Investigating Gesture and Pressure Interaction with a 3D Display

Martin Halvey, David Hannah, Graham Wilson, and Stephen A. Brewster

School of Computing Science, University of Glasgow, Glasgow, G12 8QQ, UK
{first.last}@glasgow.ac.uk, gawilson@dcs.gla.ac.uk

Abstract. We examine the use of a mobile device to provide multifunctional input and output for a stereoscopic 3D television (TV) display. Through a number of example applications, we demonstrate how a combination of gestural and haptic input (touch and pressure) can be successfully deployed to allow the user to navigate a complex information space (multimedia and TV content), while at the same time visual feedback can be used to provide additional information to the user enriching the experience. In order to investigate the usefulness of our example applications a user evaluation was conducted, where our prototypes were compared with more traditional devices for multimedia interaction. The results of the user evaluations highlight the benefits of our approach and also provide some design guidelines.

Keywords: Mobile devices, 3D, TV, pressure, gestures.

1 Introduction

Multimedia content is now an important part of everyday life and the TV is fast becoming a hub for interacting with much of this content, including increasingly complex media such as 3D video, video/photo collections, social media through the internet etc. It is important that we provide users with the best options for browsing, searching and consuming this growing and complex content. In many cases, the TV remote control is itself a limiting factor, as normally it only provides simple fixed buttons for interaction, indeed many households also have several media devices each with a different remote. As such mobile devices may provide a suitable alternative to the traditional remote, especially as a high number of phone users already interact with their phone while watching TV[1]. Initial research into using a mobile phone as a universal interaction device [1] highlighted that perhaps universal control over all appliances might not be ideal but that control over particular appliances might be beneficial. Therefore the aim of this work is to focus on the specific use of mobile phones to provide richer interactions with media on a TV screen. Currently, there are applications that run on mobile phones which provide access to content, but many of these applications simply replicate the 'look and feel' of a regular TV remote. As such, these applications do not take full advantage of all of the features that a mobile phone can offer which a normal TV remote cannot offer. These include, but and are

[1] Nielsen/Yahoo: "Mobile Shopping Framework", November 2010.

K. Schoeffmann et al. (Eds.): MMM 2012, LNCS 7131, pp. 395–405, 2012.

not limited to: (1) a local display for additional information, be it visual, audio or haptic, (2) potential for other types of input e.g. gestures on or with the device, audio input, etc. and (3) multi-user interaction, as many users in the same space would have devices that could be used to control the TV or to share information, etc.

In this paper we investigate the first two points. We examine the use of gestures and pressure input as novel input modes to control the display, the ability of users to use the visual display on the phone while interacting in this fashion and the use of different locations on the phone as locations for input. In this work we use a 3D display as it represents the current state of the art in televisions and, although it represents a more complex information visualisation to navigate, the use of 3D spatial layouts may better illustrate content structure/relationships.

2 Related Work

In this section we briefly describe research in a number of areas related to our research. Gestures are increasingly being used for interaction with mobile devices. These gestures can be loosely classified into 'discrete action' gestures and 'continuous control' gestures. The more traditional style of gesturing, discrete action control, involves the user performing an action that, once completed, the system attempts to recognise. These gestures are often used to replace one or multiple physical button clicks. Successful examples of these techniques in commercial devices often involve short movements that are fast and easy to perform such as a single stroke of a touch screen to change the view, or double tapping a phone to silence it [2]. Recently, continuous control gestures have become more prevalent. With these forms of gesture the interactions between the user and the device are closely coupled. The user provides a continuously changing stream of input, and the device adjusts the feedback constantly to respond to the user's input. Both types of gestures can be performed with a device as well as on a touchscreen e.g. by shaking or tilting the device [3]. Pressure based interaction is a relatively new area of research. Some initial work by Ramos *et al.* [4] used a pressure-sensitive stylus as a means of controlling interaction widgets. They concluded that an interaction that requires both positioning/movement and the application of pressure (specifically through the same device) should separate the two actions as much as possible so as not to interfere with either. Wilson *et al.* [5] demonstrated that eyes free pressure interaction while squeezing a mobile device was feasible and almost as accurate as when using visual feedback. Brewster and Hughes [6] asked users to input text on a pressure-sensitive screen, using light touches for lower-case letters and harder presses for uppercase letters. They found that the pressure-augmented keyboard resulted in faster but more error-prone text entry. Clarkson *et al.* [7] suggest further uses for pressure-augmented keypads such as preview zooming, 3D navigation or "affective input" where emotional state is derived from the degree of force used in an interaction.

