

# Efficient Two-View Geometry Classification

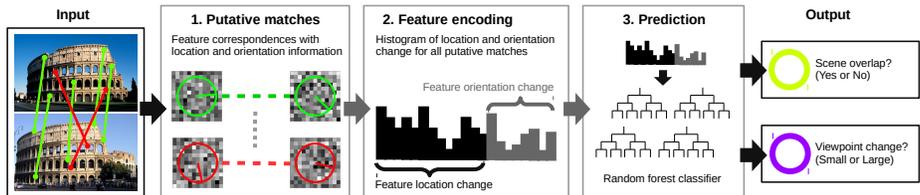
Johannes L. Schönberger, Alexander C. Berg, Jan-Michael Frahm

The University of North Carolina at Chapel Hill

**Abstract.** Typical Structure-from-Motion systems spend major computational effort on geometric verification. Geometric verification recovers the epipolar geometry of two views for a moving camera by estimating a fundamental or essential matrix. The essential matrix describes the relative geometry for two views up to an unknown scale. Two-view triangulation or multi-model estimation approaches can reveal the relative geometric configuration of two views, e.g., small or large baseline and forward or sideward motion. Information about the relative configuration is essential for many problems in Structure-from-Motion. However, essential matrix estimation and assessment of the relative geometric configuration are computationally expensive. In this paper, we propose a learning-based approach for efficient two-view geometry classification, leveraging the by-products of feature matching. Our approach can predict whether two views have scene overlap and for overlapping views it can assess the relative geometric configuration. Experiments on several datasets demonstrate the performance of the proposed approach and its utility for Structure-from-Motion.

## 1 Introduction

Over the last decade Structure-from-Motion (SfM) systems have seen tremendous evolution in terms of robustness and efficiency [1, 13, 8, 31]. Incremental SfM systems (Fig. 2) typically start with feature extraction and detection (Stage 1), followed by matching (Stage 2) and geometric verification (Stage 3) of successfully matched pairs by the assessment of the relative viewing configuration. The major computational effort is spent on Stages 2 and 3. The incremental reconstruction seeds the model with a carefully selected initial two-view reconstruction. Next, the procedure incrementally registers new cameras from 2D-3D correspondences, triangulates new 3D features, and refines the reconstruction using a non-linear optimization, known as bundle-adjustment (Stage 4). The input to the incremental reconstruction procedure (Stage 4) is typically a graph of relative, pairwise epipolar transformations. Information about the relative geometric configuration, such as small or large baseline and forward or sideward motion, is essential for SfM, since the incremental reconstruction procedure is highly dependent on the order in which cameras are registered. A suitable initial image pair and similarly a suitable next-best-view during the incremental extension depends on the relative viewing geometry, i.e. uncertainty of 3D features and camera parameters. However, assessment of the relative viewing geometry



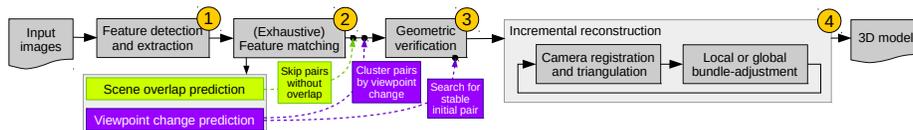
**Fig. 1.** The proposed framework for extracting PAIGE, and its application for scene overlap and viewpoint change prediction.

for every overlapping image pair in a dataset is computationally expensive. This paper presents a technique for efficiently recognizing image pairs that work well for incremental SfM – significantly improving efficiency for geometric verification as well as improving reconstruction robustness.

The relative geometric configuration of overlapping image pairs serves as the input to the incremental reconstruction procedure. Geometric verification attempts to estimate the relative viewing geometry for pairs of overlapping images. Usually, the majority of image pairs in large-scale, unordered photo-collections do not have scene overlap, and thus rejecting invalid pairs dominates execution time. Determining the relative viewing geometry for large image sets comes at significant computational expense, especially if the overlap between most images is sparse. However, it is a necessary step, as unfavorable initializations or an unfortunate order in camera registrations, e.g., pairs resulting in high camera and/or point uncertainty, can lead to failures in registration and bundle-adjustment due to weak geometry, local minima, degeneracies, etc.

