RE@CT: A NEW PRODUCTION PIPELINE FOR INTERACTIVE 3D CONTENT

F. Schweiger[1], G. Thomas[1], A. Sheikh[1], W. Paier[2], M. Kettern[2], P. Eisert[2], J.-S. Franco[3], M. Volino[4], P. Huang[4], J. Collomosse[4], A. Hilton[4], V. Jantet[5], P. Smyth[6]

[1]British Broadcasting Corporation, UK [2]Fraunhofer HHI, Germany [3]INRIA, France
[4]University of Surrey, UK [5]Artefacto, France [6]Oxford Metrics Group, UK

florian.schweiger@bbc.co.uk

ABSTRACT

The RE@CT project set out to revolutionise the production of realistic 3D characters for game-like applications and interactive video productions, and significantly reduce costs by developing an automated process to extract and represent animated characters from actor performance captured in a multi-camera studio. The key innovation is the development of methods for analysis and representation of 3D video to allow reuse for real-time interactive animation. This enables efficient authoring of interactive characters with video quality appearance and motion.

Index Terms— 3D video processing, interactive content, serious gaming, education

1. THE RE@CT PROJECT

RE@CT introduced a new production methodology to create film-quality interactive characters from 3D video capture of actor performance. Interactive in this context means that both the camera viewpoint and the 3D character's animation can be controlled live by the user.

While recent advances in graphics hardware have produced interactive video games with photo-realistic scenes, existing production pipelines for authoring animated characters with visual appeal and subtle details comparable to real actor performance as captured on film are still highly labour-intensive and involve significant costs.

RE@CT gathered a team of world leading researchers and companies in motion capture, 3D video, broadcast and interactive animation to develop a new production pipeline, including new temporal 3D matching methods and a new data representation for 3D action. The pipeline is entirely based on the video data captured by a multi-camera setup and does not require any additional marker tracking as used in traditional motion capture. This reduces the production effort tremendously, and allows actors to perform unimpeded, in a natural way, just as they would in a regular TV or film production. A textured 3D model is computed directly from the video data largely preserving nuances and producing very lifelike results. The actor's performance is segmented into logical motion units and all plausible transitions between these atomic movements are stored in a so-called 4D motion graph. In the rendering phase, it is possible to blend between existing motion units in real time allowing full control over the 3D character.

The project results are demonstrated in two application scenarios: an augmented reality application demonstrating usage for serious gaming in education and entertainment (see Fig. 5); and a production based on professional broadcast content demonstrating new synergies for developing a traditional programme and an interactive application in parallel (see Figs. 6 and 7).

	TV production	motion capture	RE@CT
inherently photo-realistic	✓	✗	✓
natural capture environment	✓	✗	✓
free-viewpoint	✗	✓	✓
live animation	✗	✓	✓

Table 1. Goals of the project in comparison with traditional approaches

2. DEVELOPED APPROACHES

The main stages in the RE@CT pipeline are depicted in Figure 1. In the following, the key achievements in the various areas are summarised, and pointers to more detailed publications given for the interested reader.

2.1. Data capture and full body reconstruction

In its latest iteration, the RE@CT studio setup comprises nine full HD cameras (1080p/25), mounted evenly-spaced on a rig approximately 280 cm above the ground, all around the capture space measuring eleven by six metres. In order to capture textures with greater detail, four additional UHD cameras (2160p/50) are positioned at chest height in the corners of the studio. All cameras are static, fixed-focus, and jointly calibrated from 3D-to-2D correspondences using a calibration object of known dimensions. A Python-based capture system records perfectly-synchronised video streams from the genlocked cameras. In order to facilitate foreground-background segmentation, the studio is equipped with retro-reflective blue screen technology.

From an actor's silhouettes, an octree-based algorithm computes the visual hull on a frame-per-frame basis and assigns texture coordinates to the resulting 3D mesh with respect to each of the cameras. The raw output of that step before temporal tracking (see 2.3) and smoothing is shown in Figure 1b.

As these computations run in real time, the 3D data can be used to automatically track an actor with an additional camera on a robotic pan-tilt-zoom head in order to obtain higher resolution textures, and

This project has been funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288369.

