Image Quality Modelling for Moving Streams

Jin-Seo Kim[1], Maeng-Sub Cho[1], Bon-Ki Koo[1], M. R. Luo[2] and Stephen Westland[3]
[1]Digital Content Research Division, ETRI, Republic of Korea
[2]Department of Colour Science, University of Leeds, Leeds, UK
[3]Centre for Colour Design Technology, University of Leeds, Leeds, UK

Abstract

Digital image reproduction is now prevalent in most image application fields such as TV, Mobile, Internet, Photograph, Digital Printing, etc. Moreover, as the network infrastructure is developing very rapidly, most people in the world are using internet services and people upload and download the contents which they create in the form of images or moving streams as the name of UCC (User Created Contents). Requirement of the high quality image reproduction therefore, is one of the main issues in the companies, service providers and users. Image quality is highly subjective concept and it is therefore difficult to quantify. However, it can be quantified by using statistical and psychophysical approaches based on the experimental data analysis. Also image quality model can be developed by previously mentioned experimental data and its analysis. In this paper, the psychophysical experiments for moving streams are introduced in order to develop an image quality model for moving streams. From the categorical judgment data, linear form of image quality model was derived with the r value of 0.9393.

Introduction

The term 'quality' is an ideal and subjective concept so it can be determined by many different attributes such as colour, resolution, sharpness and noise. A number of metrics have been published that could be used to predict image quality which includes CIECAM02[1], iCAM[2], MTFA, SNR and MSE. However, none of these metrics can easily predict certain perceptual attributes of human vision such as the naturalness of the image[3]. CIE TC8-02 is studying the calculation of colour difference using spatial characteristics.

Reproduction of moving stream is one of the major roles in image reproduction industries such as digital TV, digital cinema, animation, mobile content, etc., and the reproduction of high quality image content plays an important role to the current and the future imaging industries. With this in mind, this study covered some initial step to the development of image quality model by performing the psychophysical experiments and data analysis for the moving streams.

The aim of this study is to develop image quality model for moving streams so that it can be used to enhance the state of the art colour appearance model by adding functions which can predict the image quality attributes in both still images and moving streams. To develop image quality model for moving streams, seven image quality attributes were determined first in this study and psychophysical experiments have been conducted.

Previous Study

In the previous study[4], psychophysical experiments for the still images were conducted and the data analysis was done for the development of image quality model for still images. Six image-quality attributes (lightness, chroma, contrast, noise, sharpness, and compression) were selected in the experiments. And six different levels of transform for each attribute were applied to prepare test images. Total numbers of 36 rendered images were created. Two types of experiments, categorical judgment and pair comparison, were conducted and the total number of 63,648 observations (22,176 for categorical judgment and 41,472 for pair comparison) were performed over one month to accumulate the experimental data. Based on the categorical judgment, image quality model was derived using colour and texture as the image quality parameters. The performance of the model prediction has r value of 0.9923 which is good but still remains further study and modification.

Experimental

Setup

To collect subjective response data for the various image quality attributes, it is required to perform psychophysical experiments which consist of light source, stimuli, and observer. A BARCO Reference Calibrator®121 which is a CRT type monitor was used in a darkened room as a reference display device for the experiments. Spatial and temporal uniformity and the channel additivity of the monitor were tested before the experiments were conducted. The test results were found to be satisfactory for conducting psychophysical experiments. The gain-offset-gamma (GOG) model was used as the characterisation model for converting RGB to XYZ tristimulus values[5].

Two types of psychophysical experiments were carried out; pair comparison and categorical judgment. Pair comparison which uses paired streams in one display system was conducted for the evaluation of appearance difference between pairs of sample streams. And categorical judgment which displays single stream was also conducted for the evaluation of various image quality attributes.

Test Streams

Three MPEG test streams, Mobil, Susie, Table tennis were used in this study because these streams are official test streams for MPEG. Moreover they contain most attributes to be studied in colour imaging science. For example, Mobil was chosen as it contains high chroma, Susie contains skin tone and neutral background, and Table tennis contains high temporal frequency. Figure 1 shows the test streams used in this study.

