

Finding Better Wide Baseline Stereo Solutions Using Feature Quality

Stephen Nuske and Jay Patravali

Abstract Many robotic applications that involve relocalization or 3D scene reconstruction, have a need of finding geometry between camera images captured from widely different viewpoints. Computing epipolar geometry between wide baseline image pairs is difficult because often there are many more outliers than inliers computed at the feature correspondence stage. Abundant outliers require the naive approach to compute a huge number of random solutions to give a suitable probability that the correct solution is found. Furthermore, large numbers of outliers can also cause false solutions to appear like true solutions. We present a new method called UNIQSAC for ng weights for features to guide the random solutions towards high quality features, helping find good solutions. We also present a new method to evaluate geometry solutions that is more likely to find correct solutions. We demonstrate in a variety of different outdoor environments using both monocular and stereo image-pairs that our method produces better estimates than existing robust estimation approaches.

1 Introduction

Computing the relative geometry between a wide-baseline camera pair is an important task for many robotics problems such as loop-closure in SLAM systems or 3D reconstructions. The task can be difficult in challenging real-world scenes, mainly because it is hard to determine which point correspondences are correct (inliers) among a set containing many incorrect correspondences (outliers).

The standard approach, RANSAC [7], is to robustly estimate an epipolar geometry between the pair by testing multiple geometry hypotheses. RANSAC and its improved variants [2, 3, 6, 13, 16], are impressive in their ability to correctly estimate

S. Nuske · J. Patravali (✉)
Robotics Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh,
PA 15213, USA
e-mail: jaypatravali@cmu.edu

S. Nuske
e-mail: nuske@cmu.edu

geometry given small percentages of true correspondences and large percentages of false point correspondences. However, when the percentages of inliers and outliers are pushed further out of favor then even these robust approaches struggle to find the correct geometry. To that end, we propose a new metric for the quality of the features, which is their uniqueness amongst all other features in the image. We use this metric to weight the sampling of correspondences to help find good solutions, and we also present a new method to evaluate solutions that can better differentiate between true and false solutions. Figure 1 shows an example of our algorithm’s functioning.

Our sampling approach is different to existing weighted sampling strategies in that we first robustly quantize each feature and then calculate its uniqueness among all other features in the image, as opposed to just a feature correlation score to the nearest neighbor or a ratio of correlation scores to the second nearest neighbor.

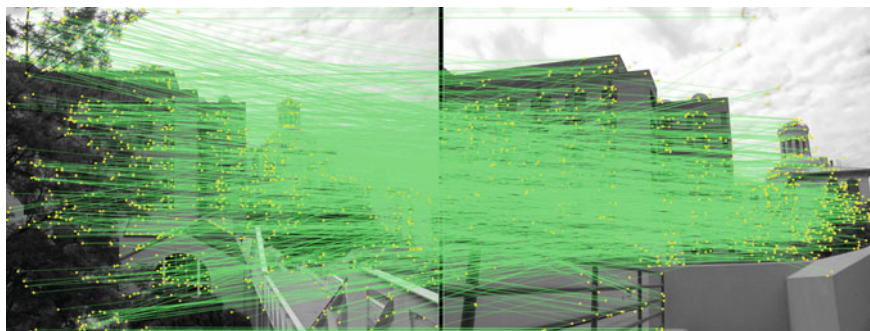
We also propose to use feature quality to evaluate geometry solutions. The traditional approach to evaluate solutions is to just use the count of supporting correspondences that pass an epilene distance threshold. With many outliers present, many false correspondences can pass the inlier test causing bad solutions to appear as good solutions. Our approach is more likely to find correct solutions because we have down weighted low-quality features whereas the traditional approach treat all features equal.

We first test our approach on a number of challenging outdoor monocular image-pairs, each with a very small fraction of inliers in the set of correspondences. The results are evaluated against manual ground-truth and compared against other robust estimation algorithms. The result is that, in comparison to the commonly used approaches, our approach finds more accurate solutions.

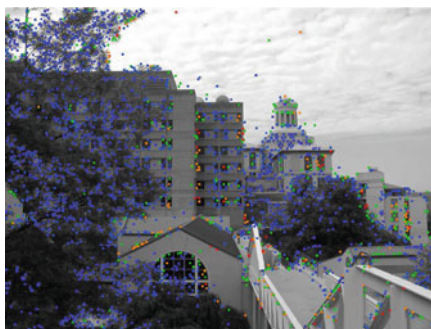
Secondly, we apply our uniqueness weighting approach for faster and accurate estimations in wide-baseline stereo visual odometry. Finally, the paper is organized by presenting the methodology, following which the results for monocular and stereo image-data are described in consecutive sections.

