Proceedings IEEE Virtual Reality 2005, pages 11–18, © IEEE Computer Society, 2005.

An Empirical User-based Study of Text Drawing Styles and Outdoor Background Textures for Augmented Reality

Joseph L. Gabbard[1]
Systems Research Center, Virginia Tech

J. Edward Swan II[2]
Computer Science and Engineering Department, Mississippi State University

Deborah Hix[3]
Systems Research Center, Virginia Tech

Robert S. Schulman[4]
Statistics Department, Virginia Tech

John Lucas[5]
Computer Science Department, Virginia Tech

Divya Gupta[6]
Nokia, San Diego, California

ABSTRACT

A challenge in presenting augmenting information in outdoor augmented reality (AR) settings lies in the broad range of uncontrollable environmental conditions that may be present, specifically large-scale fluctuations in natural lighting and wide variations in likely backgrounds or objects in the scene. In this paper, we present a user-based study which examined the effects of outdoor background textures, changing outdoor illuminance values, and text drawing styles on user performance of a text identification task with an optical, see-through augmented reality system. We report significant effects for all of these variables, and discuss design guidelines and ideas for future work.

CR Categories: H.5 [Information Interfaces and Presentation]: H.5.1: Multimedia Information Systems — Artificial, Augmented, and Virtual Realities; H.5.2: User Interfaces — Ergonomics, Evaluation / Methodology, Screen Design, Style Guides

Keywords: Outdoor Augmented Reality, Optical See-Through Display, Text Drawing Styles, Background Textures

1 INTRODUCTION

By providing direct, heads-up access to information correlated with a user's view of the real world, augmented reality (AR) has the potential to redefine the way information is presented and accessed. A challenge in presenting augmenting information, particularly in outdoor, optical see-through AR applications, lies in the broad range of uncontrollable environmental conditions that may be present, specifically large-scale fluctuations in natural lighting and wide variations in likely backgrounds or objects in the scene.

One strategy is for visual AR representations and layout to adapt to the conditions of the environment. In many cases, a carefully designed AR user interface may be easily legible under some lighting and background conditions, and moments later may be illegible when conditions change. Since lighting and background conditions may vary from moment to moment in dynamic AR usage contexts, basic research is needed to guide systems developers in understanding the relationship between real-world backgrounds and objects and associated augmenting information.

In the past decade, AR research efforts have successfully tackled several challenging hardware integration problems, so that today AR systems exist that are beginning to function robustly. As such, the field is just to the point where meaningful, systematic human-computer interaction (HCI) research can be conducted and applied. Despite the fact that this technology can fundamentally change the way we visualize, use, and interact with information, very little HCI work has been done specifically in AR. We conducted a survey [1] of seven primary publishing venues for AR research, which comprised a corpus of 880 papers. This survey revealed only 14 user-based studies of AR systems.

Usability engineering activities can be used to determine *what* information should be presented to users, or, for example, *where* information should be presented to users. Of particular interest in our work, however, is *how* augmenting information should be visually presented for optimal usability. This work has been motivated by the empirical results of our application of usability engineering activities to the design and development of the Battlefield Augmented Reality System (BARS) [2; 3]. The focus of the work reported here is studying the effect of environmental conditions on AR text legibility, with a motivation of designing text drawing styles that are optimal for dynamic environmental conditions.

2 RELATED WORK

Although it is well-known (at least anecdotally) that the color and texture of the background environment have a direct bearing on the visibility and legibility of annotations and text in AR systems, very little research has investigated and quantified these effects. Many AR systems, such as the online maintenance system described by Lipson et al. [4], for example, depict labels as white objects with solid black backgrounds. Although such backgrounds are possible for video AR displays, they cannot be used for optical see-through AR displays because see-through displays are additive, which means the color black is considered transparent and thus does not obscure the background view. The AR Quake system, developed by Piekarski and colleagues [5; 6], modified the textures of monsters in the AR Quake game to make them visible against the real world, and also provided recommendations on text color given ambient outdoor lighting conditions. However, to our knowledge, no systematic analysis has been carried out to identify the range over which information from an optical see-through AR display can be readily observed and is most legible under varying outdoor conditions.

Harrison and Vicente [7] describe an "anti-interference" font, designed to produce transparent 2D menus superimposed over different GUI background content, such as applications or the

[1] jgabbard@vt.edu (corresponding email)
[2] swan@acm.org
[3] hix@vt.edu
[4] schulman@vt.edu
[5] jolucas@vt.edu
[6] divya.gupta@nokia.com

Table 1. Summary of variables studied in experiment.

Independent Variables

subject	18	*random variable*
distance	3	near (*1 meter*), medium (*2 meters*), far (*4 meters*)
outdoor background texture (Figure 1)	6	pavement, granite, red brick, sidewalk, foliage, sky
text drawing style (Figure 1)	6	*static*: billboard, red, green *active*: complement, maximum HSV complement, maximum brightness contrast
repetition	4	1, 2, 3, 4

Dependent Variables

response time	in *milliseconds*
error	0 (*correct*), 1 (*incorrect*)
ambient illuminance	in *lux*

desktop. The work also includes an empirical evaluation of the effect of varying transparency levels, the visual interference produced by different types of background content, and the performance of anti-interference fonts on text menu selection tasks. Numerous HCI studies have examined the use of color in traditional 2D user interfaces. Some have researched the contribution of color coding in comparison to other graphical encoding techniques (e.g., Christ [8], Cleveland and McGill [9]), while others attempted to determine the optimum number of colors to use in displays, as well as which colors should be used (e.g., Cahill and Carter [10], Carter [11], Smallman and Boynton [12]). Interesting work by MacIntyre [13; 14] examined color contrast for 2D GUIs; he developed a luminance contrast metric that can be used to ensure a high degree of legibility on CRT displays. We based one of the text drawing styles we describe below, in Section 3.2.2, on MacIntyre's work.