The traditional procedure to assess the two-view geometry in geometric verification comprises fundamental or essential matrix estimation [20] followed by triangulation of 3D points [15], multi-model estimation strategies like GRIC [29], or extended RANSAC procedures for model selection such as QDEGSAC [12]. The essential matrix reveals the entire two-view geometry of calibrated cameras up to unknown scale. Triangulation of 3D points, GRIC, or QDEGSAC then determine the properties of the relative viewing geometry, e.g., the amount and direction of viewpoint change. However, while efficient on a per pair basis, these methods are computationally expensive for a large number of image pairs.

In this paper, we design an encoding of image characteristics and build a framework (Fig. 1) for the efficient recognition of image pairs with scene overlap and prediction of the stability of their two-view geometry, all without explicitly reconstructing the actual camera configuration using essential matrix estimation. The approach is based on the location and orientation properties of putative feature correspondences. In Section 6, we experimentally demonstrate the utility of the proposed framework for a variety of SfM modules, e.g. reducing the set of image pairs for which to perform geometric verification and efficient search for stable initial image pairs in large datasets.



**Fig. 2.** The stages of a typical SfM pipeline, and applications of our proposed scene overlap and viewpoint change predictor in green and purple.

## 2 Related work

Over the last years large-scale SfM systems have tremendously advanced in terms of increased robustness and reduced runtime. A variety of methods to reduce runtime in different stages of the SfM pipeline (Fig. 2) have been proposed. However, current state-of-the-art systems typically still spend major time in Stages 2 and 3. To reduce the number of image pairs in the exhaustive matching module (Stage 2), Frahm et al. [13] leverage iconic image selection through clustering of similar images, Agarwal et al. [1] employ image retrieval systems [21] to only match against similar images, Raguram et al. [26] use GPS tags to match images only to spatially nearby ones, and Wu [31] proposes a preemptive matching strategy. Recently, Hartmann et al. [16] proposed to predict the matchability of individual features (Stage 1) to reduce the number of feature comparisons during exhaustive matching (Stage 2). Most recently, Schönberger et al. [27] proposed a learning-based approach to predict scene overlap based on approximate feature correspondences. However, these techniques still yield a significant amount of image pairs that have no scene overlap, and the set of images contains many redundant viewpoints. Despite the variety of approaches, they all rely on elaborate two-view reconstructions on their potentially reduced set of images in the geometric verification stage. Apart from algorithmic advancements on estimation techniques [20, 23], only Raguram et al. [25] tried to specifically improve runtime of Stage 3 using an online learning strategy. However, their approach suffers from a significant loss of image registrations.

Complementary to these previous efforts, we propose a new method to further improve the efficiency in SfM by significantly reducing the runtime of the geometric verification module (Stage 3). Our method can detect overlapping image pairs before geometric verification and for overlapping image pairs it can efficiently classify the geometric two-view configuration in terms of the amount of viewpoint change. We achieve this by extending the method of Schönberger et al. [27] who pose the problem of scene overlap detection as a classification task. Similar to their method, we exploit the observation that when images are taken from different viewpoints, corresponding features change in scale, location, and rotation in recognizable patterns. However, instead of approximate correspondences through histogram intersection, we leverage the more reliable feature correspondences from putative matching enabling a less noisy encoding and more accurate prediction. Even though our method builds on the idea of Schönberger

et al. [27], both approaches can be used together as filters for feature matching and geometric verification in the same SfM pipeline.

### 3 Two-view geometry

#### 3.1 Estimation

Traditional techniques to derive the two-view geometry comprise feature matching (Stage 2), followed by robust essential matrix estimation (Stage 3). The essential matrix reveals the relative viewing geometry [15], but its estimation is computationally expensive [20] due to outliers and non-linearity. RANdom SAmple Consens (RANSAC) [10] or its more efficient variants [6, 7, 24, 23] are usually used for robust estimation. RANSAC can deal with large fractions of outliers, but has exponential computational complexity in the number of model parameters  $s$  and the inlier ratio  $e$ . To sample at least one outlier-free set of measurements with confidence  $p$ , one must run at least

$$d = \log(1 - p) / \log(1 - e^s) \quad (1)$$

number of iterations. Hence, the complexity quickly rises for small inlier ratios which are commonly encountered in SfM from unordered photo collections [26] (see Section 5). Moreover, RANSAC becomes infinitely expensive for image pairs without overlap since those pairs have no inliers. Hence, traditionally a minimum inlier ratio  $e_{min}$  is assumed to set an upper bound for the number of RANSAC iterations. Efficiently detecting image pairs that do not have scene overlap prior to geometric verification can significantly reduce the runtime of Stage 3.

