Effect of Copying and Restoration on Color Barcode Payload Density

Steven J. Simske, Margaret Sturgill, Jason S. Aronoff

HP Laboratories
HPL- 2009-177

Keyword(s):
Color tiles, 3D bar codes, security printing, payload density, color

Abstract:
2D barcodes are taking on increasing significance as the ubiquity of high-resolution cameras, combined with the availability of variable data printing, drives increasing amounts of “click and connect” applications. Barcodes therefore serve as an increasingly significant connection between physical and electronic portions, or versions, of documents. The use of color provides many additional advantages, including increased payload density and security. In this paper, we consider four factors affecting the readable payload in a color barcode: (1) number of print-scan (PS), or copy, cycles, (2) image restoration to offset PS-induced degradation, (3) the authentication algorithm used, and (4) the use of spectral pre-compensation (SPC) to optimize the color settings for the color barcodes. The PS cycle was shown to consistently reduce payload density by approximately 55% under all tested conditions. SPC nearly doubled the payload density, and selecting the better authentication algorithm increased payload density by roughly 50% in the mean. Restoration, however, was found to increase payload density less substantially (~30%), and only when combined with the optimized settings for SPC. These results are also discussed in light of optimizing payload density for the generation of document security deterrents.

External Posting Date: August 6, 2009 [Fulltext] Approved for External Publication
Internal Posting Date: August 6, 2009 [Fulltext]

To be Published and presented at DocEng '09, Munich, Germany. September 16-18, 2009

Effect of Copying and Restoration on Color Barcode Payload Density

Steven J. Simske
Hewlett-Packard Labs
3404 E. Harmony Rd., MS 36
Fort Collins CO 80528 USA
1.970.898.9561
Steven.Simske@hp.com

Margaret Sturgill
Hewlett-Packard Labs
3404 E. Harmony Rd., MS 36
Fort Collins CO 80528 USA
1.970.898.2463
Margaret.Sturgill@hp.com

Jason S. Aronoff
Hewlett-Packard Labs
3404 E. Harmony Rd., MS 36
Fort Collins CO 80528 USA
1.970.898.2526
Jason.S.Aronoff@hp.com

ABSTRACT
2D barcodes are taking on increasing significance as the ubiquity of high-resolution cameras, combined with the availability of variable data printing, drives increasing amounts of "click and connect" applications. Barcodes therefore serve as an increasingly significant connection between physical and electronic portions, or versions, of documents. The use of color provides many additional advantages, including increased payload density and security. In this paper, we consider four factors affecting the readable payload in a color barcode: (1) number of print-scan (PS), or copy, cycles, (2) image restoration to offset PS-induced degradation, (3) the authentication algorithm used, and (4) the use of spectral pre-compensation (SPC) to optimize the color settings for the color barcodes. The PS cycle was shown to consistently reduce payload density by approximately 55% under all tested conditions. SPC nearly doubled the payload density, and selecting the better authentication algorithm increased payload density by roughly 50% in the mean. Restoration, however, was found to increase payload density less substantially (~30%), and only when combined with the optimized settings for SPC. These results are also discussed in light of optimizing payload density for the generation of document security deterrents.

Categories and Subject Descriptors
I.4.1 [**Image Processing and Computer Vision**]: Digitization and Image Capture–*Scanning*. I.4.6 [**Image Processing and Computer Vision**]: Segmentation. K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*Authentication*.

General Terms
Algorithms, Security

Keywords
Color tiles, 3D bar codes, security printing, payload density, color compensation, image restoration.

1. INTRODUCTION
Bar codes are increasingly used for point-of-sale, mobile information gathering, and security applications. At point-of-sale, bar codes are read to identify the product, to charge the purchaser for the product, and often to track an individual's purchasing patterns and enhance customer loyalty through coupon, gaming and/or discount awards. For mobile information gathering, organizations such as the Open Mobile Alliance [1] and the GS1 Mobile Com Extended Packaging project [2] are helping to connect the consumer to the branded product through bar code enabled services. In security applications, bar codes—along with RFID chips—are used to provide EPCglobal [3] mass serialization along with other security strings (unique IDs, digital signatures of other printed information, replication of RFID information, etc.).

