CRISTAL: A Collaborative Home Media and Device Controller Based on a Multi-touch Display

Thomas Seifried[1], Michael Haller[1], Stacey D. Scott[2], Florian Perteneder[1], Christian Rendl[1], Daisuke Sakamoto[3], Masahiko Inami[4]

[1]Media Interaction Lab, Upper Austria University of Applied Sciences, Austria

[2]Department of Systems Design Engineering, University of Waterloo, Canada

[3]Department of Computer Science, University of Tokyo, Japan

[4]Graduate School of Media Design, Keio University, Japan

mi-lab@fh-hagenberg.at

ABSTRACT

While most homes are inherently social places, existing devices designed to control consumer electronics typically only support single user interaction. Further, as the number of consumer electronics in modern homes increases, people are often forced to switch between many controllers to interact with these devices. To simplify interaction with these devices and to enable more collaborative forms of device control, we propose an integrated remote control system, called CRISTAL (Control of Remotely Interfaced Systems using Touch-based Actions in Living spaces). CRISTAL enables people to control a wide variety of digital devices from a centralized, interactive tabletop system that provides an intuitive, gesture-based interface that enables multiple users to control home media devices through a virtually augmented video image of the surrounding environment. A preliminary user study of the CRISTAL system is presented, along with a discussion of future research directions.

Author Keywords

Multi-Touch, Remote Controller, Collaborative Interface

ACM Classification Keywords

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems — [H.5.3]: Collaborative computing — Computer-supported cooperative work.

INTRODUCTION

Our living rooms are often places where people share stories, play games, and where we have fun together. People watch movies together with their friends and share and view pictures with their families. More and more, we are sharing digital information. While display of this digital content is well supported by different technologies such as TVs, digital picture frames or even digital tabletops, the social process of choosing this digital content as a group is not well supported by current controllers. Most devices come with their own controllers, and usually only one person can be in control of the device. For instance, the social interaction while choosing a movie using modern digital media controllers is quite limited; everyone can watch, but only one person can control the device at a time.

Additionally, the amount of digital appliances and media found in domestic environments has risen drastically over the last decade, for example, digital TVs, DVD and Blu-ray players, digital picture frames, digital gaming systems, and robotic vacuums. It can be very confusing to determine which controller belongs to which device.

As these devices become more compatible with computer networking and as wired and wireless networking infrastructures become more prevalent in our homes, new opportunities arise for developing more advanced control of these myriad devices. However, existing centralized (or universal) remote controls lack intuitive, multi-user interfaces for mapping control functions to the target device. They often require trial and error button pressing, or experimentation with graphical user interface (GUI) controls, before a (single) user achieves an intended action.

To address these issues, CRISTAL (Control of Remotely Interfaced Systems using Touch-based Actions in Living spaces) was developed (see Figure 1). CRISTAL is a first step toward providing users with an interface to collaboratively share media content and control electronic devices in the room. The system provides a novel experience for controlling devices in a home environment by enabling users to directly interact with these devices on a live video image of their living room using multi-touch gestures (see Figure 1).

To set the context for this research, we first review the related work. We then describe CRISTAL's system design and architecture. Next, we report a preliminary study that was conducted to gather user feedback on our current design solution. Finally, we discuss some limitations of the CRISTAL system and directions of future research.

Figure 1: The demonstration living room of CRISTAL.

RELATED WORK

Many tabletop systems are designed for usage in a living room. Examples are Microsoft's Surface™[1] or the Entertaible™ [4]. To date, these platforms have primarily been used for gaming and entertainment applications. Another tabletop system that was designed to be situated in living room is the Personal Digital Historian [10]. This work facilitates collaborative viewing of pictures on a circular tabletop. Mazalek et al. [8] evaluated one such home gaming tabletop system, called TViews, in a field study. Within their study, several games were tested in a user's home environment. Users found the TViews system engaging and fun to use, especially during shared activities. Even though many tabletop concepts and applications are tailored for home-use, few applications have been developed for home automation and media controlling.

