FT-RANSAC: Towards Robust Multi-Modal Homography Estimation

Adam Barclay, Hannes Kaufmann
Interactive Media Systems Group
Vienna University of Technology, Austria

***Abstract*—As the golden standard in robust estimation, the classic RANSAC approach has undergone extensive research that contributed to further enhancements in run-time performance, robustness, and multi-structure support to name a few. Yet, the accelerating growth of multi-modal co-registered datasets requires a new adaptation of the RANSAC algorithm. In this paper, we propose a multi-modal fault-tolerant extension to RANSAC, termed FT-RANSAC, with a model-independent tolerance to degenerate configurations. Besides building on state-of-the-art RANSAC variants, such as PROSAC, our approach introduces a Hough inspired dimensionality reduction and consistency voting processes, to enable robust estimation in the presence of non-homogenous multi-modal correspondence sets. Through experimental evaluation using homography estimation of RGB-D data, we demonstrate that our approach outperforms the classic single-modality RANSAC in robustness and tolerance to degenerate configurations. Finally, the proposed approach lends itself to parallel multi-core implementations, and could be adapted to specialized RANSAC extensions found in the literature.**

Keywords*—*RANSAC, LiDAR, RGB-D, Registration, Fusion, Robust Estimation, Homography Estimation, Point Cloud.

I. INTRODUCTION

Co-registered intensity and range imaging is one example of complementary multi-modality that has been extensively explored for 3D reconstruction as in [1], yet not fully exploited in robust estimation, where degenerate configurations and varying characteristics of the different modalities pose a challenge. Yet, solutions to degenerate configurations and scene-sensor issues could benefit the most from multi-modal based approaches. The classic RANdom SAmple Consensus [5] and its variants [3,4,6,7] have been the most commonly applied approaches to robust estimation, with homogenous correspondence sets; where correspondence elements are of the same modality and share the same properties and metrics, as a precondition. Non-homogenous correspondences, with varying inlier characteristics due to modality, sensor noise or descriptor-dependent similarity measure, cannot be directly applied to RANSAC. Although various extensions have been proposed for real-time adaptation of the RANSAC algorithm, such as PROSAC[7], and increased robustness through geometric constrains [3,4], multi-modal adaptation of the RANSAC algorithm has not been as extensively researched, particularly for non-homogenous data sources. In this paper, a multi-modal fault-tolerant extension to RANSAC, termed FT-RANSAC is proposed, with PROSAC[7] inspired guided sampling, concurrent instances of the classic RANSAC[5] approach, and a Hough[16] inspired dimensionality reduction and consistency voting stages to enable the computation of the best model across the competing multi-modal solutions. Through experimental evaluation, FT-RANSAC has demonstrated its ability to exceed single modality state-of-the art approaches in stability and tolerance to degenerate configurations, despite the variable multi-modal input characteristics, and with negligible run-time overhead.

II. RELATED WORK

The classic RANSAC approach [5] consists of two stages: a hypothesis generation stage where random samples are drawn from a correspondence set to act as a seed for model generation, and a hypothesis verification stage where the generated model is verified against the correspondence set, resulting in an error cost – a cost that is to be optimized. RANSAC's standard termination criteria is based on the number of iterations required to satisfy a probability **P** of a correct model, given the inlier ratio – a ratio that is not a-priori known. Clearly, non-homogenous correspondences with varying characteristics would disrupt RANSAC's inlier vs. outlier classification performance and would not result in a correct model. Although RANSAC, by definition, is a robust estimation approach, its robustness is dependent on the techniques applied in both the sampling process during the hypothesis generation stage and the error score computation of the hypothesis verification stage. Due to the challenging aspect of sampling an inlier-only set of a size that is also sufficient for model computation, robustness oriented approaches to RANSAC has focused on single-modality model-specific geometric constraints as in DEGENSAC[3], SCRAMSAC[6], and SSCA [4].

The DEGENSAC [3] algorithm is one of the best known extensions to the classic RANSAC approach. Its goal is to generate robust models under the epipolar geometry (EG) constraint, despite degenerate (H-degenerate) conditions that are caused by the presence of a dominant plane. With this extension, the 7-point EG algorithm is used in the verification step with an additional check for H-degeneracy. If H-degeneracy is detected, a plane-and-parallax algorithm is applied to compute the fundamental matrix, along with its inlier support.

