Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

Social Image Quality

Guoping Qiu and Ahmed Kheiri
School of Computer Science, University of Nottingham, United Kingdom

ABSTRACT

Current subjective image quality assessments have been developed in the laboratory environments, under controlled-conditions, and are dependent on the participation of limited numbers of observers. In this research, with the help of Web 2.0 and social media technology, a new method for building a subjective image quality metric has been developed where the observers are the Internet users. A website with a simple user interface that enables Internet users from anywhere at any time to vote for a better quality version of a pair of the same image has been constructed. Users' votes are recorded and used to rank the images according to their perceived visual qualities. We have developed three rank aggregation algorithms to process the recorded pair comparison data, the first uses a naive approach, the second employs a Condorcet method, and the third uses the Dykstra's extension of Bradley-Terry method. The website has been collecting data for about three months and has accumulated over 10,000 votes at the time of writing this paper. Results show that the Internet and its allied technologies such as crowdsourcing offer a promising new paradigm for image and video quality assessment where hundreds of thousands of Internet users can contribute to building more robust image quality metrics. We have made Internet user generated social image quality (SIQ) data of a public image database available online (http://www.hdri.cs.nott.ac.uk/siq/) to provide the image quality research community with a new source of ground truth data. The website continues to collect votes and will include more public image databases and will also be extended to include videos to collect social video quality (SVQ) data. All data will be public available on the website in due course.

Keywords: image and video quality, image quality metric, Web2.0, social media, crowd sourcing, paired comparison, psychometric, rank aggregation

1. INTRODUCTION

Up to now, subjective image quality assessments have been done in laboratory environments, under controlled-conditions, based on limited numbers of observer participation after they underwent a training phase [5, 9, 11, 15, 17]. This is unsatisfactory because, firstly, in everyday image/photo viewing activities, such as browsing the Web, viewing family albums or view images in any other occasions, the environments under which a viewer is viewing images, such as lighting conditions and the display equipments used to display the images can vary hugely and differ significantly from typical laboratory settings; secondary, limited numbers of images being used for building the image quality matrices, ranging from several [5] to less than two thousand [15] is dwarfed by the amount of images on the Internet[1]; and the number of observers in these studies is again minuscule compared with the huge numbers of Internet users worldwide; and finally, depending on limited number of observers will increase the load on them and that could increase the probability of low attention level.

With the advent of Web 2.0 and social media technology [26], millions of Internet users are no longer passive consumers of information but rather they are now active participants contributing contents and knowledge to the Internet. These technologies enable the Internet to collect user inputs thus potentially we can use the Internet as a huge knowledge acquisition system. For example, there have already been much research in the computer vision and multimedia communities that exploits user contributed image tags to automatically annotate images [6, 27]. Other user contributed knowledge, for example, people's shopping habits are being used by Internet vendors such as amazon.com to recommend products to other shoppers.

1 According to http://techcrunch.com/2008/11/03/three-billion-photos-at-flickr/, there have been more than three billion photos uploaded to the photo sharing site Flickr as of November 2008 and illions more are being added per month. The same webpage further reported that Facebook had 10 billion photos.

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

The objective of this research is to exploit Web 2.0 and social media technology for building image quality metric. The ultimate goal can be regarded as collecting sufficiently large number of data where people from anywhere can rate the images at anytime to build a social image quality metric (SIQ).

2. RELATED WORK

From the beginning of the 21st century, the use of digital images as a means for communicating and representing information has been grown tremendously. As a result of that, computer scientists started to explore the methods for maintaining and improving the appearance of images that are processed. Nevertheless, the quality of images, whether they processed or not, is rarely perfect [19]. Images are subject to distortions during compression, synthesis, acquisition, restoration, transmission, enhancement, processing, and reproduction. In order to enhance, control or maintain the quality of images, it is important to identify and quantify image quality degradations through image quality evaluation [2, 19].

Current image quality metrics try to match the Human Visual System (HVS); and as there is no reliable mathematical model for the HVS, it's difficult to define an optimum image quality metric that perfectly match the HVS. Such metrics are classified as an objective quality metrics. A challenging task is how to evaluate these objective metrics. Usually this is performed using databases contain many distorted images for which the subjective mean opinion scores (MOS) of image quality have been experimentally collected and then correlating the objective metric scores with the subjective MOS's and if there is any improvement of the correlation between them, then it would be a proof of the success of the objective metric. IVC [11], TID2008 [15] and A57 [5] are examples of subjective databases publically available on the Internet. Figure 1 shows the procedures of evaluation the objective image quality metrics.

