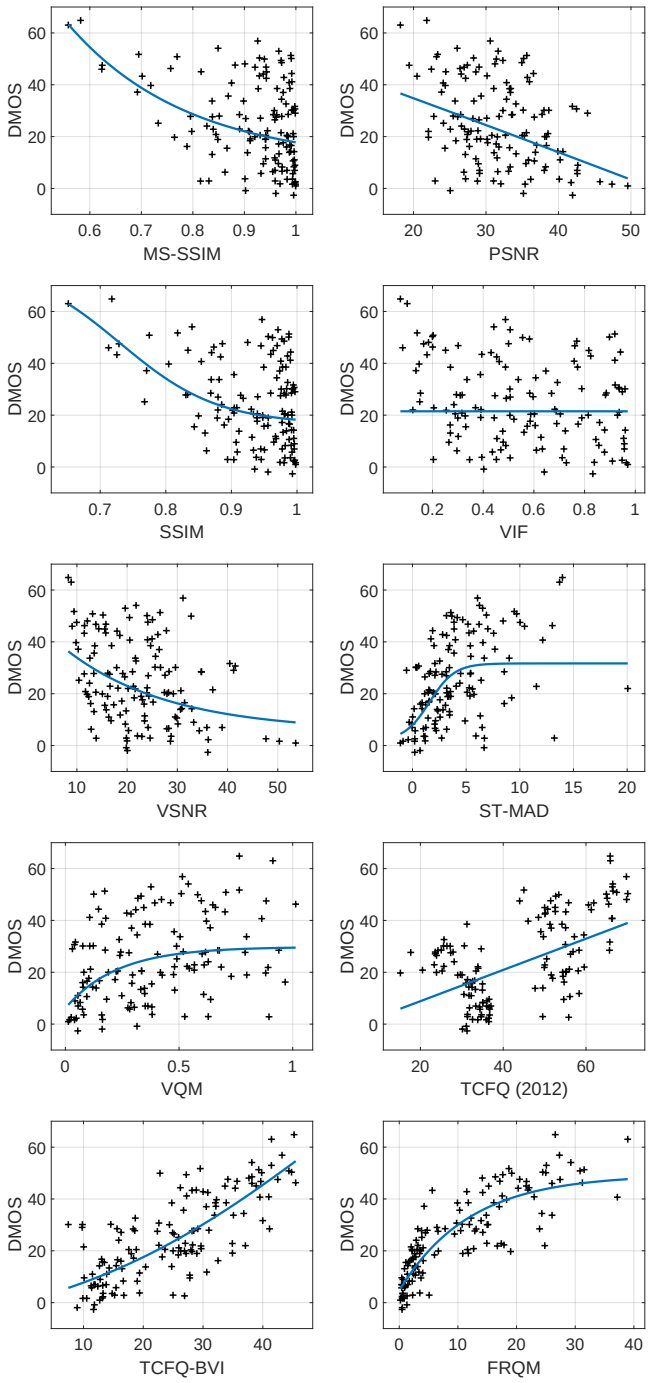



Fig. 12: Scatter plots of DMOS versus the objective valuations of various quality metrics, along with the non-linear fitting curves (blue line).

TABLE V: The statistical performance and complexity of all the tested quality metrics for BVI-HFR. The best performing metric for each row is **bold**.

| | Image Quality Metrics | | | | | Video Quality Metrics | | Frame Rate Dependent | |
|---|---|---|---|---|---|---|---|---|---|
| Metric | MS-SSIM | PSNR | SSIM | VIF | VSNR | ST-MAD | VQM | TCFQ-BVI | FRQM |
| SROCC | 0.318 | 0.347 | 0.280 | 0.250 | 0.322 | 0.507 | 0.374 | 0.739 | **0.885** |
| LCC | 0.413 | 0.399 | 0.414 | 0.264 | 0.377 | 0.520 | 0.387 | 0.767 | **0.863** |
| OR | 0.636 | 0.644 | 0.651 | 0.689 | 0.689 | 0.629 | 0.644 | 0.470 | **0.386** |
| RMSE | 15.04 | 15.07 | 15.03 | 16.62 | 15.28 | 14 | 15.16 | 10.41 | **8.191** |
| Complexity | 14 | **1** | 11 | 15 | 19 | 1256 | 108 | 1876 | 30 |

Pearson Linear Correlation Coefficient (LCC), Outlier ratio (OR) and Root Mean Squared Error (RMSE) [58], are used to appraise accuracy (LCC, RMSE), monotonicity (SROCC) and consistency (OR). A platform independent complexity score is calculated by dividing the average execution time of the quality metric by the average execution time of PSNR. Differential mean opinion scores (DMOS) are typically used for analysis, and can be calculated from MOS as follows:

$$\text{DMOS} = \text{MOS}_H - \text{MOS}_L \tag{3}$$

where $\text{MOS}_H$ and $\text{MOS}_L$ represent the high (reference) and low frame rate version of the sequence respectively.

