Color Aspects of Variable Data Proofing

Giordano Beretta
Imaging Systems Laboratory
HP Laboratories Palo Alto
HPL-2004-236
December 24, 2004*

E-mail: giordano.beretta@hp.com

variable data printing, custom printing, pre-flight checking, color perception

The Internet in combination with digital presses has allowed the geographical distribution of manufacturing printed materials. An increasing number of printed pieces is customized for the recipient; when each printed piece is different, conventional proofing fails, because it is impossible to proof the entire print job. One frequent problem in automatically generated pieces is the readability of one page element on top of another element; the color combination can be unreadable or clash. I propose simple algorithms to automatically detect and correct color discriminability problems in variable data printing.

* Internal Accession Date Only Approved for External Publication
© Copyright SPIE To be published in and presented at IS&T/SPIE's 17th Annual Symposium on Electronic Imaging, 16-20 January 2005, San Jose, CA

Color Aspects of Variable Data Proofing[*]

Giordano Beretta[†]

Hewlett-Packard Laboratories, 1501 Page Mill Road 1203, Palo Alto, CA, USA 94304

ABSTRACT

The Internet in combination with digital presses has allowed the geographical distribution of manufacturing printed materials. An increasing number of printed pieces is customized for the recipient; when each printed piece is different, conventional proofing fails, because it is impossible to proof the entire print job. One frequent problem in automatically generated pieces is the readability of one page element on top of another element; the color combination can be unreadable or clash. I propose simple algorithms to automatically detect and correct color discriminability problems in variable data printing.

1. PRINTING PARADIGMS

The Internet has enabled new digital printing workflows that are distributed, medialess, and share knowledge resources. After a brief description of these new workflows, I will discuss the new application of distributed variable data printing and focus on some color issues and their solutions. Printing here always refers to commercial printing.

Before we start, we need to understand the difference between publishing and printing. *Publishing* refers to the dissemination of contents. It mainly entails marketing (recognition of emerging markets or creation of new markets), the evaluation of business risks, logistics, and distribution. Modern publishing was invented by Aldus Manutius.[1]

Electronic publishing refers to the use of a traditional workflow where digital technologies are used in most or all stages. *Digital publishing* refers to the use of new workflows enabled by digital technologies and the Internet; it entails chunking, aggregating, syndicating, and re-purposing media-independent content. It further entails personalization and behavioral customer analysis through data mining and "round-tripping."

Printing refers to the manufacturing process of publications. It mainly entails process optimization and cost reduction, manufacturing skills, quality control, sales, finishing, and fulfillment. Modern printing was made possible by Johannes Gutenberg's moveable type technology; it flourished after the invention of modern publishing by Manutius and Luca Pacioli's invention of accounting and business forms.

Digital printing refers to the use of digital technologies in the entire manufacturing process, from layout to fonts, from color correction to pre-flighting, from process distribution to composing and printing one-of-a-kind customized pages, from finishing control to feedback via "round-tripping."

In digital printing the workflows are usually different from the traditional workflows. While in traditional printing the workflow is known when a job comes in, in digital printing the workflow is tuned as the job is executed.

Traditional commercial printing requires large capital investments, long print runs, large warehousing facilities, and care-

*. Copyright 2005 Society of Photo-Optical Instrumentation Engineers. This paper will be published in Proceedings 5670 and is made available as an electronic preprint with permission of SPIE. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

†. giordano.beretta@hp.com, http://www.hpl.hp.com/personal/Giordano_Beretta/

ful printing volume forecasting. In response to the invention of high-speed digital copiers and printers, the press manufacturers have introduced direct-to-plate presses, which allow shorter runs but still have fast printing time and allow sophisticated finishing processes.

The digital printer manufacturers have responded to these direct-to-plate presses with digital presses, which from a marking technology point of view are more like digital printers, but have the user interfaces, speed, media handling capabilities, and finishing options of traditional presses. Furthermore, they offer the flexibility inherent to digital technologies.

During the Nineties it was believed the killer application for digital presses would be *printing on demand* (POD) books, to avoid transportation and warehousing cost. This application never had a real breakthrough because conventional printing is so efficient that transport and warehousing are not an important factor, while it allows for more media flexibility and better finishing. This does not mean that POD is dead — it is well and alive when personal printers are used to print just that part of a manual that is needed at the moment, for convenience.