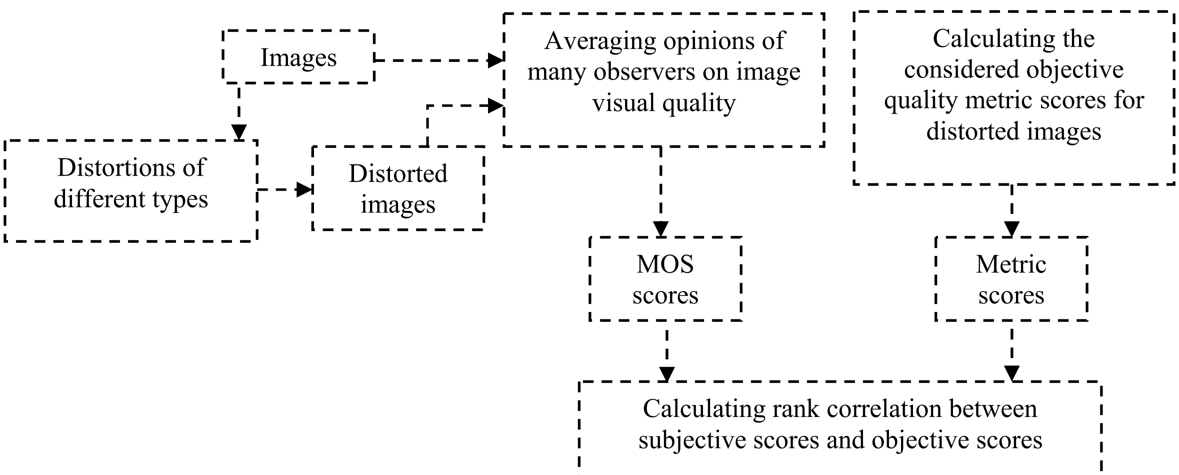

Figure 1 Creation and use of test image database for evaluation of objective image quality metrics

There are different methodologies to obtain the subjective image quality values of MOS. Depending on the strategy, the observers will be asked to evaluate the absolute quality of the image or its similarity to the reference image. In both cases the subjective evaluation will be expressed in discrete or continuous, categorical or numerical grading scale [9, 15 and 17]. As an example, in [17] five gradations with five categories, "Bad", "Poor", "Fair", "Good" and "Excellent", have been used. However, the main drawback could arise when the observer assigned "Bad" score to one distorted image but later the observer found another distorted image that is even worse. But in the used scale there is no gradation worse than "Bad"; and moreover it is often not allowed to change the previously given grade. To overcome this problem, [9] suggests to give instructions in the written form to the observers, and then they undergo a training phase; this helps the observers to get an idea of what is "Bad" and "Excellent" quality.

The methodology used by [5] to obtain the subjective ratings for the distorted images of the images in A57 database is done by placing the images on a table; the original was fixed in a position at one end of the table, and observers were instructed to position the distorted images such that the physical displacement between each distorted image and the original was linearly proportional to their subjective assessment of distortion. Thus, images which were placed further

7867 - 28 V. 1 (p.2 of 10) / Color: No / Format: A4 / Date: 2010-11-14 05:05:50 AM

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

away from the reference image were judged as the worst distorted image relative to the original. The subjective ratings for the A57 database were obtained from only seven expert observers.

Reference [15] proposed another methodology by introducing to the observers (researchers, tutors and students) two distorted images and the original image and asked them to select the distorted image that visually differs less from the reference one. The image pointed as a winner will get one point. The authors of [15] used this methodology to calculate the MOS values for each of the 68 distorted images of the images in the TID2008 database. Each distorted image participated in nine comparisons during each experiment and so the number of times each observer would ask to select the image with the higher quality between two images was 9*(68/2) which is 306 times. Points were summed up for each distorted image and, thus, each distorted image could get from 0 to 9 points in one experiment. For comparison, only representatives of the same point groups were used for comparison, making totally nine tours and that's why each distorted images participated in nine comparisons. The work of [15] also followed the suggestion by [9] of giving instructions and training to the observers before carrying out the actual experiments.

