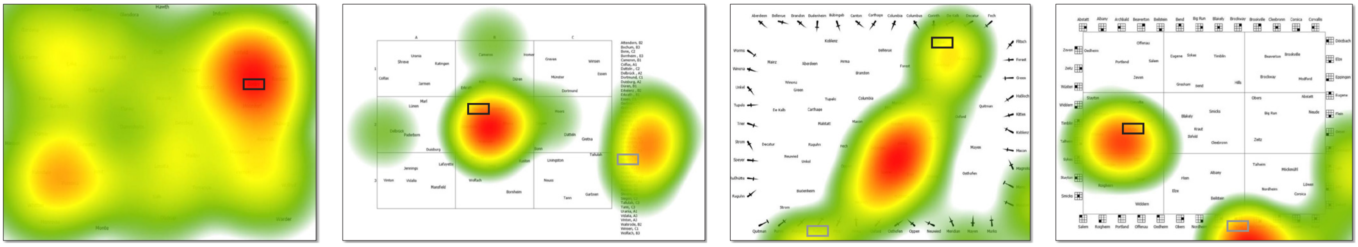
An Evaluation of Visual Search Support in Maps

Rudolf Netzel, Marcel Hlawatsch, Michael Burch, Sanjeev Balakrishnan, Hansjörg Schmauder, and Daniel Weiskopf, *Member, IEEE Computer Society*



(a) Within-image annotation (WA) (b) Grid reference annotation (GA) (c) Directional annotation (DA) (d) Miniature annotation (MA)

Fig. 1. Heatmaps of the eye-tracking data of one participant for the four evaluated variants of map annotation with labels. The different annotation variants lead to significantly different gaze characteristics. The target label is marked with a black box. In the case of the methods with support for visual search ((b)–(d)), the label is also marked with a gray box in the ordered list outside the map.

Abstract—Visual search can be time-consuming, especially if the scene contains a large number of possibly relevant objects. An instance of this problem is present when using geographic or schematic maps with many different elements representing cities, streets, sights, and the like. Unless the map is well-known to the reader, the full map or at least large parts of it must be scanned to find the elements of interest. In this paper, we present a controlled eye-tracking study (30 participants) to compare four variants of map annotation with labels: within-image annotations, grid reference annotation, directional annotation, and miniature annotation. Within-image annotation places labels directly within the map without any further search support. Grid reference annotation corresponds to the traditional approach known from atlases. Directional annotation utilizes a label in combination with an arrow pointing in the direction of the label within the map. Miniature annotation shows a miniature grid to guide the reader to the area of the map in which the label is located. The study results show that within-image annotation is outperformed by all other annotation approaches. Best task completion times are achieved with miniature annotation. The analysis of eye-movement data reveals that participants applied significantly different visual task solution strategies for the different visual annotations.

Index Terms—Visual search, laboratory study, eye tracking, map visualization

1 Introduction

There are many tasks that require visual search for objects in an image. Especially in the context of visualization, such search tasks are often supported by appropriate interaction methods: typically, the user can select an object type and the respective positions are highlighted in the visualization. However, there are scenarios in which only static representations are available or interactive selection is not provided, e.g., in printed maps or in computer-based visualizations that do not facilitate appropriate highlighting.

With this paper, we aim to evaluate and understand visual cues that support visual search in 2D visualizations. We chose geographic maps as objects of study because they are one of the oldest types of visualization, they are in wide use, and they serve as a prototypical example of 2D visualization. We use a simple version of maps: a 2D space with text labels corresponding to the objects that participants had to find visually. With this design, we avoid any confounds from additional visual elements, focusing on the actual search task.

In a controlled eye-tracking study, we compare four map annotation techniques. The analysis of the eye-gaze data allows inference about possible gaze patterns, workload, and cognitive processes of participants. The baseline is *within-image annotation* (WA; Figure 1(a)), which just contains the labels on the map without any support for visual search. Traditional *grid reference annotation* (GA; Figure 1(b)) is commonly used in city or street maps, enriching the map by an additional sorted list of labels that come with textual coordinates referring to cells in the map (e.g., grid cell "C4").

Other methods in this study also utilize a sorted list of elements indicating the positions of the labels within the map to provide search support. Nevertheless, such lists are not the only way to provide search support, e.g., maps usually contain natural landmarks, which can be used for orientation. However, we focus here on explicit annotations.

With annotations, the search space on the map is heavily reduced. However, grid reference annotation requires switching between a textual representation of a sub-area of the map and its actual visual location on the map. Therefore, we designed two new annotation variants: *Miniature annotation* (MA; Figure 1(d)) avoids this mental switching by replacing the textual coordinates with a graphical index in the form of a miniature grid that highlights the indexed cell. Finally, *directional annotation* (DA; Figure 1(c)) also uses a graphical index, yet with an arrow that points to the label within the map. Therefore, it reduces the search space to a straight line (in the direction of the arrow), instead of reducing it to the area of a cell as for grid reference or miniature annotation. For miniature and directional annotation, we place the supporting labels and graphical cues at the border of the map. All four map variants are displayed in Figure 1, overlaid with representative examples of gaze data from the study.

One result of this eye-tracking study is that maps with supporting annotations lead to better search performance than the baseline within-

• *The authors are with VISUS, University of Stuttgart. E-mail: {rudolf.netzel, marcel.hlawatsch, michael.burch, daniel.weiskopf}@visus.uni-stuttgart.de.*

Manuscript received 31 Mar. 2016; accepted 1 Aug. 2016. Date of publication 15 Aug. 2016; date of current version 23 Oct. 2016.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2016.2598898

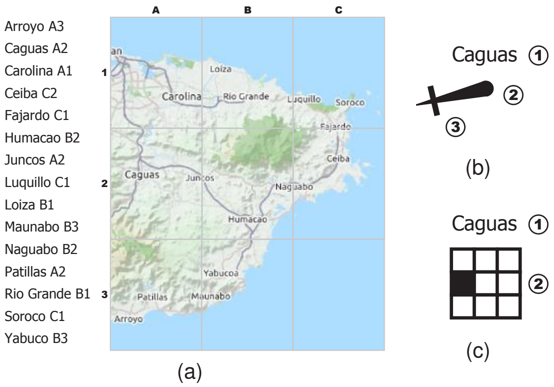


Fig. 2. (a) Traditional grid reference as commonly used in maps. It combines a geographic and a list-based representation of the labels in alphabetical order. The labels are combined with grid reference descriptions to accelerate the label search task in the map. (b) Directional annotation uses an arrow (2) to indicate the direction in which the element with the respective label (1) is located. The distance to this element is shown with a line marker (3) on the arrow. (c) Miniature annotation uses a miniature grid (2) to indicate the location of the element with the respective label (1). The grid corresponds to the grid overlaid on the map and the cell that contains the element is marked.

image annotation. Best task completion times are achieved with the miniature annotation approach. The analysis of the eye-movement data reveals interesting aspects of visual task solution strategies: participants apply significantly different reading strategies for the different annotation methods; we identified different steps that participants performed during the search task. We relate our eye-tracking results to an idealized model of the visual search steps. Based on our results, we discuss possible future directions of visual search support for visualization techniques.

2 Related Work

Visual search is studied in many fields, ranging from perceptual psychology all the way to its application in interactive visualization [12, 19, 23, 30, 34]. These studies either investigate if a visualization is useful for a given task or they focus on how visualizations can be enhanced with respect to visual cognition, perception, and attention.

