On Handling Uncertainty in the Fundamental Matrix for Scene and Motion Adaptive Pose Recovery

Sreenivas R. Sukumar, Hamparsum Bozdogan, David L. Page, Andreas F. Koschan and Mongi A. Abidi
Imaging, Robotics and Intelligent Systems Lab, The University of Tennessee, Knoxville.
Website: (http://www.imaging.utk.edu) Email: {ssrangan, bozdogan, dpage, akoschan, abidi}@utk.edu

Abstract

The estimation of the fundamental matrix is the key step in feature-based camera ego-motion estimation for applications in scene modeling and vehicle navigation. In this paper, we present a new method of analyzing and further reducing the risk in the fundamental matrix due to the choice of a particular feature detector, the choice of the matching algorithm, the motion model, iterative hypothesis generation and verification paradigms. Our scheme makes use of model-selection theory to guide the switch to optimal methods for fundamental matrix estimation within the hypothesis-and-test architecture. We demonstrate our proposed method for vision-based robot localization in large-scale environments where the environment is constantly changing and navigation within the environment is unpredictable.

1. Introduction

The fundamental matrix **F** that relates two perspective images of a single rigid object/scene is estimated by solving the epipolar constraint in Equation 1, where $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ are corresponding points in two images **I** and **I'** respectively. Assuming that the calibration matrix (**K**) of the camera acquiring images of the scene from different viewpoints is available; **F** is instrumental in the estimation of the relative rotation (**R**) and translation (**t**) of the camera relating the two images. The notation $[\mathbf{t}]_x$ refers to the anti-symmetric matrix form of the translation vector **t** in Equation 1.

$$\tilde{\mathbf{m}}_i \mathbf{F} \tilde{\mathbf{m}}'_i = 0 \quad ; \mathbf{F} = \mathbf{K}'^{-T}[\mathbf{t}]_x \mathbf{R}\mathbf{K}^{-1} \tag{1}$$

In vision-based vehicle navigation/localization applications, we note that the uncertainty about the pose (**R, t**) is directly related to the uncertainty on **F**. Unfortunately, the accuracy and the uncertainty about **F** depends on the quality of feature correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$. Though, theoretically all pixels in **I** and **I'** are eligible to contribute to $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$, for computational reasons and also for improved performance, we choose special interest points to build the list of correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$.

The special interest points extracted are inspired by perceptual heuristics from human vision theory. For example, the Harris corners [1] are intensity-gradient based interest points, the curvature corner [2] is edge based, the phase congruency corners [3] are spatial frequency inspired while SIFT [4] is a multi-resolution feature. If all these feature point detectors extracted points with 3D characteristics that can be matched across images, the estimate of **F** would be independent of the interest point detector. But, in reality we observe that feature point detectors perform better in some situations while not meeting the expectations in some others. So, how do we choose the right feature detector for the scene of interest? Is it possible to choose the feature detector as the scene changes on a mobile platform adaptively? Our goal with this paper is to provide a potential solution to this question.

If ε is the probability that a feature point is in error, and p is the probability indicative of a feature match error, the number of minimal hypothesis n of s sub-samples taken at a time to fit a motion model **M** of the fundamental matrix is shown in Equation 2. The confidence on the estimated geometry is based on an inlier bound condition modeled as a probability distribution (Equation 3).

$$n = \log(1-p)/\log(1-(1-\varepsilon)^s) \tag{2}$$

$$\Pr(\{\tilde{\mathbf{m}}_i, \tilde{\mathbf{m}}'_i\}) = \prod_i \left(\frac{1}{2\pi\sigma^2}\right) e^{-(d(\tilde{\mathbf{m}}_i, \hat{\tilde{\mathbf{m}}}_i)^2 + d(\tilde{\mathbf{m}}'_i, \hat{\tilde{\mathbf{m}}}_i)^2)/(2\sigma^2)} \tag{3}$$

Equations 2 and 3 hint at how the uncertainty on the corresponding points propagates as outliers in the subsequent matching stage corrupting the model parameters during the iterative hypothesis-verify scheme involved in solving the epipolar constraint.