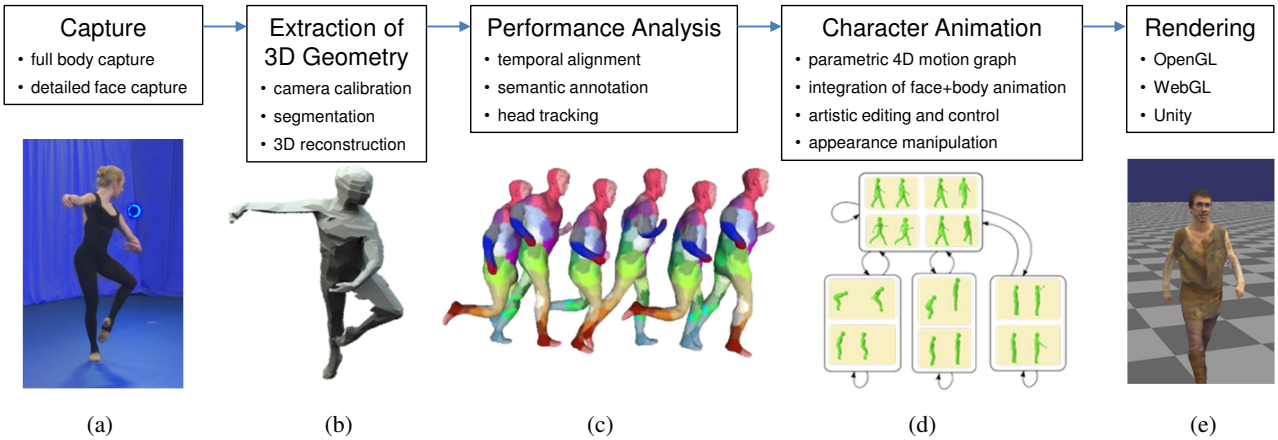

Fig. 1. The processing pipeline proposed by RE@CT (refer to text for explanations)

Fig. 2. Head model rendered with different facial expressions

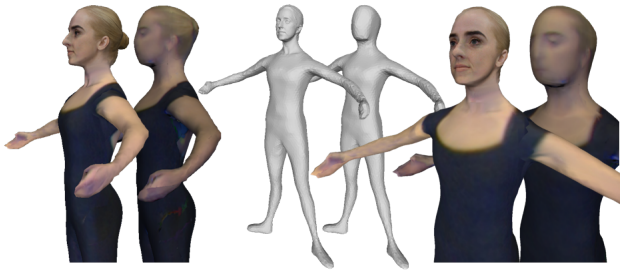

Fig. 3. Final hybrid mesh versus convex hull

to monitor the position of actors inside the capture volume triggering a warning whenever they come too close to its boundary.

More information about the full body capture and reconstruction system and can be found in [1] and in the public deliverables on the project website [2].

2.2. High-resolution face capture and hybrid mesh model

The convex hull approach works well on the body where missing concavities can be concealed by shadows in the texture map. For the face to be believably realistic, however, a greater level of detail is required. Different avenues to create high quality head models have been investigated, including a head-mounted capture system with and without markers [3], and a markerless 360° setup comprising seven DSLR stereo pairs [4].

In all cases, a static proxy mesh in neutral pose is created offline which is then enhanced with dynamic textures that are recorded either during the actual performance or in an entirely separate session – possibly using a different actor. To this end, the (original) actor's head is tracked throughout the sequence, fitting the proxy with seven degrees of freedom (six for its 3D pose and one for the jaw opening) using an analysis-by-synthesis approach [5]. Retargeting in 2D texture space ensures that the new texture stays exactly in place [6]. Not only does this approach allow arbitrary manipulation of the facial expression at any given time, its results are also highly photo-realistic (see Fig. 2).

The high-resolution head model and the convex body hull are then merged into a hybrid mesh. After various optimisation steps, including colour correction and a superresolution-based texture enhancement [7], this hybrid mesh is the basis for the subsequent performance analysis. Figure 3 shows an example in comparison with the initial convex hull.

2.3. Performance analysis and 4D motion graph

Unlike traditional motion capture, the approach described in Sections 2.1 and 2.2 yields an independent 3D mesh for every frame. In order to create a time-consistent geometry, a carefully selected reference mesh is segmented into rigid patches that are in turn tracked throughout the sequence [8, 9]. Figure 1c shows an excerpt from a tracked mesh sequence with individual patches, each highlighted in a different colour. Enforcing volumetric instead of surface constraints has proven to be more robust over an even wider range of actor motions [10].

