# Video Quality Model for Space-Time Resolution Adaptation

Dae Yeol Lee
*Department of Electrical and Computer Engineering, The University of Texas at Austin*
Austin, TX, USA
daelee711@utexas.edu

Hyunsuk Ko
*Division of Electrical Engineering, Hanyang University ERICA*
Ansan, Republic of Korea
hyunsuk@hanyang.ac.kr

Jongho Kim
*Media Coding Research Section, Electronics and Telecommunication Research Institute*
Daejeon, Republic of Korea
pooney@etri.re.kr

Alan C. Bovik
*Department of Electrical and Computer Engineering, The University of Texas at Austin*
Austin, TX, USA
bovik@ece.utexas.edu

2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS) | 978-1-7281-7574-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/IPAS50080.2020.9334940

***Abstract*— Delivering voluminous amounts of video data through limited bandwidth channels is a challenge affecting billions of viewers. Accordingly, it is becoming more important to understand the perceptual effects that arise from various dimension reduction methodologies. Towards this direction, we propose a new video quality model that predicts the perceptual quality of videos undergoing varying levels of spatio-temporal subsampling and compression. The new model is established upon the natural statistics principle of videos, which leverage the fact that pristine videos obey statistical regularities that are disturbed by distortions. We found that there exist space-time paths between video frames that best preserve the statistical regularity inherent in the spatial structure of the video frames. The distribution features extracted from frame differences displaced in the direction of these paths correlate more highly with human subjective quality opinions than those from non-displaced frame differences. Given that non-displaced frame differences are widely utilized in video quality models, the improved efficiency of spatially and/or temporally displaced (possibly by more than one frame) frame differences, is an important finding that may significantly elevate the success of studies on temporal features and video quality.**

***Keywords*— *Video quality, spatio-temporal resolution, video compression, natural video statistics, statistical regularity, displaced frame difference***

## I. INTRODUCTION

The media industry continuously progresses towards providing more realistic and immersive experiences by delivering contents having higher spatial resolution and frame rates. However, these expansions in video dimensions also dramatically increases the data volume. Consequently, the methods used to effectively compress and deliver video data through limited bandwidth channels are becoming more important. A common practice by content providers is to down-sample and encode videos prior to transmission. However, these operations may degrade the perceptual quality of the delivered contents, hence it is therefore important to understand the trade-off between data reduction amount and the perceptual effects of downscaling and compression.

Relating to this issue, in [1-3] the authors conducted human studies on the mutual effects of spatial down-sampling and compression on the perceptual quality of videos. However, these studies did not incorporate the effects of temporal down-sampling, which is a topic of recent interest. In [4-5], the authors focused on temporal resolution adaptation methods that reduce the frame rate if the contents do not perceptually benefit from a higher frame rate. However, these studies did not investigate into the combined effects of temporal down-sampling and compression. In [6], a spatio-temporal resolution adaptation method for video compression is proposed, but quality prediction and consequent decisions to down-sample are conducted separately on spatial and temporal resolution. These prior studies have helped us understand how each of spatial and temporal parameters affect the perceptual quality of videos. However, less work has been directed towards predicting the quality of videos when both spatial and temporal down-sampling is applied in conjunction with compression.

Here we propose a new video quality model that correlates highly with human subjective data collected on videos on which varying levels of spatio-temporal down-sampling and compression have been applied. The remainder of the paper is organized as follows. Recent findings on the natural statistics of space-time displaced frame differences are provided in Section 2. The details of our proposed video quality model are provided in Section 3. The experimental results are provided in Section 4. Finally, conclusions are drawn in Section 5.

## II. ON THE SPACE-TIME STATISTICS OF VIDEOS

Various studies have investigated the relationships that exist between the statistical properties of natural images and front-end processing in the visual system [7-12]. Modifications of these natural scene statistics models have been used with great success for image and video quality prediction [13-19]. The most widely accepted models involve bandpass decomposition similar to decorrelating processes that occur in the retina and area V1, followed by divisive normalization, which accounts for nonlinear gain control in retino-cortical neurons. These processes shape the distributions of the responses to be statistically regular. Image quality prediction models assume a statistical regular distribution on the pristine image, that can be used to discriminate them from a distorted image. Moreover, they can be effectively used to predict the quality of distorted

