Label Segregation by Remapping Stereoscopic Depth in Far-Field Augmented Reality

Stephen D. Peterson*
Department of Science and Technology
Linköping University

Magnus Axholt†
Department of Science and Technology
Linköping University

Stephen R. Ellis‡
Human Systems Integration Division
NASA Ames Research Center

Abstract

This paper describes a novel technique for segregating overlapping labels in stereoscopic see-through displays. The present study investigates the labeling of far-field objects, with distances ranging 100-120 m. At these distances the stereoscopic disparity difference between objects is below 1 arcmin, so labels rendered at the same distance as their corresponding objects appear as if on a flat layer in the display. This flattening is due to limitations of both display and human visual resolution. By remapping labels to predetermined depth layers on the optical path between the observer and the labeled object, an interlayer disparity ranging from 5 to 20 arcmin can be achieved for 5 overlapping labels. The present study evaluates the impact of such depth separation of superimposed layers, and found that a 5 arcmin interlayer disparity yields a significantly lower response time, over 20% on average, in a visual search task compared to correctly registering labels and objects in depth. Notably the performance does not improve when doubling the interlayer disparity to 10 arcmin and, surprisingly, the performance degrades significantly when again doubling the interlayer disparity to 20 arcmin, approximating the performance in situations with no interlayer disparity. These results confirm that our technique can be used to segregate overlapping labels in the far visual field, without the cost associated with traditional label placement algorithms.

Keywords: Label placement, user interfaces, stereoscopic displays, augmented reality, visual clutter, information layering.

Index Terms: H.5.2 [Information Systems]: User Interfaces; I.3 [Computing Methodologies]: Computer Graphics

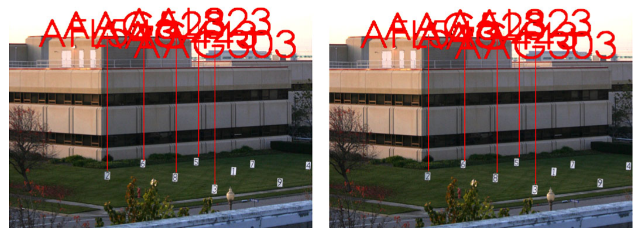
1 Introduction

Mixed and Augmented Reality (MR and AR) techniques integrate visually superimposed computer graphics with real world views to provide important data about observed scenes, reducing or removing the need to scan separate information sources such as computer monitors, diagrams and maps. Although there are many potential benefits in overlaying symbology and virtual information, care must be taken when designing such a visual user interface to avoid cluttering. Visual clutter can fragment or occlude the information, making it ambiguous.

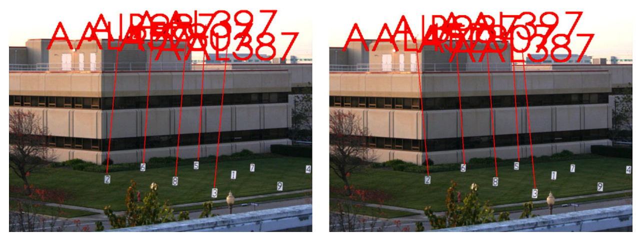
Textual labels annotating scene objects are important sources of visual clutter. Since they contain textual data, they require a minimum size to be readable. Moreover, such labels should be spatially registered with the background objects they describe. Due to such constraints, it is often unavoidable that labels obscure background objects or other labels, and thus also become elements of visual clutter in the display.

*e-mail: stepe@itn.liu.se
†e-mail: magax@itn.liu.se
‡e-mail: sellis@mail.arc.nasa.gov

IEEE International Symposium on Mixed and Augmented Reality 2008
15 -18 September, Cambridge, UK
978-1-4244-2859-5/08/$25.00 ©2008 IEEE



(a)



(b)

Figure 1: Uncrossed disparity stereograms illustrating the label layering concept: a) the default condition when labels are rendered at the same distance as the markers in the background, producing a seemingly flat label layer, and b) depth segregation of labels, which in this study was found to yield a significant performance improvement with interlayer disparities in the 5-10 arcmin range. These stereograms should however not be considered exact reproductions of the visual stimuli in the viewing conditions due to the uncalibrated stereo disparities. Furthermore these stereograms, with integrated background and overlay imagery, have largely different optical properties than the real visual stimuli produced when overlaying the far visual field using optical see-through HMDs.

The application domain explored in this work is AR displays for air traffic control (ATC) towers. This information integration application is attractive because of the current air traffic controller's needs to scan radar and weather displays, monitor physical or digital flight strips, and maintain visual surveillance through the windows of the control tower. Superimposing data onto aircraft in the window view, using see-through display technology, could reduce the attention shifts required to retrieve information in traditional head-down display devices.

Despite the high elevation of the control tower cab, the comparatively much larger distance to observed aircraft will cause vertical compression towards the local horizon. Thus, labels attached to aircraft monitored from the tower would likely overlap, especially at large airports with distant runways and taxiways.

In some applications, e.g. information visualization, the background layer that graphically describes the data can be spatially rearranged to simplify the labeling task. In AR in general, and in the tower window view in particular, there is no such flexibility; the background layer is represented by physical objects that are not under the control of the user interface.

A common approach to reduce the problem of label overlap in graphical user interfaces is to evaluate available 2D screen space, and constantly reposition labels to available locations. This technique is commonplace in, for example, cartography [6, 27], ATC radar displays [5, 16], and scientific illustration [9, 10]. This technique is, however, not constrained to 2D user interfaces; many 3D systems including MR and AR implementations utilize the same approach on the 2D view plane projection of the 3D scene [2, 3, 4, 19, 22, 23, 28]. This approach is often successful in avoiding label overlaps and thereby reducing clutter; however, if the user or the observed objects are non-stationary, the approach introduces 2D screen motion since labels rearrange according to the optimal label layout over time. This 2D screen motion may disturb or distract the system user [1], and may also cause confusion as to where a specific label is located relative its object at a specific instant.

There are alternative approaches to label management that do not introduce 2D motion. Instead they alter the appearance of the label by adjusting transparency [12, 13, 17] or color coding, or they use filtering to remove labels and other information deemed unimportant for the current task [15]. However, these approaches require automated importance classification, which in ATC applications can entail a safety risk.

