

MuFeSaC: LEARNING WHEN TO USE WHICH FEATURE DETECTOR

Sreenivas R. Sukumar, David L. Page, Hamparsum Bozdogan, Andreas F. Koschan, Mongi A. Abidi
The University of Tennessee, Knoxville, U.S.A

ABSTRACT

Interest point detectors are the starting point in image analysis for depth estimation using epipolar geometry and camera ego-motion estimation. With several detectors defined in the literature, some of them outperforming others in a specific application context, we introduce *Multi-Feature Sample Consensus (MuFeSaC)* as an adaptive and automatic procedure to choose a reliable feature detector among competing ones. Our approach is derived based on model selection criteria that we demonstrate for mobile robot self-localization in outdoor environments consisting of both man-made structures and natural vegetation.

Index Terms— feature learning, RANSAC, interest point detector evaluation

1. INTRODUCTION

There are two types of errors associated with image interest points from low-level feature detectors: (a) classification errors and (b) measurement errors. As Fischler and Bolles explain in [1], classification errors occur when a feature detector incorrectly identifies a portion of the image as an interest point, while measurement errors occur when the feature is identified correctly but not at a precise repeatable location. The measurement errors are usually smoothed by modeling the error as a normal distribution, but classification errors have a larger effect that cannot be averaged because of gross deviations.

The computer vision community has tried to minimize both these errors from two different perspectives. The first

one is by weeding out outliers using a random sample consensus (RANSAC) procedure [1] and the second by improving feature detection with newer definitions of interest points [2]. Hence, after years of research, several interest point detectors and enhancements to the RANSAC procedure [3] have come to stay in the literature. The list that includes intensity-based Harris or Shi-Tomasi corners [2], contour-based curvature features [4], phase congruent corners [5], MLESAC [6] and several others is getting longer. The recent performance evaluations of such image features [7] seem to indicate that the choice of a low-level feature detector is dependent on the application and environment. Also, the different conclusions drawn from comparisons presented in [7] and [8] only further underscore the question when to use which feature detector.

Our goal in this paper is to define a framework that can automatically decide when to use which feature. We will focus on a real world application in which a mobile robot tries to localize itself using video sensors. In such a scenario, where the usefulness and accuracy of low level features for pose recovery is significant, we will demonstrate our statistical *Multi-Feature Sample Consensus (MuFeSaC)* procedure leveraging informative characteristics of competing feature detectors and making the pose recovery reliable. With improved pose recovery, we present *MuFeSaC* that extends RANSAC as a new

- Data-driven “feature method” selection approach that improves pose recovery from scenes with many outliers caused by vegetation, and
- Information theoretic framework for learning and guiding the use of environment adaptive optimal features for camera ego-motion estimation.

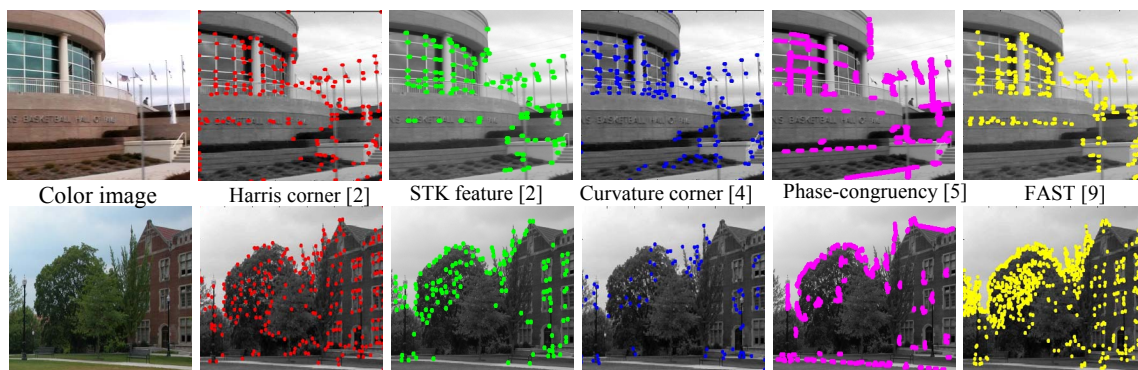


Figure 1: Commonly used interest point detectors (such as Harris) do not seem to generate repeatable features in natural scenes compared to man-made structures. The phase congruency detector and curvature corners though computationally expensive extract better features from vegetation. *MuFeSaC* can learn automatically when to use which feature in unknown unstructured environments.

