Screenfinity: Extending the Perception Area of Content on Very Large Public Displays

Constantin Schmidt[1]   **Jörg Müller**[1,2]   **Gilles Bailly**[1]

1 Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin
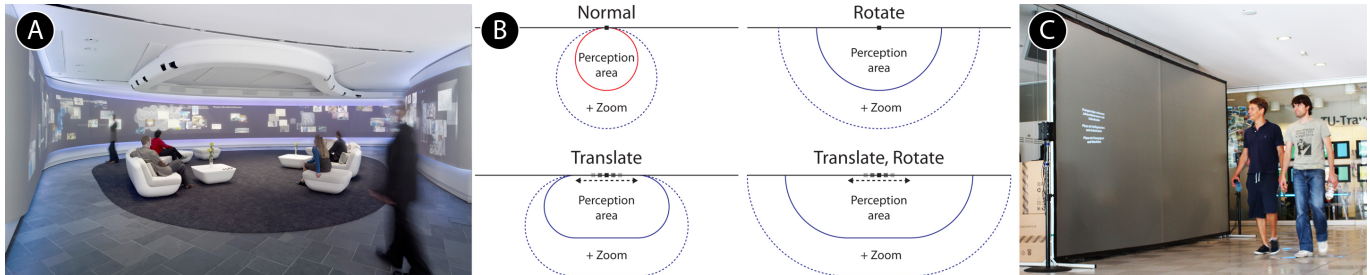
2 University of the Arts, Berlin

Figure 1. (A) Recently, *very* large public displays are being installed, like this digital display of 80m length ((c) Deutsche Telekom). This could enable passers-by to read considerable amounts of content *while walking*. (B) We propose a model for the perception area of content to analyze content legibility on public displays and three techniques (Zoom, Rotate, Translate) to increase the perception area. (C) Screenfinity is a display that tracks users and uses these techniques to enable reading while walking. A lab and a field study show that experts and novices can read while walking, while many novices stop to engage in extensive technology exploration behavior.

ABSTRACT

We propose and validate a model of the perception area of content on public displays in order to predict from where users can read. From this model, we derive *Screenfinity*, a technique to rotate, translate, and zoom content in order to enable reading while passing by *very large* displays. Screenfinity is comfortable to read when close, supports different content for different users, does not waste screen real estate and allows expert passers-by to read content while walking. A laboratory study shows that expert users are able to perceive content when it moves. A field study evaluates the effect of Screenfinity on novice users in an ecologically valid setting. We find (1) first time users can read content without slowing down or stopping; (2) Passers-by stopping did so to explore the technology. Users explore the interaction, the limits of the system, manipulate the technology, and look behind the screen.

Author Keywords
Large public displays; perception area; visual acuity

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Traditional posters or digital displays are often designed to enable content to be read while passing-by, and thus display

relatively few words [2]. However, recently very large public displays are being installed. For example, Figure 1 (A) shows a display of 80m length. This could enable passers-by to perceive significant amounts of content *while walking*. Expert users could for example pass by the same display in a subway station every day. At the beginning of the display, they could select among the top five newspaper articles (e.g., by touching the screen, performing mid-air gestures or using a profile on their mobile phone). On their walk to the subway, they could then read the article without slowing down.

However, just making the screens large alone is not sufficient. For instance, just rendering the content very large makes reading less comfortable when close to the display as it forces users to make head movements. Conversely repeating the content does not solve readability of content from far away, and it may be difficult for passers-by to read continuously while switching from one repetition of content to the next. Additionally, both solutions (1) do not allow different users to see different contents and (2) waste the entire screen space, even if users are close.

We propose a model of the *perception area* - the area where passers-by can read a piece of content - in order to analyze the readability of content for passing-by scenarios. Our model is based on the visual acuity of humans and predicts from where users can read content, and how much content they can read depending on their trajectory. From this model, we derive *Screenfinity*, a system to present content to passing-by users of very large displays.

We report on two user studies to evaluate Screenfinity for experts and novices. A laboratory study shows that expert users passing by a display are able to perceive content when it moves (e.g., is virtually rotated or translated). A field study evaluates the effect of Screenfinity on novice users in an ecologically valid setting.

We find 1) also novice users read some content without stopping; 2) if novices slow down or stop, it is mostly not for the content, but for extensive *technology exploration* behavior. Users explore the interaction, the limits of the system, inspect and even manipulate the technology, and often look behind the screen.

Our findings can help designers and practitioners to design and deploy more effective public displays that can be used while walking. In particular, Screenfinity can be built using current technology like normal or depth cameras, because only the users' positions need to be tracked.

Our two contributions are: (1) A model for the perception area of content on visual interfaces, validated through a lab study. (2) A novel public display increasing the perception area and allowing interaction while walking, validated through a lab and a field study.

RELATED WORK

Proxemic Interaction
Proxemic interaction [12] proposes that computing devices should be aware of their *spatial relationships* to other devices and users and adapt accordingly. The main concepts of proxemics are distance, orientation, movement, identity, and location. For instance, the proxemic media player [12] uses discrete zooming to fit more content when the user is close to the display. The proxemic peddler [26] moves content to attract and regain attention of passers-by depending on their orientation. The interactive public ambient display [25] follows similar ideas and e.g., automatically switches from ambient display to implicit and then subtle interaction mode when a user approaches. Range [14] is a whiteboard that proactively transitions between display mode and authoring mode depending on user distance. In this paper, we employ proxemics to adapt content to user location and distance, specifically. While many examples of proxemic interaction aim to support the user on a cognitive (and discrete) level (e.g., switching visualization mode), we focus on the perception level to optimize the users' perception of content.

