Distortion Model based on Perceptual of Local Image Content

Tung Pham Thanh[1], Duong Dinh Trieu[2], Xiem Hoang Van [2] Tien Vu Huu [3] and Ha Le Thanh[2]

***Abstract*—As humans are the ultimate receivers of the majority of visual signals being processed, the most accurate way of assessing image quality is to ask humans for their opinions of an image's quality, known as the subjective visual quality assessment (VQA). The subjective image quality scores gathered from all subjects are processed to be the mean opinion score (MOS), which is regarded as the ground truth of image quality. Due to the fact that the human visual system (HVS) is differently sensitive to features of image patch, a novel coding distortion modelling method for local image perception is proposed in this paper. An experimental quality assessment to approach database for image patch has been developed. Mean opinion score is regarded as an essential parameter meanwhile the QP-MOS sigmoid curve is determined by local image content.**

***Index Terms*—image quality assessment, coding distortion, logistic modelling**

I. Introduction

Subjective assessments are still frequently used as a benchmark for image and video quality, where a group of human viewers are asked for their opinions on quality under a range of test conditions. Subjective testing conditions must be closely controlled, with appropriate screening of observers and post-processing of the results to ensure consistency and statistical significance. Despite of being generally effective, they are costly and time consuming.

For the above reason, objective measures of video quality are taken under consideration as an alteration. They have conventionally been computed using the absolute or squared difference between the distorted version of frames and their reference version. It is however well known that the perceptual distortion experienced by the human viewers cannot be fully characterized using such simple mathematical differences. Due to the limitations of these distortion-based measures, perception-based metrics have begun to replace them. These offer the potential for enhanced correlation with subjective opinions, thus enabling more accurate estimations of visual quality.

Playing an important role in perception-based metrics, databases with subjective data facilitate metric development and benchmark have been well developed over the time. There is a number of publicly available image quality benchmark databases, including LIVE Image [1], TID2008 [2], TID2013 [3], CSIQ [4], IVC [5], IVC-LAR [6], Toyoma [7], WIQ [8], A57 [9], MMSP 3D Image [10], and Image Re-targeting Subjective Quality [11]. All of the above mentioned databases evaluate the overall picture quality. But the problem is that the perceptual quality of each image patch is different with the same level of noise. Figure 1 show that the distortions around the houses and on the sky regions (solid square) are easily observable. However, those on textural regions (dash dot square) are less noticeable.

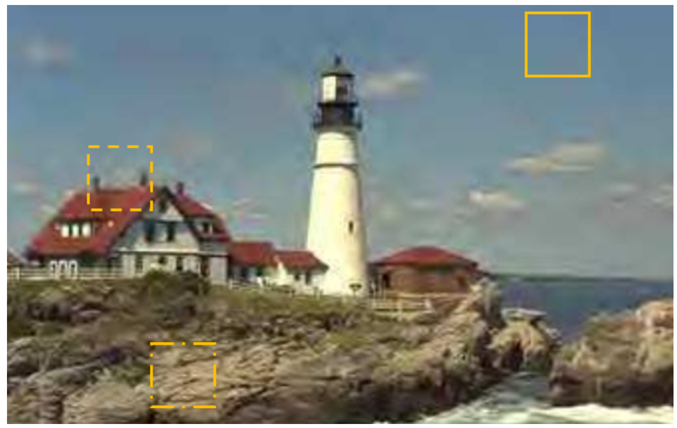


Fig. 1. Example of distorted image

In this paper, we propose a quality assessment approach database for image patch with the desire to create a new perception-based metric to apply for each region. Therefore, authors present a modeling method to assess the coding distortion at image locality based on human visual perception. In this method, authors would create a new testing database and conduct subjective test.From the data obtained, we model and analyze the coding distortion using logistic regression. The result can be customized quantization parameter (QP) block- based in video coding.

The rest of this paper is organized as follows: The subject testing is described in Section 2. In Section 3, a novel coding distortion modelling method method is proposed. Section 4 presents our conclusions.

II. Subjective testing

Since all available image quality benchmark databases are not suitable for evaluating the quality image as a whole, it is unable to investigate which parts of the testing image contribute to the testing results. In this work, we set up an experimental database to evaluate the quality of human for each image patch.

