An Investigation of Color Models' Performance between Monitor and Reflection Images

Mei-Chun Lo, Ruey-Kuen Perng and Yu Su♦

Dept. of Graphic Communications and Technology, Shih Hsin University
♦Graphic Arts Section, Master of Graphic Institute of Paper-making & Graphic Arts
Taipei, Taiwan, R.O.C.

Key Words: Device characterization, color reproduction, colorimetry, color appearance model, chromatic adaptation transform, uniform color space, softcopy, reflection original, paired comparison, category judgment, binocular simultaneous matching, color management system.

ABSTARCT

An earlier work (Lo et. al. 1999a) presented in TAGA'99 was carried out using single stimuli to study color models' performance between a self-luminous reproduced softcopy and the reflection original. The results were used to verify or modify the CIECAM97s color model found performing the best. In practice, however, complex images are frequently used in areas such as graphic arts and desktop publishing. Hence, a set of new experiments is conducted in the viewing conditions of dim surrounds using complex images.

Five color models are tested in this study. They were CIELAB, von Kries, LLAB, ZLAB, and CIECAM97s. A flatbed scanner HP 4C/T, characterized under each viewing condition tested by a forward 3rd-SVD (Singular Value Decomposition) masking type of characterization model (Lo et. al. 1999b), was used to obtain image data in terms of RGB format. In addition, the PLCC model was used to characterize the Barco monitor used in this study to display images tested.

The work was conducted closely following the CIE guidelines (Alessi 1994). The matching performance of the softcopy processed by a particular color model was judged against a standard (original) illuminated in a viewing cabinet (dim surround) by a panel of observers. Also the results will be used to confirm to the

findings obtained using single stimuli and further modify the best performing model found in this study in future works. The final aim of these works alongside a framework successively conducted is to derive a reliable color appearance model, efficiently accounting for disparate media/viewing conditions.

INTRODUCTION

CRT monitors especially HDTVs, are often used in the graphic arts industry as a soft-proofing device for previewing and editing the image before the hardcopy is produced. Practically speaking, i.e., the softcopy on a monitor is widely used to simulate and communicate how the colors will appear in the final printed reproduction. The core technique, therefore, to achieve high color-fidelity images reproduced using color monitors requires a reliable color appearance model, capable of predicting the change of perceived attributes of color appearance under softcopy and original (reflection or transparency) or hardcopy dissimilar viewing conditions.

An earlier work (Lo et. al. 1999a) was carried out using single stimuli to study color models' performance between a self-luminous reproduced softcopy and the reflection original. The color model found performing the best was verified or modified using the results obtained. A set of new experiments were further conducted in the viewing condition of dim surrounds using complex images which, In practice, are frequently used in areas such as graphic arts and desktop publishing.

COLOR APPEARANCE AND COLOR MODELS

In most practical situations, the color reproduction aims to reproduce the appearance of colors or color images and not some physical properties of stimuli. The CIE XYZ system, however, does not deal with quantities derived from the appearance of stimuli, but only basically their physical properties. Also it only deals with individual stimuli. Hence, there is a need to derive color models to predict the perceptual color appearance attributes of stimuli, considered by taking into account the influence of the environment in which they are seen.

Color models may be classified into four categories; i.e. CIE XYZ system, uniform color spaces (UCS), chromatic-adaptation transforms (CAT), and color appearance models (CAM). Five color models were tested in this study. These are as follows.

- UCS: CIELAB
- CAT: von Kries (Helson et al.)
- CAM: ZLAB (Fairchild 1997), LLAB (Lo 1995; Luo et al. 1996), CIECAM97s (Hunt 1191; Hunt 1994; Hunt and Luo 1997)

EXPERIMENTAL PREPARATION

Device Characterization

The original image data tested were obtained using a HP 4C/T flatbed scanner in terms of RGB format. A forward masking type of 3rd_SVD_lightness (Singular Value Decomposition) model previously derived (masking type) was used to characterize the scanner (Lo et. al. 1998; Lo et. al. 1999b). It applies subdivision approaches including both of highlight and shadow near-neutral-divisions, chromatic lightness-divisions, and achromatic gray-balance LUT (as shown in Fig. 1). A satisfactory prediction performance of the forward 3rd_SVD_Lightness model (predicted tristimulus values from a set of RGB primaries, i.e. RGB to XYZ) was found with mean ΔE CMC and ΔE CIE94 values of 1.2 and 0.96 respectively. The scanner RGB intensities were, therefore, transformed to the original XYZ values using the 3rd_SVD_Lightness characterization model derived.