Large Displays
Considerable research has been conducted regarding large displays as they provide several benefits such as productivity, peripheral awareness, and a more immersive experience (summarized in [6]). A number of interaction techniques have also been proposed to overcome some limitations of large displays such as target acquisition [19], task management [4], visual attention [15], or navigation [11]. While this paper is about large displays, our scenarios focus on passing-by interaction in front of public displays rather than collaborative work in meeting rooms. This means that our focus is on *immediate usability* for simple content presentation rather than performance for achieving complex tasks. Our adaptation techniques should also work while walking.

Perspective-corrected displays
Perspective-corrected displays are one specific case of large displays proposed in the context of meeting rooms. They adapt content to the perspective of the user in the context

of Multiple-Displays Environments (MDE). For instance, e-conic [21] is a MDE that zooms and rotates windows so they always appear orthogonally to the user. The size of the windows in display-space is kept constant, and the authors show that reading is 8% faster when the perspective is corrected. Larson [16] presented virtually rotated text on a flat display and found that 1) there is almost no impact on legibility for rotations up to $55°$ and 2) for small fonts reading speed decreases for larger rotations (almost half speed at $70°$). In the context of volumetric displays [13], text was rotated in yaws around the full axis ($360°$) and users were free to move their head around the display as long as they did not walk. There seemed to be no difference in reading speed for rotations up to $60°$ yaw, while reading speed at $90°$ was reduced to approximately 20%. A special case of perspective correction is Whale Tank VR [17], which divides large screens into one region for each user to provide a multiuser Fish tank VR.

Our approach is similar to e-conic [21]. However, 1) we provide a model of the perception area, which is not provided by e-conic; 2) we aim at increasing the perception area of very large public displays instead of improving work performance in office settings; 3) e-conic manipulates only two dimensions, rotation and zoom for standing users, while we also investigate translation and different combinations of these 3 dimensions; 4) we focus on passing-by scenarios, while e-conic focusses on meeting rooms with static users.

Public Displays & Media Facades
The usage scenario of public displays is usually very different from that in offices. For this reason, research on public displays has mostly resorted to field studies of actual audience behavior [22, 18, 20, 23]. These field studies often show very different user behavior from what is simulated in the laboratory. For instance, [18] found that real-life multiuser behavior around a multi-touch table was very different from scenarios of multi-touch lab studies. Looking Glass [20] shows that passers-by who notice interactivity of displays need considerable time to slow down and stop mostly after already having passed by the display (Landing Effect [20]). The authors proposed to install multiple displays in a row so users can notice interactivity on one display and stop to interact in front of another display. Our approach involves making displays so long users do not even have to stop for interacting and can do so while passing by. We call this kind of interaction *passing-by interaction*. One specific kind of very large public displays are interactive facades [7, 9]. However, facades 1) often focus more on far-away users rather than close passers-by and 2) are often low-resolution. One exception is Climate on the Wall [7]. Passers-by could stop to select some words on the facade in order to be read by far-away spectators.

Perception area
The concept of perception area was inspired by Isovist [3], viewsheds [10] and the maximal distance from where posters attract attention [2].

In architecture, Isovist [3] describes the set of all points visible from a certain vantage point. It can also be reversed to determine the number of vantage points from where a certain

point can be seen. Most Isovist research assumes the maximum visibility distance to be infinite (which may make sense for architectural structures). Nimbus (from where someone can be seen) and Focus (what someone can see) extends the concept of Isovist in the context of public displays for both displays and users [23]. Dalton et al. also investigate how the Isovist shape can impact the recall of content on public displays [8].

In geography, the concept of viewshed describes the visibility of elements such as radio antennas [10]. In contrast to most research on Isovist, viewsheds are often non-continuous, e.g., a mountaintop may be visible from a vantage point but not the valley in front of it.

Finally, advertising research mainly focused on the maximum distance from where content attracts attention rather than the shape of the perception area and the legibility of content [2]. The value of an advertising poster is usually the number of passers-by statistically passing the perception area, discounted by a factor influenced by the angle of the display, distance to walking path, environmental complexity etc. E.g., in Germany the G-Value [9] is used. In these models, the perception area is generally assumed to be pie-shaped (same distance independent of angle, up to some maximum angle).

Isovist and viewsheds provide excellent tools to determine the shape of perception area given a number of occluding obstacles, but not for the maximum viewing distance. Advertising research has focused on the maximum distance to attract attention, but not on legibility. In this paper, we focus both on the viewing distance and the legibility of content.

In summary, Screenfinity builds on proxemics by using the location, distance and walking direction of the users to adapt content. While several interaction techniques have been proposed for large displays, they generally do not focus on (1) public displays, (2) passing-by interaction and (3) legibility of content. Additionally, no model for the perception area for large public displays has been addressed in the literature.

CONTENT LEGIBILITY ON DISPLAYS

Content legibility depends on the properties of the users and properties of the display. We describe these two aspects and assume a sufficient display resolution, and an easily readable font with just black and white content.

Visual acuity

Visual acuity is a measure of the spatial resolution of the visual processing system. Because the human eye is an optical system, it is best to represent the apparent size of objects on the retina as the visual angle $\alpha = arctan(\frac{s}{2D})$, where s is the size of the object and D is the distance between the object and the eye [27]. Visual acuity is then defined as the minimum size of a white space between black spaces such that the two black spaces can be distinguished (assuming sufficient contrast). "Normal" visual acuity is about 1 minute of arc [27].

Maximal visual acuity is however only achieved in a relatively small $(1.7°)$ central area of the eye, the fovea. The overall horizontal size of the visual field is almost $200°$, but

resolution in the outermost regions is very low. For this reason, humans move their eyes, head, and body considerably to bring different regions into their field of view (active vision). Visual acuity is also strongly dependent on luminance. This is because the fovea contains almost only cone cells, which are not very sensitive to light. In low light conditions, only rods perceive light, which have a much lower spatial resolution and are not present in the fovea.

If luminance or contrast are insufficient, visual acuity decreases considerably. Therefore, all visual displays need to take these factors into account.