A new promising application for digital presses is *variable data printing* (VDP), also known as *customized printing,* where a rich template is populated with different data for each copy, typically merged from a database or determined algorithmically. Examples are:

- permission based marketing, where each copy is personalized with recipient name and the contents is chosen based on parameters like sex, age, income, or ZIP code
- do-it-yourself catalogs, where customers describe to an e-commerce vendor their purchase desires and vendors create custom catalogues with their offerings for that desire
- customized offers in response to a tender for bids, with specification sheets, white papers, and prices customized for the specific bid
- insurance and benefits plans, where customers or employees receive a contract with their specific information instead of set of tables from which they can compute their benefits
- executive briefing materials
- comic magazines, where the characters can be adapted to various cultural or religious sensitivities and the text in the bubbles can be printed in the language of the recipient
- photo albums, generated from pictures a consumer (typically “soccer moms”) or professional (typically wedding photographers) have uploaded on the provider’s Web site

In a way, variable data printing is the generalization of Gutenberg’s idea of reusable type to reusable contents.

2. INTERNET’S IMPACT ON DIGITAL PRINTING

Digital printing and workstations are intimately connected to networking. In fact, the Ethernet was originally invented to connect the first laser printer to the first workstation.[2] When local area networks are interconnected via the Internet, the various parties involved in the production of a piece can be anywhere in the world, at distances no longer limited by the immediate accessibility via courier.

With this geographical distribution, tasks are executed wherever in the world the best trained and most efficient workforce is available. The examples above suggest, that digital printing on the Internet enables new workflows, which in turn lead to new business models and new markets.

The concept is that of a market where vendors offer certain content or information. A digital publisher creates knowledge by weaving a trail through this information, in the tradition of storytelling. The digital printer manufactures a piece by traversing this trail, collecting the chunks or resources, which can be stored or cached anywhere on the Internet, and producing an artifact.

The value is when the information is variable and chosen algorithmically instead of being static. This creates a personalized artifact that is more customized and thus more valuable. The personalized print markup language (PPML), an XML

(Extensible Markup Language) language, is commonly used to define such a publication.

2.1. Pre-flighting and proofing

While publishing can be—and often is—subjective and opinionated, printing must be objective and accurate. Thus printing workflows have at each stage checkpoints and feedback, so that quality control is ensured at each manufacturing stage.

Checkpoints are implemented with pre-flight software, consisting in a set of rules a job must satisfy at a certain stage. Typically, a job is checked both at the input of a stage and at the output. Examples of rule conditions are: PDF version, use of transparency, ICC profiles, font types and embedding, image resolution and compression, …

Once a piece has been copy-edited and the content approved, the print is simulated, either on a display monitor (soft-proof) or on a proof printer. The proof is used to verify the appearance of the piece, i.e., attributes such as image quality, color fidelity or integrity, and layout. Digital presses have the advantage that a single copy costs as little as ten thousand, so the press itself can be used as the proofing device, with the benefit of using the same media and inks.

While it is easy to automate pre-flighting, this is much harder for proofing, because printing is always a compromise where human judgement is required to evaluate the appropriateness of a given set of attributes for a given job. In traditional printing the final proof is inspected visually by the customer and approved. In variable data printing, each printed copy is different, and it is not possible to proof each copy.

When there are small problems, like a little underflow or overflow, the elements on a page can be slightly nudged, scaled, or cropped. When the overflow is larger, the failure can be fatal, because elements will overlap and may no longer be readable or discriminable because the contrast is too low. Examples are shown in see Fig. 1 and Fig. 2.

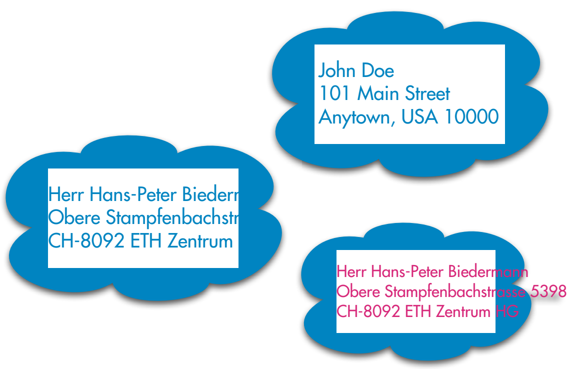

Fig. 1. Mailing label on a VDP piece. The cloud indicates that the area surrounding the address is colored. In this example the designer has chosen to print the address in the same color as the surround. While this can work well on short US addresses, it fails for example on addresses in German, where critical information for the delivery of the mail piece becomes invisible. When a contrasting color is chose to print the address, critical information is still readable to the letter carrier.