3 The Empirical User-Based Study

We conducted a study that examined the effects of outdoor background textures, changing outdoor illuminance values, text drawing styles, and distance from the user to the background, on user performance of a text identification task. We captured user response time, user error, and measured, and controlled for, variance in natural illumination levels. Table 1 summarizes the variables we examined. Our study is described in more detail by Gabbard [15].

3.1 User Task and Experimental Setup

We designed a task that abstracted the kind of short reading tasks, such as reading labels, which are prevalent in many AR applications. For this study, we purposefully designed the experimental task to be a low-level identification task. That is, we were not concerned with users' semantic interpretation of the data, but simply whether or not users could quickly and accurately read information. The user's task was to identify (find and read) a single numeric digit presented in a text string of randomized, distracting letters. Specifically, each text string contained one, and only one, numeral — either 4, 5, or 6. For each trial, users entered the numeral, using the numeric keypad of a standard extended keyboard, by pressing a key in the middle row of numbers (4 through 6). Users entered a 0 if either they could not find the text string at all (i.e., the string was effectively invisible), or if they could not see a number in the text string. Details of how we generated the distracting letters are given by Gabbard [15].

We wanted to conduct the study using outdoor illuminance values, because while indoor illuminance varies by about 3 orders of magnitude, outdoor illuminance varies by about 8 orders of magnitude [16]. However, we could not conduct the study in direct sunlight, because the graphics of our optical see-through display device (a Sony Glasstron) become almost completely invisible. Furthermore, we wanted to protect the display and other equipment from outdoor weather conditions. We addressed all of these issues by conducting our study in a greenhouse on the Virginia Tech campus. The glass roof of the greenhouse was covered with a white coating that prevented direct sunshine, which diffused and softened the entering sunlight.

We measured the amount of ambient illuminance at the user's position, both to quantify the effect of varying ambient illumination on user task performance, and to ensure that ambient illuminance fell into a pre-determined acceptable range. We used a Cooke Corporation Cal-Light 400 light meter to measure illuminance at the user's position. We established an acceptable range of illuminance between 2,000 lux (outdoor lighting no darker than that of a "cloudy dull" day) and 25,000 lux (no brighter than that of a "cloudy bright" day) [16]. We only ran subjects during the middle of the day, when the sky was clear to at most partially cloudy. In the greenhouse we never measured illuminance values beyond our acceptable range.

Our image generator was a Pentium M 1.6 GHz computer (comparable to a Pentium IV 2.4 GHz) with 772 megabytes of RAM and an NVidia GeForce4 4200 Go graphics card generating monoscopic images. The computer ran under the Linux Mandrake operating system. We used the same computer to collect user data. For the display device, we used a Sony Glasstron PLM A55 biocular, optical see-through display, at NTSC resolution. Because our user task did not require world-centered graphics, we did not use a tracking device.

3.2 Independent Variables

3.2.1 Outdoor Background Texture

We chose six outdoor background textures to be representative of commonly-found objects in an urban setting (the intended location for BARS use [2]): 'pavement', 'granite', 'red brick', 'sidewalk', 'foliage', and 'sky' (Figure 1). In order to display these textures in the greenhouse, as well as to easily control the distance between each background and the user, we created large (40" x 60") posters of each background texture. We captured the textures by taking high-resolution digital photographs, except for 'sky', which we generated using an Adobe Photoshop cloud-rendering algorithm. We made large, matte-finished prints of each texture, which we mounted onto foamcore posterboard. We scaled the prints so that texture features were life-sized, e.g. the bricks on the poster were the same size as the actual bricks on the building we photographed, the leaves were the same size as the actual leaves on the tree we photographed, and so forth.

3.2.2 Text Drawing Style

We created six text drawing styles (Figure 1) based on previous research in typography, color theory, and human-computer interaction text design [15]. Three of the text styles ('billboard', 'red', and 'green') were *static*, meaning that the text color did not change, and three of the text styles were *active,* meaning that the text color changed depending upon the outdoor background texture. For the active text drawing styles we used the average pixel

