PICOZOOM:
A CONTEXT SENSITIVE MULTIMODAL ZOOMING INTERFACE

Anonymous ICME submission

ABSTRACT

This paper introduces a novel zooming interface deploying a pico projector that, instead of a second visual display, leverages audioscapes for contextual information. The technique enhances current flashlight metaphor approaches, supporting flexible usage within the domain of spatial augmented reality to focus on object or environment-related details. Within a user study we focused on quantifying the projection limitations related to depiction of details through the pico projector and validated the interaction approach. The quantified results of the study correlate pixel density, detail and proximity, which can greatly aid to design more effective, legible zooming interfaces for pico projectors - the study can form an example testbed that can be applied well for testing aberrations with other projectors. Furthermore, users rated the zooming technique using audioscapes well, showing the validity of the approach. The studies form the foundation for extending our work by detailing out the audio-visual approach and looking more closely in the role of real-world features on interpreting projected content.

Index Terms— Mobile projection, zooming interface, proxemics, interaction techniques, augmented, and virtual realities

1. INTRODUCTION

Small mobile (pico) projectors form a compelling interface device that can enable quasi ad-hoc interaction with the world around us. Light and versatile, users can easily take them around. Yet, they remain limited in their projection abilities. It is exactly the potential and limitations of these projectors that need to be reflected in the design of interactive visualization applications. Within this paper, we follow a structural approach to designing effective techniques, by looking at the strength and limitations of pico projection characteristics as well as finding new methods for providing contextual information. In particular, we look at how the usage of the established flashlight metaphor [1] [2] [3] can be advanced to allow for a more flexible usage.

The targeted use case of our system falls within the domain of spatial augmented reality (SAR), revealing additional information about real world objects or environments by projecting spatially relevant information. Thereby, our approach draws upon zooming principles from proxemic interaction [4]: when moving closer to an object, more details are revealed, while moving away provides an overview. Yet, previous work falls short of pinpointing two relevant issues. First, how projection characteristics related to perceiving detail at at a legibility and visibility level can be used to their full extent while being affected by projection aberrations. Second, how a single mobile projector can be deployed while retaining context. Context can be an important factor to improve interpreting and relating details in a zooming interface, yet, it is difficult to achieve simultaneous contextual and detailed information by using a single projector. Within this paper, we will focus on investigating these issues through two studies.

Firstly, we report on the visibility and legibility limitations of using a typical mobile projector by performing an experiment targeting textual and graphical information. Secondly, we explore how contextual information can be provided through by advancing the flashlight approach with audio (audioscapes), and briefly look at the role of physical objects as contextual cue.

2. RELATED WORK

The provision of contextual information while zooming into details relates to the usage of focus plus context (f+c) visualization techniques [5], as well as physical f+c display systems. Previously, interface designers have focused on the combination of lower and higher resolution display devices to create an effective yet spatially constrained setup [6], while approaches using multiple mobile projectors are more flexible to convey focus and context [2]. Alternatively, a mobile projector (focus, detail) can also be moved relative to a static large projection (context) [1]. The latter system leverages an important aspect: the closer the projector gets to a projection surface, the higher the pixel density (PPI) and brightness of the projection and thus the higher the potential detail that can be displayed. Though [1] introduced different levels of detail that seemed to reflect the pixel density, no implementation details were provided: it is unknown how well the graphical and textual details were optimized for visibility and legibility while making full use of potential detail. It is this issue we targeted in the first part of our study.

