ASC: 99-0581

Further Evaluation of Methods for Producing Desired Colors on CRT Monitors

David L. Post,[1]* Christopher S. Calhoun[2]

[1] Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433-7022

[2] Logicon Technical Services, Inc., Dayton, Ohio 45437-7258

Received 13 March 1999; accepted 16 July 1999

Abstract: Three experiments were performed to evaluate methods for predicting the luminances and chromaticity coordinates produced by color CRT monitors, given known inputs. Linear and logarithmic versions of PLCC and PLVC, plus Berns, Motta, and Gorzynski's power function, were tested. Estimates are provided for the number of CRT measurement points needed to maximize each method's predictive accuracy. Correcting for unintended light from the monitor is shown to improve accuracy substantially for a case involving a seemingly small amount of light. Berns et al.*'s characterization technique, which involves measuring the monitor's neutral point, is shown to yield the same accuracy as conventional characterization while reducing the number of measurements required, and to yield improved accuracy when correction for unintended light is needed but impractical. The accuracies of the predictive methods are compared and recommendations for their use are provided.* © 2000 John Wiley & Sons, Inc.* Col Res Appl, 25, 90–104, 2000

Key words: characterization; color; colorimetry; CRT; display; modeling; monitor; predicting

INTRODUCTION

In an earlier article,[1] we compared several methods for predicting the colorimetric behavior of digital-graphics systems that display output on color CRT monitors (DG/CRT systems). We found that two methods we called *piecewise linear interpolation assuming constant chromaticity coordinates* (PLCC) and *piecewise linear interpolation assuming variable chromaticity coordinates* (PLVC) gave the most accurate predictions. PLCC uses the assumption that the chromaticity coordinates of the DG/CRT system's red, green, and blue color channels (*R*, *G*, and *B*) are constant—a simplification that makes PLCC easy to implement in software. PLVC omits the constant channel-chromaticity assumption and yields more accurate predictions of chromaticity at the expense of programming ease. (PLVC is described in more detail in Appendix A.) We also found that 16 measurements of each color channel, spaced equally between the smallest digital-to-analog converter (DAC) value that can be measured reliably and the largest possible DAC value, were sufficient to maximize PLCC and PLVC's predictive accuracies.

More recently, other researchers have reported new ideas that deserve investigation. For example, Lucassen and Walraven[2] have demonstrated a variation on PLCC in which log luminance is interpolated as a function of log DAC-value—a method we label PLGCC. This refinement might be expected to yield better accuracy than PLCC, because a log–log transform tends to linearize DAC-value → luminance transfer functions (the degree of success depends on the transfer functions[3]). Applying the same modification to PLVC, yielding a method we label PLGVC, might improve its accuracy, too.

Another important development has been Berns, Motta, and Gorzynski's[3] introduction of a predictive method that uses a power function and has become a *de facto* CIE recommendation.[4] Berns[5] later named this the *gain-offset-gamma* (GOG) model. Berns *et al.*'s analysis indicated that GOG is as accurate as PLCC, but requires a total of only eight measurements to achieve this accuracy, when it is used in conjunction with a new characterization technique. This technique involves measuring the tristimulus values of the DG/CRT system's neutral point *(N)* at five DAC triplets for which $D_R = D_G = D_B \neq 0$ (where D_R represents an *R* DAC value, etc.) and performing one measurement each of *R*, *G*,

* Correspondence to: David L. Post, AFRL/HECV, 2255 H St., Room 300, WPAFB OH 45433-7022 (e-mail: david.post@he.wpafb.af.mil)
© 2000 John Wiley & Sons, Inc. *This article is a US Government work and, as such is in the public domain in the United States of America.

and B alone to determine the channel chromaticities. Given the assumption of constant channel chromaticity, the results from the individual R, G, and B measurements can be used to decompose the five measurements of N into their assumed constituent RGB luminances, yielding five luminance values for all three channels with only five measurements. This procedure requires fewer measurements than the more common characterization technique, which involves measuring R, G, and B alone at multiple DAC values with the other two channels set to zero, and is easy to automate in software.

Still another important innovation has resulted from the realization that measuring a given screen location may yield a nonzero reading, even though the DAC values associated with that location are zero, the measuring environment is dark, and the measuring instrument is zeroed. This situation can produce large prediction errors and has several possible causes: (1) the monitor's black level, or DC offset, is too high; (2) the cutoff voltages produced by the graphics card's DACs are too high; (3) light from an adjacent screen area backscatters from the faceplate onto the location being measured (a phenomenon known as *internal* or *interreflection flare*); (4) electrons from an adjacent screen area backscatter from the shadowmask or aluminized screen and excite the measuring location; and (5) light from a source outside the intended measuring location is collected by the measuring instrument's optics (a phenomenon called *instrument flare*). Howard[6,7] was apparently the first author to correct for such problems, which we refer to collectively as *black light*, and others have followed suit.[4,5,8,9] Del Barco, Díaz, Jiménez, and Rubiño,[10] in particular, have provided a detailed analysis that shows the improvement in prediction accuracy that black correction can produce.

The present experiments were designed to update our earlier evaluation[1] by trying out the ideas discussed above. Specifically, we sought to quantify the effects of black correction and neutral-point characterization on the accuracies of the aforementioned predictive methods, obtain refined estimates of the optimal number of characterization measurements, and compare results across methods. Our overall purpose is to provide information that is useful in selecting a method, obtaining the best performance from it, and developing more accurate methods.

APPARATUS

Video Generation

Video signals were produced by a Univision model UDC7032A-10C graphics board operating under the control of our software. Each image plane on the graphics board generates 1024 × 1024 8-bit pixels at a 60-Hz noninterlaced refresh rate. The board is equipped with an Analog Devices model ADV7150 135-MHz video RAM-DAC, which contains a 3 × 256 × 10-bit lookup table that addresses three 10-bit DACs. Thus, the largest available DAC value for each color channel (D_{max}) is 1023.

Colorimetric Instrument

Colorimetric measurements were performed with an LMT model 1210 filter colorimeter, equipped with an RS-232 interface to permit external computer control. The LMT's optical head contains no lenses for focusing light and, therefore, does not yield luminance measurements; instead, it gathers light using a diffusing integrating surface and measures illuminance. Therefore, when measuring with the LMT, we placed its optical head flush with the CRT monitor's faceplate to maximize the signal. The LMT luminance measurements we report subsequently were obtained using an illuminance-to-luminance conversion constant, computed by measuring the display at $D_R = D_G = D_B = 1023$ with the LMT and a spectroradiometer and ratioing the measurements.

After loading each new set of DAC values into our graphics board, we allowed 90 s to elapse before commencing measurement to reduce the effects of short-term monitor instability. Our software then commanded the LMT to perform four measurements in succession and return the averaged results. All measurements were performed in a dark room.

Display

We tested a 19-inch (48-cm) Aydin model 9000 color CRT monitor. It was adjusted to provide a 1:1 aspect ratio, in keeping with the graphics board's 1024 × 1024 addressability, and to yield a D_{65} peak white (W_{max}) having a luminance of approximately 72 cd/m^2 at $D_R = D_G = D_B = D_{max}$ and a perceptual black when the entire screen was set to $D_R = D_G = D_B = 0$ and observed for 5 min in the dark room. The monitor was warmed up 24 h prior to data collection to improve its temporal stability.

Measuring Target

The measuring target was a solid circle drawn at the center of the CRT monitor's screen against a gray background having DAC values = (590, 590, 590). These values produced a luminance that was roughly 20% of W_{max}. The circle was 400 pixels in diameter and, therefore, occupied approximately 12% of the display's active area. This measuring target and background are similar with those used by Berns *et al.*[3] Before each measuring sequence, the LMT's optical head was positioned flush with the faceplate and centered within the circle.

PROCEDURE

Our data collection was designed to produce results that can be compared with those reported by Berns *et al.*[3] Our basic procedure was to perform a series of characterization measurements, followed immediately by measurement of a set of test colors. This process was repeated five times. The characterization data from each replication were used to predict the luminances and chromaticity coordinates of the

associated test colors, using various methods, and two dependent measures were computed for each test color: (1) absolute value of the percent luminance error; and (2) chromaticity error, expressed as distance on the CIE 1976 uniform chromaticity-scale (UCS) diagram.

Characterization

For each replication, the DG/CRT system was characterized by measuring each color channel's luminance and CIE 1976 u′v′-chromaticity coordinates at several DAC values with the other two channels set to zero. Specifically, we measured *R, G,* and *B* at 49 DAC values each that were spaced equally between the smallest DAC value that could be measured reliably (380, specifically) and D_{max}. The advantage of measuring 49 points in this way is that one can form subsets containing 25, 17, 13, 9, 7, 5, 4, 3, or 2 measurements, all of which are spaced equally between and include the smallest measurable and largest possible DAC values.

We also measured the neutral points (*N*) associated with the aforementioned DAC values. That is, we measured the 49 DAC-value triplets for which $D_{R,i} = D_{G,i} = D_{B,i}$, where *i* denotes one of the 49 DAC values. Each *N* measurement was then decomposed into its constituent *RGB* luminances, yielding a second characterization dataset.

Test Colors

Our test dataset consisted of the same DAC triplets as Berns *et al.*[3], except for (0, 0, 0).* Specifically, we measured the DAC values 0, 559, 755, 903, and 1023 in all possible three-way combinations, except for (0, 0, 0), yielding 124 measurements. The resulting luminances and chromaticity coordinates were different from Berns *et al.*'s presumably, because our monitor's peak luminance (72 cd/m^2) was somewhat higher than theirs (65 cd/m^2) and because of probable differences in the monitors' phosphors and voltage-to-luminance transfer functions.