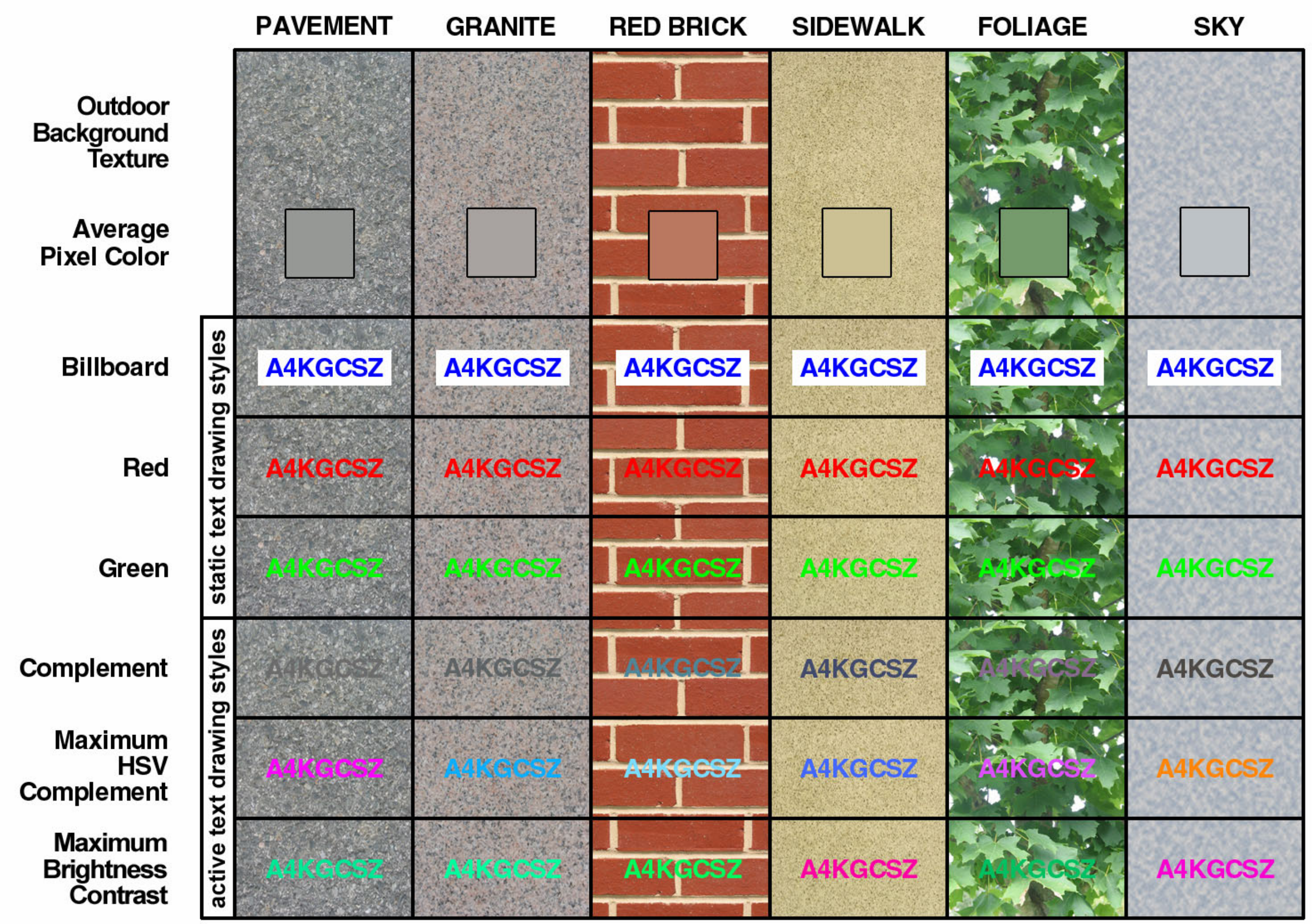


Figure 1. The six outdoor background textures, respective average pixel colors (shown in small box in each background), and six text drawing styles. When the text strings were perceived in an AR display in front of actual background posters, the perceptually experienced contrast was different than it appears in this image.

color value (Figure 1) of each outdoor background texture (calculated from the digital images), as an input to the algorithms described below. We used a sans serif font (Helvetica), and presented the text at a size that appeared approximately two inches tall at a distance of two meters. Text size did not vary during the experiment.

Billboard (static): We designed the 'billboard' style using a saturated blue text ([0, 0, 255] in RGB space) on a white ([255, 255, 255] in RGB space) rectangular background. The solid white background for 'billboard' nearly completely occluded the background texture, resulting in easily readable text regardless of the background. We chose a saturated white background both to fully occlude the background texture and because white is a commonly used background color for GUIs and print media. We chose blue text instead of black text (as most GUI-based reading studies use, or as often used in newspaper and other print media) because black is transparent in optical see-through displays.

Red (static), Green (static): Our choice of 'red' and 'green' as conditions was based on the physiological fact that cones in the human eye are most sensitive to red and green [17; 18]. Both colors were fully saturated in RGB space; 'red' was [255, 0, 0] and 'green' was [0, 255, 0].

Complement (active): We chose the 'complement' text drawing style based on color theory and graphic arts guidelines asserting that strong color contrast supports efficient text reading [19; 20]. The 'complement' style (also called 'inverse' in computer graphics) is defined in RGB space as shown in Equation 1, where (R, G, B) is the average pixel background color, and (R', G', B') is the resulting text drawing style color.

```
R' = ABS (R - 255)
G' = ABS (G - 255)          (1)
B' = ABS (B - 255)
```

A potential problem of using the complement in this setting is that some of our backgrounds' average pixel color fell into the mid-range of the RGB scale; that is, their respective R, G, and B values were closer to 128 than 0 or 255. As a result, the complement of these colors does not provide strong color contrast as compared to the source background, especially for the 'pavement' and 'granite' backgrounds. Even so, we wanted to study the complement, because it is prevalent in graphic arts and 2D GUIs.

Maximum HSV Complement (active): Our observations of problems with the 'complement' style motivated us to design this text drawing style. To achieve further contrast, we first designed a 'saturated complement' style to fully saturate the complementary color. However, pilot testing suggested that saturating the complement of our chosen backgrounds resulted in mostly dark (approaching black) text, which becomes increasingly difficult to read using optical see-through AR displays. We then refined the style with the following goals: retain the notion of employing color complements, account for the fact that optical see-through AR displays cannot present the color black, and use the HSV color model [21] so we could easily and independently modify saturation. The result is

$$\begin{aligned} H' &= (H + 180) - (((H + 180)\ \mathrm{DIV}\ 360) * 360) \\ S' &= |100\% - S| \qquad (2) \\ V' &= 100\%, \end{aligned}$$

where (H, S, V) is the average pixel background color in HSV space, and (H', S', V') is the resulting text drawing style color. This algorithm rotates H by 180 degrees, calculates the complement of S (defined as $|100\% - S|$), and maximizes V, the value (brightness) of the color.

Maximum Brightness Contrast (active): We wanted to create a text drawing style that maximized the perceived brightness contrast between the augmented text and the outdoor background texture. This style is based on MacIntyre's maximum luminance contrast technique [13; 14]. Like MacIntyre, our algorithm calculates within the *Commission Internationale de l' Éclairage* (CIE) XYZ color model [21], because the Y basis function models human brightness perception. To calculate this style, we had to convert our known average pixel colors from RGB space to CIE XYZ space. This process requires colorimeter equipment to physically measure the XYZ basis functions, which was impractical for our experimental setup[1]. Algorithmically converting between RGB colors and XYZ colors requires assuming values for two parameters: object size (measured in degrees of subtended field of view on the retina), and white point. The XYZ basis functions are only defined for 2° objects and 10° objects, and since 2° is much closer to the size of our text strings than 10°, we used the 2° standard. Because we ran our experiments outdoors with natural lighting, we used CIE standard illuminant D65 as our white point, since D65 represents sunlight. More details on these parameters and values are available in Wyszecki and Stiles [23], Foley et al. [21], and Gabbard [15].