The experiments were conducted under dim surrounds. A Barco monitor was used to display screen colors or images. The PLCC monitor model was applied to characterize the monitor tested (Lo et. al. 1998). The measurement equipment used was a PR650 spectrophotometer, manufactured by Photo Research. A set of characterization data was produced for each of white points tested in this study. The Barco monitor was internally set to each color temperature investigated. For each white point, two sets of characterization samples were produced, including 729 (cube) and 54 colors. The latter was composed of 18 samples for each of red, green, and blue channels with interval of 15 DACs ranging from 0 to 255 values. The cube data set included 9×9×9 colors, selected to give approximately an equal perceived difference between neighboring samples.

Image Preparation and Processing

Five originals (reflection prints, see Plate 1) were selected for studying models' performance . An image processing software was developed to correlate the scanner's original RGB values (in the reference field) to XYZ values (using scanner model mentioned earlier), and then via X′Y′Z′ or LCH (lightness, chroma, and hue) values obtained using each particular color model studied to

Church

Sharpness

Bride

Winter Stuff

Dining Table

Plate 1 Complex images selected for testing models' performance the monitor's RGB intensities (in the test field) (using the PLCC monitor model). It is on a pixel-by-pixel basis. The XYZ and X′Y′Z′ represent the tristimulus values for reflection original and reproduced softcopy respectively. They form a set of corresponding colors having the same color appearance when viewed under two adapting fields. The scanner and monitor characterization models described earlier are used to convert scanner's RGB values to XYZ, and X′Y′Z′ to RGB intensities, respectively.

The computational procedures for different categories of color models are varied (Lo et. al. 1999a). For chromatic adaptation transforms, the XYZ values are directly transformed to X′Y′Z′. As for each uniform color space or color appearance model, the predicted lightness, chroma, and hue are first calculated via its forward model and followed by computing its corresponding X′Y′Z′ using the model's reverse.

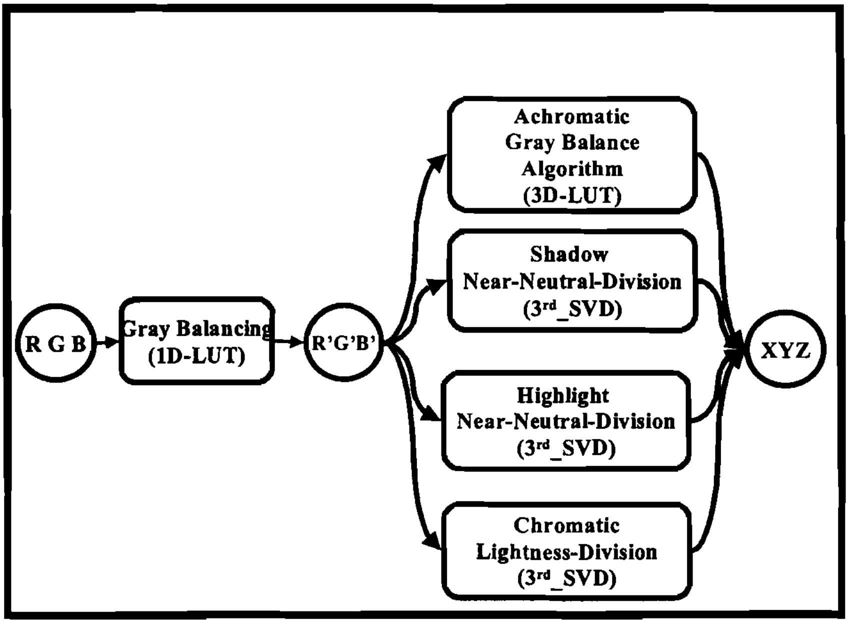

Fig. 1 Computational procedures of subdivision approaches used in the scanner 3^{rd}_SVD_lightness model derived.