Our interface approach extends the usage of the flashlight metaphor as applied in mobile augmented reality inter-
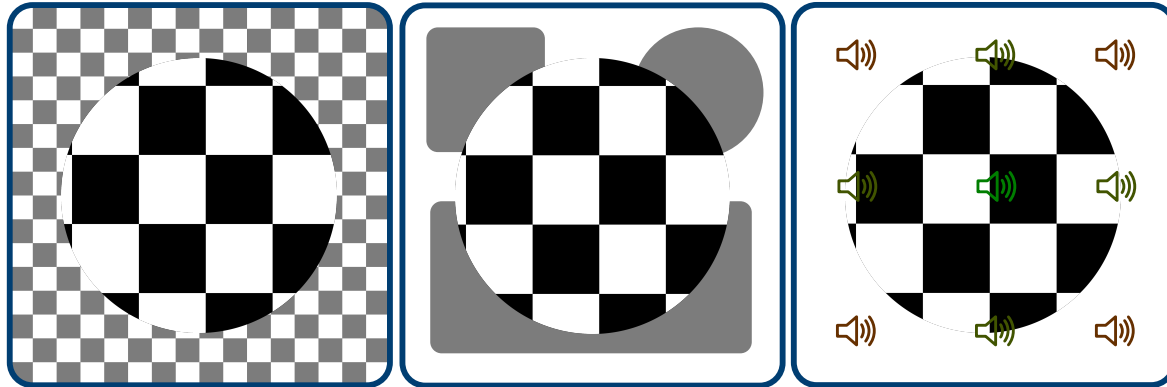
Fig. 1. The left sketch depicts a traditional visual focus and context metaphor (black - focus / grey - context) and the proposed novel concept of alternative context, using real world objects and audioscapes, is illustrated in the middle and right drafts.

faces [7] [8] [9] [10] [11]. These approaches have not studied closely the density (detail) and contextual information aspects as we are using within our approach.

Also, audioscapes (3D audio spaces generally ordered through a grid structure) have been used before in audio-only environments to communicate different levels of information in a spatial setting [12]. However, the interplay between such audioscapes and visual information has not been studied well. Finally, while studies in Augmented Reality by their very nature focus on relating overlaid information to the real world, most of these studies rather deal with issues such as, among others, depth perception or occlusion [13]: the actual role of the real world context while exploring specific (sub-) parts of the virtual content is barely studied. It is this issue we addressed in the second part of our study through the development and evaluation of alternative context in 2 application scenarios.

3. CONCEPT AND IMPLEMENTATION

The interplay between the flashlight metaphor and a zooming focus and context metaphor within a single mobile projector introduces the need for an alternative multimodal context. For instance using a zooming gesture with a single projector works well, but at detail level - if the user focuses a region of interests - the context is missing. Therefore, we propose 2 approaches of an alternative multimodal context (see Figure 1).

As first approach we introduce real world objects as context - so that the user can investigate an object without losing the context while, for instance, zooming in. The second approach enables pre-defined audioscapes as context that indicate the location of the projection while the user focuses an object. The dynamic sound volume is estimated by the distance between the projector and the audioscapes. For instance, if the user is far away from the projection area/object the volume of all audioscapes is almost the same and a mul-

timodal feedback of the whole scene is provided. If the user moves the projector next to the projection area/object to investigate further details, neighboring audioscapes adjust their volume with regard to the projector's position. The farther the position of the projector is away the lower the volume. Using audioscapes offers an additional benefit for focus and context metaphors, while almost all focus and context and zooming interfaces concentrate on a single channel feedback [1] [2], our second approach supports not only an alternative concept of contextual information, but additionally the audioscapes emphasized a multimodal focus interface. Furthermore, the concept of mulitmodal context is transferable to other augmented reality applications.

In our user study, we focused on quantifying the visual limits of mobile projection based on proximity, and explored alternative contextual information provision methods that support flexible usage of a single projector interface. While visual detail can be defined by addressing the visual acuity of the eye (for example using the Snellen test or Landolt rings [14]), in reality the projection includes aberrations that limit visibility of graphics and legibility of text. Hence, it is highly useful to obtain quantified information to design effective visualization techniques accordingly, to enable to maximum level of detail based on proximity while securing visibility and legibility.