Predictive Methods

We replicated Berns *et al.*'s[3] comparison of GOG using the conventional approach to characterization vs. the use of neutral-point measurements. (We made a correction to their neutral-point characterization procedure, though; it is discussed in Appendix B.) We also tested PLCC, PLGCC, PLVC, and PLGVC using conventional and neutral-point characterizations. The chromaticity coordinates obtained at D_{max} for each channel were used to represent the coordinates of the *RGB* channels for GOG, PLCC, and PLGCC, which assume constant channel chromaticity. PLVC and PLGVC used the coordinates that were measured at each DAC value.

The idea of using neutral-point characterization with PLVC (or PLGVC) might seem nonsensical because the decomposition process, as we have described it, relies on the assumption that the *RGB* channel chromaticities are constant. Therefore, a characterization dataset produced by this process contains *R, G,* and *B* measurements that have constant chromaticity—at least implicitly—and using such a dataset with PLVC yields the same results as PLCC.

It is possible, though, to combine information from conventional and neutral-point characterizations to obtain a decomposed neutral-point characterization dataset that contains information about measured changes in channel chromaticity. The procedure is: (1) Use PLVC and the results from a conventional characterization to compute predicted *RGB* DAC values for each *N* measurement; (2) Use PLVC and the same conventional characterization to predict the luminances and chromaticity coordinates of the individual *R, G,* and *B* DAC values computed in step 1; and (3) Substitute the DAC values at which each *N* measurement was performed for those computed in step 1. For example, an *N* measurement at $D_R = D_G = D_B = 511$ might yield computed DAC values of (509, 508, 512); the predicted luminances and chromaticity coordinates for (509, 0, 0), (0, 508, 0), and (0, 0, 512) would then be computed and the associated DAC values changed to (511, 0, 0), (0, 511, 0), and (0, 0, 511).

Step 3 is essential because, if it is omitted, the resulting characterization dataset is purely a transformation of the original, conventional characterization and any differences in the resulting predictions are due only to changing the interpolation endpoints (plus possible differences in rounding and truncation). The correct procedure, however, yields a dataset in which channel chromaticity varies with DAC value, in accordance with the measurements from the conventional characterization, but which allows PLVC to predict the measurements from the neutral-point characterization perfectly. In our implementation of this decomposition method, we used conventional characterization datasets that contained the same number of measurements as the neutral-point characterization dataset that was undergoing decomposition.

As in our earlier article, our implementations of PLCC and PLVC used the simplifying assumption that each channel's luminance is zero at zero DAC value. Our experience has been that, if the monitor's black level is adjusted to yield a perception of black for a dark-adapted viewer, this assumption yields reasonably accurate results and avoids the problems inherent in trying to measure extremely low (and possibly zero) light levels. For PLGCC and PLGVC, however, the assumption cannot be used, because zero DAC value and luminance do not exist in the log–log domain; that is, log(0) is undefined. Our solution to this problem was to use the slope defined by the two smallest measurements for a given channel to extrapolate for values below the measured range, when necessary.

* We omitted this triplet because the complete data collection effort included full replications using a MAG model MX17F monitor, as well as a Photo Research model PR-703A/PC spectroradiometer, which could not measure the luminances at (0, 0, 0). The results from these replications matched those for the Aydin + LMT combination in all important respects, and have, therefore, been omitted for the sake of brevity.

TABLE I. Mean errors from Experiment 1.

		Conventional characterization						Neutral-point characterization					
Method	Res	|%Y|	u'v'	ΔE^*_{uv}	95ptl	ΔE^*_{ab}	95ptl	|%Y|	u'v'	ΔE^*_{uv}	95ptl	ΔE^*_{ab}	95ptl
GOG	2	1.85	0.00271	2.31	6.17	1.72	4.12	1.11	0.00122	1.14	2.15	0.87	1.60
	3	1.22	0.00151	1.20	2.66	0.98	2.10	0.69	0.00089	0.82	1.49	0.66	1.29
	4	1.20	0.00147	1.19	2.68	0.96	2.08	0.66	0.00086	0.80	1.44	0.63	1.23
	5	1.23	0.00146	1.18	2.55	0.95	1.99	0.64	0.00086	0.80	1.46	0.63	1.23
	7	1.21	0.00144	1.17	2.51	0.94	1.96	0.64	0.00084	0.78	1.39	0.62	1.19
	9	1.20	0.00143	1.16	2.52	0.94	1.98	0.64	0.00086	0.79	1.45	0.63	1.22
	13	1.20	0.00142	1.16	2.48	0.93	1.92	0.63	0.00084	0.78	1.41	0.62	1.18
	17	1.18	0.00142	1.16	2.48	0.93	1.95	0.64	0.00084	0.78	1.41	0.62	1.20
	25	1.20	0.00142	1.16	2.44	0.93	1.92	0.63	0.00084	0.78	1.42	0.61	1.17
	48	1.18	0.00141	1.15	2.46	0.92	1.90	0.63	0.00084	0.78	1.41	0.61	1.18
PLCC	7	1.84	0.00179	1.48	2.85	1.19	2.19	0.87	0.00131	1.11	2.01	0.87	1.66
	9	1.55	0.00152	1.27	2.51	1.03	1.95	0.67	0.00102	0.88	1.65	0.69	1.39
	13	1.26	0.00140	1.15	2.40	0.93	1.87	0.62	0.00085	0.77	1.43	0.61	1.23
	17	1.25	0.00141	1.16	2.35	0.93	1.83	0.63	0.00088	0.81	1.43	0.64	1.22
	25	1.17	0.00135	1.11	2.34	0.89	1.80	0.63	0.00080	0.75	1.38	0.59	1.13
	49	1.15	0.00133	1.10	2.34	0.88	1.77	0.65	0.00076	0.73	1.34	0.58	1.09
PLGCC	2	4.54	0.00401	3.46	6.54	2.72	5.28	6.70	0.00368	3.68	6.34	2.91	4.90
	3	1.01	0.00159	1.21	2.67	0.98	2.25	1.90	0.00118	1.01	2.00	0.83	1.64
	4	0.78	0.00121	1.00	2.46	0.79	1.90	1.05	0.00052	0.63	1.14	0.51	0.92
	5	1.01	0.00125	1.04	2.26	0.83	1.77	0.72	0.00069	0.70	1.27	0.55	1.03
	7	1.05	0.00126	1.06	2.24	0.84	1.70	0.70	0.00065	0.66	1.23	0.52	0.99
	9	1.10	0.00130	1.07	2.28	0.86	1.81	0.64	0.00075	0.72	1.32	0.56	1.07
	13	1.11	0.00131	1.09	2.29	0.87	1.75	0.64	0.00073	0.70	1.32	0.56	1.05
	17	1.15	0.00133	1.11	2.26	0.88	1.77	0.64	0.00078	0.74	1.32	0.59	1.10
	25	1.13	0.00133	1.10	2.31	0.88	1.76	0.64	0.00077	0.73	1.34	0.58	1.10
	49	1.13	0.00132	1.09	2.32	0.88	1.74	0.66	0.00075	0.72	1.33	0.57	1.08
PLVC	7	1.84	0.00125	1.06	1.97	0.85	1.46	0.80	0.00104	0.92	1.76	0.73	1.43
	9	1.55	0.00097	0.87	1.66	0.71	1.19	0.59	0.00075	0.71	1.43	0.56	1.12
	13	1.26	0.00089	0.80	1.52	0.64	1.11	0.56	0.00059	0.64	1.21	0.52	1.00
	17	1.25	0.00087	0.79	1.50	0.63	1.10	0.58	0.00062	0.67	1.21	0.54	0.98
	25	1.17	0.00084	0.77	1.48	0.62	1.07	0.58	0.00056	0.64	1.20	0.52	0.99
	49	1.15	0.00082	0.75	1.47	0.61	1.10	0.61	0.00052	0.63	1.19	0.51	0.96
PLGVC	2	4.54	0.00380	3.68	6.30	3.01	5.19	6.54	0.00445	4.31	7.43	3.56	6.26
	3	1.01	0.00129	1.15	2.13	0.95	2.00	1.85	0.00143	1.26	2.80	1.09	2.63
	4	0.78	0.00065	0.68	1.41	0.55	1.14	1.00	0.00049	0.66	1.26	0.57	1.20
	5	1.01	0.00078	0.74	1.42	0.59	1.05	0.68	0.00048	0.63	1.14	0.52	1.02
	7	1.05	0.00075	0.73	1.42	0.59	1.08	0.65	0.00046	0.60	1.15	0.49	0.98
	9	1.10	0.00078	0.73	1.41	0.59	1.06	0.59	0.00052	0.63	1.21	0.51	1.00
	13	1.11	0.00080	0.75	1.44	0.60	1.07	0.61	0.00050	0.61	1.16	0.50	0.97
	17	1.15	0.00081	0.75	1.46	0.61	1.09	0.60	0.00054	0.64	1.18	0.52	0.99
	25	1.13	0.00082	0.76	1.47	0.61	1.08	0.59	0.00054	0.64	1.21	0.52	1.00
	49	1.13	0.00081	0.75	1.48	0.60	1.10	0.61	0.00054	0.65	1.18	0.52	0.99

EXPERIMENT 1: RESOLUTION AND CHARACTERIZATION METHOD

Our first experiment was a partial replication of Berns *et al.*,[3] extended to include PLVC, PLGCC, and PLGVC. Each method's predictions of the 124-color test dataset were tested using two repeated-measures full-factorial analyses of variance (ANOVAs) in which the error term for each effect was the interaction of that effect with replications. One ANOVA used absolute percent luminance error as the dependent measure, and the other used u'v' error. In both cases, the dependent measures were averaged across the 124 test colors. The independent variables in each ANOVA were characterization method (conventional vs. neutral-point) and the number of measurements per color channel in the characterization dataset (which we term *resolution*). The significance of all statistical tests was assessed using an alpha of 0.05.