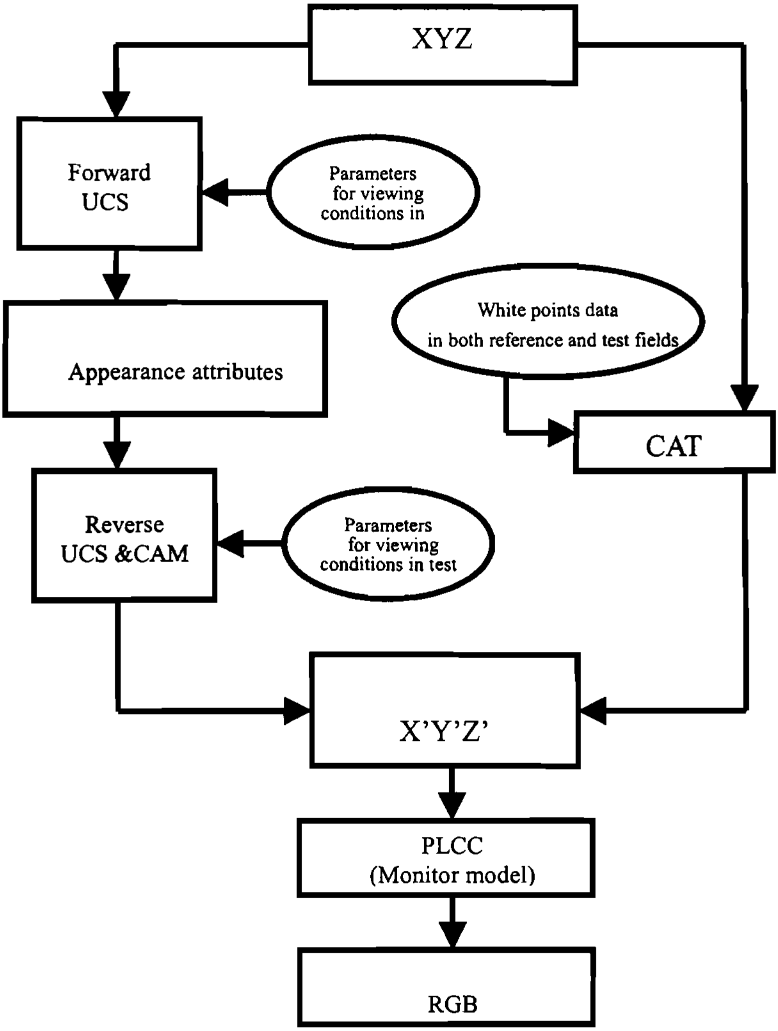

Fig. 2 Computational procedures used in the image processing correlating the original hardcopy's XYZ values to the reproduced softcopy's RGB intensities for different category of color models tested.

EXPERIMENT

Viewing Configuration

Fig. 3 illustrates the display arrangement configured in the experiment. Two tested softcopy images were simultaneously displayed; and the physical size of each screen image was the same as that of original. Both of original and two softcopies were displayed side-by-side on gray background fields with approximate L* of 50.

Fig. 3. The experimental viewing configuration.

Table 1 Summary of experimental phases.

Phase	White Points or Light Sources		No. of observers	Repetition	Pairs	No. of comparisons
	Reference Field	Test Field				
1	D50	D65	10	2	100	1,000
2	D50	D93	10	2	100	1,000
3	A	D65	10	2	100	1,000
4	A	D93	10	2	100	1,000
Total					400	4,000

Table 2 Correlated color temperatures (CCTs), luminances and colorimetric data used for both reference and test fields in each phase under a dim surround.

Phase	1	2	3	4
Reference Field				
White Point	D50	D50	A	A
CCT	4677K	4677K	2229K	2229K
Luminance (cd/m^2)	60.85	60.85	58.75	58.75
X	100.15	100.15	121.14	121.14
Y	100.00	100.00	100.00	100.00
Z	82.79	82.79	20.07	20.07
x	0.3540	0.3540	0.5022	0.5022
y	0.3534	0.3534	0.4146	0.4146
u'	0.2167	0.2167	0.2882	0.2882
v'	0.4869	0.4869	0.5353	0.5353
Test Field				
White Point	D65	D93	D65	D93
CCT	6287K	8933K	6287K	8933K
Luminance (cd/m^2)	64.52	64.87	64.52	64.87
X	94.85	94.73	94.85	94.73
Y	100.00	100.00	100.00	100.00
Z	105.10	137.37	105.10	137.37
x	0.3162	0.2852	0.3162	0.2852
y	0.3334	0.3011	0.3334	0.3011
u'	0.1986	0.1888	0.1986	0.1888
v'	0.4712	0.4485	0.4712	0.4485

