Overall Visual Quality Optimization Coding for Region of Interest

Sishuo Ma[*], *Junhua Qu*[*], *Li Zhang*[†]
[*]North China Power Electric University, Beijing
E-mail: ssma@jdl.ac.cn, qujunhua@ncepu.edu.cn
[†]Institute of Digital Media, Peking University, Beijing
E-mail: zhanglili@jdl.ac.cn

***Abstract*— Region of interest (ROI) has been widely used in the extensive applications, such as video conferencing, video surveillance, videophone etc. In this paper, an optimal bit allocation scheme between ROI and non-ROI area is proposed, which aims to improve the visual quality of ROI while guarantee the overall visual quality (OVQ) of the video is also optimal. In the proposed scheme, the visual quality of a certain area, including ROI and non-ROI, is predicted by a visual quality model, and the overall visual quality is modeled by the visual quality difference between ROI and non-ROI. Based on the proposed visual quality models, the optimal bit allocation is achieved, and experimental results show that the proposed scheme can effectively improve the overall visual quality of the video.**

I. INTRODUCTION

Decades ago, the relationship between human visual system and overall visual quality of video sequences was starting to study. Today, the scope of human visual perception and video compression is much larger and keeps expanding with the video and image compression field.

For that the applications of video and image compression are developed to be watched by human eyes, the only correct way to judge the quality of the compressed image is through subjective evaluation. But it is not feasible in practice, as that subjective evaluation is too inconvenient, time consuming and surely expensive. In order to predict perceived image quality automatically with a quantitative measure, the objective image quality assessment is extensive studied.

In order to reach this goal, many image quality metrics were developed in many fields of image applications, and these image quality metrics could be classified in two categories, one is *full-reference*, which means a complete reference image is assumed to be known while the other is *no-reference* or *reduced-reference*, which means the reference image is not or only partially available.

Besides the most widely used full-reference quality metric ---- the mean squared error (MSE) or the peak signal-to-noise ratio (PSNR), many perceptual quality metrics were proposed to enhance the objective image quality. Many of the perceptual quality assessment models were following the strategy of modifying the MSE measure to make the errors accordance with their visibility.

Most of them were error-sensitivity approaches; like visual differences predictor [1], which is a multiple-channel model for image quality; and like just noticeable distortion (JND), which is equivalent to the notion ‘‘perceptually lossless’’ compression [2]-[4], means that whether an image could be compressed without the user perceiving a difference between the compressed and original images is the only criteria of how the image was compressed. There are many other error-sensitivity approaches, like the discrete cosine transform (DCT) [5], [6] or separable wavelet transforms [7]-[9].

Recently SSIM [10] was proposed as a structural-sensitivity quality assessment. This approach has received wide agreement on its effectiveness and it is widely used for image and video compression, it is also implemented in H.264/AVC and X264 codec.

On the other hand, many psychophysical experiments were conducted to establish a HVS model, and most of them were using relatively simple patterns, like spots, bars, sinusoidal gratings, etc. But the real world image is much complex than these patterns, which could be thought of a subset of the real patterns. Recently a database called Modelfest [11] was established, it is a collaborative modeling effort where researchers have volunteered to collect simple or complex patterns, in order to provide a basis for comparing the predictions of early vision models.

To increase performance of image and video compression, many studies were focusing on the attention compression [12], [13], which provides another path to study image features through which part of image or scene naturally attracts human eye mostly. Attention compression is a significant method in image and video compression which is to identify which part or what aspect of a certain image attracts human perception mostly, and we could allocate more bitrates for this partition and less bitrates for the other, so the total bitrates of the certain image or video streams could be cut down and the bandwidth is saved. This technology is useful in many areas where requires bandwidth limit. To reach optimal result, some retina study results were brought to this area, S.Marat presents a spatio-temporal saliency model that predicts eye movement during video free viewing [14], which presents a simulation of the two pathways (magnocellular and parvocellular) of the human visual system based on their main known properties. And also, some quality assessment were proposed based on saliency maps which is similar to attention compression,

recently, an attention compression based quality assessment called quaternion frequency transform (PQFT) is proposed [15], which reveals good result.

