International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2020)
F. Argelaguet, M. Sugimoto, and R. McMahan (Editors)

A Review of Visual Perception Research in Optical See-Through Augmented Reality

Austin Erickson[†], Kangsoo Kim, Gerd Bruder, Gregory F. Welch

University of Central Florida
Orlando, Florida, United States

Abstract

In the field of augmented reality (AR), many applications involve user interfaces (UIs) that overlay visual information over the user's view of their physical environment, e.g., as text, images, or three-dimensional scene elements. In this scope, optical see-through head-mounted displays (OST-HMDs) are particularly interesting as they typically use an additive light model, which denotes that the perception of the displayed virtual imagery is a composite of the lighting conditions of one's environment, the coloration of the objects that make up the virtual imagery, and the coloration of physical objects that lay behind them.

While a large body of literature focused on investigating the visual perception of UI elements in immersive and flat panel displays, comparatively less effort has been spent on OST-HMDs. Due to the unique visual effects with OST-HMDs, we believe that it is important to review the field to understand the perceptual challenges, research trends, and future directions.

In this paper, we present a systematic survey of literature based on the IEEE and ACM digital libraries, which explores users' perception of displaying text-based information on an OST-HMD, and aim to provide relevant design suggestions based on the meta-analysis results.

We carefully review 14 key papers relevant to the visual perception research in OST-HMDs with UI elements, and present the current state of the research field, associated trends, noticeable research gaps in the literature, and recommendations for potential future research in this domain.

CCS Concepts

• ***Human-centered computing*** → ***Mixed / augmented reality;*** • ***General and reference*** → ***Surveys and overviews;***

1. Introduction

While augmented reality (AR), which seamlessly superimposes virtual content over the user's real world view, is experiencing dramatic technological advances and unprecedented public interest [KBB*18, WBSS19], optical see-through head-mounted displays (OST-HMDs) for AR, such as the Microsoft HoloLens and Magic Leap One, are becoming more and more popular and readily available for both enterprise and personal uses [Azu17]. Although a variety of display hardware and visualization technologies have been introduced and received attention from AR researchers who investigated the efficacy and efficiency of the perception of virtual content over the last decade [KSF10, HRL*16], visual perception through an OST-HMD is particularly interesting because of its typical use of an *additive light model*—which can only add light; the displayed image is the sum of the projected imagery, environmental light and background appearance [ILI*19].

When designing user interfaces (UIs) with visual elements for OST-HMDs, it is a common practice to follow the existing rules and principles, which are based on other traditional displays, e.g., flat-panel displays or immersive virtual reality (VR) HMDs equipped with near-eye opaque screens. However, previous literature has demonstrated that the effects that UI elements can have on factors such as user performance, visual acuity, visual fatigue, and subjective preference, can be vastly different between OST-HMDs and other display mediums. Due to their additive light model, OST-HMDs are comparatively more sensitive to environment light and easily lose the contrast of the perceived imagery [EKBW20]. Additionally, displaying UIs in color further causes a loss of contrast and makes virtual imagery difficult to observe. Because of these difficulties, there are many factors that must be carefully considered when displaying virtual imagery on an OST-HMD, such as the virtual UI foreground/background colors, the physical lighting conditions of the user's environment, and the capabilities of the device being used to present the virtual information.

Several methods of optimizing the appearance of the virtual imagery on OST-HMDs have been proposed for better visual perception and user experience [KBCW03, MF13, IK15]; however, to the best of our knowledge there has not been a recent survey of the research literature that could provide comprehensive knowledge and

† ericksona@knights.ucf.edu

insights in this focused research area, despite the timely importance and practical benefits for UI design on OST-HMDs. In this paper, we present a systematic survey of the key related literature, which has been collected from the IEEE and ACM digital libraries, as our initial effort to cover the visual perception research on OST-HMDs. By employing these two libraries in the domains of computer science and engineering, we aim at establishing a knowledge base for AR research focused on technological approaches to overcome the perception issues with OST displays while studying the effects.

In this survey, we aim to identify which research areas have been well covered or underrepresented in visual perception research with OST-HMDs, while understanding the current state of the research and unique challenges or opportunities in visualization techniques and UI design on OST-HMDs. We draw suggestions for designing effective UIs on OST-HMDs and future research directions, which we hope to be helpful for researchers, practitioners, and students to start or develop more impactful applications and research.

The remainder of this paper is structured as follows: Section 2 provides background information on current challenges with OST-HMDs and visual perception research. Section 3 describes the methodology of our systematic review process, and Section 4 presents a meta-analysis of how the literature is distributed in terms of the research topics. Section 5 summarizes the key points based on our in-depth reviews of the selected papers. Section 6 provides a higher level discussion of the trends and gaps discovered through analysis of the retrieved papers, and provides recommendations on how to proceed with future research. Finally, Section 7 concludes the paper.

2. Background

In this section, we cover some of the main factors and challenges at the intersection of OST displays and visual perception research.

Vergence-Accommodation Conflict When viewing an object at a close distance, the user's pupils slightly constrict and their eyes accommodate to the distance of the target object. At the same time, the muscles surrounding the eyes cause them to converge inward such that rays emanating from each eye would intersect at the focal depth. Collectively, this process is known as the *near triad* [MS90]. The near triad is of particular importance when dealing with near-eye displays, e.g., OST-HMDs, because the optics of the display are typically set to display a clear image at a single focal depth, such as two meters away from the user [SES20]. Because of this, the user must accommodate to clearly view physical objects in their environment which are placed at different depths, which has been shown to have effects on visual fatigue and user performance [GMS19]. This also introduces an effect known as vergence-accommodation conflict (VAC) when virtual imagery is presented at depths other than the focal depth of the display [Kra16]. VAC has been shown to cause eye fatigue, and has been shown to have other interesting effects on perception, such as causing errors in distance estimations [HGAB08, BASOL16, JM19].