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

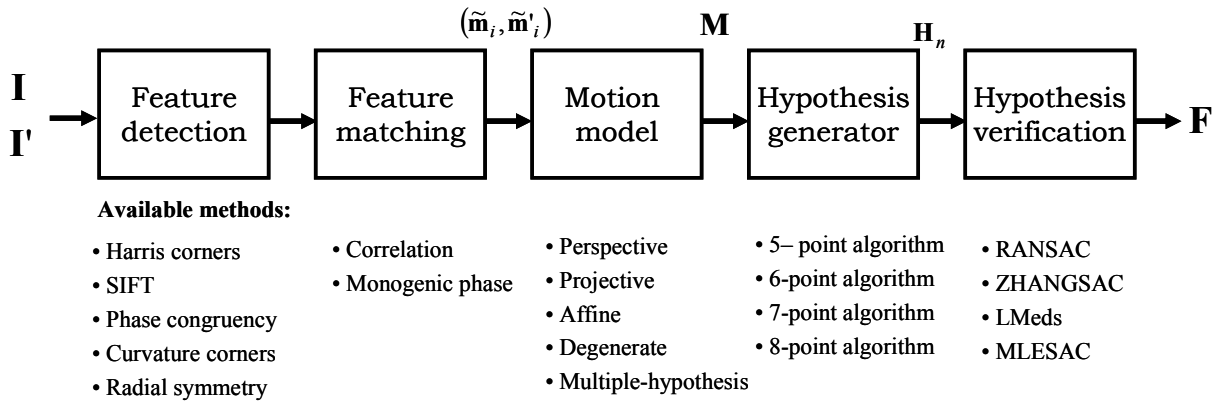


Figure 1: There are several options available in the flow process for estimating the fundamental matrix. However, these options are scene and motion specific and the choice of these methods (pipeline) in a real world situation has to be adaptive to the environment. In this paper, we propose a statistical decision procedure to increase the confidence and reliability of the fundamental matrix.

Over the years of computer vision research for feature-based 3D geometric inference summarized in Figure 1, we have seen several interest point detection methods [1, 2, 3, 4, 5] claiming the reduction of pixel localization error ε. We have seen improved matching methods based on correlation [6] and monogenic phase [7] to operate on those interest points trying to produce matches with minimal probability of error p. Furthermore, there is also a recent body of literature on the s-point (s = 8, 7, 6, 5) algorithms [8, 9, 10] for generating hypothesis model parameters **M**, along with several iterative verification improvements over RANSAC [11] to choose the best **M** and its parameters.

The vast literature has forced performance evaluation publications comparing feature detectors [12, 13, and 14], feature matching strategies [7], hypothesis generators and the verification schemes [15]. Incongruously, comparisons from different authors conclude differently. For example, SIFT features performed the best in [12], Harris in [13], and Forstner in [14]. Another example is the comparison of hypothesis generators [8] that shows Nister's 5-pt solver performing better in sideways motion while the traditional 8-pt algorithms [10] perform better for forward motion. So, for an arbitrary video sequence, in addition to the risk in the choice of the feature detector, how do we decide which minimal hypothesis generator (as in 5 pt, 6 pt, 7 pt or 8 pt solver) to use for the varying and intangible quality of feature correspondence data? This observation that the performance of feature detectors, matching algorithms and hypothesis generators are data/scene and motion specific motivates our work.

We have organized this paper to share results from our implementation of pose recovery using several interest point heuristics, competing matching methods, hypothesis generators for image sequences Further, we demonstrate our scheme being able to leverage years of computer vision research to enable mobile robots in dynamic large-scale real world environments to operate in an intelligent fashion switching pose recovery strategies while persistently considering the uncertainty about the fundamental matrix. In doing so, we would like to conclude this paper by summarizing the following contributions:

- A new statistical procedure based on information complexity that automatically decides and guides the switch to optimal feature detectors and motion models for robust estimation of the fundamental matrix.

- Formulation of the epipolar geometry estimation pipeline as a random process, generating ensemble space-ergodic characteristics to compensate for bad quality in time-stationary parameter convergence.

- A potential method for quantifying the uncertainty in the fundamental matrix that is independent of the scene, and ground truth measurements.

In the following Section, we establish the background to our approach with related work. In Section 3, we explain the theoretical inspiration for the inference engine and detail the implementation procedure in Section 4. We demonstrate our work in real world situations both in indoor and outdoor navigation applications in Section 5. Finally, we summarize our efforts and conclude with future directions in Section 6.

2. Related work

In the well researched area of camera ego-motion estimation, we begin by understanding methods discussed in [7, 15]. Out of the several methods in those surveys, our focus in this paper is on the feature-based 3D geometric inference using calibrated cameras. The seminal effort in

[16] addresses our particular concern with attention to uncertainty of the fundamental matrix explaining different methods to estimate the epipolar geometry. We use the methods described to estimate **F** and its uncertainty along with the normalization procedure *FDiff* to normalize fundamental matrices from the implementation in [16].

The methods in the literature that closely relate to our proposed work are [17] and [18]. Kanatani [17] proposes the idea of asymptotic analysis for uncertainty modeling and explains how to handle uncertainty in geometric fitting using model selection criterion. In the same paper, Kanatani modifies the Akaike information criterion for geometric inference and argues the need for a geometric extension from statistics. Using a similar model selection tool, Torr in [18] clusters feature matches for motion model determination detecting degeneracy, affine motion and even multiple motions between successive frames. Kanatani's remarks in [17] and Torr's future directions in [18] act as inspiration for our work in considering the fundamental matrix estimation process as a random process of sub-systems where the choice of the model **M** is as important as its estimated model parameters that make **F**.

3. Theoretical inspiration

The significant improvement that we provide over Torr [18] is that we perceive the fundamental matrix solver as a stochastic process and the correspondence data as the excitation signal to the process. Now, the sub-samples drawn from the large set of noisy correspondence data for model fitting is analogous to the time-shifted signal input for the model-fitting random process to produce the same or equivalent fundamental matrix as output. The *n* different hypotheses that the solver generates before making a decision is indicating uncertainty on the estimated fundamental matrix in a time-stationary sense.

