BLIND VIDEO QUALITY ASSESSMENT BASED ON SPATIO-TEMPORAL INTERNAL GENERATIVE MECHANISM

Yun Zhu, Yongfang Wang, Yuan Shuai

School of Communication and Information Engineering, Shanghai University
Shanghai 200072, China

ABSTRACT

In this paper, we present a new blind video quality assessment metric considering the characteristics of human visual system (HVS). HVS indicates that internal generative mechanism (IGM) actively predicts the visual scene (predicted information) and avoid the residual uncertainty (uncertain information). Based on spatio-temporal internal generative mechanism (ST-IGM), we employ a spatio-temporal autoregressive (AR) prediction model to disassemble the video content into the predicted part and the uncertain part. Then, we separately evaluate their quality degradations by natural video statistics (NVS) model based blind VQA method. According to the perception of distortions in two parts, different weights are assigned to yield the overall quality. The experimental results demonstrate that the proposed algorithm performs much better than the state-of-the-art blind train-free algorithm on the LIVE VQA database and shows competitive performance with the existing train-based methods.

Index Terms— natural video statistics, video quality assessment, spatio-temporal internal generative mechanism

1. INTRODUCTION

Due to the rapid growth of video application, video quality assessment becomes increasingly important. The subjective quality assessment is time-consuming and unpracticed in some cases. In general, the objective video quality assessment (VQA) is to build a mathematical model by extracting some information from the video automatically, which is more desirable compared to subjective quality assessment.

According to the availability of reference information, objective VQA is composed of three categories, namely full-reference (FR) VQA, reduced-reference (RR) VQA, and no-reference (NR) VQA. FR-VQA models always achieve excellent performance because the original videos are required as the reference to assess the visual quality of the distorted videos, such as SSIM [1], MAD [2], ST-MAD [3], MOVIE [4] and so on. RR-VQA is a compromise method between FR-VQA and NR-VQA as it only needs to extract some features from the reference video which are transmitted as auxiliary to build VQA model. Compared to the large data of the reference video of FR-VQA, RR-VQA has a good tradeoff between bandwidth occupation and superb performance [5] [6] [7]. Compared with FR-VQA, RR-VQA, NR-VQA is a great challenge for VQA research due to its complete lack of reference videos. Currently, most NR-VQA studies focus on predicting the quality of image/video by combining the estimated specific distortions, such as blockiness [8], packet loss [9], and temporal inconsistency caused by hole filling [10]. To construct a complete 'prior-information-free' VQA model, the blind VQA is proposed, which needs neither distortion type nor the codec format.

Recently, there have been some studies on blind VQA methods. In [11], natural video statistics (NVS) characteristic is proposed to quantify the distortion in video for quality prediction. However, the NVS characteristics are derived from the image quality assessment (IQA) without considering the features of the video. X.l.Li *et al.* propose a VQA model based on the spatio-temporal NVS in 3D discrete cosine transform (3D-DCT) domain [12], which explores spatial and temporal information jointly. As a result, the 3D-DCT based method is much more accurate than V-CORNIA [13], which is computed by combining frame quality scores via a hysteresis temporal pooling method. The above blind VQA methods have high consistency with the subjective rankings. However, they require training on distorted videos and human opinion scores of them, which results in time-consuming and can't meet the real-time supervision of the video quality. Therefore, A. Mittal *et al.* propose a train-free blind model called the video intrinsic integrity and distortion evaluation oracle (VIIDEO) by probing into the sub-band correlation [14], which is proved to be generalizable for all distortion types and will be applied in real-time monitoring.

In this paper, we propose a new blind video quality assessment by considering the characteristics of the human visual system (HVS). Firstly, the video content is disassembled into predicted part and uncertain part based on ST-IGM. Then, the blind video quality of each disassembled part would be assessed separately using the proposed natural video statistics model. Thirdly, the two parts are integrated to yield the overall video quality. The rest of this paper is organized as follows. Section 2 describes the details of the proposed method. Section 3 presents the experimental results. Section 4 con-

