SRQM: A Video Quality Metric for Spatial Resolution Adaptation

Alex Mackin, Mariana Afonso, Fan Zhang, and David Bull
Department of Electrical and Electronic Engineering
University of Bristol
Bristol, BS8 1UB, United Kingdom
{A.Mackin, Mariana.Afonso, Fan.Zhang, Dave.Bull}@bristol.ac.uk

Abstract

This paper presents a full reference objective video quality metric (SRQM) which characterises the relationship between variations in spatial resolution and visual quality in the context of adaptive video formats. SRQM uses wavelet decomposition, subband combination with perceptually inspired weights, and spatial pooling, to estimate the relative quality between the frames of a high resolution reference video, and one that has been spatially adapted through a combination of down and upsampling. The BVI-SR video database is used to benchmark SRQM against five commonly-used quality metrics. The database contains 24 diverse video sequences that span a range of spatial resolutions up to UHD-1 (3840×2160). An in-depth analysis demonstrates that SRQM is statistically superior to the other quality metrics for all tested adaptation filters, and all with relatively low computational complexity.

I. Introduction

The desire for high quality video experiences has led to video formats with increased spatial resolutions, higher frame rates greater dynamic ranges and wider colour gamuts [1]. However the data rates associated with these formats has meant that many of the perceptual benefits [2]–[7] are inaccessible to a large number of consumers. For example only 12% of internet users globally [8] have the recommended bandwidth of 25 Mb/s to stream UHD-1 (3840×2160) content on Netflix [9]. Moves towards 8K (7680×4320) and high frame rate (60 fps+) formats will exacerbate this issue [6]. There is then an urgent need to exploit the perceptual properties of the human visual system (HVS), beyond what is currently achieved with current video compression standards.

It is generally acknowledged that the spatio-temporal bandwidth of the HVS is limited [10], to maximum perceptible spatial and temporal frequencies of approximately 32 cpd and 60 Hz respectively for standard dynamic range displays. These limits can increase up to 55 cpd and 90 Hz for high dynamic range (HDR) displays (where peak luminance is greater than 1000 cd/m^2). This, coupled with the fact that visual sensitivity is not uniform across the range of perceptible frequencies [11], means that perceptually redundant information is being transmitted to consumers - especially as video formats begin to reach the limits of human perception [12].

Alongside incorporating perceptually inspired video coding methods [13]–[16] into video compression standards, the issue of perceptual redundancy can be addressed through adaptively select video parameters in a content-dependent manner. By reducing/eliminating redundant information within the video pipeline, we can reduce data rates and/or provide consumers with higher quality experiences than is currently realised.

Frameworks for spatial resolution adaptation have previously been proposed [17]–[20], and they typically involve videos being downsampled prior to transmission, before being upsampled to the original resolution at the receiver or the display. While downsampling usually occurs within the encoder, it could also take place during acquisition or post-processing. A key component of any such adaptive format, will be robust low-complexity quality metrics. These will be used to select the optimal resolutions given the prevailing channel conditions (among other constraints). Therefore it is important that they can successfully model visual quality across a range of resolutions and content, as salient perceptual information should not be lost during adaptation.

In this paper we present the Spatial Resolution Quality Metric (SRQM). Inspired by the Frame Rate dependent video quality metric (FRQM) [21], SRQM utilises wavelet decomposition and spatial pooling to model the reduction in visual quality associated with spatial adaptation. Our analysis demonstrates that SRQM provides best-in-class performance when characterising the subjective evaluations from the BVI-SR video database [22], and is shown to be statistically superior to all tested quality metrics when using three distinct adaptation filters - all with relatively low computational complexity.

The remainder of the paper is organised as follows: Section II outlines SRQM, while Section III benchmarks its performance. Conclusions are then provided in Section IV.

