Extended Abstract on: Minimal Structure from Motion Representation for Image Geocoding

Nuno Amorim and Jorge Gustavo Rocha

Abstract In this extended abstract we present our early work on structure from motion data compression for image geocoding. We address the advantages of image geocoding over standard trillateration solutions and identify desired characteristics for an image geocoding system such as accuracy, speed and scalability. We hypothesize that scalability impacts both speed and accuracy and should be further researched. Hence, in this thesis we would like to know which is the minimal representation of structure from motion to efficiently compute the location and orientation of new photographs.

1 Introduction

The recent explosion of images on social media has lead Computer Vision researchers to an increased interest on image processing algorithms. From those, structure from motion (SFM) has gained an increased relevance due to its applications. With a couple of photographs from the same scene, this algorithm is able to retrieve the 3D structure (point clouds) by processing the motion from photograph to photograph as stereo vision. Providing the correct focal lengths and distortion parameters, this algorithm can achieve an impressing precision on building point

Nuno Amorim
Algoritmi Research Centre, University of Minho, 4710-057 Braga, Portugal, e-mail: ntma90@gmail.com

Jorge Gustavo Rocha
Algoritmi Research Centre, University of Minho, 4710-057 Braga, Portugal, e-mail: jgr@di.uminho.pt

 In: A. Comber, B. Bucher, S. Ivanovic (eds.): Proceedings of the 3rd AGILE Phd School Champs sur Marne, France, 15-17-September-2015, published at http://ceur.ws.org.

clouds. Moreover, new photographs can be added at any time, allowing the ability to update models to this ever changing world.

The application of SFM can vary from simple visualization purposes to more complex tasks such 3D modeling and geographical location recognition. On the later, photographs with unknown location are compared with a geocoded database to compute their GPS coordinates.

2 State of Art

Early work on image geocoding started with methods based on a database of geocoded photographs [1, 2, 3]. Image features are extracted from new photographs, and feature matching is performed to retrieve similar photographs from the database. Two-view geometry is then executed to compute the pose of the query photographs.

Since two-view geometry is often computationally expensive as it usually relies in RANSAC based methods to compute the relative pose of images, there was an increased interest on using structure from motion to support the geocoding process. Image features are extracted from new photographs, but are directly compared to 3D point clouds. Rather than performing two-view geometry, it is computed a projection matrix which validates 2D (query photograph) to 3D points (database point clouds). Techniques such as nearest-neighbor feature matching [4] and data structures such as vocabulary trees [5, 6, 7] are often used on state of art work to greatly reduce the amount of matches needed to perform. Faster computational times can be achieved by resorting to parallel processing on CPU and GPU units as shown in [6] which achieved real time image geocoding with a GPU implementation of a vocabulary tree, if new photographs successfully matched the first document retrieved from their vocabulary tree queries.

3 Motivation

The advantages of image geocoding are clear when compared to other geocoding systems. First of all image geocoding does not rely on a trilateration process, which means it can compute coordinates on indoor environments as long as it has a WiFi connection to issue the geocoding request. Also, image geocoding has access to the heading of the queried photographs allowing the calculus of the direction in which the photographs were taken, in a single query. This reinforce the utility of image geocoding on supporting guiding systems. Lastly, the only (and ideal) device required for image geocoding is a smartphone, since it contains both camera and WiFi connection. As they are now omnipresent in our society, the cost for deploying an image geocoding solution is greatly decreased.

4 Problem to be Solved

In order to image geocoding replace standard trilateration solutions, three characteristics are desired: accuracy, speed and scalability. Starting from accuracy, a good pose estimation is attained when there is related data within the database to the queried photograph. Feature matching is used to ascertain which data to use when computing the pose. Assuming that we are facing the best case scenario and the focal length and distortion parameters are known, then an impressive precision can be achieved with a single photograph.

Speed is defined by how fast can we find the correct database data to geocode the queried photograph. Image processing requires expensive matrix operations as every image pixel is relevant to compute visual features. Additionally, high resolution images deliver better image features but also increase the computational time on extracting and matching those. However, the constant evolution of hardware and the parallelization of image processing algorithms is progressively breaking the barrier of real time processing.

Being scalable means that neither speed and accuracy are hindered with the growth of the geocoded database, which is not quite the case. Assuming that each SFM model contains millions of points, and each point is related to at least two image descriptors and associated 2D data, a massive amount of visual data is required to support an image geocoding system. Consequently, querying the geocoded database gets slower. Besides, more information means having an higher amount of similar features, which may confuse the image geocoding system into geocoding photographs miles way from their correct location.

5 Research Question and Future Work

Facing the scalability problem described in the previous section, in this thesis we question which is the minimal scene representation of structure from motion to allow a good geocoding rate.

Our main objective will be study and benchmark different state of art SFM based geocoding systems, to enhance existing SFM compression strategies or to develop alternative compression methods. We want compression rates able to maintain the geocoding speed and rate, while allowing the scalability of the geocoding system to wider areas. Also, rather than delivering a perfect 100% geocoding rate, we are only interested in avoiding hindering this rate due to aggressive compression.

We are aware that there is state of art research concerning SFM compression [6, 4, 8], but rather than focusing the compression into a single geocoding engine, we will generalize it to currently available engines. Since all image geocoding methods work under the same assumptions (image features and 3D pose estimation), we believe that the generalization is achievable.

References

1. Robertson D, Cipolla R (2004) An Image-Based System for Urban Navigation. Review Literature And Arts Of The Americas, doi: 10.5244/C.18.84
2. Werner M, Kessel M, Marouane C (2011) Indoor positioning using smartphone camera, International Conference on Indoor Positioning and Indoor, doi: 10.1109/IPIN.2011.6071954
3. Wang, E, Yan W (2013) iNavigation: an image based indoor navigation system, Multimedia Tools and Applications, doi: 10.1007/s11042-013-1656-9
4. Li Y, Snavely N, Huttenlocher DP (2010) Location Recognition using Prioritized Feature Matching. ECCV'10 Proceedings of the 11th European conference on Computer vision, doi: 10.1007/978-3-642-15552-9_57
5. Schindler G, Brown M, Szeliski R (2007) City-Scale Location Recognition. Computer Vision and Pattern Recognition, CVPR 07. IEEE Conference, doi: 10.1109/CVPR.2007. 383150
6. Irschara A, Zach C, Frahm J, Bischof H (2009) From structure-from-motion point clouds to fast location recognition. IEEE Conference on Computer Vision and Pattern Recognition, doi: 10.1109/CVPR.2009.5206587
7. Huitl R, Schroth G, Hilsenbeck S, Schweiger F, Steinbach E (2012) TUMindoor: An extensive image and point cloud dataset for visual indoor localization and mapping. 19th IEEE International Conference on Image Processing, doi: 10.1109/ICIP.2012.6467224
8. Cao S, Snavely N (2014) Minimal Scene Descriptions from Structure from Motion Models. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), doi: 10.1109/CVPR.2014.66