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

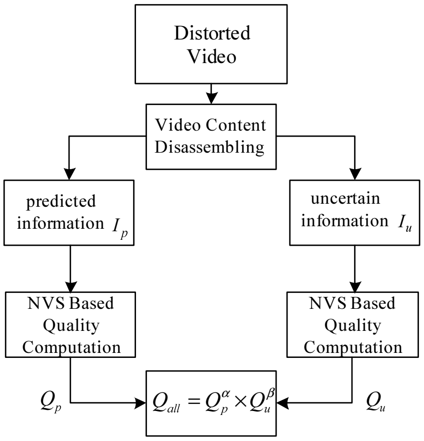
Distorted Video
Video Content Disassembling
predicted information I_p
uncertain information I_u
NVS Based Quality Computation
NVS Based Quality Computation
Q_p
$Q_{all} = Q_p^{\alpha} \times Q_u^{\beta}$
Q_u

Fig. 1. Framework of the proposed approach

cludes this paper.

2. PROPOSED APPROACH

The framework of the proposed approach is shown in Fig.1. Firstly, Inspired by internal generative mechanism theory [15], the distorted video is disassembled into the predicted information and the uncertain information by spatio-temporal prediction. Then, considering that distortions on predicted information degrade the primary visual information, while distortions on uncertain information mainly change the visual comfort, the quality of each disassembled part would be evaluated by an improved NVS based blind VQA method, which does not require the use of any external knowledge other than the distorted video being quality evaluated. Finally, we combine the two evaluation results to get the overall video quality with different weights.

2.1. Video content disassembling based on ST-IGM

Free-energy theory [16] and the Bayesian brain hypothesis [17] indicate that the brain works with an internal generative mechanism (IGM) for visual information perception and understanding. IGM can actively predict the visual scene (predicted information) and avoid the residual uncertainty (uncertain information) [16] [17]. In [15], Authors use IGM to assist image quality metric, which performs consistently with the state-of-the-art image quality assessment (IQA) metrics. In the paper, we proposed to apply spatio-temporal IGM (ST-IGM) into blind VQA. We extend a Bayesian prediction model adopted in [15] to decompose the distorted video into predicted part and uncertain part by adding temporal information. Fig.2 shows spatio-temporal prediction for Bayesian prediction model, where the circles are used to predict the central pixels, which are denoted as solid dots. To optimize the input scene, the Bayesian brain system tries to maximize the conditional function $p(x/\chi_1, \chi_2)$ between the central pixel x and χ_1, χ_2 , where $\chi_1 = \{x_1, x_2, ..., x_N\}$ are pixels surrounds x in frame t, $\chi_2 = \{x_1^{'}, x_2^{'}, ..., x_N^{'}\}$ are pixels surrounds $x^{'}$ in frame $t-1$, and x, $x^{'}$ are in same location of the frame. Hence, we apply a Bayesian prediction based autoregressive (AR) model for video content inference based on ST-IGM, which is defined as follows:

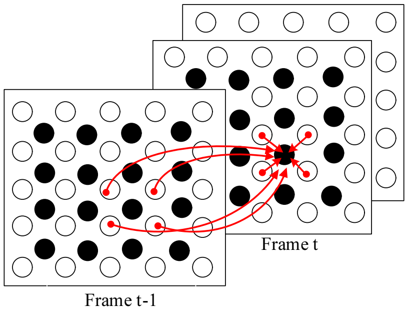

Fig. 2. Spatio-temporal prediction for Bayesian prediction model

$$x_p = l_1 \sum_{i \in \chi_1} m_i x_i + l_2 \sum_{j \in \chi_2} n_j x_j^{'} \quad (1)$$

where x_p is the prediction of pixel x, $m_i = \frac{F(x;x_i)}{\sum_k F(x;x_k)}$, $n_j = \frac{F(x;x_j^{'})}{\sum_k F(x;x_k^{'})}$, $F(x;x_i)$ and $F(x;x_j^{'})$ denote the mutual information between x and χ_1,χ_2 respectively. l_1, l_2 show different importance of spatial and temporal prediction, and $l_1 = \frac{\sum_k F(x;x_k)}{\sum_k F(x;x_k) + \sum_k F(x;x_k^{'})}$, $l_1 + l_2 = 1$.

The predicted information I_p of every frame is extracted from (1). The uncertain information of the video is denoted as:

$$I_u = I - I_p \quad (2)$$

As a consequence, the video content is disassembled into the predicted information I_p and the uncertain information I_u. Distortions in different parts of video content show different perception, where the distortions in the predicted part affect the comprehension of the content and the distortions in the uncertain part affect the comfort. The quality of these two parts is computed in the next section respectively.