The essential matrix reveals the relative transformation between two views up to an unknown scale. To derive more information about the relative viewing geometry, such as the amount of viewpoint change or the type of motion, further processing is necessary. Scene reconstruction enables to determine the amount of viewpoint change through scene analysis such as triangulation angle calculation. Alternatively, decision criterions like GRIC [29] or an extend RANSAC procedure like QDEGSAC [12] can be used to avoid degenerate viewing configurations. These methods are computationally expensive. In this paper, we propose a more efficient method to classify the amount of viewpoint change without explicit reconstruction of the scene.

#### 3.2 Uncertainty

In this section, we briefly describe the relevance of the two-view geometry for uncertainty estimation in 3D reconstruction, its relation to the baseline-length and the triangulation angles, and how this affects the search for an initial pair and the order of camera registrations in SfM. Uncertainty of the 3D feature and the camera parameter estimates in bundle-adjustment are determined by five main factors [17, 9, 11, 18]: redundancy, reliability, uncertainty of measurements, viewing geometry, and gauge. These factors have important implications for the

design of SfM systems w.r.t. the search for an optimal initial pair and a suitable next-best-view. On the one hand, for accurate reconstructions, we want to jointly maximize the number of image measurements (high redundancy and reliability) and the stability of the two-view geometry (large triangulation angles). On the other hand, we wish to achieve optimal results (uncertainty and model size) with minimal computational effort, i.e. with as few measurements and camera registrations as possible.

## 4 Feature representation

Our proposed feature representation builds upon the PAIGE feature by Schönberger et al. [27]. In this section, we describe our adaptations and extensions to their method for the efficient prediction of the two-view geometry.

PAIGE takes the extracted features from Stage 1, performs approximate feature matching through histogram intersection, and predicts scene overlap for an image pair by exploiting statistics from corresponding feature properties. Only overlapping image pairs are then forwarded to the computationally expensive pairwise image matching module (Stage 2). Analogous to their approach, we exploit the fact that corresponding features change in scale, location  $x$ , and orientation  $o$  in recognizable patterns when images are taken at different viewpoints. However, our approach leverages the more precise feature correspondences produced by feature matching in Stage 2, which enables us to produce a less noisy encoding for more accurate prediction.

For each putative feature correspondence of a matched image pair  $a$  and  $b$ , we determine the normalized image coordinates  $\mathbf{x}_a, \mathbf{x}_b$ , such that  $x_i \in [0, 1]^2$ . Normalization is necessary due to possibly different image resolutions of image  $a$  and  $b$ . Next, we calculate the displacement for each correspondence as

$$\Delta x = \|\mathbf{x}_a - \mathbf{x}_b\|_2 \quad (2)$$

We quantize the distribution of feature displacements in a  $d_{\Delta x}$ -dimensional histogram  $\mathbf{h}_{\Delta x}$  with evenly spaced bins in the interval  $[0, 1]$ . Analogously, for each feature correspondence, we calculate the change in feature orientation

$$\Delta o = |o_a - o_b| \quad \text{mod } 2\pi \quad (3)$$

and we quantize the distribution of orientation changes in an  $d_{\Delta o}$ -dimensional histogram  $\mathbf{h}_o$  with evenly spaced bins in the interval  $[0, 2\pi]$ . We normalize each of the histograms

$$\bar{\mathbf{h}}_x = \frac{\mathbf{h}_x}{\|\mathbf{h}_x\|_2}, \quad \bar{\mathbf{h}}_o = \frac{\mathbf{h}_o}{\|\mathbf{h}_o\|_2} \quad (4)$$

for invariance w.r.t. the number of feature correspondences. Finally, we use the concatenation of the normalized histograms as our proposed encoding

$$\mathcal{P}(a, b) = [\bar{\mathbf{h}}_x \quad \bar{\mathbf{h}}_o] \quad (5)$$

Similarly to PAIGE, we do not represent scale changes in the feature as it is a noisy measure. The next section describes a classification strategy leveraging this feature representation for scene overlap and triangulation angle prediction.