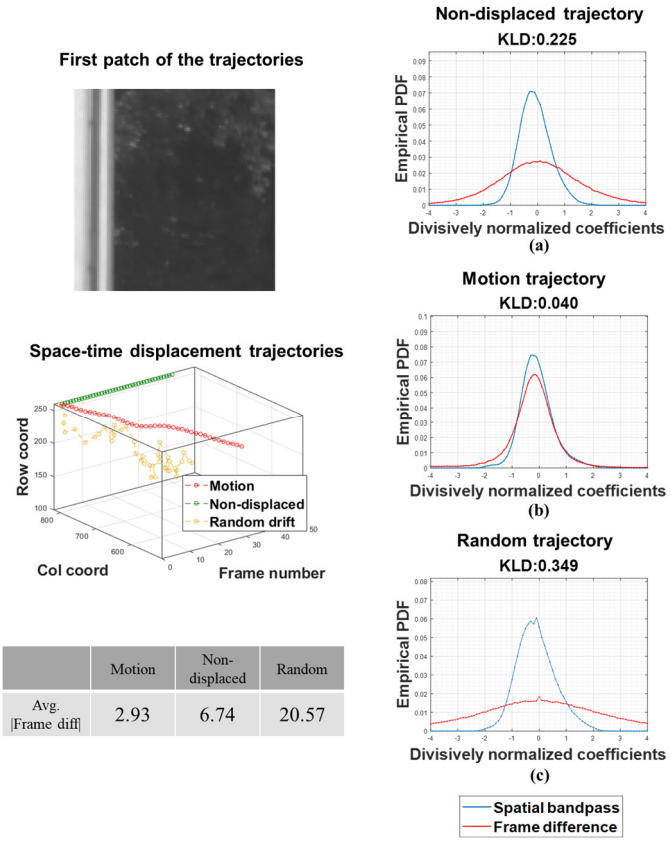
978-1-7281-7574-4/20/$31.00 ©2020 IEEE

images by measuring statistical deviations of the distorted image with respect to the pristine model. These statistical models have evolved to include videos principally, exploiting the statistical regularity of simple frame differences [18, 19].

Here we broaden and deepen modeling of frame difference statistics by introducing the use of both spatial and temporal (possibly more than one frame) displacements between frames before differencing them. This is similar, of course, to the concept of motion compensation, but with a different aim of finding space-time paths having optimal statistical regularity. We show that displaced frame differences can be used to create video quality prediction models that correlate more highly against human subjective data. This concept also relates to recent perceptual theories of very small eye movements, called microsaccades. When viewing a scene, the eyes do not remain still, but instead engage in movements such as smooth pursuit, saccades, and microsaccades [20, 21]. It is theorized that microsaccades may enhance the process of achieving efficient visual encoding in the brain. Moreover, visual signals received from the retina in visual cortex are subjected to processes of temporal lag filtering [22], which equates to smoothed temporal differencing which contributes to reduced information redundancy and improved encoding of visual signals. Likewise, differencing of frames that are relatively spatially and/or temporally displaced can achieve similar goals of space-time redundancy reduction.

Fig. 1 shows plots of the distribution of frame differences displaced along various space-time displacement trajectories as compared to spatially band-passed coefficients. We used videos from HD1K optical flow database [23], which provides ground-truth optical flow vectors. We first collected the frame patches along motion, non-displaced, and random trajectories. Then we applied spatial bandpass and divisive normalization by computing mean-subtracted contrast normalized (MSCN) [16] coefficients on the collected patch volume. The distribution plots of these divisively normalized spatial band-passed signals are depicted as blue curves in Figs. 1(a)-1(c). As the natural image statistics model suggests, the distributions of spatial bandpass coefficients form statistically regular distributions over all trajectories, when divisively normalized. We then investigated the distributions of displaced frame differences by first collecting frame-differenced patches along motion, non-displaced, and random trajectories. We applied divisive normalization on the collected frame difference patch volumes, similar to the spatial bandpass case. The distribution plots of these divisively normalized displaced frame differences are depicted as red curves in Figs. 1(a)-1(c). As shown in the figure, applying different space-time displacement paths reveals that there are inherent statistical regularities in frame difference signals that are aligned along motion directions [26]. Also, we see that the distribution of frame difference displaced along motion direction yield the lowest values of the Kullback Leibler Divergence (KLD) with respect to the statistically regular, spatially band-passed coefficients.