Figure 1 shows an example of the problem that we are trying to address in this paper. The two images are from a video sequence of a mobile robot in an outdoor environment marked with interest points detected using different feature extraction methods. The top image with structured buildings has lesser number of uninteresting gross classification errors compared to the outliers caused by the vegetation in the bottom image. In the vegetation case, if only a single detector (like Harris corner) with many such classification errors is used in the pose recovery process, the commonly used RANSAC procedure, would either converge to a sub-optimal result, or would take more iterations to ignore the outliers and converge to a solution. Our *MuFeSaC* procedure enhances RANSAC in such situations by studying competing feature definitions and feeding with informative interest points for better RANSAC performance.

We will explain the *MuFeSaC* procedure in Section 2. Then, we will demonstrate experiments on video from a self-localizing mobile platform operating in unstructured terrain and compare *MuFeSaC* with RANSAC in Section 3. Based on these experiments, we will conclude with future directions in Section 4.

2. THE *MuFeSaC* ALGORITHM

MuFeSaC is an extension of RANSAC that includes multiple feature detectors. The contribution with *MuFeSaC* is an inference engine that in addition to finding the parameters of the interest model fit based on noisy data, also evaluates the confidence in the parameter estimates. *MuFeSaC* considers the confidence score from one single interest point detector along with the information from other competing interest points, thereby reducing the risk due to the choice of the feature detector. We list the different stages of the *MuFeSaC* procedure in Table 1. The backbone of *MuFeSaC* is model selection criteria based on information complexity that is used in the computation of scores *SFOC* and *CFCS*. In the following subsections, we will explain the implementation details of using the information criteria within *MuFeSaC*.

Table 1: Pseudo code of the *MuFeSaC* algorithm

1. **For** each feature detection method FD_i , $i = 1, 2, 3 \dots N$ competing interest point detectors
 - a) Extract interest points from two successive frames.
 - b) Find the putative matches using proximity and cross correlation.
 - c) Perform RANSAC and iterate to a convergence. Collect d -estimated parameters S of model M fitted during the iterations of RANSAC.
 - d) Estimate d -variate probability distribution B_i based on n ($n > 30$) iterations of parameter estimates ($S_1 \dots S_n$) collected.
- End**
2. Score Single Feature Outlier Consensus (*SFOC*) using the model selection criterion described in Section 2.1.
3. Compute Competing Feature Consensus Score (*CFCS*) by evaluating competing distributions B_i for different hypothesis and assign values as discussed in Section 2.2.
4. Choose the optimal feature detector with minimum $SFOC_i + CFCS_i$.
5. Repeat steps 1-4 every k frames, k can be adapted based on scene complexity. (Typically $k \sim 300$)

2.1. Single Feature Outlier Consensus

Based on the RANSAC convergence consensus alone, if we were to choose the best feature detector, we would ideally want to pick the method that is indicative of maximum likelihood of the parameters with minimum uncertainty, or in simpler words B_i with minimal variance. This can be mathematically expressed as the minimizer of the criterion (Equation 1) that simultaneously considers the likelihood and also penalizes the uncertainty associated with the likelihood of the parameters of model M . This model selection criterion in the statistics literature [10] is known as *ICOMP* and derives from the Kullback-Liebler (KL) distance between estimated and unknown underlying probability density. Without much modification, we are able to apply this criterion in evaluating the confidence in the model fit during the iterations of RANSAC. We note that Equation 1 does not involve distributional assumptions and can be applied to even Parzen window estimates of B_i .