Mobil　　Susie　　Table tennis

Figure 1 Test streams

The MPEG streams can be downloaded from the MPEG site. The playing time of the test streams were set to six seconds for all of the test streams in order to avoid observer fatigue during the psychophysical experiments. Seven image quality attributes were chosen in this study. Six of them (lightness, chroma, contrast, noise, sharpness and compression) are the conventional image quality attributes which have been used for the study of still images. The remaining one (temporal) is an image quality attribute for the study of moving stream. Six different levels of transforms for each image quality attribute were applied to prepare test streams. Total numbers of 129 (3 test streams * 6 levels * 7 attributes + 3 original streams) rendered streams were generated as test material. SONY Vegas® 6.0 editing software was used to create rendered frames. All of the image quality attributes can be manipulated using the SONY Vegas® 6.0 introduced in Figure 2.

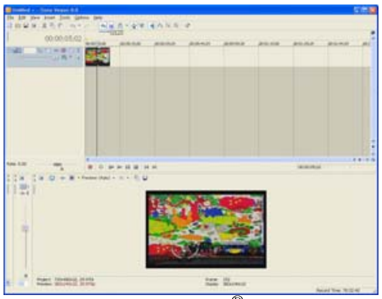


Figure 2 SONY Vegas® 6.0 Software

In each of 7 image quality attributes, the corresponding image transfer functions that the SONY Vegas® 6.0 software provides were used. The corresponding functions are summarised in Table 1.

Table 1 Image quality attributes and corresponding functions

IQ attributes	SONY Vegas functions - Video FX category-
Lightness	Brightness and contrast->Brighter->Brightness
Chroma	Saturation adjust->Reset to none->Amount
Contrast	Brightness and contrast->Brighter ->Contrast
Noise	Add noise->Color
Sharpness	Unsharp mask, Gaussian blur
Compression	Render as->Video->Cinepak format
Temporal	Render as avi->Frame rate

Procedure and Tool

Two types of psychophysical experiments were conducted in this study. To carry out the psychophysical experiments, the appropriate questionnaires should be prepared first. Different types of questionnaires were used to collect the observers' data which will be used to develop an image quality model. The corresponding questionnaires for each type of experiment are described in Table 2.

Table 2 Questions used in the experiments.

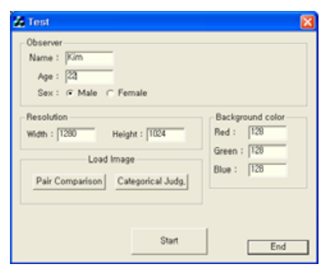
Experiment	Question
Categorical Judgment	1. How natural is the overall image? 2. How is the overall image quality? 3. How natural is the colour of the image? 4. How natural is the movement of the image?
Pair Comparison	1. Do they look the same in overall quality? 2. Do they look the same in colour? 3. Do they look the same in sharpness? 4. Do they look the same in contrast? 5. Do they look the same in noise? 6. Do they look the same in movement? 7. Do they look the same in smoothness?

14 observers consisting of 6 male and 8 female observers with ages from 22 to 31 participated in the categorical judgment experiments and 19 observers consisting of 11 male and 8 female observers with the same age distribution as categorical judgment participated in the pair comparison experiments. All the observers were tested Ishihara visual deficiency test and found to have normal colour vision.

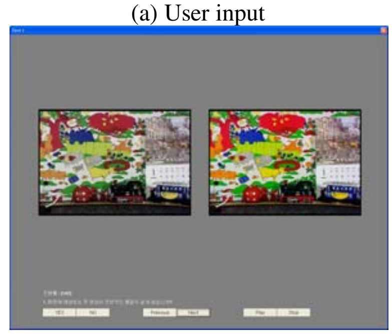
Total number of 1,176 observations (3 images × 7 attributes × 7 levels including original × 4 questions × 2 repeats) for the categorical judgment and 1,764 observations (3 images × 7 attributes × 6 levels × 7 questions × 2 repeats) for the pair comparison were carried out for each observer. For the categorical judgment experiment, single stream, either the original or one of the transformed ones, was displayed and played on a CRT in a random order, and observers were asked to assign one out of the nine equally stepped categories in which observers think the best answers to the questions. For the pair comparison experiment, the original and one of the transformed streams were displayed on a CRT and played while observers' decision for the question was completed.

