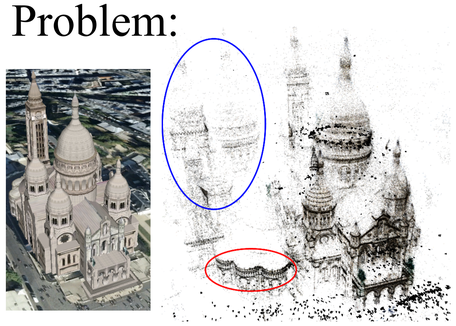
Network Principles for SfM: Disambiguating Repeated Structures with Local Context

Kyle Wilson
Noah Snavely

www.cs.cornell.edu/projects/disambig

Problem:



Repeated or ambiguous structures are a key challenge to structure from motion algorithms. Examples include:

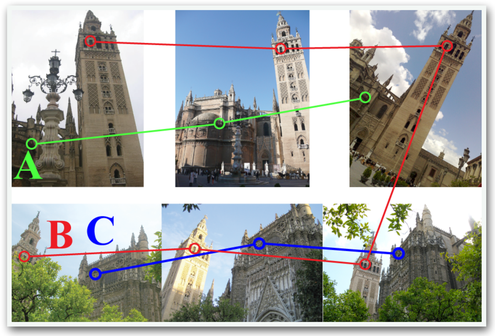
- Faces of towers (eg, Big Ben)
- Radially symmetric domes
- Buildings with strong bilateral symmetry

Ambiguities can cause major errors ranging from large duplicate structures to a complete failure to reconstruct.

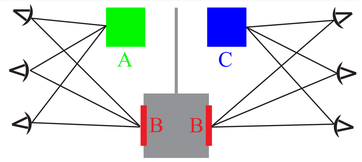
Our new method is simple, scalable, and effective on Internet photo collections with 1000s of images, a much larger scale than previous geometry-based techniques [3,4]. The key idea is to identify anomalous local structures in a visibility graph.

Model:

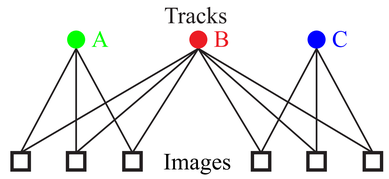
A **track** is a set of 2D observations of a 3D point matched across several images. **Good tracks** (A,C) each represent a single 3D point, but **bad tracks** (B) represent more than one, due to errors in correspondence.



Our key insight is that bad tracks often span distinct groups of background elements.



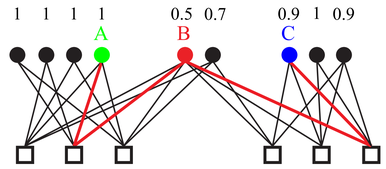
We measure this in the **visibility graph**, a bipartite graph connecting tracks with images they are seen in.



The **bipartite local clustering coefficient** is a measure of graph connectedness. It is low for tracks that bridge clusters. (This is a generalization of *triadic closure* to bipartite graphs). This is based on work in network analysis [1,2].

$$\mathrm{blcc}(t) = \frac{\#\text{ closed 4-paths centered at } t}{\#\text{ 4-paths centered at } t}$$

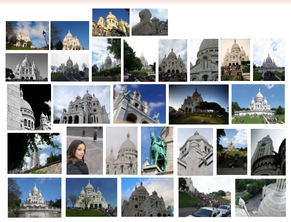
To illustrate, a sample 4-path is highlighted in red on a toy scene and blcc scores are indicated. The bad tracks have lower scores.



Method:

A preprocessing step improves our method's effectiveness: Internet photo collections tend to have some viewpoints represented vastly more than others. This artificially inflates blcc scores. We compute a **covering subgraph**, a subset of the images which more uniformly observes each track. We prefer images with large field of view because they see more background context.

The image to the right shows the first few images in the covering subgraph, computed greedily for the **SacreCoeur** dataset. Note that these are high-context images.