ROI (region of interest) based coding, which is an important application of attention compression, in these applications, users pay much more attention to the interested area, while less attention to the other areas, named non-ROI in the paper. In general, if the visual quality of ROI is good, people would feel better in the subjective quality testing. However, if more bits are allocated to ROI for high quality ROI coding and fewer bits remained for non-ROI, the overall visual quality (OVQ) of the video would drop. Therefore, enhancing the visual quality of ROI as well as maintaining the overall quality has become a major concern. To enhance the visual quality of ROI, bit allocation is adjusted by considering the complexity of ROI in [16]. In [17], a "base encoder" and a "region of interest encoder" are used for coding full view version at low resolution and ROI at high resolution respectively. However, the current ROI optimization coding schemes mostly focus on rate-distortion modeling without many considerations on the visual quality of ROI, especially the quality contrast difference between the ROI and non-ROI, which may affect the overall visual quality significantly. In [18], the relation between the quality of ROI and the overall visual quality is studied for JPEG2000, and it was found that ROI coding in JPEG2000 will only produce an overall increase in perceived image quality when three constraints are met: "the image contains a small number of region of interests (≤2); these regions are relatively small (< ¼ of the total image area); and the bit-rate is low enough to produce visible compression artifacts (<0.25 bpp)". This conclusion is drawn for JPEG2000, which uses the explicit ROI coding, where the parameters of ROI should be coded as side information. So the number of ROIs should be small and the ROI regions should be relatively small. However, these may not be true for real video applications where many ROIs exist and adaptive rate allocation for ROIs is desirable. Although the researches in [18] give a good direction on ROI optimization coding, the conclusion is very limited and only low bit rate requirement is declared. It didn't provide an optimal bit allocation scheme between ROI and non-ROI.

In this paper, overall visual quality optimization coding with ROI is studied as an optimal bit allocation problem between ROI and non-ROI. Firstly we employed Structural Similarity Index (SSIM) [19] as a subjective visual quality metric and a visual quality prediction model is established for ROI and non-ROI with SSIM. The overall visual quality is modeled with the visual quality difference between ROI and non-ROI. Based on the proposed models, the visual quality optimization coding is evolved into an optimal bit allocation problem, and it can be resolved with a Lagrangian method.

The rest of this paper is organized as follows. In section 2, the visual quality optimization coding problem is detailed, including the overall visual quality model and visual quality prediction model. In section 3, the overall visual quality optimized coding scheme is presented and experimental results are given. Section 4 indicates some disadvantages and points out some solutions. Section 5 concludes the whole paper.

II. VISUAL QUALITY OPTIMIZED VIDEO CODING

To evaluate the coding effect, a video quality metric needs to be selected. PSNR and MSE are the most prevalent metrics; however, such measurements may not in accordance with human visual quality judgment well. Hence, subjective image and video quality assessment methods have been extensively studied. SSIM [10] has received wide agreement on its effectiveness and is now implemented in H.264/AVC and X264 codec for visual quality assessment. SSIM operates based on the notion that the HVS has evolved to extract structural information from natural images, and therefore, a high-quality image is one whose structural most closely matches that of the original. SSIM metric employs a modified measure of spatial correlation between the pixels of the original and the distorted images to quantify the extent to which degree the image's structure has been distorted.

The basis of the structural similarity index as a means for image quality assessment is that we could extract structural information from the image through using HVS. The SSIM index measures three local quantities: the structural similarity, the luminance and the contrast similarity. Suppose $x = \{x_i \mid i = 1, 2, ..., N\}$ and $y = \{y_i \mid i = 1, 2, ..., N\}$ are two finite-length image signals, which have been aligned with each other. The three quantities are defined as:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

while the SSIM index is:

