VISUAL QUALITY METRIC FOR PERCEPTUAL VIDEO CODING

Long Xu[1,2], Lin Ma[3], King Ngi Ngan[4], Weisi Lin[1], and Ying Weng[5]
[1]School of Computer Engineering, Nanyang Technological University, Singapore
[2]School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China
[3]Huawei Noah's Ark Lab, Hong Kong
[4]Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong
[5]School of Computer Science, Bangor University, Bangor, United Kingdom
Emails: {lma, knngan}@ee.cuhk.edu.hk, {xulong, wslin}@ntu.edu.sg, y.weng@bangor.ac.uk

ABSTRACT

The visual quality assessment (VQA) becomes prevailing in the studies of image and video coding. It assesses the quality of image or video more accurately than mean square error (MSE) with respect to the human visual system (HVS). Toward perceptual video coding, MSE is weighted spatially and temporally to simulate the HVS response to visual signal in this paper. Firstly, the image content is depicted by edge strength to compose spatial weighting factors. Secondly, the motion strength calculated from motion vector of each block gives temporal weighting factors. Thirdly, the motion trajectory based saliency map for video signal is integrated as another weighting factor of MSE. The proposed VQM not only efficiently model HVS but also relate to quantization parameter (QP) capable of guiding perceptual video coding. A perceptual rate distortion optimization (RDO) is established on the proposed VQM. The experimental results indicate that the proposed VQM is consistent well with HVS. In addition, the better rate-distortion efficiency and accurate bit rate control can be achieved by the proposed visual quality control algorithm.

Index Terms— Visual quality assessment, perceptual rate distortion optimization, rate control, saliency

1. INTRODUCTION

In recent years, there has been an increasing interest in visual quality assessment that measures the perceptual visual quality of images and videos. Since human eyes are the ultimate receivers of the visual signal, the most accurate way of assessing image/video quality is subjective evaluation methods. The subjective experiment is actually fundamental to rank a proposed objective metric. However it is much labor consuming and time consuming. Thus, most researchers do their visual quality assessments by using the published subjective databases instead of doing their own subjective experiments. The available databases of subjective scores and test material were reported in [1] [2] for quality degradation of compression and error-prone channels. Moreover, researchers usually need to perform the subjective experiments to validate the objective visual quality metrics (VQMs) and compete with benchmarking metrics. The correlation between the subjective scores obtained by subjective experiments and objective metric outputs is analyzed to determine whether the developed objective metric is good at measuring the human perception of image/video quality. The measurements of the correlation include linear correlation coefficient (LCC), Spearman's rank order correlation coefficient (SROCC), and root mean square prediction error (RMSE).

The objective quality metric aims at automatically predicting human perceptual behavior in evaluating image/video quality. It is convenient and computationally efficient in the real-world applications. Traditionally, mean squared error (MSE) / peak signal to noise ratio (PSNR) are used to evaluate image and video qualities. Almost all image/video compression standards use MSE/PSNR to measure the visual quality of the compressed signal. However, MSE/PSNR does not represent well the response of human visual system (HVS) [3]. Thus, a host of image/video quality metrics have been proposed to replace MSE/PSNR in image/video processing in the last decade. Moving pictures quality metric (MPQM) [4], perceptual distortion metric (PDM) [5] and the Sarnoff JND [6] vision model simulate the tuning properties of HVS from visual psychology point of view. Since the HVS is highly adapted to extract structural information from the scene, structural similarity index (SSIM) was defined to measure structural similarity to approximate the perceived quality [7].

To provide an accurate VQM for video compression, Y. Ou *et al.* used an inverted falling exponential function to correlate frame rate with temporal visual quality, and a sigmoid function to correlate PSNR with spatial visual quality [8]. This metric took into account both frame rate and quantization artifact in video coding. A. Bhat *et al.* included quantization effect of video compression and spatial texture content, such as edge of object, human's skin into visual quality assessment [9]. SSIM was extended into spatio-temporal domain by introducing motion perception of human's eyes in [10]. In this paper, a new objective video quality metric is proposed for the purpose of rate distortion optimization (RDO). It concerns both modeling HVS response to video signal and building relation between perceptual visual quality measurement and quantization parameter (QP) of video compression. The rest of this paper is organized as follows. In Section II, a new objective VQM

for quality assessment is proposed. Section III illustrates the proposed perceptual RDO as well as the experimental results of perceptual RDO based video coding. Finally, a brief conclusion is given in the last section.

