*Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society (HFES), October 1994, pp. 235–239.*

# RELATIVE EFFECTS OF COLOR-, TEXTURE-, AND DENSITY-CODING ON VISUAL SEARCH PERFORMANCE AND SUBJECTIVE PREFERENCE

Gary Perlman
Department of Computer and Information Science
Ohio State University
Columbus, Ohio 43210 USA
perlman@cis.ohio-state.edu

J. Edward Swan II
Advanced Computing Center for the Arts and Design
Ohio State University
Columbus, Ohio 43212 USA
swan@cgrg.ohio-state.edu

Previously, it had been found that texture-coding was ineffective at reducing search time (Perlman & Swan, 1993). In the experiment reported here, 16 subjects searched for blank-, color-, texture-, and density-coded targets of varying complexity in a naturalistic task. The data showed that all non-blank methods were significantly and about equally more effective at reducing search time than blank-coding (no coding). The difference of outcome with previous results is explained by task simplification and by the control of possibly confounding factors. The difference suggests that coding techniques using texture, and possibly other methods, should be evaluated in context. The similar performance of color-, texture-, and density-coding is explained by the use of equal-saturation and equal-brightness colors. Recommendations for the design of effective coding methods and for future research are discussed.

*Keywords*: Screen design, coding methods, visual search, highlighting, color, tonal, density, texture fill

## INTRODUCTION

The effectiveness of color-coding for improving visual search performance has been demonstrated in many studies (Boff & Lincoln, 1988, sections 11.201–11.208) dating back to the earliest studies in human-computer interaction (Smith, 1962; Smith, 1963; Smith & Thomas, 1964; Smith *et al*, 1965). Other coding methods, such as density/tonal, texture/fill, blink, shape/symbol, line, and typographical coding have also been used to highlight items (Smith & Mosier, 1986). Effective codes make important items "pop-out" (Triesman, 1992) during visual search. If a coding method is equally effective across different numbers or set sizes (e.g., Brown (1991) found that coloring and framing text had this property), then that is evidence that the method is being processed preattentively (Doll, 1993), in parallel (Wolfe, 1993).

The *relative* effectiveness of different coding methods has been used as a basis for the design of visual displays so that the most effective coding method is assigned to the most important type/value in a display (Cleveland, 1985; Mackinlay, 1986; Perlman, 1987). However, there is a lack of empirical evidence for the rankings of coding methods. Intuitively, different visual coding methods are differently effective in different contexts, with color coding being one of the most effective and preferred (blink-coding can be irritating, even though it is highly effective). Beyond color coding, either in the absence of a color display (common on portable computers and most printers), or to code a different dimension, there is the question of which coding method is next most effective?

In a previous experiment that addressed the relative effectiveness of color versus texture coding (e.g. fill patterns such as lines and dots), Perlman and Swan (1993) found that texture coding was not an effective method for speeding visual search; subjects were no faster at finding texture-bordered windows than uncoded ones. Also, texture coding combined with color coding appeared to reduce the effectiveness of color coding (in contrast to results of other studies indicating that redundant coding improves performance (Swierenga *et al*, 1991)). For dispersed windows, the effect of color coding was a .88 second reduction of time compared with no coding, whereas the effect of texture coding was a .04 second increase; the average standard error was .042 seconds, indicating ample power for detecting differences greater than 0.1 seconds.

The surprisingly disappointing performance of texture coding motivated the experiment reported here. Texture coding is reconsidered, again in the context of color coding, but now also in the context of density/tonal coding, which has been found to be effective for communicating relative levels of a dimension (Phillips, 1982). Also, we substitute a simpler task for the window search task, which had required reading labels (a possibly confounding factor in Perlman & Swan (1993) because of the similar, and potentially interfering, spatial frequency of textures and text (Watt, 1988)).

To observe the relative effectiveness of texture and density coding, we introduce a new search paradigm that adds ecological validity to the method used previously (Perlman & Swan, 1993). In Perlman & Swan (1993), subjects searched for color- and texture-coded windows on a workstation screen (coding identified different window types). In the experiment reported here, subjects perform a task similar to that of reading a bar chart, a display in which color-, texture-, and density-codes are used extensively. If texture-coding proved to be ineffective in this experiment, it would add to the evidence that it is not an effective coding method for reducing visual search time. But if texture-coding proved to be effective at speeding search in this task, then it would suggest that the use of texture coding is only effective under certain circumstances, and should be used in a display only after evaluation in context, or after more research.

