

Research Summary (Fall 2002 - Spring 2006)

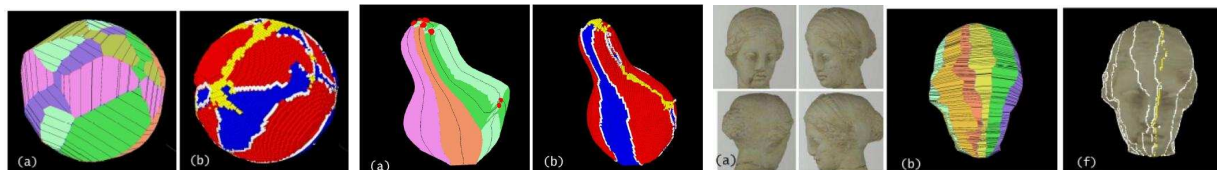
Sudipta N. Sinha, Computer Science, UNC-Chapel Hill

February 06, 2006

Vision Algorithms implemented on Graphics Hardware

My 2005 summer internship at Siemens Corporate Research (SCR), Princeton involved developing a fast GPU-based implementation of the Scale Invariant Feature Transform (SIFT) [Lowe 2004] algorithm for extracting scale invariant features from images. This is useful in various computer vision tasks such as object detection, image indexing, object recognition and structure from motion. The implementation was based on OpenGL/Cg and provided a 6-12X speedup at reasonable image resolutions on a Nvidia Geforce 7800 Gtx. As part of Urbanscape (large scale urban 3D modeling project), I am implementing low-level vision tasks, real-time KLT Feature Tracking in High-resolution Video Streams on programmable graphics hardware.

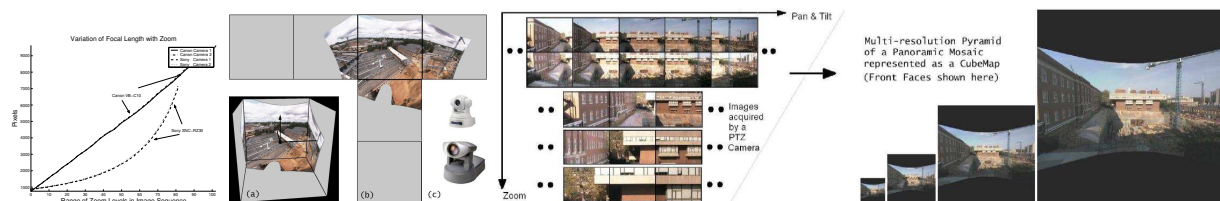
Combining Photoconsistency and Exact Silhouette Constraints



A novel approach for reconstructing a closed continuous surface of an object from multiple calibrated, segmented, color images was developed. An accurate reconstruction satisfies (1) photo-consistency and (2) silhouette consistency constraints. Most existing techniques treat these cues identically in optimization frameworks where silhouette constraints are traded off against photo-consistency and smoothness priors. Our approach strictly enforces silhouette constraints, while optimizing photo-consistency and smoothness in a global graph-cut framework. We transform the reconstruction problem into computing max-flow / min-cut in a geometric graph, where any cut corresponds to a surface satisfying exact silhouette constraints (its silhouettes should exactly coincide with those of the visual hull); a minimum cut is the most photo-consistent surface amongst them. Our graph-cut formulation is based on the rim mesh, (the combinatorial arrangement of rims or contour generators from many views) which can be computed directly from the silhouettes.

- S.N. Sinha and M. Pollefeys, *Multi-view Reconstruction using Photo-consistency and Exact Silhouette Constraints: A Maximum-Flow Formulation*, International Conference on Computer Vision (ICCV), October 2005, Beijing, China.

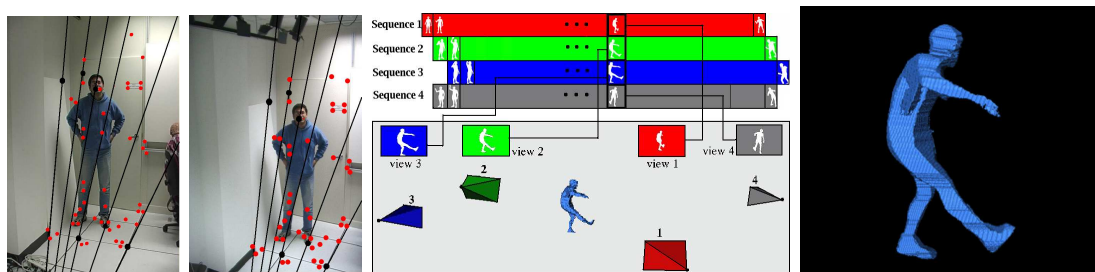
PTZ Camera Calibration and High-resolution Mosaics



