IEEE International Conference on Image Processing, Beijing, China, Sep. 2017.

PERCEPTUAL ALIASING FACTORS AND THE IMPACT OF FRAME RATE ON VIDEO QUALITY

Rasoul Mohammadi Nasiri and Zhou Wang

Dept. of Electrical and Computer Engineering, University of Waterloo
Emails: {r26moham, zhou.wang}@uwaterloo.ca

ABSTRACT

High frame rate (HFR) videos have become increasingly popular in the past few years, and frame rate is one of the major parameters for adjusting video data rate in real-world video delivery services. To achieve the best trade-off between bandwidth saving and video quality preservation by way of frame rate adaptation, it is essential to understand the impact of frame rate on video quality. In this work, we look at the problem from the viewpoint of perceptual information loss by perceptual aliasing analysis. We propose several measures, namely temporal aliasing power, temporal aliasing factor, spatiotemporal aliasing factor, and perceptual spatiotemporal aliasing factor, and use them as quality predictors for videos under frame rate changes. We also construct a database and conduct a subjective test on videos of different frame rates. Our results suggest great potentials of the proposed perceptual aliasing analysis approach.

***Index Terms*—** Video quality, frame rate, temporal aliasing, spatiotemporal aliasing

1. INTRODUCTION

Videos compose a majority of data traffic over various networks [1]. Service providers aim to deliver high quality videos while at the same time keeping the data rate as low as possible. This has become even more challenging nowadays due to the increasing popularity of higher frame rate (HFR) videos, which often have a frame rate of 60 frame/second (fps) or more, as compared to the traditional 24 fps or 30 fps in standard cinema, television, and Internet video distribution environments. In practice, frame rate reduction is often used to control data rate. However, frame rate reduction may also lead to loss in perceptual quality. To achieve the best compromise between data rate and frame rate, a video quality assessment (VQA) model that considers the impact of frame rate on quality is essential.

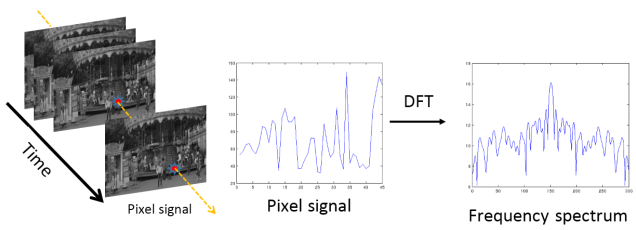
There is a remarkable growth of VQA research in recent years [2]. Most of the VQA methods compare a test video with its pristine version to find distortions and predict the perceptual quality on a per-frame basis. The final video quality score is the average [3] or the weighted average [4] of per-frame quality scores. Nevertheless, the impact of frame rate is not carefully taken into account in most of the existing VQA methods [3, 4].

Limited works have studied the impact of frame rate on video quality. In [5], frame rate, encoder type, content, bitrate, and resolution are used to build a VQA model for low bit rate QCIF and CIF videos with the frame rates from 7.5 to 30 fps. Ou, et. al. have conducted a subjective test by using QCIF and CIF videos with frame rates less than 30 fps, where the impact of frame rate is modeled by exponential terms [6]. The work had been extended by incorporating additional terms of quantization parameter [7] and spatial resolution [8]. In [9], frame rate and resolution changes are employed to estimate video quality by using spatial information (SI) and temporal information (TI) [10]. In [11] a nonlinear parametric model using frame rate as one of the video parameters is proposed. Variations of frame rate have also been explored in low bit rate, low resolution video distributions scenarios [12, 13, 14]. Only in few works, frame rates above 30fps have been exploited [15, 16]. It is shown that in a gaming environment, higher frame rates better entertain the users [15].