II. Proposed Algorithm

The SRQM video quality metric estimates the relative quality between a spatially adapted (downsampled and then upsampled) video and its original version. While the same pre-filter (e.g. bicubic) does not necessarily need to be used for both down and upsampling, it does make the associated spatial distortions easier to characterise (as they are consistent).

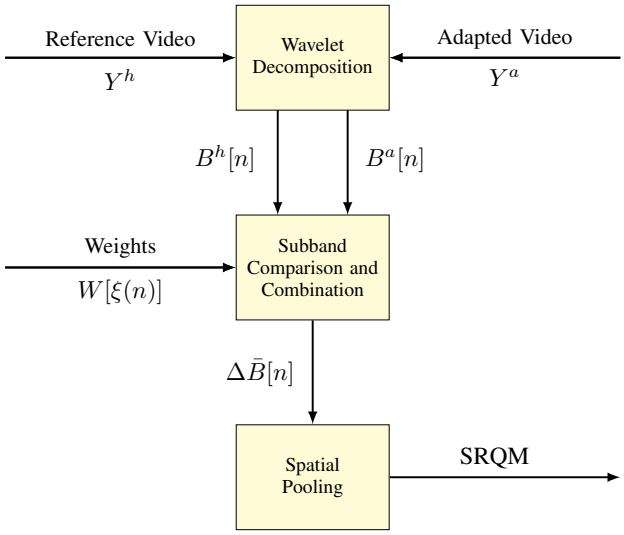

Fig. 1: An illustration of the proposed SRQM video quality metric

The architecture of the proposed video quality metric SRQM is presented in Fig. 1. The three steps required to compute the SRQM quality index are: (i) wavelet decomposition, (ii) subband comparison and combination, and (iii) spatial pooling. These steps are described in further detail below.

To ensure that SRQM is applicable across a range of bitdepths, the luma pixels in the reference (Y^h) and adapted (Y^a) video frames are scaled to the range 0-1. Given a bitdepth b, this can be achieved by dividing every pixel by 2^b-1.

A. Wavelet Decomposition

Spatial adaptation may introduce aliasing artefacts such as 'jaggies' or 'moiré' patterns into the video signal. A pre-filter is typically used to reduce the impact of these artefacts, but at the expense of blurring artefacts (due to the attenuation of high spatial frequencies). Here we attempt to estimate the magnitude of these distortions by comparing the high frequency components of the original and adapted video sequence at multiple scales (levels), and then by using perceptually inspired weights, characterise their impact on visual quality.

A 2-D Haar Discrete Wavelet Transform [12] (DWT) is applied to both the original and adapted video frames at N distinct levels (luma channel only). At each level n, the transform will decompose the video signal into one low frequency DC subband (LL), and horizontal (LH), vertical (HL) and diagonal (HH) high frequency subbands (see Fig. 2). The number of required levels (N) is calculated as follows:

$$N = \lceil \log_2 d \rceil \tag{1}$$

where d is the spatial downsampling factor, which is calculated by dividing the vertical/horizontal resolution of the reference by the corresponding resolution after spatial downsampling.

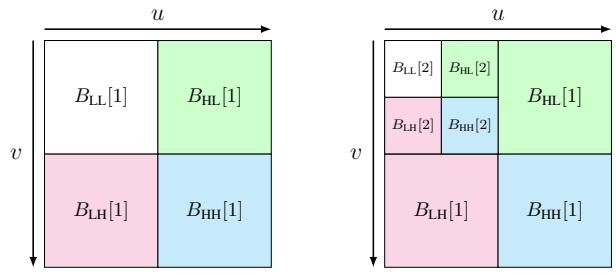

Fig. 2: A visual example of how the 2-D Haar DWT decomposes the video signal into subbands at (left) 1 and (right) 2 levels.