Motivated by these observations, we devised a method to determine appropriate displacement paths that capture the inherent statistical regularity of spatially displaced and temporally displaced (possibly more than one frame) videos. We used displaced frame differences to extract distribution features




| | Motion | Non-displaced | Random |
|---|---|---|---|
| Avg. \|Frame diff\| | 2.93 | 6.74 | 20.57 |

Fig. 1. Comparison of distributions between spatially band-passed coefficients and spatially displaced frame differences displaced along various space-time trajectories (motion, non-displaced, and random), which have been subjected to divisive normalization.

from both pristine and distorted videos, and utilized them to construct a full reference (FR) video quality model. As the space-time path of maximum regularity is also related to the motion of the video, we expect these features to capture richer information descriptive of losses of regularity, and hence, changes in perceived video quality.

## III. Proposed Video Quality Metric

An overall flowchart of our new video quality model is shown in Fig. 2. The proposed method first analyzes the initial portion of the reference video, then determines the proper displacement vector to be applied in the frame differencing procedure. The displacement vector refers to the local space-time displacement path that best preserves statistical regularity. We have demonstrated that such a path is strongly predictive of the motion direction. These displacement vectors are re-computed every second by analyzing the initial portion (200msec) of each segment. Once the displacement vector is decided, then the resulting vector, along with the frames from the reference and distorted videos, are delivered to the next processing modules, where displaced frame difference are computed followed by divisive normalization. Then, the statistical feature extraction module analyzes the processed frame difference signals to measure statistical discrepancies between the reference and distorted videos, ultimately extracting

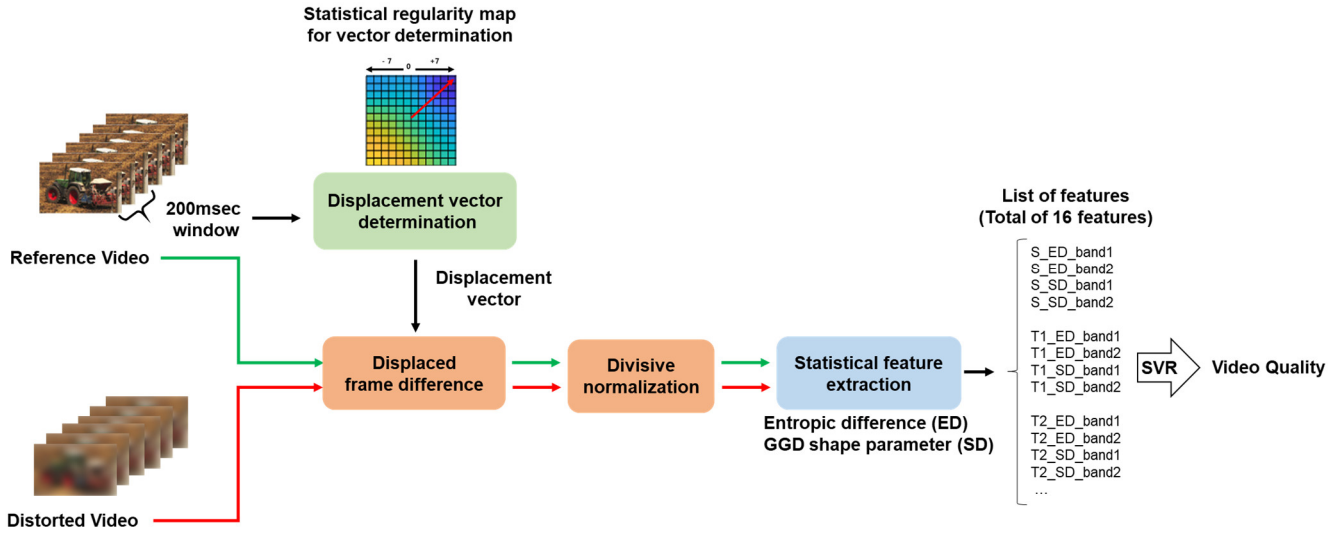



Fig. 2. Flowchart of the proposed video quality model.

multiple quality-predictive features that are combined using support vector regression.