## METHOD

### *Experimental Task*

The "bar-tracking" task introduced here requires subjects to track two target bars and decide if the bar height increases or decreases (when reading the bars from left-to-right). Each experimental trial begins with the display of a model set of N (2, 4, or 6) bars of equal height showing the codes for each bar position. An arrow below one of the bars indicates which bar (and code) should be tracked on that trial. The subject presses a key and after a pause of 1–2 seconds, two sets of target bars appear to the right of the model. The heights of the first set of target bars all differ from the model, and the heights of the second set of target bars all differ from corresponding bars in the first target set (see Figure 1). The subject's task is to indicate whether the tracked bar height increases or decreases in the target sets by pressing an "up" or "down" key. The task requires that subjects find a bar by position (for blank coding) or by position and code in other conditions, and make a judgment that would only be correct 50% of the time if guessing. A low error rate (e.g., less than 5%) indicates that subjects are finding the tracked bars and making a judgment of their relative heights (a common task when reading bar charts).

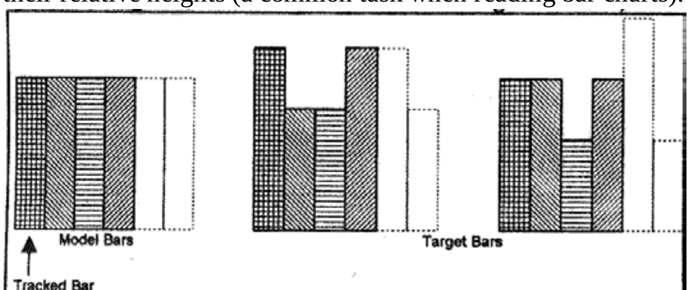


**Figure 1**: Textured bars in which first position is to be tracked in the second two sets. Correct response is "down."

Figure 1 shows Texture coding for 4 bars, with dotted bars where additional bars would be in the 6-bar condition. The texture fills are for illustration only, not the ones actually used in the experiment. The subject has viewed the model bars at the left, noted that the bar to track is the first, and requested that the target bars be shown on the right. The correct response for the trial is "down" because the tracked bar goes down in the target set.

### *Independent Variables*

*Coding Method*: Four coding methods were evaluated: Blank coding, with no coding in the bars; Color coding, with different colors in different bars; Texture coding, with different texture fills in different bars; and Density coding, with different levels of grey in different bars. Each subject saw all coding methods, with all other factors changing faster than coding method. The presentation order of coding method was controlled in a between-subjects latin square.

*Group Size*: Three group sizes were evaluated within each coding method. As the number of bars increased (from 2 to 4 to 6), using position alone (i.e., blank coding) should become more difficult, and the benefits of effective coding methods should be more apparent. Subjects first searched in groups of 2 bars, then 4, and then 6 — in increasing order to allow practice of easier tasks before the more difficult tasks.

*Position in Group*: Subjects searched for targets in each position, so there were 2 positions for groups of size 2, 4 positions for groups of size 4, and 6 positions for groups of size 6. We expected that the first and last bars would be easiest to pick out using position alone, but as the group size increased, interior bars (of which there are none for groups of size 2) would be harder to locate, especially for Blank coding. For each group size N, N codes of a coding method were chosen at random from the total set of codes (6 Color, 6 Density, 16 Texture). For example, subjects seeing N color codes would search for colors in N different positions for each replication.

*Replication*: Subjects were presented with 4 replications of each pass through all of the positions in a group; each replication was with a different set of codes sampled from the codes within a method. Thus, subjects had additional breadth of experience with different coding assignments within methods, replications better ensured a complete set of data if subjects made some errors, and replications allowed the study of short-term learning.

### *Construction of Stimuli*

Non-blank codes, although discriminable, can be similar within a coding method. For example, Figure 2 shows the complete set of texture fills used in the experiment. Similar textures (e.g., high- and low-frequency diagonal lines) are adjacent. A "restricted" random permutation of the codes ensured that two similar codes were never adjacent in a stimulus (e.g., red / purple / blue, 60% / 80% grey / black, high- / low-frequency parallel lines).