Environment Lighting OST displays are largely affected by environment lighting due to typically incorporating an additive light model. It is commonly known that using these types of displays in bright outdoor environments introduces a "washing out" effect that occurs as imagery appears more transparent and loses contrast [EKBW20]. Because of this, OST-HMDs typically have a tinted visor so that users can still view the virtual content in bright conditions, however this also comes at a cost by decreasing the luminance contrast of the user's physical environment. As light affects the imagery on the OST display, it also impacts the user's physiology by causing pupil dilation or constriction in the user's eyes, which has effects on the user's visual acuity and contrast sensitivity [CG60, SAW99]. Thus, lighting conditions are one of the most important factors to investigate for research involving visual perception on OST-HMDs.

Color Blending Considering the characteristics of light-additive OST-HMDs, rendering perceptually accurate colors on the display has been shown to be a challenging problem [GSZW10, HRISI15, MSG*16]. This is due to a phenomena known as *color blending*, where the intended color of virtual imagery is impacted by the coloration of the user's physical environment which is appearing behind it. Understanding how the the color of the light being emitted from the display mixes with the colors of the physical environment is critical for determining which colors to use when displaying UI elements on the OST device, especially since colors are commonly used to encode information.

This difficulty of OST displays is further complicated when user demographics are taken into account [SES20], as when it comes to the user's color perception ability, deficiencies in color perception are fairly common—red-green deficiency is thought to occur in 1 out of 12 males and 1 out of 200 females of Northern European descent—and affects the person's ability to distinguish between certain shades of red, yellow, and green [Dee05]. Other color deficiencies are relatively less common, such as blue-yellow deficiency, which occurs is less than 1 in 10,000 people [Dee05]. Because of the frequency with which color deficiencies occur, it is important when studying the best practices for UI design on OST-HMDs to evaluate whether or not study participants are affected by such a disorder. This is easily confirmed via use of a color vision test, such as the Ishihara color test [I*18].

3. Search Methodology

Our literature survey focused on two popular venues for AR related research: the IEEE and ACM digital libraries. These two libraries are largely dominated by computer science and engineering related papers, and should serve as a good starting point for our survey of OST displays and visual perception.

Due to the narrow focus of this survey, and domain-specific terminology, we had difficulties crafting a set of search terms that would return mainly relevant research, as certain terms such as "optical," while commonly mentioned in the domain of OST-HMDs, are also mentioned in many other fields. Because of these difficulties, we decided to "cast a wide net" by purposely including more general search terms. We concentrated on a survey of work that involved OST-HMDs while also containing user studies involving visual perception. Such user studies should directly involve the visual perception of the user, such as by measuring visual acuity, contrast sensitivity, color perception, or visual fatigue using virtual stimuli, physical stimuli or a combination of the two. This lead us to the following expression of search terms:

- *("Augmented Reality" OR Optical OR "Mixed Reality") AND*
- *("User Interfaces" OR "User Interface" OR UI OR Text OR Font) AND*
- *(Light OR Light* OR Color OR Color* OR Luminance OR Illuminance OR Acuity OR Fatigue OR Contrast)*

The search was limited to papers which included these terms within the abstract of the paper. Asterisk characters are used to depict terms that included a wildcard appended to the search term, which may include more results than searching the word on its own, such as coloration and coloring when color* is searched.

We searched for these terms using the advanced search feature on both the IEEE and ACM digital library websites, where it returned 297 results in the IEEE library and 50 results in the ACM library. The majority of these results were unrelated to our interests (e.g., the inclusion of the term "optical" returned many unrelated works in optics and photonics). Additionally, we were not interested in work which focused solely on video see-through (VST) AR, virtual reality (VR), mobile AR, or handheld AR, so all results pertaining to these topics were removed. Our search included one result in which the study took place on an in-vehicle heads-up display (HUD), and due to the similarity of methods and applications between it and the other works, we decided to include it among our results. After removing all unrelated papers, 16 papers remained that were directly relevant to our interests. As a final step, the remaining papers were examined in detail where we found that one was a duplicate, where a conference paper was extended into a journal paper. For this paper, the journal version was kept and the conference paper was removed due to the added content available in the journal version. A similar case was found where one doctoral consortium paper was returned in the search along with the conference paper it referenced. Similar to above, this doctoral consortium paper was removed and the conference paper was kept. As a result of the above pruning, we were left with 14 total papers. A graphical depiction of the above-mentioned pruning process is shown in figure 1.

4. Meta Analysis

After collecting the above mentioned papers, we proceeded to tag them with relevant keywords based on the research focus of the paper, the study methodology, the hardware used, and the application area. A complete list of the keywords along with the number of papers tagged for each word is shown in Figure 2.

These keywords can be roughly binned into eight different categories, which are shown below and are grouped by color in Figure 2. There was no limit to the number of tags which could be applied per paper, and so there is some overlap between each of these categories, for example many papers that examined text enhancements also examined color. It is also possible that the total number of tags for a category may exceed the total number of papers. This occurs if one or more papers investigate several tagged aspects that are within the same category. The distributions of tagged papers within each of these categories are described below.