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y)$$

where $l(x, y)$ compares the luminance of the two image blocks, $c(x, y)$ compares the contrast, $s(x, y)$ measures the structural correlation, μ_x and μ_y denote the sample means of x and y, respectively, σ_x^2 and σ_y^2 denote the sample variance of x and y, respectively, $\sigma_x\sigma_y$ is the sample cross covariance between x and y, and C_1, C_2, and C_3 are three constants introduced to avoid unstable behavior in the regions of low luminance or low contrast, and the three constants have the relationship that :

$$C_1 = L^2K_1^2, C_2 = (K_2L^2), C_3 = \frac{C_2}{2}$$

The block size is typically 8 × 8, and the final SSIM value is the averaged SSIM of all blocks. The maximum SSIM index is 1.The more similar two images are, the higher the SSIM index is. The SSIM takes into account properties of HVS so it is a perceptual-based image quality assessment.

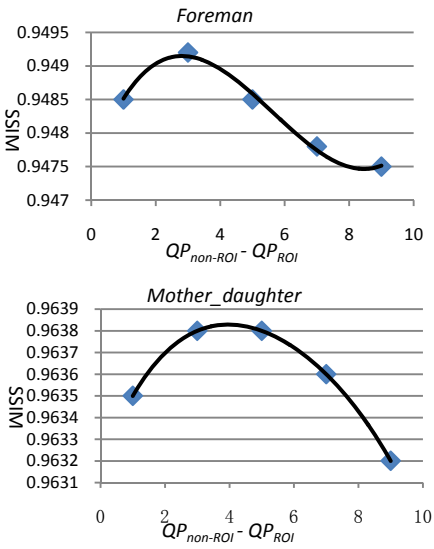


Fig. 1 Relation between quantization parameter and overall image quality

Although the local structure distortion is considered in SSIM, the ROI perception of human visual system is not included. As we know the quality of ROI would affect the overall image quality greatly, this is also verified by the experiments. Firstly we studied the relationship between quantization parameter and overall image quality. We tested 2 sequences *Foreman* and *Mother_daughter*, then we discovered that there exists a three degree polynomial relation between ($QP_{non-ROI}$-QP_{ROI}) and over all SSIM.

As shown in Fig.1, the relation could be described as:

$$OVQ = A \times ((QP_{non-ROI} - QP_{ROI}) - B)^3 + C \qquad (1)$$

where A, B and C are model parameters.

Eq.1 implies that the quantization parameter differences could affect overall visual quality in certain ways, but region visual quality cannot be restricted by quantization parameter only, many other parameters could also affect region visual quality. As a reason of that, the relation between overall visual quality and SSIM differences were studied in next step.

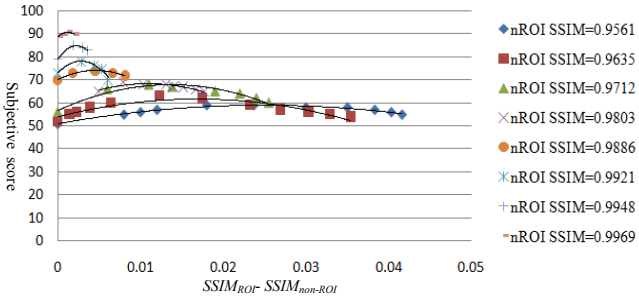


Fig. 2 The relation between the overall visual quality and the visual quality difference of ROI/non-ROI

As shown in Fig. 2, we use 7 CIF sequences to study the relation between the overall visual quality and the visual quality difference between the ROIs and non-ROIs, including *News, Paris, Mother_daughter, Foreman, Akiyo, Salesma and Deadline*. In each test sequence, 8 ROIs were labeled manually and are coded with the same quality level by adjusting the quantization parameter of ROI. Then we fixed the quality of non-ROI area and vary the quality level of encoded ROI.