Displays

The most important properties of displays for our purposes are *contrast*, *luminance*, and *angle dependence* of these variables. Luminance describes how bright a display appears. Contrast describes the ratio of luminance of bright and dark regions, that make objects distinguishable. For most display technologies, like LCD, contrast is highly dependent on the angle from which the display is seen.

In this paper, we abstract from all factors that may reduce legibility by reducing contrast. We simply assume that contrast is high independent from distance and the visual angle. If this is not the case for a certain display (e.g. LCD) or environment (e.g., bright surroundings, fog), then the model has to be adapted to the specifications of the particular situation.

STATIC MODEL OF PERCEPTION AREA

The objective of the static model is to describe the perception area of a pixel of content on a (vertical) display. The *perception area* (see Figure 1 (B)) is the set of locations from where the content is legible. Because visual acuity is based on recognition of pixels, we base the model on content pixels, where a pixel of the content can encompass many physical pixels of the display. In order to determine the perception area for a certain piece of content, one can approximate the content with one of its pixels (if content has few pixels) or consider the intersection of the perception areas of all content pixels (if content has a lot of pixels). Similarly, the perception area for an entire display with a certain content pixel size can be approximated by the perception area of a single pixel, or the intersection of the areas of all (content) pixels. A user with normal vision (20/20) can distinguish points with a difference of 1 minute of arc [27]. We can also say metaphorically that the pixel resolution of the eye is 1'.
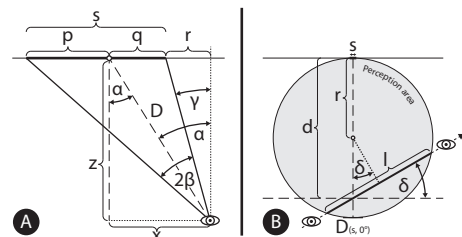

Figure 2. Calculating distance D at visual angle α from where a pixel of width s can be distinguished (A). Calculating the length l of a trajectory from where content of pixel size s is legible (B).

Based on the content pixel size s and the angle deviating from orthogonal view ($0°$) α we can then derive the distance $D(s, \alpha)$ (Figure 2(A)) from where the pixel can be recognized as (assuming contrast does not change with angle):

$$s = p + q = p + x - r \qquad\qquad p = tan(\alpha + \beta) \cdot z - x$$
$$s = tan(\alpha + \beta) \cdot z - r \qquad\qquad r = tan(\alpha - \beta) \cdot z$$
$$s = z \underbrace{[tan(\alpha + \beta) - tan(\alpha - \beta)]}_{V} \qquad z = cos(\alpha) \cdot D$$

$$D(s, \alpha) = s \cdot (V \cdot cos(\alpha))^{-1} \qquad 2\beta = 1' = 1° \cdot 60^{-1}$$
$$\approx s \cdot cos(\alpha) \cdot 3438 \qquad tan(2 \cdot \beta)^{-1} \approx 3438$$

Thus, when standing orthogonally to the display, a pixel of size 3cm can be recognized from $D(0.03m, 0°) \approx 0.03m \cdot 3438 = 103m$.

At $45°$, the distance is $D(0.03m, 45°) \approx 73m$. Thus, the perception area is a *circle* in front of the content for not perspective-corrected content (see Figure 1 (B)).
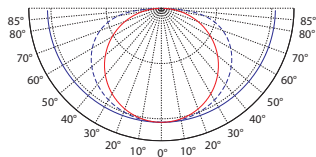
Perspective Correction
When perspective is corrected, pixels are shown so that they always appear orthogonally to the user. As long as enough display space is available, the maximum distance will always be $D(s, 0°) \approx s \cdot 3438$, independent of the angle. Therefore, the perception area for perspective corrected content is a half circle, assuming infinite display size. Of course, this technique uses more physical pixels at steep angles. However, even if the number of physical pixels is kept constant, perspective correction yields a larger perception area compared to not perspective corrected content (see Figure 3). Only at extremely steep angles ($> 86°$), perspective correction grows horizontally so that it shrinks vertically so much that it is not recognizable anymore. However, in this paper we consider content presentation on extremely long displays. Therefore, in the rest of the paper we will consider perspective correction where the *perceived size* of the content is kept constant rather than the number of physical pixels.

Dynamic Model
The dynamic model aims to predict how much content users can read if (a) they do not change their walking path, and walk straight with constant speed, and (b) they always look at the display while they can read content.The distance l (Figure 2 (B)) a user walks inside the perception area can be calculated by giving the normal distance to the screen (d) and the angle with which he passes (δ).

$$l(s, d, \delta) = 2 \cdot \sqrt{r^2 - [cos(\delta) \cdot (d - r)]^2} \quad r = \frac{D(s, 0°)}{2}$$

Figure 3. Perception area with uncorrected content (inner circle, red), perspective corrected (outer half-circle, blue) and perspective corrected with same number of physical pixels as uncorrected (dashed line between, blue).

For example, if users are passing parallel to the screen ($\delta = 0°$) at distance $d = 3m$ with a content pixel size of $s = 0.001m$, they are in the perception area for 2.3 m. Assuming a walking speed of 1.4 m/s and a reading speed of 3 words per second, passers-by can read about 5 words without slowing down. While this model is suited for the standard or zoomed perception area, it can easily be adjusted for perspective correction or translation.

Design Space of Adaptive Techniques
The core idea is now that the display is *much* larger than the content we want to present, and we can thus *translate, rotate,* and *zoom* the content according to the user position to optimize the perception area. We introduce three reference systems relative to which content can be held static or be distorted. The reference system of the *display* is the pixel coordinates of the display device. The normal case is that content is static in the display coordinate system. The reference system of the *user's eyes* is defined by individual sensor cells on the users retina. The content would be static in this reference system, for instance, if the display is attached to a contact lens. The reference system of the *user's head* is in between these two. The content is static in this reference system, e.g., if the display is attached to glasses the user wears.