We are instead investigating the feasibility of utilizing the third dimension, depth, available in stereoscopic displays, to manage the visual clutter associated with label overlap. This approach, *label layering*, does not introduce any 2D motion, dimming or filtering of the labels; it works by solely adjusting their stereoscopic disparities. Though stereoscopic disparity is mostly known as a depth cue, it is possibly more important for its role in image segregation through its support for camouflage breaking. Three-dimensional objects that share identical visual textures with their background can be revealed by disparity gradients visible in stereographic viewing systems [11]. Aerial stereography takes advantage of this fact in uncovering camouflaged military targets, invisible in a single camera view, and so does the segmentation algorithms used on stereoscopic image sequences (e.g. [18]). The random dot stereogram, invented by Julesz [14], could be considered the "perfect camouflage" since hidden objects appear only if viewed by a stereographic system such as human vision; disparity is the only means to separate objects from the background noise.

Our label layering technique uses the same concept in a cluttered AR interface: A user with stereo vision can isolate and read overlapping labels solely based on their internal disparity differences.

At the distances of interest in this work, over 100 m, the natural disparity differences between objects are very small; generally below 1 arcmin. Thus, if the labels were registered in 3D with their objects at these distances, the disparity differences between adjacent labels would be negligible; typical display devices with visual resolutions of several arcmin per pixel cannot present such small disparities. We are therefore interested in evaluating the effects of larger disparity differences, in the range between 5 and 20 arcmin. In order to produce these exaggerated interlayer disparities we need to spatially remap the labels into predefined depth layers on the optical path between the observer and the overlaid objects. The stereograms in figure 1 illustrate the label layering concept, and may be fused if viewed with uncrossed stereo vision, i.e. converging the eyes behind the physical image plane of the paper.

We have conducted a previous investigation of the benefits for visual search of label layering in laboratory conditions [21]. The experiment reported in this paper extends that work to a setting with optical conditions closer to those in an ATC tower, a far-field environment with small natural accommodation and vergence demands due to the large distance. Will disparity-based label segregation improve performance even though it introduces a larger relative vergence demand? Will the technique be affected by the accommodative demand of an optical see-through head-mounted display (HMD) which is focused closer than the background objects? Our study will evaluate various magnitudes of disparity differences between adjacent label layers, that is, different *interlayer disparity*, in order to determine if, for example, increasing depth segregation will aid in isolating and reading the individual labels, and thus affect user performance.

Section 2 presents related work to the experiment described in section 3. The results of the experiment are provided in section 4 and discussed in section 5. Section 6 gives a general conclusion, and also outlines future work on an automated label layering algorithm.

2 Related Work

Most previous work on visual clutter reduction in user interfaces through stereoscopic depth segregation can be found within information layering, a concept where information of different types is placed in different depth layers. This is slightly different from our label layering technique where we aim to segregate information of the *same type*, labels, in depth. However, these two approaches share many principles.

Information layering, using the dual physical display planes in a multilayered display, has been investigated and compared to traditional, single layer, display devices [26]. The multilayered display used in that study is composed of two LCD screens separated by a 1.4 cm transparent Perspex layer, producing both disparity and focus difference between the LCDs. This technology thus avoids the accommodation-vergence conflict inherent to most stereoscopic viewing systems, but has limited flexibility in terms of changing depth layer positions. This study found no performance benefit in the simple task conditions; however, significant performance improvements were detected using the multilayered display to segregate information under demanding task conditions.

The information layering approach was also found to effectively offset the effects of added visual clutter in a visual tracking task on a flight display [20]. The clutter was added to the display in the form of noise, making the display visually crowded, and in the form of absence of color coding for the tracking and target symbols, making the stimuli and task ambiguous. The noise, tracking and target symbols were present in only one of three segregated depth planes, reducing the noise impact and providing a cue to clearly identify the tracking and target symbols.

Multilayer interaction in an information layering context has been investigated on a tabletop display [24]. Results indicate that 4 layers spaced 4 cm apart yielded optimal user performance in a sketching application, where each layer provided different interaction and navigation functions. The results were, however, not compared to a single-layer interface.

Previous work on label layering has shown that distributing multiple overlapping labels in depth significantly improves user performance in a visual search task [21]. It was determined that the depth order of the labels must be correlated with that of the overlaid objects; a random depth distribution of labels conversely significantly degraded user performance, likely due to a mismatch in depth cues between labels and objects hampering label-object integration. That earlier work differs from the current work in that it did not evaluate varying magnitudes of depth segregation, but rather maintained fixed disparity differences at approximately 6 arcmin; furthermore it was performed in an indoor environment with smaller relative accommodation and vergence demand than the currently investigated far-field AR setting.

3 METHOD

We constructed an experimental setup which allowed us to fulfill our three design goals:

(i) Place physical markers in the far visual field (≥100 m) from the observer, where their relative stereoscopic disparity differences are small (<1 arcmin), in close proximity (∼1°) within the visual field of the observer, so that overlaid labels would partially overlap each other.

(ii) Overlay the markers with text labels in a HMD; remap the location of the labels from their default positions, co-located with the markers, to depth layers along the optical path between the markers and the observer.

(iii) Evaluate the user performance in a visual search task, requiring integration of information in background markers and overlaid labels with various magnitudes of stereoscopic disparity between the layers. Performance should then be compared with that of the default condition with no stereoscopic disparity differences between labels.

3.1 Hardware Setup

Each participant was seated on a stool on the roof of a two-story building, with an eye point approximately 10 m above ground level. The overlay graphics were presented to each participant through a Kaiser ProView 50ST, an optical see-through HMD with 74% binocular overlap. The active screen area of each display unit provided VGA resolution (640×480 pixels) with a 35°×26° Field-of-View (FoV), giving a pixel size of 3.3 arcmin. The partial overlap configuration yielded an extended horizontal FoV of 47°. The virtual image presented by the display was focused at 3 m (0.33 diopters).

Figure 2: The marker layout of the nine physical markers on the grass lawn, at an average distance of 110 m from the observer.

Nine physical markers were placed on a grass lawn, on ground level, 110±10 m from the participant's location (figure 2). The markers were placed in a random pattern within an area covering 11°×2° of visual angle. No marker was closer than 1° of visual angle to any other marker. Each marker was constructed of white foam core board of rectangular shape, 0.7×1.0 m, and firmly attached to the ground. Black gaffer's tape of width 10 cm was used to create the numeric characters 1-9, uniquely assigned to each board. The size of the numeric characters was maximized on each board, with a margin of 10 cm to the marker board edge. In terms of visual angle, the marker boards were approximately 30 arcmin in height, and glyph thickness of the numeric characters was 3 arcmin. Virtual targets intended to be spatially registered with physical targets, such as the markers, were measured with a see-through calibrated ruler to have spatial registration errors of less than 0.5 degrees within the field of regard used in the experiment.