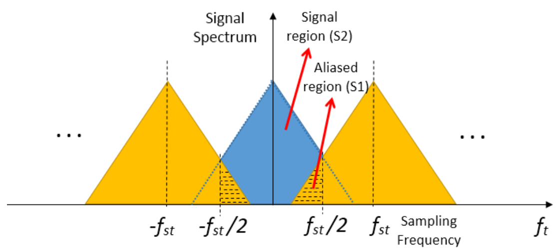
In this work, we explore the impact of frame rate changes on the perceptual quality of videos from the viewpoint of perceptual information loss. In particular, we investigate the spatiotemporal aliasing effect of frame rate reduction and incorporate human visual sensitivity models into the measurement. We also build a database and carry out a subjective test on videos across different frame rates. Our results demonstrate great promises of the proposed perceptual aliasing approach.

2. PERCEPTUAL ALIASING FACTOR ANALYSIS

A video that is continuous in space and time is a three-dimensional signal with one temporal (T) and two spatial (X and Y) dimensions. Real-world digital video is a sampled version of the video, and frame rate is the sampling rate in the temporal direction. Thus, frame rate reduction is downsampling in temporal dimension, which may potentially cause aliasing effect and information loss according to the Nyquist sampling theorem.



(a)



(b)

Fig. 1. Pixel signal representation (a) and its Fourier frequency spectrum (b) after sampling at frame rate f_{st}.

2.1. Temporal Aliasing Factor

Given gray scale video signal (V_g), the pixel value in a specific row (r_i) and column (c_i) over time constitutes a *pixel signal* (u_{r_i,c_i}) given by

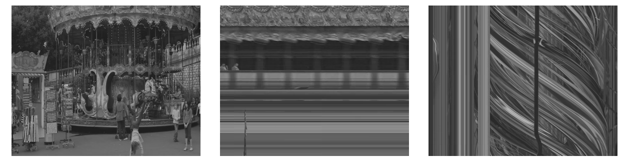
$$u_{r_i,c_i}(t) = \{V_g(r,c,t)|r = r_i, c = c_i\}. \tag{1}$$

This is a continous-time 1D signal as illustrated in Fig. 1(a). When such a signal is sampled at the frame rate f_{st}, its Fourier frequency spectrum is duplicated periodically with a period of f_{st}, as shown in Fig. 1(b). According to the Nyquist sampling theorem, if f_{st} is lower than twice of the highest signal frequency, aliasing occurs. This is evident from the illustration in Fig. 1(b), where the power of the signal from the neighboring duplication of the spectrum that overlaps with the central spectrum $s_{r_i,c_i}(f_t)$ is referred to as the temporal aliasing power (S1 in Fig. 1(b)), which can be computed as

$$P_{r_i,c_i}(f_{st}) = \int_0^{f_{st}/2} |s_{r_i,c_i}(f_{st}-f_t)|^2 df_t = \int_{f_{st}/2}^{f_{st}} |s_{r_i,c_i}(f_t)|^2 df_t \tag{2}$$

Assuming that the strength of aliasing depends on the aliasing power relative to the underlying signal power, we define a temporal aliasing factor as the temporal aliasing power normalized by the signal power (S2 in Fig. 1(b)):

$$A_{T,r_i,c_i}(f_{st}) = \frac{\int_0^{f_{st}/2} |s_{r_i,c_i}(f_{st}-f_t)|^2 df_t}{\int_0^{f_{st}/2} |s_{r_i,c_i}(f_t)|^2 df_t} = \frac{\int_{f_{st}/2}^{f_{st}} |s_{r_i,c_i}(f_t)|^2 df_t}{\int_0^{f_{st}/2} |s_{r_i,c_i}(f_t)|^2 df_t}. \tag{3}$$



(a) X-Y (b) Y-T (c) X-T

Fig. 2. X-Y, X-T, and Y-T planes constructed from 3D video volumn.