As a consequence of using a single stimulus methodology, we can compute the difference in MOS (DMOS) for a range of reference frame rates. Alongside the 120 fps source sequences, we also use 60 and 30 fps as references ($\text{MOS}_H$) e.g. 60 fps can be used as a reference for 30 and 15 fps. By doing this we increase the number of DMOS from 66 to 132.

### *B. Generic Quality Metrics*

Initially, image and video quality metrics which do not explicitly take frame rate variation into account will be evaluated.

The quality metrics considered here are: MS-SSIM [59], PSNR [60], SSIM [61], VIF [62], VSNR [63], ST-MAD [64] and VQM [65]. We artificially increase the frame rate of the down-sampled sequences through repeating frames to obtain the same frame rate as the reference (e.g. we repeat every frame twice in a 60 fps sequence to increase it to 120 fps). This enables a frame-by-frame comparison as required by most quality metrics[3].

While some of these quality metrics are designed to evaluate compressed content and/or spatial distortions, their performance will provide a useful benchmark for advanced methods.

Fig. 12 shows DMOS versus the objective valuations of these quality metrics, where the blue line is the non-linear fitting curve. The statistical performance and relative complexity of all the generic quality metrics is reported in Table V.

All the image quality metrics show similar performance across the four evaluation metrics - which may be expected as they explicitly ignore the temporal dimension (motion blur is the only spatial difference between frame rates after averaging frames). The video quality metric ST-MAD reports the best statistical performance, but at the expense of the highest complexity of all the generic quality metrics (1256 times over PSNR). All these metrics compare unfavourably to their frame rate dependent counterparts, and therefore we can conclude that they are unsuitable for this specific application.

### *C. TCFQ and MNQT (Modified)*

TCFQ [41, 47] and MNQT [42] are related video quality metrics[4] that explicitly relate reductions in frame rate to perceptual quality. Both TCFQ and MNQT are defined as:

$$\hat{q} = \text{MOS}_H \frac{(1 - e^{-cF_L/F_H})^\beta}{1 - e^{-c}} \tag{4}$$

where $F_H$ is the frame rate (fps) of the reference, $F_L$ is the frame rate we wish to down-sample to. $\beta$ is a fixed parameter, and is typically chosen to be 1 for TCFQ and 0.63 for MNQT. The model parameter $c$ is a weighted sum of video features, calculated at the reference frame rate ($F_H$):

$$c = \gamma(0) + \sum_{i=1}^{N} \gamma(i) V[i] \tag{5}$$

where $\gamma$ is a weight and $V$ is a video feature (e.g. SI).

[3]In situation where the frame rates cannot be increased in this way, e.g. when the two frame rates are not divisible, either both frame rates could be increased to the least common multiple by repeating frames, or a coarse approximation based on a smaller subset of temporally matching frames could be computed.

[4]The method proposed by Huang *et al.* [48] is intentionally ignored here, as a satisfied user ratio rather than a continuous quality scale methodology was used to collect subjective opinions - making comparisons difficult.

TCFQ and MNQT predict normalised MOS (NMOS). Therefore in order for a comparison to be made with the other metrics, NMOS needs to be first converted to DMOS:

$$\begin{aligned} \text{NMOS} &= \text{MOS}_L / \text{MOS}_H = \hat{q} \\ \text{MOS}_L &= \hat{q} \cdot \text{MOS}_H \\ \text{DMOS} &= \text{MOS}_H - \hat{q} \cdot \text{MOS}_H \end{aligned} \tag{6}$$

Three distinct methods, which all use different video features ($V$) and weights ($\gamma$), have been proposed in literature for calculating the model parameter $c$. These will be referred to as TCFQ (2011) [47], TCFQ (2012) [41] and MNQT (2014) [42]. The statistical performance when using the proposed weights (P) is reported in Table VI. All three methods exhibit poor prediction accuracy, which is likely due to the weights being calculated on video content with relatively low spatial (up to 704×480) and temporal (up to 30 fps) resolutions.