3 Hardware

We simulated a 3D TV using a PC with an Nvidia 3D Vision graphics card and active shutter glasses. Users interact with the 3D display using a Nexus One mobile phone connected to the PC via Bluetooth, providing a tangible remote controller for the 3D display. Users can use its touchscreen as proxy for interacting with the data on the TV. The touchscreen can also act as an additional display. Previous research [8] has identified a number of potential benefits to having secondary screen available via a phone interface for multimedia interaction, namely additional control, methods to enrich content, additional ways to share content, and finally the ability to transfer television content. Users can also use the Nexus device to perform device movement gestures, as it includes an accelerometer. Dachselt and Buchholz [9] have previously studied throw and tilt interactions with remote displays, they investigated continuous and discrete tilt gestures in two media interaction environments including a 3D map for Google Earth. However, these interactions did not include the TV or multiple types of media. In addition to the standard input and output modalities currently available on a mobile device, we added pressure input, which can add a z-dimension to typically 2D, x-y GUIs. Pressure input has been shown to be effective at improving interaction [5] and could be included in future mobile devices. Pressure input was provided by using two standard Force Sensing Resistors (FSR's) and a linearising amplifier [5]. This allowed users to push at two different points on the back of the device. In this way we have taken advantage of the additional space on the back of the device for controlling the television which allows interaction without obscuring the touchscreen, as outlined by Baudisch and Chu [10]. More detailed descriptions of the hardware used can be found in Hannah *et al.* [11].

4 Interfaces for Media Interaction

4.1 Browsing Image Collections

Fig. 1 (A) shows an example of a visualisation which allows the user to view a collection of images in a similar way to the iTunes Cover Flow, but with the content stretching back into the screen in 3D. Users can navigate through the image stream by tilting the device to move backwards and forwards through the image queue. As the user rotates the phone accelerometer data is filtered and commands sent to the TV to move forwards or backwards. One drawback of the visualisation outlined in Fig. 1(A) is that while it takes advantage of the 3D space it can be difficult to view images further back in the queue. Pressure input could be used to overcome this problem, as Fig. 1(B) shows a version of one of Ramos *et al.*'s [4] pressure widgets in use. Here, the stream of photos on the TV dynamically kinks as pressure is applied on the FSRs; images further down the queue can then be more easily seen. This exploits both the easy interaction with the mobile device and the additional visual space allowed by the 3D display to show more information. Again it is possible for users to see different views of the information on the local display on their mobile device. Fig. 1(C) shows an alternative solution to the occlusion problem outlined above. The same collection of images is arranged in spiral visualization with the current image in front and the tail of the spiral going into the screen in 3D. As the user rotates the phone accelerometer data is used to rotate the spiral forwards or backwards, with different photos being brought into focus at the top of the spiral. Rate of rotation is proportional to the angle of tilt of the phone. The use of the mobile phone also means that the user can see a replication of the TV display on the phone screen in 2D; such as a replication of the spiral/Cover Flow view or, alternatively (as shown in Fig. 1(D)) users can see the image at the top of the data queue on the mobile device; this allows the user to see a local version of the image as well as some metadata. This interaction example demonstrates some of the flexibility of using a phone as a remote controller; it can duplicate the TV screen, or individual users can have individual displays of information, allowing multiple views of the same data through multiple devices, realising some of the possibilities outlined by Cesar *et al.* [8].