Our proposed idea to increase the reliability on the fundamental matrix is by generating an uncertainty model in a space-ergodic sense. We excite the random solver with spatially different correspondence data generated by different interest point detectors to generate the ensemble uncertainty samples. Our method then tries to compensate the deficiencies (uncertainty) in time-stationary sense with space-ergodicity (and vice versa) with the objective of reducing the total uncertainty on the fundamental matrix.

This idea underlies our approach to uncertainty management on the fundamental matrix. We recollect that, in estimating **F** with noisy $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ using the hypothesis-test-verify paradigm, there is two inherent embedded problems. The first one is the determination of the model **M** and the second, the parameters of **M**. While there has been considerable statistical attention on the estimation accuracy of the parameters of **M**, the validity for the model choice **M** is significantly ignored in the literature. Our approach generates statistics for the model validity using several interest point detectors and combines the model validity with the model fitting accuracy in choosing optimal minimal algorithms based on the scene and motion.

Suppose correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ are ideal as in a synthetically generated case, all the *n* different hypotheses would lead us to the same or equivalent **F**. But, in the noisy real world, robust model-fitting or hypotheses verification paradigms (inspired by RANSAC) can iterate to an estimate of **F** that is sub-optimal or iterate to a non-acceptable result because of the outliers. If the quality of feature matches is good, the distribution of the model parameters during the iterative procedure will be confident and tightly bound. We model this convergence of the fundamental matrix parameters towards learning the optimal choice of methods that will lead to increased confidence in **F.** We describe the statistical procedure and the implementation details in the following section.

4. Learning the optimal minimal algorithm

The quality of feature correspondences $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ contains two types of errors, localization errors and gross classification errors. The bad matches also called outliers infiltrates uncertainty into the model estimation and fitting process. Our experience with real world environments is that the variance bound in Equation 3, needs to be adjusted drastically in images with nice structural features and ones with poor structure. Even robust outlier rejection methods like MLESAC [19] are not able to handle such situations mostly because of infinite possibilities to consider in real world dynamic situations. We will leave the model fitting method to consider modeling the error between re-projected interest points, while we consider the iterations of the consensus indicative of different hypothesis on vectorized **f** of **F** as our uncertainty sample.

A good feature detector, matching algorithm, and a robust hypothesis generation algorithm would produce equivalent fundamental matrices **F** with a majority of the *n* minimal hypotheses evaluated for the inliers bound. In other words, the confidence in the distribution of **f** measures the uncertainty over generating the model using the corrupt correspondence data. Then, we look at the parameters generated by different pipelines (choice of different feature detectors, hypothesis generation schemes and motion models) and evaluate the different estimates of **F** for model support or reliability. Our algorithm is summarized in Table 1 followed by the implementation details.

Table 1: Our proposed algorithm

1. **For** each potential competing pipeline of methods $\mathbf{P_i}$ listed in Figure 1 where *i = 1,2 ,3..k*,
   a) Use RANSAC and iterate to a convergence. Collect *d*-estimated parameters **H** of model **M** fitted during the iterations of RANSAC.
   b) Estimate *d*-variate probability distribution $\mathbf{B_i}$ based on j (j > 30) iterations of parameter estimates ($\mathbf{H_1}$…,$\mathbf{H_n}$) collected.

   End
2. Score Correspondence Outlier Consensus ($\mathbf{COC_i}$) using the model selection criterion in Section 4.1.
3. Compute Model Consensus Score ($\mathbf{MCS_i}$) by evaluating competing distributions $\mathbf{B_i}$ as a multi-sample clustering problem in Section 4.2.
4. Choose the optimal pipeline $\mathbf{P_i}$ with minimum $\mathbf{COC_i}$ + $\mathbf{MCS_i}$.
5. Repeat Steps 1-4 every *m* frames

4.1. Correspondence Outlier Consensus

As mentioned previously, we are interested in the RANSAC convergence consensus and quantifying the confidence within the convergence process. Particularly, our interest is in identifying the pipeline that is indicative of maximum likelihood of the fundamental matrix parameters with minimum uncertainty, or in simpler words $\mathbf{B}_i$ with minimal variance. This can be mathematically expressed as the minimizer of the criterion (Equation 4) 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 [20] is popularly known as *ICOMP* and derives from the Kullback-Liebler (KL) distance between estimated and unknown underlying probability density. This criterion in theory is inspired from the same source as Kanatani's geometric AIC [17] but is able to include the covariance of the model parameters. Without much modification, we are able to apply this criterion in evaluating the confidence in the model fit after the iterative convergence of RANSAC. We note that Equation 4 does not involve distributional assumptions and can be applied to even Parzen window estimates of $\mathbf{B_i}$. The parameters **f** and $\hat{\mathbf{\Sigma}}_\mathbf{F}$ in Equation 4 can be computed using one of several methods in [16], though in our implementation, we use the moments of the distribution $\mathbf{B_i}$. The correspondence outlier consensus (**COC**) is given by:

$$\mathbf{COC} = -2\log(\text{Likelihood of } \mathbf{f};\mathbf{M}) + 2C_1(F^{-1}(\hat{\mathbf{\Sigma}}_\mathbf{F})) \qquad (4)$$

where F^{-1} is the inverse Fisher information matrix, The C_1 measure and the $F^{-1}$is computed using Equations 5 and 6.