Fig. 2. In this VDP example, in-store advertisements for sales promotions are printed by enumerating information from a database and placing it on a rich template, whose main color is chosen to attract customers' attention. When the image of an item happens to be of the same or a similar color as the background, the promoted item can "disappear" in the template.

Similarly, when background and foreground color sequence through palettes, color combinations can be generated which due to insufficient contrast make a text unreadable for readers with color vision deficiencies or even for those with normal color vision. In the case of images, they can sink into a background or become too conspicuous. This problem also happens in marketing materials when elements receive indiscriminable color combinations. Examples are shown in see Fig. 3 and Fig. 4.

Fig. 3. Two HP corporate palette colors that are hard to read and impossible to read for people with deficient color vision. In permission based marketing the graphical and text elements on a page are often cycled though a corporate palette designed for manual design; color combinations are often not checked for compatibility.

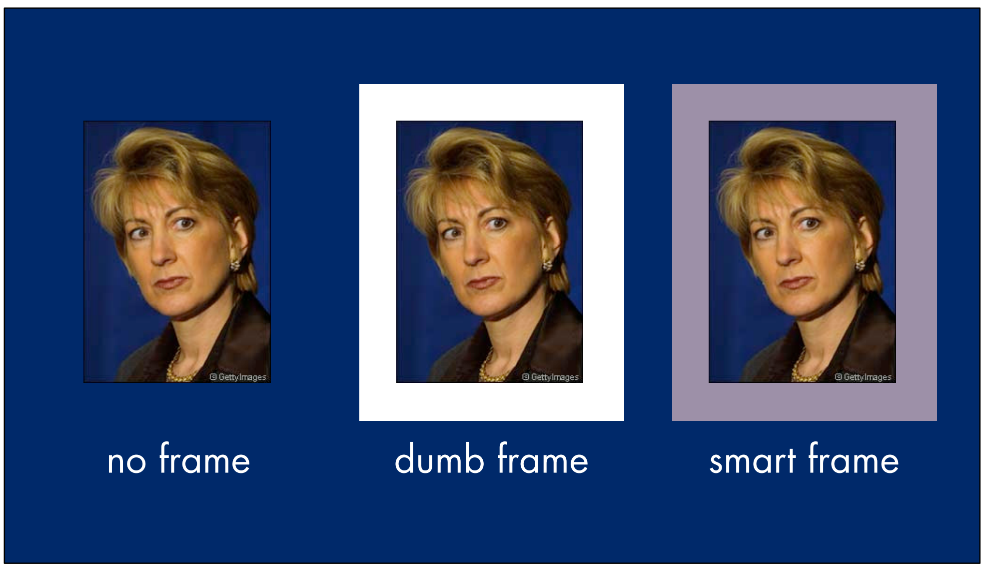

Fig. 4. When images are automatically placed in a template for a photo album, they can be perceived to sink in the background if the background color selection is unfortunate. Placing a white or black frame around images does not always work and can elicit the opposite perception. We would like a "smart" frame that lets the use perceive the image as being located just on the printed album page.

This problem can easily occur in marketing materials, where colors can be cycled through a limited palette. This problem can be very subtle; for example, corporations often update their color palettes as their branding evolves. In VDP "old" material is printed substituting with the updated color palette for uniform look, and two very different colors can be mapped into close colors.

In the following we show how we can establish rules for the robustness of color combinations for background and foreground elements, so that text remains visible and images do not vanish in the background. These rules can then be incorporated in a pre-flight system and, if appropriate, color can be nudged or substituted. Such a pre-flight step is called *variable data proofing*.

