A Study on Subjective Evaluations of Printed Color Images

Kazuhiko Tanaka
Image & Information Research Institute, Dai Nippon Printing Co., Ltd., Tokyo, Japan

Michio Sugeno
Department of System Science, Tokyo Institute of Technology, Yokohama, Japan

ABSTRACT

Fuzzy measures and fuzzy integrals are applied to build an evaluation model of printed color images. First, subjective evaluation data of seven people are collected by pairwise comparisons. The subjects give 16 kinds of evaluations for each color proof. Then a two-layer evaluation model is proposed based on the result of factor analysis. In this study, the authors adopt Choquet's integral as a form of fuzzy integral because it has good properties compared with other forms of fuzzy integrals. A relaxation-method-like procedure has been devised to identify fuzzy measures of the two-layer model. After its effectiveness is confirmed with artificial data, the algorithm is applied to actual subjective evaluation data. This gives us results, revealing that the seven subjects are divided into two groups whose evaluation characteristics are structurally different from each other.

KEYWORDS: ***fuzzy measure, fuzzy integral, structure identification, subjective evaluation, color printing***

1. INTRODUCTION

Evaluations of color printing quality have rarely been the object of scientific research because of the fuzziness of the human senses. However, in the current situation of the publication of massive quantities of visual printed matter, a more reasonable and efficient evaluation method is desired in the graphic arts

Address correspondence to K. Tanaka, Image & Information Research Institute, Dai Nippon Printing Co., Ltd., 1-1-1, Ichigaya-Kagacho Shinjuku-ku, Tokyo 162-01, Japan.

International Journal of Approximate Reasoning 1991; 5:213–222

655 Avenue of the Americas, New York, NY 10010 0888-613X/91/$3.50

and relevant industries. As the first step toward achieving this aim, we tried to construct a model suitable for subjective evaluation.

In a rare work in this field, Mishina [1] derived an equation of color-printing image evaluation from subjectively evaluated data by multiple regression analysis, where representations of material, color balance, and feeling of roughness were selected as predictor variables. In this evaluation model, predictor variables (and their values) are assumed to be additive and independent.

However, the human evaluation process with respect to reproduced images in which an evaluator subjectively selects the most preferable reproduction can be considered essentially fuzzy. Because such evaluations can be influenced by particular colors such as skin color or sky blue, fidelity to the original is not always important to its reproduction. Therefore, we tried to build a model based on the idea of fuzzy measures, where we do not have to assume additivity and independence among predictor variables.

2. COLLECTING SUBJECTIVE EVALUATION DATA

First, we collected subjective evaluation data on printed color images in order to study the relationship between overall evaluation and evaluations specific to various viewpoints. The subjects were seven people who work for a printing company. For samples we made 20 proofs from four originals (color reversal films), five slightly different proofs from each original.

The test form is shown in Figure 1. The subevaluation items (attributes) were selected as typical words that people in the printing industry frequently use in expressing their evaluations of color proofs.

Each attribute score for each proof was calculated as follows:

With respect to attribute k,

- If proof A is distinctly more favorable than proof j, then $s(j) = +2$.
- If Proof A is slightly more favorable than proof j, then $s(j) = +1$.
- If Proof A is much the same as proof j, then $s(j) = 0$.
- If proof A is slightly less favorable than proof j, then $s(j) = -1$.
- If proof A is distinctly less favorable than proof j, then $s(j) = -2$.

The $s(j)$ values for all proofs other than proof A are summed, adding in the constant 8 to make the result nonnegative.

$$\mathrm{SCORE}(A, k) = \sum_{j \neq A} s(j) + 8 \quad \geq 0$$

where SCORE(A, k) denotes the score of proof A with respect to attribute k. An example is illustrated in Figure 2.

Comparison of Proof ____ (L) and Proof ____ (R)

Which proof . . .	(L)	(L) rather than (R)	Much the same	(R) rather than (L)	(R)
1. displays 3-dimensional feeling?					
2. displays transparent feeling?					
3. displays feeling of metallic surface?					
4. displays feeling of fine texture?					
5. displays feeling of volume?					
6. has more contrast?					
7. displays feeling of sharpness?					
8. is more bluish?					
9. is more reddish?					
10. is more yellowish?					
11. is vivid, fresh in color?					
12. displays details in lighter part?					
13. displays details in darker part?					
14. is away from muddiness?					
15. is bright as a whole?					
16. do you like better?					
	+2	+1	0	−1	−2

Figure 1. Test form for pairwise comparison.

	s(A)	s(B)	s(C)	s(D)	s(E)	$\Sigma s(j)$	Score
Proof A	*	1	1	2	−1	3	11
Proof B	−1	*	1	2	−2	0	8
Proof C	−1	−1	*	2	−1	−1	7
Proof D	−2	−2	−2	*	−2	−8	0
Proof E	1	2	1	2	*	6	14

Figure 2. How to determine proof scores.

