Mackin, A., Zhang, A., & Bull, D. (2019). A Frame Rate Conversion Method Based on a Virtual Shutter Angle. In *2019 IEEE International Conference on Image Processing (ICIP)* Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/ICIP.2019.8803489

Peer reviewed version

Link to published version (if available):
10.1109/ICIP.2019.8803489

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at https://ieeexplore.ieee.org/document/8803489. Please refer to any applicable terms of use of the publisher.

# A Frame Rate Conversion Method Based on a Virtual Shutter Angle

Alex Mackin, Fan Zhang and David Bull

**Abstract**

In this paper a new method for frame rate conversion is presented. Utilising motion compensated frame prediction, our method mitigates the spatial distortions associated with traditional methods, while facilitating the conversion between a wider range of frame rates - useful when converting to/between legacy video formats. Using the concept of a virtual shutter angle, our method provides content providers with greater flexibility over the motion characteristic of their video sequences. We also propose a Gaussian weighting scheme which attempts to emulate a video sequence as if it was captured natively at the converted frame rate. A subjective experiment which compared our method to averaging frames at a range of frame rates, has shown that our method results in significantly higher visual quality across a range of content - especially for high down-sample factors.

**Index Terms**

High frame rate, frame rate conversion

## I. Introduction

Increasing video frame rates will lead to greater realism, smoother motion, improved depth perception and higher visual quality [1]–[4] through a reduction in the visibility of temporal aliasing artefacts [5]–[10] and motion blur [11], [12]. However moves towards high frame rate formats (60 fps+) will require robust techniques for frame rate conversion - especially when converting to legacy formats e.g. 24-60 fps [13].

While bespoke temporal down-sampling methods exist [14], conversion to lower frame rates is typically achieved by averaging [15] or dropping frames. The choice of method affects the motion characteristic of the video sequence (e.g. the visibility of motion blur is increased when averaging frames, but at the expense of temporal aliasing artefacts [6], [9]. The opposite is true when dropping frames), and therefore it is important that frame rate conversion methods offer flexibility over the 'look' and 'feel' of the sequence. However while current methods are simple to compute, they restrict the range of frame rates that can be down-sampled to, and introduce unwanted spatial distortions (ghosting artefacts).

In this paper we present a new method for frame rate conversion based on a virtual shutter angle. Our method mitigates ghosting artefacts through motion compensated frame prediction, which also provides greater flexibility in terms of the frame rates and effective shutter angles we can convert to (including non-integer factors). The results from a subjective experiment confirm that the proposed method leads to significant increases in visual quality and viewer preference.

The rest of this paper is organised as follows: Section II presents a framework for temporal down-sampling, Section III details our proposed method, while Section IV presents the methodology and results of a subjective experiment.

## II. Frame Rate Down-Sampling Framework

The down-sample factor $K$ is the ratio between the current frame rate $f_h$ and the desired lower frame rate $f_d$:

$$K = f_h/f_d \tag{1}$$

When converting to a lower frame rate we partition the video sequence into $K$-frame sub-sequences as shown in Fig. 1.

The authors acknowledge EPSRC for funding (grant #: EP/M000885/1). The authors are with the Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK. {a.mackin, fan.zhang, dave.bull}@bristol.ac.uk

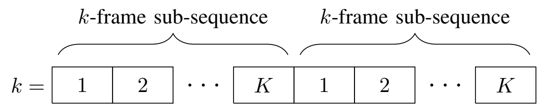


Fig. 1: How to partition a video sequence into $K$-frame sub-sequences.