In the temporally aligned, tracked mesh sequence every vertex is semantically linked to a body part, which enables a meaningful interpolation between different poses. Evaluating the transition costs between different meshes, the 4D motion graph containing all artefact-free transitions can be constructed automatically [11]. It is even possible to parametrise the intensity of a movement, such as the height of a jump or the length of a foot step, by interpolating between different recordings of the same motion [12, 13]. Figure 1d shows a parametric motion graph containing four nodes (walk, horizontal

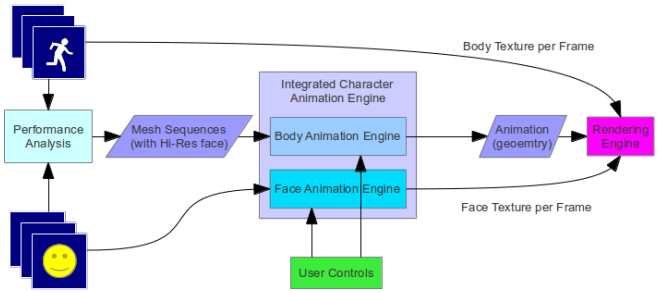

Fig. 4. Integrated body and face animation pipeline

jump, vertical jump, arm lift). Inside each node, several variations are depicted that can be continuously interpolated between (*e.g.*, straight walk slow/fast, 90° turn left/right). For even more animation flexibility that goes beyond the mixing of captured motion, hybrid skeletal-surface motion graphs can be used [14].

2.4. Character animation and rendering

The parametric motion graph described in Section 2.3, in combination with the temporally aligned mesh sequences, contains all the information necessary to link body motions together. For the face, transitions are accomplished by briefly reverting to a neutral expression in between.

Figure 4 shows the pipeline of the RE@CT animation system. The performance analysis component described in Section 2.3 provides the temporally-aligned hybrid mesh sequences with a high-resolution face. The body animation engine selects a path through the parametric motion graph and animates the character mesh accordingly. The facial animation engine provides the dynamic face textures which are applied together with the dynamic body textures in the final rendering step. Depending on the viewpoint, different texture layers are seamlessly blended using a novel 4D Video Texture technique to deliver a video-realistic result [15]. The user can control the body and face animation engines separately through a uniform interface.

The integration of 3D characters into an interactive application demands the efficient representation of free-viewpoint video texture for storage and streaming. This is especially true if the data is delivered over a network such as the Internet. A compact multi-view texture representation has been developed which re-samples the multi-camera appearance into a set of texture map layers for each frame in a captured sequence. This re-sampling is optimised to improve spatial and temporal coherence allowing compression algorithms to be applied to further reduce the required storage [16].

For less powerful devices, the computational complexity needs to be considered as well. The developed authoring tools allow subsets of motion sequences to be selected for which the animated mesh transitions are precomputed and uploaded to the device. This simplifies the rendering process turning it into a playback based on user input.

3. FINAL RESULTS

To showcase the different aspects of the project, several prototypes have been implemented and presented at various occasions. The two main demonstrators are a serious game developed in cooperation with the Museum of the Château de Montfort-sur-Meu in Brittany, France, and an educational online guide on the art of ballet in close collaboration with the producers of BBC iWonder.

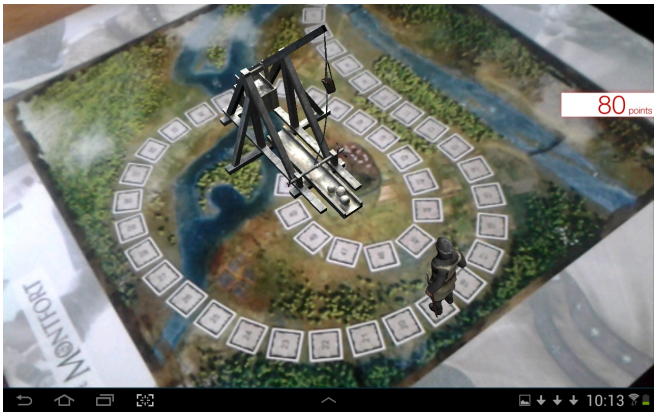

Fig. 5. Augmented reality board game

Les Seigneurs de Montfort, depicted in Figure 5, is an augmented reality game for tablet devices set in the Middle Ages. Built on the Unity engine, it renders 3D characters that were produced with RE@CT technology on top of a game board. The animation of the characters is not limited to moving across the board but also includes complex interactions with other 3D objects and between characters, such as sword fights. The game was launched in May 2013 and demonstrated publicly at several events including Mirage 2013 in Berlin.

