Stylization of Virtual Humans

Martin Breidt* Heinrich Bülthoff* Reuben Fleming† Betty Mohler*

*Max Planck Institute for Biological Cybernetics
Tübingen, Germany
†Sheffield Hallam University, UK

ABSTRACT

The growing importance of Virtual Reality (VR) in recent years has put emphasis on the topic of avatars: Especially for shared VR experiences, it is important to provide a visual representation of the user. An important question is: What will these avatars look like? Even though VR offers great artistic freedom in the design of virtual avatars, serious applications of social VR will most likely prefer a design close to the real appearance of humans. For this, a natural choice is the derivation of an avatar from the true appearance of a human person, using photographs or 3D scans, for example. While this will result in a close resemblance to the real person, achieving ultimate photo-realism is challenging, especially considering the typically uncontrolled recording conditions present in normal situations. Human avatars produced from such digital body captures are likely to be clearly distinguishable from photographs, which might result in negative reactions from observers due to conflicting cues. Instead of fighting the uphill battle to achieve ultimate photo-realism, we propose a deliberate deviation from realism by means of automatic stylization of such body captures. Here we recap previous research on this topic by us and others, summarize experimental results from an automatic stylization system, and discuss potential applications and future work.

Index Terms: Computing methodologies—Virtual reality. Computing methodologies—Computer graphics—Graphics systems and interfaces—Perception. Human-centered computing—Empirical studies in HCI.

1 INTRODUCTION

In recent years, VR has received a lot of industry attention, e.g. global venture capital investments in VR/AR reached almost 2 billion US$ annually in 2016 and 2017 [12]. It is likely that avatars (the graphical representation of users in VR) will play an important role in future VR applications, either to support anchoring a single user in VR [17] or to enable social VR applications where multiple users share the same virtual environment.

What will you look like in VR?

With this, the question immediately arises as to what those avatars should look like. While VR offers a lot of artistic freedom for avatar design, we expect that for most 'serious' applications, such as virtual business meetings, job interviews, online teaching, or customer services, a human-like appearance is desirable for natural interaction. This would also alleviate some real-time motion retargeting issues arising from large differences in body proportions between the user and the avatar. Additionally, as VR technology matures to a point where it can successfully transmit subtleties of non-verbal communication, a human-like appearance is possibly most compatible with human communication patterns. We call these human-derived avatars 'virtual humans'.

*e-mail: {firstname.lastname}@tuebingen.mpg.de
†e-mail: R.Fleming@shu.ac.uk

Virtual Humans and Crowds in
Immersive Environments 2018
19 March 2018, Reutlingen, Germany

2 RELATED WORK ON AVATAR CREATION

While there is a large body of work on creating avatar *faces* (e.g. [4,8–10,13,22]), a *full body* avatar is needed for many VR applications to avoid the 'floating head' effect. Creating such full body 3D avatars for real persons has also been the topic of a wide range of previous work, both in academia as well as in industry: Some approaches offer an interface to artistically define an avatar appearance, e.g. by selecting from a library of avatar elements such as hairstyle or head shape [20]; this is also popular in commercial large-scale VR environments such as Second Life, AltspaceVR, or VRChat. Others try to compute an avatar from still images (e.g. [5,11,14]). Finally, some approaches (e.g. [1,15,23,25]) require dense body shape and color measurements (taken by a 3D scanner or multiple photographs) to calculate an avatar, while some approaches (e.g. [3]) bypass the creation of avatars by directly streaming 3D point clouds captured by depth sensors.

The artistic creation of avatars resembling real people can be a difficult task for an average user. Therefore, a natural choice is to capture the appearance of a real person with recording hardware such as 3D scanners, depth sensors or digital cameras. Depending on the hardware, this approach is likely to produce semi-realistic virtual humans, either due to insufficient data, due to difficulties capturing the full appearance of all parts of the body, or due to a mismatch between appearance and motion of the virtual human. In addition, the lack of high-quality lighting in most recording situations will result in poor color appearance and likely clash with the real-time lighting solution used in a VR application.

With the exception of expensive recording setups in specialized 3D scanning studios using high-end multicamera photogrammetric systems (e.g. [4,6,8,21,24]), these body captures will likely result in virtual humans that are clearly distinguishable from photographs.

Such body capture approaches vary widely on the amount of data needed from the user, or the hardware required for appearance acquisition as well as for avatar calculation. Accordingly, the results produced will vary widely in terms of realism and likeness, and with that in suitability for different use cases. The high sensitivity of human perception for other human figures makes this task particularly challenging; since the box office failure of the animated movie *Final Fantasy: The Spirits Within* in 2001 the catchphrase 'Uncanny Valley', initially coined to describe the discomfort caused by human-like robots [19], has been widely used to describe the negative audience reactions to semi-realistic virtual humans in games, VR or movie productions.