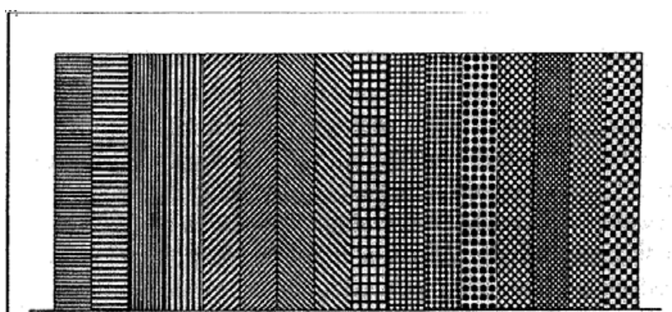

**Figure 2**: All textures used in the experiment. Bars that are adjacent in the Figure are never adjacent in stimuli.

Bars were 36 pixels wide (with a single pixel line between each) because 36 has many divisors, allowing bitmap patterns to be defined on 2x2, 3x3, 4x4, 6x6, 9x9, 12x12, or 18x18 grids. Corresponding bars were always 8 bar-widths apart so that the distance the eye had to travel would be the same for all group sizes.

*Color Codes:* Standard colors used on many computers are based on full intensity red, blue, and/or green, but these differ in psychological brightness. Because density is a factor in this experiment, we devised a set of equal-brightness, equal-saturation hues using Adobe Photoshop on a Macintosh. With approximate names, the RGB values were:

| **Color Name** | **Red** | **Green** | **Blue** |
|---|---|---|---|
| BlueViolet | .638 | .474 | .851 |
| Purple | .720 | .190 | .839 |
| Orange | .730 | .486 | .167 |
| Gold | .651 | .575 | .000 |
| Green | .388 | .652 | .307 |
| BlueGreen | .000 | .635 | .680 |

These appear different on different monitors, making them difficult to reproduce photographically, but on the monitor we used, the colors appeared to be dark pastels with equal saturation and brightness. Similar colors (those adjacent in the color table) were never adjacent in stimuli.

*Texture Codes:* Sixteen textures, shown in Figure 2, were devised to have 50-56% density (about half of the pixels were on). Two subsets of textures were devised to have different spatial frequencies, one twice the frequency of the other. The textures shown in Figure 2 juxtapose similar patterns so that during the experiment, a restricted permutation guaranteed that they would never be adjacent in a display.

For a set of trials with a display of 2, 4, or 6 bars, only the first 2, 4, or 6 bars, respectively, of a permutation were used, so subjects saw a wide variety of textures. This increased the variation of textures to improve the probability of detecting a difference among textures and to allow *post hoc* analyses of which textures were mutually discriminable.

*Density Codes:* To construct a set of six maximally-discriminable densities (of black, to grey, to white), we initially chose 0%, 20%, 40%, 60%, 80%, and 100% intensities. The 0% and 20% levels were difficult to discriminate to us, so we adjusted the levels to: (black) 0, 30, 43, 58, 75, 100 (white). Because close densities were never adjacent in a display, such minor adjustments were not considered critical.

### *Procedure*

Subjects were screened for their vision and given practice trials. Practice trials used groups of 2 bars with no cooing. Subjects were walked through 2 practice trials using verbally augmented written instructions. After this, 10 unassisted practice trials were given, followed by the experimental trials.

### *Dependent Measures*

The primary dependent measure was response time (RT), measured from the time the target bars were presented to the time the subject pressed an up or down key. The response was also recorded to detect errors. After the experiment, subjects rank-ordered the coding methods along four dimensions:

- discriminability of the coding methods,
- visual appeal,
- professionalism, and
- overall preference.

### *Apparatus*

All stimuli were presented and responses collected on an 8-bit color Sun SparcStation IPX running the X window system. The timing resolution on the workstation was about 1 msec. The experimental control program was written with the X11R4 graphics library.

