Visualization in Multimodal User Interfaces of Mobile Applications

(Draft Paper)

G. Fuchs, H. Schumann

May 04, 2005

1 Introduction

Advanced user interfaces are a crucial factor in the success of mobile information systems employed by different users on a variety of devices. They should provide state-of-the-art visualization, speech and interaction techniques tailored for specific tasks, while at the same time allowing flexible deployment of these components on a multitude of (mobile) hardware platforms. With the ongoing pervasion of mobile devices and wireless networks, the use of digital manuals, location-based services and information portals is about to become feasible with commonplace devices.

The goal of visualization is to transform large volumes of data into meaningful images that allow us to intuitively understand, to gain new insight, and to reveal previously unknown patterns or correlations. The ratio behind this is the ability of the human mind to grasp visual impressions faster, and also to remember them better than information heard or read.

The requirements differ whether a visualization is employed to explore a new data set of unknown characteristics, to verify a certain theory a user may have about the data, or to ultimately create an expressive illustration of interesting findings. During the exploratory phase, a user will start with a representation affording a good overview of the entire data, then zoom into selected areas to filter out interesting information and finally, have a detailed look at the phenomena found. Thus, visualization typically is a highly interactive process.

The transformation of abstract (i.e. inherently non-graphical) data into a visual representation is done by the so-called *visualization pipeline*. It consists of three stages: *filtering*, *mapping* and *rendering*. In the first stage, the raw data is filtered for the desired information, which can include e.g. threshold operations and the calculation of statistical values like derivation or correlation coefficients. It does also include the detection and correction of illegal values resulting e.g. from measuring errors. In the mapping stage, the abstract data is then mapped to geometric properties like position, shape or size, and visual attributes like color, opacity or saturation. Finally, this geometry with its attributes is rendered (to screen or an image) to obtain the visual representation of the information.

Above all, a visualization must meet three criteria. It must be *expressive*, i.e. show all relevant – and only these – information. It must be *effective* in that all relevant aspects of the data are perceived intuitively. Last but not least, a visualization must be *adequate*. Creating the visual representation must not require an excessive amount of time and/or system resources.

One problem that may arise is that a graphical representation becomes cluttered. In short, this will occur when it is attempted to encode too much information in too small an image (screen space), i.e. when the information density becomes too high. The expressiveness and effectiveness of cluttered representations drops sharply, to the point of being useless. This is especially an issue on mobile devices with their comparatively small displays.

With the screen size a given invariant of the sys-

Dagstuhl Seminar Proceedings 05181
Mobile Computing and Ambient Intelligence: The Challenge of Multimedia
http://drops.dagstuhl.de/opus/volltexte/2005/367

tem, a visualization must therefore manage the information density in order to maintain effectivity. This can either be achieved by filtering out information that is not relevant to the user at a certain point of time, or by communication the information via additional output modes.

Underpinning this paper is a real-world scenario, the mobile support for a maintenance worker of air-condition units. An 'electronic maintenance manual' is to provide the worker with information on schematics, operational and safety procedures and assembly instructions much like a printed manual would, albeit being much less bulky. Besides visual representation, the electronic manual also makes uses of speech as an alternate output mode. Underlying the manual is a task model [RFD04] that exactly determines which task should be performed on a component at a given time. Since e.g. schematic diagrams can become quite large and complex, and the hardware used may vary, the aforementioned challenges to visualization do apply.

Therefore the next section discusses a family of techniques used for information filtering, followed by examples on how alternate input/output modes can be used to support effective visualizations for this scenario in section 3.

2 Focus & Context

Focus and Context techniques (see e.g. [LA94]) combine a focus region that displays a part of the whole image with high detail, and a surrounding context region that presents information in lower detail to provide an overview.

Depending from the distance of the focus, the context usually is distorted to accommodate for the increased space requirements of the more detailed focus. However, it is important to note that the detail of the representation is not confined to graphical magnification. Keahey [Kea98] therefore introduced the concept of a detail axis. The degree of detail can be based on entirely semantic properties, like showing only average values or cluster centroids in the context, and individual values within the focus. Figure 1 shows two examples of focus & context displays.

Figure 1: Examples of Focus & Context techniques. Left, a distorted 'fisheye view' of a map; right, a web browser where page section in the context are collapsed into a button labeled with a meaningful phrase.

These concepts can be used for information representation on mobile handhelds as well (see e.g. [BMPW00, BHR+99]). However, a good trade off between complexity and response time is important on these devices, since computing arbitrary, global distortion functions can be very expensive. In this case a closely related approach, the so-called lens techniques, shows to be very useful.

A lens is parameterized by its position 'above' the visual representation, its shape, and the effect it has on the underlying representation [Gri04]. As shown e.g. in [SFB94] there is a wide range of effect functionalities available, including a purely graphical modification, or again by operating on a semantic level[1]. Spatial transformations (magnification, distortion), the variation of visual variables and the relocation of represented objects are graphic techniques. In contrast, an increase or a reduction of information content, as adding annotations or information hiding, are techniques with semantic effects. Since all these functions are realized for a small, confined region only, they can be provided very efficiently.