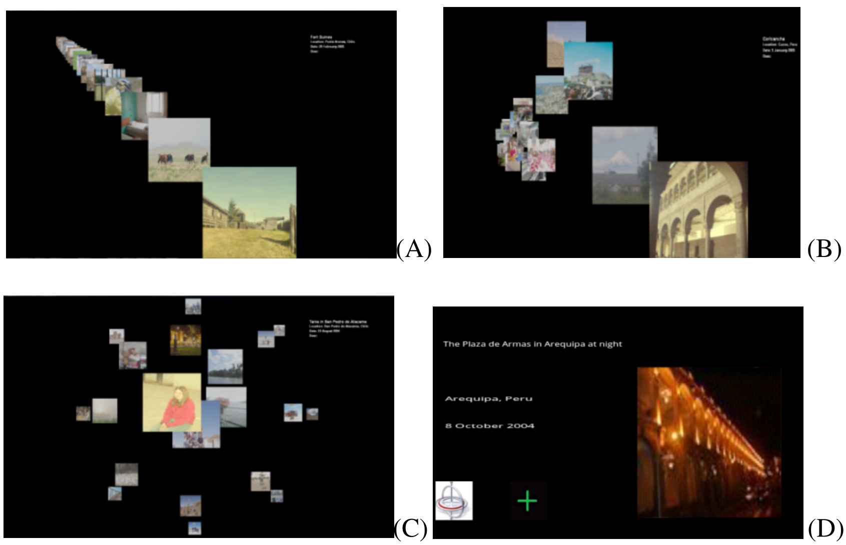

Fig. 1. 3D visualization of a collection of images (A), pressure input on the mobile device causes images on the 3D display to kink out to allow images at the back to be viewed more clearly (B), 3D visualization of a collection of images in a spiral (C) and visualisation of image at top of the queue as displayed on phone (D)

4.2 EPG Browsing

The input methods described above can also be used to navigate through more complex structures. Richer interactions can also be used to control an Electronic Programme Guide (EPG). EPGs contain a lot of dense information and, when many TV channels are available, can often be difficult and slow to navigate using a regular TV remote. There are already applications to allow users to browse EPGs on additional devices, e.g. tablets or phone; however these are not always connected to the TV. Fig. 2(A) shows a possible 3D representation of an EPG. A fisheye view in 3D gives the user an idea of programmes close to the time being viewed; users can see and compare content easily and efficiently as the visualization makes full use of the TV screen. The program information is modelled as a spiral viewed from the side, where a 24 hour period is displayed on a full 360° rotation of the spiral, with channels on the y axis. The spiral view means that there are no discontinuities between the InvINdays; the next day joins smoothly to the previous one. In many EPGs different days are presented on different screens, making the transition from one day to the next more awkward. The viewer can preview what is on earlier or later by rotating the phone side-to-side to rotate the spiral. Tilting the phone forward or back moves up or down through the channels. By squeezing on the FSRs on the back of the phone the user can control the scaling of the view, for example zooming in to a particular time/channel area. Pressing harder causes the view to "pop" through to the following day at the same time, allowing rapid skipping through different days. A particular programme can be selected by tapping on the phone screen, either through a button on the remote (Fig 2(C)) or the EPG (Fig. 2(B)). This brings up a page about the show on the phone and allows the it to be played on the 3D TV. In order to provide a comparison with more traditional EPG browsing a flat EPG (Fig. 2(B)) and a basic remote control were also implemented on the mobile device (Fig. 2(C)). Both the 3D and 2D displays could be controlled via gestures with the mobile device or by buttons on the remote control on the mobile device.