TABLE VI: The statistical performance when comparing the proposed (P) and the updated (U) weights for the video quality metrics TCFQ (2011 and 2012) and MNQT (2014). The best performing metric for each row is **bold**.

| Metric | TCFQ (2011) | | TCFQ (2012) | | MNQT (2014) | |
|---|---|---|---|---|---|---|
| | P | U | P | U | P | U |
| SROCC | 0.545 | 0.603 | 0.337 | **0.739** | 0.533 | 0.621 |
| LCC | 0.692 | 0.691 | 0.421 | **0.767** | 0.682 | 0.708 |
| OR | 0.667 | 0.629 | 0.811 | **0.470** | 0.667 | 0.629 |
| RMSE | 11.77 | 11.77 | 14.86 | **10.41** | 11.93 | 11.49 |
| Complexity | 1147 | | 1876 | | 1945 | |

The subjective evaluations provided with BVI-HFR can be used to update the weights ($\gamma$) for TCFQ and MNQT in an attempt to increase prediction accuracy. For the same video features, we calculate using a genetic algorithm the weights which minimise the squared error with DMOS during 10-fold cross-validation. In order for a fair comparison to be made, the lower frame rate version was upsampled to the same frame rates as the reference by repeating frames. As reported in Table VI, the statistical performance for all the methods is better when using the updated weights (U) - notably for TCFQ (2012). This is achieved without any additional complexity. Fig. 12 shows a comparison in the relationship between the quality metric predictions and DMOS when using the proposed weights for TCFQ (2012) and the updated weights (denoted as TCFQ-BVI). The predictions of TCFQ-BVI exhibit a greater degree of monotonicity than when using the proposed weights.

TCFQ and MNQT are clearly more suitable than the generic quality metrics in the context of frame rate variations. However they are susceptible to over-fitting, are overly dependent on the resolution of the sequence, and are too sensitive to the value of the weights in Eg. 5. This is demonstrated by the variability in statistical performance in Table VI.

### *D. FRQM*

Instead of modelling the degradation in perceived quality with frame rate (Eq. 4), FRQM [20] was designed by the authors as a low complexity, bespoke quality metric that predicts the difference in perceptual video quality between content at different frame rates through multi-level temporal wavelet decomposition, subband comparison/combination and spatio-temporal pooling. In the temporal DWT decomposition stage, all luma pixels at the same spatial coordinates over a number of consecutive frames are processed simultaneously using a 1-D Discrete Wavelet Transform (Haar wavelet). The resulting high frequency (HF) subband coefficient values of both the reference and test sequences are compared to obtain HF subband differences. These are then combined over various levels. In order to characterise the non-uniformly distributed video artefacts that arise due to frame rate reduction, FRQM employs an effective pooling strategy to calculate the sequence level quality index by taking the maximum of the local mean values of combined HF subband differences (in both spatial and temporal domains).

Fig. 12 and Table V show that FRQM offers improved performance over all other tested metrics, demonstrated by higher correlation coefficients, fewer outliers and lower prediction errors[5]. This is achieved with relatively low complexity (30 times greater than PSNR). These results indicate that the temporal HF subband energy difference

[5]The values in Table V are slightly different to those previously reported in [20], and is due to a larger number of participants being used.

calculated in FRQM correlates visually with frame rate reduction artefacts. The pooling method employed in FRQM also performs better than the simple averaging approach used by most of the other tested quality metrics.

TABLE VII: F-test results for all the quality metrics at a 95% confidence interval. A '**1**' indicates that the metric in that row is statistically superior to the metric in the column (the opposite holds for '**-1**'), while a '0' indicates that there is no statistically significant difference between the two metrics.

| | Image Quality Metrics | | | | | Video Quality Metrics | | Frame Rate Dependent | |
|---|---|---|---|---|---|---|---|---|---|
| Metric | MS-SSIM | PSNR | SSIM | VIF | VSNR | ST-MAD | VQM | TCFQ-BVI | FRQM |
| MS-SSIM | - | 0 | 0 | 0 | 0 | 0 | 0 | **-1** | **-1** |
| PSNR | 0 | - | 0 | 0 | 0 | 0 | 0 | **-1** | **-1** |
| SSIM | 0 | 0 | - | 0 | **-1** | 0 | 0 | **-1** | **-1** |
| VIF | 0 | 0 | 0 | - | 0 | 0 | 0 | **-1** | **-1** |
| VSNR | 0 | 0 | **1** | 0 | - | 0 | 0 | **-1** | **-1** |
| ST-MAD | 0 | 0 | 0 | 0 | 0 | - | 0 | **-1** | **-1** |
| VQM | 0 | 0 | 0 | 0 | 0 | 0 | - | **-1** | **-1** |
| TCFQ-BVI | **1** | **1** | **1** | **1** | **1** | **1** | **1** | - | **-1** |
| FRQM | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | - |

### *E. Significance Testing*

Table VII reports F-test results [66] for all the tested quality metrics. This test statistic can be used to ascertain whether one quality metric is statistically superior to another when tested on the BVI-HFR video database. There is generally no significant difference between the generic quality metrics. As expected, the frame rate dependent metrics are superior to their generic counterparts. FRQM is superior to every other quality metric.

## VII. Conclusions

In this paper, the influence of frame rate variation on visual quality in a number of key areas has been explored. This is achieved using the BVI-HFR video database, which contains a variety of representative scenes, motion and colours at a range of frame rates. We have concluded that it compares favourably to existing video databases, and that is representative of broadcast content.

Comparisons across frame rates has been achieved through exploiting temporal down-sampling. An in-depth analysis has shown that the choice of down-sampling methods affects the source statistics of a video sequence, the visibility of motion artefacts and has a significant impact on encoded bitrates.

A large scale subjective experiment has confirmed that frame rates have a significant impact on perceptual quality, at least up to 120 fps. Results also establish that content dependence exists - notably related to camera motion.

The subjective evaluations collected in the experiment are used to benchmark a number of generic and frame rate dependent quality metrics. The generic quality metrics (e.g. PSNR) do a poor job of predicting opinion scores, whereas the frame rate dependent video quality metrics offer mixed correlation performance. While TCFQ and MNQT are too dependent on the resolution of the video sequence, FRQM offers accurate predictions with relatively low complexity, and is shown to be statistically superior to all other tested metrics.

## VIII. Future Work

This paper has established that content dependence exists. Therefore frame rates - and other video parameters for that matter - should ideally be selected in a systematic manner, such that the source statistics of video content is exploited in order to provide perceptually optimised experiences. However before adaptive formats can be considered, interactions between video parameters and their influence throughout the entire video pipeline needs to be investigated, as the realisation of adaptive formats will represent a paradigm shift for content producers, display manufacturers and consumers alike.

The bedrock of any future adaptive format will be robust and accurate quality metrics that allow video parameters to be selected in some optimal way. However, current research methodologies in this area are somewhat flawed, in that they predominantly examine video parameters in isolation. A rate-quality analysis of current video compression schemes, over a range of video parameter values is also required, to appraise the potential advantages and feasibility of adaptive formats. Various temporal down-sampling approaches should also been investigated on their relationship to visual quality.

# APPENDIX A

## A. Temporal Down-Sampling Framework

While there are a number of prescribed methods for temporal up-sampling in literature [67–69], there are very few for temporal down-sampling [70]. Therefore we outline a commonly used method, which involves partitioning the video sequence into $k$-frame sub-sequences (Fig. 13). $k$ is usually referred to as the (temporal) down-sample factor, where:

$$k = F_H / F_L \tag{7}$$

where $F_H$ is the frame rate (fps) of the reference and $F_L$ is the frame rate we wish to down-sample to.

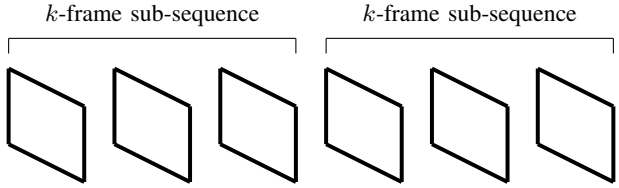



Fig. 13: A video sequence partitioned into $k$-frame sub-sequences ($k$=3).

The lower frame rate version is then generated by taking a weighted sum of all the frames in the sub-sequence:

$$\begin{aligned} Y_L &= \sum_{n=1}^{k} \alpha[n]\, Y_H[n] \\ \text{s.t.} \quad & \sum_{n=1}^{k} \alpha[n] = 1 \ \wedge \ \forall n \in [1,k] : \ \alpha[n] \geq 0 \end{aligned} \tag{8}$$

where $\alpha[n]$ is the normalised weight of frame $t$ and $Y_H[n]$ is the corresponding reference frame. This weighted sum must take place in linear light space i.e. before gamma correction.