Position and direction of the head can be tracked by external devices such as depth cameras. While pilot studies (using Optitrack) reveal promising results by using the head position to update the content geometry, using the direction of the head to translate content appears to be confusing. These results can be explained by the vestibulo-ocular reflex [5]. The human vestibular system is hard-wired to the eye muscles. Thus, any head rotation is automatically counterbalanced by eye movement, keeping the eyes focused on the same spot. This mechanism works well when translation of content is fixed in the reference system of the display, but poses problems when keeping content translation fixed in the reference system of the head. Because head direction is also difficult to track accurately, we translate content according to the user's head position only. While many possible mappings of head position to content position are possible, we initially chose orthogonal mapping as being natural and simple.

Adaptation to user reference system

- Rotation. The orientation of content is orthogonal to the axis defined by the center of the content and the head position (Figure 2 (A)). The edges of the perception area are then increased (Figure 1 (B)).

- Translation. The position of the content on the display is defined by the closest location from the user head (orthogonal projection, in Figure 2 (B) the distance d). The perception area is then elongated along the entire width of the display (Figure 1 (B)).

- Zoom. The size of the content is proportional it's distance from the user's head. The pixel size of the display is then increased, such that the perception area grows while the shape stays the same (Figure 1 (B)).
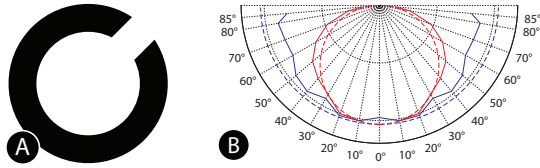
Figure 4. (A) Landolt rings are a standard stimulus for visual acuity from Ophthalmology (Study 1). (B) Perception area as determined in experiment 1 in comparison to the model (dashed). The inner circle is the perception area of the static content (red). The outer half-circle is the perception area of the perspective corrected content (blue).

STUDY 1: VALIDATING THE STATIC MODEL

The objective of the first study is (1) to validate the static model, that is, to find the maximum distance from which content can be recognized given a certain angle to the display, and (2) to validate the perception area for the rotating technique.

Stimulus and Apparatus

Because we were interested in simple legibility of content, we used Landolt rings as stimulus (Figure 4 (A)). Landolt rings are commonly used by ophthalmologists to evaluate visual acuity. In this experiment, the gap can be located on the left, right, bottom, top and the $45°$ positions in between. The stimulus was shown as a DLP back projection on a sheet of paper. A light meter was used to verify that the contrast was independent from the angle. Participants were positioned at 5m distance from the display. A curtain prevented the participant from leaning over to improve the viewing angle.

The display (projector and projection surface) was rotated around the center of the projection surface to investigate legibility at different angles without requiring participants to move between different conditions.

Task and Procedure

Participants were asked to name the direction of the opening of the Landolt ring (one ring was shown at a time). We used a classical staircase method to determine the minimum size the participants could barely recognize. The minimum size was determined as the smallest size where participants could recognize 8/10 rings. This size was found using binary search where the step width of opening size was 0.1mm. Participants started with large rings, and ring size was subsequently decreased when they could recognize at least 8/10 rings and increased when they could not.

This procedure was repeated for angles between $0°$ and $80°$ in $10°$ intervals, and $85°$ (Figure 4 (B)). The effect was assumed to be symmetric. Values were normalized for visual acuity of the participant (value at $0°$).

Perspective Conditions

We investigated perspective-corrected and non-corrected presentation. In the first condition, Landolt rings were perspective corrected for the participants perspective, such that the ring seemed to be hovering in front of the display in the same size as it would be if the display was orthogonal to the participant. In contrast to [21], the rotated stimuli's size was not decreased, such that the size remained constant in the reference system of the user rather than the one of the display.

Participants

14 participants, 3 females (aged from 23 to 48) from our institution with normal or corrected-to-normal vision participated in the experiment.

Design

Independent variables are angle to display and perspective correction (corrected & non-corrected). The dependent variable is the minimum opening size where participants can still accurately (80% success) recognize the Landolt rings. The order of angle to display was increasing for the first half of participants and decreasing for the second half of the participants to avoid possible bias due to fatigue. For each angle to display, participants started with the non-corrected condition. The design included: 12 users $\times$ 9 angles $\times$ 2 conditions (corrected & non-corrected) = 216 measurements.

Results

Distance

The measured perception area for both corrected and non-corrected content is shown in Figure 4, together with the model of the perception area. An ANOVA reveals an interaction effect for angle*condition on the distance of perception ($F_{9,117} = 19.5, p < 0.0001$) that shows that when angle increases, the change of the distance of perception occurs differently for the two techniques (as shown in Figure 4). An ANOVA executed on the non-corrected technique reveals an effect for angle on the distance of perception ($F_{9,117} = 43.2, p < 0.0001$). A post-hoc Tukey test confirmed that the distance of perception decreases with the angle for non-corrected stimuli. In contrast, an ANOVA executed on the corrected technique reveals no effect for angle on the distance of perception. An ANOVA also reveals a main effect for technique on the distance of perception ($F_{1,13} = 215.9, p < 0.0001$). A post-hoc Tukey test confirmed that the corrected technique provides a higher average perception distance than non-corrected.

Prediction Accuracy

Empirical data matched the predictions very well for uncorrected perspective (Figure 4 (B)). Regression analysis of the predicted ($Pred$) distance for the uncorrected technique against the empirical (Emp) mean shows a strong relationship: $Pred = 0.87xEmp + 13.04, R^2 = 0.98, p < 0.0001$. For the corrected technique, there seems to be a deviation between our model and the data. A linear regression of the data results in $R^2 = 0.76, p < .05$, with $D(\alpha) = -0.207 \cdot \alpha + 102$, where the model would predict $D(\alpha) = 103.12$. This means that with a rising angle, the experimental data deviate from the constant prediction. This leads to the conclusion that virtual rotation is not as effective as physical rotation in our case. One possible explanation could be the limited depth-of-field, due to the slanted display it may be slightly more difficult to focus for users than focusing on a flat display. Nonetheless, our model for perspective corrected distances seems to be sufficient for our practical purposes.