The BBC iWonder guide on ballet comprises two RE@CT-powered elements. An interactive browser game, shown in Figure 6, lets users chain together five basic dance routines in arbitrary order, creating their own choreographies out of 120 possible combinations. RE@CT technology is used to blend seamlessly between the individual moves, creating an uninterrupted dance that is rendered in the browser using WebGL. This demo was first presented publicly at CVMP 2014 in London.

The second part went online in March 2015 as part of BBC Taster, a platform for new experimental and innovative content, and consists of the 3D-enhanced video player shown in Figure 7. It presents the user with a video of a ballet solo that freezes at predefined instants, allowing free-viewpoint exploration of the dancer's movements. Interactive annotations are displayed as 3D sprites giving further information about the dance routine. Transitions between video and the WebGL rendered 3D world are nearly seamless given the level of photorealism of the 3D character and the fact that video and 3D scene share the same virtual background. Drop-shadows rendered onto the virtual floor based on RE@CT data add to the realism of the composition.

All public deliverables, videos and more information are available from the project website [2]. Furthermore, parts of the RE@CT dataset are available for download at `http://cvssp.org/projects/4d/4DVT/`.

4. REFERENCES

[1] Jim Easterbrook, Oliver Grau, and Peter Schubel, "A system for distributed multi-camera capture and processing," in *Visual Media Production (CVMP), 2010 Conference on*. IEEE, 2010, pp. 107–113.

[2] "RE@CT," `http://react-project.eu`, official project website.

[3] Oliver Grau, Edmond Boyer, Pen Huang, David Knossow, Emilio Maggio, and David Schneider, "RE@CT – immersive production and delivery of interactive 3d content," in *NEM Summit-Networked and Electronic Media*, 2012.

