# The Challenges in Modeling Image Quality in Digital Printing

***Raisa Halonen[1], Tuomas Leisti[2] and Pirkko Oittinen[1]***
***[1]Dept. of Media Technology, Helsinki Univ. of Technology, Espoo, Finland [2]Dept. of Psychology, Univ. of Helsinki, Helsinki, Finland***

## Abstract

*This study is a part of research which aims to develop a model for the visual quality of digitally printed photographs from the paper standpoint. The ongoing research has brought up many challenges in finding a universal model, such as the differences between printing methods, the contribution of paper structures and the role of image content. In this paper, the challenges are approached by analyzing the need for discrete quality models for different printing methods, papers of different quality level, and different types of image content.*

## Introduction

There is wide agreement that visually assessed quality represents the "ground truth" of image quality [1]. To get reliable visual quality assessment data is, however, time consuming. Replacing visual assessment with models which predict visual scores (such as mean opinion scores, MOS) has a high demand. Uses of such models include development of visual technologies, specification of the quality level for commercial purposes and image retrieval. Visual hardware technologies of interest currently include mobile and 3D cameras and displays, and digital printing while software technologies include image coding and processing.

Quality modeling may be discussed using a layer representation such as the hierarchy relevant for the current study (Table 1). Modeling may try to tackle all the layers from Layer 1 to Layer 5 or, say, relations within and between two adjacent layers. Relating Layer 1 factors to Layer 2 measurements, i.e. physical characteristics of paper and printing devices to instrumental measurements of print quality and further to higher level attributes, is of course a specific challenge of digital printing. It is made complex by the fact that the major digital printing methods, ink-jet and electrophotography, differ in terms of the mechanisms by which the images are formed. It has for instance the consequence that variation of a given paper property may have an adverse influence on print characteristics in the two methods.

The challenges of modeling quality of complex images across Layers 3…5 are very much similar in all technologies. Today there is quite good understanding of Layer 3 attributes. They include noise, contrast, sharpness and colorfulness. Several algorithms for the first three have been suggested with computation from the luminance channel. The performance of the algorithms - originally designed for digital images - has been found lacking in digital printing due to the prominence of noise [2].

The concept of the fourth factor, colorfulness, is less clear. In a no-reference quality modeling situation of printed image quality (as studied here) the interpretation that more colorful means better quality may function because the color gamut tends to be reduced when print quality diminishes. The assumption may not hold for image quality across image contents, and other criteria need to be found such as supported by a reduced reference approach [3].

The issue what the Layer 4 attributes are is still open. Their study is still in its infancy [4, 5] and requires that a context for image use is defined. Naturalness and usefulness are often mentioned [6], but also others have been identified [7].

Layer 4 issues are beyond the scope of the current study. It has the purpose of mapping the challenges of quality modeling by analyzing data on the relations of low level computational and visual attributes (Layer 3) and overall image quality MOS (Layer 5). This is for the purpose of developing a perceptually meaningful image quality index for digital printing in a collaborative project [8, 9, 10, 11]. The standpoint is that of paper. To meet the objective the sample set used in the study includes ink-jet and electrophotographic prints made on a wide range of papers.

## Methodology

### *Material*

A series of samples was prepared for the study. The test layout included three photographs (Fig. 1), as well as test fields. The images were captured by a semi-professional photographer using a professional digital camera. Three image contents were chosen in order to represent the typical content types of objects with details (*cactus*), a human portrait (*man*) and a landscape (*lake*). The photographs had a print size of 10 cm × 15 cm.