Discussion

From this experiment we can conclude two things. First, that our static model is a good approximation of viewing distances for displays where contrast is not dependent on viewing angle (no reflections etc.). Second, we can increase the perception area as predicted through rotation, zoom, and translation. The increase of perception area through translation follows directly from the model. The increase through zoom follows from the evaluation of the noncorrected condition. When content pixel size grows, the maximum distance grows linearly. The increase through rotation follows from the evaluation of the corrected condition.

One major issue however is that we do not know whether this increase in perception area is also valid for walking users. Because all three techniques continuously move the content while the user is walking, it is possible that content is more difficult to perceive. To verify this, we conducted an experiment for perception of content while passing by the screen by modifying the location (translation) and the perspective (rotation) of the content.

STUDY 2: VISUAL SEARCH WHILE WALKING

The objective of this study was to determine whether the content moving within the display's frame of reference while passing-by diminishes the users' ability to conduct visual search while walking. We already know from the static model that there will be areas where content is legible for rotating and translating, but not for static content (see Figure 5). Thus, for any comparison of translate, rotate, and static, one could arbitrarily make the display so long that translate and rotate would perform better than static. For this reason, we chose a passing-by task optimal for the static baseline condition (from A to B in Figure 5), where the user could already recognize the static content from the starting position. The question was then whether visual search was slower in a translation or rotation condition.

Participants and Apparatus

10 participants (1 female) aged from 23 to 29 ($\mu = 25.1, \sigma = 2.3$) participated in the experiment. A large (5x2.5m) back projected display was used (Figure 6). The head of the user was tracked with an Optitrack system. Users were provided a device with a button (Logitech Presenter). Start/goal lines were marked on the ground at both ends of the display.

Task and Stimulus

The task consisted of beginning from the starting line and to walk in normal speed towards the finishing line without stopping. As soon as the participant crossed the starting line, the stimulus was shown on the screen. The stimulus was a 5×5 grid of colored dots. The participant had to count the red dots as fast and accurately as possible and press the button.
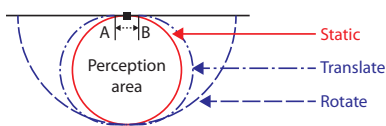

Figure 5. Perception areas for translate, rotate and static condition. In Study 2, content perception along the trajectory from A to B was tested.
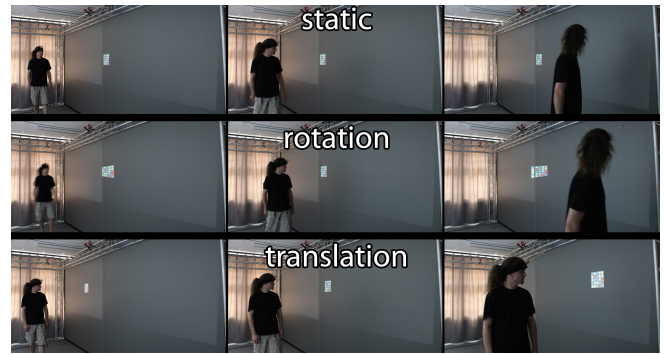

Figure 6. The three techniques tested in study 2.

As soon as the button was pressed, the selection time was measured and the stimulus disappeared. Participants then indicated to the experimenter the number of red dots.

Conditions

The three tested *techniques* were static (as a baseline), rotation and translation. The 5x5 grid of colored dots was displayed in the center of the screen for static and rotation. We introduced 6 levels of difficulty depending on the number of red dots to count in the grid (3, 5, 7, 9, 11 or 13 red dots).

Design

We used a within-subjects design. All participants tested all three techniques. The order of techniques was counterbalanced between participants. There were 8 blocks. For each block, each difficulty was tested once in random order for a total of $10 \cdot 3 \cdot 8 \cdot 6 = 1440$ selections.

Results

Selection time

An ANOVA reveals a significant effect for difficulty on selection time ($F_{5,45} = 258.9, p < .0001$). A post-hoc Tukey test shows that all pairwise comparisons are significantly confirming that selection time increases with difficulty. However, the ANOVA reveals no effect on techniques or interaction between technique and difficulty.

Angle of the head of the user

Figure 7 shows the angles of the user's head for highest level of difficulty (13 objects). The stimulus was located at 1.7m. For static and rotation, the angle starts at approx. $15°$ and rises almost continuously before dropping back near $0°$. Most participants continue to turn their head far beyond the point when they have passed the stimulus, and even beyond $90°$. We can also see that participants turn their head by only about $75°$ when they pass the stimulus, still focusing the content by
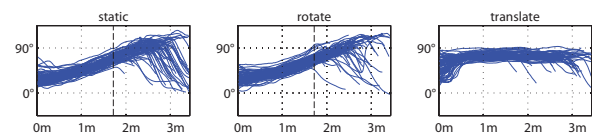

Figure 7. Head angles of participants while perceiving content for the highest level of difficulty (13 red dots), mirrored for right-to-left trials. Lines start when stimulus appears (start line) and end when stimulus disappears (button press).

rotating their eyes. Similarly, for translation, participants do not need to turn their head for the entire $90°$, but only for on average $75°$, because it is apparently more comfortable to cover the remaining $15°$ by rotating the eyes. It is also surprising for translate that participants only rotate the head towards the stimulus after having walked already for about 0.5 meters.

Discussion
From this study, we conclude that translation and rotation do not seem to affect visual search time when walking compared to static. From the model we know that the techniques increase the perception area for content. We can thus conclude that the techniques would improve the amount of information passers-by can perceive from a large public display.