While any valid weighting scheme can be used, the most commonly used are averaging [12] and dropping frames.

## B. Averaging Frames

The averaging frames method involves computing the mean of all frames in the $k$-frame sub-sequence i.e. a uniform weight is applied to each frame i.e. $\alpha(t) = 1/k$. The effective shutter angle of the lower frame rate version is unchanged, and this is because its temporal extent is the same as the high frame rate reference. The pre-requisite when using the averaging frames method, is that a fully open shutter (360°) must have been used to capture the content, as otherwise unwanted ghosting artefacts may be present in the video (see Fig. 14).

## C. Dropping Frames

Dropping frames simulates impulsive sampling (no pre-filter), and involves only a single frame from the $k$-frame sub-sequence being selected (usually $\alpha(1) = 1$). In this situation the shutter angle is effectively reduced by a factor $k$.

## D. Implementation

Temporal down-sampling methods can be utilised in tandem to achieve a desired look. For example: average frames from 300 fps to 50 fps, and then drop frames from 50 fps to 25 fps. Combining down-sampling methods may prove particularly useful when converting to legacy video formats.

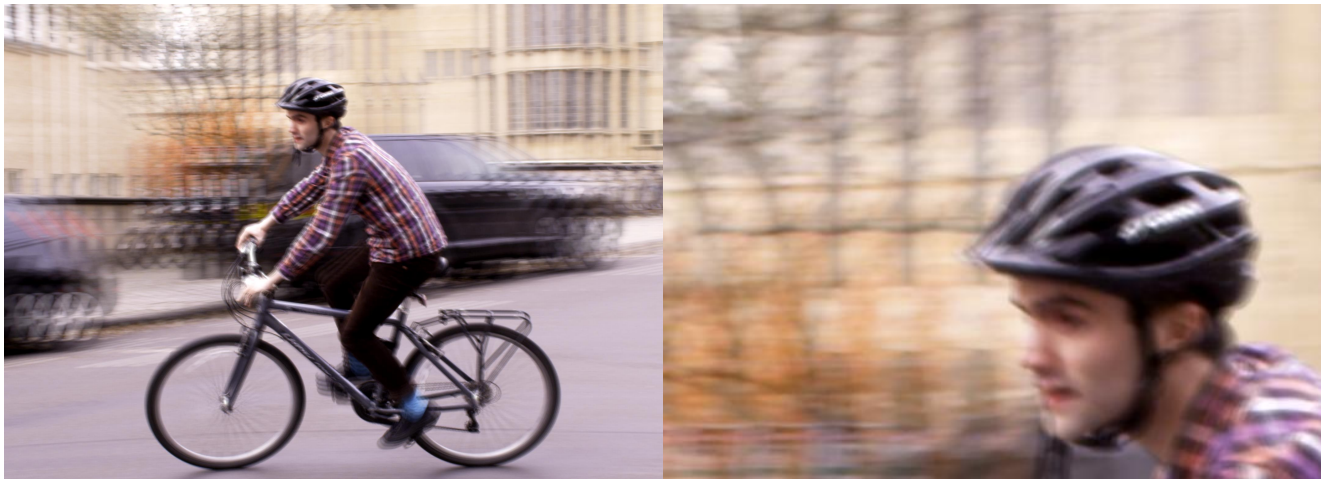


Fig. 14: The ghosting artefact that becomes apparent when a sequence with an 180 ° shutter angle is converted from 60 fps to 15 fps by averaging frames. The right frame shows a section of the original frame after 4x magnification.

## REFERENCES


[1] D. Bull, *Communicating Pictures: A Course in Image and Video Coding*. Elsevier, 2014.

[2] S. Cho, H. Kim, H. Kim, and M. Kim, "Efficient in-loop filtering across tile boundaries for multi-core HEVC hardware decoders with 4K/8K-UHD video applications," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 778–791, 2015.

[3] C. Ge, N. Wang, G. Foster, and M. Wilson, "Toward qoe-assured 4k video-on-demand delivery through mobile edge virtualization with adaptive prefetching," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2222–2237, 2017.

[4] Y. Dong, M. Pourazad, and P. Nasiopoulos, "Human visual system-based saliency detection for high dynamic range content," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 549–562, 2016.

[5] I. Hadizadeh, H.and Bajić, "Full-reference objective quality assessment of tone-mapped images," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 392–404, 2018.