$$C_1(F^{-1}(\hat{\mathbf{\Sigma}}_\mathbf{F})) = \frac{s}{2}\log\left[\frac{tr(F^{-1}(\hat{\mathbf{\Sigma}}_\mathbf{F}))}{s}\right] - \frac{1}{2}\log\left|F^{-1}(\hat{\mathbf{\Sigma}}_\mathbf{F})\right| \qquad (5)$$

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

$$F^{-1}(\hat{\mathbf{\Sigma}}_\mathbf{F}) = \begin{bmatrix} \hat{\mathbf{\Sigma}}_\mathbf{F} & 0 \\ 0 & D^+{}_p(\hat{\mathbf{\Sigma}}_\mathbf{F} \otimes \hat{\mathbf{\Sigma}}_\mathbf{F})D^+{}_p{}' \end{bmatrix} \qquad (6)$$

with $D^+{}_p$ being the Moore-Penrose inverse of vectorized $\hat{\mathbf{\Sigma}}_\mathbf{F}$, $\otimes$ representing the Kronecker product. The C_1 measure for penalizing uncertainty is obtained by maximizing mutual information in *d*-dimensions. We direct the reader to [20] for sampling bias compensating implementation details on the finite sampling form of Equation 4. Equation 7 is the reduced form of Equation 4 for a normal distribution $\mathbf{B_i}$ generated using *y* samples from the iterations. We would like to note that by choosing the pipeline with minimal **COC** score, we are able to learn the choice of methods, which require lesser number of hypotheses to evaluate at the same time penalize for the uncertainty in the estimated parameters.

$$\begin{aligned}\mathbf{COC} = {} & \frac{s}{2}\ln\left(\frac{tr(\hat{\mathbf{\Sigma}}_\mathbf{F})}{s}\right) - \frac{1}{2}\ln|\hat{\mathbf{\Sigma}}_\mathbf{F}| + \left(\frac{n_i d}{2}\log(2\pi)\right) \\ & + \frac{n_i}{2}\log|\hat{\mathbf{\Sigma}}_\mathbf{F}| + \frac{1}{2}tr(\hat{\mathbf{\Sigma}}_\mathbf{F}{}^{-1})\left(\sum_{i=1}^{n_i}(y_i - \mathbf{f})(y_i - \mathbf{f})'\right)\end{aligned} \qquad (7)$$

Though we could stop with minimizing **COC** which is a minor improvement over [18], we note that the choice of the model **M** is a significant factor in the fitting process. For example, in an urban environment one feature detector could produce correspondence data that is tracking features far away like the sky and the clouds, while another might be tracking multiple objects such as cars moving in the scene. In the former case, we will end up choosing **M** with an affine structure with 5 parameters, and in the latter case up to 18 parameters on 2 fundamental matrices. To avoid this risk in the choice of the model itself, we will quantify the model support from competing pipelines.

4.2. Model Consensus Score

The model consensus score is obtained by evaluating the data from several pipelines $\mathbf{P_i}$ in a multi-sample clustering framework using Equation 8. The idea is to cluster $\mathbf{B_i}$ of each $\mathbf{P_i}$ using information distance between the distributions inspired by the method described in [21]. The model consensus score (**MCS**) is as shown below.

$$\mathbf{MCS} = -2\log(\text{Likelihood of } \mathbf{M}) + 2g \qquad (8)$$

where *g* penalizes for the parameter parsimony for **M**.

We begin by testing for the first three hypotheses listed below and in the occurrence of Case 2 or 3 alone, evaluate combinatorial subsets in Case 4. The values of g for different cases are listed in Table 2.

Table 2: Parameter parsimony penalty in MCS.

Cases	g *of* **M**
Case 1: All models **M** equivalent (**f** and $\hat{\mathbf{\Sigma}}_{\mathbf{F}}$ are equal for all competing $\mathbf{P_i}$) Cluster : $\{\mathbf{P}_1,\mathbf{P}_2 \ldots \mathbf{P}_k\}$	$d + \frac{d(d+1)}{2}$
Case 2: Different **M** (Different **f** but equal $\hat{\mathbf{\Sigma}}_{\mathbf{F}}$ for competing $\mathbf{P_i}$) Cluster : $\{\mathbf{P}_1\}\ \{\mathbf{P}_2\} \ldots \{\mathbf{P}_k\}$	$kd + \frac{d(d+1)}{2}$
Case 3: All **M's** are different for competing $\mathbf{P_i}$ (Different **f** and different $\hat{\mathbf{\Sigma}}_{\mathbf{F}}$ for competing $\mathbf{P_i}$) Cluster : $\{\mathbf{P}_1\}\ \{\mathbf{P}_2\} \ldots \{\mathbf{P}_k\}$	$kd + \frac{kd(d+1)}{2}$
Case 4: Cluster of κ pipelines of k competing $\mathbf{P_i}$ are equivalent (κ **f's** are equivalent) Cluster : $[\{\mathbf{P}_{1\}}\ \{\mathbf{P}_2\} .. \{\mathbf{P}\kappa\}] .. \{\mathbf{P}_{k-1}\}\{\mathbf{P}_k\}$	$\kappa d + \frac{\kappa d(d+1)}{2}$

