Journal of Imaging Science and Technology® 54(1): 010501–010501-10, 2010.

Measurement and Modeling of Vividness Perception and Observer Preference for Color Laser Printer Quality

Y. J. Kim▲, Y. Bang▲ and H.-K. Choh
Digital Printing Division, Samsung Electronics, Suwon 443-742, Korea
E-mail: younj.kim@samsung.com

Abstract. *The present article investigates a particular problem: how vividness can be calculated and used to evaluate printer quality. Vividness is a term representing chromaticness of colors (conceptually similar to chroma) and has also been adopted as one of the color adjectives in Inter Society Color Council-National Bureau of Standards (ISCC-NBS) color naming and practical color coordinate (PCCS) systems. According to ISO 20462-2, a new psychophysical method (triplet comparison method) was performed. As a result, an interval scale for vividness was established, and it was modeled as a function of mean chroma, C^*_{ab}, and lightness, L^*, of printer primary and secondary colors. Pearson correlation between the metric prediction and corresponding subjective data was about 0.96. The methodology was further extended to measure observer preference (preferred-vividness). Both preferred-vividness and vividness metrics were based upon chroma and lightness, but the contribution of lightness is much higher in the former (∼40%) case than in the latter (∼10%). © 2010 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2010.54.1.010501]

INTRODUCTION

Vividness is a term representing chromaticness of colors and has been adopted as one of the color adjectives in both the ISCC-NBS color designation[1] and the PCCS system.[2] Color adjectives are substantial components in those hue-tone systems. Vividness has also been frequently used for evaluating business graphic print quality in printing industry. For instance, as the marketing strategy in the printing industry moves to business-to-business (B2B) from business-to-consumer (B2C), reproduction of higher quality vivid business graphic images in presentation slides has become strongly demanded. Nayatani[3–5] reported that the concept of degree of vividness is similar to the definition of "chroma" in the Commission Internationale de L'eclairage (CIE) International Lighting Vocabulary[6] and can be used to estimate chromatic intensity of colors using interval or ratio scales. For example, achromatic colors have zero vividness and highly saturated colors would show a high vividness value. Nayatani in 2005[5] proposed an empirical model predicting degree of vividness (DV) as a function of Munsell chroma (C) and whiteness and blackness ($[W\text{-}Bk]$) based on the observations of the NCS color chart as shown in Eq. (1).

The compound characteristics of vividness affected by chroma and lightness channels could be observed in the ISCC-NBS system as well.[1] Significantly large differences in Munsell values and chromas of the central colors were revealed for different hues;[2,7]

$$\mathrm{DV} = C[1 + 0.10(W\text{-}Bk)]. \tag{1}$$

In the field of image quality evaluation, color vividness has been understood as the degree of colorfulness[8,9] and there has been a number of efforts to predict colorfulness of images and its effects on the image quality.[10–12] Colorfulness, which is one of the perceptual attributes in color appearance modeling, is also a very similar concept to vividness. It is defined as an attribute of a visual sensation according to which an area appears to exhibit more or less light.[13,14] The colorfulness of a given color stimulus increases with luminance which is referred to as the Hunt effect[13,14] describing the perceptual difference caused by large differences in illumination and the corresponding state of adaptation.[10]

Most color appearance models include colorfulness and both luminance and chromatic information are invoked to model it [as in Eq. (1)]. For example, the Nayatani et al. color appearance model[15,16] developed in 1981 predicts colorfulness M by the chroma C of the sample multiplied by the brightness of an ideal white B_{rw} as

$$M = C\frac{B_{rw}}{100}. \tag{2}$$

In addition, the Hunt color appearance model[17] defines the colorfulness M_{94} as the product of chroma C_{94} and the luminance level adaptation factor F_L raised to a power of 0.15, derived empirically through analysis of visual scaling results as shown in Eq. (3),[13,17]

$$M_{94} = C_{94}F_L^{0.15}. \tag{3}$$

Recently, CIECAM02[18] was recommended by CIE, and a new colorfulness correlate was included based upon the Hunt model colorfulness. The CIECAM02 colorfulness M is calculated by scaling the chroma predictor, C [see Eq. (4)], by the fourth root of the luminance-level adaptation factor F_L as illustrated in Eq. (5);

$$C = t^{0.9}\sqrt{J/100}(1.64 - 0.29^n)^{0.73}, \tag{4}$$

▲IS&T Member.

Received Jan. 27, 2009; accepted for publication Aug. 22, 2009; published online Dec. 4, 2009.
1062-3701/2010/54(1)/010501/10/$20.00.

$$M = CF_L^{0.25}, \quad (5)$$

where a temporary quantity t that is related to saturation and incorporates the chromatic induction factors for surround and background as well as the eccentricity adjustment is computed, and J is the lightness correlate defined in CIECAM02 which can be computed from the achromatic response, achromatic response for white, the surround factor and the base exponent. More details about those three color appearance models discussed above can be found in Refs. 13 and 18. In a later section, prediction of the CIECAM02 colorfulness will be compared with that of the vividness model developed in this article.

There is also a publication which predicts vividness for light source evaluation. Rea and Freyssinier-Nova[19] appreciated vividness as a light source's color rendering property using gamut area index (GAI) in conjunction with color rendering index (CRI).[20] GAI can be calculated as the area of the polygon in CIELAB formed by the eight CIE standard reflectance samples' chromaticities. Because of the CRI's insufficient prediction accuracy, CRI and GAI were jointly used for measuring light sources' vividness.[19]

This article intends to propose a metric that accurately predicts the independent variable ψ representing a printer's vividness on the basis of Nayatani's empirical vividness model. Equation (6) shows a generalized form of vividness modeling using CIELAB color space units. It can be used to evaluate printer quality. It should be noted that the aforementioned works explored the vividness and colorfulness perceptions for uniform color patches but the current research focuses on the evaluation of actual color printers using complex stimuli (e.g., business graphic images);

$$\psi = \frac{1}{n}\left(\omega_C \sum_i^n C^*_{abi} + \omega_L \sum_i^n L^*_i\right), \quad (6)$$

where n denotes the number of printer primary and secondary colors to be used, e.g., $n=6$ in the case of cyan, magenta, yellow, red, green, and blue (CMYRGB), w_c and w_L represent weighting factors for C^*_{ab} and L^*, respectively, and are derived using a linear regression method based upon a set of subjective data. Since CIELAB color space has been often used in the imaging industry, our metric is based upon CIELAB rather than Munsell chroma and the NCS color chart.

