Infinite Content Within Reach: Using 3D Body-Based Interaction to Enhance Content Discovery. 3DCHI, pp. 43-52, 2012.

Infinite Content Within Reach: Using 3D Body-Based Interaction to Enhance Content Discovery

Nili Metuki
PrimeSense Ltd
28 Habarzel St.
Tel-Aviv, Israel 40500
Nili.metuki@primesense.com

Mark Bolas
Fakespace Labs, Inc.
1220 Pear Ave.
Mountain View, CA 94043
bolas@well.com

Orlee Tal
PrimeSense Ltd
28 Habarzel St.
Tel-Aviv, Israel 40500
Orlee.Tal@primesense.com

Ran Shani
PrimeSense Ltd
28 Habarzel St.
Tel-Aviv, Israel 40500
Ran.Shani@primesense.com

Author Keywords
Gestures; Embodiment; Zoomable User Interfaces; ZUI; Depth Sensing; Information Space; 3D.

ACM Classification Keywords
H5.2 [Information Interfaces and Presentation]: User Interfaces; J.4 [Social and Behavioral Sciences]: Psychology.

General Terms
Design, Human Factors, Theory, Experimentation.

Abstract
With infinite content available, remote controls no longer provide satisfying content discovery on TV. Using depth-sensing technology, we have created an alternative interface for browsing media in the living room. We utilized continuous back and forward movements on the Z axis to create a three dimensional Information Space, in which the content is controlled by grabbing media items and zooming in to both find and select them. By employing a zooming embodied metaphor and a tree-like 3D information structure, the Information Space couples content with movement-based content-related actions. Since both content and action are arranged in three dimensions, users can easily create mental models of the information and the afforded interaction, and enjoy the direct interaction with their media. Users' responses to high-fidelity prototypes highlighted some of the unique challenges of designing 3D interaction, revealing gaps between theory and practice when it comes to designing motion-based interaction. We addressed those issues by considering the shape of the Information Space itself as a design variable and carefully constructing the afforded movements in the Information Space to match them to users' expectations.

Infinite Content Within Reach: Using 3D Body-Based Interaction to Enhance Content Discovery

Nili Metuki
PrimeSense Ltd
28 Habarzel St.
Tel-Aviv, Israel 40500
Nili.metuki@primesense.com

Orlee Tal
PrimeSense Ltd
28 Habarzel St.
Tel-Aviv, Israel 40500
Orlee.Tal@primesense.com

Ran Shani
PrimeSense Ltd
28 Habarzel St.
Tel-Aviv, Israel 40500
Ran.Shani@primesense.com

Mark Bolas
Fakespace Labs, Inc.
1220 Pear Ave.
Mountain View, CA 94043
bolas@well.com

Author Keywords
Gestures; Embodiment; Zoomable User Interfaces; ZUI; Depth Sensing; Information Space; 3D.

ACM Classification Keywords
H5.2 [Information Interfaces and Presentation]: User Interfaces; J.4 [Social and Behavioral Sciences]: Psychology.

General Terms
Design, Human Factors, Theory, Experimentation.

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
With infinite content available, remote controls no longer provide satisfying content discovery on TV. Using depth-sensing technology, we have created an alternative interface for browsing media in the living room. We utilized continuous back and forward movements on the Z axis to create a three dimensional Information Space, in which the content is controlled by grabbing media items and zooming in to both find and select them. By employing a zooming embodied metaphor and a tree-like 3D information structure, the Information Space couples content with movement-based content-related actions. Since both content and action are arranged in three dimensions, users can easily create mental models of the information and the afforded interaction, and enjoy the direct interaction with their media. Users' responses to high-fidelity prototypes highlighted some of the unique challenges of designing 3D interaction, revealing gaps between theory and practice when it comes to designing motion-based interaction. We addressed those issues by considering the shape of the Information Space itself as a design variable and carefully constructing the afforded movements in the Information Space to match them to users' expectations.

