REALISTIC AND AFFORDABLE VIRTUAL HUMANS

Belen Masia[1*] João Freitas[2] Verónica Orvalho[3] Mel Slater[4] Miguel Sales Dias[2] Diego Gutierrez[1]
[1]*Universidad de Zaragoza* [2]*Microsoft MLDC* [3]*Instituto de Telecomunicações, Porto* [4]*Universitat de Barcelona*
[*]*bmasia@unizar.es*

ABSTRACT

Creating realistic virtual humans is a multidisciplinary effort, involving computer scientists specialized in computer graphics, virtual reality and human-computer interaction, as well as neuroscientists. This paper discusses, under the context of an EU-funded project, the state of the art in this regard, open challenges and future trends. Modifying how a virtual human looks, speaks, behaves or can be controlled is critical to the development of future collaborative virtual environments, while boosting other industries (games, movies…) by offering a new set of possibilities currently out of reach for most.

KEYWORDS

Virtual humans; computer graphics; realism; presence.

1. INTRODUCTION

This paper discusses the state of the art, open challenges and advances in the creation of *cheap* (both in terms of computation and financial cost) realistic virtual humans, which can be used in movie productions, games or even in neuroscience and psychology studies (we will refer to these under the general term of *presence*). We discuss these issues within the context of the GOLEM project. GOLEM is an EU-funded project involving both industrial and academic partners, which started in 2010 and is scheduled to finish in September of 2014. It is an international consortium, formed by members from both industry and academia, including Face in Motion, Universitat de Barcelona, Microsoft MLDC, the Instituto de Telecomunicações in Porto and Universidad de Zaragoza.

GOLEM proposes a novel animation pipeline to ease the character setup process and provide high quality animations, including lip-sync, and photorealistic appearance. This is currently a very expensive task, both in terms of time and money, due to the lack of a generalized system that can automatically customize different characters on the fly. Current systems require painstaking manual creation and adjustment for every single character. The main scientific and technological outcome of the project is a comprehensive system that animates spoken characters and enables avatars to respond in real-time in a variety of (virtual) environments. The system, on the one hand, speeds up the animation and rendering processes, and on the other hand aids in the study of social behaviors, like social phobia, in order to develop appropriate therapies.

2. TOWARDS REALISTIC VIRTUAL HUMANS

The overall objective of GOLEM has been to share knowledge between Industry and Academia in order to carry out multidisciplinary research that results in technology that radically improves the visual appearance and behavior of virtual characters (avatars), while streamlining the production pipeline and keeping them customizable and affordable. Virtual humans can have a relevant impact in (1) industries (such as movies or games), (2) academia (by advancing the state of the art in the related research fields) and (3) neuroscience (by using the avatars in novel experiments about presence in virtual environments). This requires a multidisciplinary approach involving human-computer interaction, computer graphics and neuroscience (Figure 1, *left*).

The specific challenges are two-fold: **First**, the development of a novel, optimized pipeline that covers capture, synthesis and animation of realistic virtual humans in real time, to enhance current state-of-the-art techniques to create virtual humans in fields such as game production or cinematography. Building on the previous challenge, the **second** one is conducting novel research on presence in virtual environments.

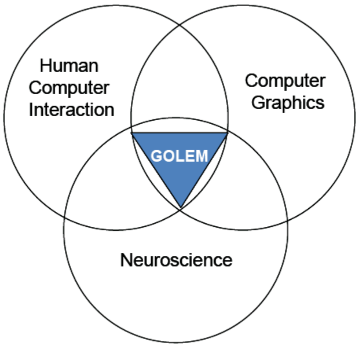

Figure 1. Left: the intersection of the three major areas of research involved in GOLEM. **Right**: An example of an existing, complex, capture hardware. The sphere (called Light Stage) is 6 meters in diameter. 150 computer-controlled LEDs and cameras illuminate the subject under very different lighting conditions to store a series of basis functions. By combining them later in novel ways, relighting can be performed. Note however that no changes in the skin appearance itself are allowed with the data captured, only lighting can be changed (image courtesy of ICT Graphics Labs, USC).

3. STATE OF THE ART AND OPEN CHALLENGES

Despite current advances in technology, producing cheap, realistic virtual humans remains a difficult task. The main research challenges arise from the synchronization and realism problems, the support for the reusability of components, and the need for an avatar-user interaction model with real time response. Traditional techniques to achieve high quality human animation are very time-consuming, expensive and laborious, and usually include key frame animation and motion capture based on facial markers. This is not only cumbersome, but unpleasant and unnatural as well. Real-time rendering techniques are also limited with respect to dynamic responses from the character or from the environment. Thus, creating realistic virtual humans is nowadays performed off-line (for example, in movies, where no real-time interaction is required) following expensive, per-character procedures. Other contexts, such as games or virtual reality, adopt cheaper approaches, at the expense of the overall visual quality and credibility of the models.