We define the first level (n=1) of the DWT transform as:

$$\langle B_{\mathrm{LL}}[1], B_{\mathrm{HL}}[1], B_{\mathrm{LH}}[1], B_{\mathrm{HH}}[1] \rangle = \mathrm{DWT}(Y) \tag{2}$$

Subsequent levels can then be generated by computing the DWT for the low frequency subband of the previous level:

$$\langle B_{\mathrm{LL}}[n], B_{\mathrm{HL}}[n], B_{\mathrm{LH}}[n], B_{\mathrm{HH}}[n] \rangle = \mathrm{DWT}(B_{\mathrm{LL}}[n-1]) \tag{3}$$

The subband coefficients (B) are scaled by a factor of 2 after each transformation (due to the use of the Haar wavelet).

B. Subband comparison and combination

The high frequency subbands at each level are upsampled by a factor of 2^n using nearest neighbour interpolation to enable pixel-level comparisons. A subband difference measure can then calculated for each level as:

$$\begin{aligned}\Delta B[n] = \frac{1}{3}\Big\{ &\left|B^h_{\text{HL}}[n] - B^a_{\text{HL}}[n]\right| + \dots \\ &\left|B^h_{\text{LH}}[n] - B^a_{\text{LH}}[n]\right| + \left|B^h_{\text{HH}}[n] - B^a_{\text{HH}}[n]\right| \Big\}\end{aligned} \tag{4}$$

An overall difference is then computed as the weighted sum of the subband differences over all N levels:

$$\Delta\bar{B} = \sum_{n=1}^{N} W[\xi(n)] \cdot \Delta B[n] \tag{5}$$

where $W[\xi(n)]$ is the weight for spatial frequency $\xi(n)$ at level n. The weighting values for SRQM are selected in Section III-A. The horizontal (u_s) and vertical (v_s) Nyquist frequencies of the display will be equivalent with 'square' pixels i.e. the spacings between pixels are equal in orthogonal directions. In this case the spatial frequency of a wavelet subband can be approximated at each level as [23]:

$$\xi(n) = \frac{u_s}{2^n} = \frac{v_s}{2^n} \tag{6}$$

Based on the Nyquist criterion, u_s is calculated by dividing the vertical resolution of the display (R_v) by twice its height (H cm):

$$u_s = \frac{R_v}{2H} \tag{7}$$

C. Spatial Pooling

The distribution of spatial distortions in a frame will generally be non-uniform - they will be concentrated around high-frequency components such as edges. Therefore rather than computing a single central tendency measure (e.g. mean) for the frame, we instead calculate multiple local measures. Here we separate each frame into K non-overlapping blocks MB_k, and compute the mean of each block as follows:

$$\mu[k] = \sum_{\forall(x,y)\in \text{MB}_k} \frac{\Delta\bar{B}(x,y)}{M^2}, \quad k = 1, 2, \dots, K \tag{8}$$

where M is the block size. A larger local window than SSIM [24] (32×32) is used to account for UHD-1 content.

We then compute a frame level quality index as the most noticeable distortion across all blocks in the frame:

$$Q_i = \max\left\{\mu[1], \mu[2], \dots, \mu[K]\right\} \tag{9}$$

Video quality is then computed using the following equation:

$$\bar{Q} = \frac{1}{L}\sum_{t=1}^{L} Q_i[t] \tag{10}$$

where L is the number of frames. $\bar{Q}$ is then converted to decibels units to form the SRQM video quality index:

$$\text{SRQM} = 20 \cdot \log_{10}\left(\bar{Q}^{-1}\right) \tag{11}$$

III. Results

The BVI-SR video database [22] is used to assess the performance of SRQM,and was chosen because it is one of very few databases that contains uncompressed video sequences that have been adapted spatially through a combination of down and upsampling. BVI-SR contains 24 source sequences with a native resolution of 3840×2160, a frame rate of 60 fps and a bit-depth of 10 bits. The source sequences were downsampled to three resolutions: HD, 540p and 270p, using 2 adaptation filters: nearest neighbour and bicubic. They were then upsampled to UHD-1 using the same filter. A CNN-based super-resolution method VDSR [25] was also used for upsampling, but only on those sequences that were downsampled using the bicubic filter (the filter used to train the CNN).