Introduction
For many years, viewers have relied on the remote control to navigate their TV content. Now, in this era of seemingly infinite content, finding and choosing content using the remote control has become an unmanageable chore. At any given moment, users may choose from hundreds of programs aired on television channels, along with the tens of thousands of titles available on-line. Media centers and smart TV interfaces that utilize the typical remote control are no longer a good solution for coping with this information overload and maximizing user satisfaction.

Media centers with a mouse and keyboard are better than a simple remote at discovering content, but have been poorly received by the marketplace. Transporting a desktop interface to the living room seems to create an awkward environment that detracts, rather than supports, an accurate and enjoyable content discovery experience.

New depth-sensing technologies provide the means to leverage body-based interaction that can intuitively handle the information flood. Previous work suggests that embodied interaction methods, especially those with metaphorical meaning, can lead to the creation of rich and intuitive interaction [e.g., 1, 16]. In the context of infinite content, body-based interaction may open new doors by affording users the opportunity to interact with their content in ways that resemble interaction with objects and people in real life, instead of typing or clicking an input device [11].

While body-tracking technologies have been used for high-end research projects, the availability of a depth camera that has sold over 10 million units in a single year provides a rare occasion to rethink user interface capabilities for markets and applications that have been practically unapproachable. The Microsoft 'Kinect' was launched in late 2010 as a game controller, and utilizing PrimeSense depth sensing technology.

We believe that depth sensing can be used to transform the user experience in the living room to create a system that makes discovering media choices as enjoyable as consuming them. We propose that this can be accomplished through the creation of a carefully constructed 3D Information Space intelligently populated with content, coupled with a set of embodied spatial gestures that naturally navigate, select, and control that content.

Designing for a new input device

Using body-based interaction via depth-sensing technology must be considered as distinct from existing input devices. Gestures performed in the air are coarse and varied. They provide no clear method of indicating an action as simple as a button click, and do not entail clear tactile feedback. Game-based UI's for the Kinect appear to be designed from scratch, so that these challenges were minimized. Kinect games avoid the need for rapid binary selection typical of hand-held controllers, while highlighting whole body motions. For example, in a full body dance game titled 'Dance Central,' when selections are required, they are accomplished by swiping the hand through a 2D menu item [5]. In other cases, such as the 'Kinect sports' game user interface, the hand is used as a cursor to navigate, while selection is triggered by hovering steadily over an item for a predefined time period [9]. Likewise, when aiming to design a media center, any ideas that rely on mouse-based interface concepts would quickly be thwarted. Our ability to achieve spatial accuracy using our hands as cursor, without inducing immediate fatigue, does not begin to match the XY resolution or the response time of a mouse moved on a desk.

Having said that, tracking hand movement using a depth sensor has an obvious strength: Z. We were able to easily map the hand movements to three dimensions, thus creating a continuous three dimensional Information Space. The Information Space contained both the content and the afforded actions (i.e., navigation and selection) as one – all controlled by abstract mapping of the hand movements to the space, as described in the following sections.

Zooming through the Information Space

Including analog movement back and forth on the Z axis created a new building block for our paradigm, enabling us to assign meaning to movements in a dimension that was thus far unavailable to traditional input devices.

Building upon embodied conventions of interaction with objects [1, 11], we discovered that movements on the Z axis map naturally to zooming in and out on information in the space. Bringing the hand toward the body (flexion movement) is naturally associated with bringing an object closer to the user, such that the object gradually takes more and more space of the visual field and more details are visible.

Applying this natural movement to the TV interface, we assigned an animated analog zoom into the screen content in response to the user's movement of their hands toward the body, and a zoom out when the hands move away.

The decision to map the Z motion of the hand to create a zooming effect takes advantage of the recognized utility of Zoomable User Interfaces. ZUI's have been proposed for displaying large amounts of content in a way that that "taps into our natural spatial and geographic ways of thinking" [12]. They do so by providing "simple methods for visually navigating complex information spaces that ease the burden of locating information while maintaining an intuitive sense of location and of relationships between information objects" [4]. As reviewed by Bederson [3], ZUI's allow users to create mental maps of information, using rich but simple visualization.