$$ICOMP = \text{Lack of fit} + \text{Profusion of uncertainty} \\ = -2 \log (\text{Likelihood of } \mu_i) + 2 C_1 (F^{-1}(\Sigma_i)) \quad (1)$$

where F^{-1} is the inverse Fisher information matrix, μ_i and Σ_i are the maximum likelihood estimates of the mean and covariance computed as the first two moments of B_i . The C_1 measure and the F^{-1} is computed using Equations 2 and 3.

$$C_1(F^{-1}(\Sigma_i)) = \frac{s}{2} \log \left[\frac{\text{tr}(F^{-1}(\Sigma_i))}{s} \right] - \frac{1}{2} \log |F^{-1}(\Sigma_i)| \quad (2)$$

with s being the rank of F^{-1} , $|\cdot|$ refers to the determinant and tr refers to the trace of the matrix.

$$F^{-1}(\Sigma_i) = \begin{bmatrix} \Sigma_i & 0 \\ 0 & D^+_p(\Sigma_i \otimes \Sigma_i) D^+_p \end{bmatrix} \quad (3)$$

with D^+_p being the Moore-Penrose inverse of vectorized Σ_i , \otimes representing the Kronecker product. The C_1 measure for penalizing uncertainty is obtained by maximizing mutual information in d -dimensions [11]. We direct the reader to [10] for sampling bias compensating implementation details on the finite sampling form of Equation 1.

The ICOMP computed on B_i is the score $SFOC_i$ that quantifies the certainty in RANSAC convergence. We will now explain how to infer feature consensus based on the model fit from other interest point detectors in the following section.

2.2. Competing Feature Consensus Score

The $CFCS_i$ quantifies the agreement between the competing models M fit by RANSAC from each feature detector. The score is obtained by first evaluating different hypothesis listed below and then choosing the optimal consensus combinatorial cluster among competing feature detectors:

Case 1: All B_i 's are maximizing likelihood of the same parameters for model M . All μ_i 's equal and Σ_i 's equal.

Case 2: All μ_i 's are unequal but Σ_i 's are equal.

Case 3: All μ_i 's and Σ_i 's are unequal, but there exists a maximal cluster of some μ_i 's equal.

The verification of these hypotheses is like performing multi-sample clustering based on information distances in an entropic sense described by Bozdogan in [12]. We follow a similar approach to verify the three cases, by considering the samples that contributed to distributions B_i to have come from the same distribution and evaluate the complexity in model-fitting as the criterion to decide which of the three cases has occurred. We use the Akaike information criterion (AIC) as shown in Equation 4 to score the different hypothesis.

$$\begin{aligned} AIC(\mu_i, \Sigma_i, \kappa) &= -(\text{Likelihood of feature cluster}) + \\ &\quad \text{Parameter parsimony after clustering} \quad (4) \\ &= -2\log L + 2m \end{aligned}$$

The evaluation of the likelihood of feature cluster L only considers the samples that contributed to the distributions B_i 's within the cluster evaluated for consensus. We evaluate the parameter parsimony factor m for the 2^N different cluster combinations based on the formulae listed in Table 2. The hypothesis that has minimum AIC is the statistical decision. Initially, we only evaluate the three cases. This initial 3-case hypothesis verification can avoid the 2^N evaluations when all methods are accurate. We assign the minimizer of the AIC for the 3-case hypothesis as $CFCS_i$ to the corresponding feature detectors. If the minimizer indicates the occurrence of Case 2 or 3, we perform the evaluation on all combinatorial "feature detector" clusters. The minimizer of the AIC score still points to the cluster with maximal sensors contributing to the same model parameters. This AIC score is assigned only to the "feature detectors" within the maximal cluster. At this point, we note that identifying the different detectors not converging on the parameters, indicates the possibility that there might not be reliable pose recovery.