The results are summarized in Table I and Fig. 1. Mean values of ΔE^*_{uv} and ΔE^*_{ab}, calculated relative to W_{max} for each replication, have been included in Table I to facilitate comparisons. The 95th percentiles for ΔE^*_{uv} and ΔE^*_{ab} are shown also, to provide information about the error ranges.

GOG

For GOG, we planned originally to test resolutions 2–49. For resolution 2, though, we had problems solving for the coefficients used in the GOG power functions. Eventually, we realized that this was because of an inherent problem—explained in Appendix C—that arises when the characterization dataset includes a measurement at D_{max}. Here, we will say simply that D_{max} measurements contribute little or nothing to solutions for the GOG coefficients. We circumvented this problem by substituting measurement 48 for measurement 49 in our GOG tests. As a result, the highest resolution we could test for GOG was 48.

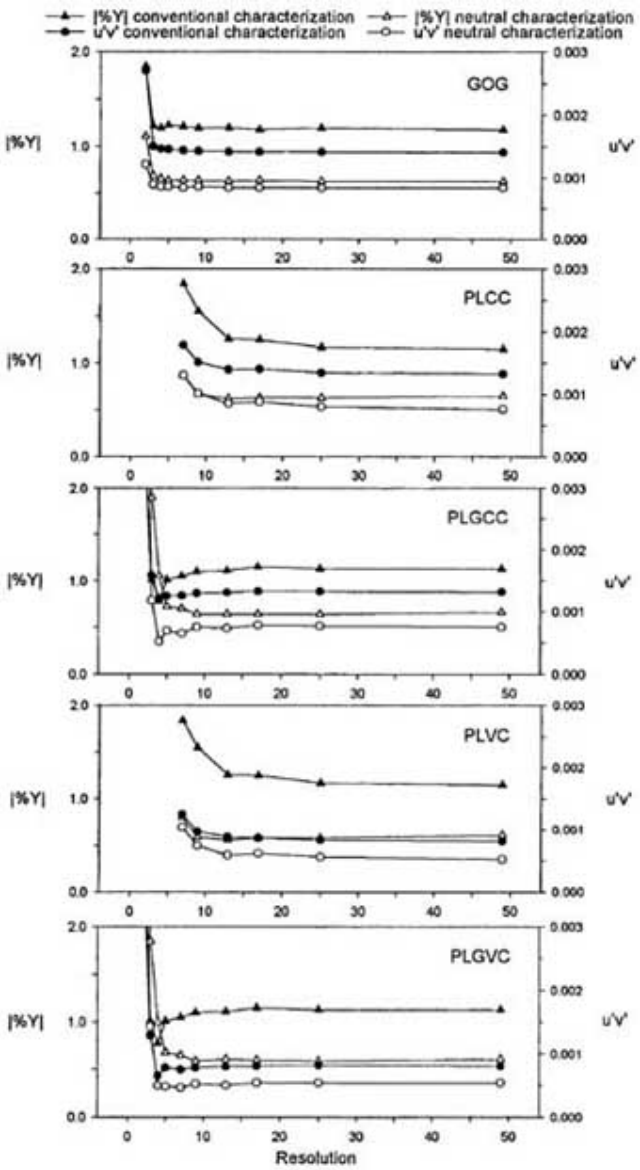

FIG. 1. Mean |%Y| and u'v' error from Experiment 1 for each predictive method as a function of resolution and characterization method.

The ANOVA on luminance error showed that the main effects of characterization and resolution are significant and account for 53.3 and 17.7% of the variance, respectively. Neutral-point characterization produced less error than conventional characterization, and increasing the resolution from 2 to 3 produced a much larger error reduction than any other resolution change.

The ANOVA on u'v' error showed that the main effects of characterization and resolution, plus their interaction, are significant and account for 43.9, 23.1, and 7.0% of the variance, respectively. Again, neutral-point characterization produced less error than conventional characterization, and increasing the resolution from 2 to 3 produced the largest error reduction. The interaction appears to reflect the fact that the initial drop is much larger for conventional characterization.

The effects of resolution are fairly consistent but quite modest beyond resolution 3. Therefore, we conclude that there is little to be gained by using resolutions beyond 3 with GOG. Also, neutral-point characterization yields less error than conventional characterization.

PLCC

For PLCC, we tested resolutions 7–49, because our previous work[1] indicated that lower resolutions would not yield competitive accuracy. The ANOVA on luminance error showed that the main effects of characterization and resolution, plus their interaction, are significant and account for 70.2, 16.0, and 4.6% of the variance, respectively. Neutral-point characterization produced less error than conventional characterization. Increasing resolution caused a larger and completely monotonic error reduction for conventional characterization, although the improvements beyond resolution 13 are trivial. For neutral-point characterization, error asymptotes somewhere around resolution 13.

The ANOVA on u'v' error showed that the main effects of characterization and resolution are significant and account for 65.9 and 27.7% of the variance, respectively. Again, neutral-point characterization produced less error than conventional characterization. Increasing resolution produced nearly monotonic error reductions, although the changes beyond resolution 13 are modest.

The effects of resolution are fairly consistent but very small beyond resolution 13. Therefore, we conclude that there is little to be gained from using resolutions beyond 13 with PLCC. Also, neutral-point characterization is clearly preferable.

PLGCC

For PLGCC, we tested resolutions 2–49. The ANOVA on luminance error showed that the main effect of resolution and the characterization *x* resolution interaction are significant and account for 91.3 and 7.9% of the variance, respectively. Increasing resolution caused a larger and more consistent decrease in error for neutral-point characterization. As a result, conventional characterization produced lower error at resolutions below 5, but neutral-point characterization is better at the higher resolutions. Error for conventional characterization reaches a minimum at resolution 4 and then increases, stabilizing somewhere around resolution 9; error for neutral-point characterization also stabilizes around resolution 9.

The ANOVA on u'v' error showed that the main effects of characterization and resolution, plus their interaction, are significant and account for 9.0, 89.9, and 0.3% of the variance, respectively. Neutral-point characterization yielded less error than conventional characterization in all cases. Error reaches a minimum at resolution 4 and then increases slightly, stabilizing somewhere around resolution 9. At the lower resolutions, there are minor differences in the effect of resolution for conventional vs. neutral-point character-

ization; these differences appear to be the source of the small interaction.

Both types of error seem to stabilize around resolution 9, so we conclude that there is no assured advantage to using higher resolutions with PLGCC. Examination of the CIELUV and CIELAB values and their 95th percentiles in Table I shows that a case can be made for using resolutions 4 or 5. This case is not as uncertain as it might seem, because we have obtained comparable results from tests involving a different monitor and colorimetric instrument. We conclude, therefore, that resolution 4 or 5 is best, but recommend 9 for those who prefer a conservative alternative. Based on the CIELUV and CIELAB values, we also conclude that neutral-point characterization is superior to conventional characterization at resolutions of 3 or more.

PLVC

For PLVC, we tested resolutions 7–49. The ANOVA on luminance error showed that the main effects of characterization and resolution, plus their interaction, are significant and account for 73.3, 13.8, and 4.6% of the variance, respectively. Examination of these effects leads to the same conclusions we reported above for PLCC. This outcome is not surprising, because our earlier article[1] showed that PLCC and PLVC produce nearly identical luminance errors.

The ANOVA on u′v′ error showed that the main effects of characterization and resolution, plus their interaction, are significant and account for 35.6, 54.6, and 0.7% of the variance, respectively. Neutral-point characterization produced less error than conventional characterization in all cases. Error decreases monotonically with resolution for conventional characterization, but seems to stabilize around resolution 13 for neutral-point characterization.

The effects of resolution are consistent but rather small beyond resolution 13 for conventional characterization; they stabilize around resolution 13 for neutral-point characterization. Therefore, we conclude that there is little to be gained from using resolutions beyond 13 with PLVC. Also, neutral-point characterization is preferable.

PLGVC

For PLGVC, we tested resolutions 2–49. The ANOVA on luminance error showed that the main effects of resolution and the characterization x resolution interaction are significant and account for 91.6 and 7.5% of the variance, respectively. Examination of the interaction leads to the same conclusions we reported above for PLGCC; thus, it seems the logarithmic versions of PLCC and PLVC also produce very similar luminance error.

The ANOVA on u′v′ error showed that the main effects of characterization and resolution, plus their interaction, are significant and account for 9.0, 89.9, and 0.3% of the variance, respectively. Neutral-point characterization yields lower error than conventional characterization at all resolutions except 2 and 3. For conventional characterization, error reaches a minimum at resolution 4 and then increases, stabilizing somewhere around resolution 9, although the changes beyond resolution 5 are trivial. For neutral-point characterization, which shows the same pattern, the corresponding numbers are 7, 9, and 4.

Our conclusions about PLGVC are the same as for PLGCC. Both types of error seem to stabilize around resolution 9, so there is no advantage to using higher resolutions. The CIELUV and CIELAB values and their 95th percentiles in Table I suggest that resolutions 4 or 5 are the best choices; our tests with a different monitor and colorimetric instrument indicate that this idea is generalizable. We recommend 9, however, for those who prefer a conservative alternative. Based on the CIELUV and CIELAB values, we also conclude that neutral-point characterization is equal or superior to conventional characterization at resolutions of 4 or more.

Summary and Discussion

For GOG, resolutions beyond 3 did not produce substantial reductions in error. This finding agrees with Berns *et al.*[3] (see their Table II), who found no differences among resolutions 5, 16, and 180. For PLCC and PLVC, there were no worthwhile error reductions beyond resolution 13. For PLGCC and PLGVC, resolutions 4 and 5 produced the best results, although the inconsistencies at the lower resolutions might justifiably cause some users to prefer resolution 9.

For all resolutions that produced acceptably small errors, neutral-point characterization produced less error than conventional characterization. This finding contradicts the results for GOG reported by Berns *et al.*[3] (see their Table II), which showed no effect of characterization method. Another noteworthy point is that the magnitudes of the errors we obtained for GOG and PLCC are larger than those reported by Berns *et al.* (see their Table III). We discuss these differences after reporting the results for Experiment 3.