To perform visual search, a reader has to actively scan the visual environment, e.g., a map filled with textual descriptions. All elements that are not of interest are so-called distractors. The search for a certain feature becomes more difficult with increasing number of distractors, unless the search target “pops out”. This phenomenon is researched in preattentive processing [11, 30].

In this paper, we focus on visual search not supported by preattentive, bottom-up processing. Instead, we consider the typical problem of searching for a target within a group of perceptually similar distractors, focusing on top-down perceptual processing. The concrete case is the search for textual objects that are used as labels in a 2D map. To speed up visual search, we have to reduce the search space from the full map to a smaller part. A typical example of this approach is the use of grid references (see Figure 2): the map is visually enhanced by an (alphabetically) ordered list of the text labels (such as street names) associated with a textual coordinate of a grid cell (e.g., “B2”). With this approach, visual search is still required but is more efficient, since it is restricted to a certain cell. Researchers argue that grids are beneficial compared to other reference line systems because grids are simple, more accurate, and facilitate map compilation [28]. We include grid references in our evaluation, since they are commonly used in maps.

Although grid references are useful and support the guidance to a smaller region in which the target is located, a viewer first needs to build the correspondence between the textual description of the grid cell (coordinates) and its spatial location, which comes with additional cognitive load. Therefore, we conjecture that a visual correspondence between annotation and target position in the map might be beneficial.

An intuitive way of establishing visual links is by drawing lines that connect two elements, e.g., as in node-link diagrams. Such diagrams typically utilize full-length straight links that perfectly guide the viewer from one location to the connected object [15]. This approach can also be used to connect boundary labels around maps with interior points [2]. Unfortunately, diagrams with many full-length links can exhibit visual clutter [27]. Therefore, crowded maps do not work with standard link representations; however, partial links [6] that point into the coarse direction of a location are a promising alternative. We adopt them for our directional annotation approach. Alternatively, directional information could also be shown by partial circles whose curvature, respectively size, indicates distance [1]. Another method that still connects labels at the border of the map with the actual position in the map with a line is presented by Kindermann et al. [16]. They rearrange the labels at the border and bend lines in order to avoid crossings and clutter.

As another variant, we use miniature annotation. Here, correspondence is not shown explicitly as for direct links, it is rather given implicitly by marking the corresponding grid cell in a simplified miniature representation of the map domain. Miniature representations are often used in visualization, for example, for sparklines [31] or as 3D miniature maps to support navigation in 3D virtual worlds [29]. Another example are TimeRadarTrees [5], which use a radial grid miniature annotation to visualize the correspondence to a large context representation of a dynamic graph. Furthermore, origin–destination (OD) maps can be visualized in a way that miniature versions of the map are placed inside the cells of a regular grid [35].

For map design in general, there are several suggestions, guidelines, and principles; see, e.g., the books by Robinson [25], Robinson et al. [26], or Tyner [32]. According to Tyner, the design goals are clarity, order, balance, contrast, unity, and harmony. Our annotation approaches do not change the content of carefully designed maps, but rather include small additional information outside of the map to support a reader in performing a search task. There are other examples of enhanced map visualization based on further graphical elements, including label placement for metro maps [37], map labeling [7], graphical legends [8], highlighting of interesting locations in tourist brochures [3], or annotation with loop lines [36].

Our paper is based on an eye-tracking study to assess the visualization efficiency, following the widely accepted need for improved evaluation in visualization research [18], including eye tracking [17].

We are not aware of any previous study that would have compared map annotations like ours. However, loosely related, there is a study of participants’ performance (latency and accuracy) for tasks with three different map legend designs [10]. In contrast to our tested methods, legends are used here to provide the meaning of a symbol on the map. Their findings show no best or least efficient map legend, but the authors came to the conclusion that the design should by influenced by the purpose of the map and also by the opinion of users. Another eye-tracking study assessed the readability of links encoding directional information in trajectory visualization [24]. The authors use here eye-tracking metrics to get insight into the cognitive workload of participants. According to their evaluation, tapered links performed best, which we use for the directional annotation.

Independent of which map or annotation is used, a viewer is supposed to establish a correspondence between the available location information and the actual position on the map. Therefore, the topic of this paper is related to the general problem of correspondence visualization, which also includes visual comparison (e.g., for multidimensional data [9]) or visual support in multiple coordinated views [21, 33]. The concepts and research results in these fields can be applied to the visual search support provided by map annotations. Here, annotations provide perceptional cues to make relations between labels and actual locations in the map clearer, and focus the attention of participants on specific areas on the map.

3 Visual Search Support

To support search tasks in maps, a connection between an alphabetically ordered list of elements and their position in the map has to be provided. Thereby, the reader first searches the ordered list for the element of interest. The list provides then a hint where the element can be found on the map.

3.1 Within-Image Annotation (WA)

The baseline for comparison is a map with labels directly placed at their corresponding positions in the display space (Figures 1(a) and 4(a)). This within-image annotation does not provide any extra visual cues to accelerate the search for labels in a map. In the worst case, the reader has to scan the entire map. Furthermore, there is no inherent ordering of labels, like in an alphabetically ordered list, i.e., no additional support for a faster search.

3.2 Grid Reference Annotation (GA)

The traditional grid reference approach is commonly used in maps (Figure 2(a)). A 2D Cartesian coordinate system divides the map into smaller cells. An alphabetically ordered list provides an index for the names and labels contained in the map. This list includes respective textual coordinates for each entry (e.g., “B2”). The reader then only has to search in the cell with the coordinates provided by the list (Figures 1(b) and 4(b)).

Based on this concept, we designed two variants of annotation techniques that support search tasks in maps: directional annotation and miniature annotation. Both techniques have in common that they use a more direct visual connection between the ordered list and the map. In contrast, the traditional grid reference uses two coordinates in a textual representation to indicate the respective cell in the map.

3.3 Directional Annotation (DA)

This annotation technique uses an arrow to point in the direction of the element of interest (see Figure 2(b)). The shape of the arrow resembles a tapered link, which was shown to be an efficient link representation [15]. We do not use a full-length link to the element of interest but rather a shortened version in the sense of a partial link [6]. With this approach, the user has to search only in the indicated direction for the element. To further improve the search, a marker on the arrow indicates the distance of the annotation to the element. Hence, this annotation technique uses direction and distance as cues for the location of the element of interest.

There are other possible encodings for the direction and distance of an element, e.g., the distance could be mapped to color. However, our design is more intuitive because we use a direct representation of both quantities, i.e., direction is represented by the angle of the arrow and distance is represented by the position of the marker.

With this approach, the duration for finding a designated element depends on the proximity of the annotation to the element. We place the ordered list of elements directly around the map to minimize the average distance between arrows and labels within the map (Figures 1(c) and 4(c)).

3.4 Miniature Annotation (MA)

Miniature annotation is more closely related to the traditional grid reference concept. We also use a grid to divide the map into different cells. Instead of using a 2D coordinate system to indicate the cell with the designated element, we use a visual miniature representation of the grid and mark the respective cell in it (see Figure 2(c)). With this approach, it is not required to read the textual coordinates and search for them along the coordinate axes of the map.

This approach has some advantages compared to directional annotation. Instead of using two different encodings for direction and distance, the miniature representation incorporates both. This might be more intuitive to understand. Furthermore, when using the directional annotation without any additional visual aids, the reading accuracy decreases with the distance of the element [4]. In contrast, the reading accuracy of the miniature annotation is independent of the location of the respective element. Finally, it is not required to place the miniature annotations close to the map.