Ground truth in the form of Mean Opinion Scores (MOS) is provided for all 240 test conditions [22]. Differential Mean Opinion Scores (DMOS) [26] are then calculated between the UHD-1 reference and the lower resolution videos as follows:

$$\text{DMOS} = \text{MOS}^h - \text{MOS}^a \tag{12}$$

where MOS^h and MOS^a are the mean opinion scores for the reference (UHD-1) and adapted videos respectively.

The performance of SRQM is compared against five quality metrics: PSNR [27], SSIM [24], VSNR [28], MS-SSIM [29] and VIF [30]. Quality metrics used for spatial adaptation should be computationally efficient - especially for UHD-1, 60 fps content. As a result we have intentionally ignored quality metrics that have high computational complexity e.g. MOVIE [31], PVM [14], ST-MAD [32] and VQM [33].

In order to reduce non-linearities, a logistic fitting curve is used to fit the quality metrics predictions to DMOS [34]. Four evaluation metrics [34] are then used to evaluate the accuracy (LCC and RMSE), monotonicity (SROCC) and consistency (OR) of the quality metrics. An F-test [35] is used to assess whether there are statistically significant differences between quality metric predictions. A platform independent complexity score for each quality metric is then calculated by dividing its average execution time by the execution time of PSNR[1].

A. *Optimal weights*

The optimal weights for SRQM were calculated using the ground truth from BVI-SR. In order to reduce overfitting, the optimal weights were obtained using a two-fold cross validation approach. This involved the subjective data being randomly partitioned into halves; the data from one half is used to find the best weights in (5) (training set), while the other half is used to benchmark performance (test set). The best weights were those which maximised the correlation between SRQM and the subjective scores after an exhaustive search. SROCC was used to assess correlation performance as it does not rely on fitting the data with the logistic curve (which can sometimes be inaccurate). In order to minimise content bias this process was repeated 100 times, with both halves being used for the test and training sets (200 partitions in total).

Table I reports the average and standard deviation of SROCC values over the 200 test sets, comparing SRQM to the five other quality metrics. SRQM and VIF are the best and most consistent performers, highlighted by the highest means and lowest standard deviations of the reported SROCC values.

TABLE I: The mean (μ) and standard deviation (σ) of SROCC values on the 200 test sets (generated using two-fold cross validation).

Metric	PSNR	SSIM	VSNR	MS-SSIM	VIF	SRQM
μ	0.823	0.8170	0.852	0.872	0.911	0.909
σ	0.026	0.032	0.032	0.021	0.018	0.013

Fig. 3 (left) demonstrates the sensitivity of SROCC values to the weights in (5) for a single partition during two-fold cross validation ($W[\xi(1)]$ is fixed). The best weights for this training set are marked with a red dot. Above a certain point SROCC is robust to changes in the weights, indicating that SRQM should offer good performance as long as a fairly representative training set is used to calculate the weights.