**Table 1. Quality levels (left to right) in this study**

| Layer 1 | Layer 2 | Layer 3 | Layer 4 | Layer 5 |
|---|---|---|---|---|
| Physical characteristics | Instrumental measurements of print quality | Computational / Visual assessment of low level attributes | Computational / Visual assessment of high levels attributes | Model prediction/ MOS |
| Factors | | | | |
| Paper, printer | MTF, dynamic range, color gamut, noise | Noise, sharpness, contrast, colorfulness | Realism, genuineness, naturalness, clarity, depth, visual appeal, usefulness | Overall image quality |
| Context | | | | |
| ICC profiled printing | Simple test fields | Photographs | Photographs with use case | Photographs with or without use case |

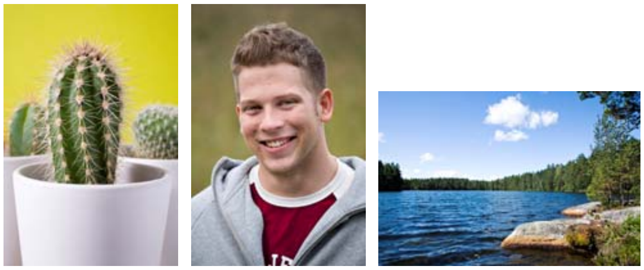

***Figure 1**. The different contents of the natural images used in the study.*

Papers of three different quality levels for both electrophotographic and ink-jet printing were selected for the study. The selection consisted of 6 multipurpose (MP), 8 electrophotographic (EPG), 7 electrophotographic photo (EPG Photo), 8 ink-jet (IJ) and 7 ink-jet photo (IJ Photo) papers. Papers marketed for printing high quality color images were considered as photo papers. EPG and EPG Photo papers as well as IJ and IJ Photo papers were printed with their respective printing methods. MP papers were printed with both methods. The variation in paper characteristics in the test series is presented in Table 2.

**Table 2. The variation in paper characteristics in the test series.**

| Paper | Whiteness (%) | Brightness (%) | Gloss (GU) |
|---|---|---|---|
| MP | 79 – 157 | 89 – 100 | 4.0 – 7.4 |
| EPG | 120 – 161 | 90 – 101 | 8.2 – 70 |
| IJ | 111 – 150 | 92 – 98 | 2.2 – 77 |
| EPG Photo | 107 – 148 | 91 – 95 | 15 – 87 |
| IJ Photo | 91 – 104 | 86 – 91 | 45 – 96 |
| Paper | Roughness PPS (μm) | Roughness Bendtsen (ml/min) | Permeability Bendtsen (ml/min) |
| MP | 4.5 – 7.2 | 91 – 247 | 392 – 706 |
| EPG | 0.9 – 4.1 | 6.5 – 51 | 0 – 213 |
| IJ | 0.7 – 5.7 | 8.1 – 270 | 26 – 193 |
| EPG Photo | 0.7 – 3.4 | 0 - 122 | 0 – 57 |
| IJ Photo | 0.6 – 4.0 | 0 – 68 | 0 – 18 |

The electrophotographic samples were printed with Xerox DC6060 (CMYK) and the ink-jet samples with Epson Stylus Pro 3800 (CMY, light C, light M, light K, light light K and photo K or matte K). On both devices, optimal print settings were chosen for each paper by identifying the print media type. For more details on the printing process, see [2].

The printed samples were digitized with a calibrated Canon EOS 5D SLR camera with a Canon EF 50 mm/2.5 lens. The digitization plane was illuminated by a filtered halogen lamp. The samples were captured by three exposures (1/10s, 1/5s and 2/5s). From these exposures a single 16 bits linear high dynamic RGB raw image was combined. The RGB image was first transformed to the absolute XYZ space and then to the LAB space.

### *Subjective Tests*

Overall quality along with sharpness, graininess, colorfulness and contrast were evaluated from the samples in subjective tests. Observers (n = 59) were university students and naïve as regards to print and image quality. The EPG samples were evaluated separately from the IJ samples: 29 of the subjects evaluated the EPG samples and 30 the IJ samples. Both tests followed the same procedure for the results to be comparable.