TABLE I
MODEL PARAMETERS ESTIMATION RESULTS BY CURVE FITTING

$SSIM_{nROI}$	*A*	*B*	*C*
0.9561	-13263	0.02482	42.72245
0.9635	-27184	0.016833	46.27736
0.9712	-58940	0.013692	45.79064
0.9803	-79796	0.010978	49.18028
0.9886	-183943	0.004585	66.33383
0.9921	-676433	0.002793	67.6548
0.9948	-893457	0.002955	71.19929
0.9969	-1000000	0.001502	86.64445

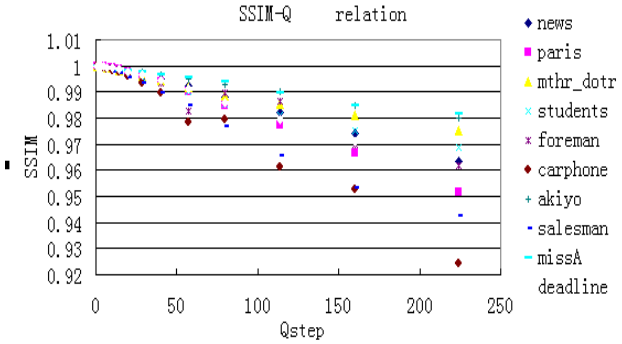


Fig. 3 SSIM-Q model on test sequences

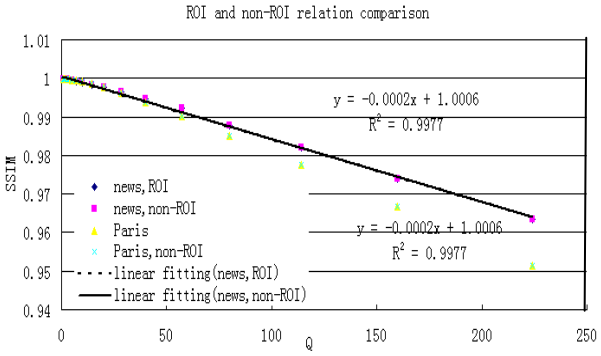


Fig. 4 SSIM-Q model curve fitting for both ROI and non-ROI

Totally we get 264 coded sequences and subjective quality assessment is done for these coded sequences. As defined in [20], SSCQE (Single Stimulus Continuous Quality Scale) method is used for subjective scoring, but in this paper, the subjective score is scaled to the range [0, 100] and 100 denote the best quality. As shown in Fig.2, we found that the overall subjective quality first increases with the increasing difference of ROI and non-ROI visual quality, then the overall visual quality would drop. So we can model the overall visual quality OVQ as:

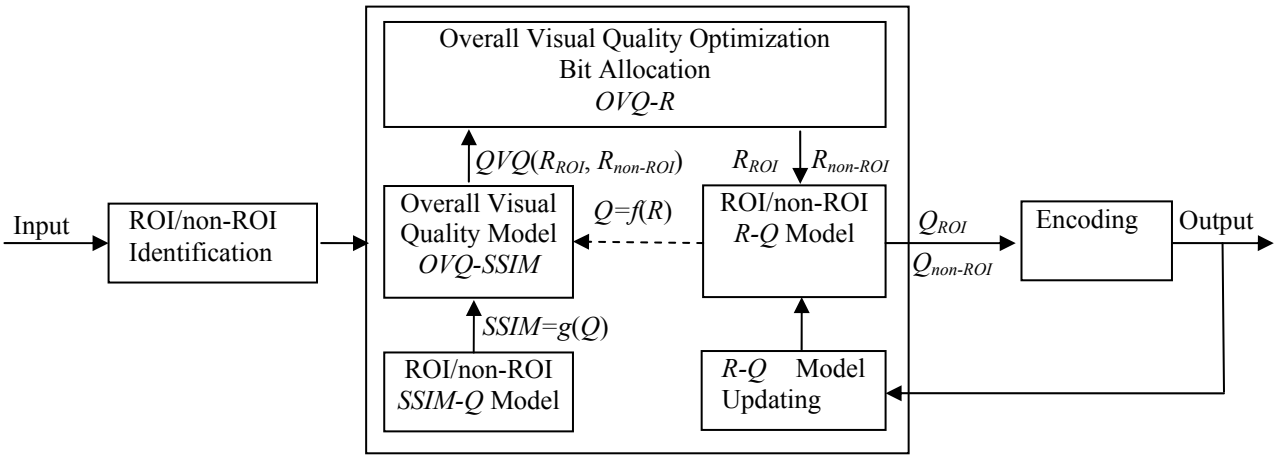


Fig. 5 The framework of the proposed bit allocation scheme

$$OVQ = A \times ((SSIM_{ROI} - SSIM_{non\text{-}ROI}) - B)^2 + C \quad (2)$$

A, B and C are model parameters. Although the SSIM values are different for the different sequences even with the same quantization parameters, it was found that the model parameters in Equation (2) vary little with the sequences and are only related with the quality of non-ROI $SSIM_{non\text{-}ROI}$. Table 1 shows the parameter values of the curve fitting results in Fig. 2.

With the overall visual quality model in Equation (2), the bit allocation between ROI and non-ROI can be modeled as a Lagrangian maximum problem, as,

$$max(OVQ)$$

with the constraint

$$R_{ROI} + R_{non\text{-}ROI} = R$$

where R_{ROI}, $R_{non\text{-}ROI}$ is the allocated bits to ROI and non-ROI respectively. R is the total bits. The problem can be resolved as,

$$max(J)$$

s.t.

$$J = OVQ + \lambda(R_{ROI} + R_{non\text{-}ROI} - R) \quad (3)$$

To resolve this problem, we need study the relation between the visual quality and the allocated bits of ROI/non-ROI. According to our observation, the SSIM-Q relation is very close to linear as shown in Fig. 3 and can be modeled as:

$$SSIM = g(Q) = m \times Q + l \quad (4)$$

Q is the quantization step. m and l are model parameters. Furthermore, it was found that we can use the same model parameters for both ROI and non-ROI for one sequence, as shown in Fig. 4. The SSIM values of ROI and non-ROI are almost the same for each quantization step. So we can use one set of model parameters in Equation (4) for both ROIs and non-ROIs in one frame.

As we know a quadratic model is usually used for rate distortion modeling [21], it is shown as,

$$R = S_2 \frac{1}{Q^2} + S_1 \frac{1}{Q} \quad (5)$$

S_2 and S_1 are model parameters. For given bit rate allocation R, the quantization step is decided as,

$$Q = f(R) = \frac{2S_2}{\sqrt{S_1^2 - 4RS_2} - S_1} \quad (6)$$

So we have,

$$SSIM = g(f(R)) \quad (7)$$

Replace the $SSIM_{ROI}$ and $SSIM_{non\text{-}ROI}$ by equation (7), J can be represented as function of R_{ROI} and $R_{non\text{-}ROI}$,

$$J = OVQ(R_{ROI}, R_{non\text{-}ROI}) + \lambda(R_{ROI} + R_{non\text{-}ROI} - R) \quad (8)$$

R_{ROI}, $R_{non\text{-}ROI}$ is the allocated bits to ROI and non-ROI areas respectively. It should be pointed out that the model parameters in (4) for the ROI and non-ROI are different. They can be updated with least square error method as [21]. So the optimal bit allocation between ROI and non-ROI region can be found by:

$$\begin{cases} \dfrac{\partial J}{\partial R_{ROI}} = 0 \\ \dfrac{\partial J}{\partial R_{non\text{-}ROI}} = 0 \\ \dfrac{\partial J}{\partial \lambda} = 0 \end{cases} \quad (9)$$

The Equation (9) can be resolved by iterative Newton method.

III. Proposed Visual Quality Optimization Coding Scheme and Experimental Results

Based on the visual quality model proposed in Section 2, an overall visual quality optimization coding scheme is proposed as shown in Fig. 5. In the proposed scheme, the ROI/non-ROI areas are first identified on the input video, which can be labeled by visual attention model or manually. In the paper, the ROI/non-ROI areas are labeled as rectangular manually and it is given as input information before encoding one frame. As referred in Section 2, the overall visual quality optimization scheme is built upon the visual quality model

and R-Q model of the ROI/ non-ROI. Once the optimal bit allocation is decided with the resolution of Equation (8) and (9), the quantization parameter for ROI/non-ROI can be calculated with the R-Q model. After encoding one frame, the R-Q model parameters are updated as [20].