2.2. Improved NVS based blind video quality assessment

Inspired by [14], we extract the frame difference to capture the temporal distortion and establish a natural video statistics (NVS) model. This NVS based blind VQA model is on the premise that the subband correlations are higher for pristine videos than distorted videos.

We defined ΔI^t as the frame difference of the adjacent frames in a block of $M \times N$:

$$\Delta I^t = I^t - I^{t-1} \quad \forall t \in \{0, 1, 2...T\} \quad (3)$$

where T is the number of frames.

We normalize the frame difference referring to the natural scene statistics (NSS) model [18]:

$$\Delta \widehat{I}^t(i,j) = \frac{\Delta I^t(i,j) - \mu^t(i,j)}{\sigma^t(i,j) + C} \tag{4}$$

where

$$\mu^t(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} \Delta I^t(i+k, j+l) \tag{5}$$

and

$$\sigma^t(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} [\Delta I^t(i+k, j+l) - \mu^t(i,j)]^2} \tag{6}$$

where $\omega_{k,l}$ is a Gaussian weighting function, $K = L = 3$.

According to the experiment [14], we find that the normalized frame differences do not always follow the generalized Gaussian distribution (GGD) perfectly, which have aberration and bad convergence. Furthermore, the distributions of four directions that used in the statistics models [14] have little difference with each other, which may improve precision slightly but increase computational complexity remarkably. Consequently, we propose to use the gradient to rectify the distribution of the frame difference, which is faster and follows GGD perfectly compared to [14].

The gradient magnitude of the normalized frame difference $|\nabla(\Delta \widehat{I}^t)|$ is defined as follows [19]:

$$|\nabla(\Delta \widehat{I}^t)| = \sqrt{(\partial_x(\Delta \widehat{I}^t))^2 + (\partial_y(\Delta \widehat{I}^t))^2} \tag{7}$$

where $\partial_x(\Delta \widehat{I}^t)$ and $\partial_y(\Delta \widehat{I}^t)$ are directional derivatives along two orthogonal directions respectively.

The gradient magnitude is used to rectify the distribution of the normalized frame difference $\Delta \widehat{I}^t$ as follows:

$$\Delta \widehat{I}^t_g = \Delta \widehat{I}^t * |\nabla(\Delta \widehat{I}^t)| \tag{8}$$

The rectified normalized frame difference $\Delta \widehat{I}^t_g$ follows zero mode generalized Gaussian distribution (GGD) [20]:

$$f(x; \alpha, \beta) = \frac{\alpha\beta}{2\Gamma(1/\beta)} \exp(-(\alpha|x|)^\beta) \tag{9}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, $\alpha = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}}$, β is the shape parameter, and α is the standard deviation.

The frame difference of the video is divided into $P \times Q$ pieces, and every piece is described by two parameters α, β, which are denoted as $\lambda^t_n = \{\alpha^t_n, \beta^t_n\}$, $n \in \{1, 2, ...P \times Q\}$. All the statistics features of the frame differences are extracted and expressed as $\Lambda^t_{P\times Q} = \{\lambda^t_1, \lambda^t_2, ..., \lambda^t_{P\times Q}\}$. To explore the correlation between the NVS statistics characteristics in multiscale, we employ a low pass filter on the frame difference:

$$\Delta J^t(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} \Delta I^t(i+k, j+l) \tag{10}$$

where $\omega_{k,l}$ is a Gaussian weighting function, $K = L = 3$. The filtered signal ΔJ^t is normalized like the former frame difference ΔI^t and rectified by the gradient magnitude similarly. This processed signal also follows the GGD and the statistics features are extracted as $\Upsilon^t_{P\times Q} = \{\upsilon^t_1, \upsilon^t_2, ..., \upsilon^t_{P\times Q}\}^T$, the variation of the statistics characteristics are modeled as:

$$\Delta\Lambda^t_{P\times Q} = \Lambda^t_{P\times Q} - \Lambda^{t-1}_{P\times Q} \tag{11}$$

$$\Delta\Upsilon^t_{P\times Q} = \Upsilon^t_{P\times Q} - \Upsilon^{t-1}_{P\times Q} \tag{12}$$

where $t \in \{1, 2, ..., T\}$. Denote $\Delta\Lambda^t_{P\times Q}$ and $\Delta\Upsilon^t_{P\times Q}$ as:

$$\Delta\Lambda^t_{P\times Q} = \{\overrightarrow{\alpha^t_\Lambda}, \overrightarrow{\beta^t_\Lambda}\} \tag{13}$$

$$\Delta\Upsilon^t_{P\times Q} = \{\overrightarrow{\alpha^t_\Upsilon}, \overrightarrow{\beta^t_\Upsilon}\} \tag{14}$$