Furthermore, the vividness metric was extended to measure observer preference (so called preferred-vividness) that can be made applicable to printer quality evaluation as well. Its prediction accuracy with respect to the subjective data, which were obtained from four sets of psychophysical experiments, was directly compared with CIECAM02 colorfulness estimate [see Eqs. (4) and (5)] in order to present merits and performance of our metric.

The psychophysical assessments performed for this article can be divided into four experiments as listed in Table I. Experiment 1 is conducted for two separate purposes. The main purpose of experiment 1 is to reduce the number of samples which are to be used in experiment 2. The other purpose of experiment 1 is to obtain a set of training data for developing a preferred-vividness metric. Experiment 2 is intended to obtain a set of training data for developing a vividness metric. More details about those experimental settings and modeling will be discussed subsequently within the section of Main Visual Assessments for Metric Training. In addition, prediction accuracy of those two metrics established in this work is tested using data from experiments 3 and 4 that will be discussed in greater details within the section Additional Visual Assessments for Metric Generalization.

Table I. List of psychophysical experiments performed.

Experiment No.	Method	Purposes
1	Category scaling	(1) To reduce the number of samples which can be used in experiment 2
		(2) To obtain a set of training data for developing a preferred-vividness metric
2	Triplet comparison	To obtain a set of training data for developing a vividness metric
3	Category scaling	To evaluate generality of the vividness metric developed in experiment 2
4	Category scaling	To evaluate generality of the preferred-vividness metric developed in experiment 1

EXPERIMENTAL

Setup

In total, 12 color laser printers produced by different manufacturers, e.g., Brother, Canon, Dell, HP, Minolta, Ricoh, Samsung, and Xerox, were selected and compared to each other in terms of vividness and observer preference of the prints. Those sampled printers show a wide range of printing performances from low end for ordinary users to high end for industrial applications. For each printer, the maximum level of the primary and secondary colors (CMYRGB) were transformed into RGB space using the specifications web offset publications (SWOP) conversion[21] and printed on A4-size Xerox Colortech+100 g/m^2 paper. CIELAB coordinates of those six colors were measured using a spectrophotometer (GretagMacbeth Spectroscan).

Figure 1 provides test images used in this article. Since vividness is strongly related to purity of colors,[3–5] business graphic images depicting highly saturated colors were chosen. Recently, the marketing strategy in the printing industry has moved to B2B from B2C, and reproduction of higher quality business graphic images in presentation slides has become strongly in demand. Test images 1 and 2 mainly include maximum cyan, magenta, yellow, red, green, and blue colors and the others contain intermediate hue levels of those printer primary and secondary colors.[22]

Main Visual Assessments for Metric Training

According to ISO 20462-2 (Photography: Psychophysical Experimental Methods for Estimating Image Quality, Part 2:

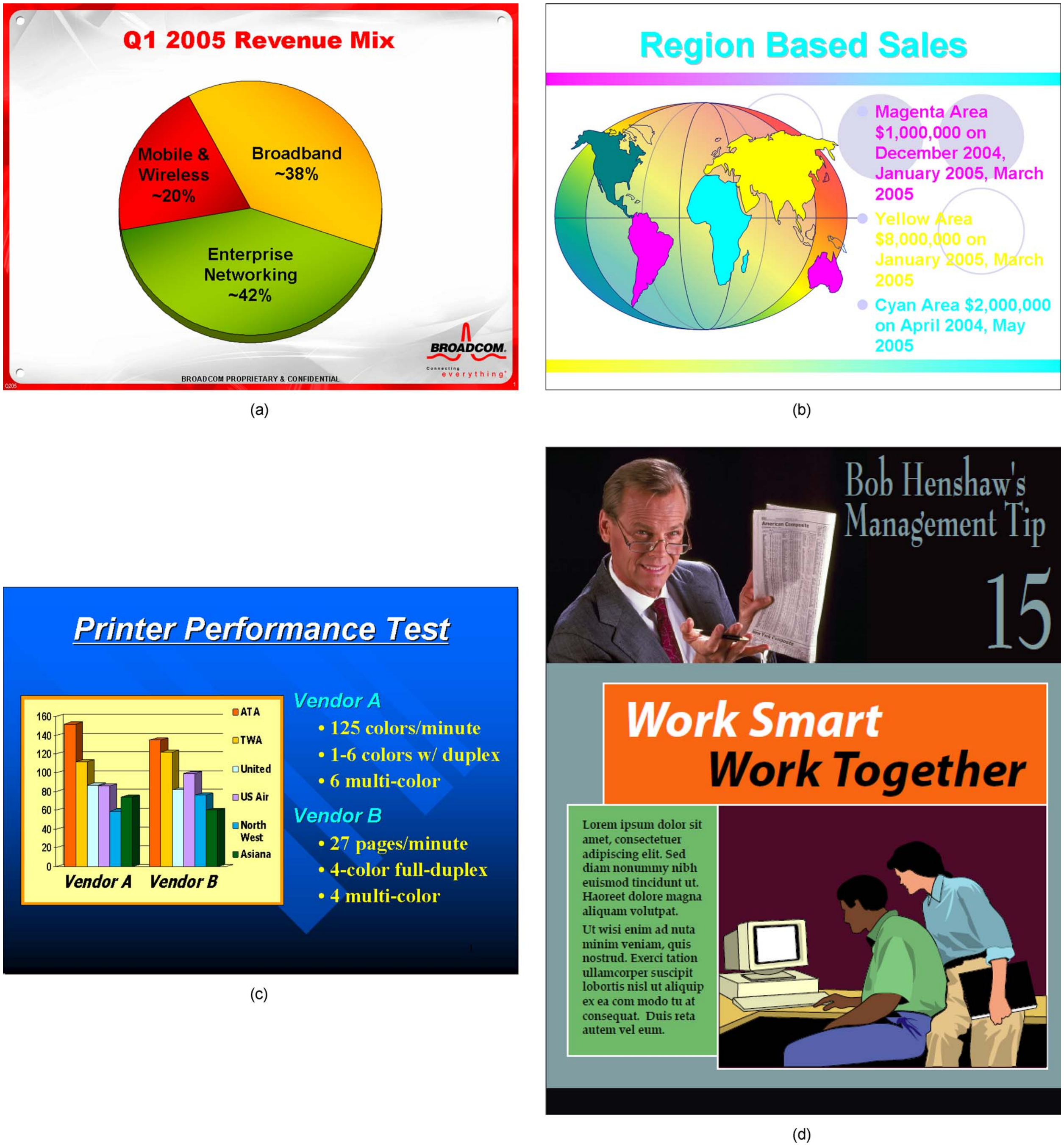

Figure 1. Test Images.