The way we interpret these hypotheses is that when all pipelines are essentially leading us to the equivalent fundamental matrix, fusing correspondence data from two feature detectors for example, is not altering the convergence process or the end result. In such cases, the model fitting process can be inferred as stationary and ergodic with high reliability. We evaluate this condition in Case 1, where we treat all samples contributing to $\mathbf{B_i}$ of each $\mathbf{P_i}$ as drawn from a single distribution $\mathbf{B_{consensus}}$. This case usually occurs when we have interesting features like corners in buildings to track.

However, our method is particularly useful only when Case 2 or Case 3 occurs. Case 2, tells us that the choice of the pipeline is influencing the estimate of **F**, and that we might have to be satisfied with the pipeline with minimal **COC** score. Case 3 forces us into the balancing act of which model and whose parameter estimates correspond to minimal uncertainty. We assign the minimal value of Equation 8 for the three hypotheses as MCS ($\mathbf{P_i}$). Then, we try to learn the most efficient model by evaluating all combinatorial subsets for this purpose in Case 4, which is actually a nested form of Case 1, 2 and 3. We evaluate all possible clusters (subsets) and assign the minimum value of MCS over all clusters only to the κ pipelines with high model support.

This procedure is comparable to modeling the convergence consensus into a mixture model and finding the order of the mixture similar to previous effort [18], the difference being that we can now accommodate for the risk in choosing a particular pipeline. The **COC** score quantifies the risk in the estimation and the **MSC** score quantifies the risk in the model. Our next step is to combine them with the objective of improving the reliability on the process of fundamental matrix estimation.

4.3. Total Uncertainty Score

The total uncertainty score U_F that we will use to make a decision about the choice of algorithms combines two probabilities (uncertainties).

$$U_F(P_i) \propto \Pr(\mathbf{f}; \mathbf{M}, P_i) + \Pr(\mathbf{M}; \mathbf{P}_j; j = 1,2,3..k) \quad (9)$$

$$U_F(P_i) = COC(P_i) + MCS(P_i) \quad (10)$$

Both the **COC** and **MCS** scores are constructed using log-likelihood approximations of information complexity and the addition of these scores to decide on optimal P_i is intuitive. In addition, **MCS** and **COC** also provide an extra benefit of informing us, if we actually can improve the uncertainty over the fundamental matrix if we fused several pipelines instead of selecting pipelines. Since the **MCS** computation already considers the reliability aspect of fusion, we are able to note that if **MCS** of selected pipeline P_i is less than its **COC** score, we can improve by fusing the κ pipelines in the maximal cluster. We make this statement following the principles in the study of fusers better than the best sensor in [22]. However, in most real world situations that we experiment in the following section, pipeline selection proves to be the promising direction.

5. Experimental Results

In this section, we will consider navigation applications in indoor and outdoor environments. We will be presenting results on two situations using mobile robotic platforms. Our robotic platforms have direct pose recovery instruments either in the form of a laser range scanner and/or GPS/INS to provide us with a ground truth on the pose. We will demonstrate our proposed framework to adapt to features based on the scene and also switch hypothesis generators adaptively based on the quality of the feature correspondences.

We begin with results from using the vision-based pose recovery module in a large scale urban mapping platform as shown in Figure 2. We demonstrate our algorithm switching interest point detectors and motion hypotheses generators based on the scene and the motion. We show the Monte-Carlo result of one switch in this dynamic environment to emphasize the uncertainty management. The graphs indicate the reliability on the recovered pose after 100 iterations of convergence on the same frames. With these results, we are able to demonstrate that our