After conversion to linear light space (gamma decoding), each down-sampled frame $\mathbf{F}_d$ is computed as a weighted sum of its respective high frame rate sub-sequence:

$$\begin{aligned} \mathbf{F}_d &= g\left(\sum_{k=1}^{K} \alpha[k] \cdot g^{-1}\left(\mathbf{F}_s[k]\right)\right) \\ \text{s.t.} \quad & \sum_{k=1}^{K} \alpha[k] = 1 \ \wedge \ \forall k \in [1, K] : \ \alpha[k] \geq 0 \end{aligned} \tag{2}$$

where $n \in \mathbb{Z}^+$, $\alpha[k]$ is the normalised weight of frame $k$, $g(\cdot)$ is the non-linear transfer function [13], [16], and $\mathbf{F}_s[k]$ is a frame in the sub-sequence (a $w \times h$ matrix of pixel values)[1].

The non-linear transfer function is typically modelled using a power-law expression (gamma correction [17]): $g(x) = x^{\gamma}$ and $g^{-1}(x) = x^{1/\gamma}$. If the non-linear transfer function is unknown, then a suitable value needs to be used i.e. $\gamma = 2$.

The most commonly used weighting schemes are referred to as dropping frames ($\alpha[1] = 1$) and averaging frames ($\alpha[k] = 1/K$) [15]. The weights in Eq. 2 can though be selected to elicit a desired frequency response if required.

### A. Shutter Angle

The length of each exposure relative to the frame time is often referred to as the *shutter angle*, which ranges from 0° to 360°. The process of temporal down-sampling may affect the shutter angle, for example if there are leading/trailing weights equal to 0, then the shutter angle $s_h$ is effectively reduced by the factor of $K/(K-L)$ ($L$ is the number of leading/trailing weights equal to 0). The shutter angle is reduced by factor $K$ when dropping frames (as there $K-1$ trailing zero weights).

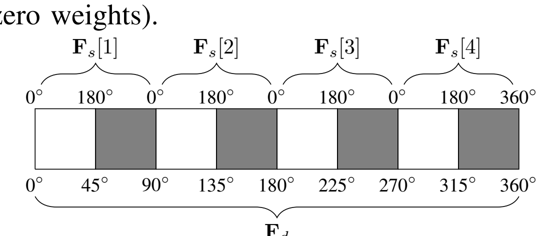


Fig. 2: In this figure the white portion of each frame is a representation of when the camera shutter is exposing/open (grey for when closed). The shutter angles shown above and below the frames are reference to the original sub-sequence $\mathbf{F}_s$ and the down-sampled frame $\mathbf{F}_d$ respectively. In this example a sequence ($\mathbf{F}_s$) with an 180° shutter angle is down-sampled by $K$=4.

### B. Limitations

This framework can lead to visible artefacts when a fully open shutter (360°) is not used to capture the content[2]. As when content is not captured with a fully open shutter, any valid weighting scheme in Eq. 2 which has more than one non-zero weight will result in a distinctive ghosting artefact for objects in motion (see Fig. 3). This artefact becomes apparent because of the discrete temporal gaps caused by the camera shutter being closed between frame periods in the original sequence. For example in Fig. 2, averaging all the frames in a sub-sequence ($\mathbf{F}_s$) would result in temporal gaps being present in the down-sampled frame $\mathbf{F}_d$. For objects in motion, these temporal gaps become spatial gaps - due to the object moving between camera exposures. The visibility of ghosting artefacts will be exacerbated by smaller capture shutter angles, faster objects and/or large down-sample factors.

The move towards higher frame rates will result in this ghosting artefact becoming more profound when converting to legacy formats (as values for $K$ will increase). Alongside reducing distortions, it is paramount that content providers have the flexibility to select from a range of shutter angles when converting between frame rates, as it controls the trade-off between motion blur and temporal aliasing artefacts [6], [9].

[1] $\mathbb{Z}^+$ is the notation used for the set of positive integers $(1, 2, 3, \ldots)$.

[2] Even at high frame rates, capturing with a 360° shutter is not practically possible, as the sensor data must be written to storage after each exposure.