Viewing Conditions and Viewing Techniques

A set of experiments is conducted under dim surrounds. It is divided into 4 phases according to different light sources and white points used in the viewing reflection originals (in the reference field) and screen softcopy images (in the test field). Table 1 summarizes the differences among all phases. For instance, a D_{50} simulator illuminates in the reference field; and the white point of the test field (monitor) sets to the chromaticities of the D_{65} simulator in Phase 1.

A panel of 10 observers attended and repeatedly made the assessments twice in each of 4 phases. In total, 4,000 comparisons were made. The binocular simultaneous matching technique, as practicing in graphic arts industry, was applied. Each observer was asked to compare two reproduced images simultaneous displayed on monitor to the corresponding original hardcopy in viewing cabinet as shown in Fig. 3, and made judgments. Each observer was instructed to use both eyes to look at one field at a time, but could switch between two fields at any time. The correlated color temperatures (CCTs), luminances and colorimetric data used for both of reference and test fields are shown in Table 2. Additionally, Table 3 tabulates some of those used for gray backgrounds in the viewing cabinet. Practically, in graphic arts industry, the reproductions produced and the originals are displayed simultaneously when making comparisons. Therefore, in this study as practicing in graphic arts industry, two tested softcopy images and the original with the same contents were compared side by side in all phases; i.e. they were displayed simultaneously on gray background fields with approximate L* of 50 as mentioned earlier, even though the asymmetrical viewing conditions wherein different chromaticities of white points were used for the hardcopy (source) and softcopy (destination) fields. The parameters accounting for effects of such as cognitive or sensory chromatic mechanism, surroundings etc. used in each color appearance tested are tabulated in Table 4.

Table 3 Colorimetric data and luminances used in reference field in each phase under a dim surround.

Phases	Sources	*x*	*y*	*L* (cd/m^2)	*X*	*Y*	*Z*
1 & 2	D50	0.3414	0.3511	15.900	25.407	26.130	22.892
3 & 4	A	0.4981	0.4151	16.160	33.004	27.506	5.755

Table 4 Parameters used for CIECAM97s, ZLAB and LLAB color appearance models in the experiments.

Phase	1	2	3	4
CIECAM97s				
Reference Filed				
L_A	15.90	15.90	16.16	16.16
F	0.65	0.55	0.55	0.55
c	0.69	0.69	0.69	0.69
F_{LL}	1.0	1.0	1.0	1.0
N_c	1.0	1.0	1.0	1.0
Test Field				
L_A	15.90	15.90	16.16	16.16
F	0.65	0.55	0.55	0.55
c	0.59	0.59	0.59	0.59
F_{LL}	1.0	1.0	1.0	1.0
N_c	0.95	0.95	0.95	0.95
ZLAB				
Reference Field				
L_A	15.90	15.90	16.16	16.16
F	0.65	0.55	0.55	0.55
$1/\sigma$	0.69	0.69	0.69	0.69
Test Field				
L_A	15.90	15.90	16.16	16.16
F	0.65	0.55	0.55	0.55
$1/\sigma$	0.59	0.59	0.59	0.59
LLAB				
Reference Field				
D	0.58	0.456	0.456	0.456
F_s	3.0	3.0	3.0	3.0
F_L	1.0	1.0	1.0	1.0
F_C	1.0	1.0	1.0	1.0
Test Field				
D	0.58	0.456	0.456	0.456
F_s	3.5	3.5	3.5	3.5
F_L	1.0	1.0	1.0	1.0
F_C	1.0	1.0	1.0	1.0

Note: L_A: Photopic luminance of adapting field
F: Factor for degree of adaptation
c: Impact of surround
F_{LL}: Lightness contrast factor
N_c: Chromatic-induction factor
D: Degree adaptation
σ: Impact of surround
F_s: Surround induction factor
F_L: Lightness induction factor
F_C: Chroma induction factor

DATA ANALYSIS

A forced-choice paired comparison was performed. Observers viewed a pair of reproduced images, processed using two of the models and displayed on the screen simultaneously, and judged which of the two gave a better appearance match to the illuminated reflection original. The order of the pairs was randomized when presented to each observer. In addition, observers also rated the color-fidelity quality of each reproduced image, against its corresponding reflection original (hardcopy), on a Chinese word category scale from "完美" (exact match) to "極差" (awful match). The visual responses were entered via radio buttons showing word category scales (Fig. 4), which were positioned underneath the corresponding viewing images accessed. The visual resulted data were first converted to a 7-point numerical type category scale anchored words as follows (Table 5).