All of these applications require information to be intentionally encoded in the elements of the barcode, suitable for later reading. Payload density (PD), typically measured in bytes/in^2, describes the amount of information for a given area that is embedded in the bar code—for security and product information purposes. Generally, the higher the PD value [4, 5, 6, 7], the better.

Our goal in this paper is to explore the impact of various factors on color barcode PD. The first, the print-scan (PS) or copy cycle, is important to security since the simplest means of counterfeiting is to make a high-quality copy of the original. To offset the PS cycle, we consider image restoration through saturation equalization (SE). We compare two logical algorithms for authentication—one based on matching to closest intensity in sRGB color space (the "RGB" method) and the other based on closest hue match (the "Hue" method). Finally, we further explore the value of applying spectral pre-compensation—which accounts for color drift in the PS cycle, and can thus decrease the need for restoration [6]—to the colors used in the color bar codes. To show the relative impact of each of these factors on PD, we confine our tests to a single all-in-one (HP 6280) and substrate (HP Multi-Purpose Plain Office Paper).

2. EXPERIMENTS PERFORMED
2.1 Color Tile Bar Code
The experiments performed use a color tile deterrent described in [4]. The color tiles consist of six data-carrying colors: red (R), green (G), blue (B), cyan (C), magenta (M) and yellow (Y). In addition, two black (K) tiles and one of each of the six color squares are placed in the upper left and lower right corners to aid in registration (non-payload indicia). White (W) surrounds the

color tile deterrent and is used for segmentation purposes. All images are printed and scanned in 8-bits/channel, 3-channel RGB-space, and the individual pixels P(i,j) for the images are designated as having R, G and B values in the triplet {r,g,b}.

The color tiles are arranged as in Fig. 1 (left). The tiles are arranged on a 10x10 grid, so that if necessary, the deterrent can support 100 tiles. In the default configuration, there are 64 tiles colored one of {R,G,B,C,M,Y,K} (the center supports a DataMatrix 2D barcode).

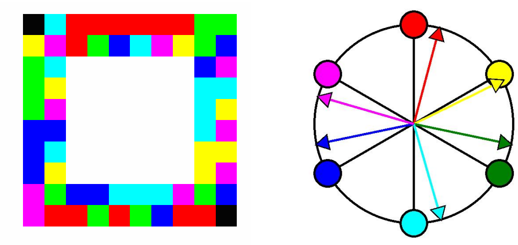

Figure 1. Color tile deterrent (left) and spectral pre-compensation colors (triangles) printed to obtain intended colors (circles) after printing and scanning.

2.2 Calibration and Density of Color Tiles

The color tile deterrent introduced elsewhere [4, 5] uses non-payload indicia (NPI) rather than error-correcting code (ECC) to provide authentication robustness (Fig. 1, left) while providing high payload density (PD), in units of bytes/inch^2. When there are six color possibilities for each payload indicia (PI), the PD is equal to (ln(6)/ln(2))*n(PI)/8, where n(PI) is the number of payload indicia in 1.0 inch^2. In the deterrent of Fig. 1, eight tiles are used to provide orientation and color calibration. Two K tiles provide "Northwest-Southeast" calibration, and the neighboring CMY ("Northwest") and RGB ("Southeast") color calibration tiles are used for 180 degree disambiguation. Note that the color opponency pairings (R-C, B-Y and G-M) are 180 degrees apart (circles in Fig. 1, right), providing the greatest possible color differences for the NPI for robustness of orientation detection. Since 8 NPI are used for the 10x10 (or 100 total) tile area, PI/(PI+NPI) = 0.92 for this deterrent—that is, 92% of the area of the deterrent is used for payload.