	Total pairs	Matched pairs	Verified pairs	$e_{all}$	$e_{geo}$	$d_0$	$d_1$	$d_2$
<b>Training &amp; Test</b>	73,542,704	1,602,996	449,207	47%	70%	2,357,586,073 (100%)	295,230,950 (12.5%)	194,851,427 (8.3%)
<b>Oxford</b>	82,944	21,574	16,303	56%	70%	72,445,847 (100%)	14,557,490 (20.0%)	9,604,943 (13.2%)
<b>Louvre</b>	693,889	252,798	4,539	27%	65%	613,625,401 (100%)	72,898,480 (11.9%)	48,212,996 (7.9%)
<b>Acropolis</b>	8,767,521	439,609	16,492	29%	78%	1,139,606,104 (100%)	117,886,481 (10.3%)	77,105,077 (6.8%)

**Table 1.** Evaluation datasets with average inlier ratio for matched ( $e_{all}$ ) and verified pairs ( $e_{geo}$ ). Number of RANSAC iterations for geometric verification without classifiers  $\mathcal{C}_A$  and  $\mathcal{C}_B$  ( $d_0$ ), after classifier  $\mathcal{C}_A$  ( $d_1$ ), and after classifiers  $\mathcal{C}_A$  and  $\mathcal{C}_B$  ( $d_2$ ).  $d_0$ ,  $d_1$ ,  $d_2$  for *Training & Test* only given for held-out test set.

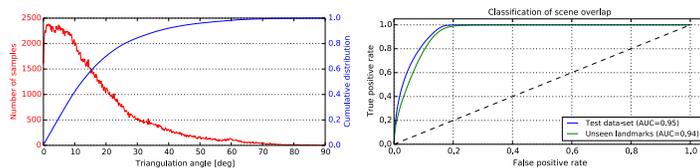
## 5 Classification

Based on the proposed encoding in Section 4, we now describe a classification strategy to answer the following two questions for any given image pair: *Is there scene overlap ( $\mathcal{C}_A$ )?* and *Is there a stable two-view geometry ( $\mathcal{C}_B$ )?* We choose random forests [3] as a classification method as it gave best results in terms of accuracy and computational efficiency.

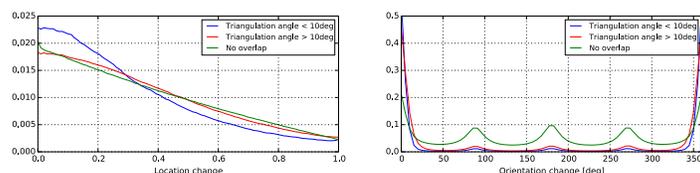
### 5.1 Training

For training, we use an existing 3D reconstruction of an image collection and its feature correspondences. Then, we calculate the mean triangulation angle  $\bar{\alpha}_{ab}$  for each image pair  $\{a, b\}$  with scene overlap as the dependent variable and extract the proposed feature  $\mathcal{P}(a, b)$  as the independent variable.

Specifically, we use 3D reconstructions of 17 unordered Internet photo-collections from different locations across the world (Rome, Notre Dame, Stonehenge, etc.) and a set of temporally sequential image sequences acquired by video cameras (to account for the orientation bias of crowd-sourced images) to serve as a training and test dataset (Table 1). The dataset consists of 1,602,996 matched ( $\geq 30$  putative feature correspondences) out of all 73,542,704 possible image pairs, of which 449,207 pairs have a geometrically verified overlap ( $\geq 15$  inliers for essential matrix estimation). Table 1 lists the minimum number of RANSAC iterations (Section 3) for essential matrix estimation of all matched image pairs with confidence  $p = 0.99$ , sample size  $s = 5$ , and minimum inlier ratio  $e_{min} = 0.28$ . As a result of these parameters, RANSAC runs for a maximum of  $d_{max} = 2674$  iterations for each image pair. The maximum number of iterations is reached for  $< 5\%$  of the pairs, since  $> 95\%$  of the pairs have an inlier ratio  $> 28\%$ . We employ SIFT features and use the ratio test for robust matching [19]; note that SIFT could be replaced by any other feature that provides location and orientation properties. The quantization of the location and orientation histograms include all 110,587,256 putative feature matches for all image pairs, including 51,968,824



**Fig. 3.** *Left:* Triangulation angle distribution for geometrically verified image pairs. *Right:* Performance evaluation for scene overlap classification  $\mathcal{C}_A$ .