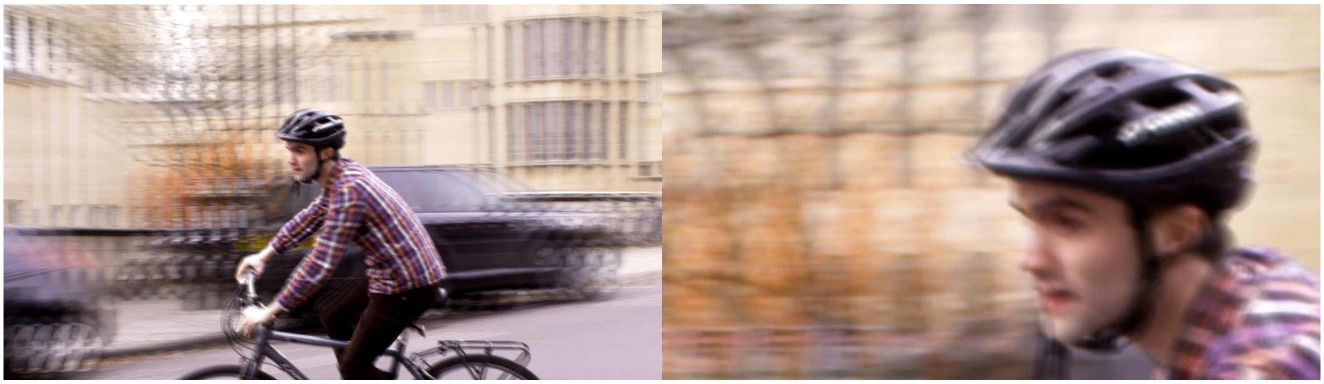

Fig. 3: The ghosting artefact that becomes apparent when a sequence [2] with an 180° shutter angle is converted from 60 fps to 15 fps by averaging frames. The right frame shows a section of the original frame after 4× magnification.

## III. Proposed Method

In this paper we propose a frame rate conversion method that simulates a virtual shutter angle using motion compensated frame interpolation (MCFI), in an attempt to provide a smoother, continuous motion characteristic (as opposed to the discrete ghosting type artefact in Fig. 3). Whereas the framework outlined in Section II only allows for integer down-sampling and shutter angles restricted to the range:

$$s_h/K \leq s_v \leq s_h \tag{3}$$

where $s_h$ is the shutter angle of the original video sequence. The method proposed in the paper enables a wider range of frame rates and shutter angles to be selected, as the video content is up-sampled prior to down-sampling.

Given enough intermediate frames, a video sequence with any non-prime integer frame rate (lower or higher than the original frame rate), and shutter angles in the range:

$$s_h/K \leq s_v \leq 360^\circ \tag{4}$$

could be generated using this method. The number of frames (original + predicted) per $K$-frame sub-sequence required for a desired frame rate $f_d$ and shutter angle $s_v$ is:

$$N = \frac{f_h}{f_d}\frac{s_v}{s_h} = K\frac{s_v}{s_h} \tag{5}$$

In the case when $N$ is an integer, we can interlock the original frames ($\mathbf{F}_s$) with interpolated intermediate frames ($\hat{\mathbf{F}}_s$). However if $N$ is non-integer, then an approximation of a video with the virtual shutter angle $s_v$ is made.

The intermediate frames $\hat{\mathbf{F}}_s$ can be predicted using any suitable frame rate up-conversion algorithm, ranging from motion-compensation (block matching [18] or optical flow [19]) to state-of-the-art CNN based methods [20]. We use block based motion compensation [18] due to the interoperability with existing video compression architectures.

One of the major benefits of this method is that once the up-sampled sequence has been created, different combinations of viable frame rates and shutter angles can easily be generated, as only Eq. 2 needs to be computed.

### A. *Interlocking Method* ($N \in \mathbb{Z}^+$)

Fig. 4 demonstrates how a video sequence with an 180° shutter angle can be used to generate a video with a virtual shutter angle of 360° by the process of interlocking frames. In this example $\hat{\mathbf{F}}_s[k]$ is predicted from the frames $\mathbf{F}_s[k]$ and $\mathbf{F}_s[k+1]$ (shown by the arrows). Then by combining all the original and predicted intermediate frames within a sub-sequence (up to $\hat{\mathbf{F}}_s[4]$), we ensure that each frame ($\mathbf{F}_d$) in the down-sampled sequence has an temporal extent equivalent to a 360° shutter angle i.e. no temporal gaps (grey regions).