2. VISUAL QUALITY METRIC TOWARDS PERCEPTUAL VIDEO CODING

Nowadays many visual quality metrics have been designed for evaluating image/video perceptual quality [3]-[17] as well as guiding the applications, such as image/video coding [11], quality monitoring [14], and so on. As discussed in [18], in order to be incorporated into the corresponding applications, the developed metric needs to be easily optimized, which can help to obtain closed-form analytic solution. In this section, a simple visual quality metric targeting perceptual video coding is proposed, which is illustrated in Fig. 1.

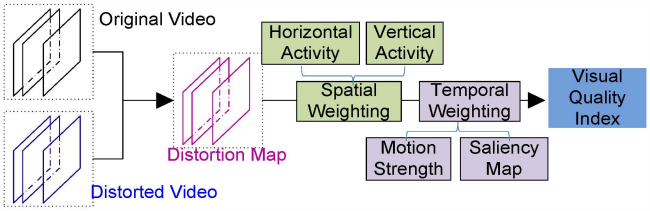

Fig. 1. Framework of the proposed visual quality metric

2.1. Visual quality metric

In order to ensure the friendliness of the quality metric for image/video coding application, mean square error (MSE) is simply weighted to simulate the HVS responses to the visual signals [9] [18]. In this paper, the weighted MSE is also employed for not only modeling the HVS property, but also relating to the QP for guiding the perceptual video coding, which is simply defined as:

$$Q_k = \omega_k \times MSE_k \tag{1}$$

where ω_k is the weighting parameter designed for the k-th macroblock (MB), Q_k is the corresponding quality value. By averaging the Q_k values over all the image and frame, the visual quality index can be obtained.

In order to accurately model the HVS property to ensure a good performance of the visual quality metric, the components comprising the weighting factor ω_k need to be carefully designed. In this paper, ω_k is researched from the spatial and temporal perspectives, respectively. Moreover, as we target at perceptual video coding, the aforementioned components for the weighting process should be block-size developed, which is convenient for the block-based intra prediction, inter motion estimation, transform, entropy coding, and so on.

2.2. Weighting from spatial perspective

From the spatial perspective, we first need to check the spatial content of the video frame. As discussed in [11] [20], different image content present different masking properties for HVS perception. Specifically, texture block can tolerate more distortions than the plain and edge block. Therefore, the spatial content of the image is depicted to compose the weighting factor, which is determined by:

$$SP_{\omega_k} = \sum_{(m,n)} \sqrt[2]{\alpha(V_k^2(m,n) + H_k^2(m,n)) + \beta} \tag{2}$$

where $\alpha = 0.5$ and $\beta = 1$ are two parameters to control the weighting level, (m,n) denotes the pixel number of the k-th MB, V and H denote the vertical and horizontal content activity of each frame, respectively, which are measured by derivatives of luminance component as [21]:

$$\begin{aligned} V_k(m,n) &= I_k(m,n) - I_k(m-2,n) \\ H_k(m,n) &= I_k(m,n) - I_k(m,n-2) \end{aligned} \tag{3}$$

From the definitions of V and H, we can see that the larger the value, the more complicated the corresponding MB is, which will generate higher HVS masking effect.

2.3. Weighting from temporal perspective

For the temporal information, the most important component is the motion vectors. It is argued that HVS is not sensitive to very fast or very low motion. The motion in medium level is most attractive. Therefore, the motion strength is first defined as:

$$MS_k = 0.5 \times \sqrt[2]{MVx_k^2 + MVy_k^2} \tag{4}$$

where MVx_k and MVy_k is the horizontal and vertical motion vectors of the k-th MB. In order to make sure that MS_i lies in the same range between different frames, MS_k is weighted by the largest MS value within the same frame. Furthermore, temporal weighting factor is defined as:

$$\begin{aligned} MS_{\omega_k} = \gamma \times (MS_k \geq 1.5 \& MS_k \leq 0.5) + \delta \times \\ (MS_k < 1.5 \& MS_k > 0.5) \end{aligned} \tag{5}$$

where $\gamma = 0.8$ and $\delta = 1.0$ to control the HVS sensitivities to different motion strengths.