Text Enhancements There were a total of 12 tags that were focused on evaluating enhancements applied directly to text-based virtual content on OST-HMDs. These enhancements affected the text in a variety of different manners and consisted of *billboards*, where text was superimposed over a solid colored rectangular region, *text outlines*, where an outline was drawn around the text in a different color, and *text drop-shadows*, where a projected image of the text appeared beneath the foreground text in a different color. It is interesting to note that eight papers investigated billboard style text enhancements, while three investigated text outlines, and only one investigated drop-shadow. These numbers indicate that there is further room for exploration in the domain of text enhancements in OST AR, including further research on the under-explored outline and drop-shadow style enhancements as well as creative new methods of enhancing the appearance of virtual text.

Lighting and Contrast A total of 19 tags were applied to papers that focused in some manner on effects of environment lighting and contrast on the users' perceptions of the virtual content. Of these papers, six focused on effects of environment lighting on the users' perception by varying the intensity of lighting in the study environment. Two measured contrast sensitivity of the users through use of contrast testing techniques such as the Pelli-Robson chart and sine wave grating test. Three investigated effects of contrast polarity by measuring differences in users' ability and perception between UIs with a darker foreground on a lighter background compared to the opposite with a lighter foreground on a darker background. Finally, eight papers investigated the effects of non color-related background appearance on user ability and perception.

Color There were 23 tags that focused on investigating the impacts of UI color choices on the user. Of these, ten papers focused on the text or foreground color of virtual imagery and seven papers focused on the effects of the color of the virtual or physical back-

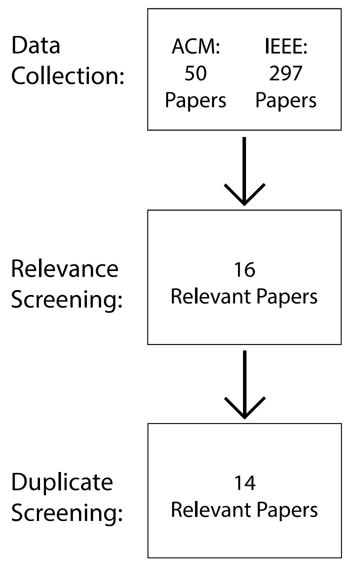

Figure 1: *This figure depicts the pruning process of returned papers from the search query performed on the IEEE and ACM digital libraries.*

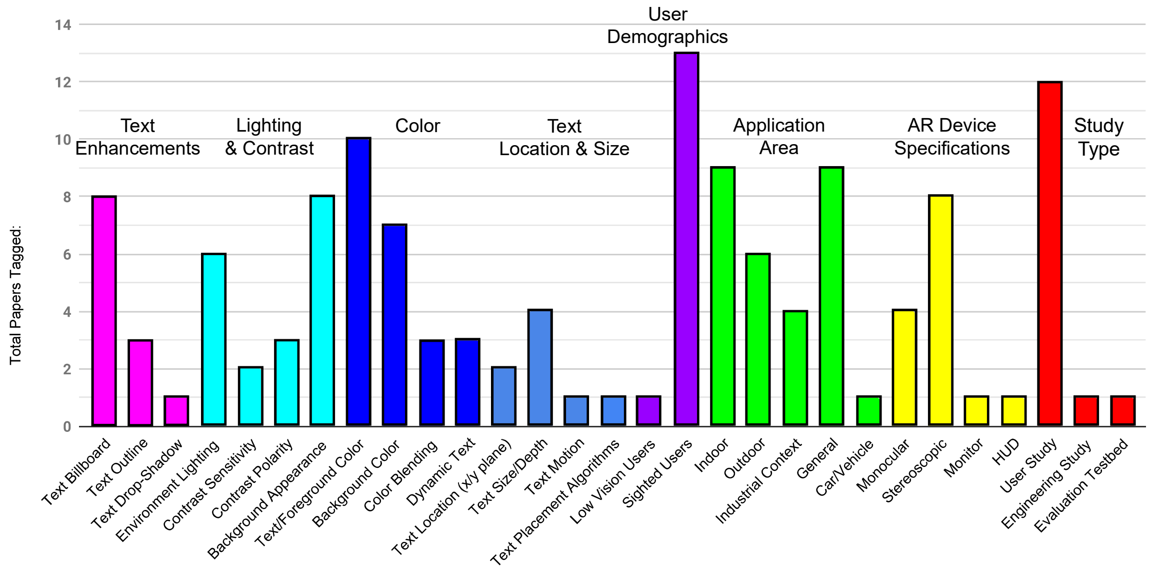

Figure 2: *This figure depicts the number of papers that were tagged for each keyword. The keywords are grouped into colored categories as described in Section 4.*

ground behind the virtual content. Three papers focused on color blending, by investigating the differences between the color which was intended to be displayed on the device and the resulting color from the users' perspective, which is influenced by the colors in the physical environment. Finally, three papers focused on dynamic enhancements of the color of the virtual foreground or background based on changes to the users' head positions or changes to the users' physical environments. Overall, this category contained the highest amount of tagged papers when compared to any of the other categories. This is likely due to several factors, including the additive nature of OST displays, which makes it difficult to predict how a developer-selected color will appear to the user due to variations in their physical lighting conditions. The majority of the papers that were tagged in this category also had tags in the text enhancement category, the lighting and contrast category, or both.

Text Location and Size There were a total of eight tags that investigated the effects of text location or text size on the user. Of these, two focused on the X/Y position of the virtual content (i.e., the OST-HMD's screen space position without regard to depth from the user), four focused on the depth at which the virtual content is displayed or the relative size of the virtual content, one investigated text motion to make it more noticeable, and one focused on text placement algorithms specifically.