The luminance of the white background markers was measured with a SpectraScan 650 photometer. It ranged between 6,300 cd/m^2 at 16:00 and 23,700 cd/m^2 at 10:30, the respective end and start times of the daily period used for experimentation. The HMD has

Figure 3: The HMD was retrofitted with thick black felt to block any unwanted ambient light to enter the eyes, which would have reduced the contrast of the overlay graphics. Three of the six cameras used for tracking can be seen as black boxes behind the participant.

Table 1: Luminance values of the background markers and HMD graphics, measured at 10:40 and 16:00, the start and end times of the trials of each experiment day.

	Time of day	
	10:40	16:00
Marker luminance (natural) (cd/m^2)	23,700	6,300
Marker luminance (transmitted) (cd/m^2)	21.61	5.74
HMD luminance (cd/m^2)	3.6	3.6
Luminance ratio (HMD/transmitted background)	0.167	0.627

a transmittance of 0.245, and was fitted with two layers of neutral density filters; each with a transmittance of 0.061. The luminance of the HMD when displaying a full red image was 3.6 cd/m^2. The transmitted luminance values and corresponding ratios are shown in table 1. Because of passing clouds and movement of the sun during the day, it was necessary to vary the number of neutral density filters used to ensure visibility both of the virtual images and the background markers during the experiment. In general it was necessary to keep the virtual display elements at least 10-15% brighter than the background. The display was retrofitted with a hood (figure 3) made of thick black felt, ensuring that no ambient light entered the HMD from the sides or below, and that the only light to reach the eyes passed through the eyepieces with neutral density filters.

Although seated, the participants were free to move their upper body and head. Head tracking data was fed through a 6-camera PhaseSpace Impulse optical tracking system, with 6 active LED markers placed on the back of the HMD mount (figure 4).

The HMD was driven by an NVIDIA Quadro FX 1400 stereo graphics card on an Intel Xeon workstation running Ubuntu Linux. The experimental code was written in C++ using OpenSceneGraph for scenario construction and VR Juggler for tracking and display configuration.

3.2 Task

We developed a task that would require integration of information in the background markers with the information in the overlay. To solve the task in each trial, the participant needed to find a specific label, and judge its horizontal positioning relative to that of a given background marker. Thus, the task was not 2D-solvable, i.e. information from both background and overlay was required for a correct judgment. Furthermore, we chose a response technique that would require as little sensory-motor effort and mental workload as

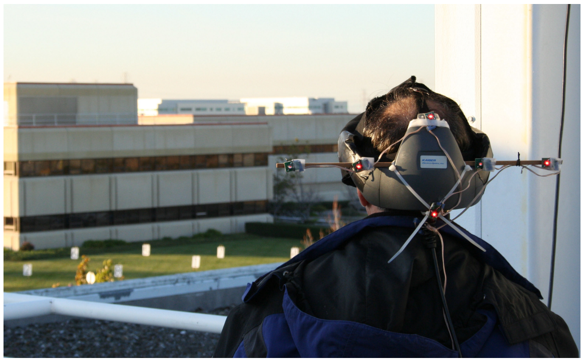


Figure 4: The back of the HMD mount was fitted with active LED markers, tracked by six cameras behind the participant. The background markers can be seen in the distance.

possible. By utilizing an input device with only two buttons, we could, accordingly, map the two left and right horizontal judgment responses to each of the two horizontally placed buttons.

The task of each experimental trial was thus to identify two given markers, and to assess whether the first "target" marker was to the left or the right of the second "reference" marker in the participant's visual field. The target marker was specified through its label, displayed in the HMD, whereas the reference marker was specified through its marker number visible in the scene. The reference marker was never chosen to be one of the horizontal edge markers, "2" or "4", since that task could be solved through methods of exclusion without using information in the HMD (see figure 2).

The labels consisted of 6-character randomized alphanumeric sequences resembling aircraft callsigns, three letters followed by three digits. The first letter and the last number were kept constant within each trial. By thus requiring the participant to read the centermost four characters we could minimize the possibilities of using methods of exclusion. By randomly generating the alphanumeric sequences the participant would not be able to recall a specific label and its location from a previous trial, thereby minimizing training effects due to scenario familiarity. The labels were approximately $6.0^\circ \times 1.5^\circ$ of visual angle in size, where the total width varied slightly depending on the character glyphs.

Only a subset of five of the nine markers was labeled in each trial, since this is the maximum number of label layers that could be accommodated considering the viewing condition with the largest depth segregation (See the section on Independent Variables below). Both the target and reference markers were included in this subset. In total, 15 such marker layouts were created with similar complexity. Each layout defined two scenarios; one with the target marker selected on the horizontal edge of the marker layout, and one with the target marker selected within the three centermost markers. Thus, a total of 30 scenarios were used for the trials. The number of times "right" versus "left" was the correct answer was equated and their presentation order counterbalanced.

3.3 Participants

We recruited 14 participants, 10 male and 4 female. Ages ranged from 21 to 60, with nine participants below 40.

All participants were staff, students or contractors at NASA Ames Research Center in Mountain View, CA. The procedure used in this study was vetted as "Low Risk" by the NASA Ames Institutional Review Board (IRB). All participation was voluntary and standard IRB precautions were taken to ensure participants' informed consent, anonymity and security during the experiment.

All participants passed the Orthorater stereo vision test, with a required retinal disparity of 1 arcmin or better.

3.4 Procedure

Each participant was given written instructions before the trials. The instructions stressed that accuracy should be prioritized over response time, but that each trial should be completed as quickly as possible.

We measured each participant's inter-pupilary distance (IPD) using a digital pupilometer. The IPD measurement was used to calibrate the stereo disparity of the HMD.

After participants completed the stereo test and signed the consent form, they were seated on a stool on the rooftop location. Each participant put on the HMD with the active tracker markers and was also given an input device, a box with two buttons: left and right. The participant was told to hold the left thumb permanently on the left button, similarly for the right thumb.

Before the experimental trials, eight test trials were given. During these trials the participants were accustomed to the task and equipment, and screened for stereoscopic depth perception to confirm that they could correctly perceive the various label layouts.

The HMD was calibrated for each participant by manually adjusting the head rotation parameters until each label leader line extended to the center of its marker.

Before each trial the target and reference markers were given in the HMD; the target marker was given through its label (e.g. `ABC123`) and the reference marker through its unique number (e.g. 5). Thus, a trial would be initialized with a question on the form "Is marker ABC123 left/right of marker 5?" The participant would then press any button on the input device in order to remove the task question, and to initiate the trial. When the experimental trial started the participant would scan through the labels, identify the target marker label, and follow its leader line to find the target marker. The participant would then judge the horizontal position of this marker relative that of the reference marker, left or right, and press the corresponding button on the input device. As soon as the button was pressed, the trial was completed, the labels were removed, and the next trial question was presented. If there was no response after 20 seconds, the trial timed out and the participant proceeded to the next trial without responding.