We created a visual testbed (see Figure 2) that makes use of Seeser M2 laser projector (25 ANSI lumen, WVGA resolution). We used a laser projector due to its ability to project sharp, in focus imaging irrespective of the projection distance. Not having this ability would have resulted in continuous adjustments of the projector focus, disturbing the study considerably. The projector and the canvas was tracked using an ART tracking system, delivering millimeter accuracy. Radial distortion was removed and the optical center computed. The user would place her head on a head support that would fix the eyes to the center-height of the projection canvas at exactly 850 mm distance. Full screen (overview, context) and maxi-
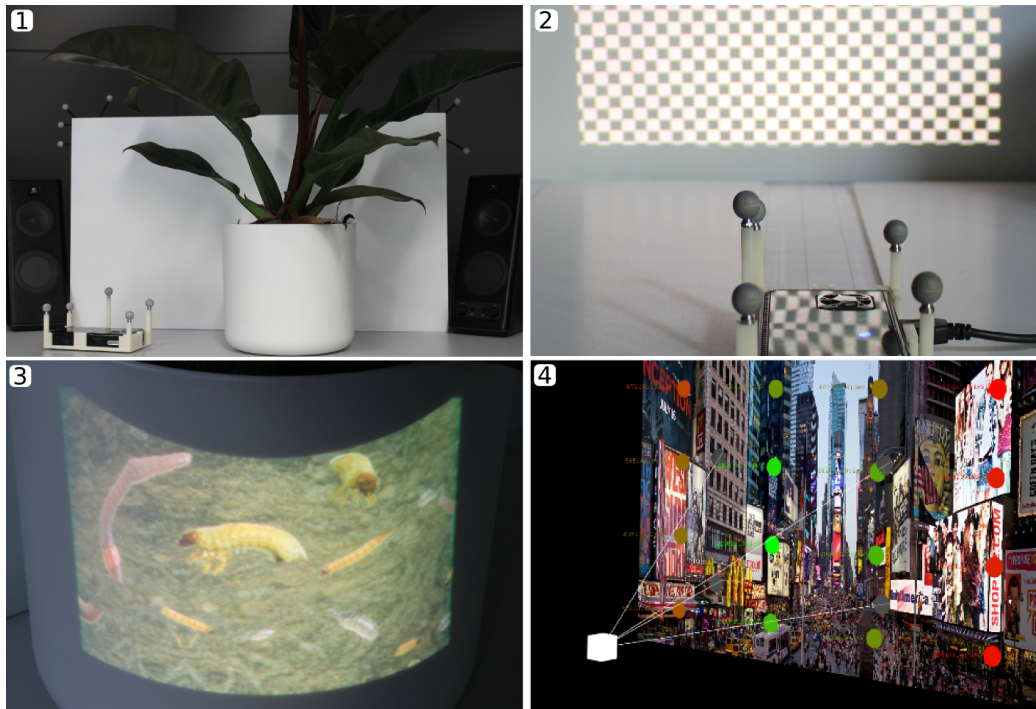
Fig. 2. System overview: Experimental setup showing projector, canvas and real world object for plant scenario (1). The designed user study to estimate visibility and legibility issues (2). Augmentation of different plant life, varying content based on the mesofauna principle in correlation to proximity (3). A projection based image viewer with contextual audioscapes (colored dots for illustrative purposes) and zooming capabilities (4).

mal focus (detail) was reachable by hand within 750 mm to 100 mm screen distance. Using the projector, different visual patterns (lines and checkerboard patterns) and fonts at different scales were displayed at a canvas (sized 600 x 335 mm). We varied the pattern scales between 0.6, 0.7, 0.8, 1.25, 2.5 and 10 mm (line and checkerboard patterns) and 2, 3, 4, 5, 6 and 12 mm respectively (font size), providing a flexible number of transitions. We used randomly generated letters and for each size asked the users to spell them out. The setup can be seen as a testbed that can be replicated for other projector setups too: the results of our study are bound to the specific projector used, hence, while other projectors may have similar aberrations, to fine-tune an interface it will be necessary to calibrate for the specific projector being used.