Our algorithm for calculating the 'maximum brightness contrast' is described by Equation 3 below. It only manipulates the CIE Y value. Let (X, Y, Z) be the CIE values of the background's average pixel color, and (X', Y', Z') be the resulting text drawing style color. The algorithm maximizes Y' if Y is less than or equal to 0.5, otherwise it minimizes Y'. This maximization (minimization) is the value closest to 1.0 (0.0) subject to the constraint that each component of the resulting (R', G', B') tuple is in the valid range of 0 to 255.

$$\begin{aligned} X' &= X \\ Y' &= \begin{cases} \max\limits_{i \in [0..1]} i : \begin{cases} 0 \le R(X',Y',Z') \le 255 \\ 0 \le G(X',Y',Z') \le 255 \\ 0 \le B(X',Y',Z') \le 255 \end{cases} & \text{if } Y \le 0.5 \\ \min\limits_{i \in [0..1]} i : \begin{cases} 0 \le R(X',Y',Z') \le 255 \\ 0 \le G(X',Y',Z') \le 255 \\ 0 \le B(X',Y',Z') \le 255 \end{cases} & \text{if } Y > 0.5 \end{cases} \\ Z' &= Z \qquad (3) \end{aligned}$$

3.2.3 Other Independent Variables

As summarized in Table 1, we varied the distance from the user to the outdoor background texture (poster) between three different levels: 'near' (one meter), 'medium' (two meters), and 'far' (four meters). Our head-mounted display had an un-adjustable accommodative focus of two meters. In addition, users saw four repetitions of each combination of independent variables.

3.3 Dependent Variables

Also as summarized in Table 1, we collected values for three dependent variables: response time, errors, and ambient illuminance at the user's position. For each trial, our custom software recorded both the user's four-alternative forced choice (0, 4, 5, or 6), and the user's response time. Whenever the distance or outdoor background texture changed, we measured and recorded the illuminance at the user's position.

3.4 Experimental Design and Users

We used a factorial nesting of independent variables for our experimental design, which varied in the order they are listed in Table 1, from slowest (subject) to fastest (repetition). We collected a total of 7776 response times and errors (18 subjects * 3 distances * 6 outdoor background textures * 6 text drawing styles * 4 repetitions), and 324 illuminance measurements (18 subjects * 3 distances * 6 outdoor background textures). We counterbalanced the presentation of the independent variables using a combination of Latin squares and random permutations [15]. Each subject saw all levels of each independent variable, so all variables were within-subject.

Eighteen subjects participated, twelve males and six females, ranging in age from 20 to 31. All volunteered and received no compensation. We screened the subjects, via self-reporting, for color blindness and visual acuity. Subjects did not appear to have any difficulty learning the task or completing the experiment.

4 Hypothesis

Prior to conducting the study, we made the following hypotheses:

(1) Because the 'billboard' style obscures the background, it will result in the fastest and most accurate task performance.

(2) The 'green' and 'red' styles will result in fast and accurate performance, because users' eyes are most sensitive to the two colors.

(3) The 'maximum HSV complement' style will result in faster and more accurate task performance than the 'complement' style, since it takes into account how optical see-through displays present the color black.

---

[1] Borrowing the terminology of MacIntyle [13; 14], there are four relevant luminance levels in this problem domain: (1) the *drive luminance*, which is how much luminance the application program tells the display to produce, (2) the *emitted illuminance*, which is how much illuminance the display actually produces, (3) the *environmental luminance*, which is how much luminance is entering the front of the display, and (4) the *perceived luminance*, which we (and MacIntyre) define to be the amount of luminance that enters the user's eye. While the perceived luminance, which is a function of emitted luminance and environmental luminance, is the relevant quantity, in practice it is not possible to precisely measure or calculate it. The drive luminance is what is easy to control from within an application, and that is what we (and MacIntyle) did in this study.

However, we have learned (post-experiment) that there are techniques that may let us physically quantify the perceived luminance [22]. As mentioned in Section 7, we are investigating these techniques, with the goal of implementing MacIntyre's original luminance contrast metric [13; 14] in an optical see-through AR context.

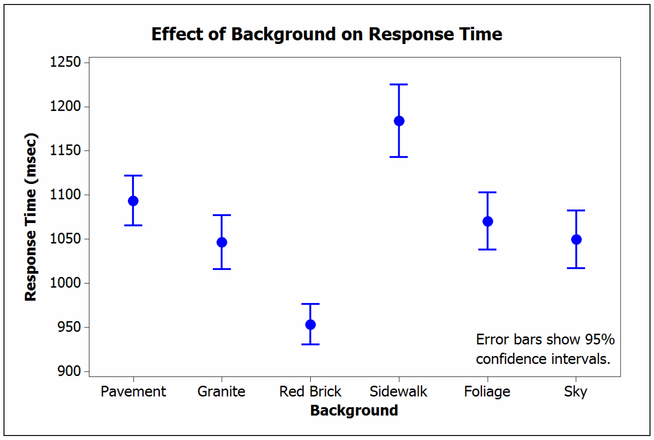


Figure 2. Effect of outdoor background texture on mean response time.