One complete set of experiment was divided into four sessions so that the observation time for any one session did not exceed 45 minutes to avoid fatigue. In total, 49,980 observations (16,464 for categorical judgment and 33,516 for pair comparison) were performed over one month and the data were accumulated.

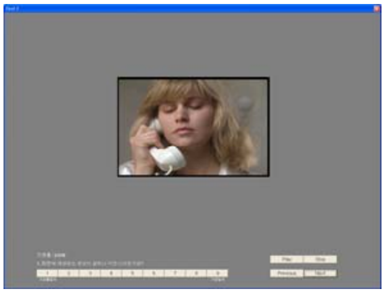
The software tool was developed to carry out psychophysical experiments. It consisted of three parts, user information input, pair comparison experiments, and categorical judgment experiments. In Figure 3, three screen shots showing the program is activating are introduced.



(a) User input



(b) Categorical judgment



(c) Pair comparison

Figure 3 Experimental software tool

Data Analysis

Before the development of image quality model, verification of the experimental data should be performed first so that the data can be used develop an image quality model with no statistical drawbacks. To do that, observer accuracy and repeatability were tested as the observer performance tests.

In categorical judgment (CJ), observer performance tests were done by calculating Pearson's correlation (r) and coefficient of variation (CV). In pair comparison (PC), wrong decision[6] analysis for the accuracy and repeatability were used. The results are shown in Table 3, 4 and 5 respectively.

Table 3 Observer performance (CJ)

r / CV	Observer Accuracy	Repeatability
Overall image naturalness	0.58 / 26	0.38 / 34
Overall image quality	0.63 / 28	0.45 / 35
Colour naturalness	0.57 / 26	0.43 / 32
Movement naturalness	0.64 / 24	0.48 / 29

Table 4 Observer repeatability (PC) – sorted by observers

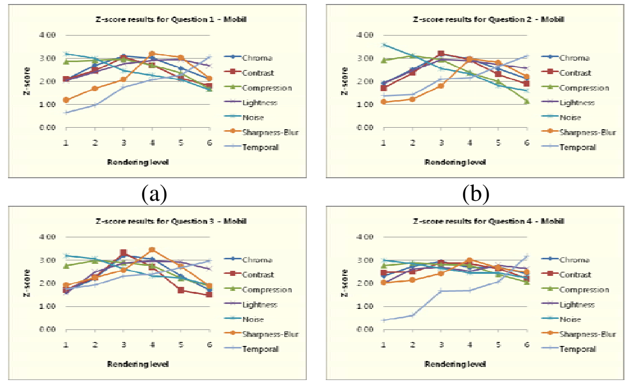
Average WD%	Min WD%	Max WD%
21	8	37

Table 5 Observer accuracy (PC) – sorted by observers

Phase	Average WD%	Min WD%	Max WD%
Phase I	21	14	30
Phase II	16	10	23

In comparison to the conventional results of the observer experiments for the stationary stimuli, the observer performance results are not quite good, and the reason can be inferred that observers are not familiar with those kinds of experiments. Therefore more training periods and the effective methods of training and experiments need to be studied and developed as the proceeding studies.

For the categorical judgment experiments, z-scores were calculated to evaluate the image quality of different levels of colour transformed streams. Figure 4 shows the z-score results of 'Mobil' for each question in categorical judgment. All of the above four plots show a similar characteristic that is observers judged level 3 and level 4 as the highest preference in case of chroma, contrast, lightness and sharpness-blur rendering attributes. In noise and compression rendering, level 1 had the highest z-scores whereas level 6 has the highest z-scores in temporal rendering. The rendering levels having the highest z-scores are those with similarly rendered as the original streams. In more detail, people prefer sharper streams than the original.