To generalise this process, the locations of frames in degrees (with respect to the down-sampled sequence) for each $K$-frame sub-sequence can be calculated as follows:

$$\mathrm{L}(s, K) = \left\{ x\frac{s}{K} \;\middle|\; x \in \mathbb{Z} : 0 \leq x \leq K\frac{s_v}{s} - 1 \right\} \tag{6}$$

The original $\mathbf{F}_s$ and predicted $\hat{\mathbf{F}}_s$ frames are located at $\mathrm{L}(360, K)$ and $\mathrm{L}(s_h, K) \setminus \mathrm{L}(360, K)$ respectively.

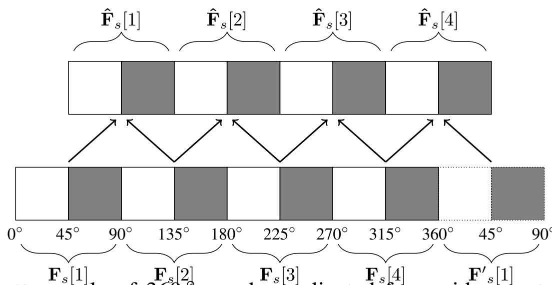


Fig. 4: An example of how a virtual shutter angle of 360° can be replicated from video content that was captured with an 180° shutter angle. The shutter angles shown are with respect to the output frame after down-sampling ($\mathbf{F}_d$).

In Fig. 4 the intermediate frame $\hat{\mathbf{F}}_s[4]$ is predicted from a frame in the next $K$-frame sub-sequence ($\mathbf{F'}_s[1]$). Therefore when this frame is not available e.g. at the end of the video sequence, a forward ($\mathbf{F}_s[4]$ as the only reference) as opposed to bi-directional frame interpolation can be used.

### *B. Hierarchical prediction* $(N \notin \mathbb{Z}^+)$

When interlocking frames is not permissible ($N$ is not an integer), a desired virtual shutter angle can be approximated. As a consequence the locations and number of frame is now a free parameter, with the only constraint being that frame locations ($y$) must lie within the following range:

$$0 \leq y \leq s_v - s_h/K \tag{7}$$

While uniform spacing may be a suitable solution, it can result in the original frames being omitted from the weighted sum in Eq. 2. Even though frame interpolation should be fairly accurate at high frame rates (due to the increased temporal correlation between frames), there may be perceptible distortions from motion compensation e.g. blocking artefacts.

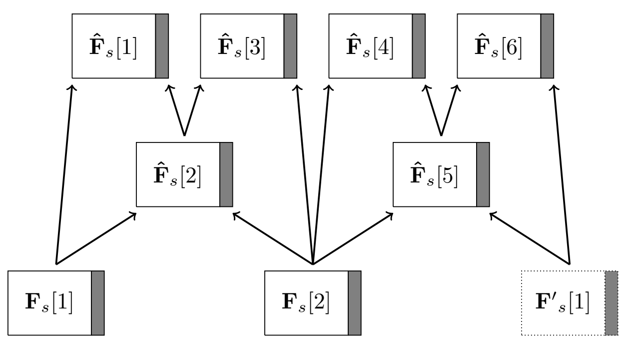


Fig. 5: An example of a hierarchical prediction structure. In this case two layers are used to down-sample a sequence by a factor ($K$) of 2.

We propose a hierarchical prediction structure (see Fig. 5), in which layers can be added to increase accuracy, or removed to reduce complexity. Any frames outside the range in Eq. 7 are zero weighted, and need not be predicted. In this structure, previously predicted frames are used as reference for the layer above, reducing the temporal distance when predicting frames. This structure also reduces computational complexity if motion compensated frame interpolation is used, as the search range $d$ can be reduced by a factor of 2 every layer.

