Dynamic user representation in video phone applications

Andreas Braun[1], Reiner Wichert[1]

[1] Fraunhofer Institute for Computer Graphics Research - IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany
{andreas.braun, reiner.wichert}@igd.fraunhofer.de

Abstract. Video phone applications are growing more commonplace with integration into mobile smart phone platforms like Apple iOS or into online social networks like Facebook. However users may desire to not show their present mood or disorderly appearance while still desiring to use such applications. Virtual user representations are an option to hide the actual appearance while still participating in video phone calls. This paper discusses different approaches to using virtual characters in video phone applications, dynamic self-representation and user interface considerations.

Keywords: Virtual Self-Representation, Emotion Recognition, Graphical user interfaces

1 Introduction

With available bandwidth increasing for both mobile and home devices connected to the internet, video phone applications that allow multiple users to communicate via video feed submission is growing more popular. The most popular smart phone platforms - Android and iOS - are supporting this feature for mobile devices. The majority of Notebooks and All-in-One PCs are factory-equipped with webcams. Facebook, currently the most popular social network is supporting video phone via plug-in.

With the extending propagation of video chat different user groups are targeted that have a different set of requirements than the tech-savvy early adopters. It is a reasonable assumption that many users that are participating in video chats do not want to share their current appearance, e.g. if they are not feeling well or have not completed their daily hygienic routine yet. The avatar principle, whereas someone uses a virtual representation that may or may not resemble the actual appearance, is commonplace in certain applications, most notably online games. Extending this principle to video phone applications poses various difficulties, ranging from computing complexity due to the real-time nature of this scenario to the control of this avatar.

This paper will discuss the problems related to video phone avatars, the state of modern virtual characters and how to control avatars in these scenarios.

2 Video phone applications

The transmission of a video signal via phone has been imagined very early in the history of telecommunication [1], leading to various commercial trials in the 20th century [2]. However the technology did not become economically feasible until the camera hardware and bandwidth became affordable enough for widespread adoption.

The current state includes ubiquitous availability of cheap camera hardware that is included in most smartphones and notebooks or can be added to regular PC systems at a minimal price. The required bandwidth for video phone applications is depending on the image quality to be achieved. Considering a high efficiency compression algorithm like H.264 any mobile device supporting UMTS and any PC connected to the internet with a speed of 1Mbit/s are sufficiently equipped for real-time video transmission. In conclusion the technical challenges for video phone applications can be considered solved nowadays.

Usage numbers of video telephony compared to regular phone and text messages is still very low [3]. Studies have shown that the interest of such systems is high for families that live far apart and have reduced opportunities for face-to-face communication and the deaf community, using such systems for communication in sign language [4].

3 Modern virtual characters and the uncanny valley

Figure 1. Facial rendering in modern game engines - from left: Half Life 2, Unreal Engine 3, Cryengine 3

Within the last decade the graphical capabilities of mobile and home computers has increased greatly, leading to the ability of rendering figures in real-time in a quality similar to early computer generated movies that may have computed several days on a single image. The result is highly detailed virtual characters with emotionally expressive and realistic facial animation and gestures available on all platforms. A few examples can be seen in Figure 1.

The uncanny valley is a hypothesis postulated by Masahiro Mori in 1970 that expects a negative emotional response on human replicas that almost approach a human being in look and action [5]. On a graph showing familiarity and human likeness the area of negative response between almost human and fully human is

called the uncanny valley (Figure 2). When designing applications with avatars this effect has to be considered.

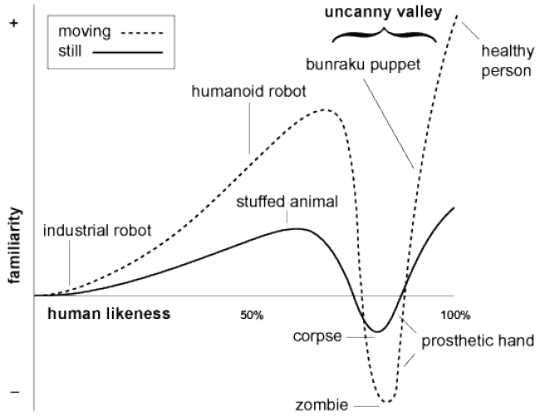

Figure 2. Human likeness in robotics and their perceived familiarity

4 Explicit and implicit avatar control

In many avatar applications, e.g. games, the user does explicitly control the movement and behavior of his virtual self. In video phone applications it is viable to offer different interaction metaphors, explicit UI based control to change the avatar behavior, as well as implicit interaction, e.g. through image analysis that provides a more immersive video phone experience without visible interface elements.