3. COLOR DISCRIMINABILITY

Color discriminability refers to the ability to recognize a colored element on top of another colored element. This concept is not to be confused with color difference, which is based on thresholds, or distinct color, which refers to robustness across media.[7]

3.1. Colored text on a colored background

The color of two overlapping elements is discriminable when they can instantly be told apart. In the case of figurative objects (see Fig. 2) this can be color contrast. However, in general it is better to rely on lightness contrast, because it is best for text readability and is robust for readers with deficient color vision. Reading performance depends not only on the available physical information delivered by a visual stimulus, but also on the ability to process this information cognitively, that is, to utilize it. Legge[3] has shown that high contrast is particularly critical for people with low vision.*

In the following I describe the algorithms in a variable data proofing module, which gives the print workflow implementors three functions. A first function returns a boolean result indicating the discriminability of a foreground element on a given background color. A second function returns a text string that gives a human readable explanation of the problem if there is one. The third function suggests a better foreground color when there is a discriminability problem.

Fig. 5 illustrates how the lightness difference between background and foreground colors is assessed and a suggestion is computed. If the lightness difference is above a *threshold*, there is no discriminability issue. To establish this threshold I started with 30 units in CIELAB, a value inspired by the largest distance between focal color† and centroid‡ for the basic color terms.[5]

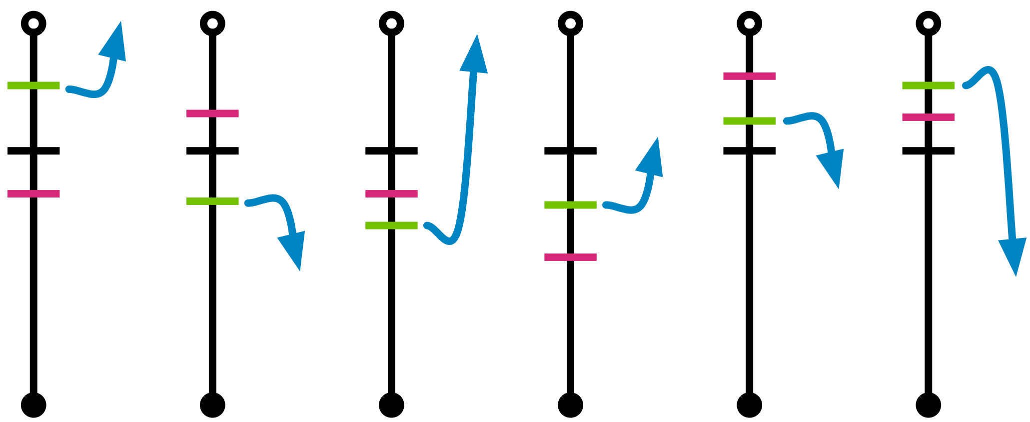

Fig. 5. An algorithm to adjust the lightness of a foreground color in relation to a background color. The vertical line represents the lightness axis and the black horizontal tick marks medium gray. The darker (violet) tick marks the background color, the lighter (green) tick at the arrow start the foreground color. The arrow indicates how the lightness of the foreground color is adjusted.

The sRGB color space has then been surveyed informally by one observer with normal color vision (the author) to scope out a lower threshold. The devices where a calibrated sRGB LCD display monitor and a dry toner color laser printer. For this empirical set-up, a value of 27 CIELAB units on the lightness axis proved to be the lowest bound for discriminability of a pair of background and foreground colors.

The second concept introduced is that of a lightness *bias*. The lightness range can grossly be divided in a a dark half with values between 0 and 50 and light half with values between 50 and 100 lightness units. In practice, even when a display

*. People with normal vision are quite tolerant to contrast loss; a tenfold reduction from maximum contrast results in only a slight decrease in reading speed (less than a factor of two). However, the reading speed of people with low vision declines with any reduction from maximum contrast.

†. Focal colors are defined as the fastest named consensus sample (all subjects name that sample consistently with that color term) for each basic color sample.

‡. Centroid values are calculated by taking an average of hue, chroma, and lightness from all responses using the same color name in a color naming experiment.

monitor is accurately calibrated, it may still be deployed in incorrect viewing conditions. The ensuing glare effectively reduces a considerable portion of the shadow range.

A similar effect happens also in printers, where low-cost papers are often substituted for the standard paper used in the calibration, resulting in soft halftones and detail loss in shadow and very vivid areas. For the equipment described above the best informal estimate for the effective mid-point between light and dark (bias) was 70 lightness units. Both the discriminability threshold and the bias are just empirical values that can change from application to application, but I believe they are typical for the real world.

With this information it is easy to understand Fig. 5, where the lightness for the foreground color is changed depending on the relative darkness and lightness of the two colors. The difference between foreground and background lightness is computed and compared to the threshold value described above. When the difference falls below this threshold, the lightness of the foreground color is changed depending on its position relative to the medium gray and the background color. The amount of the lightness change is calculated to push the new difference above the threshold.