The overall quality of the samples was evaluated on a 5-point category scale (i.e. 1: bad, 2: poor, 3: fair, 4: good, 5: excellent). Sharpness, graininess, colorfulness and contrast were evaluated as semantic differentials (e.g. blurry–sharp) on a 5-point category scale (e.g. 1: clearly blurrier than sample set average, 2: slightly blurrier, 3: about the average, 4: slightly sharper, 5: clearly sharper). The scale of graininess was of opposite direction, i.e. grainy – non-grainy. The subjects were asked to reach their ratings to both ends of the scale.

All subjects rated first the overall quality of each sample in order to avoid any bias from evaluating the quality attributes. The samples of one image content at a time were placed on the table in a random order. The evaluation order of the contents was also randomized. After evaluating the overall quality of all the samples the subject was asked to rate the sharpness, graininess, colorfulness and contrast of the samples. The samples were again presented image content at a time; the subject rated the content on all scales before moving on to the next content. Again the order of the contents was randomized along with the order of the evaluated quality attributes.

### *Computational Quality Metrics*

Blur, noise, contrast and colorfulness were computed from the digitized samples. The following algorithms were used for the computation of the quality attributes: Crete et al. [12] for blur, Tai and Yang [13] for noise, rms [e.g. 14] for contrast, and Hasler and Süsstrunk [15] for colorfulness. Blur, noise and contrast were computed from the L* channel of the CIELAB color space. The results of the algorithms were normalized within each image content to match the subjective test setting of stretching the values to the ends of the scale.

### *Quality Modeling*

Stepwise (backward) multiple linear regression analysis was used for creating the quality models relating Layer 3 attributes to Layer 5 MOS values for the discrete cases of different printing methods, papers of different quality level, and different image contents. Two sets of models predicting the subjective print quality were created. In the first set, the mean subjective ratings were the independent variables used as the predictors. The other set of models was created with the computational quality attribute values as the predictors. The goodness-of-fit of the produced models was evaluated with the coefficient of determination, $R^2$. Each image content and printing method was analyzed separately. SPSS 17.0 was used for the creation of the regression models and the assessment of their goodness.

## Results and Discussion

Table 3 presents the regression models predicting MOS by subjective Layer 3 attributes for the printing methods separately and combined. All models explained the variation in the overall quality well with coefficients of determination of over 90%. Of the attributes, graininess appeared in seven of nine the models, colorfulness and sharpness in six. Contrast contributed to the quality prediction only in the models for the whole sample set. Colorfulness appeared to be an important predictor for the quality of the IJ samples. It was also included in all three models for the image content *cactus*.

**Table 3. The regression models predicting MOS by subjective Layer 3 attributes for the EPG and IJ samples separately and combined.**

| | All samples | | EPG samples (MP+EPG+EPG Photo) | | IJ samples (MP+IJ+IJ Photo) | |
|---|---|---|---|---|---|---|
| | Model | $R^2$ | Model | $R^2$ | Model | $R^2$ |
| Cactus | -0.44 +0.25·graininess +0.37·sharpness +0.28·contrast +0.26·colorfulness | 96.1 % | -0.52 +0.84·sharpness +0.34·colorfulness | 91.2 % | -0.30 + 0.29·graininess +0.81·colorfulness | 98.2 % |
| Lake | -0.18 +0.38·graininess +0.42·sharpness +0.27·contrast | 95.9 % | -0.02 +0.38·graininess +0.62·sharpness | 93.1 % | -0.05 + 0.47·graininess +0.56·colorfulness | 98.4 % |
| Man | -0.28 +0.23·graininess +0.66·sharpness +0.21·colorfulness | 97.2 % | 0.02 +0.96·graininess | 96.2 % | -0.27 +0.50·sharpness +0.59·colorfulness | 98.6 % |

Table 4 presents the regression models predicting MOS by subjective Layer 3 attributes for the different levels of paper quality. In the case of EPG prints, the models that best predicted the overall print quality were created for the MP papers. For the middle quality level, i.e. EPG papers, all three models contained only one predictor, graininess. Overall, graininess was the most common predictor in the EPG models as it was included in five of the nine models. When comparing image contents the models of the weakest explanatory power were for *cactus* with the lowest coefficient of determination being 46.4% for the EPG Photo papers.