SCREENFINITY FOR PASSING-BY INTERACTION
Screenfinity is a combination of translation, rotation and zoom to enable passing-by interaction (Figure 8). When users are in front of the display, content is translated to be orthogonal to the users. When users stand beside the display, so that content can not be translated further, content is rotated as to appear orthogonally to the users. Depending on the distance of users, content is continuously zoomed such as to maintain a constant apparent size to the user. Screenfinity also supports multi-user interaction by presenting different content to different users. If different contents overlap, the reaction of the system depends on the difference in the distance of users. If the difference is high, the content for closer users is shown on top, because it is smaller and covers only part of the content for far away users. If the difference is small, multiple users are interpreted as a group, and the same content item is shown for them together.

Implementation
For the field study user tracking using OptiTrack was replaced by depth cameras. Screenfinity needs to track users in a considerable area ($\sim 60m^2$), especially if they are not in front of the display. In order to cover this area, we developed custom depth camera clusters able to track $205°$ field of view. Each cluster consisted of 5 Asus Xtion Pro cameras. We used one cluster at each end of the display to minimize occlusions. We used two cards with four USB controllers each to process all cameras at 30fps. We installed a mobile back projection surface ($5×2.5$m) with three DELL S500 Ultra Short Throw projectors to minimize setup depth (1m).

We processed the depth images with OpenNI and OpenCV. Depth information from all cameras are merged into a single calibrated point cloud. A simple model of the environment is learned without users and can be manually refined to remove noise and background features (floor, columns, plants etc.).
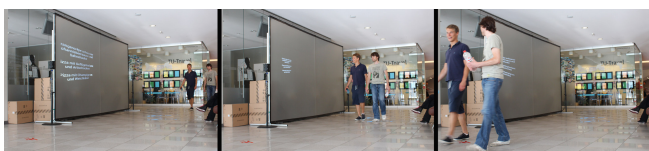

Figure 8. Two users reading text on Screenfinity while passing-by.

User tracking is performed as blob tracking on a 2D environment map (3D points are projected to the floor). Rotation, translation and zooming is realized by using a custom OpenGL renderer running at 60fps.

STUDY 3: FIELD STUDY
Screenfinity is designed to enable expert users who pass by the same display regularly to read content while walking. Screenfinity was shown to be effective for expert behavior by the model and lab study 2. However, any of these displays will also be passed by novices. The objective of this field study was to determine the general effects of Screenfinity on novice audience behavior in an ecologically valid setting. For this reason, the field study focusses on qualitative descriptions of novice audience behavior instead of quantitative measures.

Deployment
Screenfinity was installed in a well frequented lobby of a university cafeteria. It was oriented sideways along the main walking path. Passers-by were mainly students of two adjacent universities, but also included staff and guests.

Method
Our deployment was divided into two phases: *Optimization* phase and *Testing* phase. During day 1 and 2 (optimization phase), we performed several iterations of improving our system according to our on-site observations. During day 3 and 4 (Testing phase), the software was not modified and we conducted more detailed observations and conducted semi-structured interviews with users.

Data Collection
We collected anonymous videos from the 10 depth cameras, conducted continuous on-site observations (two researchers were on-site at all times), as well as semi-structured interviews with users.

Lessons learned from the optimization phase
During the first two days of deployment, we observed a general lack of attention towards the screen and its content.

Passing-by behavior
We were surprised how many passers-by continuously looked straight ahead without turning their head to observe their environment. A number of passers-by even reported not to have noticed the display at all, even though they passed it by a few meters distance. When the users walked towards the display, content was rendered almost in front of them, but when they were passing by the display, it was rendered at $90°$ angle, moving with them. Because passers-by rotated their heads so little, they rarely noticed the content.

Based on these observations, we decided to render the content at a $45°$ angle in front of the user, depending on the walking direction. Just when the users stopped, it was rendered in front of them. Transition between these states was smoothed. We also decided to pulse the content towards the user for 1s when the user walked towards the display to attract attention.

Content

Because the display was installed in front of a startup incubator, we initially showed logos of the local startups as content. While this attracted attention from a few passers-by who already knew the startups, the others said they ignored the content because it was advertising. In addition, the logos usually contained only one word of text, so that it was difficult to observe whether somebody only glimpsed at the display accidentally or read the content. From these observations, we decided to replace this "advertising" content by the menu of the cafeteria. Randomly, up to six menu choices were displayed. This content attracted more attention from passers-by. However, we do not know whether the menu attracted more attention because it looked less like advertising, because it was text, or because the information was more relevant. These three changes alleviated the problem of lack of attention, and in the two following days, much more attention was paid to the display.

Lessons learned from the testing phase

We mainly observed two user behaviors that we call *passing-by interaction* and *technology exploration*. Passing-by interaction describes the behavior the display was designed for, i.e., reading text without stopping. In contrast, technology exploration was something the system was not designed for, but rather an emergent novice behavior related to the ecological valid field study.

Passing-by interaction

Although the system was designed such that passers-by need to change their behavior as little as possible, we were surprised by just how little they changed their natural behavior when using the system in the intended way. In particular, to read content, they 1) did not stop, 2) did not slow down, and 3) even rotated their head relatively little (except when they arrived at one extremity while reading). In the following paragraph we describe passer-by behavior in detail.