### *Subjects*

Sixteen paid subjects (11 male, 5 female; ages ranging from 24 to 46 with mean = 29.5), with normal-corrected vision without color-blindness, were drawn from a pool of university students and staff. The data of one subject were replaced by that of an additional subject because of an error rate over 5% and a large number of RTs over 10 seconds (5 seconds more than any other subject); the patterns of results were the same with or without this subject's data.

## RESULTS AND DISCUSSION

### *Results*

The results showed that color, texture, and density were all significantly and equally more effective coding methods than no coding.

*Response Time:* Table 1 and Figure 3 show the average RT for the different coding methods, broken down by group size. The non-blank coding methods all improved performance significantly ($F(3,45)=15.99$, $p<.001$), although none of the coding methods was so effective that group size had no effect — the main effect of group size was significant $F(2,30)=40.33$, $p<.001$). There was a coding method x group size interaction ($F(6,90)=3.10$, $p<.01$) because the effect of group size was larger for blank-coding.

| | blank | color | texture | density |
|---|---|---|---|---|
| 2-bars | 1.31 | 1.08 | 1.07 | 1.07 |
| 4-bars | 1.50 | 1.11 | 1.19 | 1.15 |
| 6-bars | 1.80 | 1.26 | 1.27 | 1.33 |

**Table 1**: Response times for coding methods, broken down by group sizes (see Figure 3).

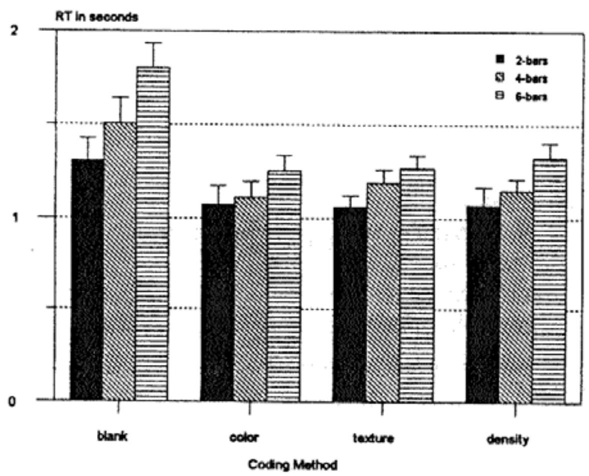


**Figure 3**: Response times for coding methods, broken down by group sizes. Error bars show one standard error.

There was also a significant effect of position in groups of bars. Pooling all positions, there was a significant effect of position ($F(5,75)=15.47$, $p<.001$). In general, it was harder to track bars surrounded by other bars. There was also a significant coding method x position-in-group interaction ($F(15,225)=3.18$, $p<.001$). There was no significant position effect for groups of 2 bars ($F(1,15)=1.69$, $p>.2$). For groups of 4 bars, there was a significant effect of position ($F(3,45)=5.70$, $p<.01$), but no significant interaction with coding method ($F(9,135)=1.39$, $p=.2$). Figure 4 and Table 2 show the average RT for searching groups of 6 bars (the trends are the same as the pooled data). Blank-coding appears to be the most affected by position, and color the least, with texture and density in between, in terms of the variation among the positions. The very best performance in Figure 4 was for the exterior positions of texture coding, although differences between those and corresponding positions in other methods are not significant.

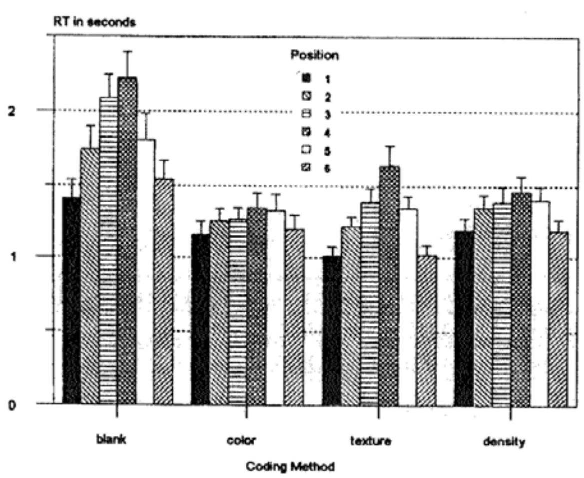


**Figure 4**: Response times for coding methods for groups of size 6, broken down by position in group. Error bars show one standard error.