Triplet Comparison Method),[23] two step psychophysical experiments were conducted. The traditional paired comparison method is one of the most common techniques for assessing image quality because of the simple and easy procedure as well as the precise scalability. However, a serious problem with the method is that the number of samples to be examined is to be relatively limited. As the number of the samples increases, the number of combinations becomes extensive which causes excessive observer stress, which, in turn, can affect the accuracy and repeatability of the results. Triplet comparison is a new psychophysical method defined in ISO 20462 that involves the simultaneous scaling of three test stimuli with respect to image quality or an attribute thereof, in accordance with a set of instructions given to the observer. It enables a large number of samples to be examined and provides precise scalability with a much lower number of assessments than paired comparison method. Accordingly, the triplet comparison method reduces the number of assessments considerably; therefore observer stress can be minimized, which, in turn, may increase the accuracy and repeatability of the results.

For all the psychophysical experiments performed in this article, the distance between an observer and given stimuli was set to 25 cm. The ambient lighting level was approximately 800 lx ($\approx$255 cd/m^2) under a typical office viewing condition. The 1931 *xy* chromaticity coordinates of the illumination are (0.334, 0.365) near both D65 and D50 but rather closer to D50. The viewing geometry was 45/0 to attenuate any glaring effects from the stimuli caused by specular reflection. The observers are all experts working in the color imaging industry.

The first step of the visual assessment methodology used in this study is a category step (experiment 1) that aims to reduce the number of samples. Equally perceived intervals between any consecutive categories are assumed and all samples are categorized into three groups defined as "3: favorable," "2: acceptable," and "1: unacceptable." Samples are selected according to the number of samples required for the triplet comparison step (experiment 2). It should be noted that experimental results so obtained will also be used as a training data set for developing the preferred-vividness model since the three categories defined above are based on preference judgment.

ISO 20462–2[23] provides examples of possible sample combinations, which can be used in triplet comparison method, when the number of samples selected is, e.g., 7, 9, 13, 15, 19, 21, 25, and 27. Then it is possible to select sample combinations that eliminate their duplication across all triplets. Therefore, the number of reduced samples T can be expressed as

$$T = 6K + 1 \quad \text{or} \quad T = 6K + 3, \tag{7}$$

where K is an integer number, e.g., 1, 2, 3, 4, 5, 6, etc. For any value of T, the number of sample combinations N is

$$N = T(T-1)/6. \tag{8}$$

Suppose a T-sided regular polygon and each apex of the polygon is assigned an integer value from 1 to T. We define the notation whereby (p,q,r) represents a triangle comprising the apices p, q, and r and where the triangle apices represent a combination of samples for the triplet comparison method. In this article, 12 samples were collected first and 7 of them were selected and used in the triplet comparison step where $K=1$ in Eq. (7) in this case.

Four expert observers, who have worked in the field of color imaging industry, participated in step 1 and divided 48 images produced by the 12 sample printers (=4 images × 12 printers) into the three categories in terms of their preference. Each observer assessed each print at a time under the typical office lighting environment. Sequence of the assessments was randomized, and the collected subjective data were averaged for each printer. This is one of the common methods for analyzing the category judgment data sets and has been recommended by ITU-R BT.500-11.[24] The mean subjective score is often referred to as mean opinion score (MOS) that can be computed as

$$\bar{u}_{jk} = \frac{1}{n}\sum_{i=1}^{n} u_{ijk}, \tag{9}$$

where u_{ijk} is a subjective score of observer i for test printer j and image k and the number of observers is n. The total number of observations is 192 (=4 images × 12 printers × 4 observers). Consequently, seven printers were selected to be used in the next step out of the total number of printers used (12) as discussed earlier.

The second step (experiment 2) is to derive a precise scaling based on an interval scale by comparing triplets of given samples. Specifically, three samples are compared simultaneously, thereby achieving high assessment accuracy while keeping the experimental scale realistic. Compared to paired comparison method, triplet comparison shortens assessment times so it is expected to improve data accuracy and reproducibility. Following ISO 20462-2,[23] Scheffe's method was applied for the statistical analysis to obtain an interval scale, and it was converted into just noticeable difference (JND) values. The interval scale relies upon Thurston's law of comparison case V[25] by computing cumulative frequency distribution and probability matrices. Precisely, probability, p, for cumulative frequency n is given by

$$p = (N + n)/2N, \tag{10}$$

where N is the total number of observations.

The amount of differentiation, Q, for the JND between two samples can be derived as given in Eq. (11) and a single JND value can be assigned for each sample by averaging the all JNDs between the given sample and the others;

$$Q = (12/\pi) \times \arcsin(\sqrt{p}) - 3. \tag{11}$$

Thirteen expert observers including the four observers who participated in the previous session (experiment 1) assessed 364 triplets (=4 images × 7 triplets × 13 observers) under the same office lighting environment. Each observer was asked to compare each triplet and rank the test stimuli in terms of vividness. Variation between observers was evaluated in terms of the Pearson correlation r and a modified version of coefficient of variance (CV). The former reflects the degree of linearity in the relationship between a pair of variables (e.g., x and y). Pearson correlation can be expressed as

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y}, \tag{12}$$

where $(n-1)$ is degree of freedom. Mean values for the x and y variables are $\bar{x}$ and $\bar{y}$ respectively and standard deviations are S_x and S_y, respectively. When the variables are perfectly linearly related, their Pearson correlation $r=+1$.