Data analysis was then carried out using both the laws of comparative (Thurstone 1927) and the categorical judgment (Torgerson 1958). Both of the analyzed results were presented in terms of z scores to represent the predicted performance of color models.

Category scaling is generally considered to yield at least ordinal level data (Meilgaard et. al. 1991). Therefore they are not normally distributed. Nevertheless, it is assumed that the data for category scale would be forced to fit the statistical normal distribution, accomplished by dealing with the accumulated, if they were transformed in Torgenson's method. However the Torgenson's law of categorical judgment correlates the relative position of a specified attribute, of the stimuli considered with respect to category boundaries on the psychological continuum rather than with respect to one another as Thurstone's law of comparative judgment does. Therefore the results obtained in this study using Torgenson's approach do not provide the measure of accuracy degree (how much) one reproduced image has more than another for the color fidelity. Only they serve as a basis for locating the models studied on the absolute scales (category) of color fidelity quality. Clearly speaking, if the relative positions of color-fidelity rendition of two models in question on the psychological continuum were with respect to the same category boundaries, then these two models are considered to have the similar rendering performance of the color-fidelity. Therefore they would be located on the same category rank.

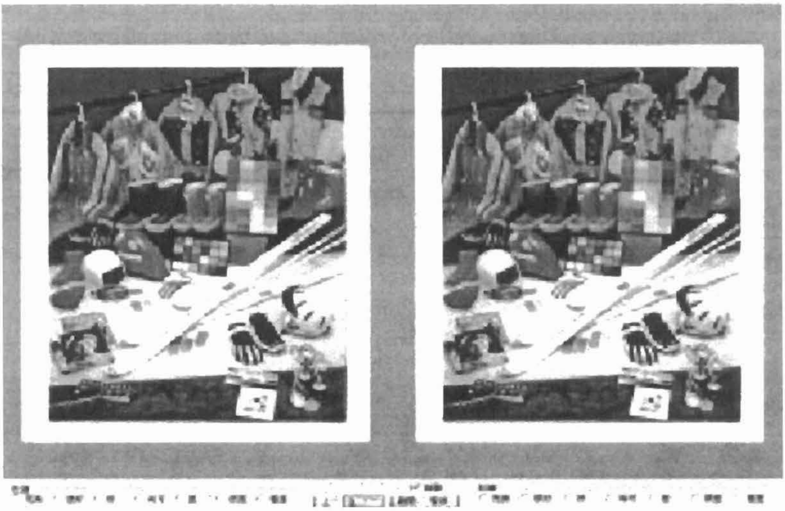

Fig. 4 Experimental arrangement configured on the Barco monitor for assessing color-fidelity quality in terms of a Chinese word category scale from "完美" (exact match), through "尙好" (acceptable match" to "極差" (awful match) (a 7-point category scale).

Table 5 7-point of category scales accessed based on color-fidelity quality anchored with both words in Chinese and English.

Category scale	English word category scale	Chinese word category scale
1	Exact Match	完美
2	Good	很好
3	Moderate Match	好
4	Acceptable Match	尙好
5	Poor Match	差
6	Bad Match	很差
7	Awful Match	極差

Table 6 Paired comparison results in terms of Z-score scale based on the judgments of the overall color-fidelity accuracy obtained from all 5 complex images combined.