EXPERIMENT 2: BLACK CORRECTION

Our second experiment examined the effects of using black correction. The black light, as measured by the LMT, was 0.06 cd/m^2, u′ = 0.0280, and v′ = 0.4529.

The procedure for black correction is fairly straightforward and identical to correcting for ambient illumination: (1) Zero the measuring target's DAC values while leaving the background color unchanged; (2) Measure the target's tristimulus values; (3) Subtract these tristimulus values from each measurement in the characterization dataset; and (4) When predicting the tristimulus values produced by a given DAC-value triplet, do the calculation in the usual way and then add the tristimulus values obtained for the target at zero DAC value. Step 3 must be performed for both the conventional and neutral-point characterization datasets, when using PLVC or PLGVC with neutral-point characterization. For predictive methods that assume constant channel chromaticity, it is important to ensure that the assumed channel chromaticities are corrected, as well as the characterization dataset.

TABLE II. Mean errors from Experiment 2.

			Conventional characterization						Neutral-point characterization					
Method	Res	Black	|%Y|	u′v′	ΔE^*_{uv}	95ptl	ΔE^*_{ab}	95ptl	|%Y|	u′v′	ΔE^*_{uv}	95ptl	ΔE^*_{ab}	95ptl
GOG	4	N	1.20	0.00147	1.19	2.68	0.96	2.08	0.66	0.00086	0.80	1.44	0.63	1.23
		Y	0.90	0.00084	0.76	1.33	0.60	0.94	0.52	0.00070	0.69	1.20	0.52	0.96
	7	N	1.21	0.00144	1.17	2.51	0.94	1.96	0.64	0.00084	0.78	1.39	0.62	1.19
		Y	0.90	0.00084	0.77	1.36	0.61	0.96	0.51	0.00068	0.67	1.20	0.51	0.95
	13	N	1.20	0.00142	1.16	2.48	0.93	1.92	0.63	0.00084	0.78	1.41	0.62	1.18
		Y	0.90	0.00082	0.76	1.32	0.60	0.95	0.50	0.00068	0.67	1.21	0.51	0.94
PLCC	7	N	1.84	0.00179	1.48	2.85	1.19	2.19	0.87	0.00131	1.11	2.01	0.87	1.66
		Y	1.54	0.00117	1.02	1.91	0.82	1.34	0.78	0.00116	1.01	1.83	0.76	1.39
	13	N	1.26	0.00140	1.15	2.40	0.93	1.87	0.62	0.00085	0.77	1.43	0.61	1.23
		Y	0.96	0.00083	0.77	1.39	0.61	0.99	0.50	0.00070	0.67	1.22	0.51	0.97
PLGCC	4	N	0.78	0.00121	1.00	2.46	0.79	1.90	1.05	0.00052	0.63	1.14	0.51	0.92
		Y	0.66	0.00065	0.65	1.26	0.50	1.01	0.97	0.00046	0.55	0.99	0.45	0.85
	7	N	1.05	0.00126	1.06	2.24	0.84	1.70	0.70	0.00065	0.66	1.23	0.52	0.99
		Y	0.77	0.00071	0.71	1.30	0.56	0.98	0.59	0.00051	0.57	1.05	0.43	0.85
	13	N	1.11	0.00131	1.09	2.29	0.87	1.75	0.64	0.00073	0.70	1.32	0.56	1.05
		Y	0.82	0.00076	0.73	1.31	0.57	0.97	0.53	0.00058	0.61	1.13	0.46	0.90
PLVC	7	N	1.84	0.00125	1.06	1.97	0.85	1.46	0.80	0.00104	0.92	1.76	0.73	1.43
		Y	1.54	0.00095	0.85	1.47	0.70	1.09	0.80	0.00105	0.95	1.84	0.72	1.43
	13	N	1.26	0.00089	0.80	1.52	0.64	1.11	0.56	0.00059	0.64	1.21	0.52	1.00
		Y	0.96	0.00056	0.56	1.10	0.47	0.86	0.59	0.00060	0.64	1.23	0.48	0.96
PLGVC	4	N	0.78	0.00065	0.68	1.41	0.55	1.14	1.00	0.00049	0.66	1.26	0.57	1.20
		Y	0.66	0.00057	0.58	1.16	0.48	1.01	1.04	0.00042	0.57	1.12	0.45	0.86
	7	N	1.05	0.00075	0.73	1.42	0.59	1.08	0.65	0.00046	0.60	1.15	0.49	0.98
		Y	0.77	0.00046	0.53	1.06	0.44	0.84	0.66	0.00046	0.55	1.07	0.41	0.78
	13	N	1.11	0.00080	0.75	1.44	0.60	1.07	0.61	0.00050	0.61	1.16	0.50	0.97
		Y	0.82	0.00048	0.52	1.07	0.43	0.82	0.61	0.00050	0.59	1.12	0.44	0.84

To ease the computational burden, we restricted attention to resolutions 4, 7, and 13 for GOG, PLGCC, and PLGVC, and resolutions 7 and 13 for PLCC and PLVC.[†] We performed ANOVAs using characterization method, resolution, and black correction (present vs. absent) as the independent variables, and again used an alpha of 0.05 for all tests. The mean luminance and u′v′ errors are summarized in Table II (wherein the values for non-black-corrected cases duplicate the corresponding values in Table I) and illustrated in Fig. 2.

The ANOVA results for effects that were tested in Experiment 1 match expectations in all important respects. One new thing we learned is that the main effect of black light is significant in all cases (error was lower when black correction was used). The most interesting outcome, though, was that the characterization *x* black-light interaction is significant in all cases, accounting for 1.3% (PLCC) to 20.1%(PLGCC) of the variance in luminance error and 10.4% (PLCC) to 15.2% (PLGVC) of the variance in u′v′ error. The interaction is visible in Fig. 2: Black correction produces a much smaller error reduction (or sometimes none) when neutral-point characterization is used. Similarly, neutral-point characterization is less effective when black correction is used. These findings imply that there is overlap in the error accounted for by neutral-point characterization and black correction—an implication we explored in the next experiment.

[†] Given the Experiment 1 results for PLGCC and PLGVC, resolution 9 might seem a better choice than resolution 7; however, Experiment 1 showed that resolution 13 should give the same results as 9 for PLGCC and PLGVC, so we thought 7 might be more informative.

EXPERIMENT 3: NUMBER OF CHANNELS

Our 124 test colors consisted of 12 single-channel colors (i.e., colors produced with only one nonzero DAC value), 48 two-channel colors, and 64 three-channel colors; thus, it was weighted heavily with two- and especially three-channel colors. Furthermore, the results from Experiment 2 suggest that the (seemingly small amount of) black light had a substantial effect on prediction accuracy, and the results reported in Appendix B show that our monitor was somewhat sub-additive. Three mechanisms could cause these facts to produce the characterization *x* black-light interaction we observed for luminance error in Experiment 2, as well as the apparent superiority of neutral-point characterization over conventional characterization in Experiment 1:

1. Each measurement for a conventional characterization dataset includes any black light that may be present. Therefore, if conventional characterization is used without black correction, luminance predictions for single-channel colors should be accurate nonetheless, whereas predictions for two- and three-channel colors should over-represent the black light by two or three times and, therefore, be less accurate. The situation for neutral-point characterization is the opposite: If black

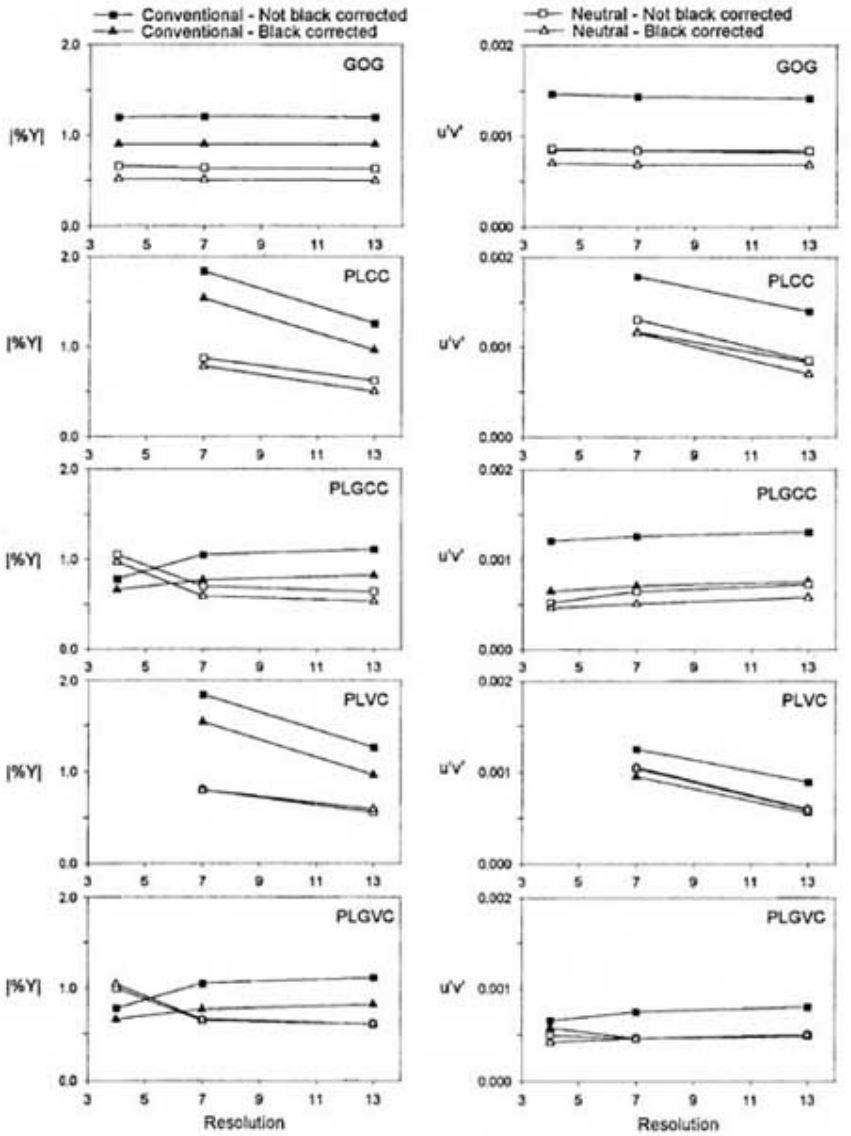

FIG. 2. Mean |%Y| and u'v' error from Experiment 2 for each predictive method as a function of resolution, characterization method, and black correction.

correction is not performed, decomposing the N measurements distributes the black light across the *RGB* channels in the characterization dataset, so luminance predictions involving three channels should be more accurate than predictions involving one or two, which should underestimate the black light.