The latter response is often used as a measure of the "observer accuracy" which represents the mean discrepancy of a set of psychophysical data obtained from a panel of observers from their mean value as illustrated in Eq. (13). This term has been widely used in color appearance and color difference studies.[26,27] The original CV is a normalized measure of dispersion for a repeated measurement but was applied to measure the degree to which a set of data points varies in this article. It is defined as the percentage of the root mean square of the difference between two variables divided by the mean value of one of the variables. The CV is normally displayed as percentage and, for a perfect agree-

Table II. Printer vs mean opinion score (MOS) from step 1.

	Group I (≥2.5)					Group II (<2.5 and ≥1.5)				Group III (<1.5)		
Printer	A	B	C	D	E	F	G	H	I	J	K	L
MOS	2.81	2.63	2.63	2.56	2.50	2.19	2.06	1.94	1.81	1.25	1.13	1.00

ment between the variables, equals 0. A range of CV from 10% to 30% is known as the acceptable level. It can also be affected by experimental methodology, so a larger CV value can be expected from complicated experimental procedures. Accordingly

$$\mathrm{CV} = 100 \times \frac{\sqrt{\sum_i (x_i - y_i)^2 / n}}{\bar{y}}, \tag{13}$$

where x_i is a subjective value of each observer for each stimulus (i) and y_i is its corresponding mean subjective value across all of the observers; n represents the number of stimuli and $\bar{y}$ represents the grand mean of all stimuli's mean subjective values.

Additional Visual Assessments for Metric Generalization

In order to verify merits and generality of the metrics developed in this article, metric generalization procedure was carried out through two additional sets of category judgment assessments (experiments 3 and 4). Predicted values from the metrics were compared with their corresponding subjective data obtained from the following psychophysical procedure. Experiment 3 is conducted in the purpose of verifying the vividness metric and the overall procedure corresponds to experiment 1. However, a five-point scale, where all categories are defined by a symmetrical design of quantitative adjectives, is used this time for a higher data scalability. The categories were defined as "5: Highly Vivid," "4: Quite Vivid," "3: Vivid," "2: Quite Unvivid," and "1: Highly Unvivid." Six expert observers rated printed test images using the five-point scale. The test images were printed by the five printers which are excluded in experiment 1 procedure and are not used for developing the vividness metric in experiment 2.

Experiment 4 is designed to test performance of the preferred-vividness model, and slightly different categories are used: "5: Favorable," "4: Fair," "3: Neutral," "2: Not Preferred" (because the image appears either too vivid or too washed out with extremely low chromaticness), and "1: Poorly Reproduced." The test images used in experiment 3 are adopted. The total number of observations is 120 (=4 images×5 printers×6 observers) for each experiment. The collected data were averaged across the observers and images for each printer as illustrated in Eq. (9).

RESULTS

Observer Variation

The mean CV of the all observers participated in this experiment ranged from 7% to 20% and the grand mean CV across all the observers and the five test stimuli was 12% which can be considered acceptable. One of the observers showed a relatively higher CV (20%) than the other observations but this observer's impact on the grand mean (12%) was not large so the data were included for further analysis and modeling. In terms of Pearson correlation, the grand mean Pearson value was 0.84, and the lowest and highest values were 0.75 and 0.93, respectively. Generally, observer variation was acceptably small and judgments of the observers were strongly correlated with each other.

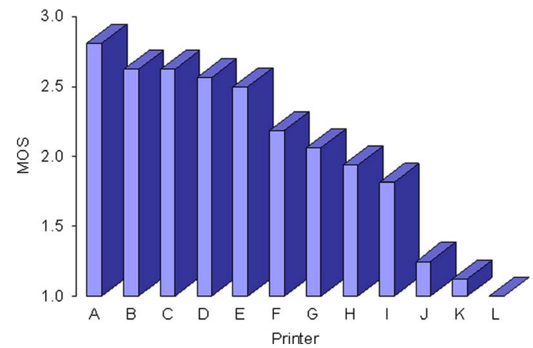

Figure 2. Graphical illustration of each sample printer's MOS obtained from Step 1 procedure. Printers A through E showed higher MOS values and MOS of printers J through L was under score of 1.5 which is much lower than others. Therefore, those data can be clearly clustered into three groups (A-E/F-I/J-L).

Results from Experiment 1 (Category Judgment)

Table II gives MOS of the 12 sample printers across the all test images and 13 observers obtained from experiment 1. Printers A through E showed higher MOS values, and MOS of printers J through L was under the score of 1.5 which is much lower than the others. Therefore, the data could be clearly clustered by three groups. The first group (group I) includes printers showing higher MOS larger than 2.5 (printers A through E) and the rest can also be divided into the middle (printers F through I: group II) and lower (printers J through L: group III) MOS groups. This clustered data distribution is graphically illustrated in Figure 2.

As Eq. (7) recommends, the number of samples should follow either $6K+1$ or $6K+3$ (where K is an arbitrary integer) in the triplet comparison method. Therefore, 5 of the 12 printers (B, C, G, I, and L) were excluded in this article so the other 7 (A, D, E, F, H, J, and K) were used only in the subsequent psychophysical experiment, experiment 2. Printer L was taken out due to its unacceptably lower color reproduction quality, and B, C, G, and I were also excluded so that that the remaining samples could represent a wide range of printing performances from low end for small business to high end for industrial applications.

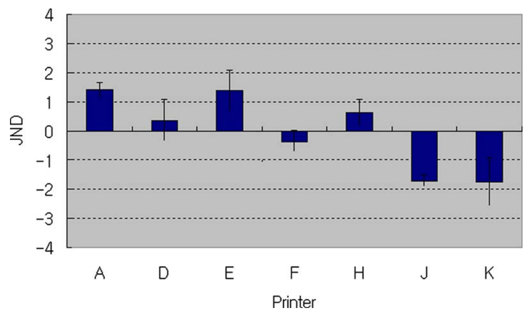

Figure 3. Mean JND computed obtained from Step 2 prcedure (triplet comparison) across the all test images. (Error bars show 95% confidence intervals.)