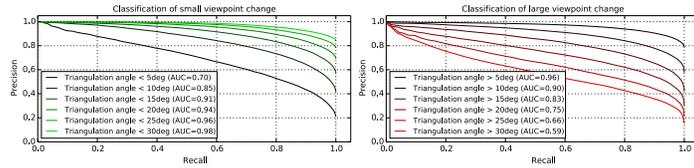
**Fig. 4.** Location and orientation change distributions of PAIGE for the entire dataset.

geometric inliers and 58,618,432 outliers, i.e. overall inlier ratio  $e_{all} = 47\%$  and  $e_{geo} = 70\%$  for geometrically verified pairs. We use a 172-dimensional feature vector  $\mathcal{P}(a, b)$  with  $d_{\Delta x} = 100$  and  $d_{\Delta o} = 72$ . Fig. 3 visualizes the distribution of triangulation angles and Fig. 4 the average feature vector  $\mathcal{P}(a, b)$  over all image pairs. We find a significant amount of pairs with only a small viewpoint change, caused by popular viewpoints of famous landmarks and less stable feature matching for large viewpoint changes. As expected, the overall location and orientation change is higher for wide than for small baselines, and the orientation change for images without overlap is significantly larger.

To answer the two binary classification problems  $\mathcal{C}_A$  and  $\mathcal{C}_B$ , we divide the set of image pairs into three different categories: small and large mean triangulation angle (using an angle threshold), and no scene overlap (pairs with failed geometric verification). Next, the dataset is split in randomly permuted training (70%) and test samples (30%). Two random forests were trained on the training dataset, using 50 decision trees each, entropy as the splitting criterion, and considering  $\sqrt{172} \approx 13$  features when looking for the best split at each node in the tree. A minimum number of three samples per leaf is enforced to avoid over-fitting. The parameters were determined with a 5-folded cross-validation on the training set. The trained random forests can efficiently decide on the two classification problems  $\mathcal{C}_A$  and  $\mathcal{C}_B$ . An embedding of the proposed classifiers in a typical SfM pipeline is demonstrated in Section 6.

## 5.2 Performance evaluation

On a conventional desktop computer the training time for both classifiers is approximately 5min, and the classification frequency averages at around 200K pairs per second including quantization and prediction, compared to around 20K



**Fig. 5.** Performance evaluation for triangulation angle classification  $\mathcal{C}_B$  using different angle thresholds. Area under curve as AUC.

pairs per second for the PAIGE approach [27]. We evaluate the classification performance on the held-out test set (30%), and three unordered photo-collections of completely unseen landmarks (see Table 1) at different geo-locations (Oxford, Louvre, Acropolis). Fig. 5 demonstrates the performance for both classifiers  $\mathcal{C}_A$  and  $\mathcal{C}_B$ . For classifier  $\mathcal{C}_A$ , we find minimal bias towards the trained landmarks, and we experience the same for classifier  $\mathcal{C}_B$ . In a subsequent evaluation, the performance of classifier  $\mathcal{C}_B$  is evaluated on the unseen landmarks w.r.t. different triangulation angle thresholds by only considering overlapping image pairs using  $\mathcal{C}_A$ . Fig. 5 shows that our method generalizes well. Next, we demonstrate the applicability of the two classifiers  $\mathcal{C}_A$  and  $\mathcal{C}_B$  within the context of SfM.

## 6 Efficient Structure-from-Motion

In the following, we show the embedding of the proposed method into a typical SfM system (Fig. 2) w.r.t. the datasets in Table 1. We demonstrate that the classifiers significantly improve the computational performance by reducing the set of images for the geometric verification module. Furthermore, we show the utility for the efficient search of stable initial image pairs in large datasets.

### 6.1 Scene overlap prediction

In Section 3, we have seen that the number of RANSAC iterations is exponentially dependent on the outlier ratio. Hence, we spend a majority of the runtime to evaluate pairs with no scene overlap. For these pairs RANSAC reaches the maximum number of iterations, leading to a significant computational burden. Our proposed method allows to filter these pairs prior to geometric verification, preventing the high computational effort for pairs that do not contribute to the final 3D model. Assuming we filtered all pairs with no scene overlap for the unseen landmarks (see Table 1) using a perfect classifier, and run RANSAC only for the remaining pairs, we can reduce the number of iterations by a factor of 35. For our classifier  $\mathcal{C}_A$ , we enforce a precision of  $\geq 0.99$  for classifying pairs with no scene overlap using an appropriate prediction confidence, and hereby lower the recall to 81%. This leads to the fact that our modified SfM pipeline only misses 1.7% of actually overlapping pairs. Please note that the majority of those images are still contributing to the final model through other pairs. Using these parameters, we achieve a 7.8x speedup for the training & test-set, and overall an 8.9x

