

PhotoSpace: A Vision Based Approach for Digitizing Props

Pravin Bhat*
Weta Digital

Sebastian Burke†
Weta Digital

Keywords: photogrammetry, turntable, modeling, texturing

1 Introduction and Motivation



Figure 1: Left image shows one of the 270 photographs of a physical prop captured using a computer-controlled turntable. The prop geometry (middle) and texture (right) are extracted automatically from the photographs using computer vision techniques and serve as 3D reference for artists creating the production asset.

This talk will concentrate on how some of the recent advances in computer graphics and vision fit into a visual effects pipeline and their application for digitizing props. The production of movies often requires digitizing physical properties of various props. At Weta Digital we model, texture, and shade thousands of props in immense detail. We present a vision-based pipeline for digitizing props that requires far less user input than previous methods. This vision based pipeline fully automates the process of creating virtual cameras, reference geometry, and texture maps, which is then used by our artists to create the final assets at a much faster rate.

Our previous in-house approach for digitizing props began with taking reference photographs and creating a Maya scene with virtual cameras that match the relative poses at which the physical camera photographed the prop. With a few hours of user input a TD would generate virtual cameras corresponding to individual photographs using interactive matchmove software. These virtual cameras were then exported to a Maya scene with the corresponding photographs attached as image-planes. This served as a multiview image reference for the artist when modelling the physical prop. Once a model was created, a 3D paint package was used to semi-automatically create a seamless texture map using a painting interface to project regions from various photographs through corresponding virtual cameras.

At Weta we also have a laser scanner to scan the geometry of a prop. However, this approach does not provide photo-reference aligned to the 3D model, which accelerates the creation of prop texture and shader. In addition, complex props can require several hours of laser guidance by a human operator to get thorough coverage. Although the dense mesh generated by a laser scanner is unsuitable for our rendering pipeline, the mesh serves as an excellent 3D reference for the modeler and considerably speeds up the modelling process.

Our vision based pipeline for digitizing props combines the best features of the approaches described above but requires far less user input. The capturing of photographs and extraction of virtual cameras, geometry, and texture is automated. The resulting data is provided as reference material inside software packages familiar to artists, thus considerably speeding up the creation of the final asset.

*e-mail: pbhat@wetafx.co.nz

†e-mail: sburke@wetafx.co.nz

2 System overview and future work

Our vision-based system for digitizing props consists of three parts that are built using readily available components.

Capture session: The prop is placed on a turntable in a studio environment. A microprocessor triggers a camera to take photographs of the prop at regular intervals as it turns the prop 360 degrees. By default the camera captures ninety photographs for one revolution of the turntable. To ensure vertical coverage of the prop, the controller can use one to three cameras at different vertical heights that photograph the prop at different vertical angles. For a single prop, the capture session runs to completion within fifteen minutes.

Photogrammetry: Once the photographs have been captured, several vision based tasks are launched on our server farm. This usually takes 2-10 hours depending on the number of photographs captured (90-270) and the complexity of the prop. For each photograph, feature points are detected and feature descriptors are extracted. Each photograph is matched to its two nearest horizontal neighbors and up to two nearest vertical neighbors, thus generating 2D matches across the entire image graph. These 2D correspondences are then processed using Snavely et al. [2010]’s structure-from-motion algorithm to extract camera parameters for each photograph. The camera parameters are then used in Furukawa et al. [2010]’s multiview stereo algorithm to extract the prop geometry as a dense pointcloud of oriented 3D points. The pointcloud is then converted into a water tight mesh using Kazhdan et al. [2007]’s octree-based Poisson reconstruction algorithm.

Reference generation: The mesh generated in the previous step is unsuitable to be used directly in our rendering pipeline. Instead the system generates a reference Maya scene that helps a human modeller speed up the modelling process. The Maya scene contains the mesh and cameras extracted in the previous step, centered at the origin, and oriented to be upright. Each photograph is attached as an image-plane to the corresponding camera in the Maya scene. A similar scene is created for our in-house 3D paint package that allows a texture artist to paint through any of the photographs thus projecting the selected region through the corresponding camera. An initial seamless texture map is automatically generated using the method described in Bhat et al. [2007]. Any residual error in the texture map is eliminated by the texture artist using a painting interface.

Future work: We hope to extend the system to also generate reference shaders for props. During the capture session we plan to use a lightstage to efficiently vary the scene lighting in a manner that will facilitate automatic extraction of the prop’s shading properties.

References

- BHAT, P., ZITNICK, C. L., SNAVELY, N., AGARWALA, A., AGRAWALA, M., CURLESS, B., COHEN, M., AND KANG, S. B. 2007. Using photographs to enhance videos of a static scene. In *EGRS 2007*.
- FURUKAWA, Y., AND PONCE, J., 2010. Patch-based multi-view stereo software. <http://grail.cs.washington.edu/software/pmvs>.
- KAZHDAN, M. M., 2007. Unconstrained isosurface extraction on arbitrary octrees. <http://www.cs.jhu.edu/~misha/Code/IsoOctree/>.
- SNAVELY, N., 2010. Bundler: Structure from motion for unordered image collections. <http://phototour.cs.washington.edu/bundler/>.