Beijar et al. [1] developed a touch-sensitive interactive tabletop system with a built-in media center. This system, called Remotable[2], enabled users to control common electronic devices on a table surface. A key difference between the Remotable system and the CRISTAL system, described in this paper, is the tabletop user interface provided for controlling the home media devices in each system. Remotable provides a minimalistic interface consisting of simple light emitting diodes (LEDs) arranged in patterns to indicate, for example, a volume scale or a simple navigation menu. While this interface is quite aesthetically pleasing, it provides little information about which device is being controlled, or what actions are possible in the room. In contrast, the design of CRISTAL attempts to provide more intuitive visual feedback of what devices can be controlled, and what control actions are possible on those devices, while at the same time providing natural, gesture-based user interaction. Furthermore collaborative use was not addressed at all in Remotable.

In contrast to using an integrated media center like the Remotable, it is more typically in today's home to see home automation and media controlling handled separately. This may be due to the use of quite different low-level communication methods in these two fields. For media control very extendable communications protocols are used that can handle large amount of data and streams. In home automation data transfer is limited but there are strong requirements for low latency transfer. Therefore only a few systems and interfaces exist that can handle both home automation and media control. AMX[3] and Crestron[4] provide touch panels for controlling all types of devices in an entire house. But their GUIs mainly consist of buttons and only a few hints are provided - besides plain text - of which device is actually controlled. Stardarw[5] provides a more customizable GUI for touch screens, but it does not offer a live video image for controlling the devices. Similarly, Yoon et al. use a marker-based setup combined with a touch panel [15].

The concept of controlling devices through a smaller image was first presented in the world-in-a-miniature interface [11]. The idea of controlling devices through a video image was introduced by Tani et al. [12] in the project Hyperplant. They explored different ways of controlling devices in a factory through a video image. They focused on ways to define interactive areas in the video image in 2D and 3D. Liao et al. [6] investigated using a video image to control an interactive shared space during remote conferencing. Their system enabled users to annotate and move presentation slides from one screen to another in the same conferencing room by dragging and dropping in a video of the meeting room. To simplify such interfaces a pen-based approach was used in "Sketch and Run". Sakamoto et al. [9] proposed a video-based Tablet-PC interface to control vacuum cleaning robots.

CRISTAL

To bring the approach of interaction through a video image to the living room a more ubiquitous interface is necessary. Desktop computers are seldom used in living rooms and they are not suitable for co-located social interaction, as they are designed to be a personal device. Mobile devices such as handheld touch screens or mobile phones seem to be a proper interface for controlling other electronic devices. Nevertheless, they are also designed to be used by a single person and do not support social interaction that well. It was our goal to create an interface that allows control of all electronic devices in one room and in the same time encourages social interaction with friends and family. Therefore we chose to explore the use of a video-based media control interface on an interactive table surface.

---

[1] http://www.surface.com

[2] http://www.remotable.se/

[3] http://www.amx.com/

[4] http://www.crestron.com

[5] http://www.stardraw.com/

A coffee table is often located on a central position in the living room. Positioned next to a couch it is always in reach of people sitting around it. CRISTAL facilitates the table and augments it with functionality to control electronic devices in the living room. CRISTAL provides an interface for both home automation and multimedia control. But in contrast to many other home automation systems the system allows more complex operations than just turning devices *on* or *off*. The movements of the vacuum cleaner are controlled by a simple path-gesture done on the table surface. Furthermore, users can choose movies they want to watch directly on the table and display them on the TV or enlarge photos on the digital picture frame. In addition, multiple users can control all devices, using one single display. In detail, the user can control the following objects in the living room:

- **Light sources:** turn on/off and dim light sources, and set a global lighting color (i.e. set a warm/cold light and all light sources are adjusted accordingly).
- **Audio:** control of volume.
- **TV/Projector/Music Player:** choose movies or music and control CD and movie playback.
- **Digital picture frame:** select a physical photo album in the video image and drag it onto the picture frame.
- **Robotic vacuum cleaner:** control the movement and position of a vacuum cleaner robot [1].

Video image interface

In contrast to other UIs, *a live video image* is used as the primary interface. The video image displayed on the coffee table shows the entire living room and also every device that can be controlled. Therefore, the video image itself is the interface. Users can control a device by simply tapping on the device's image on the coffee table. The video image also provides instant feedback. If a light is turned off it is instantly visible for the user on the screen. Since the video image shows the entire living room, users are familiar with the content of the image and recognize instantly devices they want to control.