Designing the Information Space

To allow users to access a large amount of data easily, we employed a hierarchical 3D data organization approach. Bederson noted that zoomable interfaces use hierarchical data organization [3] based mostly on 2D hierarchical trees [for example, 6, 8]. Studies on information visualization have demonstrated that hierarchical 3D trees allow display of more data nodes, in more branches, and across more levels [14]. Robertson and his colleagues demonstrated the advantages of their approach using the Cone Trees prototype, which was based on 2D input devices, similarly to the zoomable interfaces mentioned above. In a final prototype of our approach, we created a zoomable 3D hierarchical tree using media type and genre or category as node points, and thus used the third dimension to allow users to easily dive in and out of large collection of media titles or TV channels.

By creating a zoomable 3D tree, we combined a 3D hierarchical logical design, and 2D visual design, without having to visualize the entire logical structure at once, as done in the classical Cone Trees approach. Robertson and his colleagues demonstrated that animated feedback for manipulating the 3D hierarchies enables users to perceive large amount of information with less cognitive load, by providing depth cues that facilitate the visual perception of the data structures [14]. We were able to achieve similar effects, utilizing depth cues that are implicitly provided by motor feedback of movement direction, and complemented with animated transitions for zooming through the Information Space.

Hence, we designed to potentially decrease the users' mental load and allow them to easily create a mental model of the 3D tree without subjecting the model to the limitations of display – enabling the display of an infinitely and wide deep tree, and thus of infinite content.

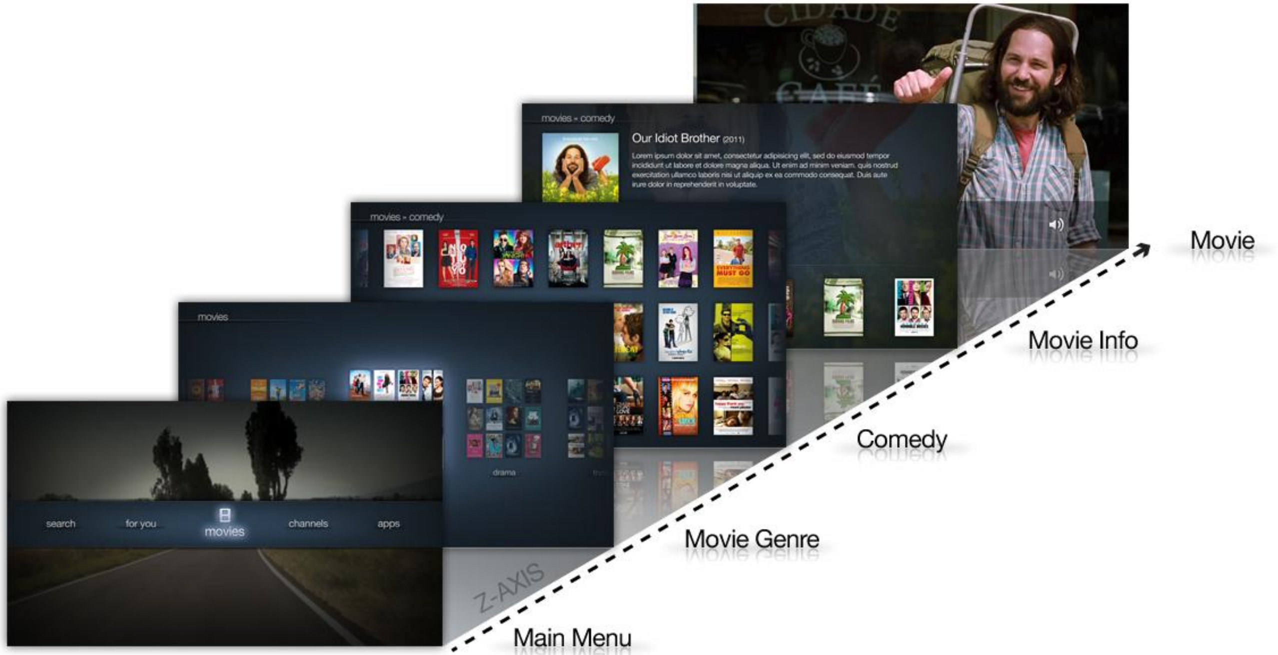

Figure 1. Zoomable hierarchical 3D tree in a high-fidelity prototype of a media center.