This temporal aliasing factor is computed for each pixel signal (as in Fig. 1(a)) extracted from the video and averaged to yield an overall temporal aliasing factor of the sampled video.

$$A_T(f_{st}) = \frac{1}{N}\sum_{i=1}^{N} A_{T,r_i,c_i}(f_{st}). \tag{4}$$

where N is the number of pixel signals involved in the evaluation.

2.2. Spatiotemporal Aliasing Factors

The temporal analysis of pixel signal can be extended to the spatiotemporal space of video. Specifically, the two-dimensional X-T or Y-T planes can be extracted by slicing a horizontal or a vertical line in the frame over time. An example is given in Fig. 2.

Using the X-T (or Y-T) plane, and the corresponding two dimension frequency spectrum $s(f_t, f_x)$, the aliasing due to temporal down-sampling is indicated as the overlap of frequency spectra with its repetition in temporal direction (f_t) as is shown in Fig. 3. Therefore, a spatiotemporal aliasing factor is calculated by

$$A_{ST}(f_{st}) = \frac{\int_0^{f_{sx}/2}\int_0^{f_{st}/2} |s(f_{st}-f_t, f_x)|^2 df_t df_x}{\int_0^{f_{sx}/2}\int_0^{f_{st}/2} |s(f_t, f_x)|^2 df_t df_x} \tag{5}$$

2.3. Perceptual Spatiotemporal Aliasing Factor

The aliasing factors stated so far assume the same importance for all frequency components, but the human visual perception has different sensitivity to different frequencies [17]. This is characterized by the visual contrast sensitivity function (CSF). CSF for moving pictures has been explored by Kelly [17] as a function of both spatial and temporal frequencies (Fig. 4). This function has been employed in many existing works [18, 19] and quantified by [18] as

$$SF(f, v_r) = kc_0c_2v_R(c_12\pi f)^2 exp(\frac{-c_14\pi f}{f_{max}}) \tag{6}$$

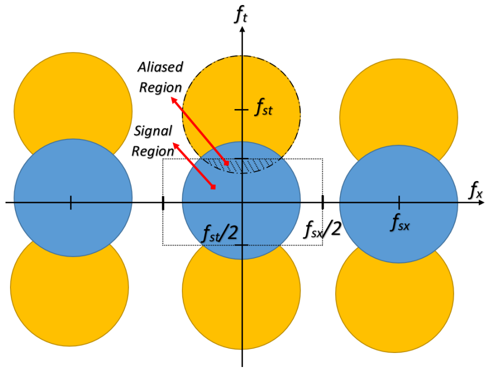


Fig. 3. Aliasing region in spatiotemporal frequency spectrum when the temporal sampling rate is lower than the Nyquist rate.

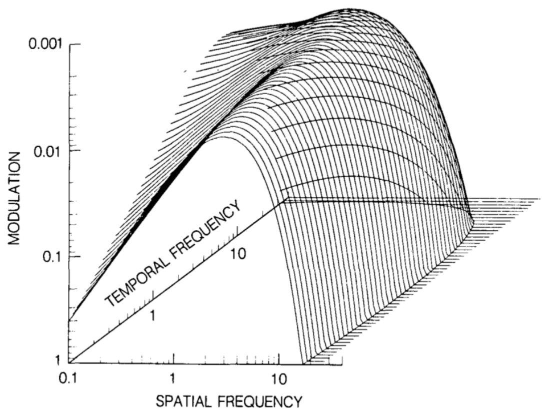


Fig. 4. Spatiotemporal contrast sensitivity function [17].