| position | blank | color | texture | density |
|---|---|---|---|---|
| **1** | 1.40 | 1.16 | 1.01 | 1.19 |
| **2** | 1.75 | 1.25 | 1.22 | 1.34 |
| **3** | 2.09 | 1.27 | 1.38 | 1.38 |
| **4** | 2.22 | 1.34 | 1.64 | 1.46 |
| **5** | 1.81 | 1.32 | 1.34 | 1.40 |
| **6** | 1.54 | 1.20 | 1.02 | 1.19 |

**Table 2**: Response times for coding methods for groups of size 6, broken down by position within group (see Figure 4).

There was no main effect of replication or any significant interactions with replication (all Fs <1).

*Subjective Rankings:* Figure 5 shows the average rankings (high values imply preferred methods) for the four coding methods and the four questions on subjective discriminability, visual appeal, professionalism, and overall preference. Subjects preferred the color- and density-coding methods over texture- and blank-coding. This is mildly surprising because subjects were as fast with texture codes as with color and density codes.

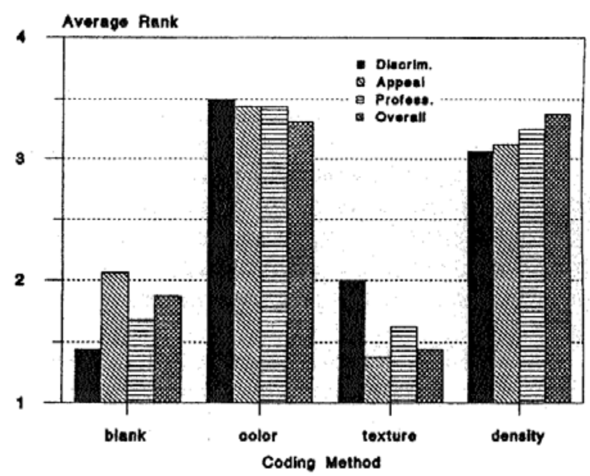


**Figure 5**: Subjective ratings of coding methods. Higher rankings indicate higher subjective discriminability, visual appeal, professionalism, and overall preference.

*Errors:* If subjects were guessing whether the tracked bars went up or down, the expected error rate would be 50%. There were a total of 48 errors in 3072 trials, or about 1.5%. This indicates that the subjects were tracking the bars before responding. No analysis of number of errors showed significant results, although there appeared to be a linear relationship between group size and errors. The average time for trials with errors was 1.38 seconds, which was not very different from the grand mean of all trials of 1.33 seconds; if there was a speed/accuracy tradeoff, it was not a large one.

*Discussion*

The results reported here seem to contradict those of Perlman and Swan (1993) in which texture coding was ineffective at reducing search time. However, the results of that study are highly reliable, so the explanation must lie in differences in the design of the two tasks. One important difference between the two tasks is that the bar-tracking task used here is simpler than the window-locating task in Perlman & Swan (1993). Subjects in Perlman and Swan (1993) had to search for a code, of which there could be many instances on a screen, and then decide which was the correct target by matching a textual label. With textures that have a spatial frequency similar to that of the labels, there could have been interference (Watt, 1988). Another difference is that the textures used here showed 36 pixels compared with 30,20, and 10 pixels in the previous experiment. Regardless of the reason for the difference in results, these findings suggest that texture-coding is not effective in all contexts, and that it and possibly other coding methods should be evaluated in context before designers rely on claims of effectiveness. Although models of preattentive texture discrimination may be a promising area of research (e.g., Sutter *et al*, 1989; Malik & Perona, 1990), we agree with the conclusion of Ware & Knight (1992): "The effective use of texture will for the immediate future be an art, rather than a science, …"

Color-coding was effective in both experiments, suggesting that it can be used as a coding technique without as much need for evaluation in context. However, the results here showed color, texture, and density to be equally effective. This is perhaps unintuitive because color is such a salient attribute. We believe the use of muted colors explains why color-coding was matched by density- and texture-coding. Our colors varied only in hue, not in saturation or intensity, the latter of which was evaluated as grey-scale density here. There is little published exploration of the design space within color, but we believe that by varying saturation and brightness in addition to hue, more effective color sets could be constructed (as is usually done in an uncontrolled fashion when full-intensity colors are used in computer displays).