3. SOCIAL IMAGE QUALITY METRIC

The methodology for carrying out subjective tests is based on the Web 2.0 and social media technology; and so the observers who participate in building the metric are the Internet users with different cultural backgrounds from anywhere and can do this at anytime. This is to overcome the drawback of the current subjective metrics where there were limited numbers of observers. Moreover and as current subjective metrics gives instructions in the written form to the observers and then they undergo a training phase, the proposed Social Image Quality (SIQ) metric has been designed in a simple way where it does not need to have a written instruction form to the observers or make them to undergo a training phase. Finally, and unlike the traditional subjective image quality metrics with limited number of observers, there is no specific time required from the observer to perform the experiment in the SIQ, this would help to reduce the load on the participants.

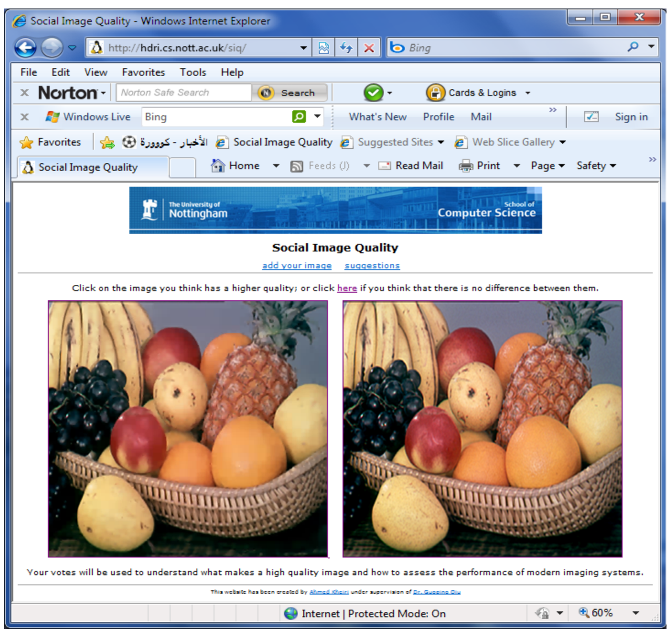

Figure 2. A screen shot of the SIQ server viewable at http://www.hdri.cs.nott.ac.uk/siq/

The way that data has been collected from participants is done by constructing a website called "Social Image Quality" or SIQ for short. At any given time, the website displays two versions of the same image of different qualities, generated by using different lossy compression factors or processed in any other different ways. The website asks the users a very simple question: "which image has a higher quality" or "I cannot tell the difference". The user simply clicks on the version that he/she thinks is the better one or click on a "no difference" link (see Figure 2). Those user inputs are then recorded in a MySQL server. Based on the pair-comparison results that have been collected, three different ranking

7867 - 28 V. 1 (p.3 of 10) / Color: No / Format: A4 / Date: 2010-11-14 05:05:50 AM

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

algorithms are implemented to rank all versions of images according to their perceived visual qualities. The first algorithm uses the naive approach [10], the second one employs a Condorcet method [7, 12, 14, 21, 22], and the last one uses the Dykstra's extension of Bradley-Terry model [3, 8]. With such ranking results, which are based on contributions from, potentially, millions of Internet users can then be used to be a social image quality metric.

The URL of the SIQ server is http://www.hdri.cs.nott.ac.uk/siq/, which is currently accessible by Internet users. For our server, the images have been taken from the standard IVC database [11], A57 database [5] and TID2008 database [15].

4. QUALITY RANKING

Based on the pair-comparison results that have been collected, three different ranking algorithms are implemented to rank all versions of images according to their perceived visual qualities.

Since there is a possibility that the outcome of the comparison could be "no difference" (that is, "draw" or "tie"), then this should be included in the methods. In the following methods it is assumed that a "draw" count as half a win and half a loss [4].