Data acquisition: Acquiring optical properties of translucent materials such as skin requires coming up with a model of the Bidirectional Subsurface Scattering Distribution Function (BSSRDF). Given its high dimensionality (the BSSRDF is an 8D function), acquisition of this function typically requires either the use of relatively complex measuring equipment, and/or sets of photographs that capture the appearance of materials from a wide range of viewing and lighting directions (Debevec et al. 2000, Goesele et al. 2004) (see Figure 1, *right*). Recent advances in reflectance acquisition have led to more sparse sets of input images being used (Ghosh et al. 2008, Donner et al. 2008), and also to the use of a single image together with knowledge of 3D geometry (Boivin and Gagalowicz 2001).

Rendering: Regarding skin rendering, Jensen and co-workers (Jensen et al. 2001), based on a dipole approximation of light diffusion, were able to capture the subtle softness that translucency adds to the appearance of skin. Donner and Jensen extended their dipole-based model to multiple dipoles, which can also capture the effects of discontinuities at the frontiers of multi-layered materials (Donner and Jensen 2005, 2006). Unfortunately, the model relies on a high number of parameters that need to be measured in advance, which makes its direct application to rendering slightly cumbersome.

Animation: Much effort has gone into retargeting or automatically creating the body rig of characters (Hecker et al. 2008) and achieved good results, but none have focused on facial rig retargeting. Most facial animation research is related to physically-based, geometric deformation or performance-driven methods

(Borshukov et al. 2006). Captured facial performances can be re-targeted to different face models (Chai et al. 2003) or blendshapes (Deng et al. 2006). These can generate realistic facial motion, but are expensive to use. Real-time markerless facial motion capture is currently an unsolved problem. Some research in this field includes work from Schreer et al. (2008) and Image Metrics^{TM} (2014); they combine different technologies (e.g., image analysis, optical flow) to achieve good results but setup is still too costly in time.

Presence: *Presence* deals with the concern of how people respond to virtual environments rather than the level of realism itself. The 'response as if real' (RAIR) has been the focus of increasing research (Sanchez-Vives and Slater 2005). It has recently been argued that there are two key components that may lead to RAIR (Slater 2009). The first is a static element that is referred to as 'place illusion' (PI), the strong illusion of actually being in the portrayed virtual place. The second major element is 'plausibility', that is, the illusion that these events are really happening. This has been found to depend on three sub-elements: the extent of correlation between the actions of the participant and events in the world, the extent to which the world directly references the participant, and the overall credibility of the scenario when depicting events that could happen in physical reality. Plausibility has been hardly addressed in the literature, and it is especially interesting in relation to interactions between real people and virtual characters, the question being: even though a participant knows for sure that a character is not a real person, do they still tend to respond to them as if they were real? There is significant evidence that this is indeed the case (Pan and Slater 2007) but there have not been detailed studies that attempt to investigate the factors that might be responsible for this. The ideas of correlation, self-reference and credibility have been derived from analysis of a large number of previous experiments, but remain to be verified by specifically devoted studies.

4. RESULTS AND CONCLUSIONS

The GOLEM project has achieved remarkable results during its four-year life span; there are significant advances in all aspects discussed in Section 3. Some examples include *Sketch Express*, an interface to create and edit facial expressions easily (Miranda et al. 2011, 2012); speech and lip-sync contributions to the state of the art (Freitas et al. 2012, Serra et al. 2012); an ultra-realistic real-time rendering algorithm for human skin (Jimenez et al. 2010, 2012); or advances in presence and embodiment in virtual environments (Borland et al. 2013, Blom et al. 2013). Figure 2 shows examples of these contributions.

Despite this large number of advances, there is still work ahead. The most important is the creation of a unified framework where *all* the technology developed can be applied, from acquisition to final render. This is mostly an engineering effort, although a considerable one. Integrating the ultra-realistic virtual humans into existing virtual reality applications poses another interesting problem, since the high degree of realism of the characters will need to be matched by everything else in the simulation. Last, we believe our virtual humans can be instrumental in future studies about the uncanny valley.

ACKNOWLEDGEMENTS

We would like to thank the reviewers. This work was funded by the European Commission through the GOLEM project.

REFERENCES

Blom, K. J. et al, 2013. Achieving Self-Likeness Between Participants and their Avatars. *Submitted to PRESENCE*.