Single & group of passers-by

For passers-by, there was different behavior for singles, communicating groups, and non-communicating groups (Figure 9). Singles had their heads mostly looking straight ahead, with very little scanning to the sides. Communicating groups often split up into pairs of two, who were looking at each other, with sometimes some remaining people scanning the environment. In this case, it was mostly the ones walking at larger distance from the screen who noticed it because they were looking in that direction, or the leftover ones. Finally, non-communicating groups behaved more like singles.
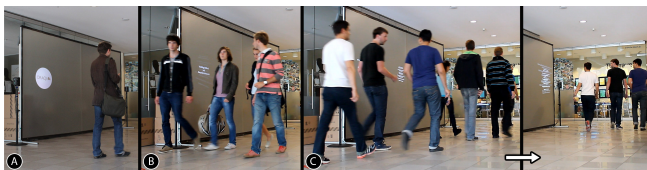
Walking speed

We were surprised by how few passers-by changed their walking speed to read more content. They often continued to look at the display until a few meters past it. When they reached the end of the display, most users aborted reading rather than slowing down. Surprisingly, we also observed one girl (Figure 10) discussing the display with the group. Rather than slowing down to show the screen, she preferred turn around and walk backwards.

Notice & Recall

Even first-time users of the display were able to read text while walking. However, they read only few words of content, and recalled only fragments when interviewed a few seconds later. Surprisingly, content was not read from beginning to end, but rather skimmed. Passers-by recalled mostly catch-words like "something with pizza" or "meatballs", although neither were these shown on top nor did the passers-by plan to eat these. Passers-by also scanned very quickly for certain features without reading the text. For example, one vegan passer-by had only scanned the menu for the "vegan" sign which was not shown, and wrongly concluded that no vegan meals were shown, looking away. However, many passers-by could accurately recall the number of menus shown. All passers-by who read content could tell that the content had moved, but almost none noticed that the content moved in relation to their own movement.

Design recommendations

Passers-by, even novices, can and will read text on displays while walking. However, similar to the landing effect [20], they often look at the display only after having already passed about 2.5m of it. Because they do not slow down, *very long displays* are necessary to enable passers-by to read larger amounts of text. Users do not remember the entire text, but only a few words from random positions inside the text, so one should *design for skimming*. Finally, when users reach the end of the display, they turn their heads, but do not slow down. Thus, such displays should be designed for *graceful abortion of interaction*. Because systems can predict that users will abort when they reach the end of the display, they can react accordingly.

Technology exploration

Besides the content-driven reading behavior, for which passers-by did not stop, many also engaged in extensive technology exploration behavior, for which they did slow down and stop (Figure 11). Technology exploration consisted of exploring interaction, testing the limits of the system, inspecting front items like cameras, manipulating the system, and investigating behind the screen.
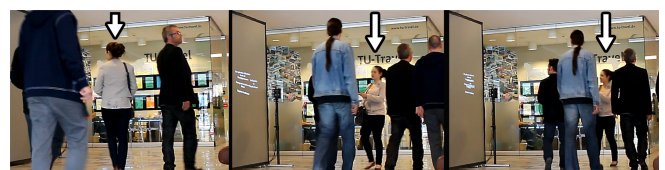

Figure 9. (A) Single user looking straight ahead. (B) The user furthest from the display notices it. (C) The last three of a group read content.
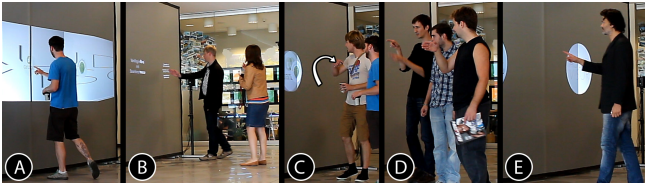

Figure 10. Girl discusses screen without stopping. To achieve this, she turns around and walks backwards.

Figure 11. Technology exploration. (A) Exploring interaction, (B) Inspecting the cameras, (C) Covering cameras with the hand, (D) Looking behind the screen.

Figure 12. Exploring interaction. (A) User touching screen. (B) Waving gesture. (C) Two users trying a punching gesture. (D) Simultaneously trying pointing and waving gestures. (E) Pointing gesture.

Exploring interaction
After noticing that the display is interactive, one of the first things many people engaged in was to explore possible interaction provided by the system (Figure 12). They often stopped and walked back and forth to explore how the display would react. Some users tried to touch the display, and some users even engaged in expressive gestures as if conducting to see whether the display reacted (Figure 12). The decision to render content at $45°$ in walking direction made the correlation between the users' movement and the content more complex. When users stopped, the content moved in front of them, and when they started moving, it moved at $45°$ in walking direction. This seemed to confuse users enough that most initially thought it was reacting to them, but then decided otherwise after playing with it.

Testing the limits
After exploring normal interaction, many users quickly proceeded to test the limits of the system. They did so by moving abruptly or quickly, or by walking to the very edges of the display to find the maximum angle for the tracking. Interestingly, when testing the limits of the system they often did so in a very theatrical way. For example, when moving quickly, they did so in a very demonstrative way, as to provide an explanation for their behavior to bystanders.

Inspecting the front
Passers-by spent considerable attention towards the camera setup of the display. Some even approached the cameras without interacting first. They did so when passing-by, noticing the cameras, but also after observing others interact from a shop at the opposite side of the hallway. They sometimes called friends and discussed the camera setup. The cameras seemed to attract almost more attention than the large screen or the interaction they supported.

Manipulating the system
Besides just looking at the cameras, many users proceeded hiding the cameras with their hands to see what happened. Because there were two camera locations, tracking mostly continued to work, apparently irritating some users. Nobody actually touched the cameras. They seemed to strongly afford blocking their view, and this seems to be not socially inappropriate. When interviewed, some users actually thought that the camera setups would project the image on the screen and wondered why they continued to see the image when they covered what they thought to be projectors.

Behind the screen
Some users spent considerable time looking *behind* the screen. Some of them seemed rather disappointed to see only projectors. Others however observed the projectors for some time. Sometimes small groups congregated at the edge of the screen, discussing the setup. In interviews, people also had a wide range of mental models of how the system worked. One user who quickly aborted interaction for example thought that it was a front projection and did not want to stand in the projection in front of the screen (although there were no projectors in front of the screen).

