Universidade do Minho
Escola de Engenharia

Semana da Escola de Engenharia
October 24 - 27, 2011

3D Reconstruction and Visualization with Mobile Devices using Sparse Images

José Henrique Brito
Departamento de Electrónica Industrial
E-mail: josehbrito@gmail.com

KEYWORDS
3D reconstruction, internet photo collections, SfM, mobile devices

ABSTRACT

This PhD project aims to do the 3D reconstruction of a scene from a collection of photos gathered from the Internet and then rendering the resulting model in a mobile device, such as a Smartphone or a Tablet PC, overlaying the rendering with the image captured by the mobile device's camera in real time.
This type of application is useful in the entertainment business as well as in a variety of scientific, engineering, industrial and cultural fields such as archaeology, architecture, e-commerce, forensics, geology, planetary exploration, movie special effects and virtual and augmented reality. (Pollefeys and Gool 2002).
The system should be able to take an image the scene and/or it's GPS coordinates and gather other images of the same scene into a database. From the image collection the system should be able to generate a 3D reconstruction of the scene. The 3D model should then be rendered in a mobile device, allowing the user to navigate in 3D space.
This project presents several challenges such as collecting and identifying images of the same scene, generating 3D models from uncalibrated images and from heterogeneous sources, extracting texture by combining different images, exploiting available image metadata, using mobile devices to visualize the geometric 3D models.
Some of the metadata to be exploited by the system includes GPS information, temporal information (date/time) and calibration data such as focal distance.
The resulting 3D model should also be aligned with georeferenced information, so that it may be presented to the end user in a way that is consistent with his/her point of view at any given time.
The development of the system is divided into several phases:

- Obtaining an image database with metadata, by identifying coherent images of the same scene
- Computing camera calibration information and generating a three-dimensional model from the picture collection, through SfM
- Generating coherent texture information for the geometry from pictures from heterogeneous sources
- Developing a server interface to feed geometric and texture information from the server to the mobile device, depending on its location at any given time
- Developing a client interface for the mobile device to receive the geometric and texture information from the server and render the 3D model according to the point of view

Although the system is primarily intended to be use in a mobile device setting, it could also be considered in a PC based environment or any other device capable of capturing images, obtaining the GPS position and relative orientation, network connectivity and reasonable rendering capabilities of 3D models.
For developing and testing purposes, a test scene was selected. This scene is a popular tourist landmark in Minho, Portugal.

Figure 1: Picture of the Selected Test Scene

To start building the image database, we performed an image search in popular image sharing websites like Flickr and Panoramio. The search was based on the GPS coordinates of the site. The initial search was subsequently complemented with image searches based on the text tags present in the metadata associated to the images gathered in the initial search. The images resulting from the secondary searches were then compared to the images from the primary search to throw away as many of the false positives as possible. This filtering of the dataset is done with image descriptors based on Gist and small thumbnails of the images (Hays and Efros 2008). It is done this way so that the process is relatively fast and the main objective is to reject a largest amount of images that really are very different from the reference set.

With the selected subset of the photo collection gathered from the internet, the system should be able to reconstruct a 3D model. The reconstruction process will detect and describe features in the individual images, so that they can be matched between different views of the same scene. The feature detectors could be SIFT (Lowe 2004), SURF (Bay at al. 2008) or similar detectors as in most works in this field (Snavely et al. 2008; Frahm et al. 2010). The feature matching process between images should also filter out images that were wrongly included in the dataset but that do not belong to the scene.

Using the feature matches, it is possible to compute the relative position and rotation of each view through Structure from Motion, which should be able to produce the camera poses and a sparse point cloud.

The metadata in the images will be used to get a first estimate of the intrinsic camera parameters, namely the focal length. In images for which no metadata is available, an estimate of the focal length will be computed from the feature matches with images for which metadata is available. The use of georeferenced metadata (GPS coordinates) in some of the images will also be useful to align the camera poses and the geometric model to a global reference frame.

The 3D reconstruction will be computed off-line, and the model will be stored for future use. Optionally, the model could be generated assyncronously, on demand by the end user.

The texture to be applied to the geometric model will be computed based on graph cut optimization and Poisson Blending, as in (Sinha, Steedly et al. 2008).

As the computational load and the data storage requirements of the 3D reconstruction process are not compatible with a complete implementation in a mobile device, the reconstruction functionalities will be implemented in a server with the necessary processing and data storage capabilities. The reconstructed model will be made available on-line and the end user will interact with the system through an interface implemented on the mobile device. The mobile device will get the model over the network and perform the rendering. Therefore, the system will be structured in a client/server architecture, with the modeling functionalities implemented in the server side and the rendering functionalities in the client side.

The client application will download the model and render it according to tis view point at any given time. We intend to estimate the view pont using some of the hardware that is usually available in modern mobile devices like GPS, digital compass and accelerometer.

REFERENCES

Pollefeys, M. and L. V. Gool (2002). "From images to 3D models." *Communications of the ACM* 45(7): 50-55.

Hays, J. and A. A. Efros (2008). "Scene completion using millions of photographs." *Communications of the ACM* 51(10): 87-94.

Lowe, D. G. (2004). "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision* 60(2): 91.

Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool (2008), "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346-359.

Snavely, N., S. M. Seitz, et al. (2008). "Modeling the World from Internet Photo Collections." *International Journal of Computer Vision* 80(2): 189-210

Jan-Michael Frahm, Pierre Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, Marc Pollefeys, (2010). "Building Rome on a Cloudless Day", *European Conference on Computer Vision*

Sinha, S. N., D. Steedly, et al. (2008). "Interactive 3D architectural modeling from unordered photo collections". *ACM SIGGRAPH Asia*

JOSÉ HENRIQUE BRITO was born in Braga, Portugal and has a degree in Electrical Engineering and Computer Science from Instituto Superior Técnico in Lisbon. He went on to obtain a Masters in Computer Graphics at Universidade do Minho and is now a PhD student in the Industrial Electronics of Universidade do Minho. His e-mail address is: josehbrito@gmail.com