The trials were divided into three blocks of 30 trials each. During the breaks between the blocks, the participants could rest, and even remove the HMD and walk around. In case participants removed the HMD during the break, it was recalibrated before recommencing. The participants were also asked to report any discomfort during the break. After the trials each participant completed a questionnaire. The total time including preparations, trials and questionnaire was approximately 60 minutes per participant.

3.5 Independent Variables

Viewing Condition Each label was rendered with a certain stereoscopic disparity in the HMD, making the label appear at a distinct depth to the participant. The depth segregation of the layers was given by the stereoscopic disparity difference between layers, the *interlayer disparity*, which was the independent variable determining the principal viewing conditions of the experiment.

There were three viewing conditions with varying amounts of depth segregation - small, medium and large - with 5, 10 and 20 arcmin interlayer disparity respectively. The depth order of the labels was correlated with the depth order of the corresponding markers, to satisfy the criterion of ordered disparity in [21]. In each of these conditions the closest label layer was located at 2.2 m, while the farthest label layer was located at 11 m in the large interlayer disparity condition. Thus, the binocular viewing conditions ensured little accommodation-vergence mismatch since the vergence demand for the rendered labels, in the most extreme case of 20 arcmin interlayer disparity, ranged from 0.45 to 0.09 meter-angles compared to the accommodative demand of 0.33 diopters.

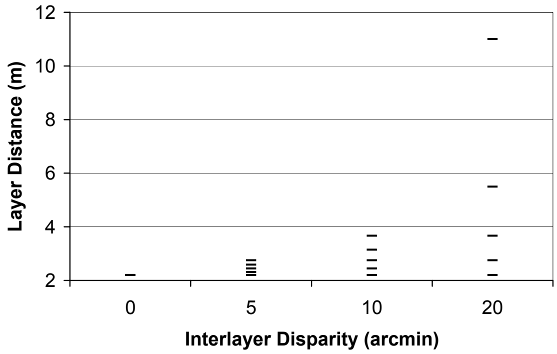


Figure 5: The distance to the five label layers in the different viewing conditions at an IPD of 64 mm. Note that there was a fifth viewing condition, with 0 arcmin interlayer disparity at a distance of 100 m, which is located out of the scale of this diagram. The 0, 5, 10 and 20 arcmin measurements correspond to the no, small, medium and large interlayer disparity conditions, respectively.

Furthermore, by placing the closest label at the same distance in each condition, the maximum vergence demand for completing a task (the vergence difference between the closest label and background imagery) would be the same, 0.45 to 0.01 meter-angles, avoiding any bias in the overall performance due to differences in total vergence demand. As a comparison, one could instead have chosen to set the farthest label in each condition at the same distance, for example at 100 m, and then place the closer labels at decreasing distances according to the required interlayer disparity. Because the HMD was focused at 0.33 diopters, this alternative approach would introduce a larger overall accommodation-vergence mismatch in all conditions, and favor the small interlayer disparity condition due to the smaller total vergence demand (0.09 meter-angles) compared to the large interlayer disparity condition (0.36 meter-angles). We chose the first approach to avoid any such bias.

In addition there were two viewing conditions with no interlayer disparity, all labels being perceived to be located in a single depth plane. In the two different conditions the depth planes were located at either 2.2 or 100.0 m respectively.

Figure 5 shows label layer distances for the different viewing conditions.

Vertical Separation The labels could have either correct, no, or inverse vertical separation with respect to each other. In the *correct vertical separation* condition each label was placed at the same height in the visual field as that of its corresponding marker, although with an offset in order to not directly overlay the marker; i.e., each label was located at a constant visual angle above its marker as seen by the participant. In the *no vertical separation* condition all labels were located at the same height in the visual field, thus removing any vertical offset and maximizing the amount of label overlap. In the *inverse vertical separation* condition the labels were separated by the same amounts as in the correct condition, but the height in the visual field of each label was inverted with respect to its corresponding marker and so the closest object, with the lowest height in the visual field, was connected to the highest label in the visual field.

Horizontal Target Position The target marker was either selected to be one of the three center markers or one of the two horizontal edge markers in a trial. If an edge marker was selected as the target, its label would only be in an overlap situation on the centermost portion, while the entire label of a center marker would be in an overlap situation.

Repetition Each combination of the other independent variables was repeated three times per participant.

3.6 Dependent Variables

Response Time We recorded the duration from the instant a trial was started until a response was given by the participants. The participants responded by pressing one of the two buttons on the input device, corresponding to the two possible responses: right or left. If no response was given within 20 seconds from the start of the trial, it timed out and the response time was capped to 20 seconds.

Response Error The response error was recorded. The response was erroneous if the participant pressed the wrong button on the input device.

3.7 Experimental Design

All participants saw all combinations of the independent variables in a repeated measures within-subject design. In total 1260 data recordings were made, 90 per participant (5 viewing conditions$\times$3 vertical separations$\times$2 horizontal target positions$\times$3 repetitions).

The viewing condition was blocked in groups of three, allowing any transient effects to abate when viewing conditions changed. The three blocked repetitions should not be confused with the independent variable repetition where all independent variables were fixed. Only the viewing condition was blocked, and all other conditions were randomized for each participant. Accordingly, the three trials in a specific block were not exact replications of the same conditions.

The blocks were constructed using a partial Latin squares design and were re-used for all participants. The presentation order of the blocks between participants was counterbalanced through random permutations.

4 Results

Initial analysis of the data revealed no transition effects on response time within the blocks of same viewing condition ($F(2,24) = 0.05, ns$). Further analysis showed no changes in the pattern of statistically significant results when excluding the first block repetition. Thus, we concluded that there were no transient effects due to the changes of ocular vergence in the different viewing conditions. Accordingly we performed all subsequent analyses on all trials within the blocks.

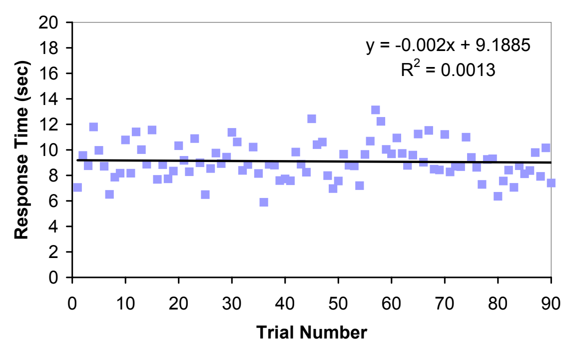


Figure 6: Mean response time per trial number. Trial number had no significant effect on response time, showing that no overall training effect was present. The equation of the linear regression through all mean values is given in the figure.