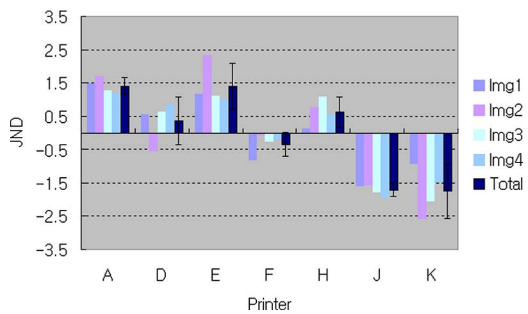

Figure 4. Independency of different test images. (Note that a lower JND in img2 for printer D is due to a huge banding artifact only shown in that case.)

Results from Experiment 2 (Triplet Comparison)

In Figure 3, mean JND values across the all test images obtained from step 2 procedure (triplet comparison) for those seven selected sample printers are depicted. The larger the JND, the higher the subjective vividness score. Printers A and E are high-end products applicable to large format prints and produced the first and second highest vividness scores. Then H, D, F, J, and K followed in order. Apparently, since printers J and K are for general users, quality of their toner would not reach that of high-end ones and their subjective vividness score were much lower. Error bars show 95% confidence intervals which can be computed as

$$\mu \pm t_{2.5\%}(n-1) \times SE, \tag{14}$$

where μ denotes the mean JND value of each printer and SE denotes its standard error of mean, i.e,, standard deviation divided by square root of number of observations, and $(n-1)$ is the degrees of freedom. The value corresponding to 2.5% of t for the given degrees of freedom is designated $t_{2.5\%}$.

Image Dependency

In Figure 4, JND values were separately computed for different test images and compared each other in a single plot. (Note: raw data are provided in Table III). Different colored bars represent different test images. The mean JND values are also given with 95% confidence interval in order to demonstrate the generally similar data trend for different image contents. As can be seen, most JND values for different scenes vary within the 95% confidence interval boundaries. However, it should be noted that there is an exceptional case of image 2 for printer D. Its JND value was much smaller than that for the other images. Considerable high frequency gray banding artifacts were observed from low chroma background colors in that specific image and may result in the lower vividness score. The periodically occurring high frequency gray bands contrasted strongly with the low chromatic pastel colors and appear obvious to human observers. The artifacts actually decrease the overall chromaticness of the image. However, vividness of the other test images was not severely affected by presence of gray bands perhaps because most colors in those images are surrounded by white background (image 1) or filled with highly chromatic colors (images 3 and 4), hence the gray bands are not easily perceptible.

Each sample printer's JND value for each test image is given in Table III, and their mean JND values across the whole image set are also provided. The Pearson correlation r was used to quantify the variation in the subjective vividness score between an image and the mean values. They all ranged from approximately 0.94 to 0.98, which represents a strong linear correlation of the results between the mean value and each test image. It seems surprising that such high Pearson correlations were achieved when image 2 suffers from the gray bands and seems to differ so much from the other images. It is still near the boundary of 95% confidence interval so its impact on the linear relation between the image contents may not be huge.

Observers locally adapt to a given complex image and

Table III. JND for each image and the mean. (Pearson: correlation of JND between mean and each image).

Printer	Image 1	Image 2	Image 3	Image 4	Mean
A	1.484	1.689	1.266	1.207	1.412
D	0.544	−0.560	0.642	0.857	0.371
E	1.175	2.324	1.108	1.015	1.406
F	−0.833	−0.078	−0.277	−0.214	−0.351
H	0.139	0.757	1.087	0.563	0.637
J	−1.585	−1.561	−1.773	−1.944	−1.716
K	−0.924	−2.571	−2.053	−1.481	−1.757
Pearson	0.941	0.944	0.980	0.973	

Table IV. Regions of Interest (ROIs) for test images.

	ROIs	No. of ROIs
Image 1	RGB colors in the pie-chart	4
Image 2	CMY continents	3
Image 3	Yellow graph, red bar, and blue background	3
Image 4	Orange, green, and purple boxes	3

focus on a particular region of interest in the image.[28] This particular region was named as region of interests (ROI) in an image. This terminology is drawn from the image processing field, where "ROI" signifies region of interest.[29] The concept of ROI seems to be arguable, but it has actually been widely used in image enhancement algorithm tests.[12,28,30–32] Because the test stimuli used in this article are complex business graphic scenes, the ROI were found out through a qualitative survey of the observers who participated in this experiment. Most observers focused on high-chroma colors corresponding to the earlier findings.[28] For example, the ROI designated by most observers for image 1 was the RGB pie chart, the CMY continents for image 2, the yellow graph, red bars, and blue background for image 3 and the orange, green, and purple colored boxes for image 4 as also given in Table IV. Each image contains different colored objects but a strong cross correlation ($r=0.94\sim0.98$) between different images comprising different contents was found as discussed earlier. It may be due to the fact that each of CMYRGB colors evenly affects perception of vividness. In other words, weights of those six colors to predict the level of vividness may equal 1/6, but a further study should be addressed to verifying this assumption.

MODELING

Modeling Vividness

Following the structure of Nayatani's empirical vividness model,[5] both chroma and achromatic intensity were selected as dependent variables in vividness modeling. Since CIELAB color space has been often used in imaging industry, metrics based upon C^*_{ab} and L^* were developed. Precisely, an independent variable ψ can be determined as a function of mean C^*_{ab} and L^* across printer primary colors, e.g., CMYRGB, as previously illustrated in Eq. (6). Those weighting factors were optimized using a linear regression method and were determined as 0.91 ($=w_c$) and 0.09 ($=w_L$). This finding is similar to the results from an earlier vividness modeling study[3] which predicts degree of vividness as a function of Munsell chroma (C) and whiteness-blackness ($[W\text{-}Bk]$) by Nayatani in 2005 as shown in Eq. (1). Figure 5 plots a linear relation between mean vividness metric predictions across the seven sample printers used in triplet comparison experiment and their corresponding subjective JND values. Pearson correlation [see Eq. (12)] between the two data sets was found to be $r=0.972$ which represents a strong linear-relation between the data sets. Equation (15) shows a matrix form of Eq. (6). The relation between vividness (**V**) and the dependent variables is determined by **T**;

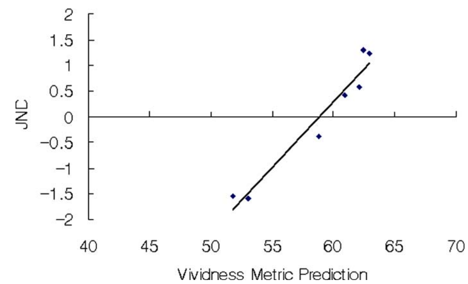

Figure 5. Relation between subjective data (JND) and vividness metric prediction for metric training ($r=0.972$).