Table 1. Factor Loadings and Cumulative Contribution Rates (CCR)

	F_1	F_2	F_3	F_4
1	0.74998	−0.51367	0.01859	−0.12555
2	0.21037	−0.83463	0.09731	−0.14157
3	−0.14265	−0.12086	0.83571	−0.12021
4	0.02056	−0.40787	0.75141	0.05090
5	0.87958	0.06078	−0.06407	−0.19624
6	0.82205	−0.38752	−0.12933	−0.14731
7	0.50481	−0.69627	0.13028	0.20734
8	−0.17877	0.24732	0.02073	0.86580
9	0.74076	0.44887	0.15436	−0.10562
10	0.58857	−0.03248	−0.00776	−0.71396
11	0.74629	−0.43571	0.17642	−0.15980
12	0.37875	0.15110	0.57723	0.25087
13	−0.11397	−0.50342	0.13149	−0.01098
14	0.23220	−0.79572	0.20052	−0.25179
15	0.01855	−0.78579	−0.09903	−0.46396
CCR	0.38632	0.56217	0.68225	0.75651

3. FACTOR ANALYSIS AND THE RESULT

Factor analysis was performed to investigate latent factors concerning human subjective evaluations of printed color images. The results were as follows. Items having a high correlation with the principal factors are

1, 5, 6, 11	with the first factor
2, 7, 14, 15	with the second factor
3, 4, 12	with the third factor

(See Table 1.)

Then three factors are interpreted as follows:

- The first factor (P factor) concerns physical and space representation.
- The second factor (T factor) concerns transparency, sharpness, and clarity of appearance.
- The third factor (Q factor) concerns representation of material constituting the main object in a reproduced picture.

These results suggested a subjective evaluation model like that of Figure 3 incorporating the concepts of fuzzy measures and fuzzy integrals. The G_P, G_T, G_Q of Figure 3 are fuzzy measures that each give the evaluation score of the respective intermediate block (P, T, and Q) from subevaluation scores 1–15. G_x is the fuzzy measure that determines the overall evaluation of a

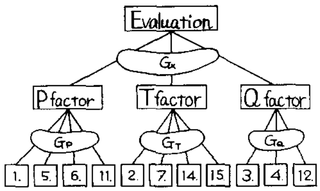


Figure 3. A subjective evaluation model for printed color images. The numbers in the lowest layer coincide with item numbers of Figures 1 and 2.

picture with the scores of the P, T, and Q factors. Attributes having relatively small correlation with the principal factors were ignored.

Generally speaking, an evaluation model like that of Figure 3, which consists of two layers, with the upper one corresponding to the principal factors derived from factor analysis, can be considered more transparent than a one-layer model, which gives the overall evaluation directly from many attributes. This is so because a two-layer model can roughly explain evaluation results in terms of fewer principal factors.

In this study, we adopted Choquet's integral as a form of fuzzy integral, since Choquet's integral is an extension of the Lebesgue integral (Sugeno and Murofushi [2]) and easy to calculate as described in the next section.

4. SUBJECTIVE EVALUATION MODEL USING CHOQUET'S INTEGRAL [2]

Suppose there are three evaluation items s_1, s_2, and s_3. Let $K = \{s_1, s_2, s_3\}$, h: K → [0, ∞), be a function giving the evaluation score for each item. In the case of $h(s_1) = a_1 \leqq h(s_2) = a_2 \leqq h(s_3) = a_3$, we have

$$(C)\int h\,d\mu = a_1\mu(K) + (a_2 - a_1)\mu(\{s_2, s_3\}) + (a_3 - a_2)\mu(\{s_3\})$$

$$= C$$

where μ represents the fuzzy measure.

The desired subjective evaluation model is obtained if we can determine the fuzzy meausure μ so that the result C is close enough to the actual overall evaluation E.

Let $\mathbf{x} = (x_1, \ldots, x_7) \in R^7$ denote the fuzzy measure μ, where $x_7 = \mu(K)$, $x_1 = \mu(\{s_1\}), \ldots, x_4 = (\{s_1, s_2\})$, $x_5 = \mu(\{s_2, s_3\})$, $x_6 = \mu(\{s_3, s_1\})$. The fuzzy measure we seek is the **x** minimizes

$$f(\mathbf{X}) = \sum_j (E_j - C_j)^2$$

under the following constraints:

$$x_1 \leqq x_4, \quad x_1 \leqq x_5, \quad x_2 \leqq x_4, \quad x_2 \leqq x_6, \quad x_3 \leqq x_5,$$
$$x_3 \leqq x_5, \quad x_4 \leqq x_7, \quad x_5 \leqq x_7, \quad x_6 \leqq x_7, \quad \text{and} \quad \theta \leqq x_1,$$

where θ is the zero vector.