Next, to study contextual cues we also implemented a zooming interface deploying the previously mentioned augmented reality (AR) flashlight metaphor. We extended the visual testbed to display different visual detail levels based on proximity, we deployed an information management method driven either by the inherent image density, or by varying the content. It is important to note that in our initial approach we did not yet use the results of the visual testbed, but rather looked into the effects of different kinds of contexts. While we briefly looked into the usage of real world objects as contextual reference for AR, our main focus was on using au-

dioscapes.

First, we created a simple SAR scenario to explore 3D visual (physical) contextual information. This was achieved by displaying different kinds of information about the plants' mesofauna, correlating size of insects with proximity, thus showing the smallest insects when closest. This information was displayed (and corrected) on the surface of a pot of a real plant: thus, users could look "inside" the pot from different sides to explore the life within. We introduced 3 levels of details, which faded in at 700, 500 and 300 mm. The three different zones simply represent three stages that are ergonomically well reachable by moving the projector close to the human body up to arm's length. The levels of detail thereby followed the simple principle of detail based on proxemity: the closer to the pot, the more detailed the information was presented. It is important to note that the representations at the three levels changed: at 500 and 300 mm additional content was blend in, instead of relying on pixel-wise detail only.

Second, we created an audioscape scenario, using a cityscape of downtown New York (14875 x 5547 pixels) the closer the user moved towards the canvas, the more details could be observed, while audio provided the user with contextual and object-centered feedback through two loudspeakers besides the projection canvas. In contrast to the plant study, we did not add additional content when the user got close to

the canvas - in this application, we solely relied on pixel density. As an example, consider hearing a church bell vaguely while observing a sub-part of the scene: it provides the user with a cue that a church may be close by. Proximity and focal point defined the sound intensity of an associated audio file through a spatially constrained audio region called a *bubble*. When a user looked at the overall image, ambient sound was played (the mixture of all sound bubbles). While a user got closer to the projection canvas, the sound volume of a specific bubble increased, while the surrounding bubbles volume decreased based on proximity. The cityscape scenario contained 16 audio bubbles evenly distributed over the canvas.

4. USER STUDY

We set up an explorative user study to quantify visibility and legibility issues, and to evaluate the zooming interface approach. In particular, the user study addressed the following research questions:

R1. Detail: how does projection distance relate to perceivable detail in text and graphics?

R2. Context: how well can audio and, to a certain extent, real objects be used to convey context while zooming into details?

The user study deployed the calibrated environment introduced in the previous section, in which the illumination conditions were kept constant. 16 subjects participated in the experiment (age 22-47, 4 female, 12 male). In the first part of experiment, which addressed R1, the projector was fixed between two bars to allow the movement of the projector forward and backward on a fixed trajectory. Here, the user study focused on validating visibility and legibility of the patterns (checkerboard and line grid) and fonts: the projector was moved towards the location where the content would be best visible first (from far to close to the projection surface). We logged the exact location in mm accuracy and calculated the associated PPI. Estimating the size of the detected pattern indicates which level of detail is visually recognizable at a given distance. All patterns (6 patterns, 6 checkerboards) and text (6 sizes) were mixed randomly. Each user had 2 runs, thus resulting in a total of 36 trials. The size of the patterns and letters 2 to 5 were estimated in an previous small experimental visibility study, patterns and letters 1 and 6 are control data (1 actually not recognizable and 6 clearly detectable from all locations).

In the second part of the study, we performed an exploratory test on the effects of the physical and auditory context using the two implemented scenarios (plant and cityscape). Users could freely explore each environment for about 2-3 minutes - there was no specific task besides exploring the various locations and details found in the environment. After the experiment, 27 questions (using a 5-point Likert scale) had to be filled out after the experiment that targeted visibility, legibility and informational quality issues, the us-