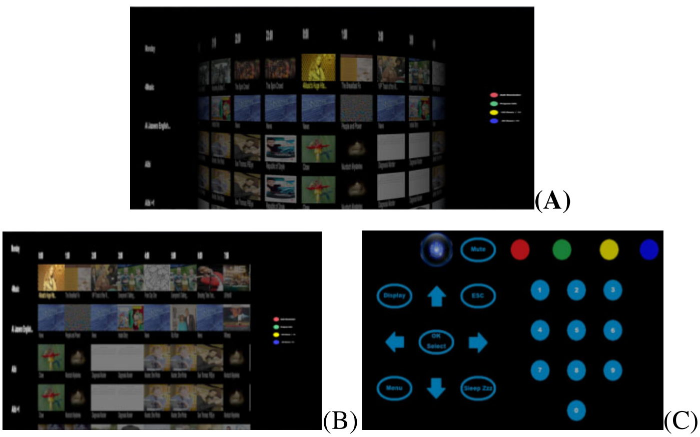

Fig. 2. An EPG displayed as a fisheye on a cylinder in 3D (A), a flat 2D EPG with the same information (B) and an example TV remote control app on a mobile phone (C)

5 User Evaluations

5.1 System Configuration

Image Browsing

For the user evaluation six different system configurations were used, based on the TV visualisation (linear or spiral), phone screen content (replication of TV content or local content) and the use of pressure (kinking or none).

- S1: Linear visualisation on 3D display (Fig. 1A), entire queue visualised on phone screen in 2D, no pressure input

- S2: Linear visualisation on 3D display (Fig. 1A), image at top of queue displayed on phone screen (Fig. 1D), no pressure input
- S3: Linear visualisation 3D display with pressure input for kinking (Fig. 1B), entire queue visualised on phone screen in 2D
- S4: Linear visualisation 3D display with pressure input for kinking (Fig. 1B), image at top of queue displayed on phone screen (Fig. 1D)
- S5: Spiral visualisation on 3D display (Fig. 1C), entire queue visualised on phone screen in 2D, no pressure input
- S6: Spiral visualisation on 3D display (Fig. 1C), image at top of queue displayed on phone screen (Fig. 1D), no pressure input.

EPG Browsing

For the EPG evaluation four different configurations were used, based on visualisation (3D or flat/2D) and input type (gestures or touchscreen remote):

- S1: 3D Spiral with gestures (Fig. 2A and gestures)
- S2: 3D Spiral with remote (Fig. 2A and Fig. 2C)
- S3: Flat with gestures (Fig. 2B with gestures)
- S4: Flat with remote (Fig. 2B and Fig. 2C)

5.2 Collection and Tasks

Image Browsing

For the experiments reported in this paper the CLEF 2007 image collection was chosen [12]. CLEF 2007 is a set of 20,000 images, 60 search topics, and associated relevance judgments. The topics were categorised into a number of different categories, including: easy/hard, semantic/visual, and geographic/general. In order to choose tasks which were as similar as possible, 6 tasks from the medium and visual categories were chosen. For each of the tasks, 25 relevant images were chosen as well as 100 irrelevant images from the remainder of the collection. For each topic the users were presented with these 125 images in a random order and were asked to navigate through the stream and mark images as being relevant.

EPG Browsing

For the EPG browsing tasks, one week of an online TV guide was downloaded. The guide contained 255 channels. The content was normalised so that each TV program corresponded to one hour in the TV guide. Twelve navigation tasks were created. For 8 of these the participant had to navigate to a particular channel on a particular day at a particular time and set a reminder for the content (by pressing a button on the mobile device), this marked the end of the task. 4 of the 8 tasks required the participants to navigate through the collection in more of a horizontal than a vertical (time based) direction. The other 4 of the 8 tasks required the participants to navigate through the collection in more of a vertical than a horizontal (channel based) direction. All of the tasks were designed so that the participants had to navigate through an equidistant amount of the EPG. The final four tasks were free browsing tasks where the participants were given a broad time period e.g. Tuesday morning, and they were asked to set a reminder for any program of their choice in that time period. For each system configuration the participants were given 1 horizontal navigation task, 1 vertical navigation task and 1free browsing task.