### A. *Space-Time Displaced Frame Difference*

Unlike older Video Quality Assessment (VQA) models that compute optical flow [24, 25] our model does not compute motion vectors, but instead finds space-time paths of maximal statistical regularity. The first step forwards determining the best spatial and/or temporal displacement between frames is to quantify the degree of resulting statistical regularity. This is accomplished by analyzing the distributions of coefficients obtained by taking between-frame differences of pristine video frames that are relatively displaced followed by a divisive normalization operation. Fig. 3 depicts the processes of constructing a statistical regularity map that is used to determine the optimal displacement vector. As we have stated already, the statistically regular displacement path is the one that yields the lowest KLD with respect to the distribution of divisively normalized, spatially band-passed signals. We use this to construct a statistical regularity map that contains information regarding how well the displaced frame differences preserve the statistical regularity inherent in spatial structures for diverse displacements. The left side of Fig. 3 depicts an example of a statistical regularity map, where the dark blue regions indicate locations where the KLD value is low. The optimal displacement path that best preserves statistical regularity is determined by taking the average of those displacement vectors yielding KLD values less than 95% (the lowest 5$^{\text{th}}$ percentile) of the values in the map. The displacement vector is then used to compute displaced frame differences from the reference and the distorted video.

### B. *Statistical Features*

The displaced frame difference computed from the reference and distorted videos are subject to a divisive normalization procedure motivated by the well-known Gaussian Scale Mixture (GSM) image model as follows [18, 19]. Partition the plane of coefficients into non-overlapping patches indexed by $m \in \{1,2,\dots,M_b\}$. Then the coefficients $C_{mb}$ of the $m^{\text{th}}$ patch in sub-band $b$ can be modeled as

$$C_{mb} = S_{mb}U_{mb}, \tag{1}$$

where $S_{mb}$ is a random scalar variable that is independent of the random field $U_{mb}$, which is distributed as $U_{mb} \sim \mathcal{N}(0, \mathbf{K}_b)$, with covariance matrix $\mathbf{K}_b$. Estimate the scalar $S_{mb} = \hat{s}_{mb}$ for each patch using the Maximum Likelihood (ML) procedure:

$$\hat{s}_{mb} = \text{argmax}_{s_{mb}} p(C_{mb}|S_{mb}) = \sqrt{\frac{C_{mb}^T \mathbf{K}_b^{-1} C_{mb}}{N}}, \tag{2}$$

where $N$ is the number of coefficients within each patch, and normalize the frame difference coefficients of each patch by the respective estimates $\hat{s}_{mb}$, whence $\text{N}_{mb} = \frac{C_{mb}}{\hat{s}_{mb}} \sim \mathcal{N}(0, \mathbf{K}_b)$. The aggregated divisively normalized coefficients $\text{N}_{mb}$ over all patches is expected to follow a statistically regular distribution for pristine videos. We quantify how much the distribution of the pristine video is affected by distortion using two kinds of statistical features.

The first is an entropic difference feature. The feature measures distortions by computing the information difference between the divisive normalization factors of the reference and distorted videos. The entropic difference (ED) is formulated as,

$$ED = \sum_{m=1}^{M} |\gamma_{mr} h(C'_{mr}|S_{mr} = \hat{s}_{mr}) - \gamma_{md} h(C'_{md}|S_{md} = \hat{s}_{md})|, \tag{3}$$

where $\gamma_x = \log(1 + s_x^2)$, and $h(C'_x|S_x = s_x) \sim \frac{1}{2}\log\left((2\pi e)^N |s_x^2 K_U|\right)$. The subscripts $x = mr$ and $x = md$ refer to the $m^{\text{th}}$ patch of the reference and distorted videos, respectively. We interpret the entropic difference as a feature that compares the information content inherent in the divisive normalization factors.

The second is the shape parameter of the Generalized Gaussian Distribution (GGD). The divisively normalized