where k and f_{max} are defined as

$$k = s_1 + s_2|log(\frac{c_2 v_R}{3})|^3, f_{max} = \frac{f_1}{c_2 v_R + 2}, \tag{7}$$

where, $s_1 = 6$, $s_2 = 7.3$, $f_1 = 45.9$, $c_0 = 1.14$, $c1 = 0.67$, and $c_2 = 1.92$ are constants selected according to [18]. In these equations, v_r is the retinal velocity and f is the spatial frequency. SF is the sensitivity as a function of f and v_r. Spatial frequency in Eq. (6) is in cycle/degree unit and can be estimated by spatial frequency of X-T plane (f_x) using

$$f \approx g(f_x) = f_x \, ppd, \tag{8}$$

where ppd is the angular resolution measured by pixel/degree unit. Retinal velocity (v_r) can be estimated by spatial and temporal frequency components by

$$v_R \approx h(f_t, f_x) = \frac{f_t \, FR}{f_x} \tag{9}$$

where, FR is the frame rate. Using Eq. (8), (9), we obtain an estimate of the sensitivity function (λ) as a function of f_t and f_x as follows

$$\lambda(f_t, f_x) = SF(g(f_x), h(f_t, f_x)). \tag{10}$$

Using λ as the visual sensitivity weighting function, we modify Eq. (5) to define a perceptual spatiotemporal aliasing factor by

$$A_{PST}(f_{st}) = \frac{\int_0^{f_{sx}/2} \int_0^{f_{st}/2} \lambda(f_t, f_x)|s(f_{st} - f_t, f_x)|^2 df_t df_x}{\int_0^{f_{sx}/2} \int_0^{f_{st}/2} \lambda(f_t, f_x)|s(f_t, f_x)|^2 df_t df_x} \tag{11}$$

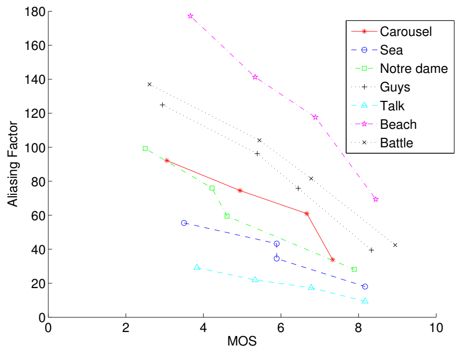
This is calculated for every X-T and Y-T slices extracted from the video, and then averaged to produce an overall perceptual spatiotemporal aliasing factor for the entire video.

3. DATABASE AND EXPERIMENTAL RESULT

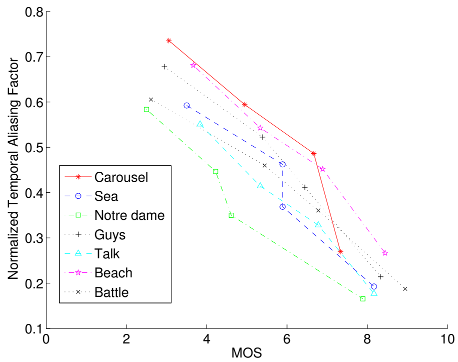
To evaluate the performance of the proposed aliasing factors in predicting video quality degradation, we construct an IVC-HFRVQA database and compare the predictions with human subjective evaluations. The variety of contents, compression parameters, frame rates (up to 60 fps), and resolutions in this database helps us perform a comprehensive evaluation of the proposed aliasing factors. Details regarding the database are given in [20]. The evaluation is based on the correlation between the aliasing factors and the mean opinion score (MOS) values. To reduce the computational complexity, for temporal analysis, we randomly select 10% of pixel signals, and the overall temporal aliasing factor is the average of all pixel signals' temporal aliasing factors. For spatiotemporal analysis, we select uniformly spaced 25% of all video lines and columns.

The aliasing power (P) defined by Eq. (2) and the temporal aliasing factor (A_T) defined by Eq. (4) are calculated for four different frame rates of 5, 10, 15, 30 fps by considering the 60 fps pristine video as the reference video. The computed aliasing power and aliasing factor against MOS values are shown in Figs. 5(a) and 5(b), respectively. It turns out that for each individual video content, the aliasing power increases monotonically with frame rate, and drops monotonically against MOS. However, the aliasing power is highly content dependent. As a result, the points in Fig. 5(a) are widely scattered. By contrast, due to the normalization factor used in the aliasing factor measurement, the resulting aliasing factor scatter appears to be better concentrated, as shown in Fig. 5(b).