Using regression analysis on the median response time of each trial we found no significant effect for trial number on response time

($r = 0.035, df = 88, ns$), meaning that participants showed no general improvement of performance throughout the trials (figure 6). This confirms not only that our measures to minimize training effects due to scenario familiarity were successful, but also that there was no significant training effect due to skill development.

One participant was removed from the data due to exceptionally large error rates (28%), over four times the median error rate. This fact, coupled with the lowest average response times of all participants, indicates that the participant misunderstood the instruction of prioritizing accuracy over response time. This exclusion from the analysis did not change the pattern of statistically significant results, but it will improve the accuracy of the data since it is clearly important that the instructions are understood and followed.

4.1 Response Time

The results were analyzed using analysis of variance (ANOVA), using a repeated measures design with participant as a random variable and a fixed model for all other independent variables.

Timeouts were treated as correct responses since the participant would likely have responded correctly had more time been given. One could argue for assigning a higher response time than the actual timeout time in such cases, but we use the more conservative approach of capping the response time to 20 s. Thus any statistically significant effects will remain even if timeouts are treated differently.

Erroneous responses are not included in the analyses for response time.

Initial analysis of the data showed a significant main effect of viewing condition on response time ($F(4,48) = 5.56, p < 0.001$). Post-hoc Scheffé comparisons showed no significant differences in response time between the two conditions with no interlayer disparity ($F(4,385) = 0.007, ns$). Hence the difference in depth of the single label layer in each of these two conditions did not significantly affect the performance. Since we are interested in the effect of depth segregation of labels compared to the traditional "2D" layout with no interlayer disparity, and this pair of means did not vary more than 0.5%, we collapsed these two conditions into one condition called "no interlayer disparity". Subsequent analyses and diagrams are based on this approach.

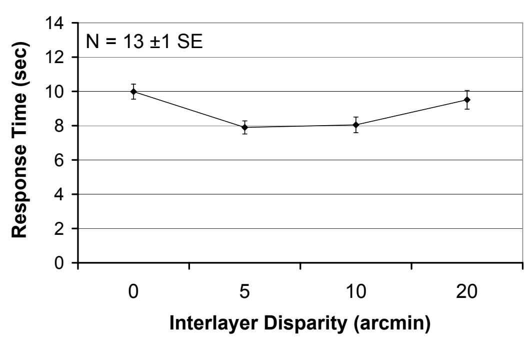


Figure 7: Viewing condition had a significant main effect on response time. The lowest average response time, 7.9 s, is found in the small (5 arcmin) interlayer disparity condition.

As the main result of this experiment we found a significant main effect of viewing condition on response time ($F(3,36) = 6.47, p < 0.001$), shown in figure 7. Post-hoc Scheffé comparisons revealed that the small interlayer disparity condition was significantly faster (20.9%, 2.1 s) than the no interlayer disparity condition ($F(3,308) = 12.9, p < 0.01$). The small interlayer disparity condition was also significantly faster (16.9%, 1.6 s) than the large interlayer disparity condition ($F(3,308) = 7.70, p < 0.1$). There was no statistically significant difference between the small and medium interlayer disparity conditions ($F(3,308) = 0.07, ns$), nor between the large and no interlayer disparity conditions ($F(3,308) = 0.67, ns$).

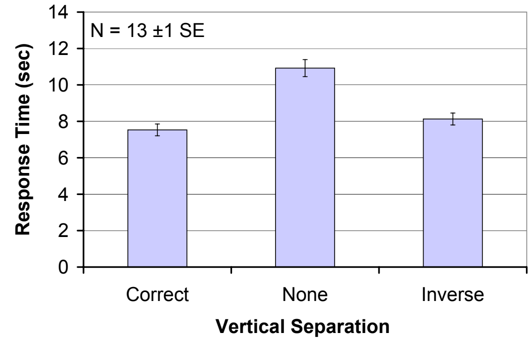


Figure 8: Vertical separation had a significant main effect on response time.

A significant main effect was also found for vertical separation on response time ($F(2,24) = 34.2, p < 0.001$), shown in figure 8. Post-hoc Scheffé comparisons showed significantly lower response times for both the correct ($F(3,312) = 60.0, p < 0.01$) and inverse ($F(3,312) = 40.7, p < 0.01$) vertical separation conditions, compared to the condition with no vertical separation. There were however no significant differences between the correct and inverse vertical separation condition ($F(3,312) = 1.86, ns$). The fact that the conditions with no vertical separation significantly degraded the performance shows that our experimental design was successful; the task becomes much more demanding when maximizing the amounts of label overlap.

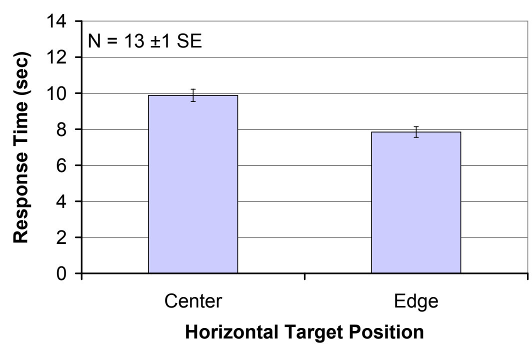


Figure 9: Horizontal target position had a significant main effect on response time.

Furthermore we found a significant main effect on horizontal target position on response time ($F(1,12) = 42.7, p < 0.001$), shown in figure 9. When the target marker was on the horizontal edge of the marker cluster, the response time was significantly faster (20.5%, 2.0 s). This further showed that we successfully made the task harder by increasing the amounts of overlap on the target marker, since in the slower center condition the label was in an overlap situation on both sides.

There were no significant interaction effects for any combinations of independent variables on response time.

Table 2: Mean error and timeout rates for each of the four viewing conditions (%).

	Interlayer Disparity			
	None	Small	Medium	Large
Error Rate	6.81	4.61	5.81	6.46
Timeout Rate	16.3	5.21	7.43	17.0

4.2 Response Error

The mean error rate, the percentage of judgments which were erroneous, was 6.1%. We also analyzed the mean timeout rate, the percentage of waived responses, which was found to be 12.4%. Thus the rate at which no correct response was given within the time limits was 18.5%. The timeout rate is also reflected in the response time data analysis since timeouts were treated as correct responses, but timeout rates are reported separately here for clarity.