2. A similar principle should apply if the monitor's additivity is imperfect, even if black correction is used. In this case, luminance predictions from conventional characterization should be most accurate for single-channel colors, for which additivity is irrelevant, and should become less accurate as the number of channels (and their DAC value) increases and, hence, additivity problems manifest themselves. Neutral-point characterization, on the other hand, is derived from three-channel measurements and should, therefore, be most accurate for three-channel colors.
3. The luminance added by black light is a constant; hence, regardless of characterization method, black-light-related luminance error assessed as a percentage should diminish as luminance increases. In our dataset, luminance tends to increase with the number of channels in use, and thus—for cases where black correction was not used—percent luminance error should decrease as the number of channels increases.

To summarize: For our dataset, mechanisms 1 and 2 should cause luminance error to increase as the number of channels increases for conventional characterization, and decrease as the number of channels increases for neutral-point characterization. Mechanism 1 should cause black correction to reduce error for two- and three-channel, but not single-channel, predictions for conventional characterization and be least effective for three-channel predictions for neutral-point characterization. Mechanism 3 should

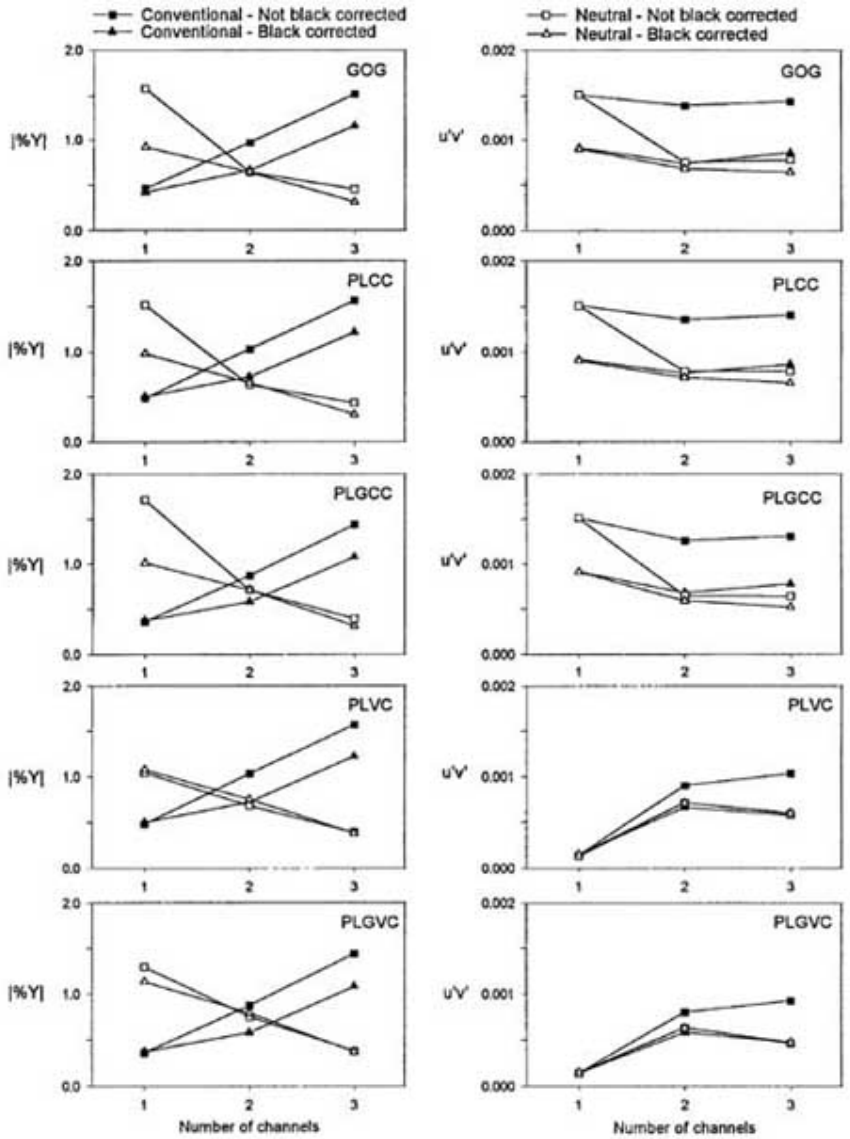

FIG. 3. Mean |%Y| and u′v′ error from Experiment 3 for each predictive method as a function of characterization method, black correction, and number of channels.

cause black correction to also reduce differences caused by the number of active channels.

As for u′v′ error, many complex mechanisms occurred to us that might contribute to the interaction. Mainly, though, we expected the same mechanisms that affect luminance error to also affect u′v′ error, because accurate chromaticity predictions often require accurate predictions of the constituent channel luminances. Overall, then, it seemed that the test dataset was weighted in favor of neutral-point characterization, especially in the cases where black correction was not performed.

The results in Table II seem to fit our reasoning fairly well, so we tested it by breaking down the resolution 13 data in Table II, according to the number of channels involved. We performed ANOVAs using characterization method, black correction, and number of channels as the independent variables, and again used an alpha of 0.05 for all tests. The results are summarized in Table III and illustrated in Fig. 3.

The ANOVA results are very similar for all predictive methods. For luminance error, the main effect of black light and the black-light x channels, characterization x channels, and characterization x black-light x channels interactions are significant in all cases; the characterization x black-light interaction is also significant for PLVC and PLGVC. The characterization x channels interaction is consistently the most important effect, accounting for 72.1% (GOG) to 83.3% (PLGVC) of the variance. For u′v′ error, all main effects and interactions are significant in all cases. For GOG, PLCC, and PLGCC, the main effects of characterization, black light, and channels are the most important, accounting for an average of 14.5, 39.5, and 21.5% of the variance, respectively. For PLVC and PLGVC, however, the main effect of channels is by far the most important, accounting for 80.1 and 73.4%, respectively.

Examination of Table III and Fig. 3 shows that luminance error behaves almost exactly as we predicted. The one

TABLE III. Mean errors from Experiment 3.

			Conventional characterization						Neutral-point characterization					
Method	Black	Chn	|%Y|	u'v'	ΔE^*_{uv}	95ptl	ΔE^*_{ab}	95ptl	|%Y|	u'v'	ΔE^*_{uv}	95ptl	ΔE^*_{ab}	95ptl
GOG	N	1	0.46	0.00151	0.66	1.76	0.91	2.86	1.58	0.00151	0.80	1.46	0.80	2.26
		2	0.97	0.00139	1.09	2.67	0.80	1.99	0.64	0.00075	0.76	1.38	0.59	1.13
		3	1.52	0.00144	1.31	2.41	1.04	1.75	0.45	0.00078	0.78	1.41	0.60	1.13
	Y	1	0.42	0.00091	0.50	1.26	0.53	1.29	0.92	0.00091	0.66	1.11	0.66	1.18
		2	0.66	0.00074	0.67	1.35	0.50	0.91	0.64	0.00068	0.67	1.23	0.50	0.93
		3	1.16	0.00086	0.87	1.32	0.69	0.95	0.31	0.00064	0.66	1.20	0.48	0.87
PLCC	N	1	0.48	0.00151	0.69	1.97	0.93	2.86	1.52	0.00151	0.80	1.63	0.81	2.33
		2	1.03	0.00136	1.08	2.56	0.80	1.97	0.63	0.00078	0.76	1.38	0.60	1.11
		3	1.57	0.00141	1.29	2.34	1.02	1.74	0.43	0.00078	0.77	1.44	0.59	1.21
	Y	1	0.50	0.00091	0.51	1.30	0.54	1.34	0.98	0.00090	0.68	1.16	0.66	1.23
		2	0.72	0.00076	0.69	1.38	0.52	0.95	0.65	0.00071	0.68	1.22	0.51	0.96
		3	1.22	0.00086	0.88	1.40	0.69	0.99	0.30	0.00065	0.66	1.23	0.48	0.90
PLGCC	N	1	0.36	0.00151	0.60	1.69	0.87	2.61	1.71	0.00151	0.80	1.42	0.79	2.23
		2	0.87	0.00126	1.01	2.47	0.73	1.81	0.71	0.00065	0.71	1.30	0.55	1.02
		3	1.44	0.00131	1.23	2.22	0.96	1.62	0.39	0.00064	0.68	1.26	0.51	1.01
	Y	1	0.37	0.00091	0.49	1.11	0.52	1.23	1.01	0.00091	0.69	1.13	0.66	1.16
		2	0.58	0.00068	0.64	1.30	0.47	0.89	0.71	0.00059	0.63	1.15	0.48	0.87
		3	1.08	0.00078	0.85	1.32	0.66	0.97	0.31	0.00052	0.58	1.10	0.41	0.78
PLVC	N	1	0.48	0.00014	0.28	0.59	0.23	0.43	1.04	0.00013	0.59	1.26	0.43	1.02
		2	1.03	0.00090	0.64	1.37	0.51	0.94	0.68	0.00071	0.71	1.23	0.63	1.04
		3	1.57	0.00103	1.01	1.57	0.82	1.15	0.39	0.00059	0.59	1.18	0.46	0.96
	Y	1	0.50	0.00015	0.30	0.59	0.25	0.49	1.07	0.00015	0.63	1.12	0.43	0.86
		2	0.72	0.00066	0.55	1.12	0.44	0.81	0.75	0.00071	0.69	1.32	0.52	0.99
		3	1.22	0.00057	0.62	1.12	0.53	0.88	0.38	0.00060	0.60	1.19	0.46	0.95
PLGVC	N	1	0.36	0.00014	0.23	0.55	0.19	0.40	1.29	0.00014	0.70	1.30	0.50	1.01
		2	0.87	0.00080	0.59	1.30	0.48	0.91	0.75	0.00063	0.72	1.22	0.65	1.07
		3	1.44	0.00092	0.96	1.52	0.77	1.13	0.38	0.00046	0.52	1.04	0.39	0.83
	Y	1	0.37	0.00014	0.24	0.55	0.20	0.42	1.13	0.00015	0.66	1.12	0.44	0.88
		2	0.58	0.00058	0.50	1.08	0.41	0.75	0.79	0.00063	0.65	1.18	0.50	0.92
		3	1.08	0.00047	0.59	1.09	0.49	0.87	0.37	0.00047	0.53	1.00	0.40	0.77