A. Testing image database creation

The goal of our study is to create a testing image database for local image perception. Due to the research orientation for video coding, testing images are extracted from the video test sequence and noise types are added to the original video by H.264/AVC compression before extracting. In each image,

[1]University of Fire Fighting and Prevention, Ha Noi
[2]The VNU University of Engineering and Technology
[3]Posts and Telecommunications Institute of Technology

we will select several patches for rating. In this database, instead of random selection, we assume that there are three attributes (smooth texture, complex texture and edge texture) in an image patch affecting the subjective image quality. These attributes are classified by a region growing algorithm (Fig. 2). Image patches (64x64) are selected in the respective regions.

Image quality ratings greatly depend on the viewing distance between the observer and the testing display. In practical, it is very difficult to change such distance so we choose two resolutions for the image including: full size image (original image) and half size image. After this procedure, we obtain 20 video sequence $\times$ 5 images $\times$ 3 positions $\times$ 2 sizes = 600 image patches.

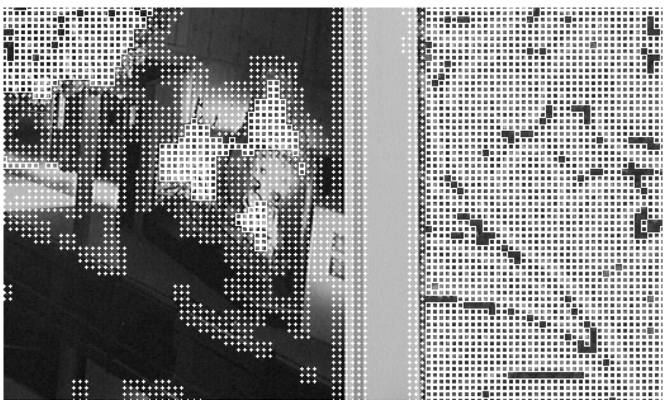


Fig. 2. Classified into three types: smooth texture (none), complex texture (o), and edge regions (+)

B. Testing methodology

For the purpose of this study, 20 subjects who have been trained and practiced quality assessing of several sample images. Because the image quality assessment methods [12] are only suitable for assess quality of image as a whole. Therefore, we modify this image selection method in the standard so that the users can only concentrate and assess the local image patch instead of the whole image. Each pair quality is assessed following the procedure of Fig. 3. After watching at least twice per image, the observers would score on scale of 5 (excellent, good, fair, poor or bad).

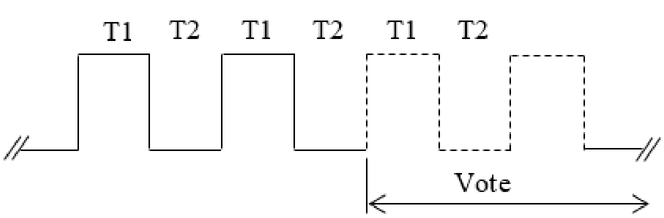


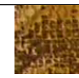
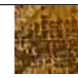
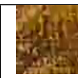
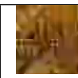
Fig. 3. Presentation structure of test material

At the end of experiment, each original image patch was compared to five distort image patches corresponding to five testing pairs. Table I shows an example of a complex texture image patch where each column is a testing pair.

C. Statistical analysis of subjective test results

After the experiment, the results of 12 000 subjective ratings to 600 image patches from 20 different subjects are obtained.

TABLE I
MOS OF A COMPLEX TEXTURE IMAGE PATCH

Patch					
QP	35	40	45	50	55
Average	4.8	3.85	2.25	1.45	1.05

Let I_o denote original image patch of an image patch and let $I_{o,qp}$ denote distortion image compressed by qp. The mean opinion score of a subjective rating under a given test condition c for subject s is $y(I_o, I_{o,qp}, c, s)$, where o ranging from 1 to 120 denote index of original patches and indicating the resolution of testing images $c \in (full, half)$. In our experiment, the mean opinion score (MOS) of each image pair is calculated by:

$$\overline{y}(I_o, I_{o,qp}, c) = \frac{1}{S}\sum_{s=1}^{S} y(I_o, I_{o,qp}, c, s) \quad (1)$$

where S is the total number of testing subjects.