In order to generate a more accurate approximation of shutter angle $s_v$, an additional frame can be inserted into the structure at the location $(s_v - s_h/k)^\circ$. The weights in Eq. 2 may have to altered slightly to reduce bias at this location.

### *C. Gaussian Weighting Scheme*

After the intermediate frames $\mathbf{F'}_s$ have been predicted using either method, they are combined with the original frames $\mathbf{F}_s$ in chronological order. The output video sequence with frame rate $f_d$ and shutter angle $s_v$ is then generated using Eq. 2.

Motion blur typically exhibits a Gaussian profile [6], therefore we propose a Gaussian weighting scheme:

$$\alpha[k] = \frac{\exp\left[-\left(\frac{k-(N+1)/2}{2\sigma}\right)^2\right]}{\sum_{i=1}^{N}\alpha[i]}, \quad k = 1, \ldots, N \tag{8}$$

where $N$ is the number of frames per sub-sequence (original + intermediate frames) and $\sigma$ is the standard deviation of the Gaussian distribution. $\sigma$ is used to trade-off the relationship between motion blur and temporal aliasing, and while any value can be used (within reason), we found that full width at half maximum (FWHM) provided good results ($2.355N$).

Fig. 6 shows the same frame from Fig. 3, but converted to 15 fps and shutter angles of (left) 180° and (right) 360° using the interlocking method and Gaussian weighting scheme. The ghosting artefacts associated with averaging frames has been mitigated, and a smooth motion characteristic replicated.

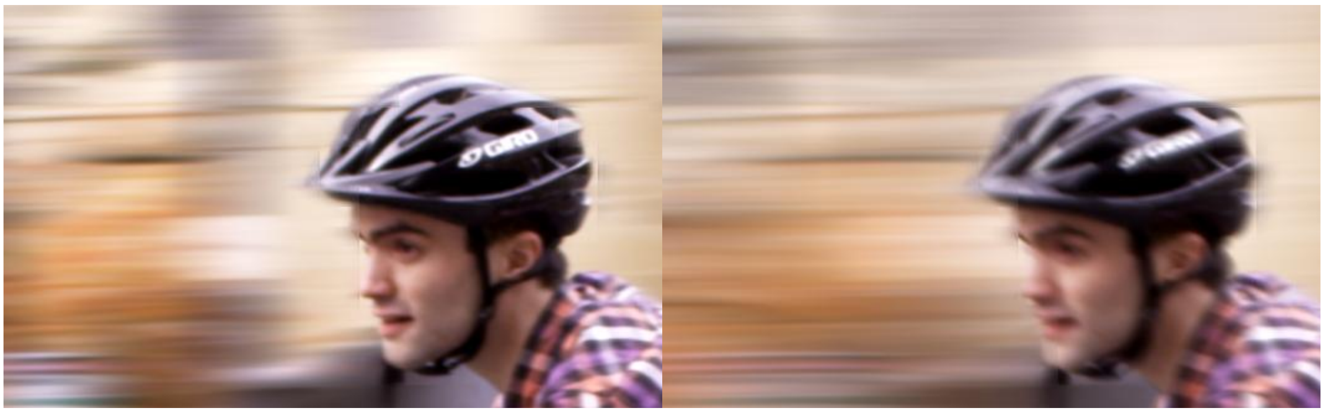

Fig. 6: A video sequence with a frame rate and shutter angle of 60 fps and 180° respectively converted to 15 fps and a shutter angles of (left) 180° and (right) 360° using the interlocking method and Gaussian weighting scheme.

### D. Unknown Shutter Angle

In the case when the shutter angle of the original video is unknown, or a good estimate cannot be made, then the hierarchical prediction structure outlined in Section III-B should be used. The value of $\sigma$ can be selected to give the desired look.