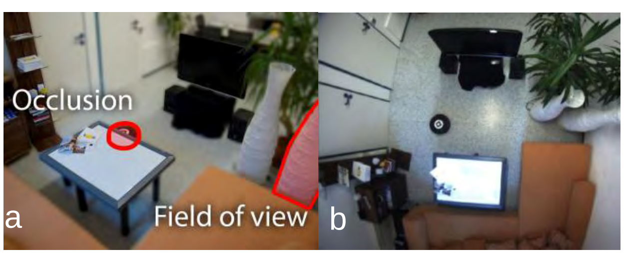

Figure 2: (a) Many occlusions occur in a perspective view. (b) A top down view with fewer occlusions.

In our testing scenario, we experimented with different camera locations (see Figure 2). The perspective view, captured from ceiling height at one corner of the room, typically introduced occlusion problems. Moreover, it was difficult to track the entire room, as some devices were located fully or partially outside of the camera's field of view (see Figure 2, a). In contrast, we found that the best tracking results were achieved using a top-down view from a ceiling-mounted camera (see Figure 2, b). This view is also useful for detecting the location of all devices. Furthermore, the top-down view does not imply any orientation and is therefore well suited for a tabletop display.

The orientation in which the camera image is displayed on the tabletop surface is an important factor for the users to orient themselves in the video image. If the camera captures the room from the ceiling, it is necessary to display the captured video in the same orientation as it is captured. From the users point of view the video image should present the room from the same orientation as perceived by the users.

Providing user feedback

CRISTAL provides two different types of interaction feedback to users: *direct video image feedback* and *augmented feedback widgets*.

Direct Video Image Feedback: Certain devices in the environments, such as light sources, provide obvious visual feedback directly in the video background image in reaction to a user's interaction in the table interface. For instance, after switching on/off a light source, users can immediately "see" the results in the video background image without any further visualization requirements.

Augmented Feedback Widget: When controlling other devices, it is necessary or desirable to obtain additional feedback visualized in the interface. Therefore, we also propose augmented feedback widgets that show additional information. For example, a slider widget for dimming the lights or controlling the audio volume, or more elaborate widgets for more complex device interactions, as described below for selecting a DVD movie to play on the TV.

Dimming the light sources

Real light sources can be dimmed by sliding over the captured light source on the table (Figure 3).

Figure 3: Dimming the light by sliding across the light source.

The advantage of using a real video input is now the automatically feedback, which is provided by the fact that users immediately *see* the status on the table. Different gestural interaction methods were implemented in CRISTAL for controlling the light devices in the environment, as described in more detail later.

Selecting a movies from the DVD rack

CRISTAL also includes a simple movie browser, from which users can freely navigate through the movies by sliding over the DVD rack. The virtual DVD rack appears when the user taps on the physical DVD shelf in the video image. The interface of the movie list is inspired by the visual appearance of Apple's *CoverFlow* control and resembles the flicking-movement approach proposed by Lucero et al. [7]. Sliding along the main axis browses through the list. A linear gesture along the minor axis initiates a drag and drop action of the current selected item (The calculation of main and minor axis is described in a later section of this paper). As choosing a movie is often a social task, CRISTAL allows multiple users to browse, move, and share virtual movie covers on the table.

Traditionally, when people choose a movie to watch, they often move the physical DVD packages around on the table, reading the description of the film, and placing movies they do not wish to watch off to one side. CRISTAL aims to support this natural interaction with virtual movie covers. Each movie cover can be independently dragged, rotated, scaled or flipped. On the back of the cover, users can get additional information about the movie, such as a short description, ratings, and genre. Alternatively, users can play a preview on the table. In order to play the movie, the user simply has to drag the movie directly onto the TV in the video image. The playback starts automatically. Tapping on the TV once pauses or restarts the movie. For seeking through the move users can use a sliding gesture similar to controlling a light. An additional menu provides further control such as jumping to certain chapters of the movie.

Controlling the vacuum cleaner

Users can also use CRISTAL to control their robotic vacuum cleaner (e.g., an iRobot Roomba).