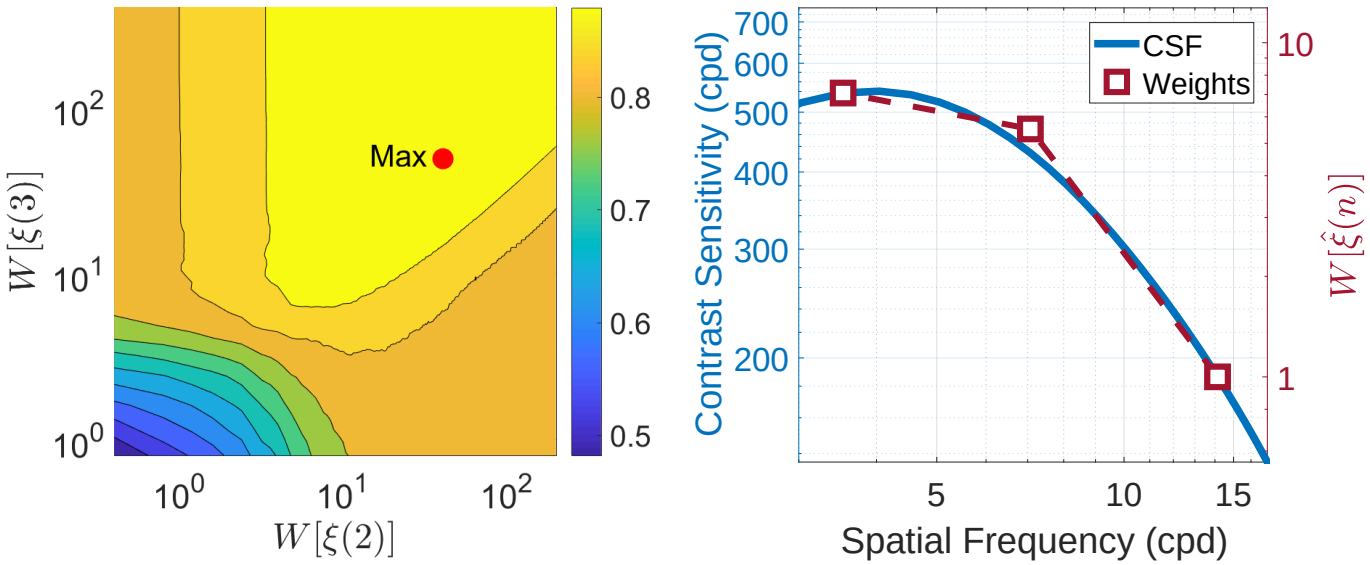

Fig. 3: (left) The sensitivity of SROCC to the weights and (right) a comparison between the contrast sensitivity function (CSF) and the optimal weights.

[1]Complexity scores obtained from a Windows 7 PC with an Intel Core i7-6800K CPU @ 3.4 GHz (on a single thread) running MATLAB 2017a.

The optimal weights for SRQM (calculated as the average of the best weights from the 200 training sets) are: 1, 5.5, and 7.1 for levels (n) 1, 2 and 3 respectively. Instead of relating these weights to the spatial frequency of the display ($\xi(n)$), we can instead calculate the spatial frequency relative to the viewer (in cycles per degree subtended) as follows:

$$\hat{\xi}(n) = \xi(n) \cdot \left[2 \tan^{-1}\left(\frac{1}{2V_d}\right)\right]^{-1} \quad \text{(cpd)} \tag{13}$$

where V_d is the viewing distance in cm. For the BVI-SR experiment: R_v = 2160, H = 36.8 cm and V_d = 1.5 H (55.2 cm).

The contrast sensitivity function (CSF) [11] models the human visual systems acuity across the range of perceptible spatial frequencies. Fig. 3 (right) compares the optimal weights to the CSF model from [36]. The clear similarity between the curves demonstrates that the optimal weights are consistent with the CSF, thus indicating a perceptual basis to them.

B. Performance

Table II reports the performance of the quality metrics on the BVI-SR video database. The evaluation metrics indicate that VIF and SRQM outperform the other tested quality metrics and that, while VIF is slightly better in terms of SROCC, SRQM is superior in terms of LCC and RMSE. The relative complexity of SRQM is related to the number of decomposition levels N. The average relative complexity of SRQM is 15, 23 and 27 when $N = 1$, 2 and 3 respectively.

TABLE II: The overall statistical performance and complexity of all the tested quality metrics. The best performing quality metric in each row is **bold**.