The sub-band correlation is denoted as follows:

$$\begin{aligned}
&\rho^t = corr(\Delta\Lambda^t_{P\times Q}, \Delta\Upsilon^t_{P\times Q}) \\
&= mean\{corr(\overrightarrow{\alpha^t_\Lambda}, \overrightarrow{\alpha^t_\Upsilon}), corr(\overrightarrow{\beta^t_\Lambda}, \overrightarrow{\beta^t_\Upsilon})\} \\
&= mean\{ \\
&\frac{P\times Q \sum_{P\times Q} \overrightarrow{\alpha^t_\Lambda}\overrightarrow{\alpha^t_\Upsilon} - \sum_{P\times Q} \overrightarrow{\alpha^t_\Lambda} \sum_{P\times Q} \overrightarrow{\alpha^t_\Upsilon}}{\sqrt{P\times Q \sum_{P\times Q} (\overrightarrow{\alpha^t_\Lambda})^2 - (\sum_{P\times Q} \overrightarrow{\alpha^t_\Lambda})^2} \sqrt{P\times Q \sum_{P\times Q} (\overrightarrow{\alpha^t_\Upsilon})^2 - (\sum_{P\times Q} \overrightarrow{\alpha^t_\Upsilon})^2}}, \\
&\frac{P\times Q \sum_{P\times Q} \overrightarrow{\beta^t_\Lambda}\overrightarrow{\beta^t_\Upsilon} - \sum_{P\times Q} \overrightarrow{\beta^t_\Lambda} \sum_{P\times Q} \overrightarrow{\beta^t_\Upsilon}}{\sqrt{P\times Q \sum_{P\times Q} (\overrightarrow{\beta^t_\Lambda})^2 - (\sum_{P\times Q} \overrightarrow{\beta^t_\Lambda})^2} \sqrt{P\times Q \sum_{P\times Q} (\overrightarrow{\beta^t_\Upsilon})^2 - (\sum_{P\times Q} \overrightarrow{\beta^t_\Upsilon})^2}}\}
\end{aligned} \tag{15}$$

For the overall quality of the video, a span of time S is chosen for temporal pooling:

$$Q = \frac{1}{S} \sum_S \rho^t \tag{16}$$

2.3. Overall quality pooling

The overall quality is a combination of the predicted part and the uncertain part that computed by the NVS based blind VQA in (16).

$$Q_{all} = Q_p^\alpha \times Q_u^\beta \tag{17}$$

where Q_p is the quality of the predicted part and Q_u is the quality of the uncertain part, the parameter α and β are used to adjust the relative importance of two parts, $\alpha + \beta = 1$.

3. EXPERIMENTS

We evaluate the accuracy of our approach and compare the results with other state-of-the-art VQA models, which are tested on the LIVE Video Quality Assessment (VQA) Database [21]. Ten kinds of pristine videos are available with the resolution of 768×432. For each kind of pristine video, 15 distorted videos are presented for four different distortions including

Table 1. SROCC of different VQA methods against DMOS on LIVE VQA database (W: wireless, I: IP, H: H.264, M: MPEG)

Method	W	I	H	M	all
MS-SSIM	0.729	0.653	0.731	0.668	0.736
MOVIE	**0.811**	0.716	0.766	0.773	**0.789**
VQM	0.721	0.638	0.652	**0.781**	0.702
STRRED	0.786	**0.771**	**0.820**	0.720	**0.802**
VIIDEO	0.532	0.612	0.674	0.556	0.624
proposed	**0.817**	**0.798**	**0.911**	**0.921**	0.779

Table 2. LCC of different VQA methods against DMOS on LIVE VQA database (W: wireless, I: IP, H: H.264, M: M-PEG)

Method	W	I	H	M	all
MS-SSIM	0.718	0.776	0.742	0.622	0.747
MOVIE	**0.842**	0.766	0.814	0.798	**0.813**
VQM	0.755	0.667	0.666	**0.813**	0.730
STRRED	0.806	**0.815**	**0.823**	0.758	**0.812**
VIIDEO	0.625	0.737	0.692	0.631	0.651
proposed	**0.822**	**0.800**	**0.918**	**0.895**	0.804

MPEG-2 distortions, H.264 distortions, wireless distortions and IP distortions. Subjective quality is provided for every video as difference mean opinion score (DMOS).