5.3 Experimental Design

The experiment lasted around 90 minutes and the participants received compensation of £10. Following a training period the participants carried out six image browsing tasks using all 6 configurations (giving 36 tasks), in a within subject design. The order of system configuration was rotated and the topic order was randomised. There was a maximum time limit for each task which was 5 minutes; participants could finish early if they wished. The participants then carried out the 12 EPG browsing tasks. Again system configuration was rotated and search tasks were randomised. Participant interactions with the display and mobile device were logged. They also filled out a number of questionnaires at different stages of the experiment. Throughout the evaluation the participants were encouraged to comment on the system and interactions, and notes were taken of any comments made. We were able to calculate precision and recall values for all of the image browsing tasks. We also counted the number of gestures required to complete each task for each system.

6 Results

17 participants took part in the evaluation. They were mostly staff and students of the University. The participants consisted of 11 males and 6 females, with an average age of 29 years old. In the entry questionnaire participants indicated that they regularly interacted with images and also regularly watched television.

6.1 Task Performance: Image Browsing

Participants successfully completed the tasks, with the average precision being quite high (Max=98.33% (S2), Min= 95.77% (S6), Avg= 96.9%). A one factor ANOVA showed no significant effect of system on precision (p=0.231) and pairwise comparisons showed no differences. The recall of the systems was slightly lower than the precision, but still high (Max= 91.17% (S5), Min= 77.05% (S2), Avg= 84.85%). A one factor ANOVA showed no significant effect of system on recall (p=0.175) and pairwise comparisons showed no differences. In terms of the average time to complete each task, the systems where users had the images on the mobile phone instead of the replicated stream took slightly more time to complete the task (Max= 255.84sec (S6), Min= 201.53 sec (S2), Avg= 233.74 sec). Pairwise comparisons shows significant differences between S6 and both S3 (p=0.03) and S4 (p=0.004). The number of gesture rotations to find items was approximately the same across all systems (Max= 171.11 (S5), Min= 142 (S2), Avg= 153.42), with no statistically significant results.

6.2 Task Performance: EPG Browsing

Due to a technical error, the interactions for the first 4 participants were not logged correctly so only the logs for the remaining 13 were used for analysis (for user preference in the next section the responses from all 17 were used). An analysis of user performance in terms of setting a reminder showed that there was little difference between systems. Out of two direct tasks (i.e. vertical reminder task and horizontal reminder task) the average number completed successfully was 1.38 (std. dev=0.63), 1.84(std. dev=0.8), 1.31(std. dev=0.38) and 1.15(std. dev=0.76) for S1, S2, S3 and S4 respectively. Surprisingly S4 the system that is most close to current EPG browsing was the worst performing. None of the differences were statistically significant. An analysis of the effort involved in terms of average time to set a reminder and number of gestures (either rotations or button presses where appropriate) revealed some differences between the systems. With respect to time to set a reminder, the average times in seconds were 81.33 (std. dev=34.43), 80.97(std. dev=56.59), 76.68(std. dev=41.84), 83.95(std. dev=54.39), and for S1, S2, S3 and S4.

Table 1. Rotation and button moves per system

	S1.Moves	S2.Moves	S3.Moves	S4.Moves
Mean	129.0000	85.8684	121.3077	80.6857
Std. Deviation	118.90536	52.16200	81.93937	55.67541

In terms of actions, it required more gestures than button presses to set reminders. The results of pair wise comparisons between the systems showed that the differences between S1 (3D with gestures) and S2 (3D with remote) was statistically significant (p=0.033). The difference in terms of time and interactions between the systems using buttons and gestures can be explained by the feedback given by users during the experiment. Many users commented that the gestures were good for moving large distances through the EPG easily, but more difficult for fine control i.e. moving one or two time slots or programs. In contrast they commented that the buttons were easy for fine control but not so good to use for moving large distances in the EPG.