Furthermore, based on the motion vectors MVx_i and MVy_i, prediction error PE_i after motion estimation (ME), and the content information I_i, a saliency map can be obtained by the quaternion Fourier transform (QFT) [22]:

$$Q_i = \zeta_{QFT}(I_i, MVx_i, MVy_i, PE_i) \tag{6}$$

where ζ_{QFT} is the QFT. The frequency response Q_i can be represented in polar form:

$$Q_i = \|Q_i\| e^{\mu \cdot p_i} \tag{7}$$

where p_i is the phase spectrum of Q_i and μ is a unit pure quaternion. As discussed in [22], only the phase spectrum is sufficient to construct the visual saliency map. Therefore, $\|Q_i\|$ is set as 1. Then by applying the inverse QFT, the reconstructed quaternion image $\acute{q}_i$ is generated. The visual saliency map is constructed by the Gaussian filtering:

$$SA(i) = g * \|\acute{q}_i\|^2. \tag{8}$$

We can further down-scale the saliency map into MB size, which means that each MB is assigned one saliency value. By considering this property, the final quality metric of each MB is obtained by

$$Q_k = \left(\frac{MS_{\omega_k}}{SP_{\omega_k}} \times SA_{\omega_k}\right) \times MSE_k, \quad (9)$$

where SA_{ω_k} is the saliency value of the k-th MB.

3. PROPOSED PERCEPTUAL RDO

In video coding, multiple modes are provided to adapt to video content variation. The different mode represents the different partitions of a MB. For the MBs with low textured video content, the large size mode is preferred. On the contrary, the MBs with high textured video content may benefit from the small partitions of a MB. The decision of the best mode recurs to RDO. RDO tries to achieve a tradeoff between the bit cost and visual quality degradation measured by MSE usually, which is formulated as:

$$J = D + \lambda R, \quad (10)$$

where D and R represent distortion and bit cost for coding a MB. λ represents the Lagrange parameter. With the same distortion, the mode with the least bit cost would be the best mode. Equivalently, the mode with the least MSE would be the best mode under the same bit cost. The prior works have proved that MSE failed to represent the perceptual quality of the video signal. In [22][24], the authors performed RDO based SSIM metric instead of MSE.

It is known that the perceptual visual quality of low textured region decreases more significantly than that of high textured region with the same MSE increased. It means that the high textured region may have more weight of distortion. It can be realized by increasing λ. For example, assume that λ equals 1, the distortions of two modes are a and $0.9a$ and the bit costs of these two modes are $0.9a$ and a, the R-D costs of these two modes are both $1.9a$. For high textured region, the first mode with distortion of a would be selected by increasing λ from 1 to 2. Inversely, to decrease λ would result in selecting the second mode with less distortion of $0.9a$, which is better for low textured region.

Assuming the original Lagrange multiplier λ_0 calculated by

$$\lambda_0 = 0.85 \times 2^{\frac{QP-12}{3}}. \quad (11)$$

From (2), the MB of high textured has a larger edge strength SP_{ω_k}. If the value of SP_{ω_k} is larger than 1, the MB is regarded as high textured. Thus, the RDO of this MB can be improved by increasing λ_0 as

$$\lambda = SP_{\omega_k} \times \lambda_0. \quad (12)$$

Further improvement of RDO can go far by taking motion activity and saliency into account.

Besides RDO, the video coding optimization can be achieved from optimal bit allocation under the bit constraint encoding. The basic idea comes from that the same bit cost increase would benefit more quality increase to the region with low textured feature than high textured region. In addition, the HVS is not sensitive to the quality change of high textured region. To have deep insight into the statement above, a picture is drawn in Fig.2 for showing the relation between quality degradation to the bit rate saving. From Fig.2, the same quality degradation ΔD can result in more bits

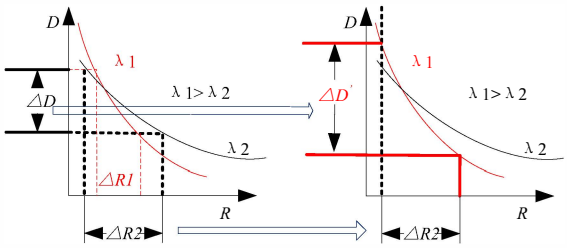

Fig. 2. Relationship between visual quality degradation and bit rate saving for different Lagrange multipliers

saving ΔR_2 for λ_2 which is less than λ_1. With such bits saving, the quality improvement of λ_1 is much better than λ_2 from the right picture of Fig. 2, i.e. $\Delta D' > \Delta D$. Such a fact holds for both region of video frame and whole video frame. In [26], the authors have exhibited that the same bits cost would produce more quality improvement to the frames with low texture complexity.