exception is the fact that, for GOG, PLCC, and PLGCC using neutral-point characterization, black correction produces a greater improvement for three-channel colors than for two-channel colors. This exception suggests that there is a fourth, as-yet unidentified mechanism influencing the results.

The results for u'v' error indicate that, contrary to our expectations, error in luminance prediction does not necessarily produce concomitant error in chromaticity prediction. For GOG, PLCC, and PLGCC, error for single-channel colors is noticeably greater than for two- and three-channel colors when neutral-point characterization is used without black correction. Otherwise, the effect of the number of channels is modest. We do not understand the latter outcome, but offer the following explanation for the former: When black light is present but uncorrected, decomposing the neutral-point measurements places erroneous luminance values in the characterization dataset, due to the use of channel chromaticities in the decomposition calculations that are correct only at D_{max} (assuming they were measured at D_{max}). Therefore, for single-channel colors, the predicted chromaticities are usually wrong, because the assumed channel chromaticities are wrong, but for two- and three-channel colors the luminance and chromaticity errors offset each other, yielding more accurate chromaticity predictions.

Another interesting point that is evident in Fig. 3 is that PLVC and PLGVC yield essentially the same single-channel u'v' error, regardless of characterization method or black correction, and this error is substantially lower than the other predictive methods in all cases. These outcomes imply that the changes in our measured channel chromaticities were caused by something more than black light, and the variable channel-chromaticity feature of PLVC and PLGVC does a good job of accounting for it, as well as black light, at least for single-channel colors. Although error is greater for two- and three-channel colors, it is often less than for the other predictive methods when black correction is used and, in this case, the complications of using neutral-point characterization in conjunction with PLVC and PLGVC do not appear to be worth the effort.

We conclude that the superiority of neutral-point characterization over conventional characterization that we observed in Experiment 1 was an artifact, caused by the presence of uncorrected black light, non-additivity, and a test dataset that emphasized multichannel colors. Although Berns *et al.*[3] used an equivalent dataset, they stated that their measuring geometry excluded black light, and evidently their additivity was better than ours (although we must note that their Table IV suggests that it was similar). If additivity is good and black light is either absent or corrected for, neutral-point and conventional characterization should yield comparable results, regardless of the number of channels in use.

If the black-corrected GOG and PLCC results for neutral-

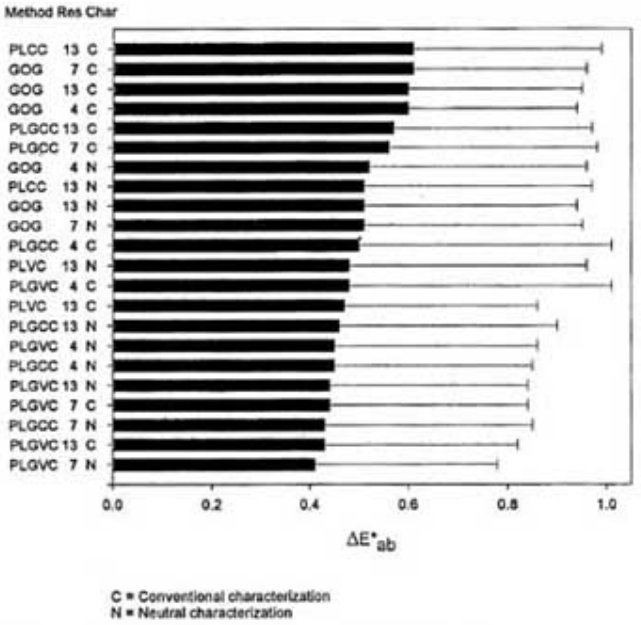

FIG. 4. Mean and 95th-percentile ΔE^*_{ab} for black-corrected cases from Experiment 2, sorted in descending order. PLCC and PLVC at resolution 7 are omitted.

point characterization in Table II are compared with those reported by Berns *et al.*[3] in their Tables II and III, the cross-laboratory differences are much less than the ones we reported for Experiment 1. Indeed, the differences are probably as small as could be expected for cross-laboratory comparisons.

CONCLUSIONS

To facilitate comparisons of predictive accuracy, Fig. 4 graphs mean ΔE^*_{ab} in descending order for the black-corrected cases from Experiment 2 (minus PLCC and PLVC at resolution 7, because these cases are not recommended for use). The 95th percentiles are shown, too. Based on this figure and consideration of our other results, we offer the following recommendations:

1. Black correction should be performed whenever there is reason to believe the area of interest on the display is luminous at zero DAC value. If a colorimetric instrument having sufficient sensitivity for this purpose is unavailable, the use of neutral-point characterization provides much, but not all, of the benefits of black correction.
2. Neutral-point characterization is to be preferred over conventional characterization, except for PLVC and PLGVC and in special cases where most or all of the colors are produced by single channels. Otherwise, neutral-point characterization yields equal or better predictive accuracy and reduces the number of characterization measurements needed.
3. If the number of characterization measurements must be minimized, GOG at resolution 3 is the best choice for a predictive method. This choice requires a total of only six measurements (three N plus one each of R, G, and B to establish the channels' chromaticity coordinates) or nine measurements, depending on whether neutral-point characterization is used, and yields very good accuracy.
4. If programming and usage complexity must be minimized, PLCC at resolution 13 is the best choice. It requires either 16 or 39 measurements (depending on whether neutral-point characterization is used), is the easiest predictive method to implement in software, does not require human intervention to ensure proper convergence, and yields accuracy comparable with GOG.
5. The best compromise between complexity and measuring time is PLGCC at resolution 4. It is only slightly more difficult to program than PLCC, requires either 7 or 15 measurements (depending on whether neutral-point characterization is used), and is a bit more accurate than GOG or PLCC.
6. If programming complexity is not an issue and accuracy must be maximized, PLGVC using conventional characterization at resolution 9 is the best choice. It requires 27 measurements, should yield the same accuracy as resolutions 13 or higher, and should also yield accuracy that is better than or comparable with PLGVC with neutral-point characterization, regardless of resolution, while avoiding the complications and added measurements required to use neutral-point characterization in conjunction with PLGVC. However, the accuracy advantage over the other methods is realized only for colors at or near the edges of the display's color gamut.

Our results indicate that, for most purposes, there is little reason to prefer one method over another on the basis of predictive accuracy, given that each is used in accordance with our recommendations. Of particular interest is the fact that PLVC's and PLGVC's allowance for variable channel chromaticity yields substantial advantages only for single-channel colors. This finding contradicts our earlier comparisons[1] against PLCC and may reflect differences in the monitors; therefore, we are uncertain how generalizable this finding is. We also note that PLVC and PLGVC may be especially valuable for predicting the colorimetry of liquid-crystal displays, which as Silverstein and Fiske[11] have shown, can exhibit substantial covariance between luminance and channel chromaticity.

It is worthwhile to consider the adequacy of the predictive accuracies we have reported. Stokes, Fairchild, and Berns[12] found that, for sequential comparisons of images on the same display, the ΔE^*_{ab} that yielded a 50% discrimination threshold ranged from 1.43–2.65 and averaged 2.15. All the methods we studied can provide 95th-percentile accuracies that are within these limits. For simultaneous comparisons of images on adjacent displays, on the other hand—where the predictive errors can sum and thresholds are presumably smaller—adequacy is less clear. We can

estimate limiting thresholds for this case by referring to MacAdam's[13] data: Brainard[14] has determined recently that Nutting's 50% discrimination thresholds, expressed as ΔE^*_{ab}, ranged from 0.30–3.37 and averaged 1.23. Therefore, none of the methods we studied may be fully adequate for matching images across adjacent displays. In such cases, the measure-and-adjust algorithm we introduced previously[1] may be useful.

Our results suggest that the last significant error source remaining to be eliminated via modeling is non-additivity. Although progress toward solving this problem has been reported,[8,15–17] a major difficulty is devising a way to incorporate channel interactions in the model while preserving reversibility (so predictions can be made of the DAC values needed to produce a given color, as well as the color produced by given DAC values) and without making the calculations forbiddingly complex. Our dataset is available by request to anyone who wishes to pursue this problem or test other ways of improving on the accuracies we have reported.

APPENDIX A: PLVC AND PLGVC

Our earlier article[1] explained PLVC by describing how it is used to calculate the DAC values that are needed to produce a desired luminance and pair of chromaticity coordinates. Communications with other workers have since indicated this explanation was not sufficiently clear. We attempt to rectify this problem here by describing how PLVC is used to calculate the luminance and chromaticity coordinates that are produced by a given set of DAC values. This case is much easier to explain and understand. Copies of our PLVC and PLGVC subroutines, coded in FORTRAN, are also available by request.