2.3 Spectral Pre-Compensation

Spectral pre-compensation [5], or SPC, has been introduced as a means to improve PD. The strategy involves printing an appropriate set of target colors, scanning them and selecting the colors from this set that, *after printing and scanning*, result in the intended color set, {RGBCMY} (e.g. 60 degrees apart depicted by the circles in Fig. 1, right). The selected colors are then used as a replacement for the original {RGBCMY} set.

The SPC color set for the HP 6280 all-in-one are indicated by the triangle-tipped arrows in Fig. 1 (right). Blue (B) is pre-compensated by adding magenta (M), M by adding B, cyan (C) by adding green (G), and G and red (R) by adding yellow (Y). This SPC implementation forces the colors {RGBCMY} to be as close as possible to 60 degrees apart after the first print-scan (PS) cycle. Successful SPC was verified by showing that the metric side-lobe entropy, introduced in [6], increased after applying the SPC colors.

2.4 Tests Performed

Test sheets of multiple color tile deterrents were printed for both uncompensated and SPC color sets. Individual tiles were sized from 5x5 to 20x20 pixels, and all printing was performed at 600 dots/inch (dpi). The 5x5 pixel tiles correspond to a PD of 4650 bytes/in^2. Due to differences in the overall deterrent size, the test sheets included 165, 150, 140, 140, 130 and 117 full deterrents, respectively, for individual tile sizes of 5x5, 6x6, 7x7, 8x8, 9x9 and 10x10 pixels, respectively. This equates to 9240, 8400, 7840, 7840, 7280 and 6552, respectively, individual tiles tested. The color tile deterrents are from 1/6 x 1/6 (5x5 pixel tiles) to 2/3 x 2/3 (20x20 pixel tiles) inches on a side, separated by 0.5 inches in each direction. For each of the 24 types of experiments performed, then, more than 100,000 individual tiles were tested. Printing was performed for uncompensated and spectral pre-compensated tile sheets using the Photosmart C62800 inkjet all-in-one and HP Multi-Purpose Plain Office Paper.

With 8% NPI, the 5x5 pixel color tile PD, if read with 100% accuracy, is 4280 bytes/in^2. For the 20 x 20 pixel color tiles, PD if read with 100% accuracy is 270 bytes/in^2.

2.5 Scanning, Segmentation and Restoration

The test sheets were copied on the 6280 (using the built in copy function with Normal settings). The copy procedure was then repeated for the resulting test sheets to provide us with a second-generation copy. The images of the original, first copy and second copy test sheets were scanned at 600 dpi, 24 bit color, default settings, using the HP ScanJet 8300 scanner. The images were then segmented using custom software. Non-white (non-W) pixels were defined as those having at least one {r,g,b} value < 128 and then run-length smeared using 1/150 of an inch run-length gap. Connected components were formed, and then prepared for authentication.

Each connected component (presumably a single 10x10 tile deterrent as in Fig. 1, left) was divided into 10x10 sections, which correspond to the individual tile zones. The corner tiles were checked for black pixels to ensure orientation was correct, and the individual tile images (100/deterrent) were eroded by 15% along each edge, resulting in, for example, 7x7 pixel, 24-bit color images for each tile in a 10x10 pixel deterrent. The colors of these tiles were determined as described next and compared to the printed sequence of tiles.

Additional restoration was also performed on all image sets. The restoration used was saturation equalization (SE), wherein the argmax{r,g,b} was scaled to 255 and the other two values from the set {r,g,b} were scaled by 255/argmax{r,g,b}.