[6] Y. Chen, X. Zhao, L. Zhang, and J. Kang, "Multiview and 3d video compression using neighboring block based disparity vectors," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 576–589, 2016.

[7] F. Ribeiro, J. de Oliveira, A. Ciancio, E. da Silva, C. Estrada, L. Tavares, J. Gois, A. Said, and M. Martelotte, "Quality of experience in a stereoscopic multiview environment," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 1–14, 2018.

[8] Gizmodo, "The Hobbit: An Unexpected Masterclass in Why 48 FPS Fails," February 2014. [Online]. Available: http://gizmodo.com/5969817/the-hobbit-an-unexpected-masterclass-in-why-48-fps-fails

[9] P. Barten, *Contrast sensitivity of the human eye and its effects on image quality*. SPIE press, 1999.

[10] A. Watson, "High frame rates and human vision: a view through the window of visibility," *SMPTE Motion Imaging Journal*, vol. 122, no. 2, pp. 18–32, 2013.

[11] A. Mackin, K. Noland, and D. Bull, "High frame rates and the visibility of motion artifacts," *SMPTE Motion Imaging Journal*, vol. 126, no. 5, pp. 41–51, 2017.

[12] M. Armstrong, D. Flynn, M. Hammond, S. Jolly, and R. Salmon, "High frame-rate television," *BBC Research & Development White Paper 169*, 2008.

[13] K. Noland, "The application of sampling theory to television frame rate requirements," *BBC Research & Development White Paper 282*, 2014.

[14] J. Wu, C. Yuen, N. Cheung, J. Chen, and C. Chen, "Enabling adaptive high-frame-rate video streaming in mobile cloud gaming applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 12, pp. 1988–2001, Dec 2015.

[15] D. Lanman, H. Fuchs, M. Mine, I. McDowall, and M. Abrash, "Put on your 3d glasses now: The past, present, and future of virtual and augmented reality," in *ACM SIGGRAPH 2014 Courses*, 2014, pp. 1–173.

[16] ITU-R Recommendation BT.2020-2, "Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange," 2015.

[17] A. Silva, M. Farias, and J. Redi, "Perceptual annoyance models for videos with combinations of spatial and temporal artifacts," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2446–2456, 2016.

[18] A. Mackin, F. Zhang, and D. Bull, "A study of subjective video quality at various frame rates," in *Image Processing (ICIP), 2015 22nd IEEE International Conference on*, 2015.

[19] A. Mackin, F. Zhang, M. Papadopoulos, and D. Bull, "Investigating the impact of high frame rates on video compression," in *Image Processing (ICIP), 2017 24th IEEE International Conference on*, 2017.

[20] F. Zhang, A. Mackin, and D. Bull, "A frame rate dependent video quality metrics based on temporal wavelet decomposition and spatiotemporal pooling," in *Image Processing (ICIP), 2017 24th IEEE International Conference on*, 2017.

[21] P. Bex, G. Edgar, and A. Smith, "Multiple images appear when motion energy detection fails," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, pp. 231–238, 1995.

[22] D. Hoffman, V. Karasev, and M. Banks, "Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth," *Journal of the Society for Information Display*, vol. 19, no. 3, p. 271, 2011.

[23] S. Daly, N. Xu, J. Crenshaw, and V. Zunjarrao, "A psychophysical study exploring judder using fundamental signals and complex imagery," in *SMPTE Conferences*, vol. 2014, no. 10. Society of Motion Picture and Television Engineers, 2014, pp. 1–14.

[24] A. Mackin, K. Noland, and D. Bull, "The visibility of motion artifacts and their effect on motion quality," in *Image Processing (ICIP), 2016 23rd IEEE International Conference on*, 2016.

[25] Y. Kuroki, T. Nishi, S. Kobayashi, H. Oyaizu, and S. Yoshimura, "A psychophysical study of improvements in motion-image quality by using high frame rates," *Journal of the Society for Information Display*, vol. 15, no. 1, pp. 61–68, 2007.

[26] R. M. Nasiri, J. Wang, A. Rehman, S. Wang, and Z. Wang, "Perceptual quality assessment of high frame rate video," in *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*. IEEE, 2015, pp. 1–6.

[27] M. Sugawara, K. Omura, M. Emoto, and Y. Nojiri, "Temporal sampling parameters and motion portrayal of television," *SID Symposium Digest of Technical Papers*, vol. 40, no. 1, pp. 1200–1203, 2009.