speedup for the unseen landmarks compared to the potential speedup of 35 using a perfect classifier. Since the computational effort for the classification is insignificant compared to geometric verification (3 to 4 orders of magnitudes faster), this speedup directly propagates to the overall geometric verification runtime. Note that the performance improves even more, if we verify very weak image pair connections, since we assume a minimum inlier ratio of 28% ( $d_{max} = 2674$ ). The reported runtimes are a vast improvement over previous efforts [25], which achieve a 70% speedup but lose 26% of image registrations, in contrast to our 9-fold speedup with 1.7% loss. Due to the less noisy encoding based on putative feature matches, our approach misses significantly fewer image pairs than the PAIGE [27] approach, which loses 38-90% of actually overlapping image pairs. Note that both approaches could be employed together, since PAIGE operates as a filter to feature matching and our approach as a filter to geometric verification. On average, we find that exhaustive matching and geometric verification spends 52% in Stage 2 using a GPU SIFT implementation and 48% in Stage 3 using a multi-threaded CPU RANSAC implementation. Ideally, the PAIGE approach [27] can eliminate the runtime of Stage 2 for sparsely connected image collections. Our proposed approach in this paper reduces the runtime of Stage 3 by a factor of 9. Combining the two approaches, we can effectively eliminate the original cost of Stages 2 and 3 compared to standard exhaustive matching.

## 6.2 Redundant viewpoint detection

In SfM systems, we achieve redundancy by tracking a 3D feature over multiple images. Corresponding features between two images cannot only be verified with direct pairwise geometric verification, but also by bridging the track using an intermediate image, that has the same point in common. Especially for small viewpoint changes, the continuation of tracks over multiple images is very likely. Beyond that, uncertainty and reliability of parameter estimates in bundle-adjustment only improve up to a certain redundancy [2, 30], i.e. the resulting 3D models do not gain from high redundancy in the same way as we spend an unproportional amount of increased computational effort. For outlier-detection in SfM, it is typically critical to have at least 3-4 observations per 3D point. Leveraging these facts and classifier  $\mathcal{C}_{\mathcal{B}}$ , we can detect clusters of images with small viewpoint change. Next, we select one iconic image in the cluster with the most points in common, and finally only perform geometric verification from the iconic image to the rest of the images in the cluster rather than exhaustive verification between all pairs. Moreover, for very large clusters, we can limit the number of images for geometric verification, and simply register the remaining images w.r.t. the final model using 2D-3D pose estimation [10]. In both datasets, we see 40% of image pairs (282,387) with small viewpoint change ( $\bar{\alpha} < 10^\circ$ ). To find clusters, we build an undirected graph of all pairs with small viewpoint change using images as nodes and small viewpoint change as edges. In this graph, we find 6,404 disjoint maximal cliques [4, 28, 5] in the training & test-set. These cliques are similar to the clusters described by Frahm et al. [13], but our clusters

are based on viewpoint change rather than GIST similarity [22]. By only considering edges from the iconic to the remaining images in a clique, we reduce the pairwise geometric verifications from 97,564 to 20,426. In addition, we further decrease this number to 14,469 by only considering images up to a maximum cluster size of 10, i.e. we improve geometric verification runtime by 30% from 282,387 to 199,292 pairs. This technique is especially beneficial for very dense datasets, as often encountered in Internet photo-collections.

### 6.3 Search for optimal initial pairs

Searching for a good initial pair as a seed for incremental reconstruction is computationally expensive, since it involves essential matrix estimation followed by triangulation of feature correspondences, and the calculation of triangulation angles or uncertainty estimates. With state-of-the-art essential matrix solvers [20] and linear triangulation [14, ch. 12.2], around 10-50 two-view reconstructions can be computed per second [25] using the parameters as in Section 5. As opposed to the traditional approach, our classifier  $\mathcal{C}_B$  enables us to efficiently search for stable pairs through an entire dataset at significantly reduced computational cost. In the unseen landmarks, we find 14,886 stable pairs (out of 17,330 true stable pairs) with  $\bar{\alpha} > 20^\circ$ , where 83% of the reported pairs are actually stable. We use these pairs as initial seeds for the incremental reconstruction by ranking the reported stable pairs based on the number of putative feature matches to attain higher initial redundancy. On the one hand, our method leads to significantly faster search for initial pairs and, on the other hand, it allows us to search for optimal initial image pairs globally.