User Demographics 13 out of the 14 papers in the survey incorporated user studies in which users were selectively chosen based on strict exclusion criteria. The exclusion criteria typically consisted of being capable of 20/25 or better visual acuity (with corrective lenses), being right eye dominant (for monocular displays only), having normal color vision (typically tested through the use of an Ishihara color plate book [I*18]), and being under the age of 35 years old to ensure that users were capable of visual accommodation. In Figure 2, these papers are tagged as "sighted users".

One paper, however, focused on the potential use of OST-HMDs for enhancing the visual capabilities of low-vision users [ZHHA17]. In this paper, the user study is of a between-subjects design, where a group of low-vision users with visual acuity worse than or equal to 20/100, was compared with a group of sighted users with an unlisted exclusion criteria. This distribution highlights the need for further research on specific demographics groups such as older age users and users with low vision.

Application Area As shown in Figure 2, there was a fairly even split between papers that focused on indoor contexts (eight papers) versus outdoor contexts (five papers). Only one paper focused on both indoor and outdoor usage, although this was a special case as the paper entailed an evaluation of a text placement algorithm that utilized images of both indoor and outdoor scenes. The distribution in this category is somewhat to be expected, as while arguably the majority of AR applications take place indoors, there are difficulties in presenting virtual content on OST-HMDs in outdoor settings due to the increase in environment illuminance, causing imagery to appear faint and difficult to distinguish [EKBW20]. It is likely because of this phenomena that a good portion of authors chose to investigate that particular context.

There was a surprising number of papers that focused specifically on industrial context applications of AR (four papers), which suggests that there is great interest in solving visual perception issues with current OST AR technology in this domain, although three of these papers were from the same group of authors. It should be noted that the results and conclusions of these industry-specific papers are not isolated to that domain, and can be easily transferred to other, more general, OST AR contexts. The majority of papers focused on more general uses of the technology and did not specify an application area (nine papers). One paper focused specifically on OST heads-up display based AR for in-vehicle contexts.

AR Device Specifications As expected, due to the increase in availability of stereoscopic OST-HMDs over the last decade, the majority of studies in the survey utilized these devices for their user studies (eight papers). Four papers used monocular OST-HMDs, however three of these papers seemed to have chosen this type of display due to limited access to commercial stereoscopic OST-HMDs at the time of publication (2013, 2014, and 2015), and offered no other justification of their HMD selection [FDUM13, DFG*14, GUFM15]. The remaining paper justified their selection of a monocular OST-HMD due to the availability of a slider-based adjustment of focal depth, which was incorporated as a crucial element of their study design [GMS19]. One paper utilized a flat-panel computer monitor in their user study, although this paper's research focus was on producing an algorithm for text placement based on the color and lighting conditions of the physical environment [OKT13]. This monitor-based user study was designed as an initial evaluation of the differences between the text positions selected by the placement algorithm and by the users, and though the results of their evaluation may have been different if performed with users wearing an OST-HMD, the authors point this out explicitly as a limitation of their study.

Study Type An interesting observation we made is that not all papers in the survey performed a human-subject study to evaluate their hypotheses. Although running a study with human participants, the appropriate hardware, and in the appropriate context is the only sure way to understand how a system will perform, this is not always possible due to limitations of current technology, limited access to participants, and extenuating circumstances such as the COVID-19 pandemic. In particular, one paper chose to perform what they deemed an engineering study, which consisted of taking repeated measurements using a colorimeter in order to objectively measure how colors presented on the OST-HMD mix with the colors of the physical environment [GSZW10].

5. State of the Field

In this section we provide a more detailed summary of the work that went on in the 14 papers in our survey as well as discuss their similarities, differences, and major findings. Subsections are used to order the papers by research focus.

5.1. Text Enhancements and Contrast Polarity

As mentioned in Section 4, several of the papers in this survey investigated different forms of text enhancements, including billboards, text outlines, and text drop-shadows.

Gabbard et al. compared billboard style text enhancements, with blue text on white billboards, to red and green colored text without billboards as well as text with dynamic color enhancement algorithms [GSH06]. Their results showed that the text enhancements had a significant main effect on participant response time in their study task, with billboard style text enhancements outperforming both the red and green plain text conditions as well as text with dynamic color enhancements. This seems to indicate that billboards are the optimal way to convey information to users of OST-HMDs, while Kim et al. [KEL*19] showed that this is not always the case.

Kim et al. also examined billboard style text enhancements in their work, however their work focused on *contrast polarity*, and compared white colored billboards with transparent text to transparent billboards with white colored text [KEL*19]. While there were no other comparisons made between other types of text enhancements, their work showed that white colored text on transparent billboards outperformed the opposite configuration in terms of user response time and visual acuity in all physical lighting conditions and physical background conditions tested. Their result suggests that *negative* contrast UIs (light text on dark/transparent backgrounds) also known as *dark mode* should be employed over *positive* contrast (dark/transparent text over light colored backgrounds) also known as *light mode*. This is also supported by subjective feedback obtained by the participants from the study by Zhao et al. where they noted that all low-vision users preferred dark colored backgrounds to light colored backgrounds [ZHHA17].

These results directly contradict the results obtained by Fiorentino et al. in 2013, where they showed that black/transparent text on a white billboard outperformed all other color combinations in terms of participant response time, including plain white text with a black/transparent billboard [FDUM13].