A drawback of this approach is that the location of the designated element is only coarsely represented, while the directional annotation points to the exact location. The accuracy of the directional annotation only depends on perceptual accuracy, but not on the resolution of the miniature discretization.

One design parameter for the miniature annotation is the number of cells of the used grid. With a small number of cells, the marked cell can be identified faster, e.g., with a 3×3 grid (see Figures 1(d) and 4(d)), it is only required to see if the marked cell is in a corner or in between. However, the size of the cells is large in such cases and contains many elements that must be scanned. With a higher number of cells, fewer elements are contained, thus, the identification of the row and column position of the marked cell requires more time. The same issue occurs with the traditional grid reference annotation. For grids with a very large number of cells, the traditional approach might perform better since the coordinates of the respective cell are directly shown.

4 Model of Visual Search in Annotated Maps

This section discusses the idealized steps that have to be taken to read the different annotations. These steps will build the basis to discuss the scalability of the variants (Section 7) and lead to hypotheses about user performance as well as reading and visual task solution strategies with the respective visualizations (Section 5.1).

The different annotation approaches require different steps that have to be performed to find an element of interest in the map (Figure 3). Our annotation variants (DA, MA) were designed to improve some of these steps in comparison to the traditional grid reference approach.

Step 1: The first step describes the search for the label of the element of interest. Here, the labels are ordered alphabetically for rapid identification. The directional and miniature annotation place graphical representations around the map in clockwise order, whereas the grid reference uses a more compact text block. Therefore, the search with the grid reference is restricted to a smaller area and possibly faster. However, with the directional and miniature approaches, the annotations are closer to the map, which can be beneficial for the subsequent search steps.

Step 2: In the second step, the search cue in the annotation needs to be recognized. The directional annotation requires the reader

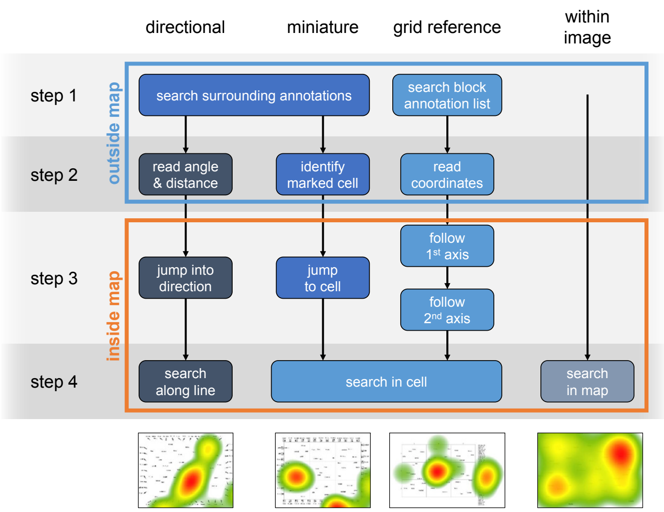


Fig. 3. Steps that have to be performed with the annotation methods. These steps are reflected in the results of our eye-tracking study (heat maps at the bottom, for one participant).

to estimate the distance to the target using the distance marker before the search can start in direction of the arrow. In the case of the miniature annotation, the marked grid cell must be identified, whereas the grid reference uses textual coordinates for this purpose. We assume that this step can easily be performed with the miniature annotation, at least for the selected size of the grid (3×3), since it uses a direct visual representation of the auxiliary grid on the map.

Step 3: In step three, the search for the area, indicated by the annotation, is performed. Miniature and grid reference annotation differ here in the way how they indicate the respective cell of the auxiliary grid. We assume that this step is faster with miniature annotation, since the reader can directly jump to the marked cell, whereas the grid reference requires additional look-ups at the two axes of the coordinate system. In the case of the directional annotation, the duration of this step depends on the reader, since there is no visual feedback if the correct area has been found. If the reader performs this step fast but inaccurately, it will take longer to find the correct element in the subsequent search step.

Step 4: The last step describes the final search for the element of interest in the search area indicated by the annotation. In the case of the within-image annotation, this is the only performed step since no additional search support outside the map is provided; the search area is not reduced and therefore the search must be performed in the full map. The miniature and the grid reference annotations benefit from the auxiliary grid on the map. In contrast, it is not feasible to combine the directional annotation with auxiliary lines, since many of them would intersect with each other in different angles leading to visual clutter. Therefore, the directional annotation provides a less exactly defined search area than the other two approaches. This might lead to a longer search process, especially in the case of larger distances between the annotation and the element of interest because the probability is higher that the reader leaves the search line defined by the arrow.

5 User Study Design

With a controlled laboratory eye-tracking study, we evaluate and compare the different annotation methods. Within-image annotation serves as baseline for the comparisons. Our study investigates task performance in terms of completion times as well as characteristics of the recorded gaze data. In the following, the design and execution of the study are discussed.

5.1 Hypotheses

Based on our theoretical considerations from Section 4, we built two hypotheses with respect to task completion times before running the experiment:

H1: Without annotations, the search for certain textual elements requires more time, since all textual elements have to be read in the worst case. Therefore, longer average completion times are expected for within-image annotation than for stimuli produced with the other annotation methods (directional, miniature, and grid reference annotations). We expect the following ordering of completion times: $WA > GA$, $WA > DA$, $WA > MA$.

H2: There will be differences concerning the completion time between directional, miniature, and grid reference annotations. We expect that $GA > DA > MA$.

Hypothesis H2 reflects that there are different perceptual sub-tasks involved with these annotations (see Section 4 and Figure 3).

Typically, the evaluation of error rates would also be considered in task-performance studies. However, our pilot studies indicated that participants were always able to find the designated label reliably, and therefore, we did not expect significant differences in task accuracy. The high accuracy is plausible because it is easy to check whether a text label is identical to a specified text, once the label is found.

The following hypotheses concern gaze characteristics and reading strategies that can be inferred from eye tracking. The reading strategies are formulated according to Figure 3. To this end, we define related areas of interest (AOIs) for the outer regions of the stimuli containing the annotations (*outside*) and the interior area where the labels are located (*inside*). The AOIs are illustrated in Figure 4. Steps 1 and 2 of Figure 3 happen in *outside*, steps 3 and 4 in *inside*.

H3: The average saccade length for directional, miniature, and grid reference annotations is larger than for within-image annotation because the ordered list of labels and the search cues allow longer jumps within the list and to the target.

H4: For directional, miniature, and grid reference annotations, the visual search starts in the outer area (*outside*), proceeds then to the interior (*inside*), and ends there at the target. We also expect that the average fixation duration for *outside* is smaller than for *inside* because the actual search for a label requires more attention than utilizing the visual aid to estimate the rough position of the label.

H5: In the interior (*inside*), the gaze movement patterns for directional annotation are different from miniature and grid reference annotations. We expect that the gaze is along a search line for directional annotation. In contrast, we expect that participants search in the rectangular area of the cell in the miniature and grid reference annotations. Therefore, the angular differences between subsequent saccades are expected to be lower for directional annotation (i.e., saccades along a line) than for the other two types of annotation.