3 STYLIZATION

When trying to overcome the lack of photo-realism in virtual humans, increasing resources and effort are needed to ensure that remaining small but perceptible deviations from a real human appearance do not result in negative reactions.

An alternative approach is the *deliberate deviation* from realism by stylization. This is similar to how early architectural designs

are often presented in a sketched drawing style, rather than photo-realistic rendering, in order to visually indicate the uncertain nature of the concept. Several companies in the animation industry, such as Pixar Animation Studios, have successfully followed this notion by intentionally avoiding too much photo-realism in the depiction of computer-generated humans, even though the computer-generated environments in such productions already approach photo-realistic levels [2].

In a systematic analysis of the effect of rendering style on virtual characters [16], we found evidence that stylized (upper body) avatars are rated higher than those rendered in a semi-realistic style (Fig. 1), when keeping shape and animation otherwise constant.

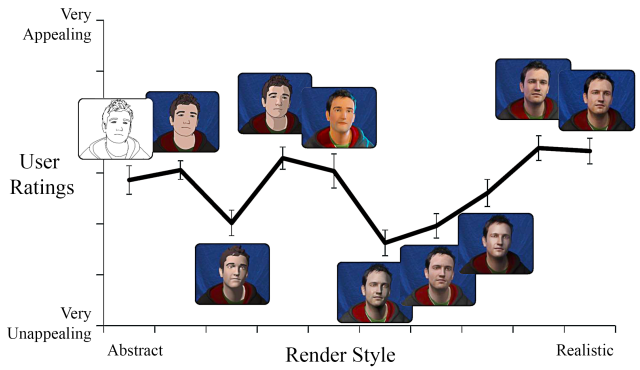

Figure 1: Effects of stylized rendering on the judgment of animated virtual humans [16].

Shape Stylization

While stylization on the rendering side (e.g. toon or sketch shading) has been investigated extensively in the field of non-photo-realistic rendering, recent work (e.g. [26]) has started to look into high-level stylization of human figures that also effect the overall appearance.

In a series of experiments [7], we explored the influence of appearance stylization on the perception of high-level character traits. For this, an automated process gradually stylized 3D body scans of real people towards popular character designs adopted from commercial productions (Fig. 2 and 3). The only manual input to the system is the one-time creation of a database of 3D style templates; all the other parts of the 3D stylization process are automated.

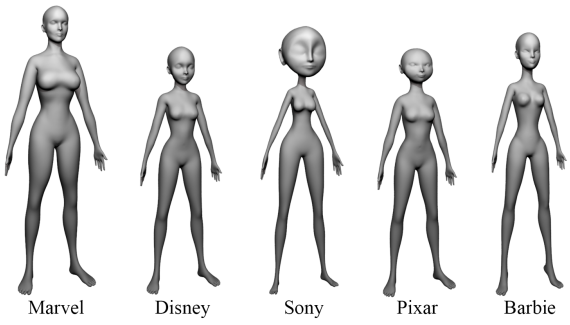

Figure 2: Style templates used in [7].

Participants were asked to rate the appearance of virtual humans produced by this process, and also to interactively adjust the amount of stylization required in order to obtain what they considered to be the most appealing virtual human. Results show (Fig. 4) that a modest amount of stylization (between 30–40% in our experiments) was able to increase *appeal* ratings of our scanned actors by about 30% over their original ratings, while higher amounts of stylization produced negative ratings. Interestingly, we also found that a higher level of rendering realism did not affect the appeal rating results. See [7] for more details and full analysis.

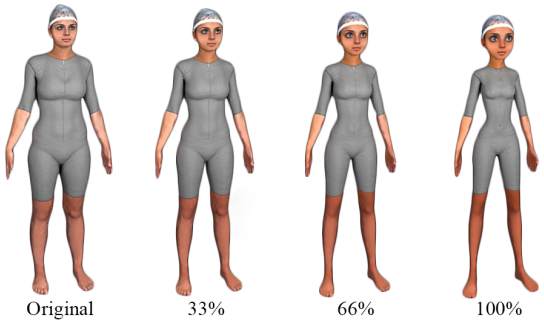

Figure 3: 3D body scan (left) gradually morphing into a 'Disney' style (right).

With such a system, potential users could have their body scanned once in a local studio and then reuse that body capture for different types of stylization depending on the context. This would minimize negative reactions, retain the resemblance of the real person, avoid the need for high-end acquisition systems, and be easy to control without the need for artistic talent. The specific stylization could be chosen depending on personal taste and context, analogous to how we dress for different occasions.