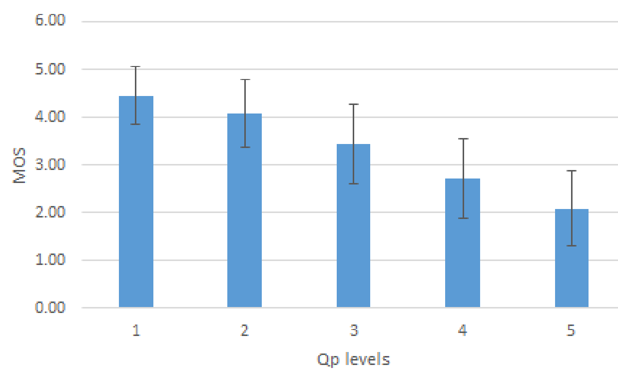


Fig. 4. Relationship between MOS and QP levels

Figure 4 shows the statistical relationship between MOS and QP levels for all tested image patches. It can be seen from the figure that average MOS decreases when QP increases. While the MOS decreases when QP increases and standard deviations increase along with QP but still in acceptable range.

Table II show average of $MOSs$ respectively for each tested image patch type, resolution and QP. There is a difference in quality among complex texture area, edge texture area and smooth texture area. The smooth texture areas have the lowest MOS score, followed by the edge and finally, the highest complex texture.

III. MOS MODELLING USING LOGISTIC FUNCTION

Subjective results are frequently used as a benchmark for establishing a relationship between MOS and QP [12]. The scores produced by the objective video quality metric must be correlated with the viewer scores in a predictable and repeatable fashion. The relationship between prediction and MOS does not need to be linear as subjective testing can exhibit non-linear quality rating compression at the extreme of the test range. The typical relationship between mean

TABLE II
AVG MOS SCORE OF EACH TYPES OF AREAS

Types	Resolution	qp=25	qp=30	qp=35	qp=40	qp=45	qp=50	qp=55
smooth texture	full	-	-	3.80	3.22	2.52	2.04	1.73
complex texture	full	-	-	4.43	3.75	2.86	1.87	1.32
edge	full	4.70	4.54	4.09	3.35	2.55	-	-
smooth texture	half	-	-	4.48	4.05	3.06	2.45	1.73
complex texture	half	-	-	4.60	4.27	3.56	2.43	1.66
edge	half	4.69	4.68	4.56	4.22	3.34	-	-

opinion score of a image patch Y and a given distortion level qp, generally exhibit a skew-symmetric sigmoid form. Hence the function $Y = f(qp)$ can be approximated by a logistic function of the form:

$$Y(I_o, qp, c) = \frac{5}{1 + e^{(qp-a)b}}, \tag{2}$$

where a denotes qp when Y at average, b is the rate of declining in image quality while increasing qp.

Two variants a and b can be derived from the experimental data. Because y depends on non-linearity of a and b, it is not possible to solve the equation to find them. The Pearson Linear Correlation Coefficient (PLCC) is used as a measure of the accuracy of fit of variants to the subjective scores. It characterizes how well the metric under test can predict the subjective quality ratings. For example, with the complex texture image patch as shown in Table I, $a = 42.7148$ and $b = -0.30482$. The chart is as follows (fig.5).

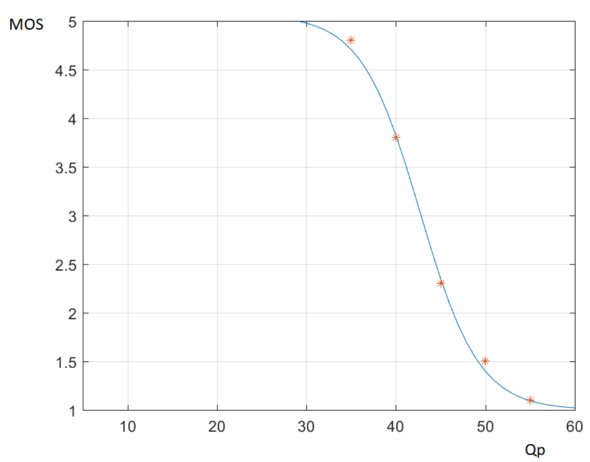


Fig. 5. Logistic function fit of data in Table I

The variable a, b can also be estimated by the visual features of image patch. Here, we denote f_i as scalar value for an visual feature of image patch. We also assume that a and b are the linear combination of the image features as follows:

$$a = a_0 + \sum_{i=1}^{N} a_i f_i \tag{3}$$

$$b = b_0 + \sum_{i=1}^{N} b_i f_i \tag{4}$$

where N is the total number of features.

