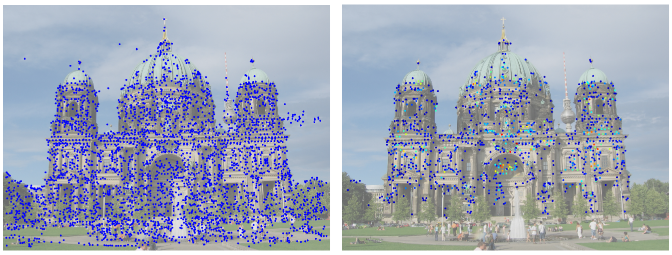
Improved Geometric Verification for Large Scale Landmark Image Collections

Rahul Raguram
rraguram@cs.unc.edu

Joseph Tighe
jtighe@cs.unc.edu

Jan-Michael Frahm
jmf@cs.unc.edu

Department of Computer Science
University of North Carolina
Chapel Hill, NC, USA.



(a) (b)

Figure 1: (a) All detected features for a single image (b) Features filtered based on the results of geometric verification – only visual words that were inliers in at least 10 previous image pairs are shown. The features in (b) are heatmap colour coded based on the inlier counts.

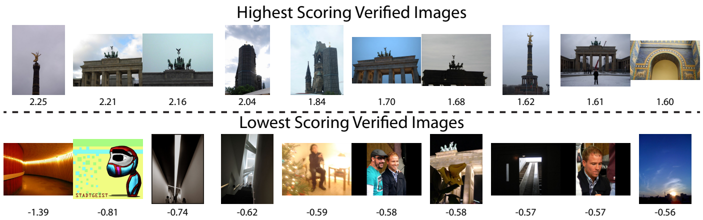


Figure 2: The top and bottom ten verified images, according to classifier scores. Our approach uses the output of geometric verification to generate a training set in an online manner, without manual labeling. The resulting classifier is able to reliably distinguish between landmark and non-landmark images.

In this work, we address the issue of geometric verification, with a focus on modeling large-scale landmark image collections gathered from the internet. In particular, we show that we can compute and learn descriptive statistics pertaining to the image collection by leveraging information that arises as a by-product of the matching and verification stages.

In designing a 3D reconstruction system for internet photo collections, one of the key considerations is robustness to "clutter" – when operating on datasets downloaded using keyword searches on community photo sharing websites (such as Flickr), it has been observed that invariably, a large fraction of images in the collection are unsuitable for the purposes of 3D reconstruction [3, 4]. Thus, one of the fundamental steps in a 3D reconstruction system is *geometric verification*: the process of determining which images in an internet photo collection are geometrically related to each other. This is a computationally expensive process, and much work in recent years has focused on developing efficient ways to perform this step. For example, Agarwal et al. [1] use image retrieval techniques to determine, for every image in the dataset, a small set of candidate images to match against. An alternate approach, adopted by Frahm et al. [2], is to first cluster the images based on global image descriptors and to then perform the verification within each cluster. While these approaches are extremely promising, there are still some limitations. For instance, even the carefully optimized approach described in [2] spends approximately 50% of the processing time simply verifying image pairs against each other. In addition, the approach in [2] suffers from "incompleteness"; due to the coarse clustering, a large fraction of images are discarded following the clustering and verification steps. In this work, we aim to overcome these limitations.

Thus far, the typical way to perform geometric verification has been to estimate the geometric relationship between pairs *independently*, which does not fully exploit the specific characteristics of the dataset being processed. Our main idea in this work is simple: as the geometric verification progresses, we learn information about the image collection, and subsequently use this learned information to improve efficiency and completeness. More specifically, since images of the same geometric structures are being repeatedly verified against each other, this process of repeated matching reveals useful information about two things:

(a) the stability and validity of low-level image features

(b) the global appearance of the various landmarks in the dataset

As a motivating example, consider Figure 1(a), which shows all detected SIFT features for a single image. Note that a large number of features lie in areas of the image that are very unlikely to pass any geometric consistency check (for e.g., features on vegetation, people, and in the sky). Now, if we have previously verified *other* images of the same scene, we can weight each visual word in the current image by the number of times that the word has previously passed the geometric consistency check in other image pairs (see Figure 1(b)). Note, in particular, that this weighting emphasizes visual words that are stable, reliable, and more likely to be geometrically consistent, while also suppressing spurious visual words. Our first contribution in this work is to integrate this learned information into a RANSAC procedure, which results in an appreciable improvement in efficiency compared to current techniques.

As a second contribution, we show that it is possible to learn additional useful information capturing higher-level information about the dataset. For instance, once we have obtained a sufficiently large set of successfully verified image pairs, we hypothesize that this set captures useful information about the *global* appearance of various landmarks present in the dataset. We observe that this information can then be used to train a classifier that distinguishes between landmark and non-landmark images (see Figure 2). We then employ this classifier during the image registration step. In this context, having a trained model of *landmark appearance* is very useful, since this allows us to only verify those images that are likely to be landmark images and discard the rest.

In summary, this work presents techniques for taking advantage of the information generated during geometric verification, to improve the overall efficiency of the process. Our approach thus integrates online knowledge extraction seamlessly into structure-from-motion systems, and is particularly relevant for large-scale image collections. Our results demonstrate both improved efficiency, as well as higher image registration performance, potentially yielding more complete 3D models for these large-scale datasets.

[1] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski. Building Rome in a Day. In *International Conference on Computer Vision*, 2009.

[2] Jan-Michael Frahm, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, and Marc Pollefeys. Building Rome on a Cloudless Day. In *European Conference on Computer Vision*, volume 6314, pages 368–381, 2010.

[3] L. Kennedy, S.-F. Chang, and I. Kozintsev. To Search or To Label?: Predicting the Performance of Search-based Automatic Image Classifiers. In *ACM Multimedia Information Retrieval Workshop (MIR 2006)*, 2006.

[4] Rahul Raguram, Changchang Wu, Jan-Michael Frahm, and Svetlana Lazebnik. Modeling and Recognition of Landmark Image Collections Using Iconic Scene Graphs. *Int. J. Comput. Vision*, 95(3):213–239, 2011.