6.3 User Preferences

Gestures

In post-search task questionnaires we solicited subjects' opinions on the use of gestural interaction to navigate both the images collections and the EPG. The following Likert 5-point scales and semantic differentials were used; some of the scales were inverted to reduce bias. The scales used were: "How easy was it to use the system" (Use), "How easy was it to learn to use the system" (Learn to use), "When interacting with the system I felt in control/not in control, comfortable/uncomfortable, confident/ unconfident". The following Semantic differentials were used: The videos I have received through the searches were: "wonderful/terrible", "satisfying/frustrating", "stimulating/dull", "easy/difficult", "flexible/rigid", "efficient/inefficient", "novel/standard", "effective/ineffective. Many of these scales and semantic differentials are used for different aspects of the system. The participants generally gave positive responses with respect to the use of gestures for all differentials and questions. It was noticeable that the participants were generally more positive for image browsing than EPG browsing. This is not totally surprising, as for the image browsing the users only had one degree of freedom for navigation, whereas the EPG browsing had two degrees of freedom, meaning that it was a more complex form of navigation. None of the differences between replies for gestures vs. buttons were statistically significant. To gain more insight into user preference for buttons or gestures, the participants were asked to judge directly between the two approaches for EPG browsing. The participants were asked, "Which of the systems did you…": "find best overall" (Best), "find easier to learn to use" (Learn), "find easier to use" (Easier), "prefer" (Prefer), and "find more effective for the tasks you performed" (Effective). These questions were also used for other direct comparisons where we compare different aspects of the systems. The users had a preference for buttons over gestures for browsing the EPG. In particular, the participants found buttons easier to both use and learn to use. When asked about their preference in more detail, many participants stated that they were familiar with using buttons, at the same time many of these participants complained that using buttons was boring.

Use of Screens

The participants had a slight preference for the 2D visualisations. When asked about this, the users stated that they found the 2D visualisation more familiar than the 3D, but at the same time they complained that they found 2D boring and the 3D 'exciting' and 'cool'. For the image browsing task two different visualisations were used on the phone screen, one showing a 2D visualisation of the 3D visualisation on the TV and the other showing the image at the top of the queue on the phone screen (see Fig. 1D). In terms of user preference the users had a preference for having the image at the top of the current queue of images on the phone instead of the visualisation of the data stream. This result is encouraging as it could have been the case that users found the different views of the same information confusing, instead many users found it beneficial. Some commented that it was useful to use the phone to validate the image at the top of the visualisation on the screen, indeed many users noted that when the entire visualisation was on the phone that many of the images were too small. It should be noted that some users stated that they did not use the phone screen at all, just looking at the TV screen; many of these users are responsible for the no difference responses. The participants were also asked about the three different visualisations used in the 3D interface. Most users had a preference for the linear or the spiral layouts. For the spiral layout, participants stated that they liked to see more images but that it was disorienting at times. For the linear layout participants stated that they liked its simplicity but that they were able to see fewer images than with the other visualisations. With respect to the use of pressure to kink the linear layout, many users found it disorienting and said they would have liked to have been able to switch focus to other parts of the visualisation than the front, this is similar feedback to that for the spiral.

Interaction with Back of Device

As the back of the device was used in both the image and EPG browsing it was possible to compare using the back of the device for both types of task. It was found that, in general, the participants were more positive about using the back of the device for EPG browsing. This is an encouraging result, as for the EPG there were two sensors on the back of the device instead of one for the image browsing. It appears that as they become more familiar with the application that the users become more positive despite the increase in complexity. Perhaps with more training users will become more familiar with this type of interaction. It should also be noted that for some users the ease with which they could interact with the back of the device was affected by the way in which they gripped or held the mobile device. Some users adjusted their grip as the tasks proceeded, this may have affected the more positive perception for the EPG browsing.

Use of Pressure

In contrast with the relative increase in positivity in feedback of users towards using the back of the device, users are generally slightly less positive about using pressure when browsing the EPG than the image collection. Again this could be because the interaction is more complex, also coupled with this, for the image browsing it was not necessary to use pressure input whereas to complete the EPG browsing tasks quickly it was essential. It should be noted that while the positivity decreases, users were still generally positive or neutral about the use of pressure. In a direct comparison between the uses of pressure input versus the use of buttons for skipping forward and back 24 hours in the EPG the feedback was split. Again when asked, users stated that they found buttons more familiar and this was the reason that some users had a preference for using buttons.