In equation 3 and 4, we can not determine the number of visual features as well as the importance of each visual feature. Therefore, visual features of image $(m \times n)$ that may affect image quality are estimated including:

- Luminance coefficient mean of 16×16 image patch μ_{16} calculated according to the formula:

$$f_1 = \mu_{16}[p, q] = \frac{1}{256} \sum_{j=1}^{16} \sum_{i=1}^{16} Y[j + 16p, i + 16q] \tag{5}$$

  where $p = 1, 2, 3 \ldots n/16$, $q = 1, 2, 3 \ldots m/16$ are row and column indices of luminance coefficient mean matrix.
- Variance of 16×16 image patch σ_{16} calculated according to the formula:

$$\begin{aligned} f_2 &= \sigma_{16}[p, q] \\ &= \sum_{j=1}^{16} \sum_{i=1}^{16} |Y[j + 16p, i + 16q] - \mu_{16}[p, q]| \end{aligned} \tag{6}$$

  where $p = 1, 2, 3 \ldots n/16$, $q = 1, 2, 3 \ldots m/16$ are row and column indices of variance matrix.
- The edge density $f_3 = \theta_{16}[p, q]$ of each 16×16 image patch is calculated based on the total number of pixels in the edge region.
- Maximum and minimum of luminance coefficient are calculated according to the formula:

$$f_4 = \max_{16}[p, q] = \max\{Y[j, i]\}, \tag{7}$$

$$f_5 = \min_{16}[p, q] = \min\{Y[j, i]\}, \tag{8}$$

  where $p = 1, 2, 3 \ldots n/16$, $q = 1, 2, 3 \ldots m/16$ are row and column indices, $i = 16q + 1 \ldots 16q + 16$, $j = 16p + 1 \ldots 16p + 16$.
- The neighboring luminance average of 16×16 image patch in terms eight neighboring patches is as follows:

$$\begin{aligned} f_6 = \mu_{lc}[p, q] = \frac{1}{8} (&\mu_{16}[p, q-1] + \mu_{16}[p-1, q] + \\ &+ \mu_{16}[p-1, p+1] + \mu_{16}[p-1, q-1] + \\ &+ \mu_{16}[p, q+1] + \mu_{16}[p+1, q-1] + \\ &+ \mu_{16}[p+1, q] + \mu_{16}[p+1, q+1]), \end{aligned} \tag{9}$$

  where $p = 1, 2, 3 \ldots n/16$, $q = 1, 2, 3 \ldots m/16$ are row and column indices.
- The neighboring variance average (f_7) and the neighboring edge density average ((f_8)) of 16×16 image patch in terms eight neighboring patches are calculated similar to the equation 9.
- Next, convert the 16×16 blocks of luminance coefficient matrix Y into the Cosin (DCT transform) frequency domain and take the absolute value:

$$Z_{pq} = |DCT(Y_{pq})|, \tag{10}$$

TABLE III
REGRESSION STATISTICS

Method	Decision Tree	Linear	k-Nearest Neighbors
Multiple R	0.7498	**0.7943**	0.5843
R Square	0.5622	**0.63091**	0.34140
Standard Error	2.382687	**1.94254**	3.45296

where Y_{pq} denote a block including row indices $p = 1, 2, 3 \ldots n/16$, row indices $q = 1, 2, 3 \ldots m/16$ in Y.

- Convert Z_{pq} into a one-dimensional matrix X_{pq} consisting of 255 elements after removing the first element.
- Mean of coefficient in frequency domain is calculated by:

$$f_9 = \mu_{dct}\left[p, q\right] = \frac{1}{255}\sum_{j=1}^{255} X_{pq}\left[j\right]. \tag{11}$$

where $j = 1..255$.

- Variance of coefficient in frequency domain is calculated by:

$$f_{10} = \sigma_{dct}\left[p, q\right] = \sum_{j=0}^{254}\left|X_{pq}\left[j\right] - \mu_{dct}\left[p, q\right]\right| \tag{12}$$

where $j = 1..255$.

- Maximum and minimum of coefficient in frequency domain are calculated by:

$$f_{11} = \max_{dct}\left[p, q\right] = \max\{X_{pq}\left[j\right]\}, \tag{13}$$
$$f_{12} = \min_{dct}\left[p, q\right] = \min\left\{X_{pq}\left[j\right]\right\}, \tag{14}$$

where $j = 1..255$.