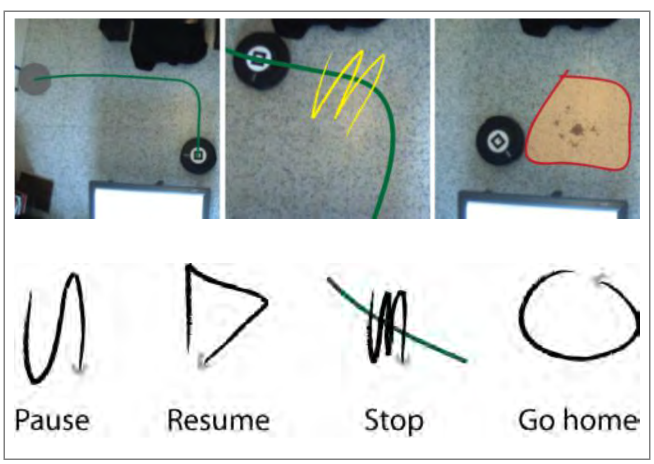

Figure 4: Different gestures used in CRISTAL.

There are three different ways to interact with the vacuum cleaner (see Figure 4). The simplest way to interact with the vacuum cleaner is by sketching a path starting at the origin of the Roomba.

Consequently, the robot follows the sketched path. The stroke can be a curve or a zigzag stroke. It must be drawn in a single stroke (cf. Figure 4, *bottom-left*). Furthermore, we use a minimal set of gestures (Figure 4, *top*) with a strong semantic correspondence to the associated operations (e.g. "stop" and "pause" functions).

In some cases moving the vacuum cleaner along a certain path does not fit the user's needs. Hence, a large lasso stroke gesture is used for defining a region of interest, where the Roomba has to clean the area (Figure 4, *top-right*). To visualize this region, we filled it with orange color.

SYSTEM HARDWARE AND SOFTWARE

Hardware implementation

As depicted in Figure 5, a single computer directly controls all devices and also provides the user interface on the table. Our multi-touch table is built on a 42" large DiamondTouch table 2, combined with a DLP projector mounted on the ceiling. To track the overall room, we mounted a Firewire camera to the ceiling of the room.

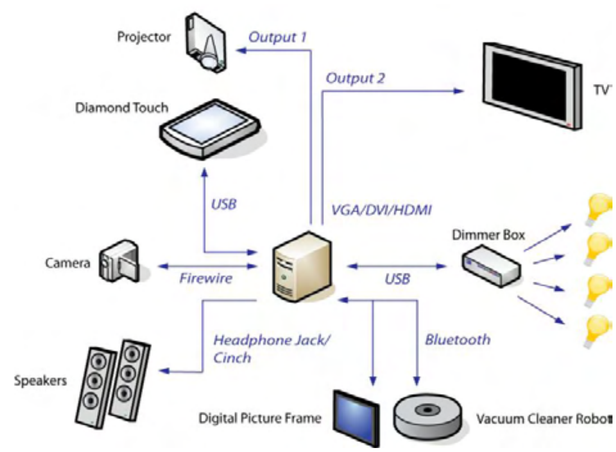

Figure 5: Prototype system.

A camera mounted next to the projector captures the live video image. The camera uses a wide-angle lens to capture the whole room and all devices that can be controlled. The camera itself is connected via Firewire to the computer, and finally, the camera image is shown in full-screen on the DiamondTouch coffee table. For the prototype setup the TV was directly connected to the computer and movies are played from the computer itself. The lights are connected to a DMX controlled dimmer box, which is connected via USB to the main computer. Finally, the iRobot Roomba was used as the vacuum cleaner and communication was established over a Bluetooth connection.

Vacuum cleaner robot

We used the iRobot Roomba model 580 which has two degrees of freedom in the base for driving with a speed of 500 mm per second at maximum. The Roomba Open Interface (ROI) also allows developers to develop customized PC-programs and wireless transmitters that communicate over the Bluetooth Serial Port Profile (SPP). In CRISTAL, we use the fiducial-marker tracking library ARToolkit [5] to detect the locations of the iRobot Roomba. The camera, mounted on the ceiling, captures the ARToolkit markers that are mounted on the vacuum cleaner. The gesture recognition engine used for controlling the robot is based on the $1-engine [14] due its insensitivity to orientation and simple configurability. Additionally to the functionality of $1 a pre-processing step was implemented to enhance the reliability of the gestures recognition. This step included the detection of drawn paths, lasso selections and stroke-out gestures that are not supported by $1.

Axis of interaction

As mentioned earlier, users can interact with a device by sliding along an interaction axis. We implemented two different interaction axis methods and compared them in a user study, described below in the "Evaluation" section. In the first approach (Figure 6, *a*), users need to slide vertically, relative to the display's coordinate system. In the second scenario, the interaction axis is defined by the device's orientation in the video image (Figure 6, *b*).