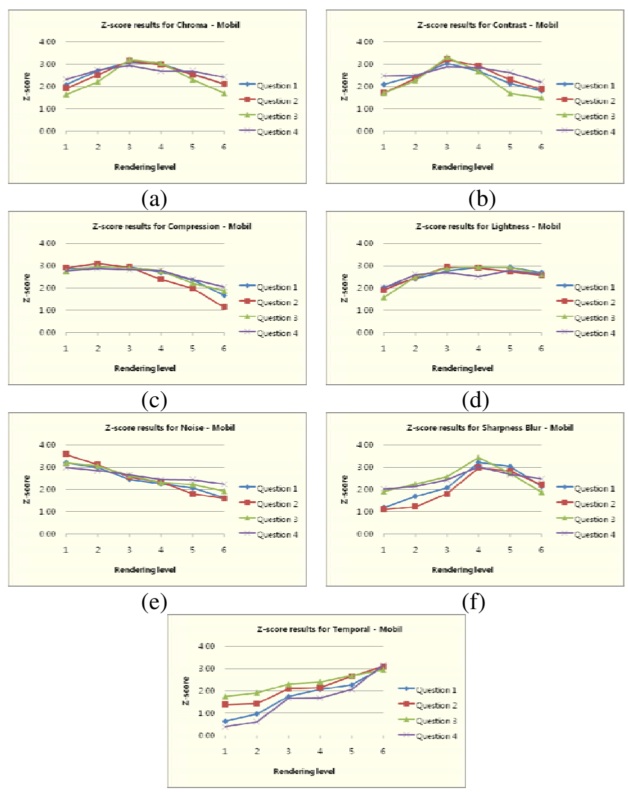


(a) (b) (c) (d)

Figure 4 Z-scores for categorical judgment (sorted by questions)

For the overall image quality, people prefer slightly less chromatic stream in case of slowly moving stimulus such as Mobil and Susie test streams whereas people prefer slightly more chromatic stream in case of fast moving stimulus such as Table tennis stream. For all of the test streams, observers could not determine their answers clearly to the questions of 'colour naturalness' and 'movement naturalness'. One possible reason is the observers were not familiar with those kinds of experiments. More training period may increase the discrimination between each rendering level.

Figure 5 shows the z-scores for the seven different types of rendering attributes for all of the four questions.

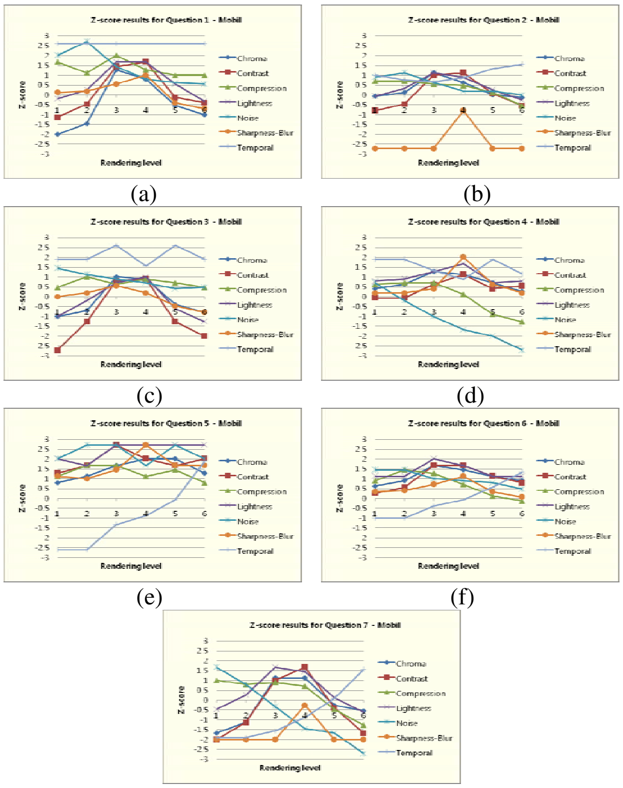


(a) (b) (c) (d) (e) (f) (g)

Figure 5 Z-scores for categorical judgment (sorted by attributes)

As can be seen from the Figure 5, the results are consistent with the results shown in Figure 4.

Similar data analysis was performed for the pair comparison. Figure 6 shows the z-score results for pair comparison of 'Mobil' stream. And Figure 7 shows the z-score results of seven image quality attributes for questions 1 to 7.

