Interactive Coffee Tables: Interfacing TV within an Intuitive, Fun and Shared Experience

Radu-Daniel Vatavu and Stefan-Gheorghe Pentiuc

University Stefan cel Mare of Suceava, Department of Computer Science,
str. Universitatii nr. 13, 720229 Suceava, Romania
vatavu@eed.usv.ro
http://www.eed.usv.ro/vatavu

Abstract. Watching television is usually a shared experience allowing family or friends that share the same viewing interests to watch, comment and enjoy programs together. The interaction part however is at the opposite end being reduced to the traditional remote control which by itself proves very limited with respect to the sharing part: although the viewing experience is shared among the group, the control part of the interface only allows one-viewer-at-a-time interaction. We are discussing in this paper a new interaction technique for controlling the TV set using one commonly available shared wide-area interface: the coffee table. By visually designating interaction sensitive areas on the coffee table surface, television control may be achieved via simple hand movements across the surface which may be performed by any of the viewers at any time. The final interface is thus fun, simple, intuitive, and very important, wide-shareable and immediately available for all the participants.

Keywords: interactive surfaces, tv, gestures, hci, computer vision.

1 Introduction

Families and friends usually gather in order to watch, comment and enjoy television shows, movies, news feeds or live sports transmissions for which they share the same interest. The most important fact to note here is that the global experience is a shared one: all the viewers share the same viewing transmission. The opposite happens however when it comes to the TV controlling interface: the traditional remote control only allows serialized one-viewer-at-a-time interactions which drastically limits the interaction with respect to the sharing part of the viewing experience. The interaction is not shared due to the fact that the interface device may be viewed as temporarily-blocked or temporarily-owned by other viewer at one particular time.

We address in this paper the very problem of allowing TV-control interactions via immediately available shared interfaces. We arrive at interactive coffee tables using simple logical implications: interaction should be shared and immediately available for all viewers, simple to achieve (should not add cognitive load or require training), fun if possible (the dimension of fun is introduced by novel

M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.): EuroITV 2008, LNCS 5066, pp. 183–187, 2008.

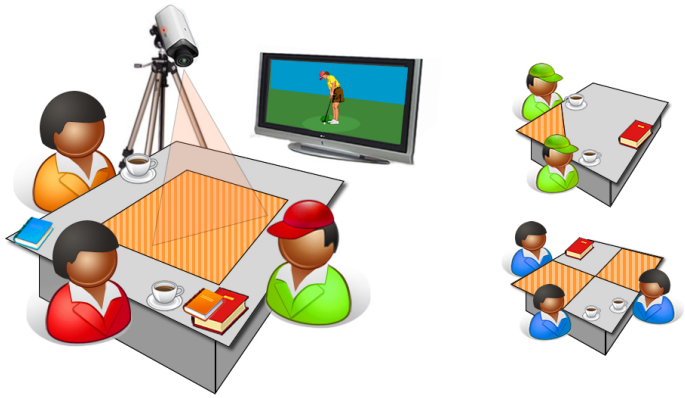

Fig. 1. Left: our vision for interactive coffee tables: all the viewers may interact with the TV set using the same wide-area interface via simple hand movements on top of the video-sensitive area of the table. The interactive area is depicted in the center of the table by using a darker color. Right: Several possible designs for the interaction sensitive area on the coffee table accommodating for 2 or multiple viewers.

user-friendly technologies). Also, TV viewing rooms usually have a coffee table. Our vision is illustrated in Figure 1: viewers are watching TV while gathered around the coffee table; a video camera is installed so that it monitors only the surface of the table; also, there is some sort of delimitation between interactive and non-interactive areas on the coffee table surface; TV control is achieved by positioning and moving one or both hands inside the interactive area. Diverse interactions may be achieved using simple hand gestures such as hand-open or hand-closed combined with the relative hands positions with respect to the sensitive area of the table.