2 Related Work

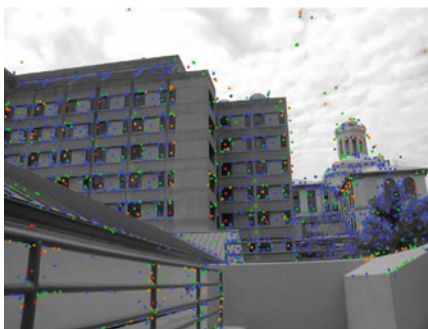
Here we present an overview of state of the art in estimation of wide baseline geometry. The task ultimately requires correlating some type of visual information between the image-pair. In the literature there are a number of different categories of information that are used for wide baseline; region based [12], edge/line based methods [9], but by far the most prevalent methods used are those using point features such as SIFT [10]. Appearance based methods, such as [8] learn to predict matchable descriptors, attaining faster geometry computation with high success rates of matching. The point-based methods are popular because firstly they have algorithms to repeatedly find the same image keypoints in a scene even when the viewing pose changes, and secondly they can form descriptions of the area surrounding the keypoints that have a reasonable level of robustness to viewing changes.



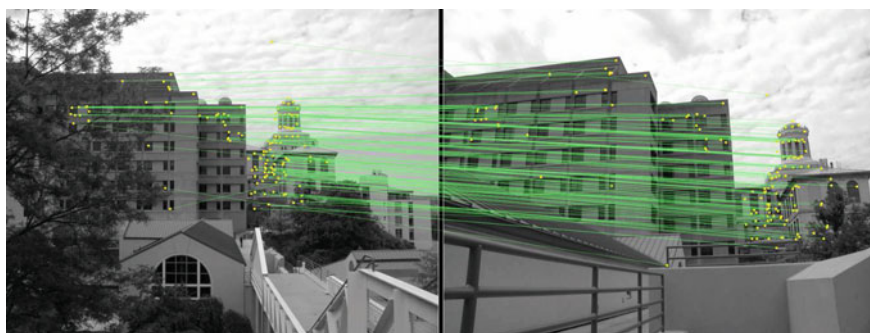
(a) Point correspondences



(b) Uniqueness left image



(c) Uniqueness right image



(d) Inliers

Fig. 1 Example of weighted epipolar geometry estimation: Campus Scene. The feature uniqueness coloring ranges from *unique* to *repetitive* as follows; red, orange, green, light-blue, blue. The low-quality repetitive foliage and many repetitive building features are downweighted

However, even though the process is comparatively robust, there are often more outliers correspondences between a wide baseline image pair than inliers. The most prevalent and successful approaches to dealing with cases where the outliers are above 50% of all correspondences is RANSAC [7] and its variants. The RANSAC process is to form a hypothesized solution to the geometry from a randomly selected minimal sample of point correspondences and evaluate the hypothesis by counting correspondences in agreement. The process is repeated with different samples until a good solution is found. To achieve higher accuracy, Locally Optimized RANSAC (LoRANSAC) [3] locally improves the estimated model hypothesis when the best model so far is found. For faster convergence, [15] precludes computation of incorrect feature matches using spatial order information in images.

There are approaches to improve RANSAC at the sampling step, where each correspondence is weighted and the likelihood of it being chosen in the sampling is adjusted [2, 6, 13, 16]. SCRANSAC [13] achieves this by ranking correspondences based on the spatial consistency of neighboring correspondences. USAC [2] combines [2, 3, 11] into a single comprehensive pipeline.

The work in [16] replaces random sampling by guided sampling, computing a weight for every correspondence based on an image correlation score. PROSAC [2] uses correspondence quality measures to speed up RANSAC based on the correlation score of keypoints. In EVSAC [4], hypothesis generation is accelerated by computing confidence values by modelling the matching scores. BEEM [6] presents another method of weighting correspondences by a correlation quality score determined by the ratio of distance of nearest to second nearest feature in the descriptor-space.

A majority of these existing approaches calculate weights from the correlation score between two features or the ratio of correlation scores to the next best correspondence. These existing weighting metrics are formed considering just one or two correspondences. This will result in a low weighting for semi-distinct features, and this weighting will be as low as the weighting for very non-distinct features. Whereas our approach is different in that it calculates the uniqueness score of a feature against all other features in the image, and will provide an appropriate weight for semi-distinct features that are similar to only a few other features. Thus our uniqueness weighting is less susceptible to the case of semi-distinct features in the scene, which are features that are similar to a small number of other features.

Furthermore the convention in these approaches is to use the number of correspondences in the consensus set that have passed an epiline distance threshold. However, when there are many outliers, many false correspondences can mistakenly pass the inlier test causing bad solutions to appear as good solutions. In our work we evaluate solutions based of the feature quality in the consensus set, which is more likely to find correct solutions because we have down weighted low-quality features whereas the traditional approach treat all features equal.