This was reexamined by Debernardis et al. in 2014 by performing a study in which 20 different color combinations were evaluated for text/billboard combinations, an amount of combinations which far exceeded those in the above-mentioned works [DFG*14]. Their results suggested that white text with blue billboards was the best performing color combination in terms of participant response time for all conditions tested, even outperforming the color combination of blue text on a white background, which is what Gabbard et al. [GSH06] demonstrated to perform best in their work, as well as white text with no/transparent billboard which is what Kim et al. [KEL*19] demonstrated was best in their work. Their results provide further support that negative contrast UIs should be used instead of positive contrast UIs in OST displays.

It is important to note that there are differences in the OST devices used between the above-mentioned studies: Gabbard et al. [GSH06] used a Sony Glasstron PLM A55, while Kim et al. [KEL*19] used a HoloLens 1, and Debernardis et al. [DFG*14] as well as Fiorentino et al. [FDUM13] used a monocular Liteye 750A display. Such variation in the OST hardware may account for some of the discrepancies between the results of these studies, as displays capable of higher luminance may achieve better user performance from color combinations which are otherwise difficult to observe on displays with lower luminance.

While slightly different in their research focus and methods, Kruijff et al. demonstrated that blue colored billboards were rated to be highest by users in terms of being most noticeable, although the text color was not specified and was not varied between conditions [KOK*19]. They hypothesize that this is due to the prominence of S-photoreceptors in the periphery of the eye, which are most sensitive to blue light [WC83]. This is interesting and may open up an avenue for future research in this area if eye tracking is employed. In this manner, new labels that require the user's attention could be shown with blue billboards, then when the user is reading a label the billboard color could be set to white to increase their performance. Its possible that such a system would improve user performance in tasks which require quick attention to labels and annotations, but it is also possible that the color changes may be distracting and have a negative impact on the user.

The effects of text outline style enhancements were investigated by Gattullo et al. where they compared varying amounts of outline thickness to text without enhancement and text with billboard style enhancements [GUFM15]. Their results indicated that the variation in the amount of text outline had no significant impact on users in terms of response time. They also showed that this variation had significant effects on the error rate of users, and that error rate decreased as text outline increased, however this was only the case when the text color was darker and more transparent than the outline color. These effects appear to be mainly due to the color choices for the UI in their study, which were white text with blue outlines and black text with white outlines. Since user performance increased with outline thickness when using darker colored or transparent text on a lighter colored outline, it can be argued that a billboard style text enhancement outperforms text outlines for this color combination, but that performance will be similar between outlines and billboards when the color combination is the opposite. Therefore, since there is similar performance between outlines and billboards when text is a lighter color than the outline, it can also be argued that outline style text enhancements should be used due to the reduced amount of screen space taken up by outlines compared with billboards.

In another work by Gabbard et al. billboard style text enhancements were compared against text with outlines as well as text with a different form of enhancement, drop shadows [GSH*07]. The results of this work were somewhat contradictory to the results obtained in the above-mentioned papers, in that users performed significantly worse, making more errors with billboard style text enhancements that they did with outlines, drop shadows, or plain text without enhancements. While this would seem to indicate that billboards should be avoided compared to other types of text enhancements, this must be carefully considered because this work employed dynamic color enhancements which actively changed the coloration of the billboard, outline, or drop shadow based on the users' current view of their environment. Because these colors dynamically change, it is difficult to see how these results compare to the previously mentioned papers, and it can only be concluded that with this type of dynamic color enhancement, billboards tend to perform significantly worse than the other text enhancements.

5.2. Foreground Color, Background Color/Appearance, and Color Blending

As shown in the preceding section, UI color configurations have a major impact on the results of studies in the domain of OST displays. In this section, we discuss the general findings of the papers in the survey in terms of foreground color, background color, and color blending.

One interesting common finding between the different papers in this survey is that red colored UIs tend to yield poor user performance compared to other colors. This was shown to be the case by Fiorentino et al. Gabbard et al. in 2006, and Gabbard et al. in 2007, and Zhao et al. [FDUM13, GSH06, GSH*07, ZHHA17]. In each of these studies, red text performed worse than the majority of other color choices, however it is worth noting that in the work by Debernardis et al. users performed relatively well with red text [DFG*14].

In terms of background appearance, Gabbard et at. investigated the impacts of physical backgrounds with varying amounts of visual complexity or noise [GSH*07]. While it was hypothesized that backgrounds with high amounts of visual complexity would yield poorer user performance due to the reduced clarity of the border between virtual content and physical content, they found that this was not necessarily the case, with backgrounds such as a brick wall and foliage yielding improved user response times over backgrounds of pavement and sidewalk. In the work by Kim et al. [KEL*19], no significant effects on user performance were found between backgrounds of varying visual complexity, however they did note that users significantly preferred plain color backgrounds to complex appearing backgrounds in several conditions.

A fair portion of the papers in the survey focused on color blending, where the research goal was to better understand how light emitted by the OST display mixes with the light from the user's environment. Gabbard et al. in 2010 performed a study on this, where measurements were taken with a colorimeter from the user's perspective in the OST display, while it displayed imagery in varying colors and against varying background posters [GSZW10]. In this work, they found that the perceived color of virtual content on the display varies substantially depending on the lighting and coloration of the user's environment. They show that colors tend to have less saturation while mostly retaining their hue for white backgrounds, however hue shifts tend to occur for colored backgrounds. While they hypothesize that these color shifts can have significant effects on user perception of the colors, confirmation of this via a user study was left to future work.