The problem of recovering a model for the calibration of active pan-tilt-zoom cameras is addressed here. The variation of the intrinsic parameters of each camera over its full range of zoom settings is estimated through a two step procedure. We first determine the intrinsic parameters at the camera's lowest zoom setting very accurately by capturing an extended panorama. The camera intrinsics and radial distortion parameters are then determined at discrete steps in a monotonically increasing zoom sequence that spans the full zoom range of the camera. Our model incorporates the variation of radial distortion with camera zoom. Both the calibration phases are fully automatic and do not assume any knowledge of the scene structure. High-resolution calibrated panoramic mosaics are also computed during this process. We describe a hierarchical approach to building multiple level of detail in panoramas, by aligning hundreds of images captured within a 1-12X zoom range. The camera intrinsics model alongwith the cubemaps will provide a calibration reference for recovering precise pan,tilt of an active camera.

- S.N. Sinha and M. Pollefeys, *Towards Calibrating a pan-tilt-zoom camera network*. In OMNIVIS 2004, 5th Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras (in conjunction with ECCV 2004), Prague, Czech Republic.
- S.N. Sinha, M. Pollefeys and S. J. Kim, *High-Resolution Multiscale Panoramic Mosaics from Pan-Tilt-Zoom Cameras*. In Proc. of the 4th Indian Conf. on Computer Vision, Graphics and Image Processing, pp. 28-33, ICVGIP Kolkata, India, Dec 2004.

Camera Network Calibration and Synchronization

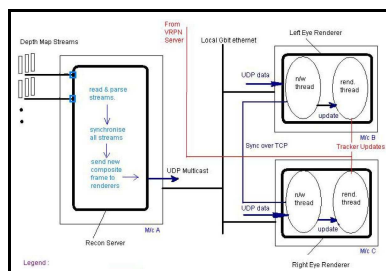


We have worked out an approach for automatic reconstruction of a dynamic event using multiple video cameras recording from different viewpoints. Those cameras do not need to be calibrated or even synchronized. All the necessary information is recovered by analyzing the motion of the silhouettes in the multiple video streams. The first step consists of computing the calibration and synchronization for pairs of cameras. We compute the temporal offset and epipolar geometry simultaneously using an efficient RANSAC-based algorithm. Next, the calibration and synchronization for the complete camera network is

computed. Finally the visual hull of the moving object is reconstructed. For unsynchronized video streams silhouettes are interpolated to deal with sub-frame temporal offsets. We are in the process of calibrating a large number of visual hull datasets captured in computer vision labs at MIT, CMU, Georgia Tech, MPI Saarbruken and UMD College Park. This approach could potentially be used to calibrate a 3D depth sensor, (as manufactured by Canesta Inc.) with respect to a color camera.

- S.N. Sinha and M. Pollefeys, *Visual-Hull Reconstruction from Uncalibrated and Unsynchronized Video Streams*, Second International Symposium on 3D Data Processing, Visualization & Transmission, 2004.
- S.N. Sinha, M. Pollefeys and L. McMillan, *Camera Network Calibration from Dynamic Silhouettes*, CVPR, pp. 195-202, 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) - Volume 1, 2004.
- S.N. Sinha and M. Pollefeys, *Synchronization and Calibration of Camera Networks from Silhouettes*, ICPR, pp. 116-119, 17th International Conference on Pattern Recognition (ICPR'04) - Volume 1, 2004.
- S.N. Sinha, *Calibration of a heterogeneous network of color and depth cameras*, one of the 10 selected entries in Canesta Design Contest, Phase I, Jan 2005.

Real-time rendering for 3D Tele-immersion



The 3D tele-immersion testbed jointly developed by UPenn and UNC consisted of a real-time distributed pipeline involving multiple video acquisition streams, stereo reconstruction, transport and interactive rendering of the reconstructed 3D scene. My contribution to the project was in developing a distributed rendering architecture which could process a larger number of depth-maps streams at interactive rates. The rendering system was able to render upto 9 incoming 640×480 depth-map streams arriving at 1-2 Hz, resulting in upto 2 million reconstructed points at 16 frames per second. The whole system was demo-ed at SuperComputing 2002, Baltimore.

- H. Towles, S.U. Kum, T. Sparks, S.N. Sinha, S. Larsen and N. Beddes. *Transport and Rendering Challenges of Multi-Stream 3D Tele-Immersion Data*, In CVRV 2003, October, 2003, Tahoe City, USA.
- N. Kelshikar, X. Zabulis, K. Daniilidis, V. Sawant, S.N. Sinha, T. Sparks, S. Larsen, H. Towles, K. Mayer-Patel, H. Fuchs, J. Urbanic, K. Benninger, R. Reddy, and G. Huntton. *Real-time Terascale Impementation of Tele-immersion*, Proc. of Intl. Conf. on Computation Science 2003, June, 2003, Melbourne, Australia.