Spatially Consistent RAndoM SAmple Concensus, termed

SCRAMSAC [6], applies a spatial consistency check by prefiltering correspondences to those where at most one potential match exists per correspondence within a given radius. The goal is to have a smaller set of high quality correspondences. The result is reported to be of similar quality to standard RANSAC, but faster than both RANSAC and PROSAC, due to the reduced set of high quality correspondences. This approach is essentially a form of non-max suppression, and does not guarantee a non-degenerate set of correspondences.

RANSAC with fusion, termed RANSAC-f [2], is a single-modality fusion approach for homogenous correspondence sets, where an ordered list of best RANSAC generated models is maintained, and the sorting is based on the error score per model. An iterative process is then initiated, where parameters of the top two best models are averaged and a new error score is computed for the whole set using the new parameters of the fused model. If the re-evaluated error score is better than that of the previous best model, then the previous best model is replaced with the new model. Otherwise, the 2nd best model is removed, and the iterative process is repeated using the current two best models. For an early termination without full evaluation of each model in the ordered list, the update ratio of the best model could be applied given a threshold that is independent of the outlier ratio.

III. FT-RANSAC: The Algorithm

As could be observed from the related work section, degenerate configurations are addressed on a per-model basis, and are computed from the correspondence set of a single modality. In this section, we describe our proposed approach for a fault-tolerant RANSAC, termed FT-RANSAC, and how the algorithm overcomes the challenges posed by the varying thresholds, noise, and differing characteristics of the modalities to be incorporated in the fusion process. The minimum requirement of the FT-RANSAC approach is that correspondence sets need to represent the same sought after model. To enable modality-independent processing and fusion, the following stages are introduced:

A. Stage 1: Guided Sampling with a Cutoff Filter

In the classic RANSAC approach, no assumption or a priori information is considered regarding correspondence elements in a given correspondence set. As a result, all correspondences of a set are assumed to be of the same quality and accuracy in the hypothesis generation stage. This assumption does not necessarily hold as pointed out by PROSAC[7], noting that a correspondence set's similarity measure embodies the quality of each element. In FT-RANSAC, we apply a similar approach to PROSAC, where elements of the correspondence set are ordered based on the correspondence estimation similarity measure. Unlike PROSAC, the progressive approach to the sampling process is replaced with a modality-independent threshold, and is used to extract the best correspondences for the hypothesis generation

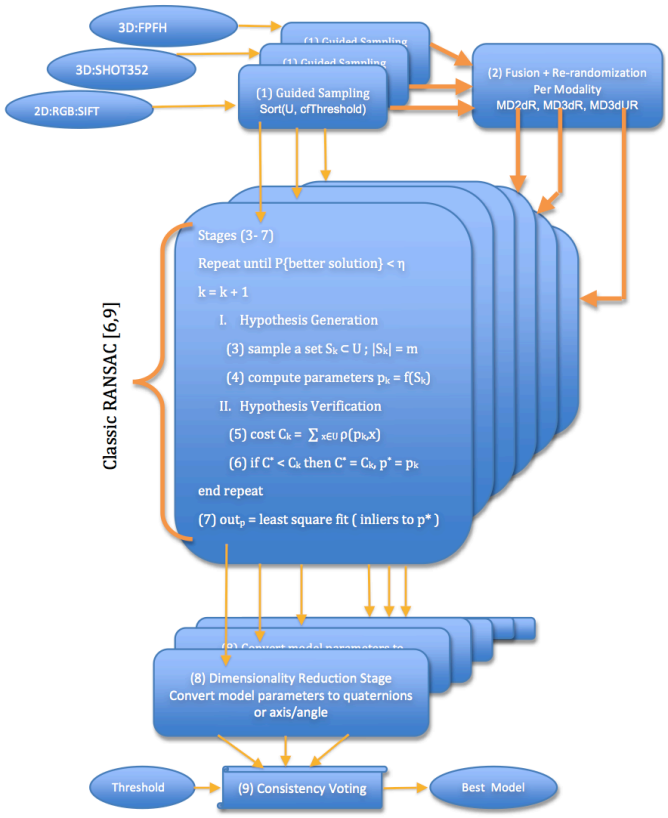

Fig. 1. Illustrative diagram of the proposed FT-RANSAC approach.

stage. This threshold is the smaller value of either (a) multiple of the minimum number of correspondences that are required to compute a model, or (b) best x% of the correspondences. In our experiments, the best 20-30% of correspondences consistently provided the best results.