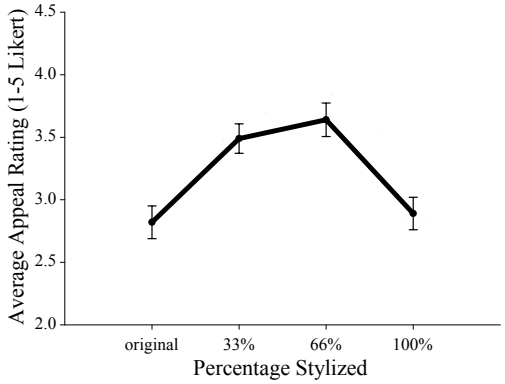

Figure 4: Appeal ratings (N=61, bars = standard error) for the avatar shown in Fig. 3 with increasing amounts of 'Disney' stylization [7].

4 Discussion and Outlook

So far, our experiments indicate that a modest amount of automatic stylization is able to improve the perception of 3D body scans. Virtual humans stylized with this approach are likely to be perceived more positively than the unmodified original. Instead of asymptotically converging on photo-realistic virtual humans with increasing effort and cost, and yet risking a negative response due to the high perceptual scrutiny of human observers, a controlled reduction in realism by stylization could be an alternative solution for improving the success of virtual humans.

With recent advances in model-based 3D body reconstruction from photographs [5], it might become even easier to capture the required body data for the stylization process, avoiding the current requirement for a full 3D body scan in order to stylize a person's appearance.

The optimal style and stylization amount is likely to depend on the task and the context of the avatar; a teaching scenario might require

a different style than a virtual sales pitch, for example. In addition, personal characteristics of the real person might be more suitable for specific stylizations. This will require further investigation.

Additional experiments are also needed to investigate the role of body motion when combined with such stylized avatars. If the proportions of the virtual human deviate from those of the user, the aspect of motion retargeting, i.e. the mapping of the user's original motion to the virtual body, is an important part to the success of a virtual humans. In particular, foot or hand contact points need to be corrected for deviations in proportions. To ensure proper contact between limps and the own body (e.g. when touching the belly or the head), a bodycentric encoding of the spatial relationship between body parts [18] could be calibrated with the dense surface reconstruction of the user's body, which is already available for the stylization process.

Also, the motion data itself needs to be considered in future work – mapping realistic motion onto stylized avatars might benefit from some degree of *motion* stylization, which could be challenging to compute due the real-time nature of social VR and the ill-defined notion of stylized motion. More generally, we expect stylized avatars to be less in conflict with sub-optimal or missing movements, an interesting aspect considering the technological challenges involved in accurately transmitting the subtleties of human movement (e.g. face, hands, hair, clothing) within a VR system.

Future work could also look into different measures of avatar success. Instead of directly collecting rating responses on specific avatar qualities such as appeal, trustworthiness, or attributed expertise, more indirect measures such as memory retention in a teaching scenario or sales success could be interesting.

Acknowledgments

The authors would like to thank Michael Black, Javier Romero and colleagues for their work on model-based scan registration and body parametrization, as well as Rachel McDonnell for her work on stylization.

References

[1] J. Achenbach, T. Waltemate, M. E. Latoschik, and M. Botsch. Fast generation of realistic virtual humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, pp. 12:1–12:10. ACM, New York, USA, 2017. doi: 10.1145/3139131.3139154

[2] A. Barillaro. Piper. Animated Short Film, 2016. Pixar Animation Studios.

[3] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, April 2013. doi: 10.1109/TVCG.2013.33

[4] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, R. W. Sumner, and M. Gross. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graph.*, 30:75:1–75:10, August 2011. doi: 10.1145/2010324.1964970

[5] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black. Keep it SMPL: Automatic estimation of 3d human pose and shape from a single image. In B. Leibe, J. Matas, N. Sebe, and M. Welling, eds., *Computer Vision – ECCV 2016*, pp. 561–578. Springer International Publishing, Cham, 2016.

[6] P. Einarsson, C.-F. Chabert, A. Jones, W.-C. Ma, B. Lamond, T. Hawkins, M. Bolas, S. Sylwan, and P. Debevec. Relighting human locomotion with flowed reflectance fields. In *Proc. 17th Eurographics Conference on Rendering Techniques*, EGSR '06, pp. 183–194. Eurographics Association, Aire-la-Ville, Switzerland, 2006. doi: 10.2312/EGWR/EGSR06/183-194

[7] R. Fleming, B. J. Mohler, J. Romero, M. J. Black, and M. Breidt. Appealing avatars from 3d body scans: Perceptual effects of stylization. In J. Braz, N. Magnenat-Thalmann, P. Richard, L. Linsen, A. Telea, S. Battiato, and F. Imai, eds., *Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 175–196. Springer International Publishing, Cham, 2017.