4.1 Naive Approach

This is the simplest method to rank the image qualities. It uses the following formula for pair-wise comparisons [10]:

$$win\_rate_i = \frac{\sum_{m:m \neq i} w_{im}}{\sum_{m:m \neq i} w_{im} + \sum_{m:m \neq i} w_{mi}}$$

Where w_{im} is the number of times that image i beats the image m.

$$\sum_{m:m \neq i} w_{im} + \sum_{m:m \neq i} w_{mi} > 0$$

4.2 Condorcet Method

Condorcet method is the implementation of the method of pair-wise comparison that Marquis de Condorcet had devised [14]. This simple method has been selected to rank the image qualities. Simply the image that defeats every other image is the winner. An image defeats another image if a majority of participants rank it higher on their voting than the other image [21]. The following pseudo-code shows how to find the majority runoff [12]:

```
count = 0
for each of the p participants pi  do
      If pi ranks i1 above i2, count++
      If pi ranks i2 above i1, count--
If count > 0, rank i1 better than i2
Else rank i2 better than i1
```

The differences between Condorcet methods occur in situations where no image is undefeated, implying that there exists a cycle of images that defeat one another, called a Condorcet paradox [21]. To overcome this problem, many versions of Condorcet method have been proposed. A simple version is Copeland's method, where after the participants pick on the image that they think has a higher quality in pair-wise elections, the method will then look for the image that has the most wins [7, 22].

7867 - 28 V. 1 (p.4 of 10) / Color: No / Format: A4 / Date: 2010-11-14 05:05:50 AM

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

4.3 Dykstra's Extension of Bradley-Terry Model

Reference [3] proposed a method of analysis for paired comparisons, [8] proposed a method that extends the Bradley-Terry model by assuming that the numbers of comparisons between the objects are unequal. This is important in our case since the number of participants and hence the number of comparisons is unknown.

Mathematically, let's assume that we have n images, $I_1 \cdots I_n$, and each pair of different quality of the same image is compared c_{ij} times where $c_{ij} \geq 0$ and $i < j$. Bradley & Terry assumed that each image I_i has a parameter, true rating π_i, such that $\sum_{i=1}^{n} \pi_i = 1$ and $\pi_i \geq 0$ for each i; and also they assumed that the probability of an image I_i being preferred to the image I_j is given by the following equation: $P_{ij} = \pi_i / (\pi_i + \pi_j)$.

The last equation described above is equivalent to $P_{ij} / P_{ji} = \pi_i / \pi_j$, that is the ratio of the number of times that the image I_i is preferred to the image I_j to the number of times that image I_j is preferred to image I_i, P_{ij} / P_{ji}, should equal to the ratio of true rating π_i to true rating π_j (π_i / π_j).

Bradley & Terry assumed that there are equal number of repetitions on each of the $n \cdot (n-1)/2$ possible pairs (i, j). However, Dykstra assumed in his model that there are unequal numbers of repetitions on the pairs, that is, c_{ij} is not the same for each of the $n \cdot (n-1)/2$ possible pairs (i, j). The model that gives the probability of the observed result in c_{ij} repetitions on the comparisons of the images i and j is given by:

$$\left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{a_{ij}} \left(\frac{\pi_j}{\pi_i + \pi_j}\right)^{a_{ji}}$$

where a_{ij} is the number of times of image I_i being preferred over the image I_j out of c_{ij} times. $a_{ij} + a_{ji} = c_{ij}$

Multiplying the appropriate expressions for all repetitions of all $n(n-1)/2$ possible pairs we obtain the expression $L(\pi_i)$ for the general likelihood function [3, 8], thus:

$$L(\pi_i) = \prod_{i=1}^{n} \pi_i^{r_i} \prod_{i<j}^{n} (\pi_i + \pi_j)^{-c_{ij}}$$

where r_i is the total number of times that image I_i is preferred in the whole comparisons. In other words, $r_i = \sum_{j \neq i} a_{ij}$

To get the maximum likelihood estimate b_i of π_i, the natural logarithm of the general likelihood function should be differentiated with respect to the π_i and then setting the result to zero [3, 8]. That means:

$$\frac{\partial(\log L(\pi_i))}{\partial \pi_i} = 0 \Rightarrow (r_i / b_i) - \sum_{j \neq i} [c_{ij} / (b_i + b_j)] = 0$$

For the purpose of computing the maximum likelihood estimates $b_i's$, we use the following equation:

7867 - 28 V. 1 (p.5 of 10) / Color: No / Format: A4 / Date: 2010-11-14 05:05:50 AM

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

$$b_i^{t+1} = \frac{r_i}{\sum_{j \neq i}^{n} \frac{c_{ij}}{b_i^t + b_j^t}}$$

The values of $b_i's$ are calculated iteratively until the difference between two consequences is sufficiently close.