Boivin, S. and Gagalowicz, A., 2001. Image-based rendering of diffuse, specular and glossy surfaces from a single image. *In ACM SIGGRAPH 2001*, pp. 107-116.

Borland, D. et al, 2013. An Evaluation of Self-avatar Eye Movement for Virtual Embodiment. *In IEEE Trans. on Vis. and Comp. Graphics*, Vol. 19, no. 4, pp. 591-596.

Borshukov, G. et al, 2006. Playable Universal Capture: Compression and Real Time Sequencing of Image-based Facial Animation. *In ACM SIGGRAPH 2006 Courses*.

Chai, J. X. et al, 2003. Vision-based Control of 3D Facial Animation. *In ACM Symp. Comp. Animation 2003*, pp.193-206.

Debevec, P. et al, 2000. Acquiring the Reflectance Field of a Human Face. *In ACM SIGGRAPH 2000*, pp. 145-156.

Deng, Z. et al, 2006. Animating Blendshape Faces by Cross-mapping Motion Capture Data. *In ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games 2006 (I3D'06)*, pp. 43-48.

Figure 2. Sample results of the project. In reading order, *SketchExpress* (Miranda et al. 2012), evaluation of self-avatar eye movement (Borland et al. 2013), an example rendering using separable subsurface scattering (Jimenez et al. 2012), and renderings obtained with an appearance model for dynamic facial color (Jimenez et al. 2010).

Donner, C. and Jensen, H. W., 2005. Light Diffusion in Multilayered Translucent Materials. *In ACM Trans. on Graphics*, Vol. 24, no. 3, pp. 1032-1039.

Donner, C. and Jensen, H. W., 2006. A Spectral BSSRDF for Shading Human Skin. *In Proc. of EGSR* 2006, pp. 409-417.

Donner, C. et al, 2008. A Layered, Heterogeneous Reflectance Model for Acquiring and Rendering Human Skin. *In ACM Trans. on Graphics*, Vol. 27, no. 5, pp. 140:1-12.

Freitas, J. et al, 2012. Towards a Multimodal Silent Speech Interface for European Portuguese. *In Intnl. Conf. on Bio-Inspired Systems and Signal Processing (BIOSIGNALS) 2012*, pp. 125-150.

Ghosh, A. et al, 2008. Practical Modeling and Acquisition of Layered Facial Reflectance. *In ACM Trans. on Graphics*, Vol. 27, no. 5, pp. 139:1-10.

Goesele, M. et al, 2004. Disco: Acquisition of Translucent Objects. *In ACM SIGGRAPH 2004*, pp. 835-844.

Hecker, C. et al, 2008. Real Time Motion Retargeting to Highly Varied User Created Morphologies. *In ACM Trans. on Graphics*, Vol. 27, no. 3, pp. 27:1-11.

Image Metrics, 2014. Image Metrics: Engineering and Consulting Services. www.imagemetrics.com

Jensen, H. W. et al, 2001. A Practical Model for Subsurface Light Transport. *In ACM SIGGRAPH 2001*, pp. 511-518.

Jimenez, J. et al, 2010. A Practical Appearance Model for Dynamic Facial Color. *ACM Trans. on Graph.*, 29(6), 141-10.

Jimenez, J. et al, 2012. Separable Subsurface Scattering. *Technical Report RR-02-12, Universidad de Zaragoza*.

Miranda, J. C. et al, 2011. Sketch Express: Facial Expressions Made Easy. *In Eurographics Symp. on Sketch-Based Interfaces and Modeling 2011*.

Miranda, J. C. et al, 2012. Sketch Express: A Sketching Interface for Facial Animation. *In Computers & Graphics 2012*.

Pan, X. and Slater, M., 2007. A Preliminary Study of Shy Males Interacting with a Virtual Female. *In PRESENCE 2007*.

Sanchez-Vives, M. V. and Slater, M., 2005. From Presence to Consciousness Through Virtual Reality. *In Nature Reviews Neuroscience*, Vol. 6, no. 4, pp. 332-339.

Schreer, O. et al, 2008. Real Time Vision and Speech-driven Avatars for Multimedia Applications. *In IEEE Trans. on Multimedia*, Vol. 11, no. 3, pp. 352-360.

Serra, J. et al, 2012. Automatic Visual Speech Animation. *In Intnl. Conf. on Comp. Proc. of Portuguese Language 2012*.

Slater, M., 2009. Place Illusion and Plausibility Can Lead to Realistic Behavior in Immersive Virtual Environments. *In Philos. Trans. R. Soc. Lond. 2009*.