age of auditory and real world context, and general usability issues. The questionnaire included the system usability scale (SUS) 10-point questionnaire to target the general attitude towards the introduced methods.
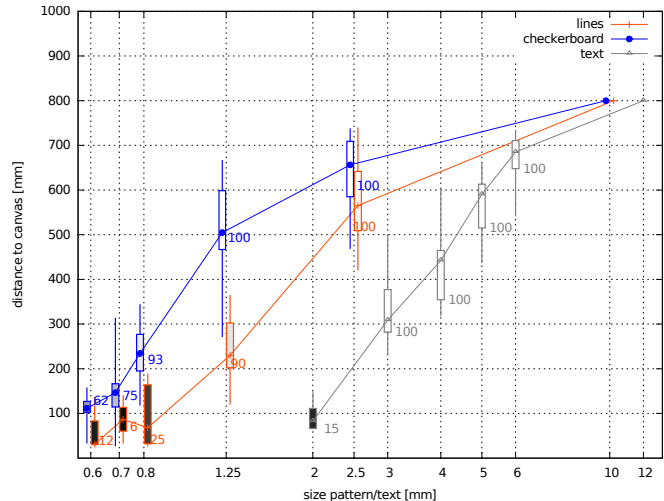
**Fig. 3**. Results of the visibility test. Numbers indicate detection rate percentage.

5. RESULTS

R1. In contrast to well-defined real world environments, in which the user can distinguish 1 mm from about 3 up to 6 m [14], projection based environments have more constraints. Figure 3 summarizes the results of our visual experiment, while Figure 4 shows clearly the offset between the maximal pattern detail afforded by the projector (reflected by a lower PPI), and the PPI at which the patterns actually where recognized (higher PPI, being close to the projection surface).

As expected through the increasing PPI, users were able to perceive increased detail closer to the projection canvas. Looking at Figure 3, the results illustrate that closest to the canvas, line grid patterns smaller than 0.8 mm were only detected by a few users (low percentage). In contrast, the detection rate for the checkerboard pattern is still good enough at around 0.7 mm. This difference is caused by the moiré effect the line grid generates at higher PPI. With our particular setup, relating detail and distance reveals that showing details larger than about 1 mm is not well feasible within a projector-canvas distance larger than 280 mm (comparable to 92 PPI) as detection rates drop below 50%.

With regards to legibility of text, the results show that text smaller than 3 mm is not readable: it is recommended that the minimum distance between projector and user should not be selected smaller than 330 mm for 3mm size (comparable to 78 PPI). Looking further, Figure 4 shows notable differences between the computed PPI of a pattern and the PPI at which it is visually afforded through projection, considering

only those users who detected the pattern (see Figure 3). Clear gaps are noticeable at the higher level details that are only visible at very high PPI, whereas with lower level details the offset between computed and legible PPI is much lower to almost non-different. In conclusion, it is crucial to consider that visual details cannot be shown at their theoretical PPI but need to be adjusted to higher PPI to be fully visible and legible. As we noted before, while the results are only valid for our specific setup, other projectors likely exhibit similar effects - the current setup can be well generalized as testbed for other projectors too. Hence, it is highly recommended to perform a similar calibration step before designing the actual visualization techniques.
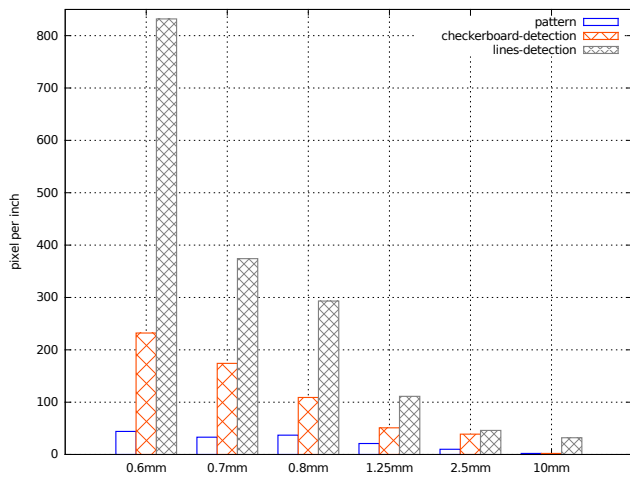
Fig. 4. Comparison between computed PPI for the pattern (black) and mean PPI when detected by users (checkerboard blue, lines orange - detection rates vary, see Figure 3).