In our experiment, the block size $M \times N$ is 36×36, P and Q are set to 22, 12 respectively. We use Spearman's rank ordered correlation coefficient (SROCC) and Pearson's linear correlation coefficient (LCC) to test the performance of VQA algorithms. A better VQA algorithm should have both higher LCC and SROCC. The following nonlinear function is applied to map the quality scores to subject DMOS before calculating LCC:

$$Q(x) = \beta_1(1/2 - \frac{1}{1 + exp(\beta_2(x - \beta_3))}) + \beta_4 x + \beta_5 \tag{18}$$

To verify the performance of our VQA method, the proposed metric is compared with other five metrics, including MS-SSIM [22], MOVIE [4], VQM [23], STRRED [5] and VIIDEO [14]. MS-SSIM, MOVIE, and VQM are FR-VQA algorithms, STRRED is a RR-VQA algorithm and VIIDEO is a blind VQA algorithm. Table1 and Table2 show SROCC and LCC comparison results, respectively, where two best algorithms have been highlighted in boldface. From Table1 and Table2, it is obvious that our proposed model outperforms the FR-VQA algorithms MS-SSIM and VQM, as these methods are based on the image quality assessment, which couldn't catch the temporal distortions. Nonetheless, our proposed method is inferior to MOVIE and STRRED. The motion-based MOVIE correlates highly with subjective quality due to its availability of pristine videos while our proposed method is completely reference free. VIIDEO is also a blind VQA method, which is an innovative train-free method. However, it can't perform better than our proposed metric as it ignores the HVS characteristics. Therefore, the proposed metric is highly consistent with the human perception owing to considering human vision characteristics. It is also shown from Table1 and Table2 that the proposed method shows incredible correlation with subjective results for H.264 and MPEG distortions, which means that our method is more sensitive to the compression distortions.

Table 3. SROCC of different train-based VQA methods and the proposed method (8 folds of videos are used for training) against DMOS on LIVE VQA database (W: wireless, I: IP, H: H.264, M: MPEG)

Method	W	I	H	M	all
V-BLIINDS	0.815	0.779	0.839	0.869	0.759
V-CORNIA	0.595	0.714	0.857	0.929	0.740
[12]	0.815	**0.826**	**0.912**	**0.961**	**0.782**
proposed	**0.817**	0.798	0.911	0.921	0.779

Table 3 compares performance of the proposed method and the train-based methods including V-BLIINDS [24], V-CORNIA [13] and [12]. The 80% of the database are chosen for training and the remaining 20% are for testing that no overlap occurs between train and test. There are 45 trials performed to generate the SROCC, among which the medians are recorded in Table 3. From Table3, we can see that the proposed method outperforms V-BLIINDS [24] and V-CORNIA [13] as it not only catches the temporal distortions but also considers the characteristics of HVS. The method proposed in [12] performs better than our method as a set of effective statistics features are trained for video visual quality prediction after all.

4. CONCLUSION

In this paper, we present a new blind VQA method based on the ST-IGM. As distortions in different parts of the video have different effect on visual perception, we disassemble the video into the predicted part and the uncertain part. For each part, the quality is gauged by an improved NVS based blind VQA method, which is reference free and distortion-unaware. Lastly, the overall quality is combined by these two parts. Experimental results show that the proposed method outperforms many existing VQA methods and its computational complexity is lower than the existing blind VQA methods, which meets the requirement of real-time monitoring of the net-work videos.

5. ACKNOWLEDGEMENT

This work was supported by Natural Science Foundation of China under Grant No. 61671283, 61301113, and Shanghai National Natural Science under Grant No. 13ZR14165.

6. REFERENCES

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[2] E. C. Larson and D. M. Chandler, “Most apparent distortion: Full-reference image quality assessment and the role of strategy,” J. Electron. Imaging, vol. 19, no. 1, pp. 1-21, Mar. 2010.