B. Stage 2: Single-Modality Fusion & Re-randomization

As pointed out in [7,8], high quality inliers tend to be close together or lie on the same degenerate structure. Alternatively, the number of high quality inliers may simply be insufficient for model extraction and verification. To enhance the tolerance to this type of failure within a single modality, correspondence sets of Stage 1 that represent different keypoint and descriptor types, therefore, less likely to be close together or lie on the same degenerate structure, are fused and re-randomized as an optional input to Stages 3.

C. Stages 3-7: Multi-Core Instances of the Classic RANSAC

For an extensible and multi-core ready design, our approach forms a wrapper around the classic RANSAC algorithm, stages 3-7 of Figure 1. Each instance of RANSAC, or one of its variants, accepts as an input a single modality correspondence set from stages 1 or 2. The output of each RANSAC instance is passed then to the dimensionality reduction stage of the same instance.

D. Stage 8: Dimensionality Reduction

The classic RANSAC approach and its variants output a model represented by its parameters, a 9-parameter

transformation matrix in the homography estimation case. With the presence of competing models, each representing a different modality, precision and noise characteristics, a direct parameter-wise comparison is not feasible. As a result, a Hough[16] inspired dimensionality reduction stage is introduced, where each 9-parameter transformation matrix is converted into a set of geometrically meaningful 3-element axis-angle values, or alternatively, a set of 4-element unit quaternion. This conversion enables an efficient and intuitive comparison, clustering, and fusion as discussed in the next section.

E. Stage 9: Multi-Modal Consistency Voting

In the classic RANSAC algorithm, the best model is formed from the least square fit of the set of inliers with the largest support, and with the implicit assumption of homogenous inlier properties. With FT-RANSAC, the competing models do not have to be the product of the same modality, but are assumed to represent the same model. As described in the dimensionality reduction stage, each of the competing models as computed by the classic RANSAC approach is converted into a set of geometrically meaningful values of the same units, and therefore, could be compared across the different modalities. In the voting stage, the competing models are classified into clusters based on the anticipated error tolerance threshold. The cluster containing the most number of models that are within the error tolerance are used to compute the best model. Other approaches that we examined include (1) An adaptive technique that superimposes a Gaussian distribution on the vote results. (2) The selection of the best quality descriptor types as cluster centroids. (3) The selection of one cluster centroid per modality. After empirical evaluation of each of the clustering schemes, a fixed error tolerance threshold of 0.2 degrees was found to best represent our sensors, and to provide the best results reported in this paper. The other mentioned clustering approaches could be valuable if large variations in the input are to be expected, or if a-priori information about the quality of the utilized descriptors or modalities are to be used.

I. EXPERIMENTAL RESULTS

A. Experimental Setup

The experimental evaluation of FT-RANSAC was performed on multi-modal image pairs, each consisting of co-registered intensity-range images. Each image pair consists of two different shots of the same scene with sensor noise. An in-plane rotation of 0-60 degrees with 5-degree incremental resolution is then applied to one image in the pair to establish the ground truth. Feature based correspondence estimation is used to generate RANSAC's input. In the 2D modality, SIFT[14], SURF[15], ORB[9] and BRISK[10] were applied to generate 4 different sets of correspondences. In the 3D modality, FPFH33[11], SHOT352[12] and SHOT1344[12,13], were used to generate the correspondence sets in the point cloud domain. Derived and fused single-modality multi-descriptor correspondences are generated in stage 2, including: MD2dR: RANSAC applied to a combined set of the top 20% of intensity based SIFT, SURF, ORB and BRISK descriptor correspondences. MD3dR: RANSAC applied to a combined set of the top 20% 3D FPFH33, SHOT352 and SHOT1344 descriptor correspondences. MD3dUR: RANSAC applied to a combined set of unique 3D FPFH33, SHOT352 and SHOT1344 descriptor correspondences. Lastly, our approach, FT-RANSAC, represents the results of dimensionality reduction and consistency voting as it is applied to both 2D intensity images and 3D point cloud correspondence sets. The evaluation error is the magnitude of the difference between the known ground truth and RANSAC's estimate with and without our approach.