[8] G. Fyffe, A. Jones, O. Alexander, R. Ichikari, and P. Debevec. Driving high-resolution facial scans with video performance capture. *ACM Trans. Graph.*, 34(1):8:1–8:14, Dec. 2014. doi: 10.1145/2638549

[9] P. Garrido, M. Zollhöfer, D. Casas, L. Valgaerts, K. Varanasi, P. Pérez, and C. Theobalt. Reconstruction of personalized 3d face rigs from monocular video. *ACM Trans. Graph.*, 35(3):28:1–28:15, May 2016. doi: 10.1145/2890493

[10] A. E. Ichim, S. Bouaziz, and M. Pauly. Dynamic 3d avatar creation from hand-held video input. *ACM Trans. Graph.*, 34(4):45:1–45:14, July 2015. doi: 10.1145/2766974

[11] itSeez3D. Realistic avatars for games, VR and AR. https://avatarsdk.com/, 2017. Accessed: 2018-01-25.

[12] B. Lang. Investors poured a record $2 billion into VR/AR in 2017…, 2018. Accessed: 2018-01-25.

[13] H. Li, J. Yu, Y. Ye, and C. Bregler. Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph.*, 32(4):42:1–42:10, July 2013. doi: 10.1145/2461912.2462019

[14] Loom.ai. Loom.ai homepage. https://www.loomai.com/, 2017. Accessed: 2018-01-25.

[15] C. Malleson, M. Kosek, M. Klaudiny, I. Huerta, J. C. Bazin, A. Sorkine-Hornung, M. Mine, and K. Mitchell. Rapid one-shot acquisition of dynamic vr avatars. In *2017 IEEE Virtual Reality (VR)*, pp. 131–140, March 2017. doi: 10.1109/VR.2017.7892240

[16] R. McDonnell, M. Breidt, and H. H. Bülthoff. Render me real: Investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.*, 31(4):91:1–91:11, July 2012. doi: 10.1145/2185520.2185587

[17] B. J. Mohler, H. H. Bülthoff, W. B. Thompson, and S. H. Creem-Regehr. A full-body avatar improves egocentric distance judgments in an immersive virtual environment. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, APGV '08, pp. 194–197. ACM, New York, USA, 2008. doi: 10.1145/1394281.1394323

[18] E. Molla, H. G. Debarba, and R. Boulic. Egocentric mapping of body surface constraints. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2018. doi: 10.1109/TVCG.2017.2708083

[19] M. Mori, K. MacDorman, and N. Kageki. The Uncanny Valley [from the field]. *IEEE Robotics Automation Magazine*, 19(2):98–100, 2012. doi: 10.1109/MRA.2012.2192811

[20] Oculus VR. A closer look at oculus avatars, oculus first contact, and asynchronous spacewarp. https://www.oculus.com/blog/a-closer-look/, 2016. Accessed: 2018-01-25.

[21] L. Perry-Smith. Infinite-realities. http://ir-ltd.net/, 2018. Accessed: 2018-01-25.

[22] S. Saito, L. Wei, J. Fursund, L. Hu, C. Yang, R. Yu, K. Olszewski, S. Chen, I. Benavente, Y.-C. Chen, and H. Li. Pinscreen: 3d avatar from a single image. In *SIGGRAPH ASIA 2016 Emerging Technologies*, pp. 15:1–15:1. ACM, New York, USA, 2016. doi: 10.1145/2988240.3014572

[23] A. Shapiro, A. Feng, R. Wang, H. Li, M. Bolas, G. Medioni, and E. Suma. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds*, 25(3–4):201–211, 2014.

[24] Ten24 Media. T170 capture stage. http://ten24.info/3d-scanning/, 2018. Accessed: 2018-01-25.

[25] A. Weiss, D. Hirshberg, and M. J. Black. Home 3d body scans from noisy image and range data. In *2011 International Conference on Computer Vision*, pp. 1951–1958, Nov 2011. doi: 10.1109/ICCV.2011.6126465

[26] E. Zell, C. Aliaga, A. Jarabo, K. Zibrek, D. Gutierrez, R. McDonnell, and M. Botsch. To stylize or not to stylize?: The effect of shape and material stylization on the perception of computer-generated faces. *ACM Trans. Graph.*, 34(6):184:1–184:12, Oct. 2015. doi: 10.1145/2816795.2818126