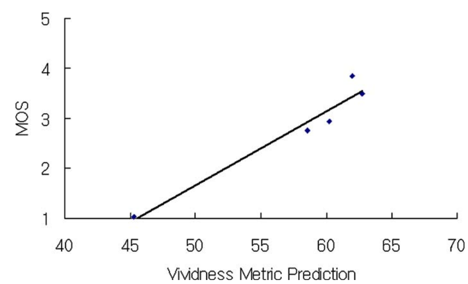

Figure 6. Relation between subjective data (MOS) and vividness metric prediction for metric generalization ($r=0.964$).

$$\mathbf{V} = \mathbf{T}\mathbf{S}, \tag{15}$$

where **S** is the corresponding mean C^*_{ab} and L^* of primary colors of a given color laser printer; it constitutes a 2×1 column matrix so the size of the transformation matrix **T** can be 1×2. This matrix **T** represents the relationship between the level of vividness achieved by a given color laser printer and the printer's physical characteristics. In other words, this relation represents how to bridge the gap between them and can let us understand the observers' taste for vividness. Mathematically, least-squares was performed to minimize residual errors between known subjective vividness scores and their corresponding metric predictions. The solution for minimizing the residual error is

$$\mathbf{T} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{V}, \tag{16}$$

where $\mathbf{S}^T$ denotes the transpose of **S**, and $\mathbf{S}^{-1}$ denotes the inverse.

In Figure 6, the relation between mean vividness metric predictions for the five sample printers used in this metric generalization procedure and their corresponding vividness scores (MOS) are shown. Pearson correlation (r) between the two data sets was found to be 0.964. This quite high correlation may be due to the fact that visual assessment for vividness is relatively easier than that for other attributes.[5] Consequently, the vividness metric developed in this article could very accurately predict subjective vividness of various color laser printers that were not used in the metric's coefficient optimization.

Modeling Observer Preference (Preferred-Vividness)

The chroma and lightness weighting coefficients, w_c and w_L, shown in Eq. (4) were reoptimized using the data set ob-

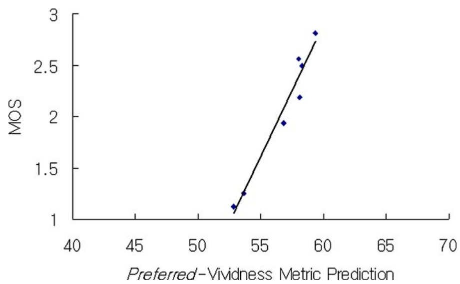

Figure 7. Relation between subjective data (MOS) and preferred-vividness metric prediction for metric training (r=0.972).

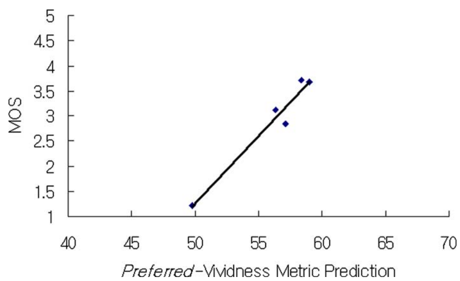

Figure 8. Relation between subjective data (MOS) and preferred-vividness metric prediction for metric generalization (r=0.978).

Table V. Comparison of weights for vividness and preferred-vividness metrics.

	w_c	w_L
Vividness	0.91	0.09
Preferred-Vividness	0.57	0.43

tained from experiment 1, which is based upon vividness-preference assessment. The original purpose of experiment 1 procedure defined in ISO 20462-2[23] was to reduce the number of sample printers. However, it was also used as a training data set for preferred-vividness[33] metric in this study. The same optimization procedure previously discussed in Eq. (6) was repeated and the weighting coefficients were determined to be 0.57 ($=w_c$) and 0.43 ($=w_L$).

Figures 7 and 8 compare the preferred-vividness metric prediction results with the corresponding subjective data for metric training and generalization, respectively. The abscissa represents prediction of preferred-vividness metric and the ordinate shows subjective data in MOS units. Pearson correlation was 0.972 for the former and 0.978 for the latter. Therefore, it can be said that accuracy of the preferred-vividness metric is high enough to estimate subjective vividness preference for various color laser printers.

Table V lists and compares the optimized weights between for vividness and preferred-vividness metrics. As can be seen, contribution of lightness is much higher for preferred-vividness (~40%) than for vividness (~10%). Apparently, a reasonably higher lightness level is also required as well as a higher chroma level to achieve a higher observer preference. Performance of printers D and H can be good examples supporting this hypothesis. According to

Table VI. Comparison of Mean C^*_{ab} and L^* and their ratio for seven sample printers.

Printer	Mean C^*_{ab}	Mean L^*	Mean L^*/C^*_{ab}	Vividness rank	Preferred-vividness rank
A	63.23	54.31	0.86	1	1
D	61.67	53.21	0.86	4	2
E	64.16	50.54	0.79	2	3
F	58.98	56.91	0.96	5	4
H	63.46	48.21	0.76	3	5
J	52.93	54.49	1.03	7	6
K	51.50	54.56	1.06	6	7

Table VII. CIECAM02 parameter settings.

	Ambient lighting (lx)	Scene white luminance (cd/m^2)	L_A (cd/m^2)	Surround
Parameter Settings	800	255	51	Average

the triplet comparison data for vividness (step 2), rankings of printers D and H were fourth and third, respectively, as indicated in bold in Table VI. However, the order was reversed in preferred-vividness data (second for D and fifth for H) due to the fact that the mean L^* of printer H is much lower than the others so its preference score was decreased despite its quite higher chroma (or vividness). It should be noted that the ratio (L^*/C^*_{ab}) of H is much lower than that of D. On the contrary, printer D obtained the second highest preference score with the aid of its high chroma and lightness values.