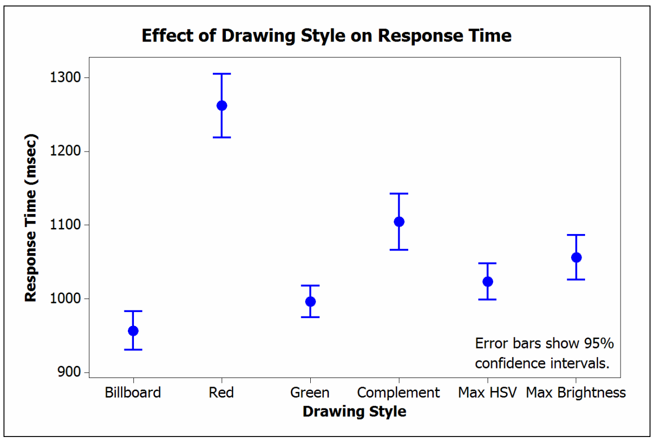


Figure 3. Effect of text drawing style on mean response time.

(4) The active styles will result in faster and more accurate task performance than the 'green' and 'red' styles, since they take the background color into account.

(5) The more visually complex outdoor background textures ('red brick' and 'foliage') will result in slower and less accurate task performance, since the complexity will interfere with the task.

(6) When the distance is 'medium' (and matches the display's accommodative focus), user performance will be faster and more accurate than with the distance is 'near' or 'far' (and does not match the display's accommodative focus).

5 Results

We analyzed our results with analysis of variance (ANOVA) and regression analysis. With ANOVA we modeled our experiment as a repeated-measures design that considers *subject* a random variable and all other independent variables as fixed (Table 1). When deciding which results to report, in addition to considering the p value, the standard measure of *effect significance*, we considered two different measures of *effect size*: eta-squared (η^2), and partial eta-squared ($p\eta^2$). Both are measures of how much variance is accounted for by an effect; η^2 is normalized across all of the experiment's independent variables, while $p\eta^2$ is normalized by each variable individually [24].

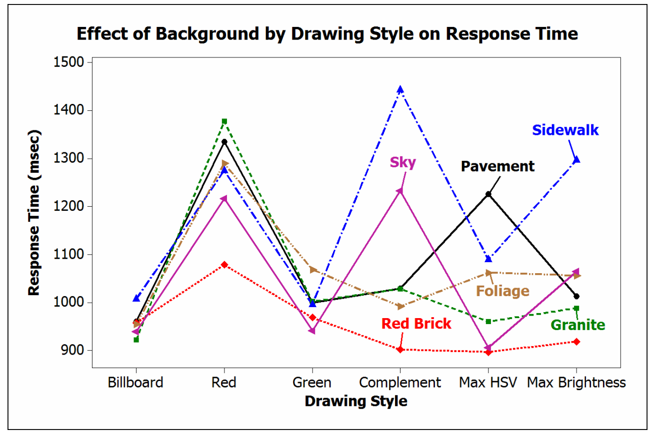


Figure 4. Effect of outdoor background texture by text drawing style interaction on mean response time.

5.1 Error Analysis

Out of 7776 total error measurements, there were 7628 correct responses, 137 incorrect responses, and 11 "target string was not visible" responses. Because the error rate was so small (~1.9%), we did not further analyze errors.

5.2 Response Time and Illuminance

Figure 2 shows, as expected, an effect of outdoor background texture on response time ($F(5,85) = 6.16$, $p < .001$, $\eta^2 = 1.36\%$, $p\eta^2 = 26.6\%$). Subjects performed fastest with the 'red brick' background, and they had comparable times for 'pavement', 'granite', 'foliage', and 'sky'. Subjects performed slowest with the 'sidewalk' background. Subjects' superior performance using the 'red brick' background may be explained by our observation that subjects adopted a strategy of moving their head slightly to center the text string within a single brick, and this framing coupled with the visual homogeneity of each brick may have enhanced legibility of the target string. These results refute hypothesis 5; the visually complex background textures performed very well ('red brick') and intermediately well ('foliage'). Clearly, factors other than visual complexity dominated background performance.

Figure 3 shows a main effect of text drawing style on response time ($F(5,85) = 19.86$, $p < .001$, $\eta^2 = 2.83\%$, $p\eta^2 = 53.9\%$). The 'billboard' drawing style supported the fastest performance, followed by 'green.' These results support hypothesis 1, and, to a lesser degree, hypothesis 2. Surprisingly, the 'red' drawing style gave the worst performance. For the active styles, 'maximum HSV complement' was significantly faster than 'complement', which supports hypothesis 3. We did not hypothesize about our 'maximum brightness contrast' style, and its performance overlapped the other active styles. These results do not support hypothesis 4: the 'green' style did at least as well as our best active style.

Figure 4 shows a response time interaction between background and text drawing style ($F(25,425) = 5.47$, $p < .001$, $\eta^2 = 2.09\%$, $p\eta^2 = 24.4\%$). Interestingly, the static styles gave much lower interaction effects than the active styles; performance varied widely depending on the combination of background and active style. Considering only the active styles, for some backgrounds ('red brick', 'granite', 'foliage') performance was relatively constant, while for the remaining backgrounds ('sky', 'pavement', 'sidewalk'), it varied considerably according to text

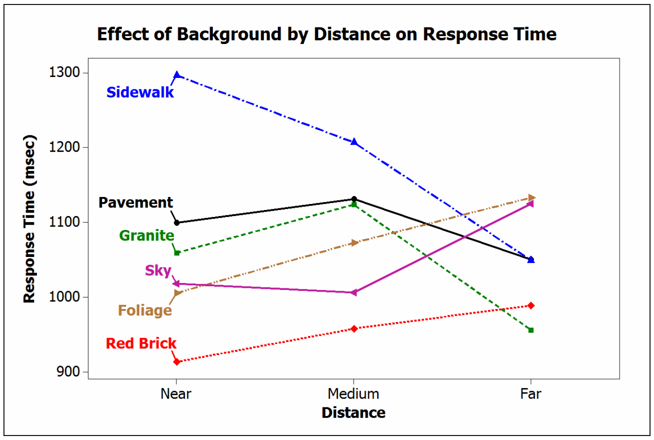


Figure 5. Effect of outdoor background texture by distance interaction on mean response time.