Phase	Reference Field	Test Field	von Kries	CIELAB	ZLAB	LLAB	CIECAM97s
1	D50	D65	-2.8853	-2.5732	1.3169	1.3079	2.8338
		Rank	**4**	**4**	**2**	**2**	**1**
2	D50	D93	-4.1363	-2.6898	1.8476	2.1253	2.8531
		Rank	**5**	**4**	**2**	**2**	**1**
3	A	D65	-5.9506	-7.5619	3.7171	4.1495	5.6460
		Rank	**4**	**5**	**2**	**2**	**1**
4	A	D93	-6.7064	-6.4338	3.6294	4.2190	5.2919
		Rank	**4**	**4**	**2**	**2**	**1**

RESULTS AND DISCUSSIONS

Table 6 summaries the results in terms of Z-score obtained using the law of the comparative judgment. The overall results are also depicted in Fig. 5 for Phases 1 to 4. Each point in these figures represents the Z-score of a model of interest, and a line drawn indicates its 95% confidence limit (CL). Two models tested are considered not to be significantly different each other if the Z-score of one model is within the 95% confidence limit (CL, i.e. 2 units of standard deviations) of the other. Hence, they were ordered in the same rank.

Clearly, from the overall results shown in both Table 6 and Fig. 5, it can be seen that the CIECAM97s model outperformed the others in all phases. This confirms to what has been found in earlier experiments conducted using single stimuli (Lo et. al. 1999b). Both the ZLAB and the LLAB gave reasonable predictions and performed similar. Both the von Kries and the CIELAB performed the worst in all phases.

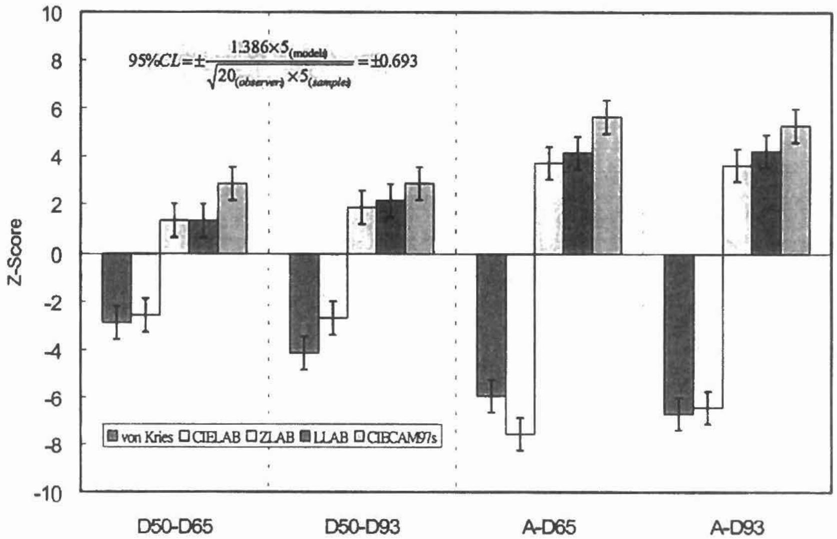

Fig. 5 Color models performance evaluated using the law of Turstone's law of comparative judgment for the overall color-fidelity accuracy of all 5 tested images from Phase 1 to 4 (including 95% confidence limit).

In addition, empirical estimates of both the category scale values of the models and the category boundaries of 7-point scale obtained using the Torgerson approach are summaries in Table 7. As mentioned earlier, they were used to locate the rendering category of the color-fidelity performance of the models considered on the psychological continuum. Still, as found from the earlier experiments tested using single stimuli, It is encouraging to see the model performing the best shown in the results obtained using the law of comparative judgment, also had top rank in those obtained using the approach of category judgment for all 4 phases studied. Moreover, the CIECAM97s model performing the best was located on the category scale of either "Moderate Match" or "good match"

Table 7 Category ranking results, including both (a) boundary estimates (T_i) and (b) category scales, analyzed using the Torgerson's law of categorical judgments, based on the data obtained from the accessment on the degree of match (category) for the color-fidelity quality (from all 5 tested complex images combined).