Metric	PSNR	SSIM	VSNR	MS-SSIM	VIF	SRQM
SROCC	0.820	0.814	0.870	0.848	**0.910**	0.905
LCC	0.841	0.831	0.887	0.858	0.928	**0.934**
OR	0.537	0.546	0.495	0.486	**0.366**	**0.366**
RMSE	13.68	14.22	11.72	12.98	9.38	**8.99**
Complexity	**1**	13	27	28	329	22

Table III reports the results from an F-test [35] which compares the predictions of the tested quality metrics. Even though SRQM is over 15 times faster than VIF, an F-test reports no statistical significance between the two. SRQM and VIF are statistically superior to the other quality metrics.

TABLE III: F-test results for 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.

Metric	PSNR	SSIM	VSNR	MS-SSIM	VIF	SRQM
PSNR	-	0	0	-1	-1	-1
SSIM	0	-	0	-1	-1	-1
VSNR	0	0	-	0	-1	-1
MS-SSIM	1	1	0	-	-1	-1
VIF	1	1	1	1	-	0
SRQM	1	1	1	1	0	-

Fig. 4 shows the relationship between the quality metric predictions and DMOS. While the predictions of all the quality metrics are fairly compact around their respective fitting curves, it is clear that the choice of adaptation filter affects the predictions of the quality metrics.

Table IV reports the performance of the quality metrics for each of the tested adaptation filters individually. The performance of SRQM is better in every evaluation metric for the nearest neighbour and bicubic filters, whereas VIF offers better performance with regard to SROCC and OR for the VDSR filter. An F-test affirms that SRQM is statistically superior to all the tested quality metrics for every adaptation filter[2] (Table V), and is therefore the most suitable for spatial adaptation, in terms of both accuracy and complexity.

IV. Conclusions

In this paper we have presented a new video quality metric SRQM for the purpose of spatial resolution adaptation (where a video sequence is downsampled before transmission, and subsequently upsampled before being displayed). The proposed method employ wavelet decomposition, subband combination and spatial pooling to estimate the perceptual impact of spatial distortions that arise during spatial adaptation. When tested on the BVI-SR video database, SRQM offers excellent performance compared to five other quality metrics, and is shown to statistically superior for each of the adaptation filters tested. The low relative complexity of SRQM indicates that it could used for spatial resolution adaptation across a range of content types, resolutions and adaptation filters.

[2]The quality scores associated with the individual adaptation filters will form a smaller distribution than if all the data is used (as is the case for Table III) - making significant differences more likely with an F-test.