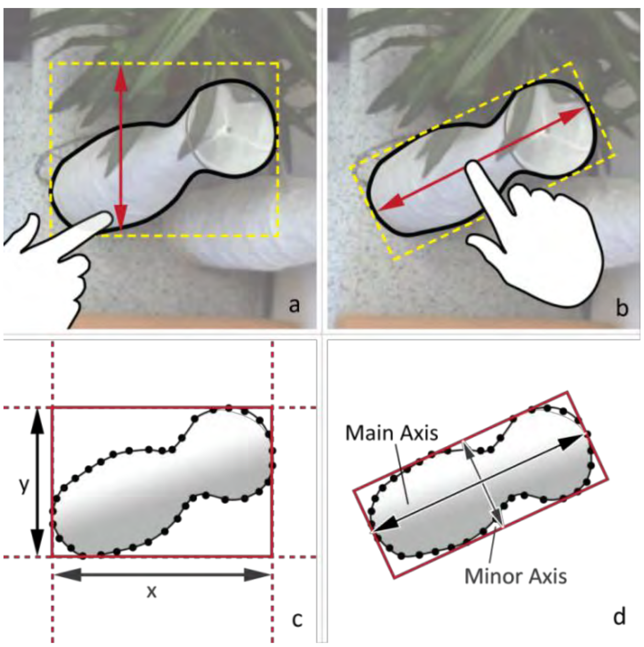

Figure 6: The direction of interaction can be along the main axis relative to the systems main coordinate system *(a, c)* or along the main axis of the bounding shape *(b, d)*.

In the display related approach we calculate an axis-aligned bounding box (AABB) and use the vertical axis (y) as an interaction vector (see Figure 6, *c*). Therefore the direction for sliding gestures is the same for every device.

In the second approach, users can slide along the main axis of the bounding box, which will be oriented according to the object's outline in the video image (Figure 6, *d*). Similar to calculating object-oriented bounding boxes (OBB) 3, we calculate the 2×2 covariance matrix *C*, with

$$C = \begin{bmatrix} \mathrm{var}_x & \mathrm{cov}_{xy} \\ \mathrm{cov}_{xy} & \mathrm{var}_y \end{bmatrix},$$

whereby we define the matrix elements with the following variables:

$$\mathrm{var}_x = \frac{1}{n-1}\sum_{i=0}^{n}(x_i - \bar{x})^2,$$

$$\mathrm{var}_y = \frac{1}{n-1}\sum_{i=0}^{n}(y_i - \bar{y})^2,$$

and

$$\mathrm{cov}_{xy} = \frac{1}{n}\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

n is the number of vertices, $\bar{x}$ and $\bar{y}$ are the mean of *x* and *y* respectively. The eigenvectors of this symmetric matrix *C* are the principal axis of the inertia. These vectors are mutually orthogonal and are normalized. After normalizing the eigenvectors, we find the extreme vertices along each axis. One of the two eigenvectors of the covariance matrix is the axis of maximum variance. This is the axis we use for defining the users' interaction direction.

Direction of Interaction

The axis of interaction itself is not enough information to control a numerical value. It is not clear how to determine that a sliding gesture was meant to decrease or increase a devices property.

In general, the direction can be defined in following ways:

- A **pre-defined direction** can be used for all devices on the whole screen.
- The **location of the individual users can be used**. In this case, a gesture towards the user could be used as decreasing the value and away from the user as increasing.
- Due to the fact that **real objects** are shown in the video image, users immediately "know" where the top and where the bottom of the object is. Therefore, the top of the object can be used as the maximum and the bottom as the minimum value.

In combination with the video image the third solution seems to be the most appropriate. However, this solution raises problems for automatically mapping user gestures to device controls, as it may be a challenge for the system to recognize the top and bottom of an object. But the complexity of this problem is reduced significantly if we consider two constraints of our system: Firstly, the camera is mounted on the ceiling and is pointing straight downwards and secondly all controlled devices are upright. This means the top and the bottom of the device are on the same vertical axis. In this case the bottom part of the device is always the area of the outline that is closer to the center of the camera image. Hence, as depicted in Figure 7, a gesture moving towards the camera center is a movement towards the bottom of the object. Consequently the value of a devices property is reduced.