[28] R. Selfridge, K. Noland, and M. Hansard, "Visibility of motion blur and strobing artefacts in video at 100 frames per second," in *Proceedings of the 13th European Conference on Visual Media Production (CVMP 2016)*. ACM, 2016, p. 3.

[29] M. Emoto, Y. Kusakabe, and M. Sugawara, "High-frame-rate motion picture quality and its independence of viewing distance," *Journal of Display Technology*, vol. 10, no. 8, pp. 635–641, 2014.

[30] R. Allison, L. Wilcox, R. Anthony, J. Helliker, and B. Dunk, "Paper: Expert viewers preferences for higher frame rate 3D film," *Journal of Imaging Science and Technology*, vol. 60, no. 6, pp. 60 402–1, 2016.

[31] L. Wilcox, R. Allison, J. Helliker, B. Dunk, and R. Anthony, "Evidence that viewers prefer higher frame-rate film," *ACM Transactions on Applied Perception (TAP)*, vol. 12, no. 4, p. 15, 2015.

[32] Y. Kuroki, H. Takahashi, M. Kusakabe, and K. Yamakoshi, "Effects of motion image stimuli with normal and high frame rates on EEG power spectra: comparison with continuous motion image stimuli," *Journal of the Society for Information Display*, vol. 22, no. 4, pp. 191–198, 2014.

[33] B. Tag, J. Shimizu, C. Zhang, K. Kunze, N. Ohta, and K. Sugiura, "In the eye of the beholder: The impact of frame rate on human eye blink," in *Human Factors in Computing Systems, 2016 CHI Conference on*. ACM, 2016, pp. 2321–2327.

[34] M. Haak, S. Bos, S. Panic, and L. Rothkrantz, "Detecting stress using eye blinks and brain activity from EEG signals," *Proceeding of the 1st driver car interaction and interface (DCII 2008)*, pp. 35–60, 2009.

[35] S. Kime, F. Galluppi, X. Lagorce, R. Benosman, and J. Lorenceau, "Psychophysical assessment of perceptual performance with varying display frame rates," *Journal of Display Technology*, vol. 12, no. 11, pp. 1372–1382, 2016.

[36] R. Nasiri and Z. Wang, "Perceptual aliasing factors and the impact of frame rate on video quality," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3475–3479.

[37] M. Vollmer and K. Möllmann, "High speed and slow motion: the technology of modern high speed cameras," *Physics Education*, vol. 46, no. 2, p. 191, 2011.

[38] K. Debattista, K. Bugeja, S. Spina, T. Bashford-Rogers, and V. Hulusic, "Frame rate vs resolution: A subjective evaluation of spatiotemporal perceived quality under varying computational budgets," *Computer Graphics Forum*, pp. n/a–n/a, 2017. [Online]. Available: http://dx.doi.org/10.1111/cgf.13302

[39] S. Winkler, "Analysis of public image and video databases for quality assessment," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 616–625, October 2012.

[40] Y. Ou, T. Liu, Z. Zhao, Z. Ma, and Y. Wang, "Modeling the impact of frame rate on perceptual quality of video," *City*, vol. 70, no. 80, p. 90, 2008.

[41] Z. Ma, M. Xu, Y. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 5, pp. 671–682, 2012.

[42] Y. Ou, Y. Xue, and Y. Wang, "Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2473–2486, June 2014.

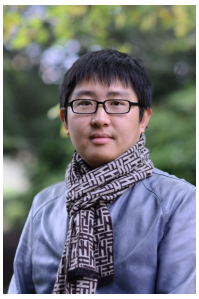
[43] F. Moss, F. Zhang, R. Baddeley, and D. Bull, "What's on TV: A large scale quantitative characterisation of modern broadcast video content," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2425–2429.

[44] E. Kurdoglu, Y. Liu, and Y. Wang, "Perceptual quality maximization for video calls with packet losses by optimizing fec, frame rate and quantization," *IEEE Transactions on Multimedia*, 2017.