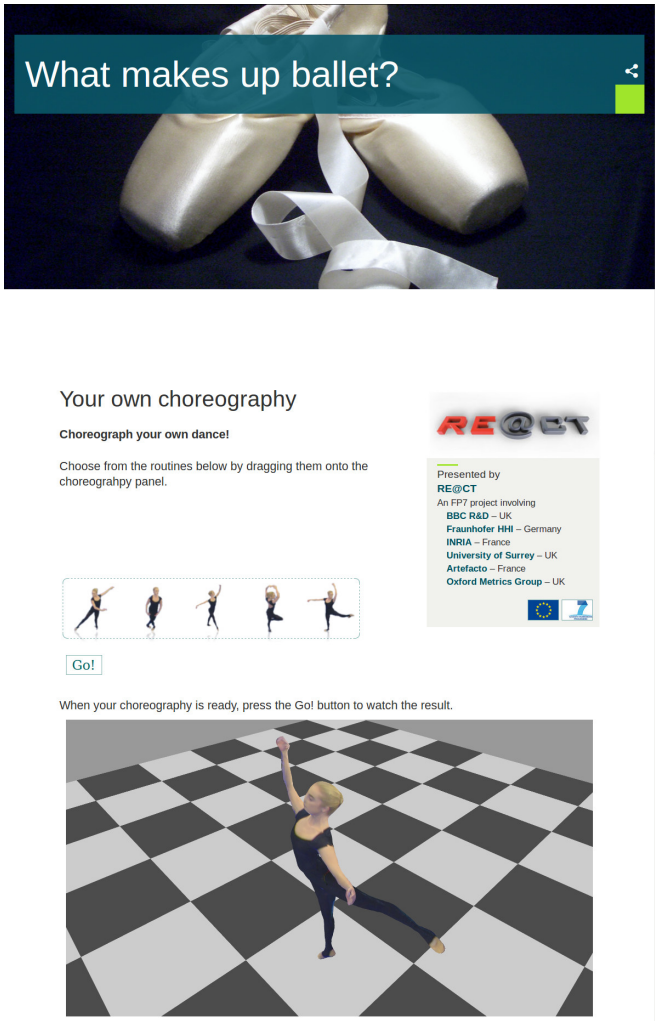

Fig. 6. Interactive 3D browser game

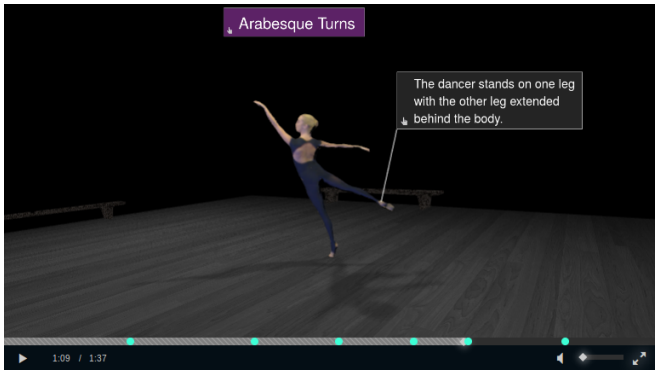

Fig. 7. 3D video augmented with interactive annotations

[4] David C Blumenthal-Barby and Peter Eisert, "High-resolution depth for binocular image-based modeling," *Computers & Graphics*, vol. 39, pp. 89–100, 2014.

[5] Markus Kettern, David Blumenthal-Barby, and Peter Eisert, "High detail flexible viewpoint facial video from monocular input using static geometric proxies," in *Proceedings of the 6th International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications*. 2013, MIRAGE '13, pp. 2:1–2:8, ACM.

[6] Wolfgang Paier, Markus Kettern, and Peter Eisert, "Realistic retargeting of facial video," in *Proceedings of the 11th European Conference on Visual Media Production*. ACM, 2014, p. 2.

[7] Vagia Tsiminaki, Jean-Sébastien Franco, and Edmond Boyer, "High resolution 3d shape texture from multiple videos," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1502–1509.

[8] Chris Budd, Peng Huang, Martin Klaudiny, and Adrian Hilton, "Global non-rigid alignment of surface sequences," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 256–270, 2013.

[9] Benjamin Allain, Jean-Sébastien Franco, Edmond Boyer, and Tony Tung, "On mean pose and variability of 3d deformable models," in *Computer Vision–ECCV 2014*, pp. 284–297. Springer, 2014.

[10] Benjamin Allain, Jean-Sébastien Franco, and Edmond Boyer, "An efficient volumetric framework for shape tracking," in *CVPR 2015-IEEE International Conference on Computer Vision and Pattern Recognition*.

[11] Peng Huang, Adrian Hilton, and Jonathan Starck, "Human motion synthesis from 3d video," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1478–1485.

[12] Dan Casas, Margara Tejera, J Guillemaut, and Adrian Hilton, "Interactive animation of 4d performance capture," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 5, pp. 762–773, 2013.

[13] Dan Casas, Margara Tejera, Jean-Yves Guillemaut, and Adrian Hilton, "4d parametric motion graphs for interactive animation," in *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. ACM, 2012, pp. 103–110.

[14] Peng Huang, Margara Tejera, John Collomosse, and Adrian Hilton, "Hybrid skeletal-surface motion graphs for character animation from 4d performance capture," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 2, pp. 17, 2015.

[15] Dan Casas, Marco Volino, John Collomosse, and Adrian Hilton, "4d video textures for interactive character appearance," in *Computer Graphics Forum (Proc. Eurographics 2014)*. Wiley Online Library, 2014, vol. 33, pp. 371–380.

[16] Marco Volino, Dan Casas, John Collomosse, and Adrian Hilton, "Optimal representation of multi-view video," in *British Machine Vision Conference (BMVC)*, Copyright 2014 The Authors.

[17] Antoine Letouzey and Edmond Boyer, "Progressive shape models," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 190–197.

[18] Jean-Sébastien Franco, Benjamin Petit, and Edmond Boyer, "3d shape cropping," in *Vision, Modeling and Visualization*. Eurographics Association, 2013, pp. 65–72.

[19] Abdelaziz Djelouah, Jean-Sébastien Franco, Edmond Boyer, François Le Clerc, and Patrick Pérez, "Multi-view object segmentation in space and time," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2640–2647.

[20] Tinghuai Wang, John Collomosse, and Adrian Hilton, "Wide baseline multi-view video matting using a hybrid markov random field," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 136–141.

[21] Guosheng Hu, Chi Ho Chan, Fei Yan, William Christmas, and Josef Kittler, "Robust face recognition by an albedo based 3d morphable model," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. IEEE, 2014, pp. 1–8.

[22] Alexandros Neophytou and Adrian Hilton, "A layered model of human body and garment deformation," in *3D Vision (3DV), 2014 2nd International Conference on*. IEEE, 2014, pp. 171–178.