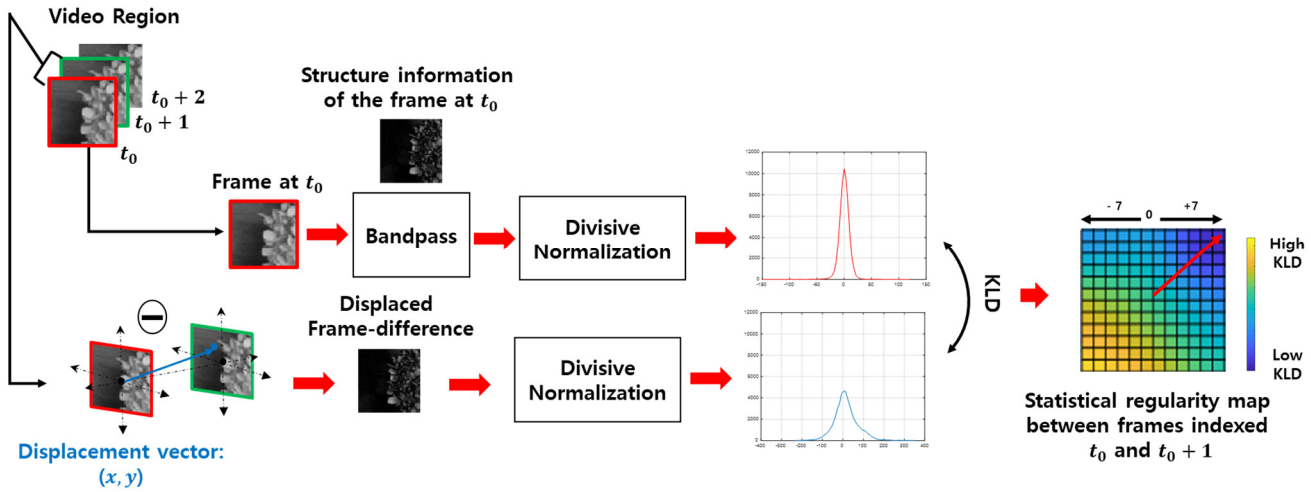



Fig. 3. Processing flow when constructing a statistical regularity map, which we use to determine the displacement vector.

bandpass frame difference coefficients are modeled using the GGD as

$$f(x;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)}\exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right), \quad (4)$$

where $\beta = \sigma\sqrt{\frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)}}$, $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$, $a > 0$ is the gamma function, $\alpha$ is the shape parameter, and $\sigma^2$ is the variance. The shape parameters of the reference ($\alpha_r$) and the distorted ($\alpha_d$) videos are compared by taking the absolute difference

$$SD = |\alpha_r - \alpha_d|. \quad (5)$$

We interpret the meaning of the GGD shape difference as a feature that captures the differences in shape of the regular distributions generated by the aforementioned divisive normalization factors.

### C. Final set of features

The statistical feature extraction module outputs the aforementioned two features (ED and SD) from a set (reference and distorted) of input planes. We have a total of four sets of input planes denoted

- $f_i$: The $i^{th}$ spatial frame,
- $D(f_i, f_{i+T_1}, v)$ : Frame difference with temporal separation $T_1$, and displacement vector $v$,
- $D(f_i, f_{i+T_2}, v)$ : Frame difference with temporal separation $T_2$, and displacement vector $v$,
- $D(f_i, f_{i+T_3}, v)$ : Frame difference with temporal separation $T_3$, and displacement vector $v$,

where $v$ refers to the spatial vector determined from the displacement determination module. The indices $T$, $T_2$, and $T_3$ refer to frame difference separations of one or more frames, chosen to correspond to micro-saccadic eye movements of durations less than 200ms [20, 21]. In addition, we also consider the aforementioned input planes at half resolution as well, to allow for the multiscale computations. Therefore, we have a total of 8 sets of input planes, and consequently, a total of 16 features which are combined by the support vector regression (SVR).

The nomenclature used for the final set of features follows the form of (Plane type)_(Feature type)_(Resolution type), where

- **Plane type:** spatial frame (S), or frame difference plane with varying temporal separation (T1, T2, and T3),
- **Feature type:** Entropic difference (ED) or GGD shape parameter difference (SD),
- **Resolution type:** Full resolution (band1) or half resolution (band2).

## IV. Experiment Result

To evaluate the prediction performance of our features and the final video quality model, we constructed a large scale video database that contains the human subjective quality scores on videos subjected to various levels of spatio-temporal distortions and compressions. The database was constructed from 15 source contents of 4K 10 bit format, of which five contents were 120Hz and ten contents were 60Hz. The distortion types that are applied to the videos are:

- **Spatial down-sampling:** 4K(orig) → 1080p/720p/540p
- **Temporal down-sampling:** Full (120/60 Hz) → Half (60/30Hz)
- **Compression:** HEVC (x265) compression, 3~4 QP points so that the spatio-temporal subsampled video falls into one of several bitrate categories, visually chosen to be generally perceptually distinguishable