2.6 Authentication

Two different authentication algorithms were used for the testing. The first, shorthanded the "RGB" approach, assigned the color of each tile as the minimum Euclidean distance between the tile sub-segment's mean {r,g,b} values and the {**r,g,b**} values of the 6 color NPI. That is, each PI tile was assigned the color of the NPI minimizing the sum |r-**r**| + |g-**g**| + |b-**b**|. The second authentication approach, shorthanded the "Hue" approach, assigned the color of each tile based on the minimum of the angular distance of the hue of the tile sub-segment and the hues of the six color NPI. That is, the minimum absolute hue difference between the tile sub-segment's mean {r,g,b} value and the hue of the NPIs' {**r,g,b**} values assigned that NPI's color to the tile. Hue angle of {R,Y,G,C,B, and M} is {0,60,120,180,240, and 300}, although the actual colors after scanning are somewhat different (although SPC preserves the 60 degree separation between the 6 colors).

Authentication accuracy was determined based on the percentage of 56-tile sequences that were correctly read in their entirety (that is, no sequence errors). Even one sequence error was considered an "authentication" failure. Thus, the reported accuracy is "full deterrent accuracy", and is appropriate since no error code checking (ECC) is used when deploying NPI. The commonly reported actual per-tile accuracy is much higher. PD based on, for example, 99.9% per-tile accuracy, was consistently higher or equal to the values reported in Section 3, which follows.

3. RESULTS

For all tests performed, authentication accuracy versus size of the individual tiles was plotted as in Fig. 2. In Fig. 2, X1 equals the intercept of the sloped element of the piecewise linear model with "0% accuracy", and X2 equals the intercept of the sloped element of the piecewise linear model with "100% accuracy". A piecewise-linear estimate for authentication accuracy versus size of tile provided a good estimate for the observed authentication accuracy (correlation coefficient $r^2 > 0.99$). The best fit was determined with 0.001 dpi accuracy for X1 and X2. The piecewise-linear curve determined by minimizing the least-squared error from the actual authentication accuracies to the piecewise-linear curves formed varying (X1, X2) from 4.0 to 21.0 pixels with a step size of 0.001 pixels.

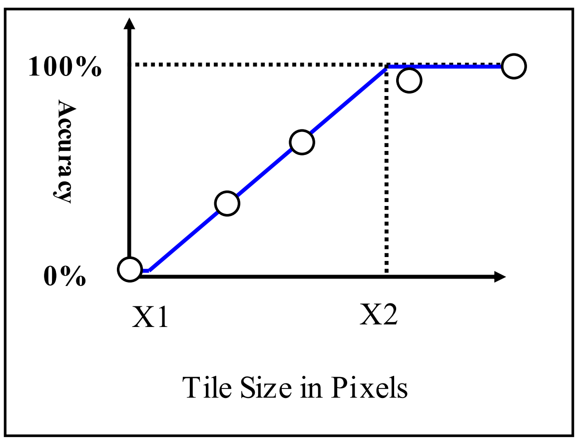

Figure 2. General piecewise-linear model for authentication accuracy (blue/solid line segments) and actual data (circles).

Eq. 1 defines SPD-fA-PL, the security payload density in bytes/in^2 at full (100%) authentication using the piecewise linear model of Fig. 2. The bits/tile is given by ln(6)/ln(2), since there are 6 different payload colors. The percentage of payload-carrying indicia is PI, and the percentage of non-payload, or calibrating, indicia, is NPI. There are 8 bits in a byte, and DPI is the printed dots per inch. X1 and X2 are as described in Fig. 2, in pixels at DPI.

$$SPD-fA-PL = \frac{\ln(6)}{\ln(2)} * \frac{PI}{(PI+NPI)} * \frac{byte}{bits} * \frac{DPI^2}{X2^2} \quad \textbf{(1)}$$

Table 1. Size of the color tile deterrents (in pixels at 600 pixels/inch) with 100% authentication (X2), using the piecewise-linear model. (Hue, RGB)=(hue-based, closest RGB value-based) authentication. SE and SPC as in the text. Orig=Original PS cycle, Copy1 and Copy2 imply 1 or 2 copy cycles after the original PS cycle