Table 2: Parameter parsimony estimation for a simple d -parameter M with $N = 3$ example.

	κ	Clustering	m
Case 1	3	(F_1, F_2, F_3)	$d + d(d+1)/2$
Case 2	1	$(F_1)(F_2)(F_3)$	$Nd + d(d+1)/2$
Case 3	2	$(F_1, F_2)(F_3)$ $(F_1, F_3)(F_2)$ $(F_2, F_3)(F_1)$	$\kappa d + \kappa d(d+1)/2$

Both the ICOMP and the AIC values being normalized information measures of complexity in our implementation; we are justified in using the sum of the two measures to choose the optimal feature detector. By establishing an information theoretic procedure for automatic "feature method detection" as a new addition to RANSAC, we present experiments deploying *MuFeSaC* in a real application.

3. EXPERIMENTS

We implemented 5 interest point detectors (Harris corner, STK features, phase-congruent corners, curvature corners and the FAST detector [9]) on a mobile robotic platform. The threshold values were automatically adjusted to generate at least 100 interest points in a given image. These interest points were fed in as input to the *MuFeSaC* algorithm described in the previous section. We show *MuFeSaC* guiding the switch between different interest point detectors in Figure 2. Also, we note that *MuFeSaC* selects detectors that generate repeatable features. In particular, for images marked 2 and 3, the switch from Harris (image marked 4) to phase congruency appears to be significant. We show the panorama in Figure 2 for visualization purposes that was taken using a separate digital camera. The marked windows show the approximate physical location of the video frame that is used for localization. We have also marked the timeline for switching interest point detectors on the panorama image for clarity.

We used a method similar to [13] for pose recovery from video for self-robot localization for a short 7 meter path. In Figure 2b, we show the result of localization and note the difference in pose recovery comparing Harris + RANSAC approach along with the ground truth measurements that were made using navigation instruments. We see that *MuFeSaC* has lesser deviation from the ground truth compared to the RANSAC procedure. In Figure 2c, we also present the computational overhead for *MuFeSaC* decision in contrast to single feature RANSAC to emphasize the real-time capability of *MuFeSaC*. Though *MuFeSaC* as a decision criterion is not a burden, the detection of multiple features along with RANSAC might pose a significant overhead. Hence, we suggest that *MuFeSaC* be executed only every k frames as previously mentioned in Table 1.

4. CONCLUSIONS

With *MuFeSaC*, we have extended the commonly used RANSAC procedure to accommodate the distinctness, stability, invariance and uniqueness of several interest point detectors in choosing an optimal one adaptively and automatically. *MuFeSaC* procedure which is inspired from MLESAC [6] to extend RANSAC, also performs better in applications where the target scenes are dynamically changing and one single feature detector might not be sufficient.

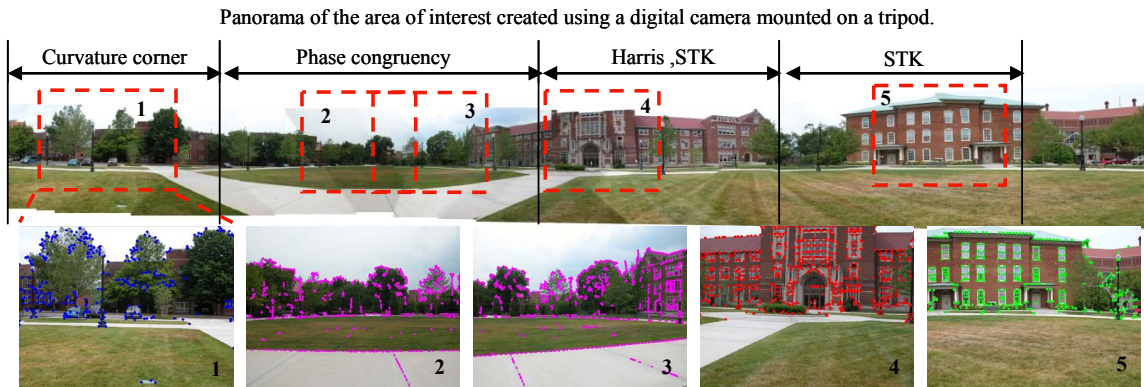
5. ACKNOWLEDGEMENTS

This work was supported by the University Research Program in Robotics under grant DOE-DE-FG52-2004NA25589.