The hue and chroma can also be nudged by a small fudge amount if they are too close, but in general, especially changing hue, is a design decision. The designer should add metadata to the elements in a template that instruct a variable data proofer whether the hue and chroma of a colored element may be changed.

This algorithm is not directly visible to the user, because it is burrowed deep into the print workflow system. For demonstration purposes I have written a small graphical user interface shown in Fig. 6 to illustrate the algorithm. Because the algorithm is based on lightness alone, it is very robust for readers with a color vision deficiency. This fact is demonstrated with a simulation based on the well-known method suggested by Viénot and Brettel.[6]

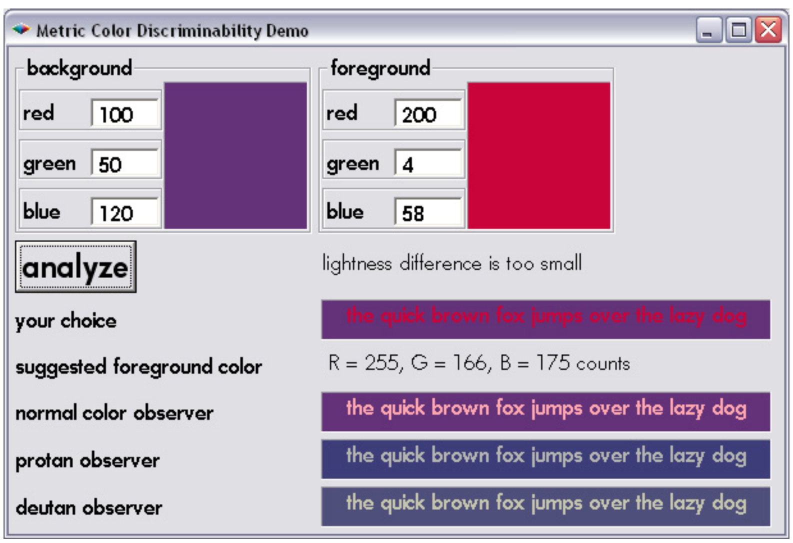

Fig. 6. A simple demo tool for the algorithm in a variable data proofing module checking the discriminability of a foreground color vs. a background color. The demonstrator can enter sRGB triplets for the two colors and click on the "analyze" button to get a diagnosis and if necessary a suggestion for a better foreground color. The demo tool shows a text sample with the original and the suggested colors, as well as a simulation for readers with a color vision deficiency.

3.2. Functional color

The described method applies to color pairs. A related problem is that of assessing, and if necessary correcting, a palette of functional colors, as for example in descriptive statistical representations such as pie charts. Many papers have been written on this subject in the past 25 years and this topic has been exhausted, so it will not be mentioned here.

Corporate materials make extensive use of color palettes, usually specified as spot colors. One problem new to VDP is that these palettes change as a corporation's branding evolves, while the corpus of marketing materials can contain pages created with an earlier palette. When such materials are mixed in a single piece, an older page will stand out by its different spot colors.

As corporate palettes contain a limited number of colors, these palettes can easily be stored, and a module in the print workflow can check the specific palettes from which each spot color originates. When a color from an obsolete palette is encountered, it can automatically be replaced by the nearest color in the current palette. Nearest is simply the smallest Euclidean distance in a perceptually uniform color space. In the implementation described here, the computation is carried out in CIELAB.

3.3. Image frames

To assess the perceived location perpendicular to the page (z-depth) of an image vs. a colored background, it is useful to recognize that the concepts of perceived z-depth, shadows, saturation, contrast, contrast-contrast, and transparency are all related. Unfortunately the perceived image location has not yet been scaled in terms of a colorimetric correlate. However, the literature yields results that can be useful to develop a heuristic approximation. The paper by Singh and Anderson is a good starting point.[4]

As known since E. Hering's work, when a color is placed between two colors and the first color is a mixture of the two adjacent colors, then the perception of transparency can be elicited. Fig. 7 illustrates that as a starting point for a frame color we compute the average color of the image center (say, 60%) in a perceptually uniform color space. The hue of the color halfway between this center color and the background is selected for the frame hue.