Integrating movement and content in the Information Space

Previous works utilized ZUI's to allow navigation in large collection of media, e.g., PhotoMesa, a zoomable image browser for home users [2]. But past ZUI designs, as well as the hierarchical 3D Cone Trees prototype, use mouse clicks to navigate and select content, resulting in fixed motions, like a metaphoric magnifying glass.

Depth-sensing technology creates an opportunity to control the zoom axis with as much fidelity as is available in the panning axes. By incorporating the natural gestures of pulling to zoom in and pushing to zoom out, we put the act of zooming in the user's hand. Navigating media, from the user perspective, is done by virtually bringing the data closer – such that moving the hand equals moving the content. Note that if the user pulls close a single item so that no others remain in view, then the content is selected. Therefore, the Information Space is fusing content and movement-based control, and the natural mapping of movement to content appears to induce a sensation of direct physical agency on the Information Space.

The fusion of movement and control enables the user to make choices through motion. By arranging the content in a tree-like structure, selections are made by reaching out to the desired content, virtually grabbing by closing the hand, and navigating through branches in the tree by pulling. Because motion along Z is continuous in both directions, choices are fluid – which is perfect for a media browser. For example, the user can pull to look at a particular genre of media and, once seeing the titles in the genre, simply reverse hand direction to push away and slide the hand over a bit to a different genre, pulling again to explore the new titles. All without a single click.

As opposed to using a mouse wheel to zoom into data, the physical act of selecting an item by pulling it toward the user in a flexion arm movement implicitly carries emotional connotations. Research of embodiment provides rich evidence for the association between movement and emotions [13]. In particular, flexion movements of the arm have been repeatedly associated with the approach motivational system and positive emotions, and extension of the arm (used in our prototype for zooming out) has been associated with avoidance motivation and negative emotions [e.g., 10, 13]. Therefore, virtually grabbing an item and pulling it in to select it is likely to evoke positive feelings and even induce approach motivation [7].

