

Digital Ira: Creating a Real-Time Photoreal Digital Actor

Oleg Alexander Graham Fyffe Jay Busch Xueming Yu Jorge Jimenez Etienne Danvoye Bernardo Antoniazzi
Ryosuke Ichikari Andrew Jones Paul Debevec* Mike Eheler Zybnek Kysela Javier von der Pahlen†
USC Institute for Creative Technologies Activision, Inc.



Figure 1: (Left) Three of eight high-res (0.1mm) light stage scans of the actor in static expressions. (Middle) Seven-camera HD performance recording. (Right) 180Hz video-driven blendshape model with screen-space subsurface scattering and advanced eye shading effects.

Overview In 2008, the “Digital Emily” project [Alexander et al. 2009] showed how a set of high-resolution facial expressions scanned in a light stage could be rigged into a real-time photoreal digital character and driven with video-based facial animation techniques. However, Digital Emily was rendered offline, involved just the front of the face, and was never seen in a tight closeup. In this collaboration between Activision and USC ICT shown at SIGGRAPH 2013’s *Real-Time Live* venue, we endeavoured to create a real-time, photoreal digital human character which could be seen from any viewpoint, in any lighting, and could perform realistically from video performance capture even in a tight closeup. In addition, we wanted this to run in a real-time game-ready production pipeline, ultimately achieving 180 frames per second for a full-screen character on a two-year old graphics card.

3D Scanning We began by scanning accommodating researcher Ari Shapiro in thirty high-resolution expressions using the USC ICT’s Light Stage X system [Ghosh et al. 2011], producing 0.1mm resolution geometry and 4K diffuse and specular reflectance maps per expression. We chose eight expressions for the real-time performance rendering, maximizing the variety of fine-scale skin deformation observed in the scans. The expressions were merged onto an artistically built back-of-the-head model. To record performances for the character, we shot seven views of 30fps video of the actor improvising lines using the same seven Canon 1Dx cameras used for the scans. We used a new tool called Vuvuzela to interactively and precisely correspond all expression texture (u,v) coordinates to the neutral expression, which was retopologized to a low-polygon clean artist mesh.

Performance Animation Our offline animation solver creates a performance graph from dense GPU optical flow between the video frames and the eight expressions. This graph gets pruned by analyzing the correlation between the video frames and the expression scans over twelve facial regions. The algorithm then computes dense optical flow and 3D triangulation yielding per-frame spatially varying blendshape weights approximating the performance.

The Game Rig To create the game-ready facial rig, we transferred the mesh animation to standard bone animation on a 4K polygon mesh using a bone weight and transform solver. The solver optimizes the smooth skinning weights and the bone animated trans-

forms to maximize the correspondence between the game mesh and the reference animated mesh.

Real-Time Rendering The rendering technique uses surface stress values to blend diffuse texture, specular, normal, and displacement maps from the different high-resolution expression scans per-vertex at run time. As a result, realistic wrinkles appear around the actor’s eyes when he squints and on his forehead when he raises his eyebrows; the color of the skin also changes with expression due to shifting blood content. The DirectX11 rendering takes into account light transport phenomena happening in the skin and eyes, from large scale events like the reflection of light of the own face into the eyes, to the shadowing and occlusion happening in the skin pores. In particular, it includes separable subsurface scattering [Jimenez et al. 2012] in screen-space, translucency, eye refraction and caustics, advanced shadow mapping and ambient occlusion, a physically-based two-lobe specular reflection with microstructure, depth of field, post effects, temporal antialiasing (SMAA T2x), and film grain.

Acknowledgements We thank Borom Tunwattanapong, Koki Nagano, Domi Piturro, Alejo von der Pahlen, Joe Alter, Curtis Beeson, Mark Daly, Mark Swain, Jen-Hsun Huang, Ari Shapiro, Valerie Dauphin, and Kathleen Haase for their important assistance and contributions to this work. This work was supported by USA RDECOM, USC, and Activision, Inc; no endorsement is implied.

References

- ALEXANDER, O., ROGERS, M., LAMBETH, W., CHIANG, M., AND DEBEVEC, P. 2009. Thedigital emily project: photoreal facial modeling and animation. In *ACM SIGGRAPH 2009 Courses*, ACM, New York, NY, USA, SIGGRAPH ’09, 12:1–12:15.
- GHOSH, A., FYFFE, G., TUNWATTANAPONG, B., BUSCH, J., YU, X., AND DEBEVEC, P. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graph.* 30, 6 (Dec.), 129:1–129:10.
- JIMENEZ, J., JARABO, A., GUTIERREZ, D., DANVOYE, E., AND VON DER PAHLEN, J. 2012. Separable subsurface scattering and photorealistic eyes rendering. In *ACM SIGGRAPH 2012 Courses*, ACM, New York, NY, USA, SIGGRAPH 2012.

*debevec@ict.usc.edu †Javier.Pahlen@activision.com