Wilcoxon signed rank tests showed no significant difference between the two viewing conditions with no interlayer disparity, either in error rate (Wilcoxon statistic$= 22.0, ns$) or timeout rate (Wilcoxon statistic$= 35.0, ns$). As discussed in the previous section we are interested in the effect of depth segregation of labels compared with the traditional "2D" layout with no interlayer disparity. Since the pairs of means did not vary by more than 3.1%, we collapsed these two conditions into one, no interlayer disparity. The subsequent analyses and diagrams are based on this approach.

In the following analyses we use the non-parametric Friedman test on the ranks of error and timeout rates to determine main effects. If they are found to be significant, we use the Wilcoxon signed rank test for pairwise comparisons of conditions.

We found no significant main effect of viewing condition on error rate ($\chi^2(3) = 1.36, ns$). There was however a significant main effect of viewing condition on timeout rate ($\chi^2(3) = 11.15, p < 0.05$). The small interlayer disparity condition had a significantly lower timeout rate than the no interlayer disparity condition (Wilcoxon statistic$= 3.00, p < 0.01$) and the large interlayer disparity condition (Wilcoxon statistic$= 4.00, p < 0.05$). There was no significant difference between the small and medium interlayer disparity conditions (Wilcoxon statistic$= 15.5, ns$). Table 2 summarizes the mean error and timeout rates for the different viewing conditions.

A significant main effect was found for vertical separation on error rate ($\chi^2(2) = 6.00, p < 0.05$). The correct vertical separation condition had a significantly lower error rate than the no vertical separation condition (Wilcoxon statistic$= 8.00, p < 0.05$). There was no significant difference between the no and inverse vertical separation conditions (Wilcoxon statistic$= 61.0, ns$), nor the correct and inverse vertical separation conditions (Wilcoxon statistic$= 26.0, ns$). There was furthermore a significant main effect of vertical separation on timeout rate ($\chi^2(2) = 15.9, p < 0.001$). Pairwise comparisons showed significant differences between all levels of vertical separation; between correct and none (Wilcoxon statistic$= 0.00, p < 0.01$), between none and inverse (Wilcoxon statistic$= 66.0, p < 0.01$), and between correct and inverse (Wilcoxon statistic$= 3.00, p < 0.05$). Table 3 summarizes the mean error and timeout rates for the different viewing conditions.

Finally we found a significant main effect of horizontal target position on error rate ($\chi^2(1) = 4.92, p < 0.05$), where a horizontal edge position of the target marker in a trial yielded a significantly lower error rate (5.23%) than the center positions (8.64%). Similarly we found a significant main effect of horizontal target position on timeout rate ($\chi^2(1) = 9.31, p < 0.01$), where the horizontal edge cases yielded a significantly lower timeout rate (8.83%) than center cases (15.7%).

Table 3: Mean error and timeout rate for each of the three vertical separation levels (%).

	Vertical Separation		
	Correct	None	Inverse
Error Rate	5.22	8.64	5.35
Timeout Rate	5.28	24.18	6.95

4.3 Questionnaire Data

In the post-exercise questionnaire the participants responded to a number of questions using a five-point Likert scale, where five represented a large impact on overall task difficulty. The participant first rated a number of factors that may have adversely affected the overall performance (see figure 10). The ratings of *concentration problems*, *headache*, *eye strain* and *neck pain* are potential indicators of excessive experiment length or equipment problems. The *overlay jitter* rating is a potential indicator of issues with tracker calibration or run-time performance in outdoor conditions. The *marker visibility problems* rating is an indicator of inadequate background marker size, brightness or contrast. The *stereo fusion problems* rating is an indicator of stereo perception issues or calibration problems with the display equipment.

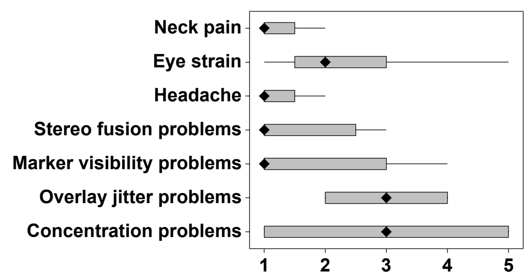


Figure 10: Questionnaire responses for potential problem indicators. A rating of 1 represents a small impact on task difficulty, while 5 represents a large impact. In this Tukey [25] boxplot the median values are indicated with a diamond.

The participants were asked to rate the effect of depth segregation and vertical separation on the task difficulty (figure 11). Depth segregation showed a significantly lower reported difficulty level in both the vertically separated case (Wilcoxon statistic$= 91.0, p < 0.01$) and the case with no vertical separation (Wilcoxon statistic$= 78.0, p < 0.01$). Note that in this question we did not distinguish the different viewing conditions with varying interlayer disparity, but for simplicity grouped them into one condition: depth segregation.

The participants were also asked to rate the effect of depth segregation magnitude on task difficulty (figure 12). Large depth segregation showed a significantly lower reported difficulty level than small depth segregation (Wilcoxon statistic$= 67.0, p < 0.05$).

Finally the participants were asked to state which of the five viewing conditions was preferable. Out of 13 participants, 7 (54%) preferred the large interlayer disparity case; one preferred each of the small and medium interlayer disparity cases; two preferred depth segregation but did not specify the interlayer disparity magnitude; and two did not provide an answer. Thus 11 out of 13 preferred depth segregation of some magnitude.

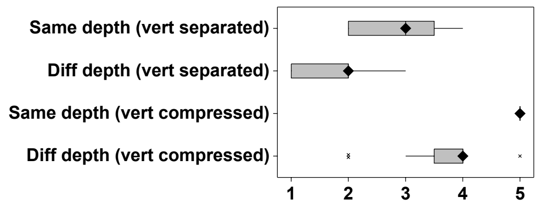


Figure 11: Questionnaire responses on the effect of depth segregation ("diff depth") and vertical separation of labels on task difficulty. A rating of 1 represents a small impact on task difficulty, while 5 represents a large impact. In this Tukey boxplot the median values are indicated with a diamond.

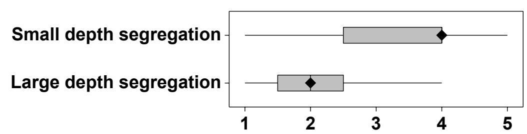


Figure 12: Questionnaire responses on the effect of the magnitude of depth segregation on task difficulty. A rating of 1 represents a small impact on task difficulty, while 5 represents a large impact. In this Tukey boxplot the median values are indicated with a diamond.