2 Related Work

Speech interfaces have been proposed for controlling the TV-set in the context of natural language processing [1] with the main challenges being related to environmental noise and speaker variations. Gesture recognition systems have been tackled before as well [2,3,4]. They make use of computer vision techniques in order to detect and recognize hand gestures performed in sometimes very complex environments. Also, several unusual interfaces have been reported such as tangible [5], affect-input devices [6] or even plush toys [7].

Controlling TV with the bare hands is very intriguing. For example, [2] use hand posture recognition combined with visual feedback of the hand position on the TV screen. Seated on a couch, users manipulate a graphical icon of a hand on the screen. Activating TV menus and performing selections are also reported by [3] in the same working scenario. The ARGUS system [4] was designed to control home appliances (TV included): users point towards a device and control

its standard functions such as power on/off, volume up/down, play or stop via hand gestures.

One major disadvantage of all the above gesture-based interfaces is that having a video camera permanently facing and monitoring the viewers while they sit in front of the TV-set may issue problems and freights with regards to intimacy and privacy. Also, as noted in [3], the current setup with subjects seated facing the TV and performing gestures with their arms in mid-air is not suitable from the articulatory and musculatory point of view causing inconvenient interactions due to fatigue settling in. Problems also arise from the complexity of the scene to be analyzed by computer vision algorithms: crowded scenes with multiple objects and colors lead to difficulties in hand tracking with immediate effect on the entire system performance.

By moving the interaction to the coffee table space the mentioned problems are instantly removed. First of all, the interaction process is not fatiguant any more as the hands rest comfortably on the surface of the table and are not held in mid-air. The privacy issue is alleviated as the camera only monitors the coffee table and not the viewers. Also, we pay attention to choosing only two easy-to-perform hand postures (hand open and hand close) which in turn prove to allow for a variety of different interactions. And fourth, very important: the scene to analyze is far less complex as it is reduced to the surface of a table (of usually one single color) instead of processing images of an unknown living-room scenario.

3 Detecting Interactions

Hands may be detected above the surface of the coffee table by exploiting the contrast difference between the light skin color of the hands and the darker color of the table. We assume that a certain amount of contrast is present so that that hands may be segmented successfully. Simple gray-level image thresholding was used in order to segment the hands against the darker background of the table inside the margins of the sensitive area as Figure 2 illustrates.

For the actual interaction process we are interested in recognizing two postures: open-hand (all the fingers are stretched) and hand-closed (all the fingers are maximum flexed). As Figure 2 clearly shows, computing the bounding rectangles for the hands gives information about the two postures: the area, width and height of the hand-open bounding rectangle are clearly greater than the ones

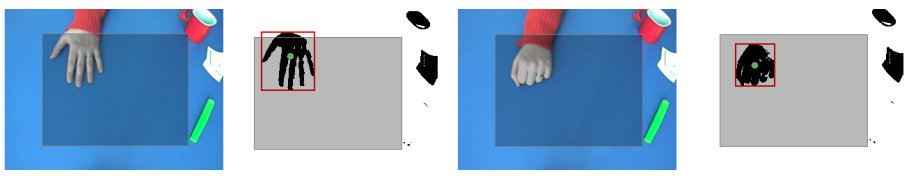

Fig. 2. Hands detection above the coffee table surface is achieved using simple color thresholding inside the margins of the interaction sensitive area

associated to the hand being closed. Hence again, simple thresholding against the geometric dimensions of the bounding boxes allows for easy and fast discrimination between the two postures.

4 Interfacing TV

The open-hand and closing the hand respectively act as a mouse-click event enabling viewers to acknowledge and issue commands. The principle is similar to the TAFFI interface [8]. Using these two hand postures several distinct interactions for interfacing TV-sets may be implemented.