7 Conclusions and Discussion

We have presented some novel prototypes for navigating media on a 3D TV-like display using a mobile device. The use of a mobile device for browsing content on a television has a number of benefits, including additional displays, gestural input and the possibility of additional input and output capabilities. Our user evaluation demonstrated that users were able to navigate through media using gestural and pressure input easily and successfully. There was very little difference in terms of performance between our prototypes and more traditional interaction methodologies, although the benefits outlined above are not available through traditional interaction methodologies. There are a number of design guidelines that can be made based on our findings:

- Users can adapt quite quickly to using positions on the back of the device to interact. However, care must be taken with the positioning so that users can grip and use the remainder of the device easily.
- Despite concerns that users would not be able to use the screen when tilting the device to gesture, users were able to use the screen if they wished. Thus, the screen on a mobile device can be used as an additional display. However, more research is needed on the impact of divided attention between two displays for

media browsing, as navigation time was impacted slightly by using the screen to display additional information.

- In general, users found gestures useful for navigating large distances quickly; however, they found fine control more difficult resulting in over- and under-shooting selections. Considerations for fine control should be taken into account.

As can be seen, a number of interesting problems and future research questions have been highlighted. The work presented in this paper is an important step to allowing more dynamic, social and natural control and navigation of multimedia which could possibly be deployed in home and public settings.

Acknowledgments. This research has been funded by the Industrial Members of MobileVCE (www.mobilevce.com), with additional financial support from EPSRC grant EP/G063427/1.

References

1. Roduner, C., Langheinrich, M., Floerkemeier, C., Schwarzentrub, B.: Operating Appliances with Mobile Phones - Strengths and Limits of a Universal Interaction Device. In: Proceedings of the International Conference on Pervasive Computing (2007)
2. Ronkainen, S., Hakkila, J., Kaleva, S., Colley, A., Linjama, J.: Tap input as an embedded interaction method for mobile devices. In: Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI 2007 (2007)
3. Eslambolchilar, P., Williamson, J., Murray-Smith, R.: Multimodal feedback for tilt controlled speed dependent automatic zooming. In: Proceedings of ACM UIST (2004)
4. Ramos, G., Boulos, M., Balakrishnan, R.: Pressure Widgets. In: Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2004 (2004)
5. Wilson, G., Stewart, C., Brewster, S.: Pressure-based menu selection for mobile devices. In: Proceedings of MobileHCI (2010)
6. Brewster, S., Hughes, M.: Pressure-Based Text Entry for Mobile Devices. In: Proceedings of MobileHCI (2009)
7. Clarkson, E.C., Patel, S.N., Jeffrey, S.P., Abowd, G.D.: Exploring Continuous Pressure Input for Mobile Phones. In: Proceedings of ACM UIST (2005)
8. Cesar, P., Bulterman, D.C.A., Jansen, A.J.: Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content. In: Tscheligi, M., Obrist, M., Lugmayr, A. (eds.) EuroITV 2008. LNCS, vol. 5066, pp. 168–177. Springer, Heidelberg (2008)
9. Dachselt, R., Buchholz, R.: Natural Throw and Tilt Interaction between Mobile Phones and Distant Displays. In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, CHI 2009 (2009)
10. Baudisch, P., Chu, G.: Back-of-device interaction allows creating very small touch devices. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009 (2009)
11. Hannah, D., Halvey, M., Wilson, G., Brewster, S.: Using Multimodal Interactions for 3D TV and Multimedia Browsing. In: Proceedings of EuroITV (2011)
12. Grubinger, M., Clough, P.: On the creation of query topics for ImageCLEFPhoto. In: Proceedings of the 3rd Workshop on Image and Video Retrieval Evaluation (2007)