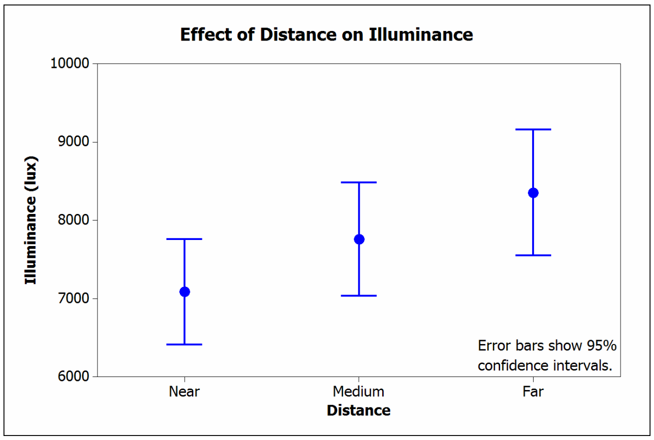


Figure 6. Effect of distance on mean illuminance.

drawing style. These results indicate that 'billboard' and 'green' were the only globally effective text drawing styles.

There was no main effect of distance ($F(2,34) < 1$), contradicting hypothesis 6. We believe this result is explained by two aspects of our task: (1) it may not require a sharp accommodative focus to read a number in a distracting text string, and (2) because our task did not require subjects to attend to the background, they may have ignored the background and just focused on the text string. But as Figure 5 shows, there was a response time interaction between background and distance ($F(10,170) = 3.59$, $p < .001$, $\eta^2 = 1.05\%$, $p\eta^2 = 17.42\%$). At the 'near' distance, background had a much greater effect on response time, both positive ('red brick') and negative ('sidewalk'). The effect of background was mitigated as the distance increased to 'medium' and then to 'far'. However, the lack of a main distance effect, combined with the reasoning above, leads us to suspect that this interaction is caused by the closer distances making the texture features more salient. At least it seems clear that the interaction is not caused by accommodative match ('medium') or mismatch ('near', 'far').

We found a main effect of repetition on response time ($F(3,51) = 25.11$, $p < .001$, $\eta^2 = .76\%$, $p\eta^2 = 59.6\%$). Subjects showed a standard learning effect: their response times dropped by 117.6 msec between the first and second repetition, and thereafter remained constant (varying by less than 3 msec). We also found a three-way response time interaction, which we do not consider because it has both low η^2 and $p\eta^2$, and does not mask any lower-order interactions. Finally, for the same reasons, we

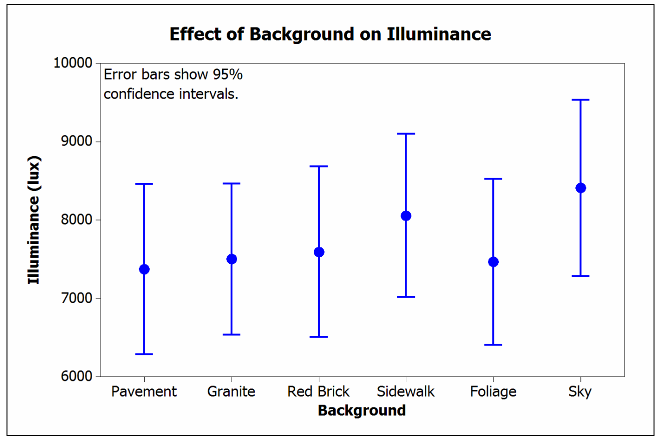
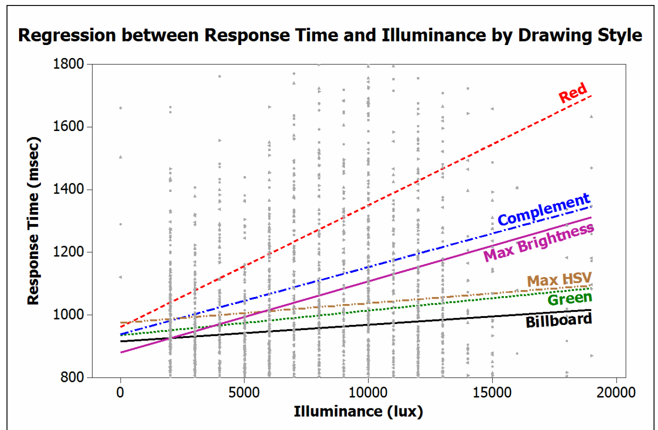


Figure 7. Effect of outdoor background texture on mean illuminance.



Billboard	$r^2 = 0.49\%$	$t(322) = 1.25$	$p = .211$
Red	$r^2 = 7.31\%$	$t(322) = 5.04$	$p < .000$ **
Green	$r^2 = 1.35\%$	$t(322) = 2.10$	$p = .0364$ *
Complement	$r^2 = 2.70\%$	$t(322) = 2.99$	$p = .003$ **
Max HSV	$r^2 = 0.63\%$	$t(322) = 1.42$	$p = .156$
Max Brightness	$r^2 = 4.95\%$	$t(322) = 4.09$	$p < .000$ **

Figure 8. Regression between response time and illuminance, grouped by text drawing style. '*' and '**' indicate significant regressions.

found but do not consider a nearly-significant two-way response time interaction.

Figure 6 shows that distance had a main effect on illuminance ($F(2,34) = 5.71$, $p = .007$, $\eta^2 = 1.79\%$, $p\eta^2 = 25.2\%$). The closer the subject was to the background posters, the less illuminance reached the subject's position. This can be explained by (1) the spatial layout of the experimental setup, time of day, and hence sun angle when we conducted the experiment, and (2) the coating on the greenhouse roof, which diffused and softened entering sunlight. The combination of geometry and diffuse sunlight resulted in the background posters casting subtle, diffused shadows in the direction of the subject.