Hincapie-Ramos et al. also examined this domain, and introduced a system called SmartColor which can be used to provide three different manners of color correction based on the user's environment [HRISI15]. Their system is capable of performing real time color corrections, contrast corrections, or smart enhancements where a billboard is displayed behind text when poor contrast is identified.

A formal user study on color blending was performed in work by Merenda et al. where users were tasked with matching colors displayed on an OST HUD to colors on a World Color Survey palette [MSG*16]. Their results are similar to what was found by Gabbard et al. [GSZW10] in that users tended to perceive colors as being less saturated than intended. They also show that there is deviation in the hues chosen by users when attempting to match the displayed color to the color on the palette, where purple colored elements are perceived by users to be more blue or red in some cases, and green colored elements are perceived without as much deviation in hue.

5.3. Environment Lighting and Contrast

A few papers investigated the impacts of environment lighting on the user performance, where in general it was noted that users tend to perform worse as the illuminance level of the environment increases. This was shown by Kim et al. for indoor lighting conditions between 10–300 Lux [KEL*19]. It was further shown for brighter conditions between 1000–4000 Lux (comparable to very bright indoor lighting or dim overcast outdoor lighting) by Gattullo et al. [GUFM15]. Similar effects were also found by Debernardis et al. although no measures of illuminance were taken in that particular study [DFG*14]. This is an expected trend, as the luminance level of current OST displays yields imagery that "washes out" and loses contrast under bright lighting conditions.

With the lighting ranges covered by the above-mentioned work, there is still room in this area for further investigation of both indoor lighting conditions up to 1000 Lux, as well as for bright outdoor lighting conditions above 4000 Lux. The instruments used to collect such illuminance measures are common and inexpensive, so future research in this domain should always report the illuminance levels of the testing environment, that way more comparisons can be made between studies.

5.4. Text Location and Size

Several papers investigated the placement and sizing of textual annotations and information on OST displays, which we discuss in this section.

Orlosky et al. created an algorithm which prioritizes the placement of virtual annotations over dark and uniform areas of the user's physical environment [OKT13]. While the system's decisions for placement did not always align to the positions deemed subjectively best by users, it did align portions of the time, and since no user study was performed with users wearing an OST-HMD and actively using the system, it is possible that the system's choices may result in increased user performance despite going against their preferred choices of locations. This is an interesting research area, and further work is required in order to better understand text placement strategies and their impacts on the user.

In terms of the sizing of information shown on OST displays, Gabbard et al. found in 2006 that the distance at which text was displayed from the users (and therefore also the size, since the font size did not scale up with distance) had no significant impacts on user performance, although this distance was only varied between one, two, and four meter depths, and was sized to be roughly two inches tall at a distance of two meters [GSH06].

Gabbard et al. also later showed in 2019 that users perform better in terms of accuracy and number of tasks completed when virtual text is positioned at a depth equal to the depth of information they are examining in their physical environment, although user performance in tasks based on both the virtual and physical content does decrease as the distance the information is displayed from the user increases [GMS19]. They hypothesised that this was due to the easier ability of the user to switch contexts between the virtual and physical stimuli when they are positioned at equal depths, whereas the user must accommodate their eye when there is a depth disparity between the virtual and physical stimuli. They further showed that user performance was better when context switching between two physical stimuli regardless of distance, than when context switching between a virtual and physical stimuli regardless of distance, indicating that AR systems in general may have performance penalties on user performance when compared with the same task in a physical environment.

Kruijff et al. noticed that in terms of noticeability, users preferred smaller sized annotations when they were being displayed in the center of the field of view of the display than they did when they were displayed toward the edges [KOK*19].

In their work with low-vision participants, Zhao et al. found that low vision users tend to prefer font sizes larger than 100 pixels when viewing content on OST-HMDs, or roughly two degrees of vertical visual angle if calculated by their OST device's field of view and resolution. They also found that there was no significant difference in visual acuity scores of the low vision users between physical visual acuity tests and virtual tests displayed on the HMD, whereas there was a significant decrease in acuity for sighted participants when tested with virtual tests as opposed to physical [ZHHA17].

6. General Discussion

In this section, we discuss the trends that occur between the literature gathered in this survey and provide recommendations on how to proceed with future work in this domain.

6.1. OST Hardware

One major difficulty in this domain is the lack of being able to easily transfer results of previous studies to current work due to variations in the available OST display hardware. There were 10 different devices used between the 14 papers in this survey, each with its own capabilities in terms of luminance, and with many different design features, with some utilizing beam splitters and LCD displays while others utilize waveguides (see [Cau95, LZW19]). These factors can have impacts on the way that the virtual content is perceived by users of the device, affecting things such as the perceived color, transparency, and image clarity or consistency [GSZW10, MSG*16, LZW19]. These difficulties are likely to continue as more OST displays are produced and used in future research [XH17], and while it is likely that there are certain results that will apply and extend across many different OST devices, we must always be aware of the individual differences and peculiarities of each device. Future research in this domain could help overcome these difficulties by incorporating study designs in which multiple OST devices are used. In this manner, we could gain further understanding of the perceptual differences that arise due to such hardware changes.

While all of the works in this paper have utilized additive light model OST displays, there is some research being done on new OST display technologies that support a *subtractive* light model, where virtual imagery is presented via the removal of light from the scene (e.g., [ILI*19, MF13]). There are still many issues to be worked out before these could become available to consumers, however when they are released, similar studies to the ones in this survey will have to be repeated for this type of device in order to understand how users perceive virtual content with this light model.