Design recommendations
Novices passing by screens like Screenfinity will probably split in three groups. Some will ignore it entirely, some will read text without slowing down, and others will engage in technology exploration behavior. Depending on whether one wants to attract attention and make passers-by stop, one should hide the technology or reveal it, and make interaction more or less deterministic. When the technology is visible, users will explore it, thereby stopping and creating a Honeypot effect. The same happens if the interaction is not very deterministic, as is the case if the content is rendered in the walking direction of users (and thus moves if the user turns around). If technology is hidden, probably less people will stop and explore it, but some may spend even more time exploring because of the bigger challenge. For both cases, revealing and hiding of technology should be carefully considered by the designer.

CONCLUSION AND FUTURE WORK
In this paper, we made two contributions. As a technical contribution, we proposed Screenfinity, a novel public display that increases the perception area by proposing a novel combination of visualization techniques as well a novel sensor merging 10 depth cameras. Screenfinity supports different content for different users, increases the perception area of content and enables passers-by to read more content while walking. Screenfinity has been deployed in a field study during four days. Our findings show that even first-time users could read content while walking. Many passers-by also engaged in extensive technology exploration behavior.

We also contributed to the theory of public displays and visual interfaces. We propose a model predicting the shape of the perception area depending on corrected or non-corrected perspective visualization which has been empirically validated. We extend this model to predict the quantity of content that

users can read when passing-by a public display. We empirically validated these models with a laboratory study.

We believe that our work can be extended to other large interactive surfaces such as interactive floors [1] or interactive ceilings [24]. We think our models are still valid for these contexts of interaction but further investigation is needed to validate them. While we focus on displays that extend in mostly one dimension (very wide), floors and ceilings are two-dimensional.

ACKNOWLEDGMENTS

REFERENCES

1. Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff, T., Becker, T., Holz, C., and Baudisch, P. Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input. ACM UIST'10 (2010), 209–218.

2. Barber, P. J., and Dickenson, A. Factors affecting visibility distances for poster panels. Technical report (1997), POSTAR, http://www.postar.co.uk/research/visibility-study/.

3. Benedict, M. L. To take hold of space: Isovists and isovists fields. Environment and Planning (1979), 6:47–65.

4. Bezerianos, A., Dragicevic, P., and Balakrishnan, R. Mnemonic rendering: an image-based approach for exposing hidden changes in dynamic displays. ACM UIST '06 (2006), 159–168.

5. Cannon, S., and Robinson, D. Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey. *Journal of Neurophysiology 57*, 5 (1987), 1383–1409.

6. Czerwinski, M., Robertson, G., Meyers, B., Smith, G., Robbins, D., and Tan, D. Large display research overview. ACM CHI EA '06 (2006), 69–74.

7. Dalsgaard, P., and Halskov, K. Designing urban media facades: cases and challenges. ACM CHI '10 (2010), 2277–2286.

8. Dalton, S. N., Marshall, P., and Dalton, R. C. Measuring environments for public displays: a space syntax approach. ACM CHI EA '10 (2010), 3841–3846.

9. Fischer, P. T., and Hornecker, E. Urban hci: spatial aspects in the design of shared encounters for media facades. ACM CHI '12 (2012), 307–316.

10. Fisher, P. F. Extending the applicability of viewsheds in landscape planning. *Photogrammetric Engineering & Remote Sensing 62*, 11 (Nov. 1996), 1297–1302.

11. Frees, S., Kessler, G. D., and Kay, E. Prism interaction for enhancing control in immersive virtual environments. *ACM Trans. Comput.-Hum. Interact. 14*, 1 (May 2007).

12. Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. Proxemic interactions: the new ubicomp? *ACM interactions 18*, 1 (Jan. 2011), 42–50.

13. Grossman, T., and Balakrishnan, R. Collaborative interaction with volumetric displays. ACM CHI '08 (2008), 383–392.

14. Ju, W., Lee, B. A., and Klemmer, S. R. Range: exploring implicit interaction through electronic whiteboard design. ACM CSCW '08 (2008), 17–26.

15. Khan, A., Matejka, J., Fitzmaurice, G., and Kurtenbach, G. Spotlight: directing users' attention on large displays. ACM CHI '05 (2005), 791–798.

16. Larson, K., van Dantzich, M., Czerwinski, M., and Robertson, G. Text in 3d: some legibility results. ACM CHI EA '00 (2000), 145–146.

17. Maksakov, E., Booth, K. S., and Hawkey, K. Whale tank virtual reality. GI '10, Canadian Information Processing Society (2010), 185–192.

18. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., and Davies, M. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. ACM CHI '11 (2011), 3033–3042.

19. McCallum, D. C., and Irani, P. Arc-pad: absolute +relative cursor positioning for large displays with a mobile touchscreen. ACM UIST '09 (2009), 153–156.

20. Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking glass: a field study on noticing interactivity of a shop window. ACM CHI '12 (2012), 297–306.

21. Nacenta, M. A., Sakurai, S., Yamaguchi, T., Miki, Y., Itoh, Y., Kitamura, Y., Subramanian, S., and Gutwin, C. E-conic: a perspective-aware interface for multi-display environments. ACM UIST '07 (2007), 279–288.

22. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. ACM CHI '08 (2008), 1285–1294.

23. Ten Koppel, M., Bailly, G., Müller, J., and Walter, R. Chained displays: configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. ACM CHI '12 (2012), 317–326.

24. Tomitsch, M. *Interactive Ceiling: ambient Information Display for Architectual Envronments*. PhD thesis, 2008.

25. Vogel, D., and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. ACM UIST'04 (2004), 137–146.

26. Wang, M., Boring, S., and Greenberg, S. Proxemic peddler: a public advertising display that captures and preserves the attention of a passerby. ACM PerDis'12 (2012), 3:1–3:6.

27. Ware, C. *Information Visualization: Perception for Design*. The Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufman, 2004.