Figure 7 shows that background had a main effect on illuminance ($F(5,85) = 2.16$, $p = 0.066$, $\eta^2 = .93\%$, $p\eta^2 = 11.3\%$). The different backgrounds reflected different amounts of light to the subject, and the brightest backgrounds ('sidewalk', 'sky') reflected the most light. Because the posters had matte surfaces, and entering sunlight was soft and diffused, the reflected light was also soft and diffuse.

Figure 8 shows that the amount of ambient illuminance at the subject's position had an effect on response time that depended on the text drawing style. Because we sampled the illuminance 324 times and the response time 7776 times, we calculated the regression in Figure 8 with a 1944-line data set, where response times are averaged over 4 repetitions, and illuminance readings are 0-order interpolated over 6 drawing styles. We considered higher order interpolations, but because both distance (Figure 6) and background (Figure 7) affected illuminance, we decided it was more accurate to model illuminance with 1st-order discontinuities at boundaries where distance and / or background change.

Figure 8 shows that subjects performed faster under the condition of less illuminance. This result can be explained by the fact that brighter illuminance tends to wash out the AR display, reducing the contrast between augmenting text and background texture. However, the strength of this effect depends on the text drawing style. In order of decreasing slope, 'red', 'maximum brightness contrast', and 'complement' showed significant ($p < .005$) regressions, while 'green' was significant at a weaker ($p < .05$) level. While 'maximum HSV complement' and 'billboard' have slight positive slopes, the *t*-test does not indicate that the regressions are different from 0. The effect of Figure 8 is why the amount of illumination reflected to the subject (Figures 6 and 7) is important.

6 Discussion

Our most important finding, which is not surprising, is clear empirical evidence that user performance on a task, which we believe is representative of a wide variety of imagined and realized AR applications, is significantly affected by background texture (Figure 2), text drawing style (Figure 3), and their interaction (Figure 4). Furthermore, the background affected the amount of ambient illuminance at the user's position (Figure 7), and the combination of this illuminance and text drawing style also affected user performance (Figure 8).

In terms of design guidelines, the current study suggests using the 'billboard' and 'green' text drawing styles, and avoiding the 'red' style (Figures 4 and 8). However, the 'billboard' style is likely effective because the solid white background of the text string obscures the real-world background, an effect that would be detrimental to many AR applications. Therefore, the main design guideline findings are evidence for the global effectiveness of fully-saturated green labels, and the global ineffectiveness of fully-saturated red labels. Interestingly, several currently manufactured monochromatic AR displays use red.

We are somewhat surprised that our active text drawing styles did not perform better relative to the static styles, refuting hypothesis 4. Nevertheless, 'maximum HSV complement', our active style that took the nature of optical see-through AR displays into account, did perform better than the 'complement' style (hypothesis 3; Figures 3 and 8). We still believe that the right active styles will result in better performance than static text drawing styles. In the current study, we only actively manipulated color, and only according to the averaged pixel color of the entire background texture. Figure 8 indicates that ambient illuminance can also affect text string reading performance. As discussed below, in the future we plan to design active styles that react to both ambient illuminance as well as to more precise measurements of background color.

Like most controlled user studies, this one had many limitations that restrict the generality of our findings. All of the limitations listed here suggest future text drawing style implementations and subsequent user-based studies.

- For the reasons discussed above, we printed our outdoor background textures onto matte posterboards. The reflective properties of the matte surface are of course different than a real surface; e.g. imagine an actual brick wall as opposed to a photograph of one.
- Furthermore, our background textures were two dimensional; many textures, such as foliage, have large depth variation, which would likely affect the results.
- Although we tested six qualitatively very different textures, we still managed to test only a small sample of the hundreds (thousands?) of possible urban textures. Furthermore, the appearance of even our tested textures varies widely with differing illumination. It may be possible to use different texture analysis techniques to systematically select or generate a more comprehensive texture set.
- Like all optical see-through AR user-based studies to date, perhaps our greatest limitation is the capabilities of the display itself. Among the serious limitations of our Glasstron display is that it does not support true occlusion and cannot display dark colors, it's shifting head-fit makes precise alignment between augmentations and real-world objects difficult, and, like all common computer displays, its dynamic range does not come close to the eight orders of magnitude of outdoor illuminance variance [16].
- Although our experiment examined outdoor illuminance values, we only sampled a fraction of the available outdoor dynamic range, which ranges from a starlit landscape to direct noon sunshine [16].
- Finally, as discussed above, our task did not require the user to integrate augmented information with real-world objects, but many potential AR tasks (such as product maintenance [4]) would require this type of integration.

In summary, what is more important than the particular effects we found is the empirical confirmation that user performance for text legibility, closely related to the fundamental AR task of reading text, was strongly affected by text drawing style, background texture, and their interaction, which motivates further research.

7 Concurrent and Future Work

This work is the beginning of a series of research efforts designed to increase legibility in outdoor AR user interfaces. At the current time, we have conducted, but have not yet reported on, a study which involves a text reading task that more tightly integrates augmented and real-world information, further studies the effects of accommodative demand, and utilizes a Microvision laser-scanning optical see-through AR display [25]. This laser scanning technology could potentially match the full dynamic range of outdoor illuminance.

In addition, we are concurrently running two empirical studies, and have several additional user-based studies planned. First, we are studying the effects of text/graphics drawing styles, environmental lighting, and real-world backgrounds on user task performance in outdoor AR tasks, using physical real-world objects, rather than posters, and more sophisticated text drawing styles. Second, we are further exploring design styles, such as altering the brightness of augmentations, without changing their fundamental color and thus preserving color encoding, and dynamically altering the opaqueness of a background rectangle.