Using Focus & Context or lens techniques generally requires a high amount of interactivity to move and

[1] These can be classified as requiring a modification of the filter stage parameters of the visualization pipeline.

redefine the focus area and/or lens parameters. However, in the outlined scenario a visualization pertains to the task currently performed.

Thus, the parametrization of either technique can be automated. The importance of information can be derived from the task description, thus creating task-dependent *Level of Details* (LoD) for different regions of the visualization. Figure 2 shows an example where different regions of a technical illustration have been rendered with varying styles relating to their importance for the current task.

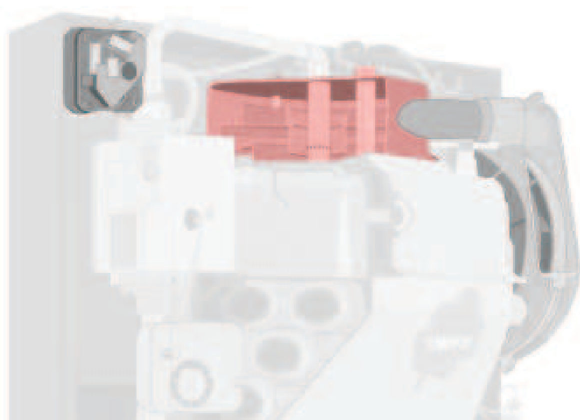

Figure 2: Example for an adapted technical illustration where components of the device have been highlighted according to their importance for a given task.

By the same token, if using a confined lens its position can be derived automatically by the task model. For example, the lens is initially positioned above the most important region or (in a technical illustration) the component that has to be handled next.

Also, if using additional output modes the information in the task model can be used to synchronize the visual focus with the information the user obtains from other the modes. For example, the lens will always center on the object the current narration is about. This aspect will be discussed in the next section.

3 Multimodality and Visualization

The purpose of employing multimodal communication during visualizations is two-fold. First, having other, non-graphical means (e.g. speech) available for output reduces the amount of information that has to be graphically encoded and placed into the visual representation of the data. This in turn reduces the information density.

Second, if providing input methods that are not dependent on graphical user interface (GUI) components, like voice commands, no screen space needs to be allocated for these components (e.g. zoom buttons), yielding more room for the visualization itself. In addition, available input devices differ with the platform used. While a laptop has a fully-featured keyboard and usually a two-button mouse, on a PDA there is only the stylus pen with a virtual keyboard, down to only a T9-keypad on some smartphones. Thus, the entering text (e.g. an arbitrary search expression) is easy on the laptop, feasible on the PDA, but cumbersome at best on the smartphone. Here, voice input would be a much better alternative.

Generally, the decision on what information is communicated by which of the three modes is based on the type of data, the goal of the visualization and the mode's suitability for that purpose. In the scope of this contribution, we consider the modes: visual (graphic), textual and verbal (speech) output.

Visualization is best suited for an initial overview of the data set, in order to find patterns, structures and correlations as well as spatial relationships. It is generally less suited[2] to gather accurate value readings.

Spoken text (by means of voice synthesis) is better suited for the description of activities or procedures, rather than reading out specific values. As such, using speech should possibly used more often in the presentation phase, e.g. to draw the audience's attention to certain aspects, provide background information on how certain results where achieved, or to give assembly instructions associated with a techni-

[2]With the exception of labeled-axis diagrams.

cal illustration in an electronic maintenance manual. While in the latter case a video might be even more appropriate, it might not be available e.g. due to bandwith constraints

As speech is a transient medium, it must also be decided when a piece of information is presented verbally, and if and how it may be repeated if the user requests it. For example, short statements should be repeated exactly the same, whereas longer texts should be modified somewhat in order to more closely resemble the way a human speaker would answer a repeated question. Again, such meta information need to be stored in the underlying task model.

Text is another graphical output mode. It may simply be printed text, or labels (annotations) within a graphical representation. In fact, a number of visualizations actually uses text labels to illustrate specific points in an image. However, the main purpose of using visualization is to relieve the user from the cognitive effort to read and parse single data items and then to interpret the findings as a whole (e.g. analyzing a spreadsheet).

Therefore in the context of this contribution, labels can be seen as a way to communicate additional information, e.g. precise value readings that are hard to obtain from e.g. color shades, and tiresome to listen to if read out. However, an indiscriminate labeling of too many data points would defeat the sense of visual representation.

The main benefit of labels (as a graphical element) can therefore be seen in the ability to create links between the visualization and non-graphical modes. An example is to use labels containing keywords in conjunction with speech I/O. This indicates to the user that there is additional, spoken information available for a certain element, and that it can be obtained by saying the keyword. The font size might also be used to indicate the relative importance of labeled elements, cf. figure 3.

Labeling is an important aspect in cartography. Very few maps can do without proper annotations. As such, algorithms for (automatic) label placement are legion. There are, however, differences between labeling maps and visualizations.