[45] J. Wu, C. Yuen, M. Wang, J. Chen, and C. Chen, "Tcp-oriented raptor coding for high-frame-rate video transmission

over wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2231–2246, 2016.
[46] J. Wu, C. Yuen, N. Cheung, J. Chen, and C. Chen, "Modeling and optimization of high frame rate video transmission over wireless networks." *IEEE Trans. Wireless Communications*, vol. 15, no. 4, pp. 2713–2726, 2016.
[47] Y. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 286–298, 2011.
[48] Q. Huang, S. Jeong, S. Yang, D. Zhang, S. Hu, H. Kim, J. Choi, and C. Kuo, "Perceptual quality driven frame-rate selection (PQD-FRS) for high-frame-rate video," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 640–653, 2016.
[49] B. Butterworth, "History of the 'BBC Redux' project," *BBC Internet Blog*, 2013.
[50] F. Moss, K. Wang, F. Zhang, R. Baddeley, and D. Bull, "On the optimal presentation duration for subjective video quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 1977–1987, 2016.
[51] A. Watson, A. Ahumada, and J. Farrell, "Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays," *JOSA A*, vol. 3, no. 3, pp. 300–307, 1986.
[52] S. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," in *Photonics West '98 Electronic Imaging*. International Society for Optics and Photonics, 1998, pp. 180–191.
[53] A. Torralba and A. Oliva, "Statistics of natural image categories," *Network: computation in neural systems*, vol. 14, no. 3, pp. 391–412, 2003.
[54] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
[55] F. Bossen, "Common HM test conditions and software reference configurations (JCTVC-L1100)," *Joint Collaborative Team on Video Coding*, 2013.
[56] M. Kleiner, D. Brainard, D. Pelli, A. Ingling, R. Murray, C. Broussard *et al.*, "Whats new in Psychtoolbox-3," *Perception*, vol. 36, no. 14, p. 1, 2007.
[57] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," 2012.
[58] International Telecom Union. (2004) *Tutorial: Objective perceptual assessment of video quality: Full reference television*.
[59] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2. IEEE, 2003, pp. 1398–1402.
[60] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.
[61] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
[62] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.
[63] D. Chandler and S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.
[64] P. Vu, C. Vu, and D. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 2505–2508.
[65] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on broadcasting*, vol. 50, no. 3, pp. 312–322, 2004.
[66] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," *IEEE transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.
[67] H. Lim and H. Park, "A region-based motion-compensated frame interpolation method using a variance-distortion curve," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 518–524, March 2015.
[68] H. Kaviani and S. Shirani, "Frame rate upconversion using optical flow and patch-based reconstruction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1581–1594, Sept 2016.
[69] Y. Huang, F. Chen, and S. Chien, "Algorithm and architecture design of multirate frame rate up-conversion for ultra-hd lcd systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 12, pp. 2739–2752, Dec 2017.
[70] K. Templin, P. Didyk, K. Myszkowski, and H. Seidel, "Emulating displays with continuously varying frame rates," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, p. 67, 2016.

**Alex Mackin** obtained his MEng in Engineering Mathematics (2013) and PhD in Electrical and Electronic Engineering (2017) from the University of Bristol. He is currently a research associate in the Visual Information Laboratory Department of Electrical and Electronic Engineering, University of Bristol. His research interests include high frame rates, human visual system modelling, adaptive acquisition and immersive video formats.



**Fan Zhang** (M'12) received the B.Sc. and M.Sc. degrees from Shanghai Jiao Tong University (2005 and 2008 respectively), and his Ph.D from the University of Bristol (2012). He is currently working as a Research Associate in the Visual Information Laboratory, Department of Electrical and Electronic Engineering, University of Bristol, on projects related to perceptual video compression. His research interests include perceptual video compression, video quality assessment and immersive video formats including HDR and HFR.

**Professor David R. Bull** (M'94-SM'07-F'12) PhD, FIET, FIEEE, CEng obtained his BSc form the University of Exeter (1980), his MSc from the University of Manchester (1983) and his PhD from the University of Cardiff (1988). He currently holds the Chair in Signal Processing at the University of Bristol and is head of the Visual Information Laboratory. His previous roles include: Lecturer at the University of Wales (Cardiff) and Systems Engineer for Rolls Royce. He was Head of the Electrical and Electronic Engineering Department at Bristol between 2001 and 2006 and is now Director of Bristol Vision Institute (BVI). In 2001 he co-founded ProVision Communication Technologies Ltd. David has worked widely in the fields of 1 and 2-D signal processing. He has won two IEE Premium awards for this work and has published numerous patents, several of which have been exploited commercially. His current activities are focused on the problems of image and video communications and analysis for wireless, Internet and broadcast applications. He has published some 500 academic papers, various articles and 3 books and has also given numerous invited/keynote lectures and tutorials.