### E. Limitations

The major limitations of the method proposed in this paper are as follows: there is increased computational complexity, as the intermediate frames have to be predicted using motion interpolation; and dependent on the motion model used, the motion compensated frames may contain additional artefacts - although the magnitude of these distortions should be limited when using original frames in the weighted sum (Eq. 2).

## IV. Subjective Testing

Due to the lack of a reference, the performance of the proposed frame rate conversion method cannot be scrutinised using traditional objective metrics such as PSNR. This is because it would be nearly impossible to capture exactly the same scene at the converted frame rate and shutter angle. Therefore using a double stimulus methodology [21], we designed a subjective experiment to characterise differences in visual quality between the proposed and traditional methods.

### A. Methodology

Six sequences from the Ultra Video Group video database [3] were used in the experiment (see Fig. 7). The source sequences have a 3840×2160 resolution, are 10 bits per colour channel, were captured at 60 fps with a 180° shutter angle, and are 5 seconds in length (cut where necessary). These sequences were converted to 60, 30 and 15 fps using two methods: (i) by averaging frames and (ii) the interlocking method with the Gaussian weighting scheme (Section III-C). An 180° virtual shutter angle was used to ensure that the temporal extent was equivalent between the two methods.

[3] http://ultravideo.cs.tut.fi/

Fig. 7: Sample frames from the Ultra Video Group video database.

18 participants with an average age of 29 took part in the experiment. All of them had normal or corrected-to-normal vision acuity. After a brief training session, participants viewed two versions of the same sequence - one generated using averaging frames, the other using the proposed method. These two versions were presented in a random order, and after viewing both, participants scored their perceived quality for both on separate continuous quality scales (0 to 100) [21].

A Panasonic BT-4LH310 LCD reference monitor with a peak luminance of 210 cd/m$^2$, a contrast ratio of 400:1, and a 3840×2160 spatial resolution (measuring 65.4×36.8 cm) was used. The viewing distance was chosen as 1.5 H [22]. The viewing environment conformed to BT.500-13 [21].

### B. Results

Mean Opinion Scores (MOS) were calculated for each test condition [21]. The results of the experiment are shown in Fig. 8 (left). While there is little perceptual difference at 60 fps, as the down-sample factor $K$ increases, our proposed method results in greater visual quality (higher MOS). Fig. 8 (right) shows the relationship between the percentage of participants who preferred the proposed method to averaging frames with frame rate for each test condition, and indicates that the higher the down-sample factor, the greater the preference for our method over averaging frames - with over 80% of participants preferring the proposed method at 15 fps.

A one-tailed paired t-test on the raw opinion scores demonstrates that our method is statistically superior in visual quality at a 0.025 significance level: $t(323) = 6.7$, $p \approx 0$.

## V. Conclusions

In this paper we have proposed a frame rate conversion method based on a virtual shutter angle. Unlike traditional down-sampling frame works, our method can reproduce content over a range of frame rates and shutter angles, while also reducing perceptible ghosting artefacts. By using a Gaussian weighting scheme, we attempt to emulate the blur profile associated with capturing the content with a video camera.

Through a subjective test we demonstrate that the proposed method frame rate conversion method offers (significantly) higher visual quality compared to traditional methods.

## References


[1] 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.

[2] 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.

[3] 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, pp. 15, 2015.

[4] 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. 60402–1, 2016.