In the case of IJ prints, the models with the best fit were for the IJ Photo papers. Again graininess was of great importance in the models, as it was included in seven of the nine models, i.e. in all models for *lake* and *man*, as well as all models for the IJ Photo papers. In general, the models for the IJ prints were better in predicting the overall quality of the samples than the models for the EPG prints. However, in all but four cases, the models for the different levels of paper quality had a weaker explanatory power compared to the models for the whole sets of EPG and IJ samples (Table 3).

**Table 4. The regression models predicting MOS by subjective Layer 3 attributes for the different levels of paper quality.**

| EPG samples | | | | | | |
|---|---|---|---|---|---|---|
| | MP model | $R^2$ | EPG model | $R^2$ | EPG Photo model | $R^2$ |
| Cactus | 0.54 +0.71·graininess | 83.7 % | 1.40 +0.52·graininess | 68.6 % | 1.35 +0.67·sharpness | 46.4 % |
| Lake | -1.55 +0.86·sharpness +0.61·colorfulness | 99.5 % | 0.34 +0.88·graininess | 68.5 % | 0.48 +1.34·sharpness -0.55·colorfulness | 94.9 % |
| Man | -1.10 +0.44·sharpness +0.71·contrast +0.28·colorfulness | 99.8 % | -0.12 +0.99·graininess | 86.4 % | 0.19 +0.92·graininess | 76.8 % |
| IJ samples | | | | | | |
| | MP model | $R^2$ | IJ model | $R^2$ | IJ Photo model | $R^2$ |
| Cactus | -0.37 +0.76·sharpness +0.38·contrast | 91.1 % | -0.003 +0.99·colorfulness | 92.3 % | -1.19 +1.28·graininess | 86.6 % |
| Lake | 0.13 +2.76·sharpness -2.03·graininess | 98.9 % | -0.51 +0.68·graininess +0.52·colorfulness | 81.7 % | 0.19 +0.46·graininess +0.53·contrast | 98.1 % |
| Man | -0.77 +1.71·graininess | 76.1 % | -0.54 +0.45·graininess +0.80·contrast | 94.8 % | 0.10 +0.96·graininess | 97.1 % |

**Table 5. The regression models predicting MOS by computational Layer 3 attributes for the EPG and IJ samples separately and combined.**

| | All samples | | EPG samples (MP+EPG+EPG Photo) | | IJ samples (MP+IJ+IJ Photo) | |
|---|---|---|---|---|---|---|
| | Model | $R^2$ | Model | $R^2$ | Model | $R^2$ |
| Cactus | 0.43 -11.51·blur -2.84·noise +17.22·contrast | 70.2 % | -96.92 +84.10·blur +21.65·noise | 48.9 % | -11.07 +15.64·contrast | 94.3 % |
| Lake | -6.61 -2.35·noise +12.83·contrast | 80.8 % | -13.30 +17.55·contrast | 58.8 % | -21.36 +17.86·blur +1.41·noise +7.10·contrast | 97.6 % |
| Man | -2.87 -3.02·noise +8.96·contrast | 70.3 % | -8.74 +12.57·contrast | 28.8 % | -10.82 +5.66·blur +9.96·contrast | 95.9 % |

Table 5 shows the regression models predicting MOS by computational Layer 3 attributes for the two printing methods separately and combined. The overall quality of the IJ prints could be predicted well with the computational attribute values, whereas the models for the EPG prints were not as strong. Of the attributes, contrast was included in eight of the nine models while colorfulness did not contribute to the quality prediction in any of the cases. Clearly the color metric used does not capture the right color properties perceived as important by subjects. Contrast on the other hand was considered difficult to evaluate by the subjects. This could also be seen in the deviation of the contrast ratings between the subjects.

## Conclusions

The purpose of the study was to consider the challenges in modeling the quality of digitally printed images, which rise from the differences between electrophotographic and ink-jet printing, the contribution of paper structures and the role of image content.