The aforementioned distortion types were jointly applied to the source videos. As a result we generated a total of 437 distorted videos, of which 227 are full frame rate with spatial

TABLE I. PREDICTION PERFORMANCE OF TEMPORAL MODELS/FEATURES BASED ON ENTROPIC DIFFERENCE COMPUTATION.

| | Model/Feature | Full frame rate | | Half frame rate | | Overall (full + half frame rate) | |
|---|---|---|---|---|---|---|---|
| | | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| Utilizing non-displaced frame differences | T-RRED | 0.697 | 0.672 | 0.213 | 0.128 | 0.393 | 0.229 |
| | T-SpEED | 0.773 | 0.715 | 0.201 | 0.102 | 0.397 | 0.175 |
| Proposed, utilizing space-time displaced frame differences | T1_ED_band1 | 0.868 | 0.854 | 0.336 | 0.323 | 0.515 | 0.441 |
| | T2_ED_band1 | 0.874 | 0.860 | 0.337 | 0.325 | 0.523 | 0.446 |
| | T3_ED_band1 | 0.877 | 0.860 | 0.331 | 0.320 | 0.524 | 0.445 |

TABLE II. PREDICTION PERFORMANCE RESULTS FOR 5-FOLD CROSS VALIDATION ON 1000 ITERATIONS.

| | Full frame rate | | Half frame rate | | Overall (full + half frame rate) | |
|---|---|---|---|---|---|---|
| | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| PSNR | 0.67±0.15 | 0.64±0.15 | 0.50±0.18 | 0.38±0.21 | 0.51±0.13 | 0.46±0.15 |
| MSSSIM | 0.75±0.14 | 0.70±0.14 | 0.47±0.18 | 0.31±0.21 | 0.53±0.13 | 0.38±0.19 |
| S-RRED | 0.85±0.08 | 0.84±0.10 | 0.54±0.20 | 0.42±0.24 | 0.63±0.14 | 0.55±0.18 |
| T-RRED | 0.80±0.13 | 0.77±0.14 | 0.40±0.18 | 0.20±0.16 | 0.47±0.11 | 0.28±0.14 |
| ST-RRED | 0.83±0.11 | 0.73±0.15 | 0.49±0.19 | 0.26±0.22 | 0.53±0.14 | 0.27±0.20 |
| S-SpEED | 0.89±0.07 | 0.89±0.07 | 0.52±0.18 | 0.41±0.22 | 0.58±0.13 | 0.50±0.16 |
| T-SpEED | 0.84±0.09 | 0.81±0.12 | 0.40±0.18 | 0.17±0.15 | 0.43±0.11 | 0.21±0.13 |
| SpEED | 0.87±0.08 | 0.77±0.14 | 0.43±0.18 | 0.25±0.19 | 0.49±0.12 | 0.21±0.17 |
| VIF | 0.76±0.12 | 0.73±0.13 | 0.53±0.18 | 0.44±0.24 | 0.60±0.13 | 0.53±0.17 |
| VMAF | 0.78±0.16 | 0.76±0.15 | 0.62±0.20 | 0.59±0.23 | 0.67±0.16 | 0.66±0.17 |
| Proposed | 0.87±0.08 | 0.88±0.08 | 0.65±0.16 | 0.64±0.17 | 0.75±0.11 | 0.73±0.10 |

down-sampling and compression applied, and 210 videos are half frame rate videos with spatio-temporal down-sampling and compression applied. The ACR-HR (ACR with Hidden Reference removal) method was used as the subjective protocol, and total of 34 participants rated the videos. We will be making this video database and the detailed reports public soon.

Table 1 shows the prediction performances of the temporal models/features that utilize simple, non-displaced and displaced frame differences. T-RRED [18] and T-SpEED [19] compute entropic differences on non-displaced frame difference, while our proposed temporal features compute entropic differences on spatio-temporally displaced frame difference. As may be seen from the results, the prediction performance of the proposed model increased as compared to T-RRED and T-SpEED. This indicates that the statistical information contained within the frame differences displaced in the motion direction better predicts perceptual quality than do simple non-displaced frame differences.