Analyzing the data has determined the relationship between parameters of the logistic model with the features by three regression methods. Table III shows that Linear method is the best match to the tested data. Parameters of the logistic model are calculated by 15, 16.

$$\begin{aligned} a = {} & 36.2247 - 0.0581 f_2 + 0.0153 f_3 + 0.0248 f_4 + \\ & + 0.1874 f_7 - 0.0127 f_8 - 0.0052 f_{11} \end{aligned} \tag{15}$$
$$\begin{aligned} b = {} & -0.2003 - 0.0009 f_1 + 0.0003 f_3 + 0.0002 f_5 + \\ & + 0.0006 f_6 \end{aligned} \tag{16}$$

The results of regression method show that some features affect the model and some does not. In equation 15, the average quality of compressed image patch a depends on the features of original image patch including: edge density, variance, maximum of luminance coefficient, maximum of coefficient in frequency domain. In addition, a contrasts with features of neighboring area including: edge density and variance. This is consistent with the experimental results that complex regions have subjective score higher other regions. The highest score of complex region occurs when neighboring area is in another region (smooth or edge). It can be seen from equation 16 that the rate of declining b depends on the brightness (luminance coefficient) and edge density. The higher the value of brightness is, the faster the image quality decreases. For example, with the original complex texture image patch as shown in Table I, we can predict $a = 43.0352$ and $b = -0.31143$ following equation 15, 16. This result match the model in figure 5 which can be used to select an appropriate qp for the image quality in video coding.

IV. CONCLUSIONS

This paper presents a coding distortion modelling method for local image perception, which is able to predict objective evaluation from the perceptual point of local image content. There are 600 distortion samples in quality assessment database rated by 20 subjects with standards ITU-R BT.500-11. Experimental results show that compressed image quality decreases depending on the visual features of image. The intrinsic visual features of the image patches are: edge density, average brightness, variance, maximum coefficient, minimum coefficient on the pixel domain and frequency domain. At the same time, the research also considers neighborhood patches visual features that affect the quality of the image. A method is used for determining the intrinsic visual features of image patches that can affect the sensitivity of the human eye to noise. The patches in the image are compressed at different levels depending on the sensitivity to the human eye: the less sensitive the patch, the less the encoder will be used compared to the other patches. In video compression, the QP is used to change the number of image bit encoders, the higher the port number, the lower the bitrate.

In addition, the author models the relationship between QP and visual quality measured by human visuals in the form of logical curves. According to this model, image quality decreases as the QP increases. However, the average image quality and the quality loss rate are different among the patches depending on the sensitivity of the patch. For noise-sensitive patches, the average image quality is lower and the speed is reduced faster. Experimentally, the control coefficients of the imaginary loop are calculated based on the characteristics that affect the noise sensitivity of the image patch.

REFERENCES

[1] LIVE Video Quality Database. http://live.ece.utexas.edu/research/quality /livevideo.html
[2] Tampere Image Database. http://www.ponomarenko.info/tid2008
[3] Ponomarenko N, Jin L, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F, Kuo C-CJ (2014) Image database TID2013: peculiarities, results and perspectives. Signal Process Image Commun 50:5777.
[4] Categorical Image Quality (CSIQ) Database. http://vision.okstate.edu/csiq
[5] IVC Image Quality Database. http://www2.irccyn.ec-nantes.fr/ivcdb
[6] IVC-LAR Database. http://www.irccyn.ecnantes.fr/autrusse /Databases /LAR
[7] Toyoma Database. http://mict.eng.utoyama.ac.jp/mictdb.html
[8] Wireless Imaging Quality (WIQ) Database. http://www.bth.se/tek/rcg.nsf/pages/wiq-db
[9] A57 Database. http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr
[10] MMSP 3D Image Quality Assessment Database. http://mmspg.epfl.ch/cms/page-58394.html.
[11] Image Retargeting Subjective Database. http://ivp.ee.cuhk.edu.hk/projects/demo/retargeting/index.html
[12] R-REC-BT.500-11-200206
[13] Ke Gu, Min Liu, Guangtao Zhai, Xiaokang Yang, Wenjun ZhangQuality "Assessment Considering Viewing Distance and Image Resolution", IEEE Transactions on Broadcasting Volume: 61, Issue: 3, Sept. 2015