The proposed visual quality optimization bit allocation scheme has been implemented into H.264 reference software JM10.0. The test sequences are coded with IPP format, RDO on and 3 reference frames. The ROI region is first labeled manually and for simplicity the case of only one ROI is studied in the paper. Subjective quality assessment is done for these coded bitstreams with SSCQE method as defined in [6].

TABLE 2. SUBJECTIVE VISUAL QUALITY TESTING RESULTS

Sequence	Target bitrate (kb/s)	Subjective Score		
		Original	Proposed	Improvement
Akiyo	128	80	84	4
	96	69	74	5
News	128	76	78	2
	96	65	68	3
Foreman	128	74	76	2
	96	63	65	2
Mother_ daughter	128	77	78	1
	96	68	72	4
Deadline	128	71	73	2
	96	61	65	4

Table 2 shows the subjective visual quality testing results. From the table it can be seen that visual quality is improved for all test sequences with the proposed scheme. Especially, at low bit rate, the improvement is more obvious.

The subjective visual quality comparison is provided in Fig. 6 and Fig.7. It can be found that the quality of ROI is improved while that of non-ROI has not been degraded.

IV. FUTURE WORKS

As this paper is just to explore ways to increase overall visual quality, the ROI regions in this experiment were labeled manually. However, there appear many automatic methods to label ROI regions recently, as [14] described; hence, automatic ROI region detection is a useful method that needs to be studied and applied desperately next step.

On the other hand, some new quality assessments as [15] could be brought to this model to enhance the performance.

V. CONCLUSION

In this paper, a bit allocation scheme that aims to improve the visual quality of region of interest, while the overall visual quality of the entire frame is guaranteed as well. In the proposed scheme, the subjective visual quality of the whole frame is optimized with respect to the bit allocation on ROI area.

The experimental results could be reflected from the subjective visual quality comparison between print-screens of two sequences *Deadline* and *Foreman*. The region in the rectangle is manually labeled as the region of interest. Fig.6 shows a comparison between the print-screen of sequence *Deadline* using original and proposed schemes separately. We could draw from the comparison that visual quality of ROI and overall visual quality are both effectively improved although the visual quality of other region is little degraded.

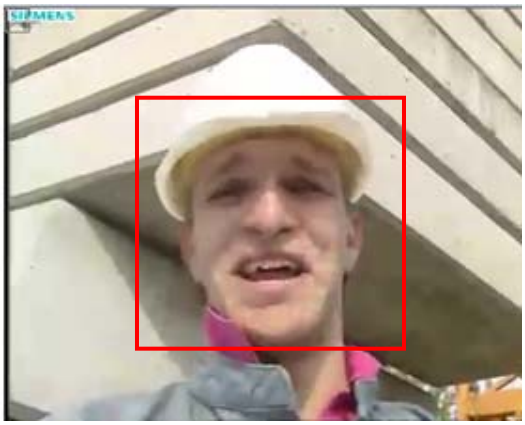
Fig.7 shows a comparison between the print-screen of sequence *Foreman* using original and proposed schemes separately. We could draw from the comparison that the object edge which is distorted in low bit rate is well recovered by our proposed scheme. Many details in the manually labeled ROI are also well preserved and overall visual quality is also increased.

Experimental results show that the proposed scheme can effectively increase the visual quality of ROI and the overall visual quality as well. As said before, the ROI is labeled manually in the paper; in the future adaptive bit allocation for multiple ROIs labeled with attention model will be researched.