Comparison with CIECAM02 Colorfulness

Color vividness has been understood as the degree of colorfulness according to some references,[8,9] and there have been a number of efforts to predict colorfulness of images and its effects on the image quality.[10–12] We decided to adopt a colorfulness correlate, M, from the most recent color appearance model recommended by CIE, CIECAM02,[18] in order to compare the prediction accuracy to the vividness and preferred-vividness with that of our metric. The CIECAM02 colorfulness correlate is defined in Eq. (5). Table VII provides CIECAM02 parameter settings for this study. The ambient lighting level was approximately 800 lx ($\approx 255\ cd/m^2$) under typical office viewing conditions. The 1931 xy chromaticity coordinates of the illumination are (0.326, 0.331) near both D65 and D50. Because only hardcopy stimuli are dealt with in this study, the scene white luminance equals the surround white luminance. Thus, the luminance of the surround white is greater than 20% of the scene white, and the surround condition can be thought of as "average."[18]

There are four separate psychophysical data sets used for the test. Data 1 represents the category judgment results

Table VIII. Pearson correlation between subjective data sets and vividness metric (*VM*) and CIECAM02 *M*.

	VM	CIECAM02 *M*
Data 1	0.972	0.964
Data 2	0.971	0.989
Data 3	0.987	0.988
Data 4	0.974	0.867
Mean	0.984	0.945

from experiment 1 and data 2 comes from the triplet comparison procedure in experiment 2. The data used for generalizing vividness (experiment 3) and preferred-vividness (experiment 4) modeling, respectively, are referred to as data 3 and 4. In Table VIII, Pearson correlation coefficients between the model predictions and the four test data sets and the grand mean across the whole data are given. The grand mean of Pearson correlation for our vividness metric (0.984) was slightly higher than that of CIECAM02 *M* (0.945). Especially for data 4, our vividness metric (0.974) outperformed CIECAM02 *M* (0.867). The CIECAM02 also adopts a lightness factor, *J*, in modeling of *M*, as shown in Eqs. (4) and (5), so prediction accuracy for subjective vividness was high enough (data 2 and 3) and its performance was quite similar to our vividness metric. However, for preferred-vividness (data 1 and 4), performance of CIECAM02 M could not reach that of our metric. More comprehensive experiments should be conducted to compare the vividness metric with CIECAM02 directly in the future.

Measurement of ROI for Image Based Prediction

As previously discussed in image dependency section and Table IV, particular ROIs were reported for each image through a qualitative survey. Colorimetric values, i.e., CIELAB, of those ROIs were measured directly from the prints using a spectrophotometer, and we replaced the input parameters of our vividness metric by them. Only small area in each ROI was measured and the rest was masked so that the identical position may be repeatedly measured for different prints.

A variable n in Eq. (6) can denote the number of ROIs in the given image; in this case and C^*_{ab} and L^* represent the ROIs' colorimetric values. For example, n for image 1 equals 4 because the number of ROIs reported for the particular image was 4 as can be seen in Table IV, and weighted sum of their mean C^*_{ab} and L^* can be the vividness or preferred-vividness scores. Pearson correlation between mean metric-prediction and the MOS was 0.852 for vividness and 0.860 for preferred-vividness, so it can be said that image based ROI measurement also predicted MOS quite accurately. In case of CIECAM02 colorfulness correlate *M*, the corresponding Pearson correlations were 0.837 and 0.801. These relatively lower correlations may be due to the fact that the metric dimension is lower than for the previous procedure. In other words, six primary and secondary colors were used for computation previously, but a smaller number of ROI colors (3∼4) were used this time. However, it should be noted that Pearson correlation value of 0.8 can still be acceptable in subjective data analysis.

CONCLUSIONS

The present article develops a metric to measure vividness and observer preference (so called preferred-vividness) that can be applicable to printer quality evaluation. A number of psychophysical assessments were carried out following a procedure recommended by ISO 20462-2.[23] First, category judgment procedure was performed and the number of printer samples was reduced to seven with the consideration of performance of the sample printers. Second, a triplet comparison experiment, using the seven printers, was established to derive a precise scaling based on an interval scale by comparing triplets of given samples. There was a very strong cross correlation of subjective results between the five different image contents. Consequently, vividness and preferred-vividness were quantified as a function of mean C^*_{ab} and L^* of printer primary and secondary colors, e.g., CMYRGB. The merits and performance of both metrics were evaluated by comparison with corresponding subjective results ($r>0.96$). The metric developed in this study is based upon both chroma and lightness defined in CIELAB color space, but contribution of lightness is much higher for preferred-vividness (∼40%) than for vividness (∼10%). Contribution of lightness to the vividness metric also agrees with earlier findings by Nayatani.[5] Consequently, a reasonably higher lightness level is required as well as a higher chroma level to achieve a higher observer preference. The overall results were confirmed again using image based ROI measurement data. Prediction accuracy was compared with CIECAM02 colorfulness correlate *M*. Since CIECAM02 also adopts a lightness factor *J*, in modeling of *M*, the prediction accuracy for subjective vividness was high enough, and its performance was quite similar to that of our vividness metric. However, for preferred-vividness, performance of CIECAM02 *M* could not reach that of our metric.

For future studies, more effective colors affecting the vividness and preferred-vividness will be identified. Currently, we have used mean chroma and lightness values of printer primary and secondary colors under the assumption that each of CMYRGB colors evenly affects vividness, so weights of those six colors to predict the level of vividness may correspond to 1/6. However, more psychophysical experiments will be conducted to separate their effects based upon more uniformly designed color appearance spaces with advanced vision theories. Relation between gloss and vividness can be another interesting consideration. Glossy paper is known to enlarge the color gamut volume of prints,[34] so gloss seems very likely to be an additional significant dimension for determining vividness perception.