3 Sampling Using Feature Uniqueness: UNIQSAC

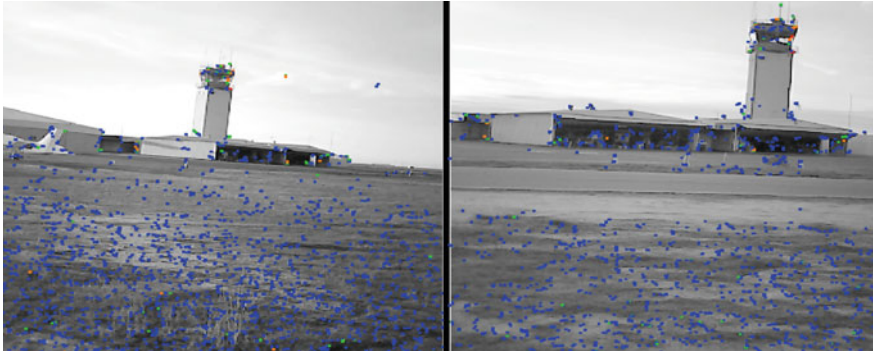
We first detail our approach to weight the sampling of correspondence subsets for the RANSAC procedure. We reiterate the necessity for weighting the sampling of correspondences to achieve timely convergence in difficult image pairs as previously demonstrated in [2, 6, 16] and propose an alternative strategy for determining the weights. The first step is to quantize the feature descriptors into distinct labels, similar to the bag-of-words approach [14]. We count how many times each type of feature appears in an image, giving us an estimate of that feature’s uniqueness. The bag-of-words approach requires a large prior database relevant to the current scene at hand and a large amount of preprocessing to cluster the database of features. We propose an alternative feature-quantization approach that requires no learning and is efficient.

Our approach is to first take a N -dimensional feature descriptor and randomly sample M -dimensional sub-features from it K times. Then, we quantize these K sub-features into K integers. The reason for a set of identifiers for each feature is that quantizing large N -dimensional features is difficult and error prone, because if any one dimension is disrupted (from any number of potential viewing disruptions; lighting, sensor noise, viewing pose changes, non-planarity, partial-occlusions etc.) then the resulting integer identifier will be wrong. So by choosing an appropriate value for M and K , the resulting set of K integers will still each on their own be a reasonable means for identifying a feature, and even though it is still likely that a number of the K integer identifiers will be erroneous because of viewing disruptions, it is likely that there will be enough identifiers for each feature that are still valid.

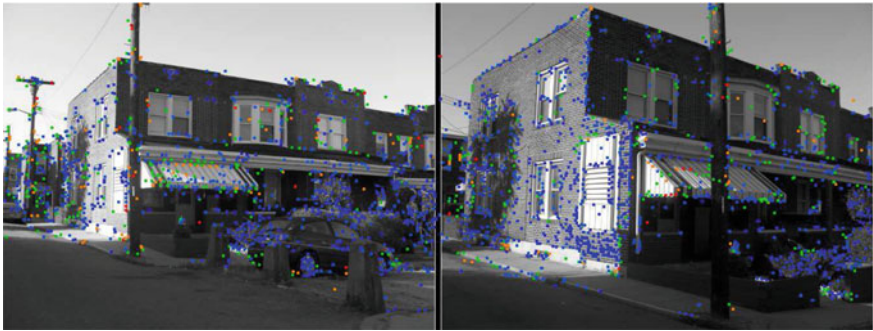
The procedure is as follows. Let f_i be the value of the i th dimension of feature, \mathbf{f} , which is a normalized value between 0 and 1. Then let $r(k, m, N)$ be a function that returns a random hasher between 1 and N , which will return the same random hasher for specific k and m values (i.e. it will pick the same dimensions to sample for each feature we are quantizing). Given a feature, \mathbf{f} , we randomly generate a sub-feature, \mathbf{s} , several times to form the set of sub-features, \mathbf{S} , where \mathbf{s}_j is the j th sub-feature. We then quantize the sub-features into an integer identifier, d , with the quantization function $q(v, p, z)$, which takes a feature value v between 0 and 1, an exponent value, p , and the number of quantizations for each dimension, z ;

$$q(v, p, z) = \lfloor v^p z \rfloor \quad (1)$$

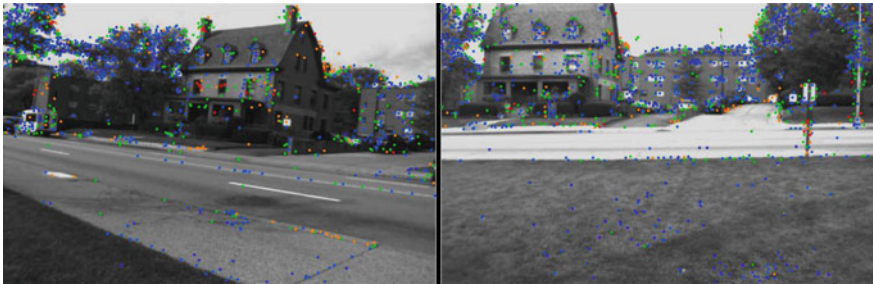
Combining the sub-features we form a set of integer identifiers, \mathbf{D} , for each feature. Here \mathbf{D}_l is the l th integer identifier. This process is defined as the QUANTIZE() function, given in Algorithm 1. Therefore by creating K integer identifiers for each feature descriptor, \mathbf{f} , we can count each time one of these identifiers appears in an image, $C(d)$, compute the most commonly appearing feature identifier, C_{MAX} , and then compute an average over all K identifiers to compute a score of how unique that type of feature is amongst the other features in that image, U_f . This details our method for calculating the uniqueness scores of a set of feature descriptors.