We believe that firm recommendations on coding techniques are premature because:

- There appears to be a large design space within color. By factoring out density, varying density within different hues might allow encoding independent variables.
- Textures are useful in some contexts (but not all) so the exploration of spatial frequency and other attributes within textures would be worthwhile.
- Density is a useful dimension, but there might not be much design latitude in it because there are not many discriminable levels between 0 and 100%.
- Combinations of methods (e.g., hue+density, hue+texture) should be explored.

## ACKNOWLEDGMENTS

We would like to thank several people at ACCAD for their help: Paul Badger for his help in the construction of the equal-saturation, equal-brightness colors, Phil Ritzenthaler and Stephen N. Spencer for their help with equipment and logistics. We also thank Caroline Palmer for her assistance.

## REFERENCES

Boff, K. R. & Lincoln, J. E. (1988) *Engineering Data Compendium: Human Perception and Performance*. Wright-Patterson AFB, OH: AAMRL.

Brown, T. J. (1991) Visual Display Highlighting and Information Extraction. **Proceedings HFS 35th Annual Meeting**, 1427–1431.

Cleveland, W. S. (1985) **The Elements of Graphing Data**. Monterey, California: Wadsworth.

Doll, T. J. (1993) Preattentive Processing in Visual Search. **Proceedings HFES 37th Annual Meeting**, 1291–1294.

Mackinlay, J. (1986) Automating the Design of Graphical Presentations of Relational Information. **ACM Trans. on Graphics**, 5:2,110–141.

Malik, J. & Perona, P. (1990) Preattentive Texture Discrimination with Early Vision Mechanisms. **Jl. Optical Soc. of America**, 7, 5, 923–932.

Perlman, G. (1987) An Axiomatic Model of Information Presentation. **Proceedings HFS 31st Annual Meeting**, 1229–1233.

Perlman, G. & Swan, J. E. II (1993) Color versus Texture Coding to Improve Visual Search Performance. **Proceedings HFES 37th Annual Meeting**, 343–347.

Phillips R. J. (1982) An Experimental Investigation of Layer Tints for Relief Maps in School Atlases. **Ergonomics, 25**, 1143–1154.

Smith, S. L (1962) Color Coding and Visual Search. **Journal of Experimental Psychology, 64**, 434–440.

Smith, S. L (1963) Color Coding and Visual Separability in Information Displays. **JI. of Applied Psych., 47**, 358–364.

Smith, S. L., & Thomas, D. W. (1964) Color versus Shape Coding in Information Displays. **Jl. of Applied Psych., 48**, 137–146.

Smith, S. L, Farquhar, B. B., & Thomas, D. W. (1965) Color Coding in Formatted Displays. **JI. of Applied Psych., 49**, 393–398.

Smith, S. L. and Mosier, J. N. (1986) *Guidelines for Designing User Interface Software* (ESD-TR-86-278). Bedford, MA: The MITRE Corporation.

Sutter, A., Beck, J., & Graham, N. (1989) Contrast and Spatial Variables in Texture Segregation: Testing a Simple Spatial-Frequency Channels Model. **Perception and Psychophysics, 46**, 4, 312–332.

Swierenga, S. J., Boff, K. R. & Donovan, R. S. (1991) Effectiveness of Coding Schemes in Rapid Communication Displays. **Proceedings HFS 35th Annual Meeting**, 1522–1526.

Triesman, A. (1992) Perceiving and Re-Perceiving Objects. **American Psychologist, 47**:7, 862–875.

Ware, C. & Knight, W. (1992) Orderable Dimensions of Visual Texture for Data Display: Orientation, Size and Contrast. **Proceedings of CHI'92 Human Factors in Computing Systems**, 203–209.

Watt, R. (1988) **Visual Processing: Computational, Psychophysical and Cognitive Research**. Hinsdale, NJ: Lawrence Erlbaum Associates.

Wolfe, J. M. (1993) Guided Search 2.0: The Upgrade. **Proceedings HFES 37th Annual Meeting**, 1295–1299.