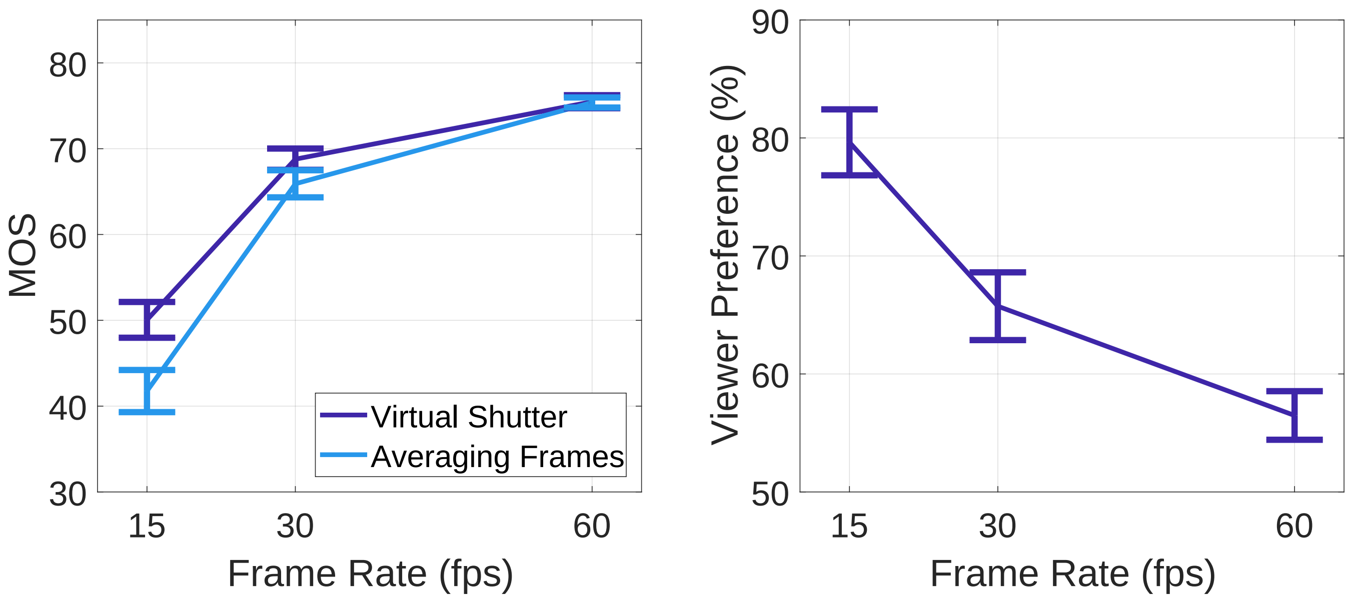


Fig. 8: The relationship between (left) MOS / (right) viewer preference and frame rate. Error bars represent standard error of the mean.


[5] 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.

[6] 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.

[7] S. Daly, N. Xu, J. Crenshaw, and V. Zunjarrao, "A psychophysical study exploring judder using fundamental signals and complex imagery," in *SMPTE Conferences*. Society of Motion Picture and Television Engineers, 2014, vol. 2014, pp. 1–14.

[8] 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.

[9] 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.

[10] A. V. Katsenou, A. Mackin, D. Ma, F. Zhang, and D. R. Bull, "Exploring the challenges of higher frame rates: from quality assessment to frame rate selection," *IEEE COMSOC MMTC Communications - Frontiers (E-Letter)*, 2018.

[11] 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.

[12] 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.

[13] ITU-R Recommendation BT.709-6, "Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange," 2015.

[14] 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, pp. 67, 2016.

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

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

[17] C. Poynton, *Digital video and HD: Algorithms and Interfaces*, Elsevier, 2012.

[18] D. Wang, A. Vincent, P. Blanchfield, and R. Klepko, "Motion-compensated frame rate up-conversion – Part II: New algorithms for frame interpolation," *IEEE Transactions on Broadcasting*, vol. 56, no. 2, pp. 142–149, 2010.

[19] G. Lu, X. Zhang, L. Chen, and Z. Gao, "A novel frame rate up conversion using iterative non-local means interpolation," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1–5.

[20] J. van Amersfoort, W. Shi, A. Acosta, F. Massa, Jo. Totz, Z. Wang, and J. Caballero, "Frame interpolation with multi-scale deep loss functions and generative adversarial networks," *arXiv preprint arXiv:1711.06045*, 2017.

[21] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," 2012.

[22] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," 2008.