4.1 Dynamic user representation

Dynamic user representation is the feature of letting the user select in real-time different levels of detail regarding his avatar for video phone applications. E.g. it can be envisioned to use a system providing four different levels of representation, an iconic representation, a two-dimensional static representation, a three-dimensional animated character and the actual unaltered video feed of the user. This gives the user explicit control over his appearance in the video call and therefore potentially increases the acceptance of video phone applications by providing choice.

4.2 Emotion Recognition

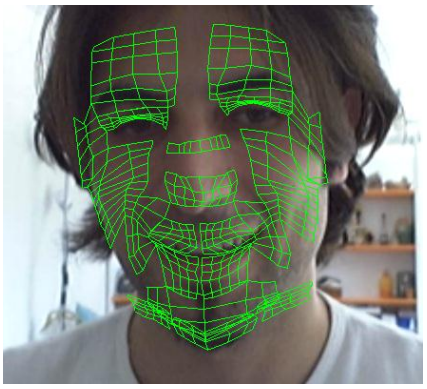

Figure 3. Detection of facial expressions to gather information about the user's mood.

The recognition of emotion through analysis of images taken of the face has been researched extensively [6]. The required computations do require a certain hardware specification that is however matched by most modern smartphones and every common PC. The recognized emotion or mood can be used to implicitly control the user avatar, e.g. a detected smile of the user can be applied to the virtual character directly.

4.3 Low cost 3D hardware - the impact of Kinect

Figure 4. 3D reconstruction of objects and false color texturing

In late 2010 Microsoft made the Kinect system commercially available, a combination of infrared depth tracking and RGB camera that is powerful while being inexpensive. Researchers have shown examples how this technology can be used to create a 3D model from a regular person in real-time [7]. System like this may be used for generating avatars of the user for video phone applications with varying and dynamically changing levels of detail.

5 Discussion

It can be concluded that the increasing bandwidth together with advancing computing capabilities and algorithms does allow using avatars in video phone applications that modify appearance and behavior according to the data gathered by sensors attached. Emotion recognition can be used to select from a variety of pre-determined animations while 3D hardware can be used to directly map recognized facial and skeletal features to the model. Both approaches can help to overcome the uncanny valley effect in video phone applications.

References

1. du Maurier, G. Punch magazine, December 9th, 1878 (1878)
2. Daly, Edward A. & Hansell, Kathleen J. Visual Telephony, Artech House, Boston. (1999)
3. Press Release: Smartphone Video Call Users to reach 29 million by 2015 Globally, finds Juniper Research. http://juniperresearch.com/viewpressrelease.php?pr=209 (Retrieved August 28th, 2011)
4. Kentucky School for the Deaf (KSD). DEAFinitely Connected: Bridging the Language Divide with Telecommunications, Computerworld Honors Program, Computerworld Information Technology Awards Foundation, Worcester, MA, 2010. (Retrieved August 28th, 2011)
5. Mori, M. Bukimi no tani The uncanny valley (K. F. MacDorman & T. Minato, Trans.). Energy, 7(4), 33–35. (1970)
6. De Silva, L.C., Miyasato, T., Nakatsu, R. Facial emotion recognition using multi-modal information. Proceedings of 1997 International Conference on Information, Communications and Signal Processing. (1997)
7. Zollhöfer M., Martinek M., Greiner G., Stamminger M., Süßmuth J. Automatic reconstruction of personalized avatars from 3D face scans. Computer Animation and Virtual Worlds, Volume 22, Issue 2-3, pages 195–202. (2011)