[3] P. V. Vu, C. T. Vu, and D. M. Chandler, “A spatiotemporal most apparent distortion model for video quality assessment,” Proc. IEEE ICIP, pp. 2505-2508, Sep. 2011.

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

[5] R. Soundararajan and A. C. Bovik, “Video quality assessment by reduced reference spatio-temporal entropic differencing,” IEEE Transactions in Circuits, Systems and Video Technology, vol.23, no. 4, pp.684-694, 2012.

[6] L. McLaughlin, S. S. Hemami, “Reduced-reference quality assessment with scalable overhead for video with packet loss,” Image Processing (ICIP), pp. 1622-1626, Sept. 2013.

[7] L. Ma, K. N. Ngan, L. Xu, “Reduced reference video quality assessment based on spatial HVS mutual masking and temporal motion estimation,” IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, 2013.

[8] Z. Wang, A. C. Bovik, and B. L. Evans, “Blind measurement of blocking artifacts in images,” in Proc. IEEE Int. Conf. Image Process., vol. 3. pp. 981-984. Sep. 2000.

[9] L. Anegekuh, L. F. Sun, E. Jammeh, “Content-based video quality prediction for HEVC encoded video streamed over packet networks,” IEEE Transactions on Multimedia, Volume: 17, Issue: 8, pp.1323-1334, Aug. 2015.

[10] H. G. Kim, Y. M. Ro, “Measurement of critical temporal inconsistency for quality assessment of synthesized video,” Image Processing (ICIP), pp. 2381-2386, August 2016.

[11] M. A. Saad, A. C. Bovik, and Christophe Charrier, “Blind Prediction of Natural Video Quality,” IEEE Transactions on Image Processing, pp.1352-1364, vol. 23, no. 3, March 2014.

[12] X. L. Li, Q. Guo and X. Q. Lu, “Spatiotemporal Statistics for Video Quality Assessment,” IEEE Transactions on Image Processing, pp.3329-3341, July 2016.

[13] J. Xu, P. Ye, Y. Liu, and D. Doermann, “No-reference video quality assessment via feature learning,” in Proc. IEEE Int. Conf. Image Process. pp. 491-495. Oct. 2014.

[14] A. Mittal, M. A. Saad, and A. C. Bovik, “A Completely Blind Video Integrity Oracle,” IEEE Transactions on Image Processing, pp.289-300. January 2016.

[15] J. J. Wu, W. S. Lin, G. M. Shi, and A. Liu, “Perceptual Quality Metric with Internal Generative Mechanism,” IEEE Transactions on Image Processing, pp. 43-54, August 2012.

[16] K. Friston, “The free-energy principle: a unified brain theory?” Nat Rev Neurosci, vol. 11, no. 2, pp. 127-138, Feb. 2010.

[17] D. C. Knill and R. Pouget, “The bayesian brain: the role of uncertainty in neural coding and computation,” Trends In Neuroscience, vol. 27, pp.712-719, 2004.

[18] L.P.Yang, H.Q.Du, J.T.Xu,and Y.Liu, “Blind Image Quality Assessment on Authentically Distortion Images With Perceptual Features, ” Image Processing (ICIP), pp.2042-2046, Sept 2016.

[19] Q. H. Li, W. S. Lin, and Y. M. Fang, “No-Reference Quality Assessment for Multiply-Distorted Images in Gradient Domain,” IEEE Signal Processing Letters, pp.541-545, April 2016.

[20] K. Sharifi and A. Leon-Garcia, “Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video,” IEEE Trans. Circuits and Systems for Video Technology, pp. 52-56, 1995.

[21] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, “Study of subjective and objective quality assessment of video,” IEEE Transactions in Image Process., vol. 19, no. 6, pp. 1427-1441, June 2010.

[22] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in Proc. 37th Asilomar Conf. Signals, Syst. Comput., vol. 2. pp. 1398-1402. Nov. 2003.

[23] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” IEEE Trans. Broadcast., vol. 10, no. 3, pp. 312-322, Sep. 2004.

[24] M. A. Saad and A. C. Bovik, “Blind quality assessment of natural videos using motion coherency,” in IEEE Asilomar Conference on Signals, Systems, and Computers, pp.332-336, November 2012.