Table 2 and Fig. 4 present comprehensive performance comparisons of our proposed model against other high-performing models such as VIF [13], spatio-temporal RRED [18], SpEED [19], and VMAF [27]. Our proposed video quality model uses just 16 features, which were used to train a SVR with radial basis function (RBF). The SVR-RBF parameters were determined using cross validation within the trainset, as presented in [28]. Our database consists of videos afflicted by various type of distortions applied on the same source content. Therefore, special care must be taken to separate the train/test sets 'content-wise.' This means that the train/test sets do not share videos having the same source contents. For performance evaluation, we used 5-fold cross validation. Since our database consists of 15 source contents, this means that the SVR model was trained using videos from 12 source contents and tested on

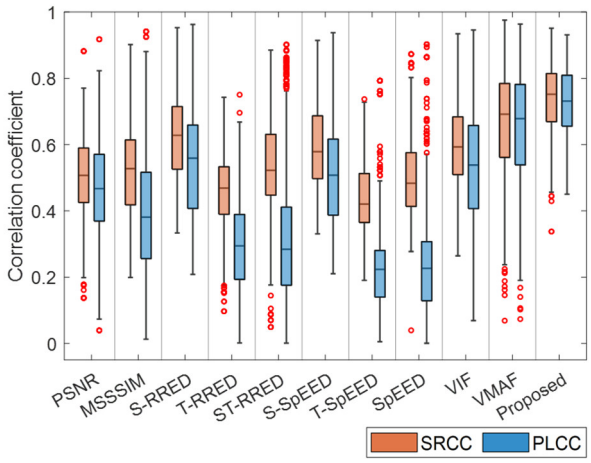



Fig. 4. Box plots of SRCC and PLCC for all videos over 1000 iterations of random train-test splits. For each box, the center line indicates the median, the edges of the box represent the 25th and 75th percentiles, and outliers are indicated by red circles.

the videos from the other 3 source contents. We ran 1000 train/test iterations, where the train/test sets were randomly split over each iteration while abiding by the content-wise separation.

The results in Table 2 show the medians and standard deviations of prediction performance across the 1000 train/test splits. For fair comparison, we also measured the performances of the other models on the same splits. As seen in the results, our proposed model was able to out-perform the other ones in most cases.

A tendency we see is that, for most models, the full frame rate results are higher than the half frame rate results. The full

frame rate results are computed from the subset of the database containing only compression and spatial down-sampling distortions. Most of the models can cope with mixtures of the two distortions. S-SpEED delivered very high performance when only considering the full frame rate case, where it slightly outperformed the proposed method. However, we see that the performances of most of the models fall considerably when temporal down-sampling is introduced as a distortion type. Most of the models either consider distortions at a frame level or utilize simple frame differences, but apparently, these features are inadequate to capture the perceptual losses of temporal down-sampling. Among the various models, VMAF achieved relatively high performance overall and on the half frame rate case, but its performance was not very high on the full frame rate case. Our proposed model, which utilizes the statistical information derived from spatially and temporally displaced frame differences (optimized to maximize space-time statistical regularity), yielded the best prediction performance overall, by a wide margin.

## V. Conclusion

We have presented our findings on the usefulness of the space-time natural statistics of videos for quality prediction. By finding space-time displacement paths between frames that best preserve statistical regularities inherent in videos, we derive features from the optimally displaced frame differences which better correlate with human percepts of quality. We used these features to create a new video quality model that can account for the joint perceptual effects of spatio-temporal subsampling and compression. The findings from this study may be fruitfully utilized to assist optimal space-time resolution adaptation strategies for perceptual video coding.

## Acknowledgment

This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No. 2017-0-00072, Development of Audio/Video Coding and Light Field Media Fundamental Technologies for Ultra Realistic Teramedia).

## References

[1] J. Y. Lin, R. Song, C. H. Wu, T. Liu, H. Wang, and C. C. J. Kuo, "MCL-V: A streaming video quality assessment database," J. Vis. Commun. Image Represent., vol. 30, pp. 1–9, July 2015.

[2] M. Cheon and J.-S. Lee, "Subjective and objective quality assessment of compressed 4K UHD videos for immersive experience," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 7, pp. 1467–1480, Jul. 2018.

[3] A. Mackin, M. Afonso, F. Zhang, and D. Bull, "A study of subjective video quality at various spatial resolutions," IEEE Int. Conf. Image Process., Oct. 2018, pp. 2830–2834.

[4] Q. Huang et al., "Perceptual quality driven frame-fate selection (PQD-FRS) for high-frame-rate video," IEEE Trans. Broadcast., vol. 62, no. 3, pp. 640–653, May 2016.

