Directable Stadium Crowds from Image Based Modelling for "Bohemian Rhapsody"

Ted Waine
DNEG
ted@dneg.com

May Leung
DNEG
myl@dneg.com

Paul Norris
DNEG
pdn@dneg.com

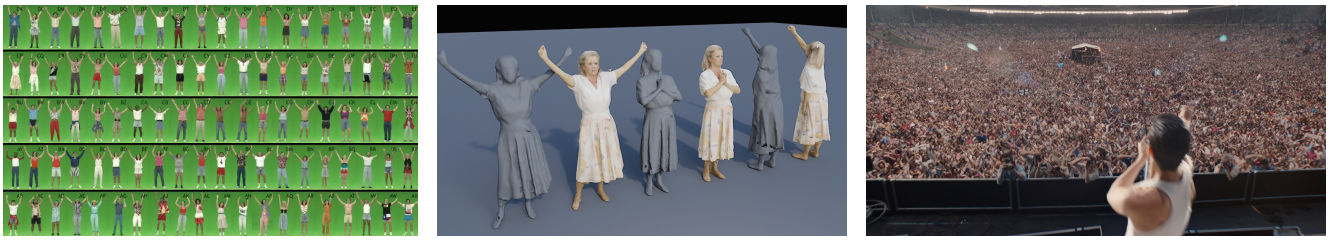

Figure 1: *Contact sheet of performance capture takes, Gray shaded and textured 3D sprites, The Wembley crowd responding to Freddie Mercury*

ABSTRACT

To deliver photoreal, dynamic and directable rock concert audiences for "Bohemian Rhapsody" to a demanding client brief, lead VFX vendor DNEG developed a novel crowd simulation solution based on multi-view video capture and image based modeling. Over 350 choreographed performances by individual crowd extras, totalling more than 70 hours of footage, was acquired on set using a video camera array. A system was developed to convert the video data to lightweight 3D sprites that could be quickly laid out, synchronised, edited and rendered on a large scale. Efficient artist workflow tools and scalable video processing technology was developed so that crew with little previous experience in crowd simulation could fill a virtual Wembley Stadium with a dancing, cheering crowd responding to Freddie Mercury's electrifying performance.

CCS CONCEPTS

• **Computing methodologies** → **Computer graphics**; *Computer vision*; *3D imaging*; *Image processing*; *Motion capture*; • **Applied computing** → **Media arts**.

KEYWORDS

image based modelling, performance capture, crowd simulation

ACM Reference Format:
Ted Waine, May Leung, and Paul Norris. 2019. Directable Stadium Crowds from Image Based Modelling for "Bohemian Rhapsody". In *Proceedings of SIGGRAPH '19 Talks.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3306307.3328170

SIGGRAPH '19 Talks, July 28 - August 01, 2019, Los Angeles, CA, USA

ACM ISBN 978-1-4503-6317-4/19/07.
https://doi.org/10.1145/3306307.3328170

1 INTRODUCTION

The award winning feature film 'Bohemian Rhapsody' culminates in a 15 minute sequence recreating one of rock's most iconic live performances: Queen's set at the 1985 Wembley Stadium Live Aid concert. The stage itself was carefully reconstructed for the production but creating the stadium, which has since been demolished, and the crowd within was an entirely VFX effort. The client insisted on a photography based solution as they felt that established crowd technologies, typically using fully digital animated crowd agents driven by AI logic networks, would not quite deliver the energy of a stadium audience critical to the success of the sequence. Sweeping camera moves and a requirement for lighting decisions in post precluded a simple approach of projecting video elements of crowd members onto cards. A very restricted timeframe for specifying a solution precluded significant R&D for novel capture hardware, however. The performance capture system would need to be based on tried and tested technology that film crew professionals would be comfortable with. DNEG therefore proposed a solution employing image based modeling techniques: a multi-camera rig would capture actors' performances choreographed to an audio edit of Queen hits. Each take could then be converted using dense stereo matching into a lightweight 3D poly-mesh that could be scattered and instanced through the Wembley model. An image based modeling approach has already been presented by DNEG for in-camera capture of a single hero performance for a short action sequence[Waine and James 2009]. For this project, however, a completely new data pipeline and artist workflow was required to scale the techique to deliver 100,000 'directable' crowd agents for a 15 minute sequence with a shot count pushing into the hundreds.