(a) Airport



(b) Building 1



(c) Building 2

Fig. 2 Visualization of Uniqueness Computation. The calculated uniqueness are colored for the features detected in a set of challenging outdoor image pairs. Coloring ranges from *unique* to *repetitive* as follows; red, orange, green, light-blue, blue. The airport scene noticeably has many low-quality ground features downweighted by our technique. Building scenes have many repetitive low-quality features from brickwork and foliage. Whereas most of the high-quality features have been correctly identified and given higher weights

Algorithm 1 Uniqueness Sampling and Consensus (UNIQSAC) Algorithm.

<p>Input: Set of features \mathbf{F}</p> <p>Output: Set of uniqueness values \mathbf{U}</p> <p>forall \mathbf{f} in \mathbf{F} do</p> <p style="padding-left: 20px;">$\mathbf{D} = \text{QUANTIZE}(\mathbf{f})$</p> <p style="padding-left: 20px;">$\mathbf{U}_{\mathbf{f}} = \sum_{k=0}^K \frac{1 - C(\mathbf{D}_k)}{C_{\text{MAX}}}$</p>	<p>function $\text{QUANTIZE}(\mathbf{f})$</p> <p>for $k = 0$ to K do</p> <p style="padding-left: 20px;">$d = 0$</p> <p style="padding-left: 40px;">for $m = 0$ to M do</p> <p style="padding-left: 60px;">$i = r(k, m, N)$</p> <p style="padding-left: 60px;">$\mathbf{s}_m = \mathbf{f}_i$</p> <p style="padding-left: 60px;">$d = d + q(\mathbf{s}_m, e, z)z^m$</p> <p style="padding-left: 40px;">$\mathbf{D}_k = d$</p> <p>end function</p>
---	---

Figure 2 provides a visualization of uniqueness computation in three monocular wide-baseline settings. The coloring follows a *heat-map* color scheme, where from red is a unique feature and blue is a frequently occurring feature. Here it is clearly identified the features are that are unique and which are low-quality. In the airport scene it is noticeable that most of the ground features have been correctly down-weighted as they are low-quality features, and many of the features on the hangers and tower have correctly been identified as the useful *unique* features. In the building scenes, most of the foliage features and many of the features on the brickwork, repeated windows have been suitably down-weighted.

4 Evaluating Hypotheses Using Feature Uniqueness

The previous section described the sampling stage of geometry estimation, the next stage of the RANSAC framework is to evaluate the resulting model hypothesis. Traditionally, almost all methods use the cardinality of the consensus set [7] (*inlier* set), \mathbf{c} , as the metric to decide the quality of the model, which has limitations in being able to determine good hypotheses from bad.

The common approach is to test if a feature correspondence, i , is in consensus and therefore an *inlier*, is if its distance, ϵ , away from the epipolar line computed from the n th model is less than the threshold, τ :

$$\mathbf{c}_n = \forall \mathbf{c}(i) \in \mathbf{C}, \epsilon_i < \tau \quad (2)$$

Then given N RANSAC iterations, the final estimated model, \hat{m} , is simply the model with the largest consensus set as follows:

$$\hat{m} = \arg \max_n (|\mathbf{c}_n|) \quad (3)$$

where \mathbf{c}_n is the consensus set from the n th RANSAC iteration. However, there can be many correspondences that pass the inlier test that are in fact outliers. Therefore, there can be poor geometry hypotheses that appear good because of many false correspondences in the consensus set.

As a result of the limitations of using just the cardinality of the consensus set as the evaluation metric, we have developed a new metric to evaluate RANSAC hypotheses, that is a measure of the quality of the features in the consensus set. Our approach is less likely to find false solutions because we down-weight low-quality features in the evaluation process, whereas the traditional approach treats all features as equal. We propose to use feature weighting approach described in the previous section as a measure of quality, and use the sum of the inliers' feature weights as the metric for hypothesis evaluation. So our formulation for the best model becomes:

$$\hat{m} = \arg \max_n \left(\sum_{\forall \mathbf{f}_i \in \mathbf{c}_n} w(\mathbf{f}_i) \right) \quad (4)$$

Now we plot two graphs comparing our new metric against conventional inlier count as shown in the Fig. 3. The first graph Fig. 3a, shows the correlation between size of the consensus set (inliers) and the actual true inliers (validated inliers), recorded for 1,000,000 iterations of RANSAC (computing 1,000,000 iterations would most likely take too long in practice, but used here to give a clear picture of the relationship to the true number of inliers) on the wide baseline image pair seen in Fig. 2a. Similarly, the second graph in Fig. 2b shows the correlation between summation of inlier weights and validated inliers.

