

Virtual Headcam: Pan/tilt Mirror-based Facial Performance Tracking

Xueming Yu, Shanhe Wang, Jay Busch, Thai Phan, Tracy McSheery², Mark Bolas, Paul Debevec
USC Institute for Creative Technologies ²PhaseSpace Inc.



Introduction High-end facial performance capture solutions typically use head-mounted camera systems which provide one or more close-up video streams of each actor's performance. These provide clear views of each actor's performance, but can be bulky, uncomfortable, get in the way of sight lines, and prevent actors from getting close to each other. To address this, we propose a virtual head-mounted camera system: an array of cameras placed around the performance capture volume which automatically track zoomed-in, sharply focussed, high-resolution views of the each actor's face from a multitude of directions. The resulting imagery can be used in conjunction with body motion capture data to derive nuanced facial performances without head-mounted cameras.

Virtual Headcam: Pan/tilt Mirror-based Facial Performance Tracking

Xueming Yu, Shanhe Wang, Jay Busch, Thai Phan, Tracy McSheery², Mark Bolas, Paul Debevec
USC Institute for Creative Technologies ²PhaseSpace Inc.



Figure 1: (Left) A stationary camera with an instrumented zoom & focus lens attached to motorized pan/tilt mirrors. (Middle) Two units of our virtual head-mount camera setup in a performance capture stage for facial tracking. (Right) Frames from a tracked facial performance.

Introduction High-end facial performance capture solutions typically use head-mounted camera systems which provide one or more close-up video streams of each actor’s performance. These provide clear views of each actor’s performance, but can be bulky, uncomfortable, get in the way of sight lines, and prevent actors from getting close to each other. To address this, we propose a virtual head-mounted camera system: an array of cameras placed around the performance capture volume which automatically track zoomed-in, sharply focussed, high-resolution views of the each actor’s face from a multitude of directions. The resulting imagery can be used in conjunction with body motion capture data to derive nuanced facial performances without head-mounted cameras.

System Design Our approach keeps the weighty camera and zoom lens in a fixed place and frames the view of the actor through a two-axis mirror pan/tilt apparatus as in [Okumura et al. 2011], and adds remote zoom & focus control of the lens. The light weight of the glass mirrors eliminates the need for heavy mechanics, and dramatically improves the system response time and accuracy so that cameras can be quickly retasked to different actors as they move within the performance volume. Unlike [Okumura et al. 2011], we design our system to work with fast off-the-shelf video lenses with a large exit pupil, and we avoid expensive mirror galvanometers. Instead, we adapt brushless gimbal motors (Fig. 2) in the 36N42P (N for stator arm; P for magnet pole) configuration to drive mirrors, providing smooth rotation as well as constant torque over variant speed. A 10K-count optical encoder is used for motor position feedback. With this setup, we are able to achieve rotary resolution as accurate as 0.009° for the mirror.

We use Fujinon-TV.Z 7.5mm-105mm lenses to cover a subject’s movement inside a performance capture stage. This lens includes on-board gear & motor box for remote control of zoom and focus. We designed a circuit board (Fig. 2) with a networking interface and an on-board micro-controller to control the lens box and the gimbal motors. 3D printed parts and 80/20TM aluminum framing modules are used for the mechanical construction. We use Pt. Grey Flea3 USB3 cameras running at 1280×1024 pixels at 60fps to record

our capture results.

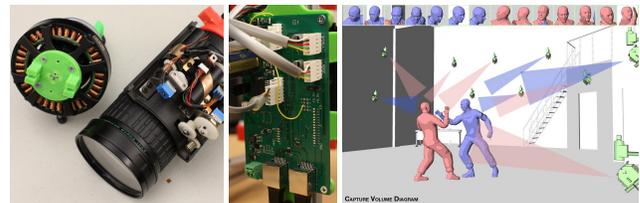


Figure 2: (Left) 36N42P gimbal motor and zoom lens mounted with gear & motor box. (Middle) Circuit boards for motor drivers and networking. (Right) System diagram illustrating multiple camera units tracking multiple subjects.

Capture Volume Integration We install the tracking cameras around the volume of a PhaseSpace motion capture system, and use 120Hz position and orientation data of lightweight active tracking markers on top of the head (out of view of the actor) to calculate the proper mirror angles and zoom/focus positions to aim each camera at each actor (Fig. 2). The resulting view of each camera is firmly locked on the face of the actor with tight framing, allowing nuanced performances to be seen. The use of active focus allows for a wide lens aperture, minimizing the requirements for ambient light. The cameras in each unit stream the image data to storage media. We can track two actors simultaneously and switch which camera is looking to each actor in less than 0.1s. This will allow multiple cameras to record performances of multiple actors, dynamically switching between faces as actors turn around and walk in front of each other to ensure that each actor’s face is seen from multiple angles throughout their performance.

Ongoing Work We are developing 3D reconstruction techniques to convert the multi-view video data into 3D models of dynamic facial performances. We are constructing additional facial tracking cameras in order to provide sufficient views. More units will be built in order to yield a robust input data into the software pipeline for successful 3D reconstruction.

References

OKUMURA, K., OKU, H., AND ISHIKAWA, M. 2011. High-speed gaze controller for millisecond-order pan/tilt camera. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 6186–6191.