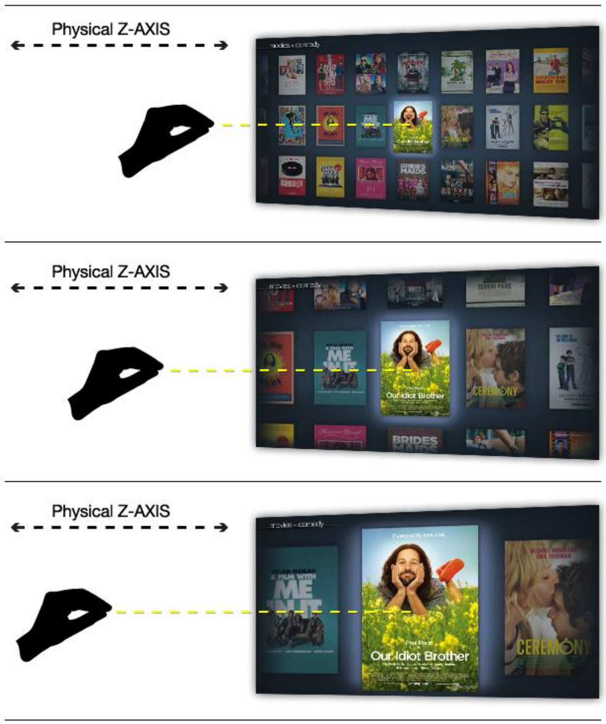

Figure 2. Fused movement of hand and content in the Information Space

From theory to practice
Throughout the development process, we have observed numerous users exploring high-fidelity prototypes of media centers based on our approach. Over 50 adults participated one-on-one user testing sessions in a living-room like laboratory settings, and hundreds were observed using the prototype in open exhibitions. During these observations, we have encountered interesting issues concerning the 3D interaction, stemming from the use of continuous flexion and extension movements on the Z axis on top of the traditional interaction in XY. When considering the use of three dimensions for user interfaces, we found it highly useful to consider non-linear and asymmetric mappings between a user’s motions and the content, thus considering the shape of the Information Space itself as a design variable, and constructing the afforded movement to fit users’ perceptions and expectations of moving content.

Constrained construction of space

One challenge we encountered was the freedom paradox: While users enjoyed the fluent content browsing afforded by continuous movements in all dimensions, in practice they encountered difficulty in planning the movement in the lack of constraints. Thus, users would report uncomfortable positions and many unintentional transitions between the zoom levels, resulting in frustration and reports of perceived lack of control and physical discomfort. In virtual reality studies, researchers concluded that constraining the degrees of freedom assists the user [15]. This was successfully applied in the Information Space, improving usability even if restricting some of the fluency of content discovery.

For instance, we constrained long continuous Z movements by segmenting the afforded motions according to the tree hierarchy. Under this constraint, users could only complete a transition to the next level by opening their hand. Creating the segmentation led to using smaller motions to control the application and reducing the load of planning movement length. In addition, at any given moment, movement was constrained to a single plane: XY (to navigate within a level) or Z (to move between levels). This was done by (1) providing different methods to navigate within a level (moving the hand on XY while open) than between levels (moving the hand on Z while grabbing), and (2) discarding below-threshold movements in one plane (XY or Z) while significant movement was detected in the other. Creating a distinction between movements in the different planes significantly reduced the load of planning 3D movements. Introducing these constraints led to increased usability and satisfaction, giving users the feeling that they operate the prototype freely.

Asymmetric construction of space

Another interesting paradox we encountered when testing our prototypes with users was the asymmetry paradox. While technically, moving one level down (zooming in) or up (zooming out) in the Information Space required the exact same movement length and could be performed in a perfect symmetry, users perceived and performed these unidirectional movements in different manners. Analyzing users' behavior and reports, we recognized that underlying this mismatch between required and actual behavior is an interesting interaction between users’ mental model, motivations, and ergonomics stemming from the inherent integration between movement and content in the Information Space. When zooming in, the synergy between the mental model (zooming in is in fact item selection), and the implicit positive emotional and motivational connotations of the inward movement, leads to higher engagement of the users directed at the selected item. Thus, users are more motivated to be accurate, which presents ergonomically as fine-tuned movements and steep learning curves. However, when zooming out, the mental model of getting rid of the content as well as the extension movement seem to reduce the motivation to perform accurate motions.

Ergonomically, this is typically presented as coarse movements with longer extension of the arm, demanding more effort than really necessary. Moreover, in their subjective reports users often complain that pushing to move one level up demands longer movement that desired. Hence, to promote symmetrical small and accurate movements when zooming in and out, we had to induce asymmetry in the way we detected movement in both directions, and alter the on-screen feedback to facilitate the inherent differences. This new asymmetry fostered equal reports of learnability, satisfaction, and effort for both directions.

Conclusions

Our prototype provides a proof of concept to using body-based interaction in three dimensions to enable efficient and enjoyable content discovery in the era of infinite content, in living room settings. This was done by fusing the content and its control affordances in a zoomable, tree-like, constructed three dimensional Information Space, allowing users to manipulate content by moving their hand in analog Z movements and bringing the content closer.