## 7 Conclusion

In this paper, we adapt the PAIGE feature for efficient two-view geometry classification to further improve the computational efficiency and robustness in SfM. Experiments demonstrate a speedup for geometric verification by an order of a magnitude over the traditional exhaustive approach, while only losing less than 1.7% of the valid image pairs. Compared to PAIGE, our method provides an order of magnitude faster prediction performance, while achieving significantly better prediction accuracy. PAIGE and our approach are complementary methods that can both be integrated into the same SfM pipeline to speedup feature matching and geometric verification. Furthermore, the framework significantly reduces runtime for very dense photo-collections and we demonstrate the utility for the efficient, global search of optimal initial pairs.

## Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1252921, IIS-1349074, IIS-1452851, CNS-1405847, and by the US Army Research, Development and Engineering Command Grant No. W911NF-14-1-0438.

## References

1. Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S., Szeliski, R.: Building rome in a day. ICCV (2009)
2. Baarda, W.: Statistical concepts in geodesy (1967)
3. Breiman, L.: Random forests. Machine Learning (2001)
4. Bron, C., Kerbosch, J.: Algorithm 457: Finding all cliques of an undirected graph. Commun. ACM (1973)
5. Cazals, F., Karande, C.: A note on the problem of reporting maximal cliques. Theoretical Computer Science (2008)
6. Chum, O., Matas, J.: Matching with prosac-progressive sample consensus (2005)
7. Chum, O., Matas, J., Obdrzalek, S.: Enhancing ransac by generalized model optimization. In: ACCV (2004)
8. Crandall, D., Owens, A., Snavely, N., Huttenlocher, D.P.: Discrete-Continuous Optimization for Large-Scale Structure from Motion. CVPR (2011)
9. Criminisi, A.: Accurate Visual Metrology from Single and Multiple Uncalibrated Images (2001)
10. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM (1981)
11. Förstner, W.: Uncertainty and projective geometry. In: Handbook of Geometric Computing (2005)
12. Frahm, J.M., Pollefeys, M.: RANSAC for (quasi-) degenerate data (QDEGSAC). CVPR (2006)
13. Frahm, J.M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.H., Dunn, E., Clipp, B., Lazebnik, S., Pollefeys, M.: Building Rome on a Cloudless Day. ECCV (2010)
14. Hartley, R., Schaffalitzky, F.:  $L_\infty$  minimization in geometric reconstruction problems. CVPR (2004)
15. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision (2003)
16. Hartmann, W., Havlena, M., Schindler, K.: Predicting matchability. In: CVPR (2014)
17. Kanatani, K.: Statistical Optimization for Geometric Computation: Theory and Practice (1996)
18. Kanatani, K., Morris, D.D.: Gauges and gauge transformations for uncertainty description of geometric structure with indeterminacy. IEEE Trans. Inf. Theor. (2006)
19. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV (2004)
20. Nister, D.: An efficient solution to the five-point relative pose problem. In: CVPR (2003)
21. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: CVPR (2006)
22. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV (2001)
23. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.: Usac: A universal framework for random sample consensus. IEEE PAMI (2013)
24. Raguram, R., Frahm, J.M., Pollefeys, M.: Arrsac: Adaptive real-time random sample consensus. In: ECCV (2008)
25. Raguram, R., Tighe, J., Frahm, J.M.: Improved geometric verification for large scale landmark image collections. In: BMVC (2012)

26. Raguram, R., Wu, C., Frahm, J.M., Lazebnik, S.: Modeling and recognition of landmark image collections using iconic scene graphs. *IJCV* (2011)
27. Schönberger, J.L., Berg, A.C., Frahm, J.M.: PAIGE: PAirwise Image Geometry Encoding for Improved Efficiency in Structure-from-Motion. *CVPR* (2015)
28. Tomita, E., Tanaka, A., Takahashi, H.: The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science* (2006)
29. Torr, P.H.: An assessment of information criteria for motion model selection. *CVPR* (1997)
30. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment - a modern synthesis. In: *Vision Algorithms: Theory and Practice* (2000)
31. Wu, C.: Towards linear-time incremental structure from motion. *3DV* (2013)