[5] A. V. Katsenou, D. Ma, and D. R. Bull, "Perceptually-aligned frame rate selection using spatio-temporal features," IEEE Pict. Coding Symp., June 2018, pp. 288–292.

[6] M. Afonso, F. Zhang, and D. R. Bull, "Video compression based on spatio-temporal resolution adaptation," IEEE Trans. Circuits Syst. Video Technol, vol. 29, no. 1, pp. 275–280, Oct. 2018.

[7] E. P. Simoncelli, and B. A. Olshausen. "Natural image statistics and neural representation," Annu. Rev. Neuroscience, vol. 24, no. 1, pp. 1193-1216, Mar. 2001.

[8] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," J. Math. Imag. Vision, vol. 18, no. 1, pp. 17-33, Jan. 2003.

[9] D. L. Ruderman, "The statistics of natural images," Netw.: Comput. Neural Syst., vol. 5, no. 4, pp. 517-548, July 1994.

[10] D. J. Heeger, "Normalization of cell responses in cat striate cortex," Vis. Neuroscience, vol. 9, no. 2, pp. 181– 197, 1992.

[11] W. S. Geisler, and D. G. Albrecht, "Cortical neurons: Isolation of contrast gain control," Vision Res., vol. 32, pp. 1409–1410, Aug. 1992.

[12] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, "Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons," In R. Rao, B. Olshausen, and M. Lewicki, Eds. Probabilistic Models of the Brain: Perception and Neural Function, Cambridge, MA: MIT Press, 2002, pp. 203–222.

[13] H. R. Sheikh, and A. C. Bovik, "Image information and visual quality," IEEE Trans. on Image Process., vol. 15, no. 2, pp. 430-44, Feb. 2006.

[14] R. Soundararajan, and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," IEEE Trans. on Image Process., vol. 21, no. 2, pp. 517-526, Feb. 2012.

[15] A.K. Moorthy, and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," IEEE Trans. on Image Process., vol. 20, no. 12, pp. 3350-3364, Dec. 2011.

[16] A. Mittal, A. K. Moorthy, and A.C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Trans. on Image Process., vol. 21, no. 12, pp. 4695-708, Dec. 2012.

[17] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209-212, Mar. 2012.

[18] R. Soundararajan, and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 4, pp. 684-694, Aug. 2012.

[19] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "SpEED-QA: Spatial efficient entropic differencing for image and video quality," IEEE Signal Process. Lett., vol. 24, no. 9, pp. 1333-1337, Sept. 2017.

[20] B. Fischer, and E. Ramsperger, "Human express saccades: extremely short reaction times of goal directed eye movements," Exp. Brain Res., vol. 57, no. 1, pp. 191-195, Jan. 1984.

[21] W. M. Joiner, and M. Shelhamer, "Pursuit and saccadic tracking exhibit a similar dependence on movement preparation time," Exp. Brain Res., vol. 173, no. 4, pp. 572-586, Sep. 2006.

[22] D. W. Dong and J. J. Atick, "Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus," Netw.: Comput. Neural Syst., vol. 6, no. 2, pp. 159-178, Jan. 1995.

[23] D. Kondermann et al., "The HCI benchmark suite: stereo and flow ground truth with uncertainties for urban autonomous driving," in Proc. IEEE Conf. Comput. Vision Pattern Recog., 2016, pp. 19-28.

[24] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," IEEE Int. Conf. Acoust., Speech, Signal Process., April, 2007, pp. I-869-I-872.

[25] K. Seshadrinathan and A. C. Bovik, "Motion-tuned spatio-temporal quality assessment of natrual videos," IEEE Trans. on Image Process., vol. 19, no. 2, pp. 335-350, Feb. 2010.

[26] D. Lee, H. Ko, J. Kim, and A. C. Bovik, "On the space-time statistics of motion pictures", J. Vision, submitted.

[27] Z. Li et al., "VMAF: The journey continues," The NETFLIX tech blog, 2018. [Online]. Available: https://medium.com/netflix-techblog/ vmaf-the-journey-continues-44b51ee9ed12

[28] C. W. Hsu, C. C. Chang and C. J. Lin, "A Practical Guide to Support Vector Classification," Technical Report, Department of Computer Science and Information Engineering, University of National Taiwan, Taipei, pp. 1-12., 2003.