R2. With respect to the qualities of the two contextual information methods, we assessed the subjective results with following results. Overall, the SUS analysis revealed that the zooming interface was well usable by all users, scoring 84.

Visual quality. Most subjects reported positively on the visual quality of the projector. Users were satisfied with the overall sharpness of the projection (avg 3.75/sd 1.90), and noted that the projection quality was reasonable overall (avg 3.00/sd 1.41) and particularly well when the projector was brought closer to the projection surface where the projection notably increased the PPI (avg 3.85/sd 1.39). The higher PPI was found to be very important when interpreting detailed information (avg 4.50/sd 1.41). In contrast to the positive rating of brightness and sharpness, many users noted disturbing effects (avg 1.65/sd 1.56). The laser technology we used generates a minimal speckle effect and in particular regular line patterns smaller than 1.25 mm generate a moiré pattern. Nonetheless, the aberrations did not seem to have affected the subjective view on the system too much. In general, the size of projection (dependent on distance to the canvas) was noted

positively (avg 3.56/sd 1.72), which is an important aspect for the flashlight metaphor.

Representation. Most users noted that information available at different distances was well chosen (avg 4.06/sd 1.49). While this mostly refers to the switching content in the plant scenario, it also addresses the transition to details by smoothly increasing the resolved details in the cityscape image. The general approach of increasing detail in representation (visual and informational level) was rated very well (avg 4.50/sd 1.41), similar to the usefulness of this information for search tasks (avg 4.56/sd 0.99). This was in line with the suitability of the noted level of detail (PPI) based on the distance (avg 3.81/sd 1.62). As such, results support the usefulness of the flashlight approach in combination with proximity, as introduced in our implementation.

Real world context. With respect to the real world context, users noted that the blending of different levels of information was very useful (avg 4.13/sd 1.85) while the real world object was, not surprisingly, suitable as direct context for the shown content (avg 4.80/sd 0.78). This was supported by user behavior observation, with most users exploring the plant and its digital contents from different sides. However,with current statistic and observational data, it is not possible to make any further (and important) conclusions on issues such as consistency and the building up of mental images, sourcing the role of the real and virtual.

Auditory context. The usefulness of the auditory feedback was rated very positively (avg 4.50/sd 1.41), which in part is also likely due to the fact sound enriches the overall application. Subjects noted they could reasonably well differentiate between different sounds (avg 3.68/sd 1.41), while also the neighboring sounds were still well observable and understandable by most but not all users (avg 3.43/sd 2.23). Similarly, to most but not all users the sound bubbles were well localizable (avg 3.68/sd 2.42). These results support our approach and show that auditory contextual cues can enhance spatial interpretation of the scene.

6. CONCLUSION AND FUTURE WORK

Within this paper, we quantified legibility and visibility boundaries of level of detail in pico projector based environments and validated the quality and usefulness of our novel context-sensitive zooming interface. Results showed that indeed there is a considerable offset between the theoretical projected PPI and the details that can be conceived by the user. There are limitations in text and graphical visualization affected by projection aberrations that need to be taken reflected when defining the best level of detail at different proxemic levels. The quantification can aid in designing interfaces that make use of a similar projector, while the methodology also gives guidance when a different projector is used. Different types of projectors will produce different levels of aberrations, hence, the testbed should be seen as a guideline for

testing with other projectors.

With respect to contextual cues, the user study showed that subjects rated the audioscapes very positively, while also noting that the plant give relevant cues for interpreting the projected information. Interestingly, users did not seem to have noticed that the audio in the interface only communicated proximity and not directional cues, as per current implementation. However, since the study was explorative, we only have obtained a general notion on the usability of our approach. While the reception of the techniques was positive, further testing is needed to pin down the actual performance of the techniques.