	Hue	Hue+SE	RGB	RGB+SE
Orig	9.4	9.8	10.1	10.1
Orig+SPC	8.0	7.1	9.5	8.4
Copy1	14.2	16.1	17.9	17.6
Copy1+SPC	10.3	10.1	14.3	12.5
Copy2	19.6	22.2	35.5	24.3
Copy2+SPC	14.5	16.1	16.7	22.1

Table 2. SPD-fA-PL for each set of deterrents in Table 1

	Hue	Hue+SE	RGB	RGB+SE
Orig	1260	1160	1090	1090
Orig+SPC	1740	2210	1240	1580
Copy1	550	430	350	360
Copy1+SPC	1050	1090	550	710
Copy2	290	230	90	190
Copy2+SPC	530	430	400	230

Table 3. Relative impact of each of the four factors (authentication algorithm, copying, restoration and spectral pre-compensation) on SPD-fA-PL

Factor	Percent Change ($\mu \pm \sigma$)
First Copy	-56.9% ± 9.2%
Second Copy	-52.5% ± 14.7%
Two Copies	-79.3% ± 7.9%
Copy (1,2) + Hue Authentic.	-51.7% ± 7.8%
Copy (1,2) + {rgb} Authentic.	-57.8% ± 15.2%
Copy (1,2) + No restoration	-52.2% ± 15.0%
Copy (1,2) + SE restoration	-57.2% ± 8.5%
Copy (1,2) + no compensation	-58.7% ± 10.9%
Copy (1,2) + SPC	-50.8% ± 12.5%
SPC (Spectral Pre-Comp.)	93.5% ± 84.2%
Hue vs. {rgb} Authentic.	57.2% ± 58.4%
SE Restoration	11.2% ± 39.7%
SE Restoration, SPC Original	27.2% ± 0.3%
SE Restoration, All Other	3.5% ± 42.5%

Eq. 1 therefore provides a single parameter suitable for comparison amongst the other test variables (number of PS cycles, SPC, authentication approach, restoration.). SPD-fA-PL provides a simple means for comparing different bar code density approaches and symbologies on different substrates, using different printers, and between different research teams.

Table 1 provides the X2 values for the 24 sets of experiments performed. Table 2 provides the corresponding SPD-fA-PL data. Table 3 provides the results for the different experiments in terms of their percent impact on payload density (PD).

4. DISCUSSION

The data in Tables 1-3 show that copying (PS cycle) produces a consistent reduction of PD by approximately 55% under all tested conditions. SPC positively impacts PD in 12 out of 12 comparisons (Table 2), though with higher variance than the PS cycle effects. SPC nearly doubled the payload density, while selecting the better authentication algorithm had half the impact of SPC in the mean—increasing PD by roughly 50% in the mean. Restoration, however, was found to increase payload density less substantially (~30%), and only when combined with the optimized settings for SPC. Interestingly, the highest PD reported in these experiments was 2210 bytes/in^2 for "Orig+SPC+Hue+SE" in Table 2. It is worth noting that its non-SPC, non-Hue, non-SE counterpart (1090 bytes/in^2) has 51% less PD—a percentage comparable to an entire PS cycle. Thus, advanced printing and imaging optimizations such as combined SPC+Hue+SE can effectively "remove" an entire PS cycle. The implications of these findings to document security are that advanced printing and imaging techniques should be deployed for security-related color barcodes. Otherwise, document counterfeiters are effectively given a "free pass" of one copy.

The techniques described in our paper can be used to increase the effective payload density (PD) of color tile deterrent. The resulting improved PD matches or exceeds the PD of other barcode types and barcode research efforts. Table IX of reference [7], for example, cites the following bar code PD in bytes/in^2: Data Matrix, 1555; Aztec Code, 1888; QR Code, 1941; Multilevel 2D Bar Code (p_1 - s_1), 2211; and Multilevel 2D Bar Code (p_5 - s_1), 2397. The latter two are similar to the highest density of 2210 bytes/in^2 achieved for the "Orig+SPC+Hue+SE" case as presented here. However, it is unknown how these previous studies determined "100%" authentication, nor what substrates were used. The per-tile accuracies for the 24 cases in Table 1 were computed using the nearest size to the X2 values reported. For these, the per-tile accuracies were 99.9±0.1%, 99.9±0.1%, and 99.7±0.1%, respectively, for the Orig, Copy1 and Copy2 tiles. Thus, the size at 100% accuracy as predicted by the linear model corresponds to 99.9% exact per-tile accuracy.