REFERENCES

[1] D. Nickerson and S. M. Newhall, "Central notations for ISCC-NBS color names", J. Opt. Soc. Am. **31**, 587–591 (1941).

[2] H. Kawasaki and A. Kodoma, "A study of practical color coordinate

system", *Proc. AIC Color 97*, (AIC, Kyoto, 1997) Vol **I**, pp. 467–470.
[3] Y. Nayatani, "Proposal of an opponent-color system based on color-appearance and color-vision studies", Color Res. Appl. **29**, 135–150 (2004).
[4] Y. Nayatani, "Why two kinds of color order systems are necessary?", Color Res. Appl. **30**, 295–303 (2005).
[5] Y. Nayatani, "Relationships among chromatic tone, perceived lightness, and degree of vividness", Color Res. Appl. **30**, 221–234 (2005).
[6] Commission Internationale de l'Eclairage (CIE), *International Lighting Vocabulary*, CIE Publication 17.4 (CIE, Vienna,1987).
[7] Y. Nayatani, "Adequateness of a newly modified opponent-colors theory", Color Res. Appl. **28**, 298–307 (2003).
[8] E. A. Fedorovskaya, H. de Ridder, and F. J. J. Blommaert, "Chroma variations and perceived quality of color images of natural scenes", Color Res. Appl. **22**, 96–110 (1997).
[9] A. Tremeau and C. Charrier, "Influence of chromatic changes on the perception of color image quality", Color Res. Appl. **25**, 200–213 (2000).
[10] C. Cui, "Which color is more colorful, the lighter one or the darker one?", Color Res. Appl. **28**, 168–174 (2003).
[11] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Color reproduction and the naturalness constraint", Color Res. Appl. **24**, 52–67 (1999).
[12] Y. J. Kim, M. R. Luo, P. Rhodes, V. Cheung, S. Westland, W. Choe, S. Lee, and C. Kim, "Image color-quality modeling for mobile LCDs", *Proc. IS&T/SID 14th Color Imaging Conference* (IS&T, Springfield, VA, 2006) pp. 159–164.
[13] M. D. Fairchild, *Color Appearance Models*, 2nd ed. (Wiley, New York, 2005).
[14] R. W. G. Hunt, *The Reproduction of Color*, 6th ed. (Wiley, New York, 2004).
[15] Y. Nayatani, K. Takahama, and H. Sobagaki, "Prediction of color appearance under various adapting conditions", Color Res. Appl. **11**, 62–71 (1986).
[16] Y. Nayatani, K. Takahama, and H. Sobagaki, "A nonlinear color appearance model using Estevez-Hunt-Pointer primaries", Color Res. Appl. **12**, 231–242 (1987).
[17] R. W. G. Hunt, "An improved predictor of colorfulness in a model of color vision", Color Res. Appl. **19**, 23–26 (1994).
[18] Commission Internationale de l'Eclairage (CIE), *A Color Appearance Model for Color Management Systems: Ciecam02*, CIE Publication 159-2004 (CIE, Vienna, 2004).
[19] M. S. Rea and J. P. Freyssinier-Nova, "Color rendering: A tale of two metrics", Color Res. Appl. **33**, 192–202 (2008).
[20] Commission Internationale de l'Eclairage (CIE), *Method of Measuring and Specifying Color Rendering Properties of Light Sources*, CIE Publication 13.3 (CIE, Vienna, 1995).
[21] International Digital Enterprise Alliance (IDEAlliance), *Specifications Web Offset Publications (Swop) 2007 Spec*, 11th ed. (IDEAllinace, Alexandria, 2007).
[22] Commission Internationale de l'Eclairage (CIE), *Colorimetry*, 2nd ed., CIE Publication 15.2 (CIE, Vienna, 1986).
[23] International organization for standardization (ISO), *Photography: Psychophysical Experimental Methods for Estimating Image Quality, Part 2: Triplet Comparison Method, ISO 20462-2* (ISO, Geneva, 2005).
[24] International telecommunication union (ITU), *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Rec. BT. 500–10 (ITU, Geneva, 2002).
[25] L. L. Thurstone, "A law of comparative judgment", Psychol. Rev. **34**, 273–286 (1927).
[26] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 color-difference formula: CIEDE2000", Color Res. Appl. **26**, 340–350 (2001).
[27] M. R. Luo, G. Cui, and C. Li, "Uniform color spaces based on CIECAM02 color appearance model", Color Res. Appl. **31**, 320–330 (2006).
[28] Y. J. Kim, M. R. Luo, P. Rhodes, V. Cheung, S. Westland, W. Choe, S. Lee, S. Lee, Y. Kwak, D. Park, and C. Kim, "Image-color quality modeling under various surround conditions for a 2-inch mobile transmissive LCD", J. Soc. Inf. Disp. **15**, 1–8 (2007).
[29] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using Matlab* (Pearson-Prentice Hall, Upper Saddle River, NJ, 2004).
[30] K. M. Braun, "Memory color enhancement algorithm", *Proc. International Congress of Imaging Science, ICIS 2006* (IS&T, Springfield, VA, 2006) pp. 277–279.
[31] Y. Kwak, J. Hong, D. Park, and C. Y. Kim, "Preferred memory and accent colors shown on the display and their size effect", *Proc. IS&T/SID 14th Color Imaging Conference*, (IS&T, Springfield, VA, 2006) pp. 224–229.
[32] Y. J. Kim, M. R. Luo, P. Rhodes, W. Choe, H. S. Kim, S. O. Park, Y. Baek, S. Lee, and C. Kim, "Factors affecting the psychophysical image quality evaluation of mobile phone displays: the case of transmissive liquid-crystal displays", J. Opt. Soc. Am. **25**, 2215–2222 (2008).
[33] Y. J. Kim, Y. Bang, and H. Choh, "Modeling vividness perception for color laser printer evaluation", *Proc. IS&T/SID 16th Color Imaging Conference* (IS&T, Springfield, VA, 2008) pp. 299–303.
[34] E. Perales, F. M. Martinez-Verdu, V. Viqueira, J. Fenandez-Reche, J. A. Diaz, and J. Uroz, "Comparison of color gamuts among several types of paper with the same printing technology", Color Res. Appl. **34**, 330–336 (2009).