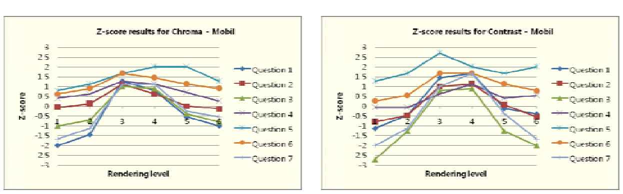


(a) (b) (c) (d) (e) (f) (g)

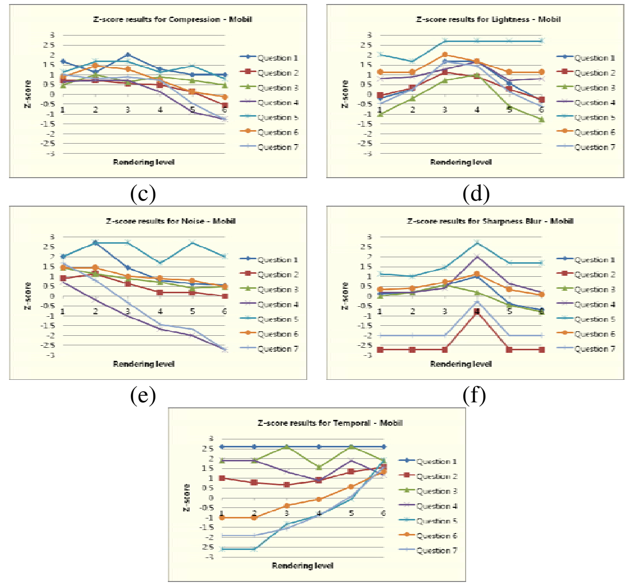
Figure 6 Z-scores for pair comparison (sort by questions)

It can be seen from Figure 6 that most of the rendering attributes have the similar shape with the case of categorical judgment. That is the peak z-scores locate around the middle rendering levels out of six except compression, noise and temporal attributes. Most observers couldn't discriminate the overall image quality difference in case of temporal attribute, so the resulting z-scores didn't change according to the change of rendering levels. From the z-score results for moving stream, it can be seen that observers' visual perception of moving stimuli is similar to that of stationary stimuli in most of the image quality attributes in which the conventional colour science have used. However in the attributes of movement and smoothness which are more related to the moving stimuli, observers tend to give higher z-scores when the contrast becomes stronger.

Z-cores for pair comparison results were also sorted by rendering attributes and Figure 7 shows the results.



(a) (b)



(c) (d) (e) (f) (g)

Figure 7 Z-scores for pair comparison (sort by attributes)

In Figure 7, overall characteristics are similar to Figure5 that is people prefer similarly rendered streams as the original. In chroma and lightness rendering, people prefer slightly less chromatic and darker streams when they answer most of the questions except the question of noise difference. This means people couldn't discriminate the noise difference when the stimulus is moving.

Model Development

To develop image quality model for moving streams, the most important image quality attributes should be determined first, then the development of a statistical optimisation model can be derived based on that. By keeping this in mind, determination of important image quality factors were performed by the calculation of inter comparison between image quality attributes applied to pair comparison using Pearson's correlation coefficient analysis. Table 6 shows the results.

Table 6 Inter comparison results

	IQ	Naturalness	Colour	Movement
IQ	1.00	**0.91**	**0.83**	0.73
Naturalness	0.91	1.00	0.74	0.89
Colour	0.**83**	0.74	1.00	0.53
Movement	0.73	0.**89**	0.53	1.00

In Table 6, naturalness has highest correlation with image quality (IQ) followed by colour and movement. Linear image quality model for moving streams is derived by the following form.

$$IQ = a + bN + cC + dM$$

where, IQ is overall image quality, N is naturalness, C is colour, M is movement, and a, b, c, d are coefficients. By using MATLAB optimisation tool, the coefficients were optimised with the multiple regression function and the final image quality model has the following equation.

$$IQ = 0.0417 + 0.9579N + 0.2316C - 0.2096M$$

Next, the performance of the model was tested and the result is shown in Figure 8.

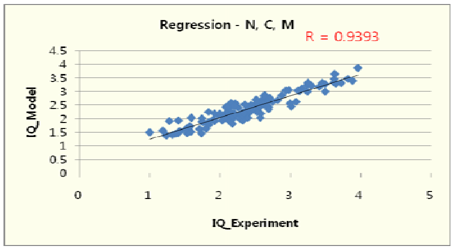


Figure 8 Model performance test – Complete model

In Figure 8, the y-axis represents the experimental data and x-axis represents the model prediction of the overall image quality. As can be seen from the figure, the simple linear model has high prediction accuracy with r value of 0.9393, therefore when predicting image quality of moving streams, the proposed linear model works fine.