5.2 Stimuli and Task

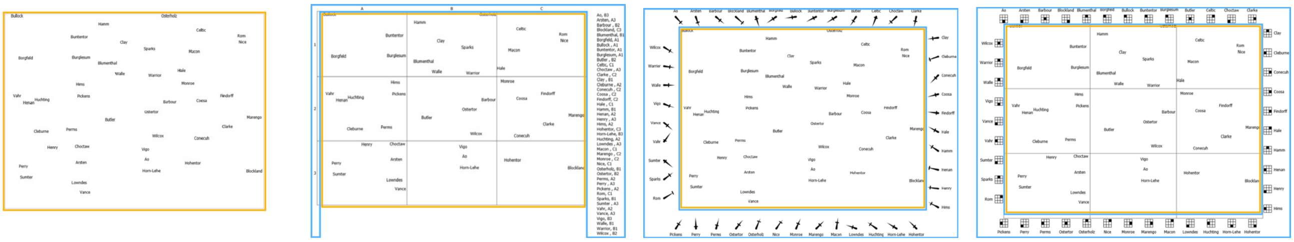
We generated artificial datasets to avoid that participants would have prior knowledge about a map used in the study. The maps contained the names of major cities of the US, France, Germany, and the United Kingdom. Each label was displayed with the same text representation parameters, i.e., font size, font face, text orientation, and color. We utilized a 3×3 grid structure to distribute the labels within the stimuli. In each grid cell, five different labels were randomly placed, leading to 45 labels per stimulus to achieve an initial equal distribution of labels. To avoid a too regular distribution, we randomly add or remove up to three labels to each cell. This process did not change the total number of 45 labels.

The size and position of the map area were identical for all stimuli. The generated images had a size of 1000×900 pixels. Figure 8 shows examples of the stimuli.

We generated 80 maps and applied each of the four methods, resulting in a total number of 320 stimuli. These numbers were adjusted based on the results of a pilot study, in order to meet the time constraints for the whole procedure. A subset of these stimuli was presented to each participant in a way that each individual stimulus was read at least by eight participants.

Overall, a participant was confronted with 80 stimuli (20 for each visualization method). We performed two-staged counterbalancing to compensate for learning and fatigue effects. By using two stages, the counterbalancing can be achieved with a smaller number of participants. First, the 80 stimuli of one participant were divided into four blocks each consisting of 20 images, containing visualizations from all four methods. The block order was counterbalanced. Second, each block was further subdivided into four sub-blocks (in counterbalanced order), each containing 5 stimuli generated with the same visualization technique. Their order within each sub-block was randomized.

The task that the participants had to perform was as follows: Find the specified label within the map. The label was shown to the participants before the stimulus was shown, and was varying from stimulus to stimulus. We asked the participants to perform the task as fast and accurately as possible, but with the focus being on high accuracy.



(a) Within-image annotation (WA) (b) Grid reference annotation (GA) (c) Directional annotation (DA) (d) Miniature annotation (MA)

Fig. 4. The rectangles indicate the AOIs used for the analysis of the eye-tracking data (blue for *outside* and orange for *inside*).

5.3 Environment Conditions and Technical Setup

The study was conducted in our institute's laboratory isolated from outside distractions. The room was artificially illuminated. Apart from the study participant, only one experiment operator was in the room.

The Tobii T60 XL system was used for eye tracking. The stimuli were presented on the 24-inch LCD screen of the eye tracker, with a resolution of 1920×1200 pixels. The participants were sitting at a distance of about 60 cm in front of the screen, which led to good calibration of the eye tracker. We used the standard filter parameters of the eye-tracking software: minimum of 10 pixels covering and a minimum of 30 ms fixation duration. The head of the participants was not fixated, and therefore the distance to the screen was not constant, resulting in a slightly varying visual angle for a constant distance in the stimulus. However, the influence of head motion was small; 1° of visual angle corresponds in our set up to approximately 35 pixels.

5.4 Participants

For this study, we recruited 32 participants (27 male and 5 female). The age of the subjects ranged from 20 to 32 years with an average age of 22.8 years. All participants were enrolled as students at our university: 29 of them were studying computer science or software engineering, the others had a major in communication engineering or aerospace engineering. Gender, age, and background were not considered confounding factors. All subjects had normal or corrected-to-normal vision, which was tested with a Snellen chart. Each session was designed to not exceed 60 minutes and each participant was compensated with EUR 10. We excluded the data of two participants due to erroneous eye-tracking recordings from technical problems, resulting in data of 30 participants remaining for the evaluation. Removing two participants did not affect the counterbalancing in this case, since each permutation was still tested at least once. Nevertheless, the number of participants who were at least reading a stimulus was reduced to seven.

5.5 Study Procedure

We applied a consistent operational procedure in the following order: First, the participants signed a consent form and were asked to provide some demographic information. Then, visual acuity was tested using a Snellen chart, and the participants were instructed about the task and went through a tutorial. The tutorial consisted of an explanation of each annotation method followed by example tasks that the participants had to solve.

During the actual task execution for the study, the presentation of a stimulus was decoupled into two parts: first, only the name of the target label was presented, and after pressing a key, the stimulus itself. The subjects were not allowed to use auxiliary means, like the mouse or their fingers, during the search for the correct target, since this could affect the performance of directional annotation. After pressing a key, all measurements were stopped, and participants were able to select the found target using the mouse. Afterward, the next stimulus was presented. After the execution of the task, the participants were asked to fill out a questionnaire.

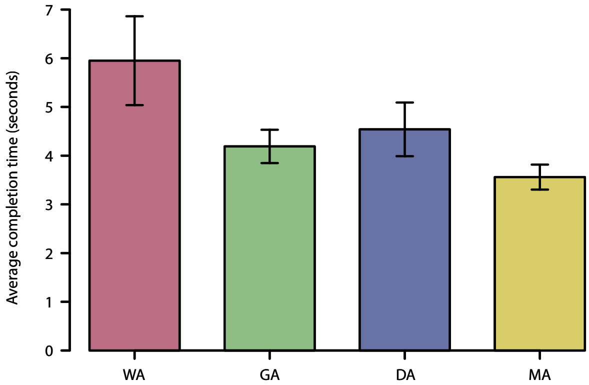


Fig. 5. Completion times for within-image annotation (WA), grid reference annotation (GA), directional annotation (DA), and miniature annotation (MA): average times along with error bars that show the standard error of the means (SEM).

6 Evaluation

This section presents the statistical evaluation of task performance, the analysis of the recorded eye-tracking data, and the results of the subjective feedback obtained through the questionnaire.

For our statistical evaluations, we performed a Kruskal-Wallis test, followed by pairwise Wilcoxon tests for post-hoc comparison when the initial test indicated any significance. The results are presented in the form of $\chi^2(d) = e$, where d is the degree of freedom and e the associated chi-square value. To account for multiple comparisons, we applied a Holm-Bonferroni p-value adjustment. Performing an ANOVA was in many cases not possible because the initial tests for normal distribution (Shapiro-Wilk test) and homogeneity of variance (Bartlett test) were not successful, even after logarithmic transformation. In the cases where ANOVA could be used, the results are reported as $F(d,r) = a$, where r corresponds to the residuals and a to the F-value. The post-hoc analyses were then conducted with pairwise t-tests.

6.1 Task Performance

We performed our evaluation of task performance based on completion times. The best results were achieved with miniature annotation (MA) followed by grid reference annotation (GA), directional annotation (DA), and within-image annotation (WA) at last. The participants performed the task on average in 3.56 s (MA), 4.19 s (GA), 4.54 s (DA), and 5.95 s (WA), see Figure 5. The average task completion time over all techniques was 4.56 s. Expressed in percentage compared to the baseline WA, MA was 40.2% faster, GA 29.6%, and DA 23.7%.