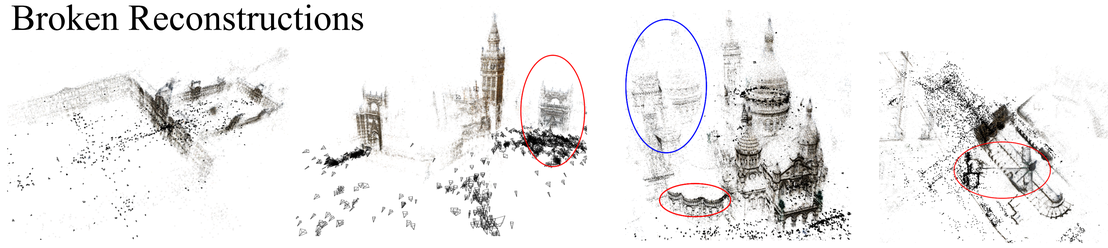
Random Sampling from all Images

Images Selected First for Subgraph

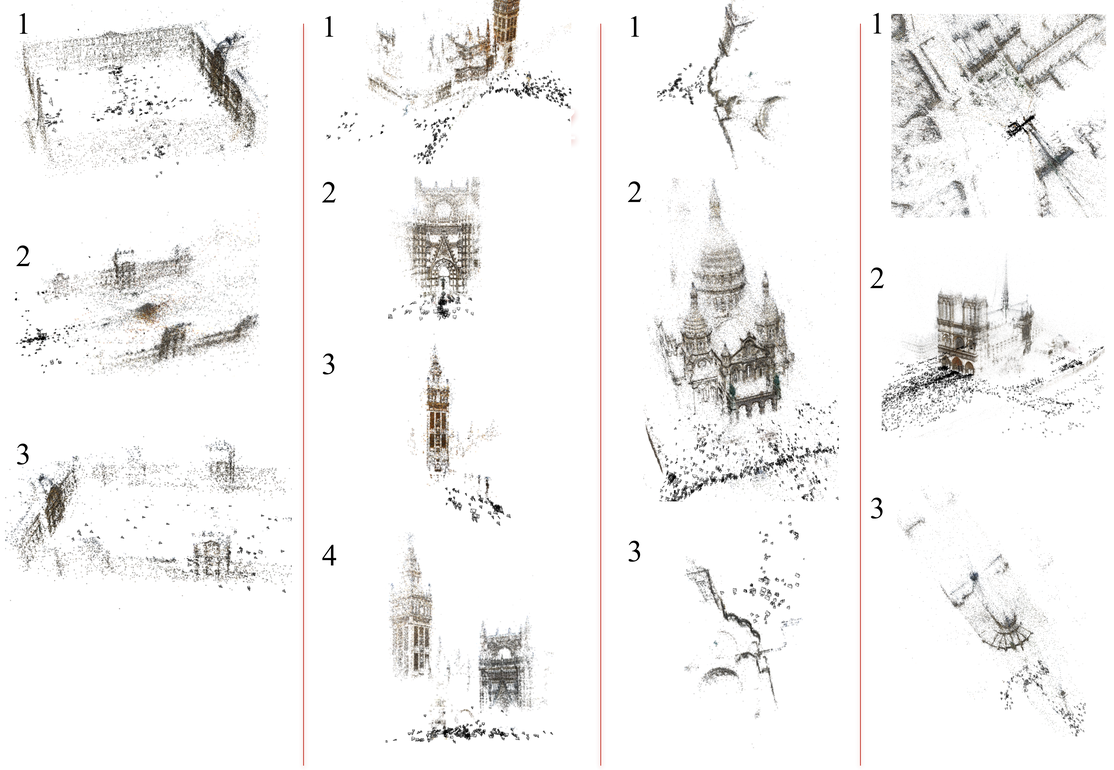
Algorithm to Disambiguate a Scene:

1. Compute a covering subgraph
2. Compute blcc for each track (using subgraph)
3. Remove tracks scoring lower than a threshold
4. Iteratively choose that threshold as the lowest which gives a specified number of components
5. Run reconstruction on each component
6. Rigidly merge components if possible

Broken Reconstructions



Disambiguated Components



Limitations:

- Some scenes break into unnecessary components
- Based around assumptions of lots of redundant data —does not work on small, sparse toy problems

References:

1. S. H. Strogatz. Exploring complex networks. Nature, 410(6825):268–276, March 2001.
2. T. Opsahl. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Social Networks, 2011.
3. C. Zach, M. Klopschitz, and M. Pollefeys. Disambiguating visual relationships using loop constraints. In CVPR, 2010.
4. N. Jiang, P. Tan, and L. Cheong. Seeing double without confusion: Structure-from-motion in highly ambiguous scenes. In CVPR, 2012.