2 DATA ACQUISITION

Figure 1: The capture stage as seen through the 6 video cameras

An array of 6 networked Arri AlexaXT cameras was rigged around a greenscreen stage built near to the main unit shoot for the Wembley sequence. Camera positions were carefully designed to provide maximum coverage and to allow the extraction of dense stereo depth information and the cameras were programmed to shoot 48fps to cover requirements for slo-mo shots. A guide track composed of key verse and chorus sections of the Queen songs that might feature was painstakingly edited to ensure that the full variety of crowd actions could be captured whilst also minimising the take duration. Voiceover instructions were added prompting performers to carry out specific actions required for each song. Crowd extras were called to the stage one-by-one where they performed the choreographed routine in time to the guide track.

3 IMAGE BASED MODELLING

Image based modeling is already widely employed in VFX for 'scanning' static assets and characters using multi-viewpoint photography. Popular products like RealityCapture and PhotoScan which are used for this purpose rely on dozens and sometimes hundreds of closely overlapping image viewpoints to solve camera positions and generate dense feature matches. For this project DNEG needed completely bespoke software that could surface-model moving human subjects from only 6 camera viewpoints and scale efficiently to convert many hundreds of thousands of frames of video into lightweight animated 3D sprites on a PC cluster. The solution relied on a separate calibration stage to solve camera positions from video of a reference object captured during the shoot. The video could then be procedurally keyed and de-spilled before a depth matching algorithm was applied to the resulting stereo video pairs to form a surface point cloud. The point cloud would then be converted to a surface mesh using the Poisson method[Kazhdan et al. 2006]. Further processing steps carved away any garbage geometry, reduced noise and reduced the poly count in the mesh. Finally, de-spilled texture data was projected onto polygon faces before writing to an Alembic cache.

4 LAYOUT TOOLS

Computing the 3D sprite was both CPU and IO intensive and front-ending this process in the production schedule would have drawn on resources very heavily. Since it was likely that only some fraction of the entire video data set needed to be converted to sprites for final renders it was decided that the full 'extraction' of the 3D sprites from the video should be done on-demand instead: the layout team were provided with very lightweight 2D proxy sprites, generated from a single camera view, that would be suitable for developing initial crowd scatters. Only following layout approval would an extraction task be initiated.

Figure 2: Very low-res 2D sprites could be used for layout tasks (left), before 'extracting' the full 3D sprite (right)

Since each performance was done to a common audio track every capture could be synchronised to provide layout artists with a library of individual sprites doing specific actions. By composing the crowd of a mix of selected actions this allowed a certain 'directability' of the crowd's behaviour even though every peformance was, of course, captured in-camera. Our solution allowed less experienced artists to develop crowd layouts using the designated lighting and rendering tool, Isotropix Clarisse, instead of being done by experienced crowd TDs in 3D animation and effects simulation tools like Houdini or Maya. Supervisors built template node networks leveraging Clarisse's native scattering tools while a custom sprite importer was designed that would allow artists to select an action from the performance capture sessions and synchronise it to the audio track. By adjusting parameters on control nodes in the network, the artist could change the mix of different actions within the crowd to deliver a mix of synchronised and variated motion for a convincing look. Further refinements could be made to apply a randomised hue shift to texture data to mitigate obvious duplication of certain sprites with more noticeable, highly coloured costumes. Time delays were also introduced to simulate the effect of the speed of sound as the audience visibly 'rippled' when clapping in time to the band's music.

5 CONCLUSION

Departing from established crowd simulation technology for a new system based on video capture was a major technical challenge for the DNEG crew as they sought to create convincing stadium crowds for 'Bohemian Rhapsody'. This novel approach, however, allowed DNEG to deliver to the client's demanding spec within a tight budget and with a final result that gave the movie's key concert sequences a realistic, dynamic and responsive crowd opposite Freddie Mercury and the band on stage.

REFERENCES

Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. 2006. Poisson Surface Reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing (SGP '06)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 61–70. http://dl.acm.org/citation.cfm?id=1281957.1281965

Ted Waine and Oliver James. 2009. Dense Stereo Event Capture for James Bond, Quantum of Solace. In *SIGGRAPH 2009: Talks (SIGGRAPH '09)*. ACM, New York, NY, USA, Article 27, 1 pages. https://doi.org/10.1145/1597990.1598017