BMVA technical meeting: Dynamic Scene Reconstruction

June 21, 2017

Michael Zollhöfer started proceedings, presenting the achievements of his group at the Max Planck Institute for Informatics in real-time reconstruction, making the case for the importance of real-time performance for VR/AR applications. This included the parameterised face model behind the famous face2face demo featuring George W. Bush and others, enabling real-time reconstruction and re-targeting from monocular video. Many other impressive results were also presented including hand/object tracking from RGB-D, reconstruction of non-rigid deformable objects, and large scale scene reconstruction avoiding loop closure issues, again all in real-time. Michael introduced Opt, a programming language developed by the group specifically for these kind of optimizations, which allows energy minimisation problems to be simply defined, and compiled to produce high performance GPU enabled code. This enables rapid development and iteration of research prototypes.

Christian Richardt, from the University of Bath, showed work recently published at the International Conference on 3D Vision (3DV) performing dynamic 3D scene reconstruction from just 2 handheld video cameras. They achieve impressive results from such challenging input sources through innovation to the scene flow refinement process, leveraging edge information in the source images to overcome occlusions.

Armin Mustafa, from the University of Surrey, presented an approach to semantic 4D reconstruction of dynamic scenes, to be published in The Conference on Computer Vision and Pattern Recognition (CVPR) this year. She introduced concept of semantic tracklets to enable co-segmentation and reconstruction of dynamic objects, without the need for a priori models. The technique was demonstrated on indoor and outdoor sequences, including challenging human motion, achieving impressively complete, semantically and temporally coherent reconstruction of the scene and its dynamic contents.

Jonathan Starck, Head of Research at Foundry, provided a valuable industry perspective. He introduced Nuke, an internally developed compositing tool for VFX artists that is capable of single view scene reconstruction. Nuke uses a node based image processing pipeline that allows a team of artists and engineers to work concurrently on different parts of a complex task. Cutting edge computer vision methods can be implemented by his team and made available to artists within Nuke as new nodes ready to be added to a processing pipeline. Jonathan also talked about the continuing industry excitement around VR productions. He discussed the production techniques of 360 video and its limitations in providing an immersive VR experience when compared to true positional VR. It is much more challenging however to use live action footage in a positional VR experience, requiring highly accurate and detailed reconstruction of real scenes and performers into a rendered 3D world. He discussed some of his experiences working with Lytro, developers of the first commercial light field camera, and on the Innovate UK project ALIVE, in collaboration with the University of Surrey and Figment Productions. ALIVE is an exciting looking project aiming to address these challenges and produce an immersive cinematic VR experience.

Nadejda Roubtsova, from the University of Bath, presented a highly sophisticated reflectance modelling technique recently published in the International Journal of Computer Vision (IJCV). The technique is an advance of white-light helmholtz stereopsis to allow the reconstruction of dynamic scenes. In their experimental setup, coloured light sources are employed to capture multiple helmholtz stereopsis pairs simultaneously over a dynamic sequence. These are processed in a novel wavelength multiplexing technique to effectively decouple the geometry and reflectance properties of the target object.

Lourdes Agapito, of University College London, started the afternoon session, noting the progress in the field since the technical meeting on dynamic reconstruction she held 5 years ago. She has observed the development of increasingly sophisticated and robust reconstruction techniques. Some of her latest work, for example, incorporates reflectance modelling in non-rigid surface reconstruction to account for specular highlights. The general focus of Lourdes' work has been 3D reconstruction from monocular video, without the advantages of multiple cameras, fiducial markers or prior models. More recent work has started to include models of different forms however, such as work appearing at CVPR later this year which uses a distribution of human pose data from mocap to inform a deep learning framework predicting 3D human pose from a single image. Lourdes also presented work that appeared recently at the International Conference on Robotics and Automation (ICRA) in which they have developed a dynamic SLAM system for robotics which uses semantic and motion information from a single RGB-D camera to track, segment and reconstruct a scene containing multiple dynamic objects.

Benjamin Biggs, from the University of Cambridge, presented his early PhD work developing a novel deep-learning system for animal skeletal tracking. Rather than learning body-part labelling from images, as CNNs are often employed to do, Ben's system aims to learn skinning weights for a generic quadruped model. This would allow the full body shape of an animal to be generated, which would be of much more use than just a skeleton, for example in veterinary analysis.

Dan Casas, of the Universidad Rey Juan Carlos, then presented his work focussing on the challenge of outdoor human performance capture, which appeared at 3DV last year. Their method starts by fitting a volumetric colour model, based on a set of 3D Gaussians, to the source RGB images to provide a coarse skeletal position estimate. A full mesh based model of the subject is then more closely fitted to the image data, again employing colour-based gaussian distribution applied to each vertex. Points that appear on the edge of the model from each camera view are given special processing to allow the vertices to align precisely with the actor's extent, overcoming the need for explicit background segmentation, which can be difficult to obtain in unconstrained, outdoor environments.

Rilwan Basaru, of City University, presented his PhD work towards hand depth estimation from stereo images. He gave a brief introduction to the challenges specific to hand reconstruction - the capturing of fast, complex non-rigid motion and the relatively textureless appearance of skin. His approach is to build a model-mapping between depth and stereo using a conditional regressive random forest, a combination of conditional random field and random forest methods.

Stamatia Giannarou, from Imperial College London, was another speaker who presented at the preceding dynamic reconstruction technical meeting. Stamatia presented her work in medical imaging, identifying the extent of tumour tissue during surgery. Her work focuses on the field of neurosurgical oncology and is supported by the NEURO grant, aiming to develop a surgical vision platform capable of robotic mapping of tissue surface to aid tumour resection surgery.

The meeting concluded with a panel discussion by the keynote speakers. They observed the maturing of the field in the 5 years since its first technical meeting. We are now able to achieve dense, 3D reconstructions with methods that can generalise well and perform in real-time, helped greatly by crossovers into the graphics and optimisation fields. The availability of cheap scanners, such as Kinect, and code libraries and datasets has notably accelerated research. We have also witnessed the rapid democratisation of techniques from cutting edge to common place, Jonathan Starck noting that static reconstruction is now used every day in the VFX industry. Collaborative projects, such as those sponsored by Innovate UK, that bring businesses and academia together to bring innovation to industry help greatly in this process. We are also seeing the spread of technology outside of the performance capture/VFX industries, as demonstrated by Benjamin and Stamatia's work.

Looking to the future, the panel foresee the likely benefits of applying the power of deep learning to model tracking and correspondence regression tasks, and accordingly the need for more challenging datasets, but acknowledge the difficulty of ground truth generation that comes with that. With the advancement of scene understanding, Lourdes also noted the converging of vision and robotics challenges, and anticipates the benefits of collaborative efforts tackling issues such as semantic modelling and perception of scene interactions.

It was a fascinating meeting, bringing together a stellar cast of leaders in the field, able to provide a comprehensive overview of the state of the art and a historical perspective of how far the field has come, plus some thought-provoking insights of future possibilities!