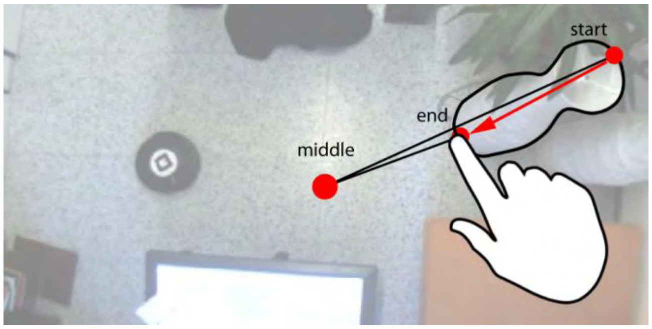

Figure 7: A gesture towards the center of the camera image results in decreasing a devices value (e.g., Light intensity).

Handling extreme lighting conditions

The exposure setting of camera, used for capturing the room, was configured manually to ensure a good visibility of all devices. But during the day the ambient lighting conditions may vary to a great extent and expose the camera with too much or too less light to capture a good image. In this case, some objects are indistinguishable from the background (e.g. at night when all lights are switched off the image is black). Therefore, we also visualized the outlines of all controllable devices using semi-transparent colors of medium brightness. Under normal lighting conditions the outlines are nearly invisible, but can be easily seen if the image is over- or underexposed.

SYSTEM EVALUATION

In order to begin evaluating how well the CRISTAL design addresses our two main design goals,

1. to enable simple, intuitive control of different types of home media and digital devices, and
2. to support collaborative interaction with these media and devices,

we conducted a preliminary laboratory-based user study. The study aimed to gather user feedback to elucidate which aspects of our design should be improved and which aspects were appropriate and appreciated by users.

Participants

Sixteen volunteers (12 male and 4 female), between the ages of 17 and 46, were recruited from the local university and nearby companies. All participants were frequent (4+ hours per day) computer users. Two participants had little to no exposure to interactive surfaces (including PDAs and Tablet PCs). Their educational level varied from high-school to post-graduate. No compensation was offered.

Apparatus

The study took place in a controlled laboratory setting. Participants completed the study while seated at a horizontal, top-projected 42-inch DiamondTouch table, with a 1024×768 pixel projected display. Participants could interact with two light sources, a DVD rack, a picture rack, a TV, a set of speakers, a digital picture frame, and a vacuum cleaner robot. As depicted in Figure 8, users were asked to sit either at the long side (*front*) or at the short side (*side*) of the table.

Figure 8: Evaluation setting.

Task

Participants were asked to perform a series of device control tasks. Task 1 was to switch on a floor lamp. In task 2, participants were asked to browse the entire movie collection and choose a particular movie and play it on the TV. Task 3 involved changing the audio volume of the TV from silent to loud. In task 4, participants were asked to select one representative photo out of a set of 20 photos, explain why they had chosen it, and enlarge it on any display they wanted. Finally, in task 5, they were asked to clean a certain spot on the floor by controlling the vacuum cleaner robot.

Procedure and Design

Participants performed the study in pairs. Each pair was asked to complete four trials of the sequence of five tasks described above. They were given as much time as needed to complete each trial. Each trial presented a different video image view or interaction method, including perspective or top-down video image view, and OBB or AABB axis-alignment interaction. The order of presentation of the different views and interaction methods was counterbalanced across pairs. When the pair was finished with their trials, they were interviewed by the experimenter to gather their opinions and experiences on using the CRISTAL system. Once the interview was complete, participants were asked to complete a post-study questionnaire which asked participants to assign an overall rank to each of the different interaction metaphors. The participants rated the ease of use of each interaction metaphor, using a 5-point Likert-scale (1 = totally disagree, 5 = totally agree). During the study, participants' interactions were observed by the experimenter, who took notes about their interaction with the system during the experimental tasks.

RESULTS

Overall, participants responded very positively to the CRISTAL system. All but one participant rated the system positively (rating it 4 or higher). Most participants also found the video interaction metaphor highly useful for controlling all proposed devices. Only one person reported preferring a conventional user interface. Many participants also liked the look and feel of the interface; 82% had no problems using the system after only one minute and thought that the system was easy to use. In contrast, some participants had problems finding the interactive areas without guidance from the experimenter (7/16 rated 3 or lower).