It is obvious from Fig. 3b, that the new metric produces a far better correlation to true inliers when compared to the conventional inlier count. Although we see an upwards trend in Fig. 3a, relating number of inliers to number of validated inliers,

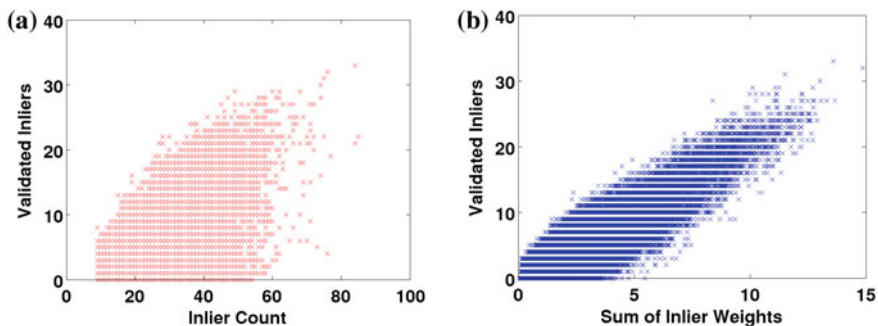


Fig. 3 Hypothesis Evaluation and Accuracy score for a million iterations. **Left:** At higher inlier counts, RANSAC is inaccurate with few validated inliers. **Right:** Feature Uniqueness provides a linear correlation between summation of inlier weights and validated inliers

there is an alarming variance in this distribution, where some hypotheses with many overall inliers have very few validated inliers. This illustrates, that our new metric is far less likely at wrongly selecting a poor hypothesis.

5 Monocular Wide-Baseline Geometry Estimation

In this section, we provide a detailed description of implementing feature uniqueness weighting for sampling correspondences in a RANSAC-type geometry estimation for wide-baseline monocular images. The correspondence set, \mathbf{C} , is the set of matched features between \mathbf{F}_1 and \mathbf{F}_2 , the two features sets from the image pair. From this correspondence set, we take a minimal subset, \mathbf{s} , of the entire correspondence set using a weighting function that computes the likelihood of \mathbf{f} being sampled, $w(\mathbf{f})$;

$$w(\mathbf{f}) = U_{\mathbf{f}}^2 \quad (5)$$

The minimal sample set is used to estimate our hypothesized model which for us estimating epipolar geometry is the fundamental matrix. The cardinality of \mathbf{s} being 8, because we are using the 8-point algorithm [7]. When selecting the 8 point-correspondences of \mathbf{s} from \mathbf{C} we use a Monte-Carlo sampling strategy based on the weights of the two features $\mathbf{f}_1(i)$ and $\mathbf{f}_2(i)$ of a specific correspondence, i :

$$p(\mathbf{C}(i)) = w(\mathbf{f}_1(i))w(\mathbf{f}_2(i)) \quad (6)$$

6 Experimental Results: Monocular Image-Data

We collected a variety of particularly challenging wide baseline image pairs in a number of different outdoor settings to evaluate our approach. Here we present results collected from four different scenes, shown previously in Figs. 1 and 2.

We evaluate our UNIQSAC approach against three standard epipolar geometry estimation methods; RANSAC [7], LoRANSAC [2] and BEEM [6]. All these methods sample a small number of sparse feature point correspondences, fit an epipolar geometry model, evaluate the model, store if it is the best found and iterate. As an input, we use the SIFT features [10] for keypoint detection and SIFT feature descriptor for each of these keypoints.

The parameters for feature quantization, given in Eq. 1, are set as per the following: Dimension of sub-feature M as 6, Number of sub-features K as 6, Dimension of feature N as 128, Exponent p as 1, Number of quantization z as 2. These parameters are set for a coarse quantization making it suitable to distinguish frequent features from the more unique features. Taking the quantizations and computing a uniqueness for each feature according to Algorithm 1, yields a visualization of uniqueness shown in Fig. 2.

Table 1 Epipolar geometry estimation results

Image-pair	Statistic	RANSAC	LoRANSAC	BEEM	UNIQSAC
Airport	Correspondences	1237	1237	1237	1237
	Inliers	82.5	98.0	102.4	96.66
	Validated inliers	22.3	29.9	36.4	43.2
	Epiline error (pix)	20.0	11.4	4.7	4.1
Building 1	Correspondences	1204	1204	1204	1204
	Inliers	187.7	214.5	216.9	217.0
	Validated inliers	88.9	113.1	124.5	125.8
	Epiline error (pix)	7.9	5.4	5.2	3.6
Building 2	Correspondences	1192	1192	1192	1192
	Inliers	146.6	181.3	177.7	182.5
	Validated inliers	87.8	116.0	115.3	125.0
	Epiline error (pix)	4.1	2.8	2.3	1.7
Campus	Correspondences	1451	1451	1451	1451
	Inliers	143.7	157.2	162.8	175.7
	Validated inliers	103.3	121.6	130.7	128.7
	Epiline error (pix)	14.9	5.9	8.0	4.5

Once uniqueness has been computed, the next step is to find point correspondences. We compute the correspondences between the pair using a KD-tree [1] (using 250 as the maximum number of leaves to visit). In each of the image pairs the number of inliers is low, somewhere between 10 and 15%. We evaluate the performance of the algorithm by manually designating a set of groundtruth correspondences (usually around 30) between image pairs. We compute a *groundtruth* fundamental matrix using the standard 8-point algorithm and use this fundamental matrix to validate how many of the inliers found also agree with the groundtruth fundamental matrix (we call these the *validated inliers*). We also compute an epipolar line from the estimated fundamental matrix and compute the mean distance of the groundtruth correspondences to this line (we call this the *epiline error*).

To maintain a fair comparison we fix the number of iterations for all algorithms to the same number: 50,000. For LoRANSAC we use an inner loop of 5,000 iterations, where each inner iteration contributes to 50,000 iterations in total. We run 50,000 iterations for each algorithm, 10 times on each image-pair, compute the mean and present these in Table 1.