Correspondence Modality	**Mean and variance of homography estimation Error**	
	Mean	***Variance***
MD2dR	0.0807	0.0883
MD3dR	0.0857	0.0280
MD3dUR	0.0873	0.0545
FT-RANSAC	0.0460	0.0263

Fig. 2. Mean and variance of the magnitude of FT-RANSAC's homography estimation error as compared with those of 2D-only and 3D-only measures over a 0-60 degree range.

B. Parameter Settings

In the 2D modality, RANSAC threshold applies to the inlier rate in the pixel domain, and is set to 0.1. In the 3D modality, RANSAC threshold applies to the inlier rate of the point cloud domain and is set to 0.01, indicating the higher noise level to be expected. Our proposed approach introduces a new threshold that represents the desired error tolerance of the anticipated model transformation, in the form of an axis-angle rotation. This error tolerance threshold, which is tuned to the scene-sensors' properties by setting it to 0.2 degrees, is applied to the consistency voting process, to cluster the best models from the competing solutions.

C. Evaluation and Analysis

In the evaluation process, FT-RANSAC was compared against state-of-the-art single-modality descriptor based approaches, for (1) Estimation stability and (2) Tolerance to degenerate conditions.

As shown in Figure 2, our multi-modal FT-RANSAC approach has been tested against a set of combined single modality pixel space correspondences, MD2dR, a set of combined single modality point cloud correspondences, MD3dR, and a set of combined single modality, high quality, unique correspondences in the point cloud space. Our approach, where the consistency voting stage enables a multi-modal clustering of the best models across the 2D pixel space and the 3D point-cloud space, clearly exhibits a much lower mean error than any of the other approaches. Furthermore, as shown in figure 2, the error variance of our proposed approach is lower than that of the classic RANSAC derived models in the pixel space or the point cloud space for the same data set. The smaller mean error and variance, reflects the stability of our approach given the variable multi-modal noise sources, and contributes to smaller drift in mapping and navigation applications. One noteworthy observation is the increase in the mean error and variance of the MD3dUR set; where the

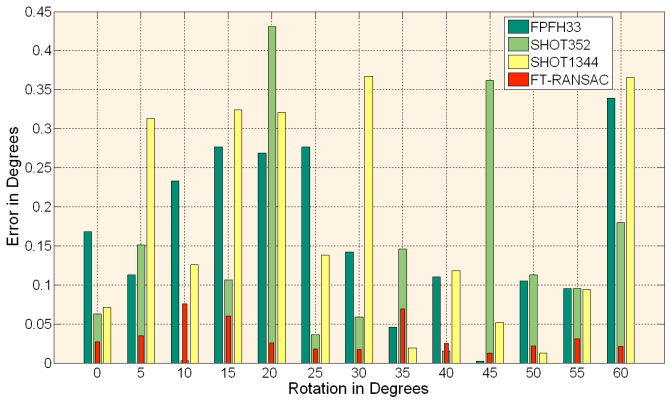

Fig. 3. A comparison of FT-RANSAC's estimation to 3D state-of-the-art.

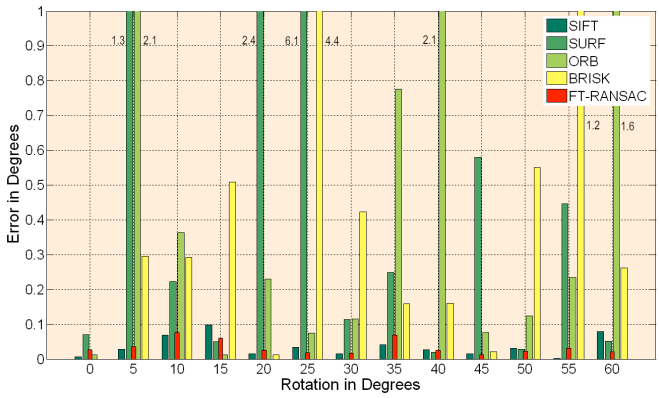

Fig. 4. A comparison of FT-RANSAC's estimation to 2D state-of-the-art.