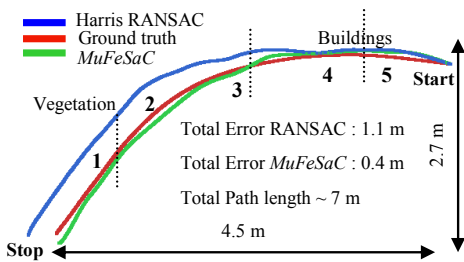
6. REFERENCES

[1] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". *Communications of the ACM*, 24(6), pp. 381-395, 1981.
 [2] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors," *Int'l J. Computer Vision*, 1(60), pp. 63-86, 2004.
 [3] 25 years of RANSAC, Workshop, in the *Proceedings of the Intl. Conf. on Computer vision and Pattern recognition, CDROM*.
 [4] X. C. He and N. H. C. Yung, "Curvature Scale Space Corner

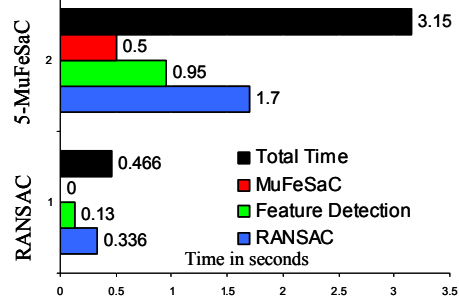
Detector with Adaptive Threshold and Dynamic Region of Support", in the *Proc. of the 17th International Conference on Pattern Recognition*, 2, pp. 791-794, 2004.
 [5] P. Kovess, "Image Features From Phase Congruency," *Videre: A Journal of Computer Vision Research*, MIT Press. 1(3), 1999.
 [6] P.H.S Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding*, 78, pp. 138-156, 2000.
 [7] K. Mikolajczyk and C. Schmid, "Performance evaluation of local descriptors," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10), pp. 1615-1630, 2005.
 [8] M. Zuliani, C. Kenney and B.S. Manjunath, "A mathematical comparison of point detectors", in the *Proc. of Second IEEE Image and Video Registration Workshop*, Washington DC, 2004.
 [9] E. Rosten and T. Drummond, "Machine learning for high speed corner detection," in *Proc. of the 9th European Conference on Computer Vision*, 1, pp. 430-443, 2006.
 [10] H. Bozdogan, "Akaike' Information Criterion and Recent Developments in Information Complexity," *Journal of Mathematical Psychology*, 44, pp. 62-91, 2000.
 [11] Van Emden, "An Analysis of Complexity", *Mathematical Centre Tracts*, Vol. 35, 1971.
 [12] H. Bozdogan, "Multi-sample cluster analysis as an alternative to multiple comparison procedures", *Bulletin of Informatics and Cybernetics*, 1-2, pp. 95-129, 1986.
 [13] D. Nistér, O. Naroditsky and J. Bergen, "Visual Odometry," in *Proc. of the Intl. Conf. on Computer vision and Pattern Recognition*, 1, pp. 652-659, 2004.



(a) Adaptive feature method selection using *MuFeSaC* in a challenging outdoor environment for robot self-localization.



(b) *MuFeSaC* performs better than Harris + RANSAC



(c) Timing results: 5-feature *MuFeSaC* vs. Harris + RANSAC

Figure 2: Results after deploying *MuFeSaC* for the self-localization of a robotic mobile platform with a video sensor. *MuFeSaC* helps choose detectors that generate repeatable interest points and hence reduce error in self-localization.