We highlight in bold the best performing algorithm for the number of validated inliers and the groundtruth epilene error. Overall our *uniqueness* approach to weighted sampling performs better than all other algorithms, as it provides more accuracy in terms of epilene distance and has more validated inliers.

We can see that in some of the tests that some of the algorithms produce more inliers, however as we have discussed throughout this paper, the number of inliers

does not necessarily mean a better solution (as false correspondences can be counted as inliers). The important statistics are the number of validated inliers and the epilene distance.

7 Wide-Baseline Stereo Camera Pose Estimation

In this section we extend the monocular method to solve geometry between sets of stereo pairs. The results presented in previous section provides significant motivation to apply our Uniqueness Sampling and Consensus (UNIQSAC) technique to compute a 6 Degree-of-Freedom pose for a wide-baseline stereo camera.

Our Stereo image dataset comprises of plant images captured in a wide-baseline stereo setup from Grape Vineyards, Sorghum Fields and Apple Orchards. Due to repetitive features present in plant foliage and background clutter, sampling unique features becomes essential in computing motion estimates accurately within a feasible number of iterations. To compute a complete 6 DoF motion, we implement a stereo visual odometry pipeline consisting of three stages: Feature Extraction and Matching, Uniqueness score computation and Geometry estimation.

7.1 Feature Extraction and Matching

With a moving stereo camera setup, we obtain four images at any given time: left and right images of two consecutive frames. To compute camera transformation, we require feature matches between all four images. Starting from Previous Left image, we look for correspondences in Current Left Image. After establishing correspondences between Previous and Current Left image pair, we utilize these matched features as templates to search for corresponding features in the Previous and Current right images along the epipolar scanlines. The template image matching uses a Normalized Cross Correlation Score. For feature matching process, keypoints are detected using SIFT features. Subsequently, SIFT descriptors are extracted and matched using K-nearest neighbor technique.

7.2 Uniqueness Score Computation

Before computing the uniqueness score, features are quantized keeping the same parameters as described in the earlier section. In addition to the Uniqueness score U_{score} , we also consider depth consistency and distance-ratio of matched features in form of a new hybrid score T_{score} . This hybrid score is used to assign weights to all the matched features as given in the Eq. 7. Depth consistency D_{score} is computed by calculating the differences in depth values from Previous Left- Current Left and

Current Left-Previous Left, and then normalizing the difference values. The distance-ratio R_{score} of matched features is determined by a process described in [6].

$$\mathbf{T}_{score} = U_{score} + D_{score} + R_{score} \quad (7)$$

7.3 Geometry Estimation

Once we have pixel coordinates of all matched features, we can apply triangulation to acquire two sets of 3D coordinates. The camera motion (r,t) can be computed by minimizing the sum of reprojection errors using a Gaussian-Newtonian optimization. Our motion estimation scheme from sparse feature matches is similar to the method presented by Geiger et. al [5]. Instead of randomly drawing 3 correspondences, we replace RANSAC with our Uniqueness sampling technique to increase the probability of selecting high quality features. For testing model hypothesis, we use our metric given in Eq. 4 to obtain the final inlier set.

8 Experimental Results: Stereo Image-Data

To test the performance in grape and sorghum image dataset, we consider both time-efficiency and accuracy for evaluation. Since the run-time for execution of one cycle of an algorithm can be relative, a good way to measure the time-efficiency is to check the number of iterations the algorithm takes to converge.

In our setup, the robot is equipped with a stereo camera moving at an average velocity of 0.45 m/s, capturing images at 5 frames/sec facing a plantation wall at a distance of 3 feet. Figure 4 illustrates our complete Stereo Visual Odometry pipeline for the grape dataset Fig. 4a and sorghum image dataset Fig. 4b. We can see that the raw images have considerable portions that will yield low quality features—ground, sky and plant foliage. As a result, many feature correspondences are incorrect. Computing a hybrid Uniqueness score downweighs a majority of features extracted from ground and plant foliage. Only a tiny fraction of these correspondences are correctly identified as inliers (inlier ratio <9%). Similar results from Apple orchard dataset are illustrated in Fig. 5.

Our algorithm UNIQSAC is tested against the three standard methods: RANSAC, LoRANSAC and BEEM. The metric for evaluation is the validated inlier score (total validated inliers), and computed for 1000 and 10,000 iterations. To ensure robustness and repeatability, the scores are averaged over 1000 times. The results presented in Fig. 6 show that our proposed technique provides 25% (Grape) and 80% (Sorghum) improvement in Accuracy-Time Efficiency as compared to RANSAC.