(a) Boundary estimates

Phase	Reference Field	Test Field	T_1	T_2	T_3	T_4	T_5	T_6
1	D50	D65	0.0000	1.0384	1.8706	2.6365	3.5091	4.9062
2	D50	D93	0.0000	0.9517	1.7825	2.5896	3.6164	5.0187
3	A	D65	0.0000	0.5184	1.6287	3.0908	4.0547	5.0475
4	A	D93	0.0000	0.8919	2.0772	3.0504	4.2148	5.2207

(b) Category scales

Phase	Reference Field	Test Field	von Kries	CIELAB	ZLAB	LLAB	CIECAM97s
1	D50	D65	2.0516	1.9646	2.8322	2.8857	3.3266
		Rank	**4**	**4**	**3**	**3**	**3**
2	D50	D93	1.4637	1.6696	2.8396	2.9636	3.2157
		Rank	**5**	**5**	**3**	**3**	**3**
3	A	D65	1.6331	1.1921	3.7638	3.7468	4.1459
		Rank	**4**	**5**	**3**	**3**	**2**
4	A	D93	1.1645	1.0114	2.9789	2.9687	3.2337
		Rank	**5**	**5**	**4**	**4**	**3**

SUMMARY

The work presented in this extended abstract was conducted closely following the CIE guidelines (Alessi 1994). The matching performance of the softcopy processed by a particular color model was judged against a standard (original) viewed under in a viewing cabinet (dim surrounds) by a panel of observers. Also the results were used to confirm to the findings obtained using single stimuli and further the best performing model found in this study will be modified in a future works found in this study. The final aim of these works alongside a framework successively conducted is to derive a reliable color appearance model, efficiently accounting for disparate media/viewing conditions.

ACKNOWEDGEMENTS

This research has been mainly financed by both the National Science Council and the Industrial Technology Research Institute, Taiwan, R.O.C..

REFERENCES

1. Alessi, P. J., CIE guidelines for coordinated research on evaluation of colour appearance models for reflection prints and self-luminous display image comparisons, Central Bureau of the CIE, *Color Res. Appl.*, **19**, 48-58 (1994)
2. Fairchild, M. D., The ZLAB colour appearance model for practical image reproduction applications, *CIE Expert Symposium '97 Colour Standard for Imaging Technology*, 89-94 (1997)
3. Helson, H., Judd, D. B., and Warren, M. H., Object-color changes from daylight to incandescent filament illumination, *Illum. Eng.*, **47**, 221-233 (1952)
4. Hunt, R. W. G. and Luo, M. R., The structure and use of the CIE colour appearance model (CIECAM97s), *CIE Expert Symposium '97 Colour Standard for Imaging Technology*, 81-88 (1997)
5. Hunt, R. W. G., An improved predictor of colourfulness in a model of colour vision, *Color Res. Appl.*, **19**, 23-26 (1994)
6. Hunt, R. W. G., Revised colour-appearance model for related and unrelated colour *Color Res. Appl.*, **16**, 146-165 (1991)
7. Lo, M. C. and Chiang, J. C., The characterization models for multi-colored CMYKRGB printing process, TAGA Proc., 242-254 (1998)
8. Lo, M. C., Chen, Y. L., and Lee, T. T., The characterization models and quantification of color appearance on HDTV, *IPC Proc.*, 461-463 (1998).

9. Lo, M. C., Perng, R. K. and Chiang, J. C., A study of color models performance using single stimuli, *TAGA Proc.*, 606-622 (1999a)
10. Lo, M. C., Perng, R. K. and Chen, Y. L., The characterization of color scanner devices, Optics and Photonics Taiwan'99, 1052-1054 (1999b).
11. Lo, M. C., Perng, R. K. and Chen, Y. L., Characterization models of color monitor devices, TAIPEI International Conference on Graphic 1999 communications, 000-000 (1999c).
12. Lo, M. C., The LLAB Models for Quantifying Colour Appearance, Doctoral Thesis of Loughborough University (1995)
13. Luo, M. R., Lo, M. C., and Kuo, W. G., The LLAB (*l:c*) color model, *Color Res. Appl.*, **21**, 412-429 (1996).
14. Post, D. L. and Calhoun, C. S., An evaluation of methods for producing desired colours on CRT monitors, *Color Res. Appl.*, **14**, 172-186 (1989).
15. Thurstone, L. L., A law of comparative judgement, *Psycholog. Rev.*, **34**, 273-286 (1927).
16. Torgerson, W. S., Theory and Methods of Scaling, John Willey & Sons, New York, 159-246 (1958).
17. Meilgarrd, M, Civille, G. V. and Carr, B. T., Sensory Evaluation Techniques, 2^{nd}, CRC Press LLC, 53 (1991).