The statistical test revealed significant differences between some of the methods ($\chi^2(3) = 45.42; p < 0.001$). The results of the post-hoc analysis are summarized in Table 1.

There are significant differences between WA and all other annotation methods ($p < 0.001$). Furthermore, we can find significant results between MA-GA ($p = 0.01$) and MA-DA ($p < 0.001$). There are no

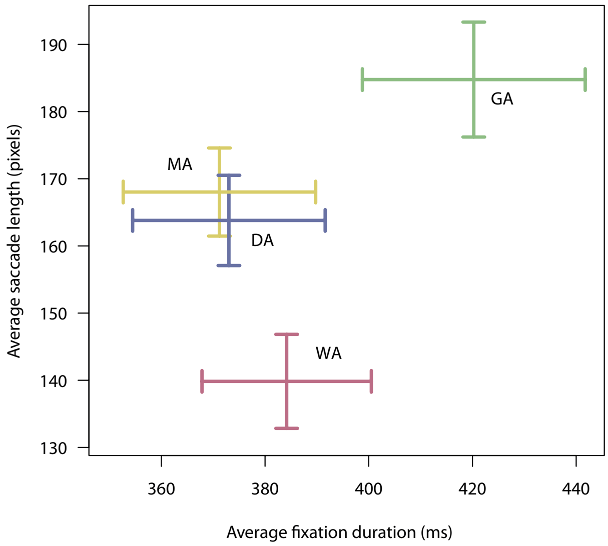


Fig. 6. Average fixation duration (x-axis) and average saccade length (y-axis) for all four annotation methods. Error bars show the standard error of the means (SEM) for the eye-tracking data.

Table 1. P-values of post-hoc comparisons of completion times.

Method	GA	MA	DA
MA	0.01	-	-
DA	0.26	<0.001	-
WA	<0.001	<0.001	<0.001

significant differences between DA and GA.

These findings confirm hypothesis H1: as expected, visual search support improves task performance compared to the baseline WA. However, we can only partially accept hypothesis H2, because there is no significant difference between GA and DA, but MA exhibits a better perform than both.

6.2 Eye-Tracking Data Analysis

In addition to a traditional task performance analysis, gaze data are used for a more detailed evaluation with inference of, e.g., visual interaction patterns of subjects, their workload, or cognitive processes while they were solving a task. The first point is often based on the evaluation of AOI sequences, while the latter two utilize different metrics for quantification.

To this end, we utilize the average fixation duration and average saccade length, which are the most common eye-tracking metrics [14]. The average saccade length is used to discriminate between explorative and localized eye movement. Larger values indicate a more explorative movement, while smaller values could correspond to increasing task difficulty, when information needs to be collected for the enrichment of the cognitive processes. The average fixation duration was used as an indicator for the cognitive processing depth. High values mean typically that a participant investigated an area more carefully, since the task has demanded it or due to a high degree of complexity. Lower values, occurring in a restricted area in combination with a small average saccade length, could be an indicator for stress. These measures were previously used by Netzel et al. [24] in a similar context. They investigated effects of different line direction encoding of trajectories on the gaze behavior of participants.

We begin the analysis of the eye-tracking results with a statistical evaluation of average fixation duration and average saccade length as dependent quantities. Figure 6 summarizes the results. ANOVA indicated that there is a significant impact of annotation type on average fixation duration ($F(3,116) = 5.63$; $p = 0.002$) and saccade length ($F(3,116) = 25.12$; $p < 0.001$). In Figure 6, we can see that DA and MA build a cluster due to their similar design. Furthermore, GA and WA are located in separated areas. The post-hoc analysis of the fixation duration reveals that there are differences between GA and all the other methods ($p < 0.05$). For the saccade length, all pairwise combinations show significant differences ($p < 0.007$), except between MA-DA. This confirms hypothesis H3, since MA, DA, and GA in fact lead to higher average saccade lengths.

With the AOIs for the outer region (*outside*) and inner region (*inside*), we can analyze visual solution strategies; see Figure 4 for illustrations of the AOIs. We will show that there are two phases of visual search for GA, DA, and MA: in the first phase, subjects focus on the annotation area *outside*, and switch then to *inside* in order to find the required label in the second phase. To extract these two phases, we looked at the mean timestamps of the fixations, separately for each AOI. The mean timestamp is here the sum of all timestamps within an AOI divided by the number of fixations in the AOI. Figure 7(a) shows that the two phases are clearly separated: fixations in *outside* happen much earlier than fixations in *inside*. For all annotation methods except for WA, participants start searching in the outer regions because the mean timestamp of *outside* has a significantly lower value. For WA, there is no first phase and therefore no *outside*. This confirms the first part of hypothesis H4.

For the second part of H4, we analyze the average fixation durations, separately for the two AOIs. Figure 7(b) shows the statistics plots. We recognize the impact of the two phases. During the second phase, the average fixation duration is higher, since this corresponds to the search of the actual label by inspecting labels in a target area. For the first phase, GA shows higher fixation duration in comparison to MA and DA. For GA, participants not only had to find the reference to the grid cell where the label is located, but also had to memorize the grid coordinates. Therefore, they had to focus more. For MA and DA, this is not the case because participants quickly identify the rough position and start the transition into the second phase, searching for the actual label. The significance test for *outside* indicates differences ($\chi^2(2) = 19.67; p < 0.001$). The post-hoc test shows that there are only differences between GA-MA ($p < 0.001$) and GA-DA ($p < 0.001$). GA achieved an average fixation duration of 406.17 ms, MA 311.92 ms, and DA 303.81 ms. For *inside*, there are also significant results ($\chi^2(3) = 25.89; p < 0.001$). In particular, there are differences between: GA-WA, MA-WA, and DA-WA with $p < 0.001$. GA achieved an average fixation duration of 451.76 ms for *inside*, MA 485.58 ms, DA 476.95 ms, and WA 384.08 ms. Based on these outcomes, we can confirm the second part of hypothesis H4 as well.

For hypothesis H5, we analyzed the average direction change along scanpaths at each fixation in the second phase. We assume that searching along a line results in smaller angular differences than searching in a rectangular area. In order to handle inversion of search direction (180° turns), we use the smaller angle of the change in direction.

The statistical tests indicate significant results with respect to the annotation type ($F(3,116) = 34.84; p < 0.001$) and the post-hoc tests report differences for the following pairs: DA-WA, DA-GA, DA-MA, and GA-WA with $p < 0.001$. The average change in direction was 41.74° (GA), 39.34° (MA), 31.39° (DA), and 36.82° (WA); see Figure 7(c). For DA, we can see participants were changing the angle of direction less strongly, resulting in a more directed eye movement compared to the other methods. This indicates that the design of DA has an impact on the visual search behavior of the subjects, and therefore we can confirm hypothesis H5.