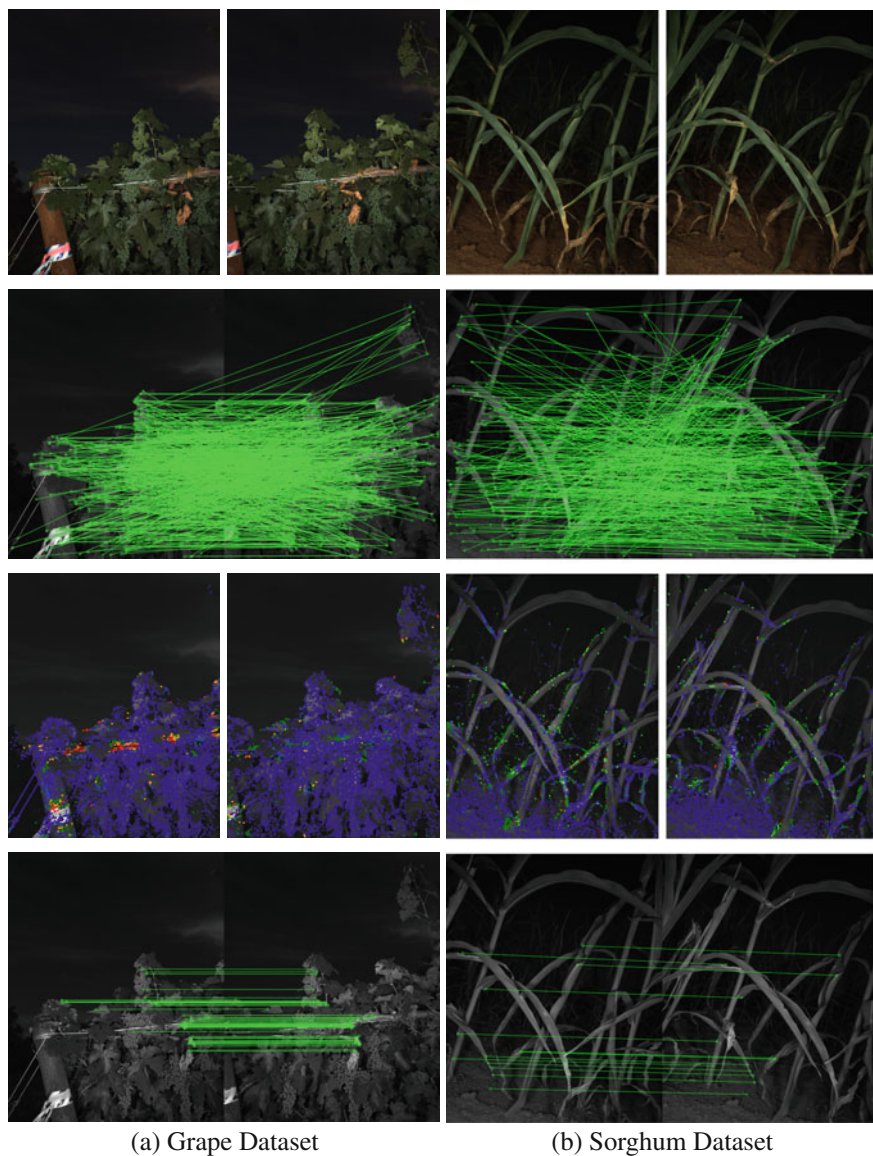


Fig. 4 Wide-baseline stereo geometry estimation. An example of wide-baseline geometry estimation on stereo image-pairs. Illustrations on the left column are obtained from the Grape dataset and right column from the Sorghum dataset. **First Row**: Raw image-pairs as captured from the Stereo Camera. **Second Row**: Correspondence Matching using K-nearest neighbors, between Previous Left and Current Left Image. **Third Row**: Feature uniqueness computation and colored heat map for Previous Left and Current Left Image. **Fourth Row**: Plotted Inliers are obtained using our Uniqueness sampling technique

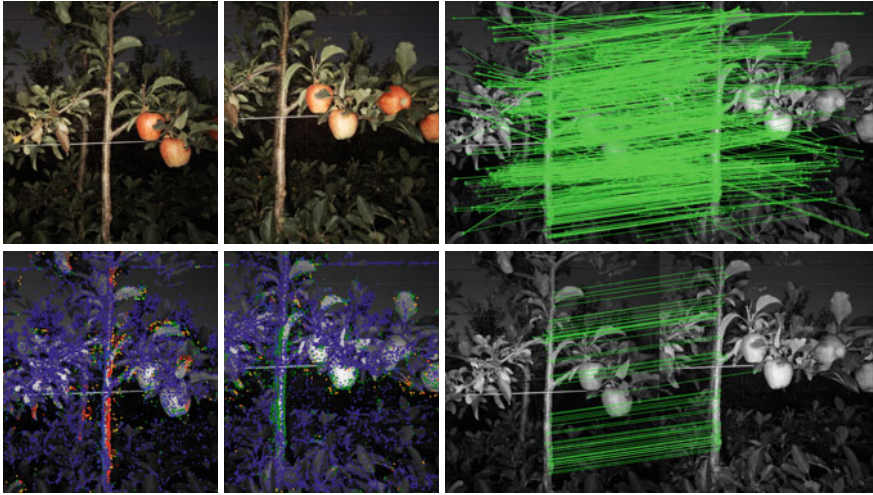


Fig. 5 Apple Orchard Dataset. **First Row-Left to Right:** Raw image-pairs, Correspondence Matching using K-nearest neighbors. **Second Row-Left to Right:** Uniqueness computation as colored heat maps, plotted inliers using UNIQSAC, between Previous Left and Current Left Image