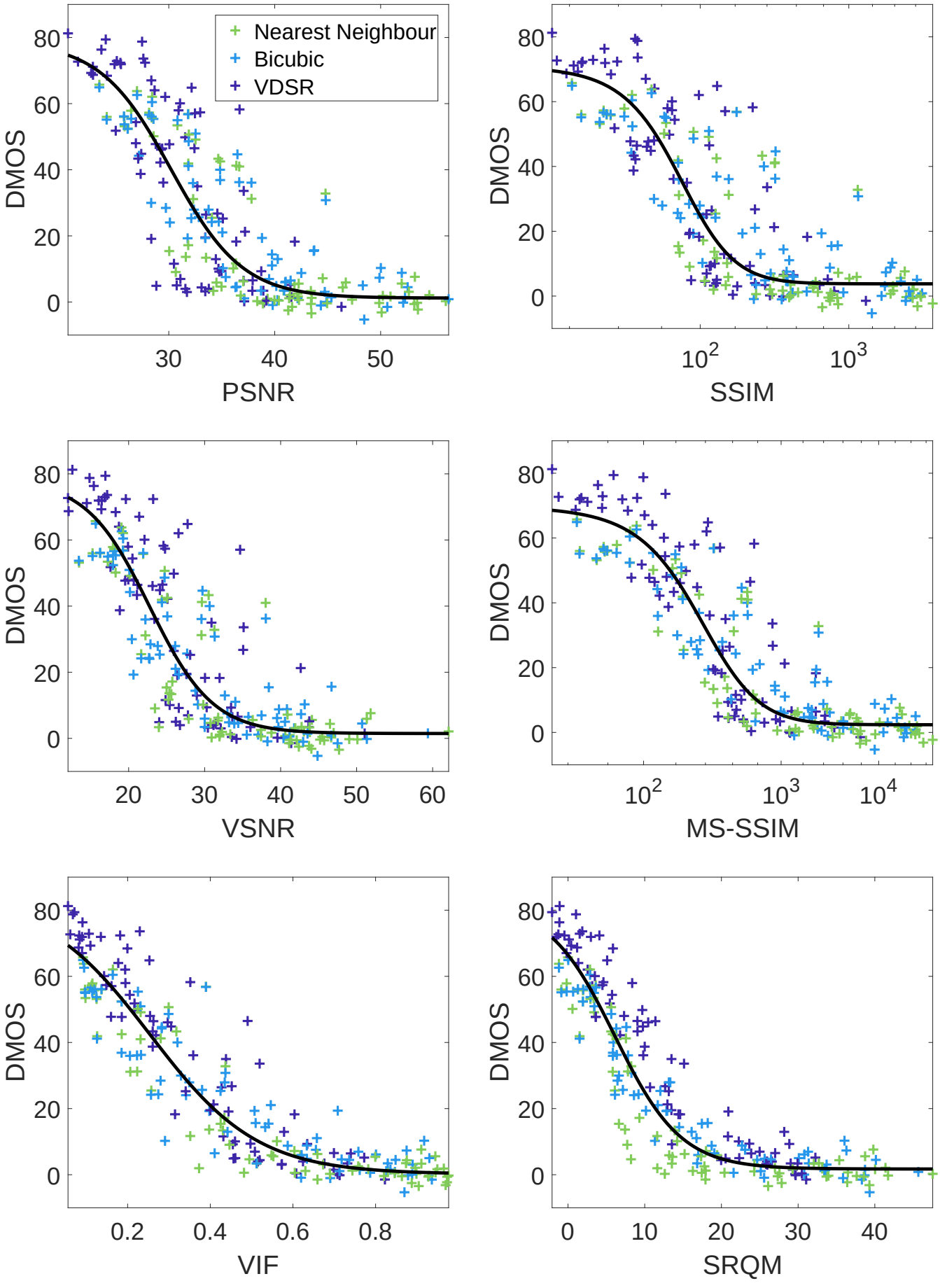

Fig. 4: The relationship between the quality metric predictions and DMOS. The (black) line is the four parameter logistic fitting curve.

TABLE IV: The statistical performance of the quality metrics for each of the adaptation filters. The best performing quality metric in each row is **bold**.

Metric	PSNR	SSIM	VSNR	MS-SSIM	VIF	SRQM
Nearest Neighbour						
SROCC	0.780	0.805	0.862	0.826	0.932	**0.958**
LCC	0.786	0.823	0.871	0.833	0.942	**0.972**
OR	0.6110	0.597	0.500	0.556	0.347	**0.278**
RMSE	17.10	16.04	13.82	15.45	9.25	**6.48**
Bicubic						
SROCC	0.826	0.794	0.870	0.858	0.895	**0.936**
LCC	0.864	0.830	0.906	0.880	0.919	**0.963**
OR	0.458	0.444	0.3610	0.417	0.333	**0.222**
RMSE	10.66	11.95	8.92	10	8.27	**5.59**
VDSR						
SROCC	0.777	0.777	0.848	0.810	**0.871**	0.854
LCC	0.852	0.816	0.899	0.869	0.937	**0.961**
OR	0.4170	0.4030	0.306	0.361	**0.250**	0.306
RMSE	11.89	13.22	9.92	11.26	7.89	**6.22**

TABLE V: F-test results for the quality metrics at a 95% confidence interval for each adaptation filter. The first, second and third elements of each cell correspond to the nearest neighbour, bicubic and VDSR filters respectively.