Most of the perceptual rate controls have achieved the R-D improvements by revising the QP of MB based on the similar idea mentioned above. In [19], Xu *et al.* have proposed a consistent visual quality control algorithm based on a window-level rate control scheme. The QP regulation scheme of frames and MBs in [19] conformed to the analyses above. However, the RDO process in [19] was not optimized. In this paper, the perceptual video coding concept is enhanced against [19] by introducing perceptual RDO additionally.

The rate control scheme in this work employs Xu's algorithm in [19]. Assuming that the distortion model of residual signal is $QP^2/12$, (9) can be rewritten as

$$Q_k = \left(\frac{MS_{\omega_k}}{SP_{\omega_k}} \times SA_k\right) \times {QP_k}^2. \quad (13)$$

It can be observed that there exists a quadratic relationship between Q_k and QP_k. For consistent Q_k among MBs, the QP of each MB is revised by

$$\begin{cases} QP_k = p_k \times QP_f \\ p_k = 1/\sqrt{\left(\left(MS_{\omega_k}/SP_{\omega_k}\right) \times SA_k\right)} \end{cases}, \quad (14)$$

where QP value of each frame denoted by QP_f is derived from the rate control algorithm in [19].

4. EXPERIMENTAL RESULTS

The proposed metric is evaluated on the LIVE subjective video database [2]. LCC, SROCC, and RMSE, are employed to illustrate the effectiveness of the corresponding quality metric. We compare the proposed quality metric with PSNR, SSIM, MSSIM, VSNR VIF, Xu's VQM [19] and MOVIE. The experimental results are illustrated in Table. 1. It can be observed that PSNR performs poorly, because it is not related to the HVS perception. Also VSNR performs badly, which

can be attributed to two reasons. The first is that VSNR analyzes the HVS perception of the distortion in the wavelet domain. But the MPEG-2 and H.264 compression schemes introduce the distortions during quantization process in DCT domain. The second one is that VSNR is an image quality metric designed to capture the spatial distortions. For video quality assessment, the temporal information accounts for significantly. This is also the reason why SSIM, MSSIM and VIF perform successfully in image quality evaluation, but not so well on the video quality assessment. It is the main reason why the proposed method outperforms PSNR, VSNR, SSIM, MSSIM and VIF. The proposed metric improves the LCC of 0.03 against [19] by introducing spatio-temporal saliency additionally. Although the proposed method is slightly poorer than MOVIE, MOVIE is of great complexity. In order to find a clear relationship between QP and the proposed VQM, the performance is compromised a little. Therefore, the proposed VQM can be easily integrated into video coding to guide perceptual video coding.

Table. 1. Performance comparisons between different VQAs

(a) Linear Correlation Coefficient

Algorithm	Wireless	IP	H.264	MPEG-2	All Data
PSNR	0.677	0.478	0.589	0.409	0.569
SSIM	0.473	0.537	0.611	0.582	0.503
MSSIM	0.684	0.684	0.692	0.632	0.676
VSNR	0.680	0.737	0.614	0.507	0.688
VIF	0.593	0.636	0.649	0.673	0.577
MOVIE	0.855	0.798	0.853	0.806	**0.829**
Xu's [19]	0.762	0.735	0.709	0.556	0.741
Proposed	0.772	0.736	0.729	0.626	0.771

(b) Spearman Rank Order Correlation Coefficient

Algorithm	Wireless	IP	H.264	MPEG-2	All Data
PSNR	0.671	0.430	0.477	0.394	0.553
SSIM	0.539	0.474	0.659	0.569	0.533
MSSIM	0.729	0.645	0.734	0.681	0.735
VSNR	0.694	0.693	0.641	0.587	0.672
VIF	0.538	0.553	0.638	0.635	0.558
MOVIE	0.834	0.735	0.837	0.739	**0.804**
Xu's [19]	0.753	0.724	0.664	0.564	0.721
Proposed	0.763	0.735	0.674	0.613	0.748

We also integrate the proposed perceptual RDO into JM14.0 reference software with the configurations: Profile/Level: 100/40; Reference frames: 2; Full search; Search range: 32; RDO: on and CABAC; IPPP encoding structure. The rate control is enabled to compete R-D efficiency on same target bit rates. The PSNR, VQM and bit control accuracy are listed in Table 2 for the proposed perceptual RDO based rate control and Xu's visual quality control algorithm in [19]. From Table 2, the obvious R-D improvement can be observed for the proposed algorithm with 1.58% VQM gain and about 0.2dB PSNR gain. In addition, the proposed rate control algorithm is superior to the benchmark in terms of bit control accuracy. Regarding computational complexity, for a 250-frame 768x432 sequence on a 3G Hz quad-core CPU with 6G RAM, the computing times of PSNR, SSIM, MSSIM, VSNR, VIF, NTIA-VQM, MOVIE and the proposed metric are 4, 24, 60, 26, 636, 57, 6320 and 43 seconds respectively. The proposed metric is superior to MSSIM, VIF, NTIA-VQM and MOVIE with respect to computing time.