6.3 Qualitative and Subjective Evaluation

For the qualitative and subjective evaluation, we asked the participants to fill out a questionnaire with the following questions regarding the different annotation methods:

1. Have you used a search strategy? If yes, please state it.
2. Have you used the annotations for orientation purposes?
3. Have you searched the annotations for a label in alphabetical order? (only for MA, GA, and DA)

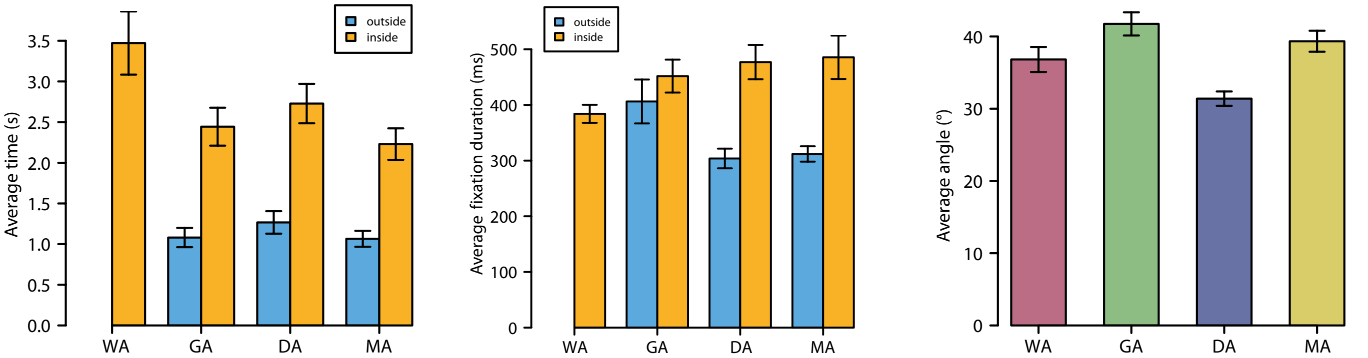


Fig. 7. Diagrams used to identify different subject behavior. (a) There are two phases during the solution of the task. (b) The fixation behavior is also different between the phases. The average direction change of saccades along scanpaths for all four methods is shown in (c).

4. Have you used the distance line marker for orientation purposes? (only for DA)
5. Which method allowed you to perform the visual search task most quickly?
6. Which method allowed you to identify the grid cell that contains the target label most quickly: MA or GA?

For questions 2 and 4, we used a Likert scale from 1 (never) to 4 (always). For questions 1 and 3, we counted the yes and no answers. For questions 5 and 6, we counted the votes for each individual map annotation method. In the following, the answers will be presented with respect to the four methods.

Within-image annotation (WA): 19 participants stated that they used a search strategy. The most common strategies were horizontal or vertical scanning, or starting in the middle of the screen and progressing outward in a spiral.

Grid reference annotation (GA): 25 participants used a search strategy: They started at the annotations to search for the label and the associated grid coordinates, and then moved to the grid cell. This correlates with the average answer for question 2, which is 3.7 on the Likert scale. Also, 23 participants were looking for the target label in the legend in alphabetical order.

Directional annotation (DA): Here, 26 subjects used a search strategy. First, they were searching for the directional annotation at the border (23 in alphabetical order), and next, they followed the arrow. An average Likert value of 3.73 was achieved for question 2, and a Likert value of 2.8 for question 4 (use of the distance line marker).

Miniature annotation (MA): A common search strategy was used by 25 subjects. Similar to DA, they first searched in the annotations (average Likert value of 3.8) for the correct label, and then jumped to the associated grid cell. 23 subjects stated that they benefited from the alphabetical order.

General questions: Questions 5 and 6 were dedicated to summarize the impression of the subjects about the map annotation methods. 21 stated that they could perform the search task fastest while using MA, followed by GA (7 votes) and DA (2 votes). Question 6 was supposed to compare MA and GA, since they are both grid-based. Here, the participants favored MA with 23 votes.

7 Discussion

Based on the results of the statistical evaluation in the previous section, we are able to confirm several hypotheses from Section 5.1. We can confirm H1, since we could verify that there are significant differences between WA and the other three map annotation methods. Participants were 23–40% faster in locating a specific label using the annotation methods with visual search support.

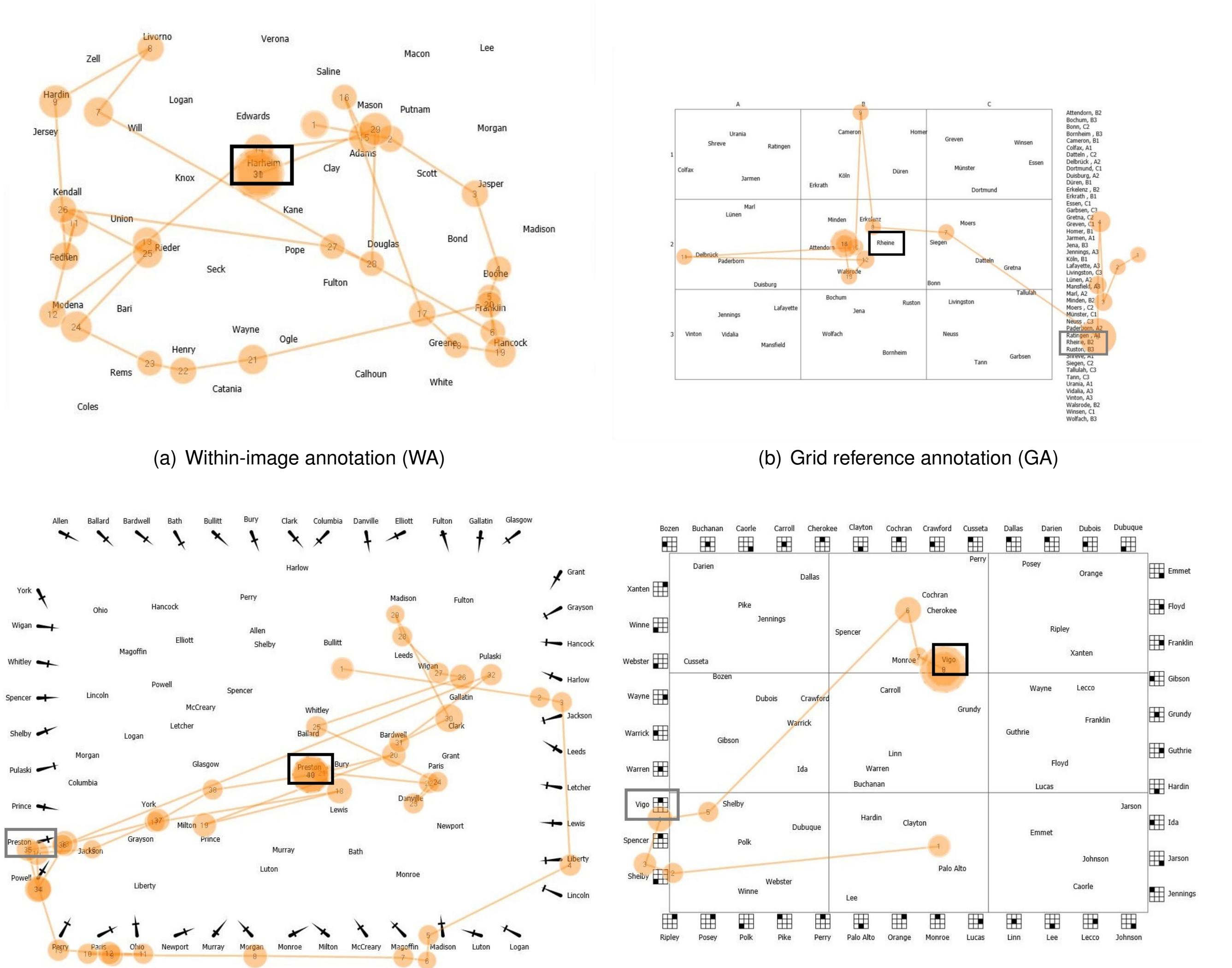
We can confirm hypothesis H2 in parts. The post-hoc analysis of completion times allows us to rank MA better than GA and DA. From our theoretical considerations from Section 4, the following potential outcomes could have been expected: On the one hand, DA could have performed better than MA, since DA provides directional and distance information, and therefore the search area should be more focused compared to the size of a grid cell. On the other hand, the performance of MA and GA could have been better than that of DA. Here, subjects might be more familiar with grid-based approaches, as they are used commonly in maps and atlases. Perceptual issues could also be a reason for the achieved performance of DA. It is well known that direction and distance estimation especially for long distances depend on many factors, e.g., arrow length or the number of distractor elements in the target area [4, 13, 20, 22]. For example, an arrow points to a specific point in a cluster. If the arrow is drawn with the full length to the target, it can be identified accurately. Shortening the arrow would introduce an error for the perceived direction in which the arrow is pointing. As a result, one of the distractors might be perceived as the target.