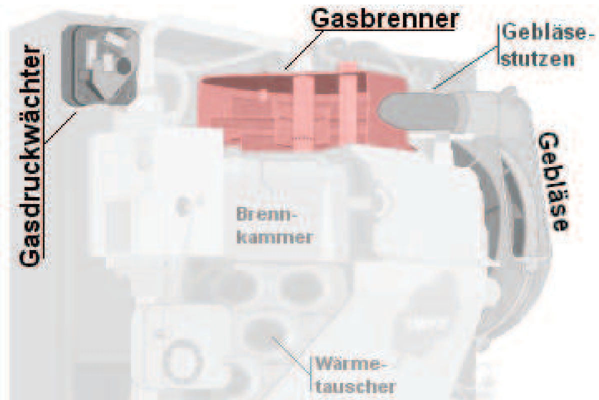

Figure 3: The illustration from figure 2 with labels. More important components of the device are labeled more prominently.

Most important is the competition between the graphical elements of the visualization and the labels. On a map, features (e.g. lines) may be partially covered or interrupted by labels and still be legible. Contrary to this, the elements of a visualization encode data, so altering shapes or area sizes by occlusion might result into erroneous perception, violating expressiveness. Thus, the label placement is even more crucial in information visualization than it is in cartography. Adding to this is the requirement that visualization generally be interactive, so label placement must be done in real-time with acceptable quality.

An aspect that must be considered with any multimodal presentation is to establish the correct linking between the visual representation and the other modes, especially when using focus & context or lens interaction techniques. The 'visual focus' should usually match that of the other mode(s). For example, the 'acoustic focus', i.e. the verbal information being read out to the user should relate to the object that is currently in the visual focus. Likewise, the perceived importance of text labels (e.g. by font size) within a visualization should be the same as the importance values used to create the focus & context distortion (cf. figure 3).

If this condition is not met, that fact must be obvious to the user in order to avoid erroneous association of visual, vocal and textual information. Otherwise, adding speech and/or text would cause confusion instead of helping the user to understand.

4 Summary

Mobile information systems suitable for a variety of user groups and hardware platforms require, among other things, state-of-the-art, interactive visualization techniques. The main challenges for creating useful visualizations in this context are the vastly different capabilites of the mobile devices in terms of computational power, storage capacity and input/output facilities as well as their generally very limited screen size and resolution.

In order to maintain expressiveness and effectiveness, the information density of created representations must be carefully managed. One way to to this is to employ focus & context or lens techniques which allow the user to interactively explore regions of interest in the data at maximum detail, yet affording him a comprehensive overview for orientation.

Another solution is to link the visualization with other input/output modes like speech. Communication information verbally reduces the information density of visualizations.

The availability of voice commands, on the other hand, enables the designer to omit GUI elements like buttons, in turn reserving more screen space for the visualization. Also, dictating words to the machine is very helpful if a keyboard is not available.

Linking the visualization with other modes is critical. The user must be able to correctly associate information presented visually with the information fragments communicated by sound. One way to do this is using text labels in the visualization. However, labeling as a complex task is problematic in its own.

A combination of the three modes visualization, speech and text has been examined for a scenario on mobile mantenance of air-condition units. A software prototype for this 'electronic maintenance manual' was presented on the CeBIT 2005 fair.

Future works will include further studies on how speech, visualization and labeling are combined best. This includes questions on what kind of information is best delegated to graphical, textual and vocal output.

References

[BHR+99] S. Björk, L.E. Holmquist, J. Redström, Bretan, R. I. Danielson, J. Karlgren, and K. Franzen. West: a Web Browser for small terminals. In *ACM Conference on User Interface Software and Technology 1999*. ACM Press, 1999.

[BMPW00] O. Buyukokten, H.G. Molina, A. Paepcke, and T. Winograd. Power Browser: Efficient web browsing for PDAs. Technical report, Stanford University, Stanford, 2000.

[Gri04] Henning Griethe. Einsatz von Linsentechniken für Ikonen über interaktiven Kartendarstellungen. MSc thesis, University of Rostock, 2004.

[Kea98] T. Alan Keahey. The Generalized Detail-in-Context Problem. In *Proceedings of the IEEE Symposium on Information Visualization*, IEEE Visualization, October 1998.

[LA94] Y. K. Leung and M. D. Apperley. A Review and Taxonomy of Distortion-Oriented Presentation Techniques. *ACM Transactions on Human-Computer Interaction*, 1(2):126–160, June 1994.

[RFD04] Daniel Reichart, Peter Forbrig, and Anke Dittmar. Task Models as Basis for Requirements Engineering and Software Execution. In *Proceedings of Tamodia*, pages 51–58, Prague, Czech Republic, 2004.

[SFB94] Maureen Stone, Ken Fishkin, and Eric Bier. The Movable Filter as a User Interface Tool. In *Proceedings CHI 94*, pages 306–312. ACM, 1994.