We found this problem to be one of quadratic programming. It can be solved by applying the Lemke method (Kojima [3]).

IDENTIFYING AN ACTUAL EVALUATION MODEL

When we consider a two-layer model as in Figure 3, we cannot apply the result of the previous section unless we have all scores of the intermediate blocks. Therefore, we devised a relaxation-method-like procedure to identify fuzzy measure that infers scores of the intermediate blocks as well as the fuzzy measures, repetitively, eliminating incompatibilities among data. An outline of this algorithm is as follows. (For relaxation methods, refer to Rosenfeld et al. [4].)

Each block score of the ith sample ${}_iP$, ${}_iT$, or ${}_iQ$ is regarded as an object to be labeled, where the label is a kind of block score of the ith sample that can take only one discrete value, for instance, any one of $\Lambda = \{0, 1, 2, \ldots, 10\}$, according to a certain probability distribution.

To represent the above assumption for each ${}_iP$, ${}_iT$, or ${}_iQ$ labeling vector, we define ${}_iPP$, ${}_iTP$, or ${}_iQP$. ${}_iPP(\lambda)$, which denotes the λth component of ${}_iPP$, has the real number equal to the probability that ${}_iP$ takes λ as its block score.

STEP 1 Initialize labeling vectors ${}_iPP$, ${}_iTP$, and ${}_iQP$ for all i. Set the initial scores of temporary blocks ${}_iPtmp$, ${}_iTtmp$, and ${}_iQtmp$ based on the labeling vectors. Expectations of labeling vectors are taken.

Let $K = 1$.

STEP 2 By applying the Lemke method for these temporary block scores and the data obtained from subjective evaluations, obtain fuzzy measures Gp, Gt, Gq, and Gx. Let M^k denote the result.

STEP 3 Check the degree to which actual evaluations agree with calculated scores on M^k for all combinations of labeling of ${}_iP, {}_iT, {}_iQ$. This is called compatibility. The compatibility for ${}_iP = \lambda$, ${}_iT = \lambda'$ on model M^K, which is denoted ${}_i\mathrm{COM}_{PT}(\lambda, \lambda')$ is calculated as the residual of the ith sample over the P block,

$$ {}_i\mathrm{RES}_P = \left| \lambda - (C) \int hd\mu \int_i (P_1, P_2, P_3)\, dG_P \right| $$

The second term on the right-hand side denotes Choquet's integral result with evaluation scores P_1, P_2, P_3 with respect to the fuzzy measure G_P.

$$ {}_i\mathrm{RES}_T = \left| \lambda' - (C) \int hd\mu \int_i (T_1, T_2, T_3)\, dG_T \right| $$

$$ {}_i\mathrm{RES}_{XPT} = \left| {}_iE - (C) \int hd\mu \int_i (\lambda, \lambda', iQtmp)\, dG_x \right| $$

where ${}_iE$ denotes the overall evaluation score of the ith sample.

$$ {}_i\mathrm{ORES}_{PT} = 0.5\,{}_i\mathrm{RES}_{XPT} + 0.25({}_i\mathrm{RES}_P + {}_i\mathrm{RES}_T) $$

(overall residual of the ith sample), and

$$ {}_i\mathrm{COM}_{PT}(\lambda, \lambda') = 1 - \frac{2\,{}_i\mathrm{ORES}_{PT}}{\max_j({}_j\mathrm{ORES}_{PT})} $$

${}_i\mathrm{COM}_{PQ}(\lambda, \lambda')$, ${}_i\mathrm{COM}_{TQ}(\lambda, \lambda')$ can be calculated in a similar manner.

STEP 4 If the compatibility of a label is big, then the probability assigned to the label is increased and if it is small, then the profitability is decreased. In this way, the modification quantity of each assigned probability is determined. For instance, the modification quantity of ${}_iPP(\lambda)$ is calculated from the equation

$$ \begin{aligned} \Delta_i PP(\lambda) &= 0.5 \sum_k \left[{}_i\mathrm{COM}_{PT}(\lambda, k) * {}_iTP(k) \right] \\ &\quad + 0.5 \sum_k \left[{}_i\mathrm{COM}_{PQ}(\lambda, k) * {}_iQP(k) \right] \end{aligned} $$

STEP 5 Modify the labeling vectors. Calculate new temporary block values from the new labeling vectors.

Let $K = K + 1$. Go to step 2.

Figure 4 shows experimental results obtained with the above procedure. Three hundred artificial data sets, each consisting of P_1, P_2, P_3, T_1, T_2, T_3, Q_1, Q_2, Q_3, and E, were given to the test program. Each E was calculated through arbitrary fuzzy measures G_x, G_P, G_T, and G_Q.