Assume that *XYZ*-tristimulus values have been measured for *R*, *G*, and *B* at several DAC values each, with only one nonzero DAC value at a time. This yields a table of measured tristimulus values for each channel. To predict the tristimulus values that will be produced for $D_R = i$, $D_G = 0$, $D_B = 0$, where $i > 0$, enter the table for *R* and interpolate between the two measurements that lie immediately above and immediately below *i* to get the predicted X_R. Repeat this procedure to get Y_R and Z_R. These predicted tristimulus values for *R* are the final solution, because D_G and D_B are zero, which we assume to mean that the tristimulus values for the *G* and *B* channels are also zero. This solution technique differs from PLCC in that three independent piecewise linear interpolations are used to obtain the three tristimulus values. Thus, variations in the measured *R* chromaticity coordinates as the *R* DAC value changes are accounted for in the resulting predictions. With PLCC, on the other hand, interpolation is used only to obtain the predicted *Y*-tristimulus value. The predicted *X* and *Z* are obtained from the predicted *Y* plus a single pair of measured chromaticity coordinates for the *R* channel, which are assumed to be constant.

A slightly more complicated example involves $D_R = i$, $D_G = j$, $D_B = k$, where $\{i, j, k\} > 0$. In this case, the procedure described above must be performed for each channel, resulting in a total of nine piecewise linear interpolations. Each interpolation yields a predicted tristimulus value. The final solution is obtained by summing these tristimulus values across each channel; that is, computing $X = X_R + X_G + X_B$, etc. PLGVC works the same way as PLVC, but interpolation is performed for $\log(X_R)$ as a function of $\log(D_R)$, etc. Thus, PLVC and PLGVC perform trilinear interpolation based solely on measurements of single channels, and are similar with (but simpler than) trilinear interpolation methods that include multichannel measurements and have become popular for predicting the output of color printers.

In our implementation of PLVC (and PLCC, for that matter), predictions for DAC values that lie below the smallest measured values in the tables are obtained by assuming that the tristimulus values become zero at zero DAC value. That is, we interpolate between zero and the smallest measured values. We follow this procedure because it is difficult ordinarily to obtain reliable colorimetric measurements at $D_R = D_G = D_B = 0$, especially because we do our best to assure that the monitor does not emit at this setting.

A PLVC algorithm to predict tristimulus values, given DAC values, is only slightly more difficult to implement in software than PLCC. The opposite operation—that is, predicting the DAC values that yield a given set of tristimulus values—is more troublesome. One method is iterative search: Compute *XYZ* for every possible combination of D_R, D_G, and D_B and select the D_R, D_G, D_B triplet that comes closest to yielding the desired *XYZ* (using whatever measure of color error one prefers). This method is inefficient and, depending on the application and the number of bits in one's graphics DACs, can be unacceptably slow. Appendix A in our previous article describes a far more efficient algorithm that we have used routinely for many years.

Our PLVC algorithm does not yield valid solutions for PLGVC, however. That is, one cannot simply log-transform the DAC values and measured *XYZ*, use our PLVC algorithm to find $\log(D_R, D_G, D_B)$, and then transform back. The problem is that a closed-form solution to the equation our PLVC algorithm needs in order to perform its iterations does not exist in the log–log domain.

The best solution to this problem we have devised so far takes advantage of the facts that the correct solution subspace for PLVC must also be the correct subspace for PLGVC, and furthermore the values of D_R, D_G, and D_B that constitute the PLGVC solution must be greater than or equal to the values for the corresponding PLVC solution. Therefore, we compute a solution using PLVC, thereby identifying the correct subspace for PLGVC, and then perform an iterative search of all D_R, D_G, D_B triplets lying between the PLVC solution and the top of the solution subspace. Thus, we obtain solutions using iterative search, but reduce the required number of iterations greatly by restricting the search area. With this approach, the required number of iterations decreases as the resolution of the characterization

dataset increases, because the size of the solution subspaces decreases.

APPENDIX B: CORRECTED NEUTRAL-POINT CHARACTERIZATION

There is an oversight in the description given by Berns *et al.*[3] (and the CIE[4]) for implementing neutral-point characterization. One of the equations used by Berns *et al.* to predict the CIE 1931 *XYZ*-tristimulus values that is produced by specified DAC values is

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_{R,D_{max}} & X_{G,D_{max}} & X_{B,D_{max}} \\ Y_{R,D_{max}} & Y_{G,D_{max}} & Y_{B,D_{max}} \\ Z_{R,D_{max}} & Z_{G,D_{max}} & Z_{B,D_{max}} \end{bmatrix} \begin{bmatrix} \hat{T}_R^{D_{max}} \\ \hat{T}_G^{D_{max}} \\ \hat{T}_B^{D_{max}} \end{bmatrix} \tag{1}$$

where $X_{R,D_{max}}$, etc. are the *XYZ*-tristimulus values obtained from independent measurements of the *R*, *G*, and *B* channels at the maximum-possible DAC value (D_{max}) and $\hat{T}_R^{D_{max}}$, $\hat{T}_G^{D_{max}}$, and $\hat{T}_B^{D_{max}}$ are *RGB* monitor tristimulus values, which represent the outputs of the *RGB* channels and are normalized with respect to $T_{R,D_{max}}$, $T_{G,D_{max}}$, and $T_{B,D_{max}}$, respectively. For purposes of the present discussion, it may be convenient to think of $\hat{T}_R^{D_{max}}$, etc. as normalized luminances produced by the *R*, *G*, and *B* channels, but they can be calculated using any units that are proportional with the radiances produced by the *RGB* channels.

When GOG is used to predict the *XYZ*-tristimulus values that are produced by specified DAC values, values of $\hat{T}_R^{Dmax}$, $\hat{T}_G^{D_{max}}$, and $\hat{T}_B^{D_{max}}$ are obtained by evaluating the GOG power functions for those DAC values and then Eq. (1) is used to compute *XYZ*. When GOG is used to predict the DAC values that yield desired *XYZ*, the inverse of Eq. (1) is used to compute the values of $\hat{T}_R^{D_{max}}$, $\hat{T}_G^{D_{max}}$, and $\hat{T}_B^{D_{max}}$ that yield *XYZ*, and inverse GOG power functions are then used to compute the DAC values that yield $\hat{T}_R^{D_{max}}$, etc.

If channel additivity is perfect, evaluating the inverse of Equation 1 for W_{max} yields $\hat{T}_R^{D_{max}} = \hat{T}_G^{D_{max}} = \hat{T}_B^{D_{max}} = 1$. Typically, though, channel additivity is imperfect, and these values differ from unity. In such cases, if neutral-point characterization is used, Berns *et al.* and the CIE recommend renormalizing all values of $\hat{T}_R^{D_{max}}$, $\hat{T}_G^{D_{max}}$, and $\hat{T}_B^{D_{max}}$ that are obtained via the inverse of Eq. (1) with respect to the values obtained for W_{max}. Renormalization yields

$$\frac{\left(\dfrac{T_R}{T_{R,D_{max}}}\right)}{\left(\dfrac{T_{R,W_{max}}}{T_{R,D_{max}}}\right)} = \hat{T}_R^{W_{max}}, \quad \frac{\left(\dfrac{T_G}{T_{G,D_{max}}}\right)}{\left(\dfrac{T_{G,W_{max}}}{T_{G,D_{max}}}\right)} = \hat{T}_G^{W_{max}},$$

$$\text{and} \quad \frac{\left(\dfrac{T_B}{T_{B,D_{max}}}\right)}{\left(\dfrac{T_{B,W_{max}}}{T_{B,D_{max}}}\right)} = \hat{T}_B^{W_{max}}, \tag{2}$$

So, when neutral-point characterization is used with GOG, the measurements of *N* are decomposed using the inverse of Eq. (1), the results are renormalized if necessary using Eq. (2), and coefficients are fit to the GOG power functions to predict the resulting monitor tristimulus values. The outputs of the power functions are used subsequently in conjunction with the 3 × 3 matrix in Eq. (1) to predict *XYZ* for specified DAC values.

This procedure produces an inconsistency, however, when renormalization is used. The inconsistency can be illustrated by substituting the renormalized monitor tristimulus values into Eq. (1), expanding the 3 × 3 matrix, and rearranging terms to yield

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \dfrac{x_R}{y_R} & \dfrac{x_G}{y_G} & \dfrac{x_B}{y_B} \\ 1 & 1 & 1 \\ \dfrac{z_R}{y_R} & \dfrac{z_G}{y_G} & \dfrac{z_B}{y_B} \end{bmatrix} \begin{bmatrix} \hat{T}_R^{W_{max}} * Y_{R,D_{max}} \\ \hat{T}_G^{W_{max}} * Y_{G,D_{max}} \\ \hat{T}_B^{W_{max}} * Y_{B,D_{max}} \end{bmatrix} \tag{3}$$

where x_R, etc., are the chromaticity coordinates obtained from the individual measurements of *R*, *G*, and *B*. Thus, it can be seen that the use of renormalized monitor tristimulus values with Eq. (1) implicitly multiplies monitor tristimulus values that have been normalized with respect to W_{max} by $Y_{R,D_{max}}$, $Y_{G,D_{max}}$, and $Y_{B,D_{max}}$. That is, monitor tristimulus values that have been normalized with respect to the peak white are multiplied by the peak luminances obtained from measuring the channels individually, rather than the luminances the tristimulus values were actually normalized against (i.e., the peak channel luminances implied by the peak white).