Concerning the eye-tracking evaluation, we were able to accept hypothesis H3. This is supported by the statistical tests and also visible in Figure 6. Furthermore, we were able to extract two phases during the solution of the task when annotations were provided. The first phase corresponds to the lookup of the target label *outside* of the map, while during the second phase, the participants were searching *inside* the map (see Figure 3). Since there is a separation between the two phases, we could further investigate the behavior characteristics by comparing the average fixation durations for both phases. During the first phase, GA exhibits a higher fixation duration compared to the other methods, which could indicate that subjects required more time for reading and memorizing the respective coordinates (e.g., “C2”).

During the second phase, all participants exhibited basically the same behavior: They were looking for the target label at the rough position provided by the position hint. Therefore, the average fixation durations look alike, since reading a label and comparing it mentally with the target label should take for all annotation methods approximately the same amount of time.

We were also able to show that the design of the annotation method influences the behavior of the subjects (hypotheses H4 and H5). This is illustrated in Figure 1. Here, the attention is distributed differently for each annotation method. Looking at scanpaths also indicates different solution behaviors. Some example scanpaths are shown in Figure 8. Furthermore, we were interested in if the participants followed the arrow that points in the direction of the label for DA. Our evaluation of the average relative direction change of the scanpaths shows that DA has the lowest values, indicating that subjects tend to move their eyes more linearly compared to the other methods.

The main drawback of miniature and directional annotations con-



(a) Within-image annotation (WA)

(b) Grid reference annotation (GA)

(c) Directional annotation (DA)

(d) Miniature annotation (MA)

Fig. 8. Scanpath examples of different participant behaviors for all four annotation methods tested in the study. Random search is shown in (a). In (b), there is first a lookup in the list followed by identifying the grid cell, and finally searching for the label. In (c), first the annotation was used to get information about the direction and distance of the target, followed by a search in the respective direction. The initial step of (c) is also present in (d). In the second part, a search in the target cell is performed. The rectangles indicate the target label.

cerns the available space for placing the annotations at the border of the map. With an increasing number of labels, the size of the annotations at the border must be scaled down in order to avoid overlapping labels, leading to symbols that are hard to recognize. Although grid reference annotation also needs more space to show an increasing number of labels, the scalability issue is less pronounced because the annotation region is independent from the size of the map.

To increase the number of labels around the map, one could think of adding additional layers of annotations, instead of a single one as used in this study. The label has a huge impact on the space used by an annotation. Longer labels would require more horizontal space and, therefore, fewer annotations can be placed horizontally. Shortening their lengths, e.g., by using the shortest prefixes, would allow us to increase the number of labels. However, there is still a limit on how many layers and labels can be placed around the map. For miniature annotations, a hybrid approach can overcome this problem completely. Here, the annotations can be displayed separately on additional pages similar to GA. This is illustrated in Figure 9. We did not test a hybrid approach in this study, since we wanted to isolate the effect of the different visual representations of positional information in our developed annotation methods. To this end, we used the same layout to arrange the annotations around the map.

As for most laboratory user studies, we also had to limit the design space and range of independent factors due to constraints for the total procedure time. First, we used a constant number of labels (45 labels within each stimulus). Assuming that the lookup of a label is in general fast enough due to, e.g., an alphabetical ordering of the labels, we did not test the effect of different label numbers in this study. However, the number of used labels was sufficient to test the ability of the annotations to support visual search in maps. Second, we tested one grid resolution (3×3) for MA and GA. Even with these limitations, we concluded from the results of a pilot study that we could only test 80 stimuli with each subject. Nevertheless, we have obtained baseline measurements for the comparison with future study results with a subset of annotation methods and an extended parameter space (number of labels, grid size, etc.). Another interesting factor that could influence the results is the use of realistic instead of artificial maps, since details like roads, contour lines, or other features could complicate or support visual search.

Although we could obtain significant results from this study, the differences in task solution times are only within a couple of seconds. Therefore, based on the tested parameter space, an infrequent

Fig. 9. Hybrid approach between GA and MA. Labels at the right side are combined with miniature representations.

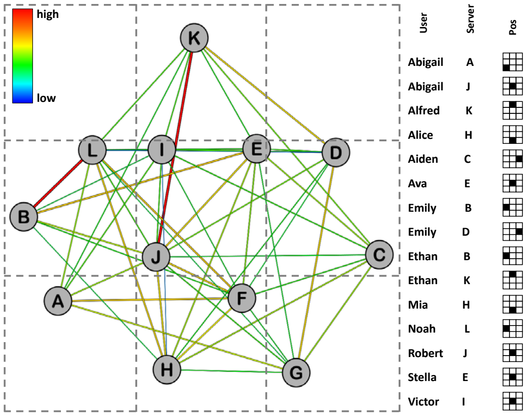


Fig. 10. Graph visualization showing the connections between computer servers. Colors encode the connection bandwidth. The annotation shows which user is assigned to which server and the respective position in the network topology.

use would not benefit from annotations in practice. Furthermore, a potential drawback of MA and DA compared to GA is the loss of the ability to quickly communicate positional information verbally.

As a general guideline, we conjecture that miniature annotations are a suitable concept for supporting visual search. It seems to be easier to follow a miniature representation of the domain compared to textual coordinates as in the case of GA. Furthermore, the search in the target area seems to be faster compared to DA. Potential applications that benefit from such annotations might not be restricted to maps and related scenarios. Possible use cases are discussed in the next section.

8 Future Directions

We think that annotation concepts could support other visualization applications besides maps. There are many other applications that require visual search. Especially settings that do not allow user interaction can benefit from the evaluated visual search support, e.g., any kind of printed visualization. But even with non-static output devices like displays, it is not always possible or desirable to enable user interaction. For instance, in a collaborative visualization setting with multiple users standing in front of a large display like a Powerwall, interaction might not be provided for all users to avoid conflicts. Even in an interactive visualization scenario, it might be useful to use annotations to point to interesting areas in the visualization. For instance, in applications with multiple coordinated views, interaction is typically required to establish the connection between the views, e.g., selecting elements in one view highlights them in the other views. Using annotations, a link between the views can already be provided without interaction. For example, extrema in a plot can be augmented with annotations pointing to their locations in another view. This could reduce the number of interactions when exploring the data.

A visualization could also benefit from visual search support if it exhibits a larger number of points of interest. Similar to reading maps, it could be easier for the viewer to find a specific point of interest when the search is supported with an annotation. For example, for graph visualizations, the annotation methods can help find specific nodes. Figure 10 shows the example of a computer network where different users are connected to servers in the network. The annotation list is ordered according to the username and shows the position of the respective server they are connected to. This approach supports finding the position of the server in the network to which a certain user is connected and can be used if interaction might not be available, e.g., on a large overview display in the control room of a network provider. In the case of 2D scalar fields, which are often visualized with color maps, the annotations can help find important points, e.g., points with local maximum or minimum values. In matrix visualizations, different clusters are often of interest. Annotations can be used to guide the viewer to these clusters.