4.4 Control Case

We performed a separate experiment on eight of the participants in a different display condition: the see-through optics of the HMD was blocked by opaque plastic. The background markers were instead rendered in the HMD at the same distances as the real markers, yielding equal vergence demand as in the main experiment. The purpose of this control case was to determine if and how the label layering technique was affected by the large accommodative demand in the far-field AR setting; in the completely virtual environment of the control case there was a fixed accommodative demand, 0.33 diopters, since all imagery was rendered on the virtual image plane at 3 m. Furthermore there were no registration errors since all objects in the visual field were virtual.

When comparing the results with the main experiment we treated the display condition (see-through or opaque) as an independent variable in a within-subject design. The order of the display condition was counterbalanced for the eight subjects. Preliminary analyses of the data using ANOVA showed a significantly lower response time (~ 1 s) in the opaque condition than in the see-through condition ($F(1,7) = 11.7, p < 0.05$), but no other interaction effects with the other independent variables. These initial data indicate that the increased accommodative demand in the far-field see-through condition does yield a longer overall response time, but it does not change the pattern of statistically significant results with respect to the other independent variables. The full analysis of this experiment will be reported in a separate paper.

5 Discussion

Viewing condition showed a significant effect on response time, with the small interlayer disparity case of 5 arcmin depth segregation yielding the best performance. This condition was not significantly different from the medium disparity condition, with 10 arcmin depth segregation. Thus, labels should be placed within the 5-10 arcmin interlayer disparity range for optimal performance. Likely a more detailed and focused experiment could find the optimum with a better resolution, or even produce a model describing the exact behavior of this function if needed. Interestingly, increasing the depth segregation by setting the interlayer disparities to 20 arcmin does not further improve response times or error rates; on the contrary the performance degrades and is not significantly different from the case without any depth segregation. Likely the extreme stereoscopic depth differences in this condition hampers stereo fusion due to the large vergence demand, making the detection of individual depth layers more difficult and time consuming. The practical effect of this finding is that under the requirement of smaller depth separation, more layers could theoretically be accommodated for; however, the specific effects of adding more overlapping labels requires empirical investigation.

The effect of vertical separation shows that our experiment design was successful; the task complexity significantly increased when removing any natural vertical label offsets and so increasing the magnitude of label overlap. Similarly the task complexity increased by choosing a central target marker, producing overlaps on both sides of the target label. The absence of interaction effects between these two variables and the principal independent variable, viewing condition, shows that the decluttering behavior of depth segregation does not change significantly as the magnitude of overlap changes. This might indicate that the perceptually "simple" conditions, with correct vertical label separation and an edge target marker, were already visually cluttered above a certain saturation threshold so that additional clutter did not change the effect of depth segregation. Alternatively our experiment was unable to detect such effects, perhaps due to random factors such as registration error due to tracker jitter making these interactions fade.

The correct and inverted vertical separation conditions showed similar performance. This indicates that the height of the label in the visual field does not need to match the height of its object, and thus the label does not need to be elevated at a constant angle. A more controlled and focused experiment should, however, investigate this effect further, and also consider layouts with random vertical separation.

If timeouts had been assigned a response time value over 20 seconds the main effect of viewing condition would likely have been even stronger. The performance in the large and no interlayer disparity conditions would likely have deteriorated further since the timeout rates in these conditions were significantly higher. Importantly, the main effect of viewing condition is significant without imposing any punishment for timeouts on the actual response times.

When analyzing the questionnaire data we found a relatively high overall difficulty rating due to tracker jitter. The optical tracking system was chosen since it would not interfere with or be affected by the metal walls and other electronics equipment on the rooftop location. However, the optical system may have been affected by glints due to sunlight in the marker LEDs. Perhaps a more stable tracking would have been achieved with sunlight protection, but the jitter did not compromise the overlay-background registration and thus the tracking quality was deemed sufficient for this experiment.

Like the empirical response time data, the questionnaires reported significantly lower difficulty levels with depth segregation of the labels, both with and without vertical separation (see figure 11). Interestingly, large depth segregation was judged to simplify the task compared with small depth segregation (see figure 12), in contrast to the response time data that showed significantly better performance for small depth segregation. Furthermore, when participants were asked to select their overall preferred viewing condition, the large depth segregation was the most frequent response. Some of this effect might be attributed to participants not correctly remembering the precise characteristics of the different viewing conditions when completing the questionnaire; however, the results indicate that the large disparity condition was not subjectively judged to be a more difficult condition, contrary to the measured performance.

6 Conclusion & Future Work

We have shown that remapping the stereoscopic depth of multiple overlapping labels in far-field AR significantly improves user performance. Response times in a visual search task integrating labels with background information improved over 20% or 2.1 seconds on average, compared to leaving labels at their default location. Placing labels at their default locations in far-field AR, co-located with the objects they annotate, produces little or no disparity difference (< 1 arcmin) between the labels. Our results show that setting the stereoscopic disparity differences in the range 5-10 arcmin yields significantly improved performance, both in terms of response time and response error. Our data also shows that the disparity differences must not be exaggerated; setting the disparity differences at 20 arcmin significantly degraded performance to levels similar to the condition without any disparity differences at all.

The current results add to the previous findings [21] on the label layering technique; we now know that i) the label order should be correlated with the distance order of the background objects, ii) the height in the visual field of the labels does not need to be correlated with that of the background objects, iii) the interlayer disparity should be in the range 5-10 arcmin, iv) the label layering technique works not only in a laboratory environment but also in a far-field AR setting with optical properties similar to those in an ATC tower. This empirical baseline data is important in the future development of an automated label layering algorithm using stereoscopic disparity, which effectively expands the design space and could alleviate the need for compromises required by traditional decluttering methods, such as filtering, dimming, aggregation, or label size reduction. This algorithm would subsequently be evaluated against traditional 2D label placement algorithms, which could be performed in two steps, each aiming at one of the identified potential benefits of label layering: i) improved label-object integration and ii) absence of distracting label motion in 2D screen space.

Firstly, the potential benefits of label layering for label-object integration would be analyzed. How would users performing a visual search task benefit from the knowledge that labels are constantly located at a specific angle and screen distance from the objects they annotate? I.e., how does segregating labels in depth instead of 2D screen space affect human performance, and how does this compare to the traditional approach of dispersing labels in 2D screen space but not in depth?

Secondly, the benefits of continuous label segregation over time without motion in 2D screen space would be analyzed. Label layering indeed introduces label motion, but only in stereoscopic depth. How does this label movement in stereoscopic depth affect performance in a visual search task, and to what extent could it distract and capture unnecessary attention compared to the motion in 2D screen space caused by traditional label placement algorithms? There have been previous studies showing that pure stereomotion does not cause pop-out effects from disparity noise [8]. Other research has shown that looming objects are much more likely to capture attention than receding ones [7]. The looming stimuli used in that study were however provided through relative size adjustments, with stereoscopic disparity fixed. Conversely, our system would vary the stereoscopic disparity of labels while keeping relative label size fixed. The practical effect of depth motion in our system requires empirical investigation.