Metric	PSNR	SSIM	VSNR	MS-SSIM	VIF	SRQM
PSNR	-	0,0,0	0,0,0	0,0,0	-1,-1,-1	-1,-1,-1
SSIM	0,0,0	-	0	0,-1,-1	-1,-1,-1	-1,-1,-1
VSNR	0,0,0	0,0,0	-	0,0,0	-1,0,-1	-1,-1,-1
MS-SSIM	0,0,0	0,1	0,0,0	-	-1,0,0	-1,-1,-1
VIF	1,1,1	1,1,1	1,0,1	1,0,0	-	-1,-1,-1
SRQM	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	-

V. Acknowledgements

The authors received funding from the EPSRC 'Vision for the Future' platform grant (EP/M000885/1).

References

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

[2] J. Li, Y. Koudota, M. Barkowsky, H. Primon, and P. Le Callet, "Comparing upscaling algorithms from HD to Ultra HD by evaluating preference of experience," in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*. IEEE, 2014, pp. 208–213.

[3] 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.

[4] 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.

[5] G. Van Wallendael, P. Coppens, T. Paridaens, N. Van Kets, W. Van den Broeck, and P. Lambert, "Perceptual quality of 4K-resolution video content compared to HD," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*. IEEE, 2016, pp. 1–6.

[6] 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.

[7] 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.

[8] Akamai, "State of the internet report," Q1 2017.

[9] Netflix, "Internet connection speed recommendations," https://help.netflix.com/en/node/306, [Accessed Jan. 2018].

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

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

[13] J. Balle, A. Stojanovic, and J. Ohm, "Models for static and dynamic texture synthesis in image and video compression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1353–1365, 2011.

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

[15] P. Ndjiki-Nya, D. Doshkov, H. Kaprykowsky, F. Zhang, D. Bull, and T. Wiegand, "Perception-oriented video coding based on image analysis and completion: A review," *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 579–594, 2012.

[16] U. Thakur, M. Bhat, M. Bläser, M. Wien, D. Bull, and J. Ohm, "Synthesis of fine details in b picture for dynamic textures," in *Image Processing (ICIP), 2017 24th IEEE International Conference on*, 2017.

[17] R. Wang, C. Huang, and P. Chang, "Adaptive downsampling video coding with spatially scalable rate-distortion modeling," *IEEE transactions on circuits and systems for video technology*, vol. 24, no. 11, pp. 1957–1968, 2014.

[18] G. Georgis, G. Lentaris, and D. Reisis, "Reduced complexity superresolution for low-bitrate video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 2, pp. 332–345, 2016.

[19] M. Afonso, F. Zhang, A. Katsenou, D. Agrafiotis, and D. Bull, "Low complexity video coding based on spatial resolution adaptation," in *Image Processing (ICIP), 2017 24th IEEE International Conference on*, 2017.

[20] D. Bull, F. Zhang, and M. Afonso, "Video Processing Method," *UK Patent: GB1714791.9*, 2017.

[21] 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.
[22] A. Mackin, M. Afonso, F. Zhang, and D. Bull, "A study of subjective video quality at various spatial resolutions," in *Image Processing (ICIP), 2018 25th IEEE International Conference on*, 2018 (Submitted).
[23] P. Abry, "Multirésolutions, algorithmes de décomposition," *Ondelettes et turbulence*, 1997.
[24] 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.
[25] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.
[26] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," 2012.
[27] 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.
[28] 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.
[29] 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*. IEEE, 2003, vol. 2, pp. 1398–1402.
[30] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.
[31] K. Seshadrinathan and A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE transactions on image processing*, vol. 19, no. 2, pp. 335–350, 2010.
[32] 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.
[33] 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.
[34] International Telecom Union, "*Tutorial: Objective perceptual assessment of video quality: Full reference television*," 2004.
[35] 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.
[36] S. Westland, H. Owens, V. Cheung, and I. Paterson-Stephens, "Model of luminance contrast-sensitivity function for application to image assessment," *Color Research & Application*, vol. 31, no. 4, pp. 315–319, 2006.