6.2. Dynamic Backgrounds

The majority of studies in the survey were done with participants facing either background posters or static environments, which only covers a small portion of use cases for OST displays. These devices are likely to be used in a variety of locations, many of which involve dynamic motion of people and other things in the user's environment. Only one study specifically tested against dynamic backgrounds, and incorporated a video that played behind the virtual content being displayed on the OST device [KOK*19]. While this is a step in the right direction, future work in this domain should consider how to perform research studies in physical dynamic environments, as it is possible that the additional depth information gained from being in a real environment versus in front of a screen may have impacts on user performance and subjective preferences. Such environments also introduce other interesting research questions, as annotations which follow moving objects in the scene may require vastly different presentation methods due to issues tied to refresh rate of the OST device as well as the changing background appearance.

6.3. User Demographics

The vast majority of work in this domain has involved participants that met strict exclusion criteria for their perceptual capabilities, such as having 20/25 visual acuity or better and having normal color vision. These works have also focused on a younger demographic of users (typically all under the age of 40), likely due to the convenience of obtaining participants from university communities. While studying this type of demographic covers arguably the majority of current use cases of OST displays, it is possible that there are differences in how users from different demographics perceive content on these devices, and therefor different guidelines for displaying virtual content to different demographics may be required.

Only one paper in the survey investigated a different demographic, by recruiting low-vision participants in order to evaluate if OST displays could be used as assistive devices for users with visual impairments. Their results are promising, and the authors show that virtual content must be presented in larger sizes and with careful regard to color choices when users come from this specific demographic [ZHHA17]. This illustrates the need to evaluate how users perceive virtual content on OST displays for many different demographics, especially including user age and perceptual capability.

6.4. Evaluating Results of User Studies

It has been a dream of researchers in this field that one day these devices may become as ubiquitous as the smart phones we commonly use today. If this occurs, it means that there will be a much larger population of people using these devices than has been evaluated for in user studies. We typically evaluate results in this domain in terms of what is statistically significant and what is not, and focus our discussions and recommendations only on the effects which are shown to be significant. Many of the papers in this survey did not report effect sizes for non-significant results, and many did not even include these statistics for the significant effects. This will become problematic as the number of users of OST displays increases, since effects which were not considered significant in the presentation of a study may become significant to a subset of users in this expanded population. Because of this, it is important to consider how we evaluate the results of future studies in this domain, and we should be careful to place more emphasis on reporting the effect sizes with statistics as opposed to focusing solely on statistical significance.

7. Conclusion and Future Work

This paper has presented a systematic survey of research which investigates how users perceive virtual content on augmented reality optical see-through displays. The papers presented here were gathered solely from the IEEE and ACM digital libraries, and provide a basis for the computer science centered work in this domain, however it is likely that additional relevant work exists in other technology centered libraries as well as libraries which focus more on vision sciences. For this reason, we hope to expand our review in the near future to include such libraries.

Acknowledgments

This material includes work supported in part by the Office of Naval Research under Award Number N00014-17-1-2927 (Dr. Peter Squire, Code 34) and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch).

References

[Azu17] AZUMA R. T.: Making augmented reality a reality. In *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)* (2017), Optical Society of America. 1

[BASOL16] BRUDER G., ARGELAGUET-SANZ F., OLIVIER A.-H., LECUYER A.: CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups. *Presence: Teleoperators and Virtual Environments 25*, 1 (2016), 1–16. 2

[Cau95] CAUDELL T. P.: Introduction to augmented and virtual reality. *Proceedings of SPIE*, 1 (1995), 272. 7

[CG60] CAMPBELL F. W., GREGORY A. H.: Effect of Size of Pupil on Visual Acuity. *Nature 187* (1960), 1121–1123. 2

[Dee05] DEEB S.: The molecular basis of variation in human color vision. *Clinical Genetics 67*, 5 (2005), 369–377. 2

[DFG*14] DEBERNARDIS S., FIORENTINO M., GATTULLO M., MONNO G., UVA A. E.: Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. *IEEE Transactions on Visualization and Computer Graphics 20*, 1 (2014), 125–139. 5, 6, 7

[EKBW20] ERICKSON A., KIM K., BRUDER G., WELCH G.: Exploring the Limitations of Environment Lighting on Optical See-Through Head-Mounted Displays. In *Proceedings of the ACM Symposium on Spatial User Interaction* (2020). 1, 2, 4

[FDUM13] FIORENTINO M., DEBERNARDIS S., UVA A. E., MONNO G.: Augmented Reality Text Style Readability with See-Through Head-Mounted Displays in Industrial Context. *Presence: Teleoperators and Virtual Environments 22*, 2 (2013), 171–190. 5, 6

[GMS19] GABBARD J., MEHRA D., SWAN J. E.: Effects of AR Display Context Switching and Focal Distance Switching on Human Performance. *IEEE Transactions on Visualization and Computer Graphics 25*, 6 (2019), 2228–2241. 2, 5, 7

[GSH06] GABBARD J. L., SWAN J. E., HIX D.: The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence: Teleoperators and Virtual Environments 15*, 1 (2006), 16–32. 5, 6, 7

[GSH*07] GABBARD J. L., SWAN J. E., HIX D., KIM S., FITCH G.: Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications. In *Proceedings of the IEEE Virtual Reality Conference* (2007), pp. 35–42. 6

[GSZW10] GABBARD J., SWAN J. E., ZEDLITZ J., WINCHESTER W. W.: More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In *Proceeding of the IEEE Virtual Reality* (2010), pp. 79–86. 2, 5, 6, 7