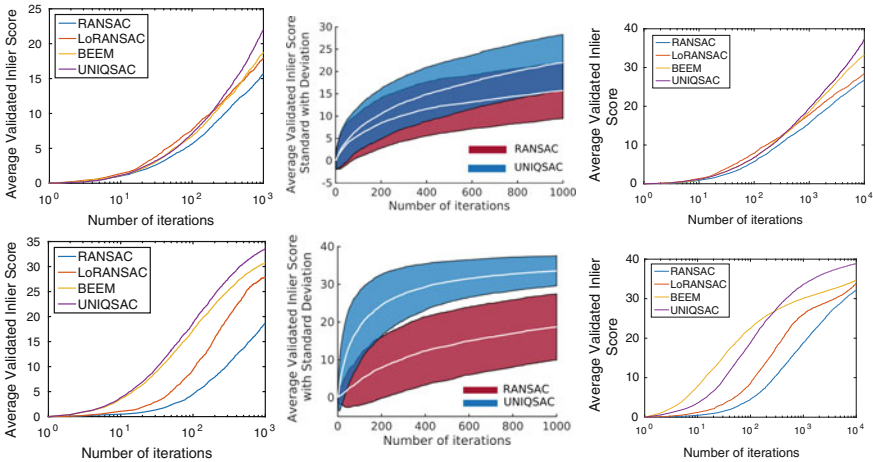


Fig. 6 Time Efficiency-Accuracy Score. Each figure in top row left-to-right are results acquired from the Grape Dataset. **Left:** Comparison for 1000 iterations. **Centre:** Comparison for RANSAC versus UNIQSAC, with standard deviation from the mean validated inlier score. **Right:** Comparison for 10,000 iterations. Similarly, each figure in the bottom row left-to-right, are results obtained from the Sorghum Dataset

9 Conclusion

This paper presented a new weighted sampling method called UNIQSAC, based on feature quality for computing epipolar geometry between wide baseline image pairs. Also presented is a new metric for evaluating model hypothesis based on the feature quality of the consensus set that performs better than simply counting the cardinality of the consensus set. Together, these two methods were demonstrated in a variety of different outdoor environments, where the low-quality repetitive features are correctly down-weighted and the more unique structural features are correctly up-weighted, resulting in more accurate solutions.

To evaluate the performance of our methods, a set of monocular and stereo image-pairs was ground-truthed and the results were compared against standard robust estimation approaches like RANSAC, LoRANSAC, and BEEM. It was demonstrated in monocular settings, that our methods produces in all but one case the most accurate estimates. These results are further validated in the stereo settings, where we acquire the most accurate and time-efficient estimates while operating at very low inlier ratios ($<9\%$). Utilizing feature quality to guide random samples, opens up several possibilities for future work. In particular, we are interested to extend our methods to obtain faster and accurate registration of dense 3D point clouds.

Acknowledgements This work is supported under award numbers; USDA 2015-51181-24393, USDA 2016-67021-24535 and ARPA-e 1830-219-2020937.

References

1. Beis, J.S., Lowe, D.G.: Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. Proceedings, pp. 1000–1006. IEEE (1997)
2. Chum, O., Matas, J.: Matching with prosac-progressive sample consensus. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 220–226. IEEE (2005)
3. Chum, O., Matas, J., Kittler, J.: Locally optimized ransac. In: Joint Pattern Recognition Symposium, pp. 236–243. Springer (2003)
4. Fragoso, V., Sen, P., Rodriguez, S., Turk, M.: Evsac: accelerating hypotheses generation by modeling matching scores with extreme value theory. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2472–2479 (2013)
5. Geiger, A., Ziegler, J., Stiller, C.: Stereoscan: Dense 3d reconstruction in real-time. In: Intelligent Vehicles Symposium (IV) (2011)
6. Goshen, L., Shimshoni, I.: Balanced exploration and exploitation model search for efficient epipolar geometry estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(7), 1230–1242 (2008)
7. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer vision*. Cambridge University Press (2003)
8. Hartmann, W., Havlena, M., Schindler, K.: Predicting matchability. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9–16 (2014)
9. Lee, J.A., Yow, K.C., Chia, A.Y.S.: Robust matching of building facades under large viewpoint changes. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1258–1264. IEEE (2009)

10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
11. Matas, J., Chum, O.: Randomized ransac with sequential probability ratio test. In: Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005, vol. 2, pp. 1727–1732. IEEE (2005)
12. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. *Image Vis. Comput.* **22**(10), 761–767 (2004)
13. Sattler, T., Leibe, B., Kobbelt, L.: Scramsac: improving ransac’s efficiency with a spatial consistency filter. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2090–2097. IEEE (2009)
14. Sivic, J., Zisserman, A.: Video google: efficient visual search of videos. In: *Toward Category-level Object Recognition*, pp. 127–144. Springer (2006)
15. Talker, L., Moses, Y., Shimshoni, I.: Using spatial order to boost the elimination of incorrect feature matches. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1809–1817 (2016)
16. Tordoff, B., Murray, D.W.: Guided sampling and consensus for motion estimation. In: *European Conference on Computer Vision*, pp. 82–96. Springer (2002)