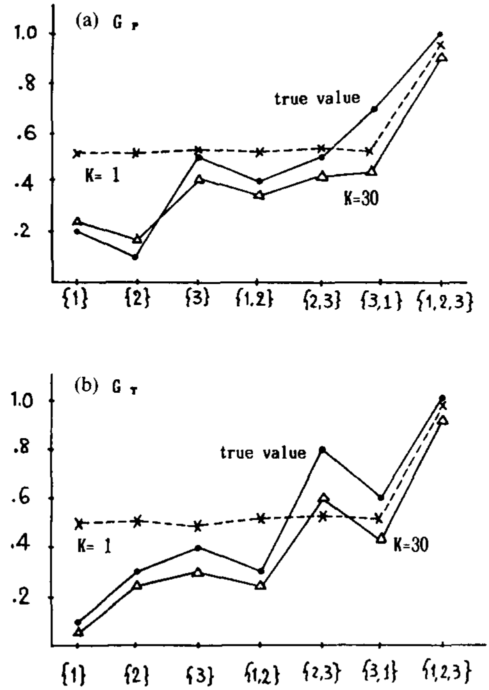


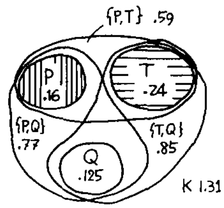
Figure 4. Experimental results with the relaxation-method-like procedure.

Since the algorithm does not guarantee convergence, we have to stop execution of the program at the most appropriate times of iteration. Experiments with several data sets show that the best identified result for G_x tends to be obtained after about 20 iterations, whereas those for G_P, G_T, and G_Q tend to be obtained after about 50 iterations. Thus, we conclude that 30 iterations is best for identifying the four sets of fuzzy measures.

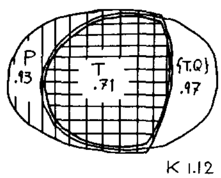
Comparing the result of 30 iterations ($K = 30$) with that of $K = 1$, it is seen that the accuracy of the calculated overall evaluation of model M^K was improved approximately threefold. Furthermore, it is found that the fuzzy measure values of G_P, G_T, G_Q of $K = 30$ clearly show their characteristic tendencies, whereas those of $K = 1$ appear almost trivial.

6. IDENTIFIED RESULTS

Before devising the relaxation-method-like procedure described above, we had a prediction that the seven subjects could be divided into two groups whose approaches to evaluation of the color prints were distinctly different. This



(a)



(b)

Figure 5. (a) G_x of four-person model. Any set $K = \{P, T, Q\}$ corresponds to a closed loop whose area is proportional to its fuzzy measure value. (b) G_x of three-person model.

prediction was obtained by applying subjective evaluation data of each subject to a linear model and comparing the results.

Therefore, we applied the relaxation-method-like procedure for each of the two groups.

Figure 5 shows the identified results of the Figure 3 models. Only G_x's are illustrated. The four-person model reveals that when members of this group evaluate printed color images, they give almost equal importance to P, T, and Q factors. Results obtained with the other group (three-person model) reveal that they almost ignore the Q factor and that the P and T factors are interdependent in their evaluations.

Figure 5 not only shows the differences between corresponding fuzzy measures (which are often seen in linear models as differences between corresponding coefficients), but also reveals, so to speak, structural differences, which appear as differences in interdependency among the three factors.

CONCLUSION

In order to build a model that explains the evaluation mechanism more understandably, we used fuzzy measures rather than linear models. The reason we consider a two-layer model based on the result of factor analysis is that we hope to devise a still more transparent evaluation model.

We have shown a concrete procedure to construct those models. Differences in the ways in which pictures are evaluated are revealed that are not merely quantitative but also structural.

References

1. Mishina H., Ohno, Y., Niikura, M., Irie, H., Mikami, S., Ohtsuki, K., and Kominami, T., Evaluation of print quality, *Proceedings of the 72nd Spring Conference of Japanese Society of Printing Science and Technology*, 1-4, 1984 (in Japanese).
2. Sugeno, M., and Murofushi, T., Choquet's integral as an integral form for a general class of fuzzy measures, *Proceedings IFSA'87*, 408-411, 1987.
3. Kojima, M., *Souhosei to Hudouten*, Sangyou Tosyo, Tokyo, 1981 (in Japanese).
4. Rosenfeld, A., Hummel, R., and Zucker, S., Scene labeling by relaxation operations, *IEEE Trans. Syst., Man Cybern.* SMC-6(6), 420-433, 1976.
5. Onisawa, T., Sugeno, M., Nishiwaki, Y., Kawai, H., and Harima, Y., Fuzzy measure analysis of public attitude towards the use of nuclear energy, *Fuzzy Sets Syst.* 20, 259-289, 1986.