5. CONCLUSIONS

In this paper, we first proposed a new metric for visual quality assessment of video signal. It is proved to be with competitive performance comparing with the state-of-the-art visual quality metrics of videos. Secondly, a perceptual RDO is proposed based on the proposed visual quality metric. The significant R-D improvement exhibits the advantage of the proposed perceptual RDO. The further study will go far on the theoretical and analytical foundation of optimized video coding on the basis of the proposed perceptual video metric.

Table. 2. Performance comparisons in terms of bit control error and PSNR/VQM gain

Sequence	Target bit rate (kbps)	Traditional rate control				Proposed rate control			
		Bit rate (kbps)	PSNR	VQM	Error (%)	Bit rate (kbps)	PSNR	VQM	Error (%)
Foreman	*2000*	1996.98	42.18	0.985	-0.15	1997.44	42.35	0.989	-0.15
	1000	1000.56	39.31	0.971	0.06	1001.59	39.32	0.985	0.06
	500	500.41	36.62	0.949	0.08	500.94	36.70	0.965	0.08
	300	300.97	34.70	0.925	0.32	302.78	34.82	0.943	0.32
News	*2000*	1998.33	48.10	0.997	-0.08	1998.74	48.17	0.998	-0.06
	1000	999.91	44.94	0.991	-0.01	1001.55	45.02	0.995	0.15
	500	499.9	41.90	0.982	-0.02	500.84	41.97	0.990	0.17
	300	300.17	39.49	0.973	0.06	300.44	39.54	0.983	0.15
Silent	*2000*	2000.39	45.92	0.990	0.02	2001.86	46.01	0.998	0.09
	1000	999.78	42.20	0.966	-0.02	1000.00	42.28	0.970	0.00
	500	499.72	38.46	0.912	-0.06	500.11	38.47	0.929	0.02
	300	299.97	35.96	0.865	-0.01	300.23	35.96	0.879	0.08
Tennis	*2000*	1999.83	42.16	0.983	-0.08	2001.79	42.43	0.987	-0.08
	1000	1001.03	38.72	0.967	-0.01	1002.11	38.86	0.968	-0.01
	500	501.32	35.40	0.918	-0.02	503.06	35.42	0.933	-0.02
	300	299.79	33.31	0.875	0.06	301.05	33.44	0.894	0.06
Average			**39.96**	**0.953**	**0.03**		**40.05**	**0.964**	**0.05**
Night	*10000*	9992.06	39.29	0.980	-0.08	9993.23	39.42	0.988	-0.08
	8000	7994.46	38.61	0.975	-0.07	7995.11	38.79	0.982	-0.07
	5000	4995.81	37.20	0.961	-0.08	4996.76	37.31	0.973	-0.08
	2000	2000.69	34.01	0.913	0.03	2000.90	34.22	0.937	0.03
Crew	*10000*	9996.95	41.39	0.962	-0.03	9997.15	41.49	0.973	-0.03
	8000	7995.43	40.84	0.958	-0.06	7996.05	40.99	0.969	-0.06
	5000	4996.46	39.77	0.948	-0.07	4998.23	40.07	0.951	-0.07
	2000	1994.48	37.30	0.917	-0.28	1995.10	37.58	0.934	-0.28
Harbour	*10000*	9997.14	37.41	0.970	-0.03	9997.60	37.62	0.984	-0.03
	8000	7997.97	36.53	0.966	-0.03	7999.30	36.66	0.969	-0.03
	5000	4999.71	34.70	0.955	-0.01	5000.90	34.92	0.961	-0.01
	2000	1996.56	31.28	0.917	-0.17	1997.18	31.47	0.924	-0.17
Average			**37.36**	**0.952**	**-0.07**		**37.55**	**0.967**	**-0.07**

6. ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China under Grants 61202242, 61301090 and 61363054, and Singapore National Research Foundation under its IDM Futures Funding Initiative and administered by

the Interactive & Digital Media Programme Office, Media Development Authority.