To simplify the model, the least correlated attribute with IQ was determined to be removed. To do that, M attribute was removed and a new model had the following equation.

$$IQ = a + bN + cC$$

Again, multiple regression method was applied to determine the coefficients. The final image quality model had the following form.

$$IQ = -0.149 + 0.7587N + 0.2905C$$

Finally, image quality model for single most correlated attribute with IQ was derived by following the same methodologies described above. The result of optimisation had the form of,

$$IQ = 0.1245 + 0.9402N$$

The performances of the two simple image quality models are shown in Figure 9 and Figure 10 respectively.

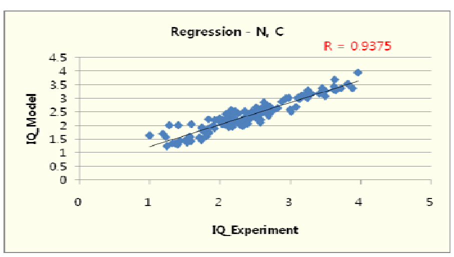


Figure 9 Model performance test – Naturalness, Colour

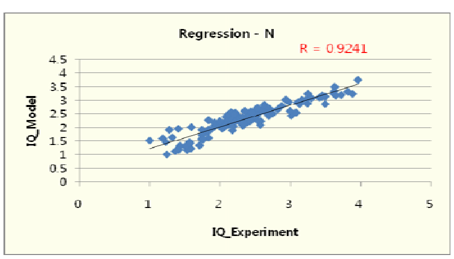


Figure 10 Model performance test – Naturalness

It can be seen from the performance results, simple version of image quality models have quite acceptable results with r value of 0.9375 in case of using naturalness and colour, and r value of 0.9241 in case of using only naturalness. And these results are close to the complete image quality model. Therefore image quality can be modelled by describing the single image quality attribute, naturalness.

Conclusion

Psychophysical experiments were conducted to develop image quality model for moving streams. According to the data analysis, different observer characteristics were found compared to the data analysis of still images. The reason is not clear and further analysis is required. For example, methodology for applying conventional colour difference formulae for stationary stimuli to the moving stimuli needs to be studied. Also new approach for the calculation of differences between original and rendered streams is required to develop reliable image quality model. In this paper, three types of linear image quality models were derived according to the observer data and the models had the reasonable performance of prediction with the r values of 0.9393 for complete IQ model, 0.9375 for removing M, and 0.9241 for removing C and M. However, there remain further improvements because only limited image quality attributes were used in this study. Further development of the model is needed as the future study. The future study will also cover the study of extending colour appearance model for predicting image quality for images and streams.

References

[1] Moroney N., Fairchild M.D., Hunt R.W.G., Li C., Luo M.R., Newman T.: The CIECAM02 Colour Appearance Model, Proceedings of the tenth Colour Imaging Conference. IS&T/SID, Scottsdale, Arizona, (2002).

[2] Fairchild M.D, Johnson G.: Meet iCAM: A next-generation colour appearance model, Proceedings of the tenth Colour Imaging Conference, IS&T/SID, Scottsdale, Arizona, (2002)

[3] Yendrikhovskij S.: Towards perceptually optimal colour reproduction of natural scenes, Colour Imaging Vision and Technology, Chapter 18, (Wiley, 1999)

[4] J-S Kim, M-S Cho, S.Westland and M.R.Luo, Image quality assessments for photographic images, AIC Colour 05, (2005)

[5] Berns S.: Methods for characterizing CRT displays, Displays vol. 6, no.4, (1996)

[6] Y.J.Kim, M.R.Luo, P.Rhodes, W.Choe,S.Lee, and C.Kim, Affective attributes in image quality of a mobile LCD, Proc. Third European Conference on Colour in Graphics, Imaging, and Vision, (2006)

Author Biography

Jin-Seo, Kim received his M.S. degree in electrical engineering from Polytechnic University, Brooklyn, NY, USA, in 1993, and he is a Ph.D student in department of colour science, the University of Leeds, UK. He is a senior research engineer at ETRI since 1993. His research interest includes colour science, colour reproduction, colour management system, and digital cinema