We have to choose initial values for $b_i's$ to start the iteration. The resulting first estimates are substituted into the right side of the above equation and second estimates are then obtained, the second estimates being resubstituted, and so on, until the equalities hold. Dykstra suggested that reasonable first estimates may be obtained by assuming that the $b_i's$ are not too different from each other.

Dykstra assumed that $b_j^0 = b'$ for all $j \neq i$ then $b' = (1 - b_i^0)/(n-1)$.

$$\Rightarrow b_i^0 = \frac{r_i}{\sum_{j \neq i}^{n} \frac{c_{ij}}{b_i^0 + (1 - b_i^0)/(n-1)}} \Rightarrow b_i^0 = \frac{r_i}{(n-1)\sum_{j \neq i}^{n} \frac{c_{ij}}{(n-2)b_i^0 + 1}} \Rightarrow b_i^0 = \frac{r_i[(n-2)b_i^0 + 1]}{(n-1)\sum_{j \neq i}^{n} c_{ij}}$$

Therefore:

$$b_i^0 = \frac{r_i}{(n-1)\sum_{j \neq i}^{n} c_{ij} - (n-2)r_i}$$

However, for simplicity this assumption has been considered, in this project, to calculate the $b_i's$; and then we need to normalize those values to satisfy the condition $\sum_{i=1}^{n} b_i = 1$.

5. RESULTS

5.1 Ranking Algorithms

The Naive Approach calculates the win_rate to each image by looking for the total number that the image won in the votes. Therefore, this approach does not consider opponents' image qualities.

The Condorcet Method considers the direct paired comparisons; but still it has a drawback since it does not consider the indirect paired comparisons. As an example to that, imagine that we have four objects A, B, C and D. A defeats B and C, B defeats C, C defeats D and D defeats A. According to the Condorcet Method and more particularly the Copeland method, A is the winner. The method wouldn't consider the fact that A defeated by D.

The Dykstra's extension of Bradley-Terry model algorithm differs from that of the Condorcet Method algorithm in that it doesn't simply consider the direct paired comparisons but also consider the indirect paired comparisons. This algorithm is much complicated and it works slowly in terms of complexity comparing to the other methods as it needs a lot of calculations to rank the images. However and from the definitions of the three ranking algorithms, Dykstra method has the best performance to rank the images among the three ranking algorithms. In the following sections, mathematical performance measures have been implemented to make sure that the Dykstra method has the best performance.

5.1.1 Kendall Tau Rank Correlation Coefficient

The Kendall tau rank correlation coefficient evaluates the degree of similarity of the orderings of the data when ranked by different methods [1]. Table 1 presents the average correlation coefficient for the whole groups of images used in the experiment. (The results on Table 1 are based on the data collected from the SIQ server on 24th, August 2010).

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

Table 1. Kendall tau correlation coefficients for each two ranking approaches.

	Naive Approach	Condorcet Method	Dykstra Method
Naive Approach	1.000	0.875	0.947
Condorcet Method	0.875	1.000	0.882
Dykstra Method	0.947	0.882	1.000

From Table 1, the large correlation coefficient values of 0.947, 0.875 and 0.882 demonstrate that the three different approaches lead to very similar rankings.

5.1.2 Violations and Hits

Let's assume that we have a set of objects and those objects are ranked using a specific algorithm. Then for each pair of objects x and y, the Violation occurred when x ranked better than y but y beat x; and the Hit occurred when x ranked better than y and x beat y. A good ranking algorithm should make many hits while causing few violations, so a good algorithm would minimize the evaluation criterion value (number of violations/number of hits) [10].

Table 2 shows the number of violations, number of hits and the ratio number of violations/number of hits of each of the three ranking approaches. (The results on Table 5.2 are based on the data collected from the SIQ server on 24th, August 2010).

Table 2. Number of violations, number of hits and number of violations/hits for each ranking approaches.