In general, the annotation approaches should work with all types of 2D visualization, although not all of them might benefit from them. An extension to 3D visualization should be possible, but typical issues of 3D visualizations like occlusion and perspective distortions make this more difficult. In all cases, the presented annotation methods have the advantage that the visualization is not cluttered by overlaying it with additional graphical objects. However, miniature representations of the visualization domain seem to be most suitable to provide visual search support.

9 Conclusion

In this paper, we have evaluated three annotation methods that support visual search in maps: miniature annotation (MA), directional annotation (DA), and grid reference annotation (GA). As baseline, we used within-image annotation (WA). Using such annotations is beneficial if traditional encoding mechanisms, like color, cannot be applied, since they are already used to encode another information. The results show that each of the three approaches (MA, DA, GA) leads to significantly reduced completion times. We have found significant differences between the three annotation methods: the lowest and therefore the best completion times were achieved by using MA, followed by GA, and DA. Correctness was not evaluated, since the task could be solved accurately in any case. By evaluating the eye-tracking data, we were able to show that participants solved the task in two phases: first, they used the provided annotation to get a hint for the position of the label; then, they performed the actual search. We could also show that the fixation duration changed after the transition into phase two. For phase one, GA had the highest average fixation duration, indicating deep cognitive processing. In phase two, in contrast, the fixation durations are approximately the same because similar search is performed for all annotation techniques. However, we have seen that DA triggers search along straight lines, whereas the other techniques lead to an area-oriented search.

Overall, considering task performance and eye-tracking data, the MA technique seems to be the best annotation method. Generalizing our findings, we think that the use of visual miniature annotations is likely to be beneficial for other visualization applications as well. We are convinced that the application of annotation concepts to other visualizations besides maps is of interest for future work.

Acknowledgments

We would like to thank the German Research Foundation (DFG) for financial support within project B01 of SFB/Transregio 161. The used maps are courtesy of OpenStreetMap Foundation and are licensed under the Creative Commons Attribution-ShareAlike 2.0 license (please see `http://www.openstreetmap.org/copyright`).

REFERENCES

[1] P. Baudisch and R. Rosenholtz. Halo: A technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 481–488, 2003.

[2] M. A. Bekos, M. Kaufmann, A. Symvonis, and A. Wolff. Boundary labeling: Models and efficient algorithms for rectangular maps. *Computational Geometry*, 36(3):215 – 236, 2007.

[3] M. Birsak, P. Musialski, P. Wonka, and M. Wimmer. Automatic generation of tourist brochures. *Computer Graphics Forum*, 33(2):449–458, 2014.

[4] H. Bouma and J. Andriessen. Perceived orientation of isolated line segments. *Vision Research*, 8(5):493 – 507, 1968.

[5] M. Burch and S. Diehl. TimeRadarTrees: Visualizing dynamic compound digraphs. *Computer Graphics Forum*, 27(3):823–830, 2008.

[6] M. Burch, C. Vehlow, N. Konevtsova, and D. Weiskopf. Evaluating partially drawn links for directed graph edges. In *Proceedings of Graph Drawing (GD)*, pages 226–237, 2011.

[7] H. A. do Nascimento and P. Eades. User hints for map labeling. *Journal of Visual Languages & Computing*, 19(1):39–74, 2008.

[8] J. Dykes, J. Wood, and A. Slingsby. Rethinking map legends with visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):890–899, 2010.

[9] C. Ebert. Correspondence visualization techniques for analyzing and evaluating software measures. *IEEE Transactions on Software Engineering*, 18(11):1029–1034, 1992.

[10] I. Golebiowska. Legend layouts for thematic maps: A case study integrating usability metrics with the thinking aloud method. *The Cartographic Journal*, 52(1):28–40, 2015.

[11] C. G. Healey and J. T. Enns. Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Transactions on Visualization and Computer Graphics*, 5(2):145–167, 1999.

[12] C. G. Healey and J. T. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2012.

[13] R. F. Hess, S. C. Dakin, and N. Kapoor. The foveal 'crowding' effect: physics or physiology? *Vision Research*, 40(4):365–370, 2000.

[14] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 2011.

[15] D. Holten, P. Isenberg, J. J. van Wijk, and J. Fekete. An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, pages 195–202, 2011.

[16] P. Kindermann, B. Niedermann, I. Rutter, M. Schaefer, A. Schulz, and A. Wolff. Multi-sided boundary labeling. *Algorithmica*, 74:1–34, 2015.

[17] K. Kurzhals, B. D. Fisher, M. Burch, and D. Weiskopf. Evaluating visual analytics with eye tracking. In *Proceedings of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV)*, pages 61–69, 2014.

[18] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.

[19] L. Lischke, S. Mayer, K. Wolf, N. Henze, A. Schmidt, S. Leifert, and H. Reiterer. Using space: Effect of display size on users' search performance. In *Proceedings of the ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA)*, pages 1845–1850, 2015.

[20] J. M. Loomis, J. A. Da Silva, J. W. Philbeck, and S. S. Fukusima. Visual perception of location and distance. *Current Directions in Psychological Science*, 5(3):72–77, 1996.

[21] M. Maciel, B. Meiguins, R. Loureno, A. Meiguins, and P. Godinho. The impact of multiple coordinated views on the visual data exploration and analysis. In *Proceedings of the International Conference on Information Visualisation (IV)*, pages 113–119, 2008.

[22] D. McCready. On size, distance, and visual angle perception. *Perception & Psychophysics*, 37(4):323–334, 1985.

[23] J. Müller, R. Rädle, H.-C. Jetter, and H. Reiterer. An experimental comparison of vertical and horizontal dynamic peephole navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1523–1526, 2015.

[24] R. Netzel, M. Burch, and D. Weiskopf. Comparative eye tracking study on node-link visualizations of trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2221–2230, 2014.

[25] A. H. Robinson. *The Look of Maps: An Examination of Cartographic Design*. Esri Press, 2010. 1st reprint.

[26] A. H. Robinson, J. L. Morrison, P. C. Muehrcke, A. J. Kimerling, and S. C. Guptill. *Elements of Cartography*. Hardcover, 3rd edition, 1995.

[27] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. Feature congestion: A measure of display clutter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 761–770, 2005.

[28] G. S. Srivastava. *An Introduction to Geoinformatics*. McGraw-Hill Education, 2014.

[29] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a WIM: Interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 265–272, 1995.

[30] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[31] E. Tufte. *Beautiful Evidence*. Graphics Press, 2006.

[32] J. A. Tyner. *Principles of Map Design*. Guilford Press, 2010.

[33] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pages 110–119, 2000.

[34] C. Ware. *Visual Thinking for Design*. Morgan Kaufmann, 2008.

[35] J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal*, 47(2):117–129, 2010.

[36] H.-Y. Wu, S.-H. Poon, S. Takahashi, M. Arikawa, C.-C. Lin, and H.-C. Yen. Designing and annotating metro maps with loop lines. In *Proceedings of the International Conference on Information Visualisation (IV)*, pages 9–14, 2015.

[37] H.-Y. Wu, S. Takahashi, C.-C. Lin, and H.-C. Yen. Voronoi-based label placement for metro maps. In *Proceedings of the International Conference on Information Visualisation (IV)*, pages 96–101, 2013.