This inconsistency adds unnecessary error to the GOG predictions. For example, if Eq. (3) is used to predict the *XYZ*-tristimulus values produced by $\hat{T}_R^{W_{max}} = \hat{T}_G^{W_{max}} = \hat{T}_B^{W_{max}} = 1$, the resulting *XYZ* are the sums of the values obtained from the individual measurements of *R*, *G*, and *B* at D_{max} rather than the values obtained for W_{max}. If *XYZ* are calculated for $\hat{T}_R^{W_{max}}$, $\hat{T}_G^{W_{max}}$, and $\hat{T}_B^{W_{max}}$ corresponding with other *N* measurements, the results do not match the values that were actually obtained, either.

One way to correct the problem is to replace Eq. (1) with

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} Y_{R,W_{max}} \dfrac{x_R}{y_R} & Y_{G,W_{max}} \dfrac{x_G}{y_G} & Y_{B,W_{max}} \dfrac{x_B}{y_B} \\ Y_{R,W_{max}} & Y_{G,W_{max}} & Y_{B,W_{max}} \\ Y_{R,W_{max}} \dfrac{z_R}{y_R} & Y_{G,W_{max}} \dfrac{z_G}{y_G} & Y_{B,W_{max}} \dfrac{z_B}{y_B} \end{bmatrix} \begin{bmatrix} \hat{T}_R^{W_{max}} \\ \hat{T}_G^{W_{max}} \\ \hat{T}_B^{W_{max}} \end{bmatrix} \tag{4}$$

whenever channel additivity is imperfect, where

$$\begin{bmatrix} Y_{R,W_{max}} \\ Y_{G,W_{max}} \\ Y_{B,W_{max}} \end{bmatrix} = \begin{bmatrix} \dfrac{x_R}{y_R} & \dfrac{x_G}{y_G} & \dfrac{x_B}{y_B} \\ 1 & 1 & 1 \\ \dfrac{z_R}{y_R} & \dfrac{z_G}{y_G} & \dfrac{z_B}{y_B} \end{bmatrix}^{-1} \begin{bmatrix} X_{W_{max}} \\ Y_{W_{max}} \\ Z_{W_{max}} \end{bmatrix} \tag{5}$$

and $X_{W_{max}}$, etc. are the *XYZ*-tristimulus values of W_{max}. Alternatively, replace Eq. (1) with

TABLE IV. Mean errors for Equations 3 vs. 4 at resolution 13.

Method	Eq.	|%Y|	u'v'	ΔE^*_{uv}	ΔE^*_{ab}
GOG	3	1.41	0.00119	0.83	1.01
	4	0.63	0.00084	0.62	0.78
PLCC	3	1.44	0.00119	0.83	1.01
	4	0.62	0.00085	0.61	0.77
PLGCC	3	1.31	0.00109	0.77	0.95
	4	0.64	0.00073	0.56	0.70

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{x_R}{y_R} & \frac{x_G}{y_G} & \frac{x_B}{y_B} \\ 1 & 1 & 1 \\ \frac{z_R}{y_R} & \frac{z_G}{y_G} & \frac{z_B}{y_B} \end{bmatrix} \begin{bmatrix} \hat{T}_R^{W_{max}} * Y_{R,W_{max}} \\ \hat{T}_G^{W_{max}} * Y_{G,W_{max}} \\ \hat{T}_B^{W_{max}} * Y_{B,W_{max}} \end{bmatrix} \tag{6}$$

The choice depends on whether one finds it easier to compute the 3×3 matrix in Eq. (4) or switch between normalized and unnormalized monitor tristimulus values, as required by Eq. (6). For methods like PLCC and PLGCC, which do not require the use of monitor tristimulus values, Eq. (6) can be replaced with

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{x_R}{y_R} & \frac{x_G}{y_G} & \frac{x_B}{y_B} \\ 1 & 1 & 1 \\ \frac{z_R}{y_R} & \frac{z_G}{y_G} & \frac{z_B}{y_B} \end{bmatrix} \begin{bmatrix} Y_R \\ Y_G \\ Y_B \end{bmatrix} \tag{7}$$

which is definitely easier to work with.

We compared Eqs. (3) and (4) by performing calculations both ways for GOG, PLCC, and PLGCC at resolution 13 for all five replications. The results are summarized in Table IV. The error reductions due to Eq. (4) are similar for all three predictive methods. The values of $\hat{T}_R^{D_{max}}$, $\hat{T}_G^{D_{max}}$, and $\hat{T}_B^{D_{max}}$ that were obtained for W_{max} averaged 0.99450, 0.97659, and 0.98378, respectively, across the five neutral-point characterization datasets. Therefore, it seems that our correction yields useful error reductions, even for seemingly small departures from additivity.

APPENDIX C: GOG AND CHARACTERIZATION MEASUREMENTS AT D_{max}

There is a problem with using characterization measurements performed at D_{max} in conjunction with GOG. The GOG power function for predicting the monitor tristimulus value for a given channel has the form

$$T = T_{max}(k_o + k_g D/D_{max})^\gamma, \tag{8}$$

where D is the selected DAC value for a given color channel; D_{max} is the maximum possible DAC value; T_{max} is the channel's monitor tristimulus value at D_{max}; k_o, k_g, and γ are coefficients representing the channel's offset, gain, and gamma, respectively; and T is the predicted monitor tristimulus value. Berns *et al.*[3] showed that $k_o + k_g = 1$; therefore, by substitution, Eq. (1) can become

$$T = T_{max}(1 - k_g + k_g D/D_{max})^\gamma \tag{9}$$

or

$$T = T_{max}(k_o + (1 - k_o) D/D_{max})^\gamma. \tag{10}$$

Either way, only two free coefficients appear, so only two measurements are needed to uniquely determine the values of these coefficients for each channel. To the best of our knowledge, no closed-form solution is possible, so nonlinear regression (i.e., search techniques) is the only available means of obtaining solutions. We used the Marquardt method, as implemented in the Statistical Analysis System's NLIN procedure, for all GOG tests reported in this article.

By definition, a measurement at D_{max} yields T_{max}; therefore, setting $D = D_{max}$ in Eq. (9) or (10) yields the identity $T_{max} = T_{max}$. If one attempts to solve for the coefficients using only two measurements, one of which was obtained at D_{max}, the D_{max} measurement, therefore, places no constraints on the coefficients. One is left with two free coefficients to fit the one remaining measurement, and an infinite number of solutions becomes available. If three or more measurements are used instead, only one solution is possible, but it is the same one that would have been obtained if the D_{max} measurement had been omitted.[‡] In other words, measurements at D_{max} contribute nothing to solutions to Eqs. (9) and (10).

The same problem can arise in a more complicated form with Eq. (8), which demands a minimum of three measurements to obtain a unique solution. As the regression program converges toward values of k_o and k_g that sum to unity, a D_{max} measurement contributes progressively less to the solution, and so the estimates of k_o, k_g, and γ become unstable. If more than three measurements are used, a D_{max} measurement does not interfere with convergence, but this is because it contributes to the solution only in the early stages.

ACKNOWLEDGMENTS

Gary Beckler (Logicon Technical Services, Inc.) wrote much of the software needed to conduct this research. Chuck Goodyear (Chuck's Discount Stats) assisted with the statistical analyses. Patrick Monnier (Sytronics, Inc.) final-

[‡] If more than three measurements are used, no true solution is possible, but an optimal solution can be obtained by minimizing a statistic such as the sum of squared error.

ized the figures. Roy Berns (Munsell Color Science Laboratory) verified the correction reported in Appendix B. Louis Silverstein (VCD Sciences, Inc.) provided many helpful comments on the entire manuscript. We thank these individuals for their contributions to the concepts and results described in this article.

1. Post DL, Calhoun CS. An evaluation of methods for producing desired colors on CRT monitors. Color Res Appl 1989;14:172–186.
2. Lucassen MP, Walraven J. Evaluation of a simple method for color monitor recalibration. Color Res Appl 1990;15:321–326.
3. Berns RS, Motta RJ, Gorzynski ME. CRT colorimetry, part I: theory and practice. Color Res Appl 1993;18:299–314.
4. CIE. The relationship between digital and colorimetric data for computer-controlled CRT displays. CIE Publ 122. Vienna: CIE Central Bureau; 1996.
5. Berns RS. Methods for characterizing CRT displays. Displays 1996; 16:173–182.
6. Howard CM. Display characteristics of example light-valve projectors. Tech Rept No AFHRL-88-44. Brooks Air Force Base, TX: Air Force Human Resources Laboratory; 1988.
7. Howard CM. Color performance of light-valve projectors. Proc Soc Photo-Opt Inst Eng (SPIE) 1989;1081:107–114.
8. Takaoka H, Ogata Y, Minato S. A new input modulation method for generating expected colors on a CRT monitor. SID Int Symp Digest 1991;57–60.
9. Fiske G, Silverstein LD. Accurate color rendering for display simulation. SID Int Symp Digest 1993;133–136.
10. del Barco LJ, Díaz JA, Jiménez JR, Rubiño M. Considerations on the calibration of color displays assuming constant channel chromaticity. Color Res Appl 1995;20:377–387.
11. Silverstein LD, Fiske TG. Colorimetric and photometric modeling of liquid crystal displays. Proc First IS&T/SID Color Imag Conf 1993; 149–156.
12. Stokes M, Fairchild MD, Berns RS. Precision requirements for digital color reproduction. ACM Trans Graph 1992;11:406–422.
13. MacAdam DL. Visual sensitivities to color differences in daylight. JOSA 32:247–274.
14. Brainard DH. Personal communication. 18 June 1999.
15. Motta RJ. An analytical model for the colorimetric characterization of color CRTs. M.S. Thesis. Rochester Inst Tech; 1991. p 161–173.
16. Bodrogi P, Schanda J. Testing a calibration method for colour CRT monitors. A method to characterize the extent of spatial interdependence and channel interdependence. Displays 1995;16:123–133.
17. Woolfe GJ, Berns RS, Alessi PJ. An improved method for CRT characterization based on spectral data. CIE Expert Symp '97; 1997. p 38–45.