Violations and Hits	Naive Approach	Condorcet Method	Dykstra Method
Number of Violations	196	373	157
Number of Hits	2838	2661	2877
Ratio (#Violations/#Hits)	0.069	0.140	0.055

From Table 2, it is clearly observable that Dykstra Method performs better than the two others approaches; while Condorcet Methods is the worst since it has the largest number of violations and the worst ratio of number of violations to the number of hits.

5.2 SIQ Metric

The analysis part of comparing the three different ranking approaches showed that the Dykstra Method performs better than the Naïve Approach and the Condorcet Method. Accordingly, the scores obtained by Dykstra Method have been selected to be the Social Image Quality (SIQ) metric.

5.2.1 Evaluation of SIQ for Images Taken from the IVC Database

To evaluate the SIQ metric, its values are recorded for each distorted images [http://hdri.cs.nott.ac.uk/siq/score.xls] and then compared to the corresponding MOS scores using the Spearman rank correlation.

The Spearman's rank correlation coefficient is a statistical measure that assesses the degree of the monotonic relationship between two variables [20, 25]. The online software [20] has been used to calculate the Spearman rank correlation for the SIQ with MOS. The performance of the SIQ metric in term of the Spearman correlation has the correlation value of 0.97 which is very high correlation and it demonstrates that both methodologies MOS and SIQ lead to very similar results.

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

Now, to plot the MOS against SIQ, a psychometric function should be used in order to transform SIQ values S in the range of MOS. This methodology is recommended and approved by [18]. The psychometric function used in our case is the function with 3 parameters given by the equation:

$$f(S) = \frac{b_1}{1 + e^{-b_2(S - b_3)}}$$

An Excel solver has been used to obtain the three parameters of the psychometric function which are learned by minimizing the Root Mean Square Error (RMSE) between the transformed metric outputs f(S) and the corresponding MOS. Figure 3 depicts scatter plot of the MOS values (on the vertical axis) plotted against transformed SIQ output (on the horizontal axis).

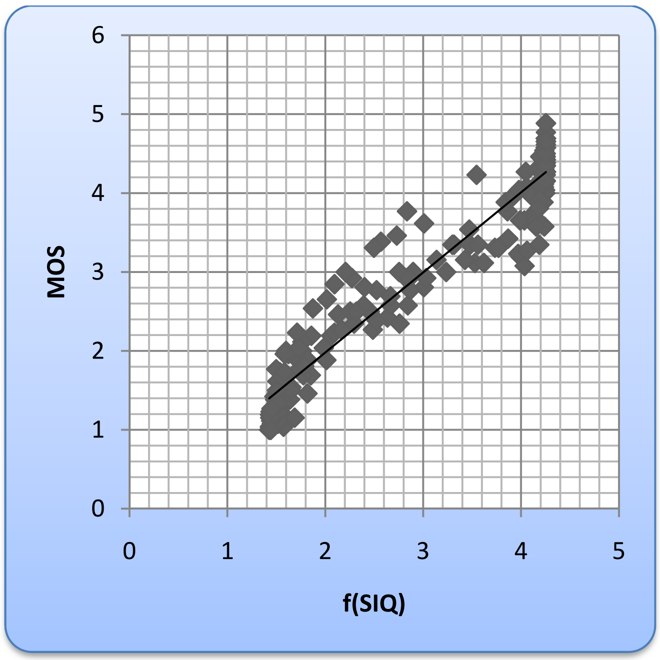

Figure 3. MOS values plotted against transformed SIQ values for the images from IVC database, each sample point represents one image

From Figure 3, it is clearly observable that the data-points are densely scattered around the fitted line. From Figure 4, the images that have been taken from IVC database give excellent results where the transformed SIQ and the MOS almost overlap.

