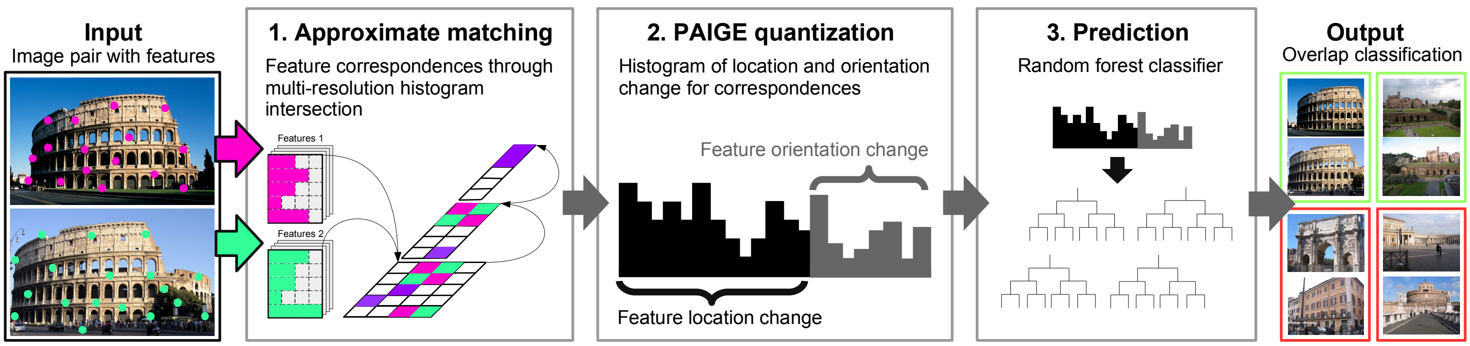
PAIGE: PAirwise Image Geometry Encoding for Improved Efficiency in Structure-from-Motion

Johannes L. Schönberger, Alexander C. Berg, Jan-Michael Frahm
Department of Computer Science, The University of North Carolina at Chapel Hill



Over the last years, large-scale Structure-from-Motion (SfM) has seen tremendous evolution in terms of robustness and speed in all stages of processing [1, 2, 3, 4, 5, 6, 7, 8, 9]. Incremental SfM commonly starts with feature detection and extraction (Stage 1), followed by matching (Stage 2), and geometric verification (Stage 3) of successfully matched pairs. After the matching and verification stage, typical SfM seeds the model with a carefully selected initial two-view reconstruction, before incrementally registering new cameras from 2D-3D correspondences, triangulating new 3D features, and refining the reconstruction using bundle-adjustment (Stage 4).

Generally, major computational effort is spent on Stages 2–4. In Stages 2 and 3, it is essential to discover a sufficient number of image correspondences that link together all parts of the scene to obtain complete and large-scale reconstructions. In addition, robust and accurate alignment is aided by finding multiple redundant image-to-image connections across the entire scene. However, exhaustively searching for these overlapping pairs is infeasible for large-scale image collections due to quadratic computational complexity in the number of images and features. Moreover, as the number of registered images grows, the scalability of bundle-adjustment algorithms becomes a significant performance bottleneck.

This paper evaluates existing techniques for reducing the cost of Stages 2 and 3, feature matching and geometric verification. Usually, the majority of image pairs in unordered Internet photo-collections do not have scene overlap, so rejecting those pairs dominates execution time, even though such pairs are not useful for 3D reconstruction. Consequently, various approaches have been proposed to efficiently find overlapping pairs in noisy datasets and only forward those pairs to Stages 2 and 3. A downside of sending fewer image pairs to Stages 2 and 3 is that enough images with overlapping geometry must be processed to produce accurate camera alignment and complete reconstructions. Hence, it is essential to find the right trade-off between computational efficiency and sufficient image connectivity.

Despite the impressive progress in reducing the cost of the matching (Stage 2), relatively little attention has been paid in comparing the techniques. The goals of this paper are therefore twofold: First, we present a comprehensive analysis and evaluation of various state-of-the-art matching techniques; second, we use the insights gained from this evaluation to propose the PAirwise Image Geometry Encoding (PAIGE) to build a scalable framework for the efficient recognition of the relative viewing geometry, all without explicit feature matching and without reconstructing the actual camera configuration using geometric verification. The proposed encoding is based on location and orientation properties efficiently inferred from approximate feature correspondences. A subsequent classification strategy leverages the encoding to only perform matching and geometric verification for image pairs that are identified as overlapping. As demonstrated in comprehensive experiments, this novel approach leads to a further speedup of large-scale SfM than the existing state of the art.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

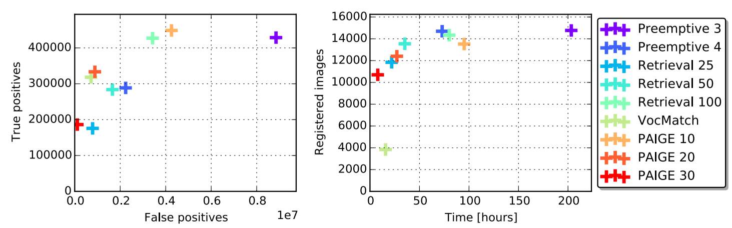


Figure 1: Evaluation results for preemptive matching [9], retrieval-based matching [1], VocMatch [4], and our proposed approach. Left: ROC figure for classification of overlapping (positives) and non-overlapping (negatives) image pairs. Right: Reconstruction performance in terms of matching time and number of registered images.

[1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Commun. ACM*, 54(10):105–112, October 2011.

[2] D. Crandall, A. Owens, N. Snavely, and D. P. Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *Proc. CVPR*, 2011.

[3] J.M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, and M. Pollefeys. Building rome on a cloudless day. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Proc. ECCV*, volume 6314 of *Lecture Notes in Computer Science*, pages 368–381. 2010.

[4] Michal Havlena and Konrad Schindler. Vocmatch: Efficient multiview correspondence for structure from motion. In *Proc. ECCV*, volume 8691, pages 46–60. 2014.

[5] J. Heinly, J. L. Schönberger, E. Dunn, and J.-M. Frahm. Reconstructing the World* in Six Days *(As Captured by the Yahoo 100 Million Image Dataset). In *Proc. CVPR*, 2015.

[6] J. L. Schönberger, F. Radenović, O. Chum, and J.-M. Frahm. From Single Image Query to Detailed 3D Reconstruction. In *Proc. CVPR*, 2015.

[7] N. Snavely, S. M. Seitz, and R. Szeliski. Skeletal graphs for efficient structure from motion. In *Proc. CVPR*, 2009.

[8] K. Wilson and N. Snavely. Robust global translations with 1dsfm. In *Proc. ECCV*, volume 8691, pages 61–75. Springer International Publishing, 2014.

[9] Changchang Wu. Towards linear-time incremental structure from motion. In *Proc. 3D Vision*, pages 127–134, June 2013.