uniqueness characteristic contributed to a much smaller set of high quality correspondences - a goal similar to that of [6,7] among others. Although having a smaller set of high ratio inliers could be viewed as advantageous, particularly if the applied correspondences are robust against degenerate configurations, noise tends to influence the final estimate as the size of the set gets smaller. Figures 3 and 4 show further comparisons against state-of-the-art 2D and 3D descriptors, and how FT-RANSAC outperforms in both stability and noise tolerance, with consistent results in the 0.0X error range, a level of stability that single-modality estimation, including those based on SIFT were not able to consistently achieve throughout the tested range. One such example is SIFT at 15 degrees of Figure 4, with a 0.1 degrees of error, and the potential impact on accumulated error during mapping and navigation.

For increased tolerance to degenerate configurations, we've tested FT-RANSAC against a combined intensity-range environment with a degenerate intensity modality. Both MD3dR, and MD3dUR were able to extract the correct model, albeit with higher levels of noise as could be expected, whereas the 2D-only approach simply failed. Furthermore, the multi-descriptor approach has demonstrated its benefits in degenerate configurations through increased accuracy of the homography estimation, when the re-randomization stage incorporates a 2-3x multiple of the highest quality correspondences, aiding the sampling process.

I. Conclusion

In this paper, we have presented a novel fault-tolerant extension to RANSAC; termed FT-RANSAC, that enables robust homography estimation using multi-modal non-homogenous correspondence sets. With two layers of model-independent enhanced tolerance to degenerate configurations, in the form of multi-descriptor support, and multi-modal support, FT-RANSAC addresses the challenges common to approaches of single-type sensor and single-type descriptor correspondences, particularly for smaller sets of high quality inliers, that are potentially degenerate. Through experimental evaluation, FT-RANSAC has demonstrated its ability to exceed single-modality state-of-the art approaches in stability, and error variance, despite variable input noise, and with negligible run-time overhead. Future work items include incorporating support for RANSAC variants in the literature, along with specialized sensors, while leveraging the scalability offered by our multi-core parallel design.

Acknowledgment

This research has been sponsored in-part by Toyota in collaboration with Open Perception Inc.

References

[1] J. Li-Chee-Ming, C. Armenakis, „Fusion of Optical and terrestrial Laser Scanner Data", in CGC, 2010.

[2] A.Lacey, N.Pinitkarnand, N.Thacker, ‚An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibration", in BMVC, 2000.

[3] O. Chum, T. Werner, and J. Matas. „Two-View Geometry Estimation Unaffected by a Dominant Plane", in CVPR, 2005.

[4] A. Buch, D. Kraft, H. Petersen, N. Krueger, "Pose Estimation using Local Structure-Specific Shape and Appearance Context", ICRA'13

[5] M. A. Fischler, R. C. Bolles. Random Sample Consensus: „A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", in Communications of the ACM, 24(6):381–395, 1981.

[6] T. Sattler et al. „SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter" in ICCV, 2009.

[7] O. Chum and J. Matas. „Matching with PROSAC - Progressive Sample Consensus", in CVPR, 2005.

[8] Raguram, R., Frahm, J.M., Pollefeys, M.: „A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus" in ECCV, 2008.

[9] E. Rublee, V. Rabaud, K. Konolige, G. R. Bradski. „ORB: An efficient alternative to SIFT or SURF", in ICCV, 2011.

[10] S. Leutenegger, M. Chli, R. Siegwart. „BRISK: Binary Robust invariant scalable keypoints", in ICCV, 2011.

[11] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in ICRA, 2009

[12] F. Tombari, S. Salti, L. Di Stefano, "Unique Signatures of Histograms for Local Surface Description", in ECCV, 2010.

[13] F. Tombari, S. Salti, L. Di Stefano, „A Combined Texture-Shape Descriptor For Enhanced 3D Feature Matching", in ICIP, 2011.

[14] D. G. Lowe. „Object recognition from local scale-invariant features", in ICCV,1999.

[15] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool. „SURF: Speeded Up Robust Features", in CVIU, 2008.

[16] D. Ballard „Generalizing the Hough transform to detect arbitrary shapes", SPIE Proc. on Vision Geometry, 1981.