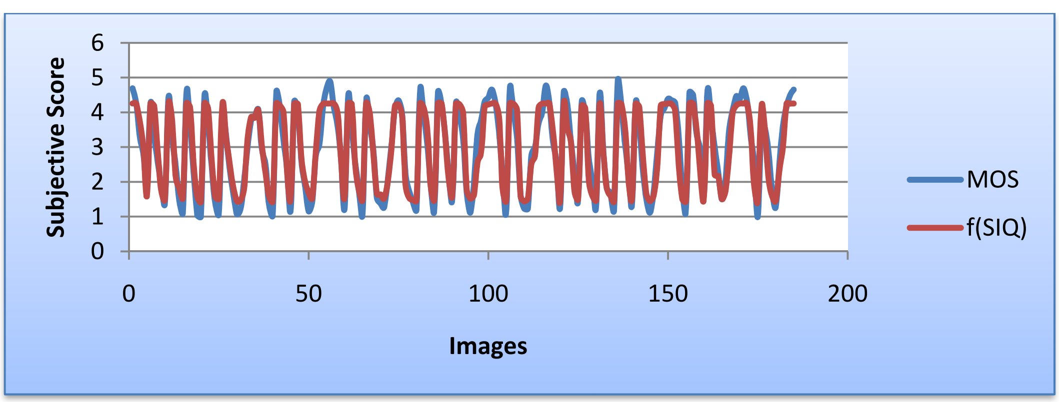

Figure 4. MOS and transformed SIQ for the IVC database (185 images)

7867 - 28 V. 1 (p.8 of 10) / Color: No / Format: A4 / Date: 2010-11-14 05:05:50 AM

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

Pearson correlation coefficient is another performance measure has been used in the experiment. It measures the prediction accuracy by reflecting the degree of linear relationship between the two metrics [20, 23]. The online software [20] has been used to calculate the Pearson's correlation coefficient for the SIQ with MOS. The performance of the SIQ metric has the correlation value of 0.96 which is very high correlation.

Another measure used to evaluate the SIQ with MOS is the Outlier Ratio (OR), which measures the consistency of the SIQ after mapping the SIQ values in the range of MOS values. The outlier ratio is define as (OR = nfalse/n) where nfalse is the number of SIQ values outside twice of the standard deviations of the MOS values; and n is the total number of objective quality metric values. OR indicates how often an algorithm predicts subjective quality values within a given range [13, 18]. The threshold range of the twice of the standard deviations of the MOS for defining outliers has been chosen as recommended by [18]. From the definition, the smaller OR is the more consistent SIQ will be with regard to MOS. An Excel sheet has been used to find the OR value. It gives a value of 0/185 (%0.0) and that shows a perfect consistent SIQ with regard to MOS.

6. CONCLUDING REMARKS

This research presented a new subjective image quality metric called social image quality (SIQ). The methodology followed for carrying out the subjective tests is done by using the Web 2.0 and social media technology. The SIQ provided very similar results to the currently used subjective image quality metric (MOS).

To the best of our knowledge, the work presented in this research is the first work that employs Web 2.0 and social media principles for image quality assessment and for building image quality metrics. This approach to harnessing the power of vast number of Internet users and collecting very large amount of data in a variety of realistic viewing conditions will likely have the potential of transforming the way image quality is assessed and collecting very large volume of data for building better image quality metrics. More image databases and video data will be added to the website in the near future and data will again be made available online.

The SIQ data for the IVC database is available online: http://hdri.cs.nott.ac.uk/siq/score.xls

REFERENCES

[1] Abdi, H., "The Kendall Rank Correlation Coefficient," Retrieved August 2, 2010, from http://www.utdallas.edu/~herve/Abdi-KendallCorrelation2007-pretty.pdf.

[2] Aydin, T. O., Mantiuk, R., Myszkowski, K., Seidel, H.-P. "Dynamic Range Independent Image Quality Assessment," ACM SIGGRAPH 2008 Papers held in Los Angeles, California, SIGGRAPH '08, pp. 1-10, New York, NY, USA: ACM, (2008).

[3] Bradley, R. A. and Terry, M. E. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," Biometrika, 39(3/4), 324-345 (1952).

[4] Butler, K. and Whelan, J. T. "The Existence of Maximum Likelihood Estimates in the Bradley-Terry Model and its Extensions," ArXiv Mathematics e-prints, (2004).

[5] Chandler, D. M. and Hemami, S. S Online Supplement to "VSNR: a Visual Signal-to-Noise Ratio for Natural Images Based on Near-Threshold and Suprathreshold Vision," Retrieved July 15, 2010, from http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.

[6] Chua, T.-S., Tang, J., Hong, R., Li, H., Luo, Z. and Zheng, Y. "NUS-WIDE: A Real-world Web Image Database from National University of Singapore," In Proceedings of ACM Conference on Image and Video Retrieval held in Santorini, Fira, Greece, CIVR 09, pp. 1-9, New York, NY, USA: ACM, (2009).