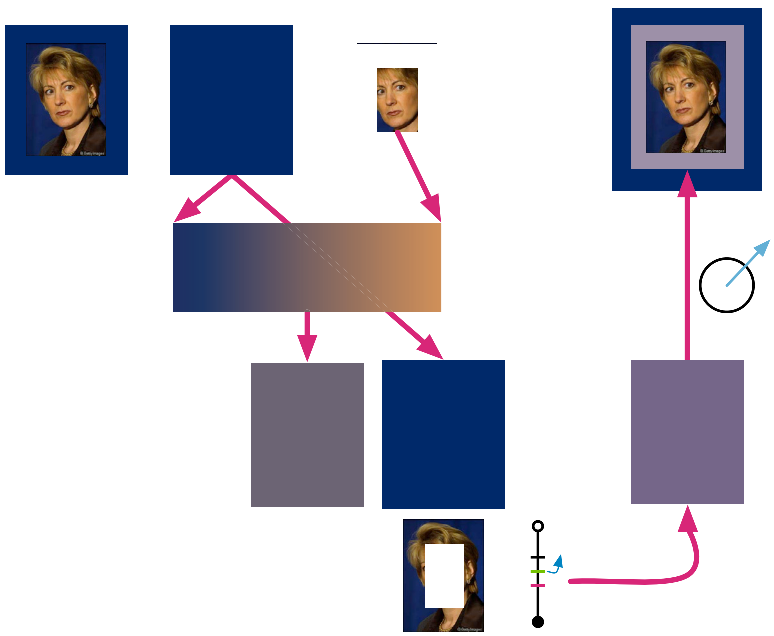

Fig. 7. A heuristics to compute the color for a frame for an image on a colored background. The hue is chosen between the center's average color and the background color. The lightness is chosen as described in Fig. 5, where the page color is the background and the average color of the image periphery (the image less the center) is the foreground color. Finally, the chroma is selected in function of the contrast-contrast, comparing the Michelson contrast in the image center and the contrasts introduced by the frame.

In a second step the average color is similarly computed for the peripheral area of the image. The method described earlier for suggesting the lightness for colored text on a colored background (see Fig. 5) is then used to chose the frame lightness.

Choosing the chroma or saturation is the tricky part. The following heuristic has been implemented. First the contrast between background and image periphery, contrast between frame and background, and contrast between periphery and frame are calculated. The maximum of these three contrasts is chosen as an estimate for the contrast introduced by the frame.

When the contrast of frame is larger than the Michelson contrast of the image center, then the latter perceived contrast is reduced; therefore, in this implementation the frame vividness is heuristically increased above the mean between background and average image center. More precisely, when the contrast between frame and background is larger that the contrast between frame and image periphery, the frame is made very vivid, otherwise just more vivid. When the image contrast is not reduced by the frame, then the vividness is left at the mean value.

4. CONCLUSIONS

The Internet and digital presses are the enablers for new variable data printing applications, such as personalized marketing materials and photo albums. This introduces a new challenge for the design of printing workflows, because a job can no longer be proofed. The solution is to create new pre-flight tools that check each page for visual flaws and corrects flows when possible.

This paper has described heuristic methods to check and correct the colors of an element on a background and to pick the frame color for an image on a colored background.

5. REFERENCES

1. G. Beretta, "Internet's impact on publishing," *IS&T/SPIE Symposium on Electronic Imaging Science & Technology,* Technical Conference **3019**, San José (California, USA), 200–211, February 1997.
2. D. Boggs, Personal Communication.
3. G.E. Legge, "Reading: effects of contrast and spatial frequency," *Applied Vision*, OSA Technical Digest Series, **16**, 90–93, 1989.
4. M. Singh and B.L. Anderson, "Toward a Perceptual Theory of Transparency," *Psychological Review*, **40**, 23, 492–519, 2002
5. J Sturges and T.W.A. Whitfield, "Locating Basic Colours in the Munsell Space," *Color Research and Applications*, **20**, 6, 364–376, December 1995.
6. Françoise Viénot and Hans Brettel, "Color display for dichromats," *IS&T/SPIE Symposium on Electronic Imaging Science & Technology*, Technical Conference **4300**, San José (California, USA), 199–207, January 2001.
7. Xerox. *Color Encoding Standard*, **XNSS 288811**. Xerox Systems Institute, 475 Oakmead Parkway, Sunnyvale, CA 94086, 1989.