In conclusion, the visual experiment showed the importance of calibrating pico projector setups to adjust for optimal visibility and legibility. Furthermore, the system validation illustrated the potential of the audioscape as contextual cue, while the real world object was found, as expected, to be an important asset to align digital and real content. As such, we showed that an audiovisual zooming interface is a valid approach that hopefully will be taken up further in SAR usage scenarios. Informed by the experiments, our next steps will be to optimize the levels of detail per proxemic step, by taking in the exact on aberrations guidelines obtained from the first part of the study. In addition, we will extend our work on spatializing audioscapes and investigating further the exact role of real world objects through a series of implementations and experiments. Thereby, we will specifically target search task performance and spatial recall while addressing different contexts, as well as look deeper at user behavior at different proximities.

7. REFERENCES

[1] M. Weigel, S. Boring, N. Marquardt, J. Steimle, S. Greenberg, and A. Tang, "From Focus to Context and Back: Combining Mobile Projectors and Stationary Displays," in *In Proceedings of GRAND Network Centres of Excellence Meeting*, 2013.

[2] Xiang Cao and Ravin Balakrishnan, "Interacting with dynamically defined information spaces using a handheld projector and a pen," in *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*, New York, New York, USA, Oct. 2006, p. 225, ACM Press.

[3] Karl D. D. Willis, Takaaki Shiratori, and Moshe Mahler, "Hideout: Mobile projector interaction with tangible objects and surfaces," in *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, New York, NY, USA, 2013, TEI '13, pp. 331–338, ACM.

[4] Nicolai Marquardt, Robert Diaz-Marino, Sebastian Boring, and Saul Greenberg, "The proximity toolkit," in *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, New York, New York, USA, Oct. 2011, p. 315, ACM Press.

[5] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson, "A review of overview+detail, zooming, and focus+context interfaces," *ACM Computing Surveys*, vol. 41, no. 1, pp. 1–31, Dec. 2008.

[6] Patrick Baudisch, Nathaniel Good, and Paul Stewart, "Focus plus context screens," in *Proceedings of the 14th annual ACM symposium on User interface software and technology - UIST '01*, New York, New York, USA, Nov. 2001, p. 31, ACM Press.

[7] Sunjun Kim, Jaewoo Chung, Alice Oh, Chris Schmandt, and Ig-Jae Kim, "iLight," in *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, New York, New York, USA, Apr. 2010, p. 3631, ACM Press.

[8] Brett Ridel, Patrick Reuter, Jeremy Laviole, Nicolas Mellado, Nadine Couture, and Xavier Granier, "The Revealing Flashlight," *Journal on Computing and Cultural Heritage*, vol. 7, no. 2, pp. 1–18, July 2014.

[9] S. Rapp, G. Michelitsch, M. Osen, J. Williams, M. Barbisch, R. Bohan, Z. Valsan, and M. Emele, "Spotlight navigation: Interaction with a handheld projection device," in *International Conference on Pervasive Computing*, 2004.

[10] Tao Ni, Amy K. Karlson, and Daniel Wigdor, "Anatonme: facilitating doctor-patient communication using a projection-based handheld device.," in *CHI*, Desney S. Tan, Saleema Amershi, Bo Begole, Wendy A. Kellogg, and Manas Tungare, Eds. 2011, pp. 3333–3342, ACM.

[11] Johannes Schning, M. Rohs, S. Kratz, M. Lchtefeld, and Antonio Krger, "Map torchlight: a mobile augmented reality camera projector unit," in *Proceedings of The 27th International Conference Extended Abstracts on Human Factors in Computing Systems*. ACM New York, NY, USA, 2009, pp. 3841–3846, ACM.

[12] G. Eckel, "Immersive audio-augmented environments: the LISTEN project," in *Proceedings Fifth International Conference on Information Visualisation*. 2001, pp. 571–573, IEEE Comput. Soc.

[13] D Kalkofen, E Mendez, and D Schmalstieg, "Interactive Focus and Context Visualization for Augmented Reality," in *Proceedings of 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2007.

[14] E Goldstein, *Sensation and Perception*, Brooks Cole, 5th edition, 2002.