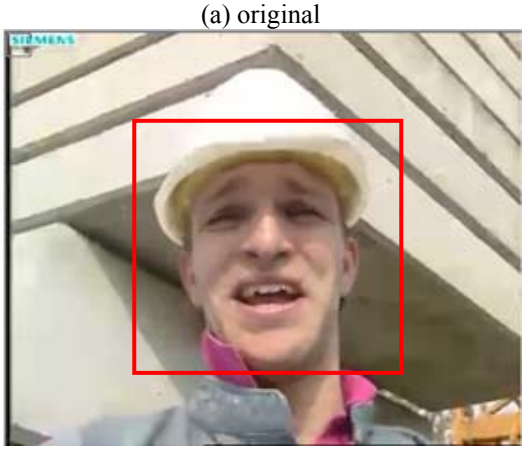
(a) original

(b) proposed

Fig. 6. Visual quality comparison between the original and the proposed visual quality optimization coding method of sequence *Deadline*



(a) original



(b) proposed

Fig. 7. Visual quality comparison between the original and the proposed visual quality optimization coding method of sequence *Foreman*

REFERENCES

[1] S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital images and human vision* (A. B. Watson, ed.), pp. 179–206, Cambridge, Massachusetts: The MIT Press, 1993.

[2] V. Ramamoorthy and N. Jayant, ''On transparent quality image coding using visual models,'' in *Human Vision, Visual Processing, and Digital Display, Proc. SPIE* 1077, 146–154 ,1989

[3] S. Daly, ''Application of a noise adaptive contrast sensitivity function to image data compression,'' in *Human Vision, Visual Processing, and Digital Display, Proc. SPIE* 1077, 217–227 ,1989

[4] C. Stein, A. Watson, and L. Hitchner, ''A psychophysical rating of image compression techniques,'' in *Human Vision, Visual Processing, and Digital Display, Proc. SPIE* 1077, 198–208 ,1989

[5] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE*, vol. 1913, pp. 202–216, 1993.

[6] A. B. Watson, J. Hu, and J. F. III. McGowan, "DVQ: A digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[7] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing*, vol. 6, pp. 1164–1175, Aug. 1997.

[8] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Processing*, vol. 5, pp. 717–730, May 1999.

[9] Y. K. Lai and C.-C. J. Kuo, "A Haar wavelet approach to compressed image quality measurement," *Journal of Visual Communication and Image Representation*, vol. 11, pp. 17–40, Mar. 2000.

[10] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[11] A. B. Watson, "Visual detection of spatial contrast patterns: Evaluation of five simple models," *Optics Express*, vol. 6, pp. 12– 33, Jan. 2000.

[12] A. Treisman and G. Gelade,''A feature integration theory of attention,'' *Cogn. Psychol.* 12, 97–136, 1980.

[13] B. Julesz, ''AI and early vision—Part II,'' in *Human Vision, Visual Processing, and Digital Display, Proc. SPIE* 1077, 246–268 ,1989

[14] Sophie Marat, Tien Ho Phuoc, Lionel Granjon, Nathalie Guyader, Denis Pellerin, Anne Guérin-Dugué, "Modelling Spatio-Temporal Saliency to Predict Gaze Direction for Short Videos", *International Journal of Computer Vision,* Vol. 82, Issue: 3, Pages: 231-243 2009

[15] Qi Ma, Liming Zhang, and Bin Wang," New strategy for image and video quality assessment", *Journal of Electronic Imaging,* Vol.19, Issue 1, pp. 011019-011019-14 ,2010

[16] A. Pietrowcew, A. Buchowicz and W. Skarbek, "Bit-Rate Control Algorithm for ROI Enabled Video Coding," *Lecture Notes in Computer Science*, vol. 3691, pp514-521, 2005.

[17] P. Wu and H. Chen, "Frame-Layer Constant-Quality Rate control of Region of Interest for Multiple Encoders With Single Video Source," *IEEE Transaction on Circuits and Systems for Video Technology*, 17(7): 857-867, 2007.

[18] A. Bradley, "Can Region of Interest Coding Improve Overall Perceived Image Quality?" *APRS Workshop on Digital Image Computing,* pp 41-44, 2003.

[19] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[20] ITU-R Recommendation BT.500-10:"Methodology for the Subjective Assessment of the Quality of Television Pictures." ITU, Geneva, Switzerland, 2000.

[21] T. Chiang and Y. Zhang, "A New Rate Control Scheme Using Quadratic Rate Distortion Model," *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1): 287-311, April. 1997.