Based on the models predicting MOS by subjective attributes, creating a single quality model seems feasible provided that the chosen set of papers covers a wide range of quality variation. The models of different image contents varied to some extent in terms of the included variables and their weights. This suggests that there is a real need for quality models which are adaptive to the content.

The challenges of the computational models are in predicting the quality of electrophotographic prints. Furthermore, it appears that the subjective evaluation of colorfulness depends on the image content, which notably complicates the computational approach.

The studied samples gave no indication whether the quality level of the papers affects the decision making criteria of the subjects. To address this question, papers of different quality levels should be evaluated subjectively as separate sets. Now the samples were evaluated as a single set and the data was divided afterwards based on the different quality levels.

## References


[1] H.R. Sheikh, M.F. Sabir & A.C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms", IEEE Transactions on Image Processing 15, 3440 (2006).

[2] R. Halonen, T. Leisti & P. Oittinen, The Influence of Image Content and Paper Grade on Image Quality Attributes Computed from Printed Natural Images. Proc. IS&T NIP24, pg. 459. (2008).

[3] M. Nuutinen, R. Halonen, T. Leisti & P. Oittinen, Reduced-reference quality metrics for measuring the image quality of digitally printed natural images, (to be published).

[4] A. Loui, M.D. Wood, A. Scalise, et al., Multidimensional Image Value Assessment and Rating for Automated Albuming and Retrieval, Proc. ICIP 2008, pg. 97. (2008).

[5] J. San Pedro & S. Siersdorfer, Ranking and Classifying Attractiveness of Photos in Folksonomies", Proc. WWW 2009, pg. 771. (2009).

[6] T.J.W.M. Janssen & F.J.J. Blommaert, "Predicting the Usefulness and Naturalness of Color Reproductions", Jour. Imaging Sci and Technol. 44, 93 (2000).

[7] T. Leisti, J. Radun, T. Virtanen, et al., Subjective Experience of Image Quality: Attributes, Definitions and Decision Making of Subjective Image Quality. Proc. SPIE 7242, pg. 72420D-1. (2009).

[8] P. Oittinen, R. Halonen, A. Kokkonen, et al., Framework for modeling visual printed image quality from the paper perspective. Proc. of SPIE 6808, pg. 68080L-1. (2008).

[9] J. Pulla, M. Mettänen, K. Konkarikoski, et al., Bayesian Network Model as an Overall Image Quality Measurement System. Proc. IMEKO TC1-TC7, pg. 67. (2008).

[10] T. Eerola, J.-K. Kämäräinen, T. Leisti, et al., Is there Hope for Predicting Human Visual Quality Experience? Proc. IEEE SMC, pg. 725. (2008).

[11] T. Eerola, J.-K. Kämäräinen, T. Leisti, et al., Finding Best Measurable Quantities for Predicting Human Visual Quality Experience. Proc. IEEE SMC, pg. 733. (2008).

[12] F. Crete, T. Dolmiere, P. Ladret, et al., The blur effect: perception and estimation with a new no-reference perceptual blur metric. Proc. SPIE 6492, 11 p. (2007).

[13] S.-C. Tai & S.-M Yang, Fast Method For Image Noise Estimation Using Laplacian Operator and Adaptive Edge Detection. Proc. ISCCSP, pg. 1077-1081. (2008).

[14] E. Peli, *Contrast in complex images*, Jour. Optical Society of America, 7, 10, pg. 2032-2040. (1990).

[15] D. Hasler & S. Süsstrunk, Measuring colourfulness in natural images, Proc. SPIE 5007, pg. 87-95. (2003).


## Author Biography


*Raisa Halonen received her M.Sc. degree in Graphic Arts Technology from Helsinki University of Technology (TKK), Finland, in 2008. Since 2006 she has worked in the Department of Media Technology at TKK where her work has focused on the determination of quality from printed natural images.*