The color bar code deterrent described here uses 8% of the tiles as color calibration. These are designated non-payload indicia (NPI), and effectively comprised an error-correcting code (ECC) of 8%. This is substantially lower than the usual ECC percentage of 25-50%. In fact, the relative effectiveness of ECC in comparison to using NPI depends on the slope of the piecewise-linear approximation to the authentication curve shown in Fig. 2—more specifically, on the ratio of X1/X2 in this approximation. Suppose, for example, we use ECC=50%. At the midpoint between X1 and X2, the error rate is 50% and so there must be twice as many tiles as a non-ECC, or NPI, guided bar code. This means the midpoint between X1 and X2 is the square root of (0.5), or 0.7071 times X2. This makes X1 = 0.4142*X2. The general form of the equation is:

$$\frac{X2 - X1}{X2} = \frac{1-\sqrt{\eta}}{1-\eta} \qquad \textbf{(2)}$$

Where η=1-ECC, with 0.0 ≤ ECC ≤ 1.0. When ECC=0.25 (25% error correction), (X2-X1)/X2 = (1-0.866)/(1-.75) = 0.536, or X1 = 0.464*X2. The limit of Eq. 2 as ECC→0.0 is 1.0, meaning X1=0.0, and the limit as ECC→1.0 is 0.5, since by L'Hôpital's rule:

$$\lim_{\eta\to 1} \frac{1-\sqrt{\eta}}{1-\eta} = \lim_{\eta\to 1} \frac{-0.5}{-\sqrt{\eta}} = 0.5 \qquad \textbf{(3)}$$

Interestingly, Eq. 3 shows that as the amount of ECC is added, the greater the relative difference between X1 and X2 needs to be to justify using ECC. Regardless, for any amount of ECC to merit the effort, X1 must be ≤ 0.5*X2. In fact, we consistently measured X1 to be 0.6*X2 ≤ X1 ≤ 0.7*X2. These values indicate that NPI should be as effective as ECC for any non-trivial ECC rate for color barcodes.

Color barcodes are likely to become ubiquitous in the next few years. Attention-getting, customizable, and data rich, they are also therefore suitable for secure document workflows. This is an important current area of our research on color barcodes. Future work will focus on more thorough direct comparisons of NPI vs. ECC, as well as the creation of new NPI strategies. We will also investigate alternative color combinations, authentication approaches and image restoration pipelines in order to, potentially, further increase PD.

5. REFERENCES

[1] http://www.openmobilealliance.org/.

[2] http://www.gs1.org/productssolutions/mobile/.

[3] http://www.epcglobalinc.org/home.

[4] Simske, S.J., and Aronoff, J.S. 2007. Qualification of a layered security print deterrent. JIST, 51(1):86-95, 2007.

[5] Simske, S.J., Aronoff, J.S., Sturgill, M.M., and Golodetz, G. 2008. Security printing deterrents: a comparison of thermal ink jet, dry electrophotographic, and liquid electrophotographic printing. JIST, 52(5):50201:1-7, 2008.

[6] Simske, S.J., Aronoff, J.S., Sturgill, M.M., and Villa, J.C. 2008. Spectral pre-compensation and security deterrent authentication. Proc. NIP24, 24:792-795, 2008.

[7] Villán, R., Voloshynovskiy, S., Koval, O., and Pun, T. 2006. Multilevel 2D bar codes: towards high capacity storage modules for multimedia security and management. IEEE Transactions on Information Forensics and Security, 1(4):405-420, 2006.