[GUFM15] GATTULLO M., UVA A. E., FIORENTINO M., MONNO G.: Effect of Text Outline and Contrast Polarity on AR Text Readability in Industrial Lighting. *IEEE Transactions on Visualization and Computer Graphics 21*, 5 (2015), 638–651. 5, 6, 7

[HGAB08] HOFFMAN D. M., GIRSHICK A. R., AKELEY K., BANKS M. S.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8*, 3 (2008), 33–33. 2

[HRISI15] HINCAPIÉ-RAMOS J. D., IVANCHUK L., SRIDHARAN S. K., IRANI P. P.: SmartColor: Real-Time Color and Contrast Correction for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics 21*, 12 (2015), 1336–1348. 2, 6

[HRL*16] HARDING T. H., RASH C. E., LATTIMORE M. R., STATZ J., MARTIN J. S.: Perceptual issues for color helmet-mounted displays: luminance and color contrast requirements. In *Degraded Visual Environments: Enhanced, Synthetic, and External Vision Solutions* (2016), vol. 9839. 1

[I*18] ISHIHARA S., ET AL.: Tests for Color Blindness. *American Journal of Ophthalmology 1*, 5 (1918), 376. 2, 4

[IK15] ITOH Y., KLINKER G.: Light-Field Correction for Spatial Calibration of Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics 21*, 4 (2015), 471–480. 1

[ILI*19] ITOH Y., LANGLOTZ T., IWAI D., KIYOKAWA K., AMANO T.: Light attenuation display: subtractive see-through near-eye display via spatial color filtering. *IEEE Transactions on Visualization and Computer Graphics 25*, 5 (2019), 1951–1960. 1, 7

[JM19] JAMIY F. E., MARSH R.: Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing 13* (April 2019), 707–712(5). 2

[KBB*18] KIM K., BILLINGHURST M., BRUDER G., DUH H., WELCH G.: Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics 24*, 11 (2018), 2947–2962. 1

[KBCW03] KIYOKAWA K., BILLINGHURST M., CAMPBELL B., WOODS E.: An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *Proceeding of the IEEE/ACM International Symposium on Mixed and Augmented Reality* (2003), pp. 133–141. 1

[KEL*19] KIM K., ERICKSON A., LAMBERT A., BRUDER G., WELCH G.: Effects of Dark Mode on Visual Fatigue and Acuity in Optical See-Through Head-Mounted Displays. In *Proceedings of the ACM Symposium on Spatial User Interaction* (2019), pp. 9:1–9:9. 5, 6, 7

[KOK*19] KRUIJFF E., ORLOSKY J., KISHISHITA N., TREPKOWSKI C., KIYOKAWA K.: The Influence of Label Design on Search Performance and Noticeability in Wide Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics 25*, 9 (2019), 2821–2837. 5, 7, 8

[Kra16] KRAMIDA G.: Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics 22*, 7 (2016), 1912–1931. 2

[KSF10] KRUIJFF E., SWAN J. E. I., FEINER S.: Perceptual Issues in Augmented Reality Revisited. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (2010). 1

[LZW19] LEE Y., ZHAN T., WU S.: Prospects and challenges in augmented reality displays. *Virtual Reality & Intelligent Hardware 1*, 1 (2019), 10–20. 7

[MF13] MAIMONE A., FUCHS H.: Computational Augmented Reality Eyeglasses. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (2013), pp. 29–38. 1, 7

[MS90] MYERS G. A., STARK L.: Topology of the near response triad. *Ophthalmic and Physiological Optics 10*, 2 (1990), 175–181. 2

[MSG*16] MERENDA C., SMITH M., GABBARD J., BURNETT G., LARGE D.: Effects of real-world backgrounds on user interface color naming and matching in automotive AR HUDs. In *Proceedings of the IEEE Virtual Reality Workshop on Perceptual and Cognitive Issues in AR* (2016), pp. 1–6. 2, 6, 7

[OKT13] ORLOSKY J., KIYOKAWA K., TAKEMURA H.: Dynamic Text Management for See-through Wearable and Heads-up Display Systems. In *Proceedings of the ACM International Conference on Intelligent User Interfaces* (2013), pp. 363–370. 5, 7

[SAW99] STRANG N. C., ATCHISON D. A., WOODS R. L.: Effects of defocus and pupil size on human contrast sensitivity. *Ophthalmic and Physiological Optics 19*, 5 (1999), 415–426. 2

[SES20] SINGH G., ELLIS S. R., SWAN II J. E.: The Effect of Focal Distance, Age, and Brightness on Near-Field Augmented Reality Depth Matching. *IEEE Transactions on Visualization and Computer Graphics 26*, 2 (2020), 1385–1398. 2

[WBSS19] WELCH G., BRUDER G., SQUIRE P., SCHUBERT R.: *Anticipating Widespread Augmented Reality: Insights from the 2018 AR Visioning Workshop*. Tech. rep., University of Central Florida and Office of Naval Research, August 2019. 1

[WC83] WILLIAMSON S. J., CUMMINS H. Z.: *Light and Color in Nature and Art*. John Wiley and Sons, 1983. 5

[XH17] XU M., HUA H.: High dynamic range head mounted display based on dual-layer spatial modulation. *Optics Express 25*, 19 (2017), 23320–23333. 7

[ZHHA17] ZHAO Y., HU M., HASHASH S., AZENKOT S.: Understanding Low Vision People's Visual Perception on Commercial Augmented Reality Glasses. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (2017), pp. 4170–4181. 4, 5, 6, 7, 8