In the future, we intend to examine other potential active text drawing styles, such as a more accurate maximum luminance contrast style as it was originally designed [22], the anti-interference font [7], as well as other text drawing styles taken from graphics arts and the television and movie industries, such as drop shadows, halos, and so forth. This work will help identify static text rendering styles that not only preserve color-coding, but are flexible and robust enough to use in varying outdoor conditions. This includes developing methods to quantify the luminance of real-world objects and augmentations within the AR display, and using these methods to better measure the actual luminance contrast between augmentations and real-world objects.

Acknowledgements

BARS is supported by the Office of Naval Research (ONR) and the Naval Research Lab, and our usability engineering research is also supported by ONR under Program Manager Dr. Larry Rosenblum.

References

[1] Swan, JE II and Gabbard, JL, *Perceptual and Ergonomic Issues in Mobile Augmented Reality for Urban Operations*, Naval Research Laboratory Technical Memorandum Report, 2003.

[2] Gabbard, JL, Swan II, JE, Hix, D, Lanzagorta, M, Livingston, MA, Brown, D, Julier, S, "Usability Engineering: Domain Analysis Activities for Augmented Reality Systems", *The Engineering Reality of Virtual Reality 2002*, A Woods, J Merritt, S Benton, M Bolas, Editors, Proceedings SPIE Volume 4660, Stereoscopic Displays and Virtual Reality Systems IX, 2002, pages 445–457.

[3] Livingston, MA, Rosenblum, L, Julier, SJ, Brown, D, Baillot, Y, Swan, J.E. II, Gabbard, JL, Hix, D, "An Augmented Reality System for Military Operations in Urban Terrain", In *Proceedings of the Interservice / Industry Training, Simulation, & Education Conference (I/ITSEC '02)*, Orlando, FL, December 2–5, 2002.

[4] Lipson, H, Shpitalni, M, Kimura, F, Goncharenko, I, "Online Product Maintenance by Web-Based Augmented Reality", In *New Tools and Workflows for Product Development*, Berlin, 1998, pages 131–143.

[5] Piekarski, W, Thomas, B, "ARQuake: The Outdoor Augmented Reality Gaming System", *Communications of the ACM,* Volume 45, Issue 1, 2002, pages 36–38.

[6] Thomas, B, Close, B, Donoghue, J, Squires, J, De Bondi, P, Piekarski, W, "First Person Indoor/Outdoor Augmented Reality Application: ARQuake", *Personal and Ubiquitous Computing*, Volume 6, 2002, pages 139–146.

[7] Harrison BL, Vicente KJ, "An Experimental Evaluation of Transparent Menu Usage", In *Proceedings CHI '96*, 1996, pages 391–398.

[8] Christ, RE, "Review and Analysis of Color Coding Research for Visual Displays", *Human Factors*, Volume 17, Issue 6, 1975, pages 542–570.

[9] Cleveland, WS, McGill, R, "Graphical Perception and Graphical Methods for Analyzing Scientific Data", *Science*, Issue 229, 1985, pages 828–833.

[10] Cahill, MC, Carter, RC, "Color Code Size for Searching Displays of Different Density", *Human Factors*, Volume 18, Issue 3, 1976, pages 273–280.

[11] Carter, RC, "Visual Search with Color", *Journal of Experimental Psychology: Human Perception and Performance*, Volume 8, Issue 1, 1982, pages 127–136.

[12] Smallman, HS, Boynton, RM, "Segregation of Basic Colors in an Information Display", *Journal of the Optical Society of America*, Volume 7, Issue 102, 1990, pages 1985–1994.

[13] MacIntyre, B, *A Constraint-Based Approach To Dynamic Colour Management For Windowing Interfaces,* Master's thesis, University of Waterloo, Available as Department of Computer Science Research Report CS-91-55, 1991.

[14] MacIntyre, B, Cowan, W, "A Practical Approach to Calculating Luminance Contrast on a CRT", *ACM Transactions on Graphics*, Volume 11, Issue 4, 1992, pages 336–347.

[15] Gabbard, JL, *Real-world Scenes, Objects, and Textures as a GUI Background: Designing Effective Text Styles for Outdoor Augmented Reality*, PhD Dissertation Preliminary Report, Virginia Tech, 2003.

[16] Halsted, CP, "Brightness, Luminance and Confusion", *Information display*, March, 1993.

[17] Hecht, E, *Optics (2nd edition)*, Addison Wesley, 1987.

[18] Williamson, SJ, Cummins, HZ, *Light and Color in Nature and Art*, Wiley and Sons, NY, 1983.

[19] Legge, G, Parish, D, Leubker, A, Wurm, L, "Psychophysics of Reading XI: Comparing color contrast and luminance contrast", *Journal of the Optical Society of America*, Volume 7, 1990, pages 2002–2010.

[20] Knoblauch, K, Arditi, A, "Choosing Color Contrasts in Low Vision: Practical Recommendations", In AC Kooijman, PL Looijestijn, JA Welling, GJ van der Wildt (Eds.), *Low vision: Research and new developments in rehabilitation. Amsterdam*, The Netherlands, IOS Press, 1994, pages 199–203.

[21] Foley, JD, van Dam, A, Feiner, SK, Hughes, JF, Phillips, RL, *Introduction to Computer Graphics (2nd edition)*, Reading, MA, Addison-Wesley, 1993.

[22] MacIntyre, B, Personnel Communication, Autumn 2003.

[23] Wyszecki, G, Stiles, WS, *Color Science – Concepts and Methods, Quantitative Data and Formula*, John Wiley, NY, 1982.

[24] Cohen, J, "Eta-Squared and Partial Eta-Squared in Fixed Factor ANOVA Designs", *Educational and Psychological Measurement*, Volume 33, 1973, pages 107–112.

[25] Gupta, D, *An Empirical Study of the Effects of Context-Switch, Object Distance, and Focus Depth on Human Performance in Augmented Reality*, Masters Thesis in Industrial and Systems Engineering, Virginia Tech, Spring 2004.