Figure 9 depicts the reported preferences of the interaction techniques in CRISTAL. In general, all interaction methods were found to be at least somewhat useful. Participants found the interaction methods for controlling the light sources, audio volume, and the TV the most useful, while the methods for browsing movies and photos were found to be less useful, though still with a positive trend for utility.

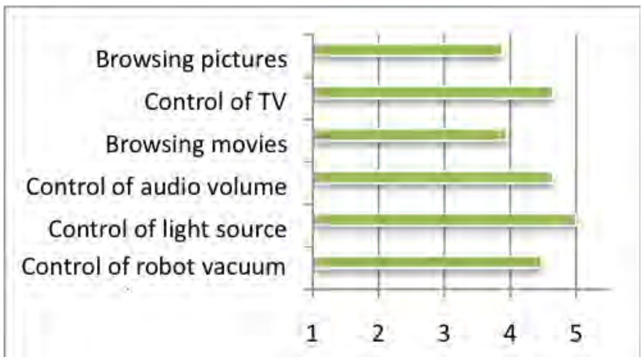

Figure 9: Participants ratings of the interaction metaphors, from useless (1) to highly useful (5).

75% of the participants preferred the top-down view of the camera to the perspective view. In fact, only a few participants seated at the *front* side (see Figure 8) of the table preferred the perspective view. No participant at the short side liked the perspective view. Many of them said that this view looks awkward. One participant also noted during the tasks, that it would be more convenient if the image would be a little smaller and all devices more easily reachable.

Most of the participants (75%) dimmed the light by sliding along the object's main axis, thus using the OBB approach. During the interview, most people said, that this mode is more intuitive than sliding along the system's main axis. Users, who preferred the AABB interaction mode, reported that they are used to handling horizontal or vertical sliders in conventional GUIs. One user, who also preferred the AABB interaction mode, also stated: "When the slider is axis-aligned then the direction should be the same for every user, because if this interaction mode was shown to me sitting on another side of the table I would think I have to do it the same way."

In the interview, people were asked if they usually control audio volume and dimmable lamps by adjusting them to certain values (e.g., 50% brightness or volume of 2 dB) or just intuitively. The reason for this question was to see what kind of feedback is necessary or if the video background feedback itself would is already enough. The majority (11 out of 16) said that they adjust those devices intuitively. In contrast, some participants preferred controlling per values. They said that they like to have objective values to be able to restore a certain setting later on.

DISCUSSION AND FUTURE WORK

In general, the results of the usability study confirmed the simplicity of CRISTAL; the video-based user interface was reported as being easy to use. Furthermore the tabletop interface provides support for fluid group interaction. Nevertheless, various aspects of the user interface warrant improvement and also the choice of the control device is not entirely certain.

Usability and Utility of a Tabletop Control Device

The outlines of the interactive areas were barely visible in the interface and most users did not recognize them immediately. However, one participant noted: "If this was my living room, I would know where my stuff is". Nevertheless, stronger visual feedback would likely support non-expert users in finding interactive objects (e.g. a visiting friend who is less familiar with the contents of the home). Furthermore, an additional feedback for drop-enabled areas would guide users through drag-and-drop interactions (e.g. when dragging a picture in the table interface, the TV and the digital picture frame could be visually highlighted to show users that those devices can receive the currently selected item).

The questionnaire results showed that most users are happy with the size of the video image and but a few had problems reaching distant controls. Generally, it appears that the table itself is too large for being used as an interactive coffee table. This problem could be solved by showing a smaller video image (i.e. not full-screen) and allow users to move it around on the table. Alternatively, zooming and panning of the image may also reduce this problem. Furthermore, multiple video images from different angles could be offered to the user. Although this would reduce occlusion and reach problems, it will also increase the complexity of the system.

For controlling movable devices (e.g. robot vacuum cleaner) the path-sketching approach seems to be more natural and easy-to-be used than setting individual checkpoints. In practice, a robotic vacuum like the iRobot Roomba would be an unlikely candidate for such direct control, as they are designed to be fully autonomous. The current system design uses a Roomba simply as a proof of concept to test the associated interaction mechanism. However, this style of interaction may work well for future helper robots in both domestic and work environments.

Beyond the home, interactive surface control systems may also be useful in corporate multimedia meeting rooms, where many different devices are used. For such rooms it is difficult to create a master control because not all devices are typically needed in every meeting and there are often different ways to use a single device or display. Instead, a system similar to CRISTAL could facilitate a quick and simple way to configure the room for a meeting.