As a further extension of the current work one could envision the development of a fully three-dimensional label placement algorithm, which could evaluate both depth layering and traditional view-plane based label position for optimal label placement in any specific situation.

Acknowledgements

Stephen D. Peterson and Magnus Axholt were supported by PhD scholarships from the Innovative Research Programme at the EUROCONTROL Experimental Centre, Brétigny-sur-Orge, France. These authors were also supported through the NASA Grant NNA 06 CB28A to the San José State University Research Foundation. The experiment was conducted at NASA Ames Research Center in Mountain View, CA (corresponding author: Stephen R. Ellis, sellis@mail.arc.nasa.gov).

References

[1] K. R. Allendoerfer, J. Galushka, and R. H. Mogford. Display system replacement baseline research report. Technical Report DOT/FAA/CT-TN00/31, William J. Hughes Technical Center, Atlantic City International Airport (NJ), 2000.

[2] R. Azuma and C. Furmanski. Evaluating label placement for augmented reality view management. In *Proceedings of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003)*, pages 55–75, Tokyo, Japan, October 2003.

[3] B. Bell and S. Feiner. Dynamic space management for user interfaces. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pages 238–248, San Diego, CA, November 2000.

[4] B. Bell, S. Feiner, and T. Höllerer. View management for virtual and augmented reality. In *Proceedings of the symposium on User Interface Software and Technology (UIST'01)*, pages 101–110, Orlando, Florida, 2001.

[5] A. Dorbes. Requirements for the implementation of automatic and manual label anti-overlap. Technical Report 21/00, EUROCONTROL Experimental Centre (EEC), 2000.

[6] S. Edmondson, J. Christensen, J. Marks, and S. Shieber. A general cartographic labeling algorithm. *Cartographica*, 33(4):13–23, 1996.

[7] S. L. Franconeri and D. J. Simons. Moving and looming stimuli capture attention. *Perception & Psychophysics*, 65(7):999–1010, 2003.

[8] J. M. Harris, S. P. McKee, and S. N. J. Watamaniuk. Visual search for motion-in-depth: stereomotion does not pop-out from disparity noise. *Nature Neuroscience*, 1:165–168, 1998.

[9] K. Hartmann, K. Ali, and T. Strothotte. Floating labels: Applying dynamic potential fields for label layout. In *Proceedings of 4th International Symposium on Smart Graphics*, pages 101–113, Berlin, 2004. Springer Verlag.

[10] K. Hartmann, T. Götzelmann, K. Ali, and T. Strothotte. Metrics for functional and aesthetic label layouts. In *Proceedings of 5th International Symposium on Smart Graphics*, pages 115–126, Berlin, 2005. Springer Verlag.

[11] M. B. Holbrook. Breaking camouflage: stereography as the cure for confusion, clutter, crowding, and complexity - three-dimensional photography. *Photographic Society of America Journal*, 8, 1998.

[12] M. S. John, B. A. Feher, and J. G. Morrison. Evaluating alternative symbologies for decluttering geographical displays. Technical Report 1890, Space and Naval Warfare System Center, San Diego, CA, 2002.

[13] M. S. John, H. Smallman, D. I. Manes, B. A. Feher, and J. G. Morrison. Heuristic automation for decluttering tactical displays. *The Journal of the Human Factors and Ergonomics Society*, 47(3):509–525, 2005.

[14] B. Julesz. *Foundations of Cyclopean Perception*. The University of Chicago Press, Chicago, 1971. ISBN: 0-226-41527-9.

[15] S. Julier, M. Lanzagorta, L. Rosenblum, S. Feiner, and T. Höllerer. Information filtering for mobile augmented reality. In *Proceedings of ISAR 2000*, pages 3–11, Munich, Germany, October 2000.

[16] S. Kakos and K. J. Kyriakopoulos. The navigation functions approach for the label anti-overlapping problem. In *Proceedings of the 4th EUROCONTROL Innovative Research Workshop*, Paris, France, 2005.

[17] W. Kienreich and M. Granitzer. Visualising knowledge maps for encyclopedia articles. In *Proceedings of the Ninth International Conference on Information Visualisation (IV05)*. IEEE, 2005.

[18] R. Koch. 3-d surface reconstruction from stereoscopic image sequences. In *Proceedings of Fifth International Conference on Computer Vision (ICCV)*, pages 109–114, 1995.

[19] S. Maass and J. Döllner. Efficient view management for dynamic annotation placement in virtual landscapes. In *Proceedings of the In-*

ternational Symposium on Smart Graphics, volume 4073 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin, 2006.

[20] R. V. Parrish, S. P. Williams, and D. E. Nold. Effective declutter of complex flight displays using stereoptic 3-d cueing. Technical Report 3426, NASA, 1994.

[21] S. D. Peterson, M. Axholt, and S. R. Ellis. Managing visual clutter: A generalized technique for label segregation using stereoscopic disparity. In *Proceedings of the IEEE Virtual Reality Conference (VR'08)*, pages 169–176, Reno, NV, March 2008.

[22] E. Rose, D. Breen, K. H. Ahlers, C. Crampton, M. Tuceryan, R. Whitaker, and D. Greer. Annotating real-world objects using augmented reality. In *Computer Graphics: Developments in Virtual Environments (Proceedings of CG International '95 Conference)*, pages 357–370, Leeds, UK, June 1995.

[23] E. Rosten, G. Reitmayr, and T. Drummond. Real-time video annotations for augmented reality. In *International Symposium on Visual Computing*, 2005.

[24] S. Subramanian, D. Aliakseyeu, and A. Lucero. Multi-layer interaction for digital tables. In *Proceedings of the symposium on User Interface Software and Technology (UIST'06)*, pages 269–272, October 2006.

[25] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA, 1977.

[26] B. L. W. Wong, R. Joyekurun, H. Mansour, P. Amaldi, A. Nees, and R. Villanueva. Depth, layering and transparency: Developing design techniques. In *Proceedings of the Australasian Computer-Human Interaction Conference (OZCHI)*, Canberra, Australia, 2005.

[27] M. Yamamoto, G. Camara, and L. A. N. Lorena. Tabu search heuristics for point-feature cartographic label placement. *GeoInformatica*, 6(1):77–90, 2002.

[28] F. Zhang and H. Sun. Dynamic labeling management in virtual and augmented environments. In *Proceedings of the 9th International Conference on Computer Aided Design and Computer Graphics (CAD/CG)*, 2005.