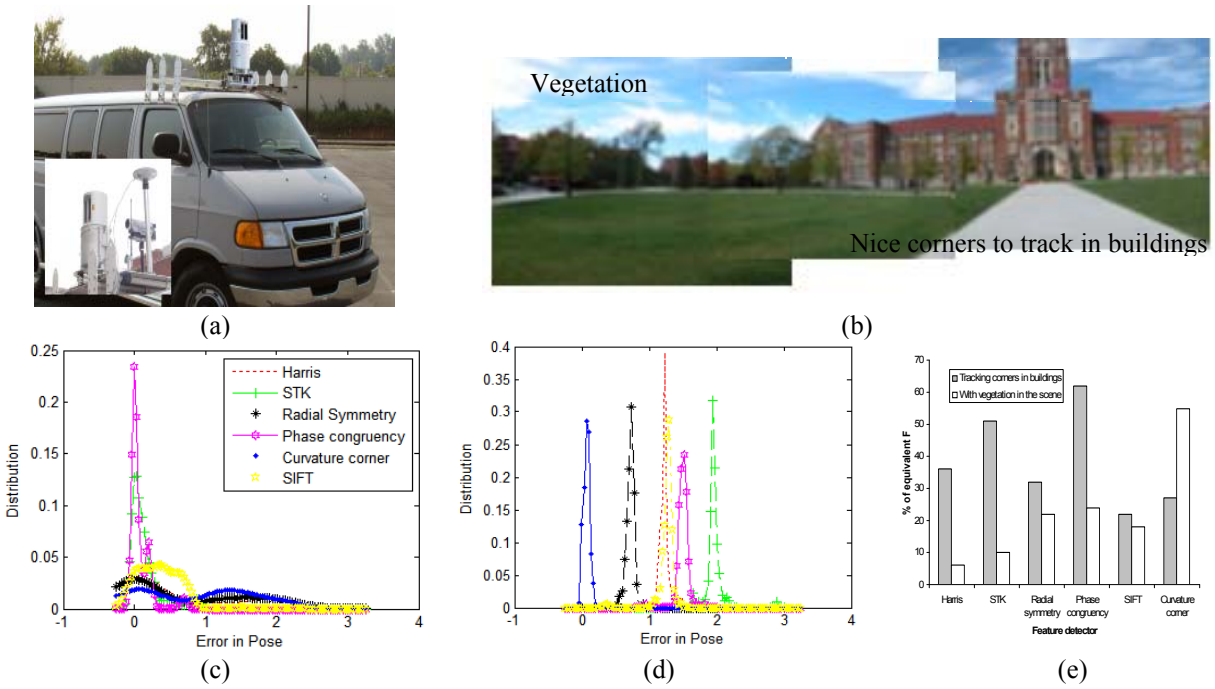


Figure 2: As the mobile platform with vision-based pose recovery navigates in this environment, tracking nice corners in buildings and then encounters vegetation, our method guides the switch to the better feature detector. (a) Our data collection system with GPS, INS ground truth measurement. (b) The challenging scene of interest with buildings and vegetation. (c) The Monte-Carlo analysis on the recovered pose and the error plotted as a distribution when tracking good corners. (d) Monte-Carlo analysis on the recovered pose when encountering vegetation. Our method guided the switch from STK/ Phase congruency feature with the buildings to the curvature corners when the scene dealing with vegetation. (e) Percentage of equivalent **F** generated in 100 trials.

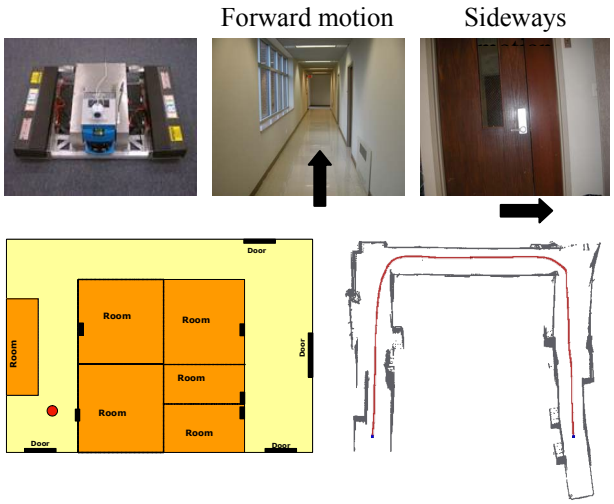
method chooses feature detectors that guarantee reproducibility with better accuracy in pose recovery. From Figure 2(e), we deduce that, if we did not guide the switch, there is 50% greater chance that we did not estimate the optimal fundamental matrix.

For the indoor case, our robot was equipped with a camera and SICK laser range finder and was intended to traverse a corridor as shown in Figure 3. During the course, the robot had to maneuver at different orientations during which our algorithm decided to use 5pt algorithm or 8 pt algorithm based on the data. For comparison sake, on offline processing of the image data using each one of the algorithms separately, we realize the significant error we would have accumulated had we not countered for Nister's dilemma in choosing between 5pt and 8pt hypothesis generators [8]. In Figure 3, we also tabulate the average error in recovered pose per frame following the definition of error provided in [8] using the essential matrix.

Our next experiment was to evaluate the robustness of the framework proposed in this paper. We selected 16 video sequences (Figure 4), each with 2000 frames containing both man-made structures and natural vegetation. We show a plot of the average uncertainty score on each of these datasets with our adaptive method and a standard method (Harris+8 pt + RANSAC) with timing comparisons in Figure 4. From Figures 4 (b) and (c) we are able to visualize how the uncertainty management using our framework translates to reliability on the fundamental matrix. Our method is able to do as good as the standard method when the standard method is sufficient, and is able to further improve the reliability in other situations.