Future Work

CRISTAL was a first attempt to develop an intuitive domestic control device that supports the social nature of the home. However, we recognize that a tabletop interface has limitations, especially in the casual environment of a living room. In many homes, coffee tables are often cluttered with magazines, board games, food, etc. Thus, we are currently investigating design variations of the CRISTAL system that are more accommodating of the limited space in some homes. For instance, one version that we developed, and are currently comparing to the CRISTAL tabletop system in an ongoing user study, is a mobile version that runs on one or more interactive Smartphones. Our goal in this line of investigation is to better understand the design tradeoffs between supporting individual control and group sharing of media and devices in casual, social environments.

CONCLUSION

In this paper, we presented the design, implementation and an initial evaluation of a tabletop system for controlling devices in a living room. CRISTAL provides an intuitive, gesture-based interface that enables users to control their home media devices through a virtually augmented video image of the room in which they are situated. The results of the initial user study of the system supported many of our design decisions and also provided further design recommendations for future iterations of CRISTAL and for similar video-based control systems.

ACKNOWLEDGEMENTS

We gratefully acknowledge the comments provided by our anonymous reviewers. Their advice was extremely helpful for improving the motivation and context of this work. We would also like to thank our study participants for their time and effort in providing valuable user feedback.

REFERENCES

1. Beijar, J., Leinerud, D., Nilsson, R., Thorin, P., and Weimar, G. (2007). Remotable: Managing a Built-in Media Center with the Table Top Surface. *Conference Supplement of Tabletop'07*, October 2007.
2. Dietz, P.H. and Leigh, D.L. (2001). DiamondTouch: A Multi-User Touch Technology. *Proc. of UIST'01*, 219-226.
3. Gottschalk, S., Lin, M. C., and Manocha, D. (1996). OBBTree: a hierarchical structure for rapid interference detection. *Proc. of SIGGRAPH '96*, 171-180.
4. Hollemans, G., Bergman, T. and Buil, V. (2006). Entertaible: Multi-user multi-object concurrent input. *Adjunct Proc. of UIST*, 55-56.
5. Billinghurst, M., Kato, H., Poupyrev, I. (2001). The MagicBook–Moving Seamlessly between Reality and Virtuality. *IEEE CG&A*, 21(3), 6-8.
6. Liao, C., Liu, Q., Kimber, D., Chiu, P., Foote, J. and Wilcox, L. (2003). Shared interactive video for teleconferencing. *Proc. of Multimedia'03*, 546-554.
7. Lucero, A., Aliakseyeu, D., and Martens, J.B. (2007). Augmenting Mood Boards: Flexible and Intuitive Interaction in the Context of the Design Studio. *Proc. of Tabletop'07*, 147-154.
8. Mazalek, A., Reynolds, M. and Davenport, G. (2007). TheTViews Table in the Home. *Proc. of Tabletop'07*, 52-59.
9. Sakamoto, D., Honda, K., Inami, M. and Igarashi, T. (2009). Sketch and Run: A Stroke-based Interface for Home Robots. *Proc. of CHI'09*, 197-200.
10. Shen, C., Lesh, N., and Vernier, F. (2003). Personal digital historian: story sharing around the table. *Interactions,* 10(2), 15-22.
11. Stoakley, R., Conway, M. J., and Pausch, R. (1995). Virtual reality on a WIM: interactive worlds in miniature. *Proc. of CHI'95,* 265-272.
12. Tani, M., Yamaashi, K., Tanikoshi K., Futakawa M., and Tanifuji S. (1992). Object-oriented video: interaction with real-world objects through live video. *Proc. of CHI'92*, 593-599.
13. Wigdor, D., Penn, G., Ryall, K., Esenther, A. and Shen, C. (2007). Living with a Tabletop: Analysis and Observations of Long Term Office Use of a Multi-Touch Table. *Proc of Tabletop'07*, 60-67.
14. Wobbrock, J. O., Wilson, A. D. and Li, Y. (2007). Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. *Proc. of UIST'07*, 159-168.
15. Yoon, H and Woo, W. (2008). Design and implementation of a universal appliance controller based on selective interaction modes. *IEEE Transactions on Consumer Electronics*, 54(4), 1722-1729.