Placing one open-hand in the sensitive area of the coffee table makes the channels selection menu appear as Figure 3 illustrates. Moving the hand across the surface determines the same movement of a hand icon on the TV screen with a permanent visual feedback. Closing the hand actually performs menu selection and the channel is changed. The same principle may act in selecting options from a general TV menu. Placing two hands inside the sensitive area of the coffee table allows controlling slide-like actions, for example modifying the audio volume or changing the brightness level. Similarly, the actual change is only triggered when the hands are closed and the effect is proportional to the relative distance between the two hands. Turning on/off the TV-set may be achieved by simply waving one open-hand inside the sensitive area.

Fig. 3. Selecting TV channels (left) using one hand and modifying the audio volume (right) in accordance with the relative distance between the hands

The system was implemented on a P4 2.66GHz desktop computer with digital mms live streaming TV. Video was processed at 25fps with a resolution of 320x240 pixels. The total CPU load was around 45% out of which the video processor took 35% while the rest was due to the Windows Media Player control. A demonstrative video of the system running is available at *http://www.eed.usv.ro/∼vatavu/*.

5 Conclusions and Discussion

We described in this paper an interaction technique for interfacing the TV-set using simple, intuitive and comfortable gestures. Gestures are detected inside a sensitive area on the coffee table. The sensitive pattern we implemented is not unique and other versions may be imagined as Figure 1 illustrates. We have not yet performed actual usability tests as they will be the subject of further work. However, the first results are positive due to the system simplicity and people seem impressed of the technology and its ease-of-use which stimulates continuation of our work.

The main advantages that our system brings with respect to related work are given by: simple video processing (the scene is known in advance and very simple) which leads to better tracking and recognition accuracy; there are no privacy or intimacy issues related to video cameras monitoring viewers as the camera only faces the coffee table; using two easy-to-perform hand postures (open hand and hand closed) a diversity of interactions may be implemented; and, very important, the coffee table allows for a wide-area interface which is evidently shareable and immediately available for all the viewers at any time.

As future work it may be interesting to investigate new options for the coffee table scenario: adding and recognizing objects (such as coffee-cups) inside the sensitive area may lead to new interactions. The coffee table may also act as a personal DVR with different areas of the table 'hidding' the recorded TV shows.

Acknowledgments. This work was supported by the national funding grant Research of Excellence, Ref. No. 131/2006 CEEX, INTEROB.

References

1. Fujita, K., Kuwano, H., Tsuzuki, T., Ono, Y., Ishihara, T.: A new digital TV interface employing speech recognition. In: IEEE Int. Conf. on Consumer Electronics, pp. 356–357 (2003)
2. Freeman, W.T., Weissman, C.D.: Television Control by Hand Gestures. In: IEEE International Workkshop on Automatic Face and Gesture Recognition, Zurich (1995)
3. Lenman, S., Bretzner, L., Thuresson, B.: Using marking menus to develop command sets for computer vision based hand gesture interfaces. In: The 2nd Nordic conference on Human-computer interaction, pp. 239–242. ACM Press, New York (2002)
4. Kohler, M.: Special topics of gesture recognition applied in intelligent home environments. In: Wachsmuth, I., Fröhlich, M. (eds.) GW 1997. LNCS (LNAI), vol. 1371, pp. 285–296. Springer, Heidelberg (1998)
5. Tahir, M., Baily, G., Lecolinet, E.: ARemote: A tangible interface for selecting TV channels. In: ICAT 2007: the 17th Annual Conference on Artificial Reality and Telexistence, pp. 298–299. ACM Press, New York (2007)
6. Lee, C.-H.J., Chang, C., Chung, H., Dickie, C., Selker, T.: Emotionally Reactive Television. In: ACM IUI 2007, pp. 329–332 (2007)
7. Kawasaki, Y., Igarashi, T., Ajioka, T., Honda, I.: Vision-based Gestural Interaction Using Plush Toys. In: ACM UIST 2005 (2005)
8. Wilson, A.: Robust Vision-Based Detection of Pinching for One and Two-Handed Gesture Input. In: ACM UIST 2006 (2006)