Thus far, we presented results on real world environments both indoors and outdoors using the same mobile platform and pre-programmed vision module. In the experimental phase we made two interesting observations. The first observation was that, when we replaced our high resolution camera with a noisy cheap one, our method automatically chose a different set of feature detectors and hypothesis generation algorithms adapting to the noise characteristics of the cheap camera.

The second observation exposed a caveat. We realized that the time that we gain by minimizing the number of hypothesis to evaluate using our proposed method depends on how many competing models we evaluate and how often do we evaluate for a scene change. Typically in our experiments, our inference engine feeds the vision system once every 300 frames to operate real-time. A drastically changing scene environment or a fast moving



	Forward motion	Sideways motion
	Average error	Average error
8 pt	0.8	2.6
5 pt	1.6	0.9

Figure 3: Robot in a corridor with a camera and a laser sensor. We show the intended path for the robot along with a rough map that was integrated with the redline indicating the path of the robot. We also tabulate error results in the frames to show shat by switching methods, we maintain the uncertainty levels across frames.

mobile robot might require more frequent updates demanding more computational overhead. In Figure 5, we show how much of an overload our framework requires compared to believing one single feature detector and one estimation algorithm. The graph shows that our proposed method is not going to challenge real-time operation in most cases when evaluated every 50 frames.

6. Summary

We proposed a statistical procedure for uncertainty management when dealing with images from a calibrated camera in pose recovery using the fundamental matrix. We demonstrated results on pose recovery for navigation applications in indoor and outdoor environments. Our method combines the uncertainty in the convergence of the parameters of the fundamental matrix with the uncertainty of the model itself. The different models generated in the hypothesis-test framework of pose recovery provide the statistics for the confidence in the matches while the models generated by different feature detectors provide the statistics for quantifying the uncertainty in the model. By combining both these uncertainties, we have formulated a generic scheme using model selection theory that will help us choose methods for reliable estimation of the fundamental matrix at the same time acting as a performance measure of pose recovery using image features.

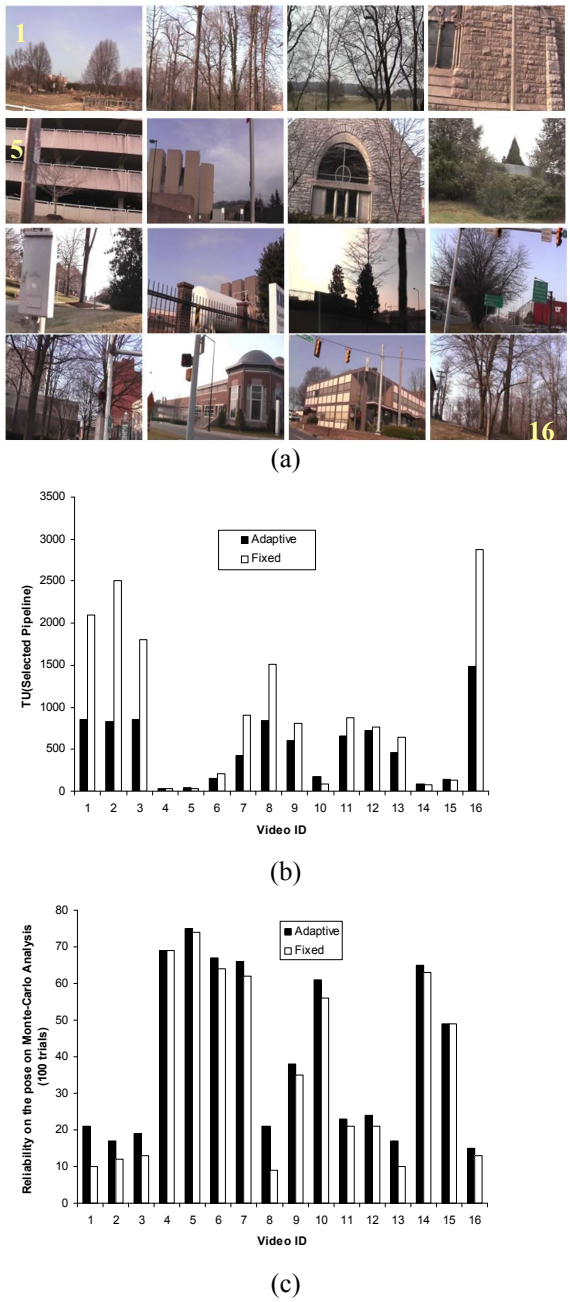
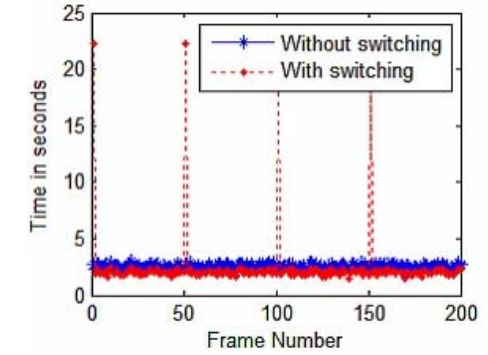
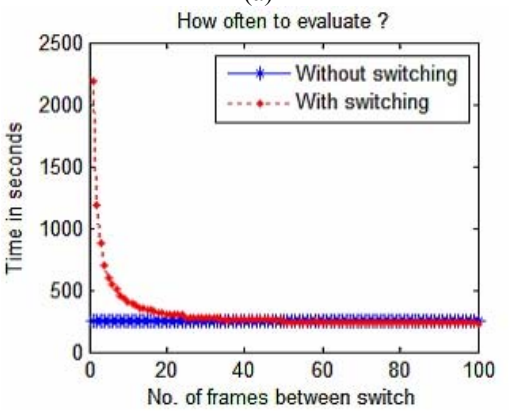