The spatiotemporal aliasing factor (A_{ST}) as in Eq. (5) and the perceptual spatiotemporal aliasing factor (A_{PST}) as in Eq. (11) against MOS for all test video contents across four frame rates are shown in Fig. 6(a) and Fig. 6(b), respectively.

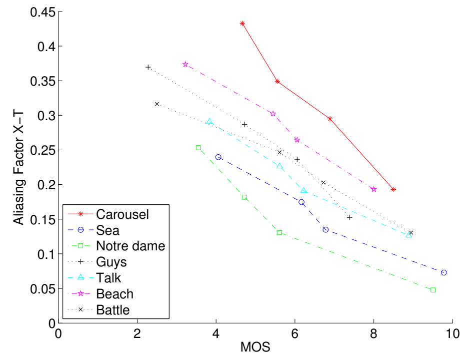


(a)

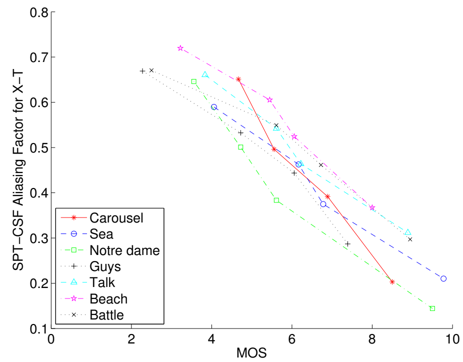


(b)

Fig. 5. Temporal aliasing power (a) and temporal aliasing factor (b) versus MOS for different video contents at four different frame rates.



(a)



(b)

Fig. 6. Spatiotemporal aliasing factor (a) and perceptual spatiotemporal aliasing factor (b) versus MOS for different video contents at four different frame rates.

Compared with Fig. 5, the relationship between aliasing factors and MOS are more linear, indicating the benefit of joint consideration of spatial and temporal aliasing effects. Without considering the effect of human visual sensitivity effect, the spatiotemporal aliasing factor still exhibits strong content-dependency in Fig. 6(a), as the data points spread for different video contents. With the perceptual factor incorporated, the perceptual spatiotemporal aliasing factor largely reduces the effect and the scatter plot in Fig. 6(b) appears to be tightly concentrated. This suggests that the proposed perceptual spatiotemporal aliasing factor is very promising to serve as a key factor in assessing video quality across frame rates.

For quantitative evaluation of the proposed aliasing factors, we calculate the Pearson linear correlation coefficient (PLCC) and Spearman's rank correlation coefficient (SRCC) between different aliasing factors and MOS, and the results are summarized in Table 1. These results confirm our observations in Figs. 5 and 6.

Table 1. Correlations between the proposed aliasing factors and MOS

Method	PLCC	SRCC
Temporal Aliasing Power	0.625	0.626
Temporal Aliasing Factor	0.883	0.884
Spatiotemporal Aliasing Factor	0.764	0.775
Perceptual Spatiotemporal Aliasing Factor	0.934	0.942

4. CONCLUSION

We investigate video quality degradation due to frame rate reduction from the viewpoint of perceptual information loss in terms of various perceptual aliasing factors. Our results suggest that modeling spatial and temporal aliasing effects jointly and taking into account spatiotemporal perceptual sensitivities of the visual system lead to notable success at improving the prediction performance of subjective video quality. It needs to be aware that aliasing is only one of many factors that may affect the perceptual quality of videos. We are currently working towards combining the proposed perceptual aliasing measurements with other perceptual factors to construct a comprehensive model of perceptual video quality and use the model to guide perceptual optimization of video coding and video delivery systems.

5. REFERENCES

[1] Cisco Visual Networking Index, "Forecast and methodology, 2014-2019 white paper," *Technical Report, Cisco, Tech. Rep.*, 2015.