[7] Crowley, M. "Election Theory: How Voters and Parties Behave Strategically in Democratic Systems," Retrieved July 1, 2010, from http://www.cs.ubc.ca/~crowley/academia/papers/gtdt-election-jan302006.pdf.

[8] Dykstra, O. J. "Rank Analysis of Incomplete Block Designs: A Method of Paired Comparisons Employing Unequal Repetitions on Pairs," Biometrics, 16(2), 176-188 (1960).

[9] Egiazarian, K., Astola, J., Ponomarenko, N., Lukin, V., Battisti, F. and Carli, M. "Two New Full-Reference Quality Metrics Based on HVS," Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona, USA, 4 p (2006).

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are acceptable, (3) all fonts and special characters are correct, and (4) all text and figures fit within the margin lines shown on this review document. Return to your MySPIE ToDo list and approve or disapprove this submission.

[10] Huang, T.-K., Lin, C.-J. and Weng, R. C. "Ranking Individuals by Group Comparisons," Journal of Machine Learning Research, 9, 2187-2216 (2008).
[11] Le Callet, P. and Autrusseau, F. "Subjective Quality Assessment IRCCyN/IVC Database," Retrieved June 4, 2010, from http://www.irccyn.ec-nantes.fr/ivcdb/.
[12] Montague, M. and Aslam, J. "Condorcet Fusion for Improved Retrieval," In Proceedings of the Eleventh International Conference on Information and Knowledge Management held in McLean, Virginia, USA, CIKM '02, pp. 538-548, New York, NY, USA: ACM, (2002).
[13] Ninassi, A., Le Callet, P. and Autrusseau, F. "Pseudo No Reference Image Quality Metric Using Perceptual Data Hiding," In proceedings of SPIE Human Vision and Electronic Imaging, San Jose, CA, USA, HVEI'06, Vol. 6057, pp. 146-157 (2006).
[14] O'Connor, J. J. and Robertson, E. F. "Marie Jean Antoine Nicolas de Caritat Condorcet," Retrieved July 1, 2010, from The MacTutor History of Mathematics Archive: http://www-history.mcs.st-andrews.ac.uk/Mathematicians/Condorcet.html.
[15] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M. and Battisti, F. "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," Advances of Modern Radioelectronics, Vol. 10, pp. 30–45. Retrieved July 17, 2010, from http://www.ponomarenko.info/tid2008.htm.
[16] Seshadrinathan, K., Soundararajan, R., Bovik, A. C. and Cormack, L. K. "Study of Subjective and Objective Quality Assessment of Video," IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1427–1441 (2010).
[17] Sheikh, H. R., Sabir, M. F. and Bovik, A. C. "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3441-3452 (2006).
[18] VQEG "Final report from the video quality experts group on the validation of objective models of video quality assessment". Retrived August 26, 2010, from http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseII/downloads/VQEGII_Final_Report.pdf.
[19] Wang, Z. and Bovik, A. C. "Modern Image Quality Assessment," Synthesis Lectures on Image, Video, and Multimedia Processing, 2(1), 1-156 (2006).
[20] Wessa, P. "Free Statistics Software, Office for Research Development and Education, version 1.1.23-r6," Retrieved August 30, 2010, from http://www.wessa.net/.
[21] Wikipedia, "Condorcet Method," Retrieved June 28, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Condorcet_method.
[22] Wikipedia, "Copeland's Method," Retrieved July 1, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Copeland%27s_method.
[23] Wikipedia, "Pearson Product-Moment Correlation Coefficient," Retrieved August 30, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Pearson_correlation_coefficient.
[24] Wikipedia, "Root Mean Square Deviation," Retrieved August 28, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Root_mean_square_deviation.
[25] Wikipedia, "Spearman's Rank Correlation Coefficient," Retrieved August 30, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient.
[26] Wikipedia, "Web 2.0," Retrieved June 28, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Web_2.0.
[27] Zhou, N., Cheung, W., Qiu, G. and Xue, X. "A Hybrid Probabilistic Model for Unified Collaborative and Content-based Image Tagging," IEEE Transactions on Pattern Analysis and Machine Intelligence, to appear.