Figure 4: (a) Sample frames from several video sequences considered in our experiments. (b) The average uncertainty score after choosing optimal algorithms vs. using Harris corners + RANSAC + 8 pt algorithm. Our approach shows significant improvement over believing on a single pipeline (c) Reliability (Average count on convergence on true pose between frames) on the fundamental matrix after conducting the same experiment 100 times.



(a)



(b)

Figure 5: Timing analysis for learning competing models. (a) How much of a burden depends on the number of competing pipelines to evaluate. For this particular case, we had 20 competing pipelines, evaluated once every 50 frames. (b) This graph indicates using our framework every 50 frames is not as much as a burden compared to the pose recovery.

Acknowledgements

This research was supported by the DOE University Research Program in Robotics under grant DOE-DEFG02-
86NE37968.

References

[1] C. Harris and M. Stephens. A Combined Corner and Edge Detector. Proc. Alvey Vision Conf., Univ. Manchester, 147-151, 1988.

[2] X. C. He and N. H. C. Yung, Curvature Scale Space Corner Detector with Adaptive Threshold and Dynamic Region of Support, ICPR, 2:791-794, 2004.

[3] P. Kovesi, Image Features from Phase Congruency, Videre: A Journal of Computer Vision Research, MIT Press. 1(3), 1999.

[4] David G. Lowe, Distinctive image features from scale-invariant key points, IJCV, 60(2):91-110, 2004.

[5] G. Loy and A. Zelinsky, Fast Radial Symmetry for Detecting Points of Interest, IEEE PAMI,25(8):959-973,2003.

[6] T.S Huang, A.N. Netravali, "Motion and structure from feature correspondences: a review," Proc. of the IEEE, 82(2), 252-268, 1994.

[7] T. Kunio, Feature Point Correspondence of Stereo Images by Monogenic Phase, Communications, Computers and Signal Processing, Proc. of IEEE Pac.Rim Conf, 272-275, 2007.

[8] D. Nistér, An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 26(6):756-770, 2004.

[9] H. Stewenius, C. Engels and D. Nister, Recent developments on direct relative orientation, ISPRS Journal of Photogrammetry and Remote Sensing, 60(4):284-294, 2006.

[10] R. I. Hartley, In Defense of the Eight-Point Algorithm, IEEE PAMI, 19 (6): 580-593, 1997.

[11] 25 years of RANSAC, Workshop, *in the Proceedings of the Intl. Conf. on Computer vision and Pattern recognition, CDROM.*

[12] K. Mikolajczyk and C. Schmid, “Scale and Affine Invariant Interest Point Detectors,” Int'l J. Computer Vision, 1(60), pp. 63-86, 2004.

[13] M. Zuliani, C. Kenney and B.S. Manjunath, A mathematical comparison of point detectors, In the Proc. IEEE Image and Video Registration Workshop, 2004.

[14] V. Rodehorst and A. Koschan, Comparison and evaluation of feature point detectors, in Proc. 5th International Symposium Turkish-German Joint Geodetic Days "Geodesy and Geoinformation in the Service of our Daily Life", Berlin, Germany, March 2006.

[15] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision, Cambridge University Press, ISBN: 0521623049, 2000.

[16] Z. Zhang, Determining the Epipolar Geometry and its Uncertainty: A Review, IJCV, 27(2):161-195, 1998.

[17] K. Kanatani, Uncertainty modeling and model selection for geometric inference, IEEE PAMI, 26(10):1307-1319, 2004.

[18] P.H.S. Torr, Geometric motion segmentation and model selection, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 356(1740):1321-1340, 1998.

[19] P.H.S Torr and A. Zisserman, “MLESAC: A New Robust Estimator with Application to Estimating Image Geometry,” *Computer Vision and Image Understanding*, 78: 138-156, 2000.

[20] H. Bozdogan, “Akaike’ Information Criterion and Recent Developments in Information Complexity,” Journal of Mathematical Psychology, 44:62-91, 2000.

[21] H. Bozdogan, “Multi-sample cluster analysis as an alternative to multiple comparison procedures”, Bulletin of Informatics and Cybernetics, 1-2: 95-129, 1986.

[22] N. S. V. Rao, “On fusers that perform better than best sensor”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(8): 904-909, 2001.