[2] Maria Torres Vega, Maria Torres Vega, Vittorio Sguazzo, Vittorio Sguazzo, Decebal Constantin Mocanu, Decebal Constantin Mocanu, Antonio Liotta, and Antonio Liotta, "An experimental survey of no-reference video quality assessment methods," *International Journal of Pervasive Computing and Communications*, vol. 12, no. 1, pp. 66–86, 2016.

[3] Shyamprasad Chikkerur, Vijay Sundaram, Martin Reisslein, and Lina J Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[4] Abdul Rehman, Kai Zeng, and Zhou Wang, "Display device-adapted video quality-of-experience assessment," in *SPIE/IS&T Electronic Imaging*. International Society for Optics and Photonics, 2015, pp. 939406–939406.

[5] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316–1324, Nov. 2008.

[6] Y. Ou, T. Liu, Z. Zhao, Z. Ma, and Y. Wang, "Modeling the impact of frame rate on perceptual quality of video," in *Proc. IEEE Int. Conf. Image Proc.*, San Diego, CA, Oct 2008, pp. 689–692.

[7] Y. Ou, Z. Ma, and Y. Wang, "Modeling the impact of frame rate and quantization stepsizes and their temporal variations on perceptual video quality: A review of recent works," in *Proc. IEEE Int. Conf. on Information Sciences and Systems*, Princeton, NJ, March 2010, pp. 1–6.

[8] Y. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 21, no. 3, pp. 286–298, 2011.

[9] L. Janowski and P. Romaniak, "QoE as a function of frame rate and resolution changes," in *Future Multimedia Networking*, pp. 34–45. 2010.

[10] P ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," Apr 2008.

[11] Jose Joskowicz and J Ardao, "Combining the effects of frame rate, bit rate, display size and video content in a parametric video quality model," in *Proceedings of the 6th Latin America Networking Conference*. ACM, 2011, pp. 4–11.

[12] Ming-Chen Chien, Ren-Jie Wang, Chien-Hsun Chiu, and Pao-Chi Chang, "Quality driven frame rate optimization for rate constrained video encoding," *Broadcasting, IEEE Transactions on*, vol. 58, no. 2, pp. 200–208, 2012.

[13] J. Joskowicz, R. Sotelo, and Ardao J. C. L., "Towards a general parametric model for perceptual video quality estimation," *IEEE Trans. on Broadcasting*, vol. 59, no. 4, pp. 569–579, Dec. 2013.

[14] A. Khan, L. Sun, and E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks," *IEEE Trans. on Multimedia*, vol. 14, no. 2, pp. 431–442, Apr. 2012.

[15] M. Claypool, K. Claypool, and F. Damaa, "The effects of frame rate and resolution on users playing first person shooter games," in *Proc. SPIE 6071, Multimedia Computing and Networking*, San Jose, CA, Jan. 2006.

[16] A. Banitalebi-Dehkordi, M. T. Pourazad, and P. Nasiopoulos, "The effect of frame rate on 3D video quality and bitrate," *3D Research*, vol. 6, no. 1, pp. 1–13, Dec. 2014.

[17] D.H. Kelly, "Motion and vision. II. stabilized spatio-temporal threshold surface," *Journal of Optometry Society of America*, vol. 69, no. 10, pp. 1340–1349, 1979.

[18] Scott J 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.

[19] A Murat Demirtas, Amy R Reibman, and Hamid Jafarkhani, "Full reference video quality estimation for videos with different spatial resolutions," in *International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 1997–2001.

[20] Rasoul Mohammadi Nasiri, Jiheng Wang, Abdul Rehman, Shiqi Wang, and Zhou Wang, "Perceptual quality assessment of high frame rate video," in *Multimedia Signal Processing (MMSP), 17th International Workshop on*. IEEE, 2015, pp. 1–6.