7. REFERENCES

[1] H. R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "Live image quality assessment database release 2," Available [Online]: http://live.ece.utexas.edu/research/quality.

[2] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "LIVE Video Quality Database," Available [Online]: http://live.ece.utexas.edu/research/quality/live_video.html.

[3] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? - A new look at signal fidelity measures," IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[4] C. van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatio-temporal model of the human visual system," in Proc. SPIE, vol. 2668, pp. 450–461, 1996.

[5] S. Winkler, "A perceptual distortion metric for digital color video," in Proc. SPIE HVEI, volume 3644, San Jose, CA, January 1999.

[6] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in Digital Images and Human Vision, A. B. Watson, Ed. The MIT Press, 1993, pp. 163–178.

[7] Z. Wang, H. R. Sheikh, A. C. Bovik and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, pp. 600-612, Apr. 2004.

[8] Y.-F. Ou, Z. Ma, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," IEEE Trans. Circuits Syst. Video Techno., vol. 20, no. 3, pp. 286-298, Mar. 2010.

[9] A. Bhat, et al., "A full reference quality metric for compressed video based on mean squared error and video content," IEEE Trans. Circuits Syst. Video Techno., vol. 22, no. 2, pp. 165-173, Feb. 2012.

[10] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," J. Opt. Soc. Am. A Opt. Image Sci. Vis., vol. 24, no. 12, pp. B61-B69, Dec. 2007.

[11] L. Ma, K. N. Ngan, F. Zhang, and S. Li, "Adaptive block-size transform based just-noticeable difference model for images/videos", Signal Process.: Image Comm., vol. 26, no. 3, pp. 162-174, Mar. 2011.

[12] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," IEEE Trans. Multimedia, vol. 13, no. 5, pp. 935-949, Oct. 2011.

[13] D. M. Chandler, and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images", IEEE Trans. Image Process., vol. 16, no. 9, pp. 2284-2298, Sep. 2007.

[14] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," IEEE Trans. Multimedia, vol. 13, no. 4, pp. 824-829, Aug. 2011.

[15] H. R. Sheikh, and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006.

[16] S. Li, L. Ma, and K. N. Ngan, "Full-reference video quality assessment by decoupling detail losses and additive impairments," IEEE Trans. Circuits Syst. Video Techno., vol. 22, no. 7, pp. 1100-1112, Jul. 2012.

[17] K. Seshadrinathan, and A. C. Bovik, "Motion tuned spatio-temporalquality assessment of natural videos," IEEE Trans. Image Process., vol. 19, no. 2, pp. 335–350, Feb. 2010.

[18] F. Zhang, L. Ma, S. Li, and K. N. Ngan, "Practical Image Quality Metric Applied to Image Coding," IEEE Trans. Multimedia, vol. 13, no. 4, pp. 615-624, Aug. 2011.

[19] L. Xu, K. N. Ngan, S. Li, and L. Ma, "Video Quality Metric for Consistent Visual Quality Control in Video Coding," APSIPA-ASC, 2012.

[20] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference video quality assessment of compressed video sequences," IEEE Trans. Circuits Syst. Video Techno., vol. 22, no. 10, pp. 1441-1456, Oct. 2012.

[21] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," CVPR 2005, vol. 1, pp.886-893 vol. 1, 25-25 June 2005.

[22] L. Ma, S. Li, and K. N. Ngan, "Motion trajectory based visual saliency for video quality assessment," in Proc. ICIP, 2011.

[23] S. Q. Wang, Rehman. A, S. W. Ma and W. Gao, "SSIM-Motivated Rate Distortion Optimization for Video Coding," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no.4, pp. 516-529, Apri. 2012.

[24] Y.-H. Huang et al., "Perceptual rate-distortion optimization using structural similarity index as quality metric," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 11, pp. 1614-1624, Nov. 2010.

[25] J.-S. Lee and T. Ebrahimi, "Perceptual video compression: A survey," IEEE Journal of Selected Topics in Signal Processing, 6(6), pp. 684-697, Oct. 2012.

[26] L. Xu, D. Zhao, X, Ji, L. Deng, S. Kwong, and W. Gao. "Window level rate control for smooth visual quality and smooth buffer occupancy", IEEE Trans. Image Process., vol. 20, no. 3, pp. 723-734, Mar. 2011.