Using a depth sensor as input device has great promise for alternative interactions utilizing the additional depth dimension. Designing with this technology is challenging, requiring thorough understanding of technical affordance and limitations, as well as ergonomic, emotional, and cognitive aspects of interaction, combined with sheer exploration of visualization and feedbacks to support all those aspects. Due to the inherent fusion among all the outlined factors, designing the 3D Information Space can only be done in a process of formative evaluation. In our work, we have constructed the afforded movement in the Information Space based on users' feedback, to match the way users perceive the content and the afforded control. Our experience shows that exploring this new territory in HCI requires iterative processes exposing users to high-fidelity prototypes, because many challenges are revealed only when concepts meet the users. While many issues in this field are yet to be explored, and many challenges are ahead, we see great potential in introducing the third dimension into new interactions in the living room and beyond.

Acknowledgements

The described UX approach is the fruit of the work of many collaborators. We would like to thank Martin Frey, Marcus Hauer and the entire project team at IDEO for conceiving this concept and guiding its first steps, and to our collaborators at PrimeSense UX team (Adi Berenson, Amir Hoffnung, Arnon Yaari, Asaf Kiselnik, Assaf Segal, Dana Cohen, Daniel Shine, Eran Weissenstern, Ido Inbar, Jonathan Pokrass, Len Zhong, Micha Galor and Ron Slossberg) for giving it life.

References

[1] Beckhaus, S. and Kleesiek, J. Intuitive Interaction: Tapping into body skills to find rich and intuitive interaction methods for Virtual Reality. In *Proc. CHI 2011 Workshop 'Embodied Interaction: Theory and Practice in HCI'*, ACM Press (2011), 128-131.

[2] Bederson, B.B. PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. In *Proc.UIST 2001,* ACM Press (2001), 71–80.

[3] Bederson, B.B. The promise of zoomable user interfaces, *Behaviour & Information Technology 30*, 6 (2011), 853-866.

[4] Bederson, B.B. and Hollan, J.D. Pad++: A Zoomable Graphical Interface System. In *Proc. CHI 1995*, ACM Press (1995), 23-24.

[5] Dance Central, Harmonix, 2011. http://www.dancecentral.com/

[6] Grosjean, J., Plaisant, C., and Bederson, B.B. 2002. Space-Tree: design evolution of a node link tree browser. In *Proc. InfoVis,* IEEE (2002), 57–64.

[7] Haeffel, G.J. Motion as Motivation: Using Repetitive Flexion Movements to Stimulate the Approach System. In *Behavior Therapy 42* (2011), 667–675.

[8] Hightower, R.R., Ring, L.T., Helfman, J.I., Bederson, B.B., and Hollan, J.D. Graphical Multiscale Web Histories: A Study of PadPrints. In *Proceedings of ACM Conference on Hypertext 1998*, ACM Press (1998), 58-65.

[9] Kinect Sports, Rare, 2010. http://marketplace.xbox.com/en-GB/Product/Kinect-Sports/66acd000-77fe-1000-9115-d8024d5308c9

[10] Maxwell, J. S., and Davidson, R. J. Emotion as motion: Asymmetries in approach and avoidant actions. In *Psychological Science, 18* (2007), 1113–1118.

[11] Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Treacy, E. S., and Zigelbaum, J. Reality-Based Interaction: A Framework for PostWIMP Interfaces. In *Proc. CHI 2008*, ACM Press (2008), 201-210.

[12] Perlin, K. and Fox, D. Pad: an alternative approach to the computer interface. In *Proc. SIGGRAPH '93*, ACM Press (1993), 57–64.

[13] Price, T.F., Peterson, C.K and Harmon-Jones, E. The emotive neuroscience of embodiment. In *Motivation and Emotion* (in press).

[14] Robertson, G. G., Mackinlay, J. D. and Card, S. K. Cone trees: Animated 3D visualizations of hierarchical information. In *Proc. CHI '91*, ACM Press (1991), 189–194.

[15] Stuerzlinger, W. and Wingrave, C.A. The Value of Constraints for 3D User Interfaces. In *Virtual Realities*, S. Coquillart, G. Brunnett and G. Welch (Eds.), SpringerWienNewYork, Germany, 2008, 225-238.

[16] Tucker, D. Toward a theory of gesture design. M.F.A. dissertation, University of Southern California, United States – California.