Real-time Full-Body Motion Capture from Video and IMUs
Charles Malleson, Marco Volino, Andrew Gilbert, Matthew Trumble, John Collomosse and Adrian Hilton
CVSSP, University of Surrey, Guildford, U.K.

Overview
A real-time full-body motion capture system is presented which uses input from a sparse set of inertial measurement units (IMUs) along with images from two or more standard video cameras. A real-time optimization framework incorporates constraints from the IMUs, cameras, and a prior pose model. The combination of video and IMU data allows the full 6-DOF motion to be recovered including axial rotation of limbs and drift-free global position. Tests on indoor and outdoor captured data show the effectiveness of the approach for tracking a wide range of human motion in real-time.

Motivation
- Applications in entertainment (film, TV, games, VR, AR), life sciences
- Real-time full-3D kinematic motion capture with low encumbrance, flexible capture configurations
- Overcoming limitations of previous methods

Flexible input configuration
The method can trade accuracy against capture hardware and setup complexity (number of cameras and IMUs) as well as framerate. In the experiments, IMU input was from Xsens MTWs [13], while 2D joint detections are obtained from a state-of-the-art convolutional pose machine (CPM) detector [19].

Frame packing for increased detection throughput
The 2D joint detection is a bottleneck (requiring > 150 ms per image on our system). In order to maintain video rate operation while detecting on multiple cameras, we pack regions of interest from several input images into a single image for detection and resolve the results to their originating frame and camera.

Results
Indoor capture
The total capture indoor motion capture dataset [14] has 360 camera coverage and 13 IMUs. It also contains a marker based ground truth reference motion for evaluation.

Outdoor capture
A new outdoor dataset Outdoor 1 was captured with cameras in a 120 degree arc along with 13 IMUs.

Conclusion
This work demonstrates the effectiveness of the approach for real-time full-body motion capture with a wide range of applications. Future work will focus on improving the accuracy and robustness of the method, particularly in challenging environments such as outdoor capture with a larger number of cameras and IMUs.

References

Confidentiality Statement
This manuscript is based on the work presented at the 3DV conference in 2017. The authors wish to maintain the confidentiality of certain aspects of the research to ensure the protection of intellectual property and proprietary information. The details of the implementation, algorithms, and specific datasets used in the experiments are proprietary and will not be disclosed in this abstract. The authors confirm that the research was conducted in accordance with ethical guidelines and that the results are based on original data.

Acknowledgements
The authors would like to thank the team at CVSSP and the University of Surrey for their support and encouragement. This work was supported by the Innovate UK Total Capture project (grant 102685) and in part by the EU H2020 Visual Media project (grant 687800). We wish to acknowledge the contributions of all team members, especially for their hard work and dedication to the success of the project.