

BLOGS: Balanced Local and Global Search for Non-Degenerate Two View Epipolar Geometry

Aveek Shankar Brahmachari, Sudeep Sarkar
Department of Computer Science and Engineering
University of South Florida
Tampa, FL-33613

abrahmac@cse.usf.edu, sudeep@cse.usf.edu

Abstract

This work considers the problem of estimating the epipolar geometry between two cameras without needing a pre-specified set of correspondences. It is capable of resolving the epipolar geometry for cases when the views differ significantly in terms of baseline and rotation, resulting in a large number of features in one image that have no correspondence in the other image. We do conditional characterization of the probability space of correspondences based on Joint Feature Distributions (JFD) [21]. We seek to maximize the probabilistic support of the putative correspondence set over a number of MCMC iterations, guided by proposal distributions based on similarity or JFD. Similarity based guidance provides large movements (global) through correspondence space and JFD based guidance provides small movements (local) around the best known epipolar geometry the algorithm has found so far. We also propose a simple and novel method to rule out, at each iteration, correspondences that lead to degenerate configurations, thus speeding up convergence. We compare our algorithm with LO-RANSAC [2], NAPSAC [7], MAPSAC [19] and BEEM [9], which are the current state of the art competing methods, on a dataset that has significantly more change in baseline, rotation, and scale than those used in the current literature. We quantitatively benchmark the performance using manually specified ground truth corresponding point pairs. We find that our approach can achieve results of similar quality as the current state of the art in 10 times lesser number of iterations. We are also able to tolerate upto 90% outlier correspondences.

1. Introduction

One of the problems that form the backbone of vision research is establishing correspondences, particularly within the context of estimating motion and 3D structure. Corre-

spondences are required to estimate structure and motion, and knowledge of structure and motion can help establish correspondences. We can refine correspondences based on epipolar constraints induced by the camera geometry. Deviance from the epipolar constraint is the negative log likelihood of correctness of either the motion estimate or the correspondences in consideration [20]. We can refine camera and scene geometry by minimizing the re-projection errors for the inferred 3D scene geometry. Bundle adjustment is a non-linear technique to minimize the re-projection error iteratively by refining the motion estimates, the 3D scene structure and the intrinsic camera calibration parameters, all at the same time. Re-projection error gives the negative log likelihood of correctness of 3D scene structure, motion parameters and intrinsic camera calibration parameters taken together [5]. Again, a good estimate of the epipolar geometry could effectively bootstrap this iterative estimation process. Thus, estimation of epipolar geometry, as captured by the fundamental matrix, is central to motion and structure estimation. Current research works on this problem consider situations with wide baseline, which result in significant amount of features in the scene with no correspondence.

There is significant amount of work on fundamental matrix estimation [12]. Salvi et al. [1] compares the performance of fundamental matrix estimation algorithms classified as linear, iterative and robust algorithms. Among the linear algorithms, seven point algorithm and the normalized eight point algorithm [11] have been the most popular. These linear methods are used in almost all robust algorithms to get an estimate of the fundamental matrix. Robust algorithms like RANSAC [8] have been very popular for fundamental matrix estimation. Algorithms involving random sampling and consensus work in general by randomly picking up a minimal set of correspondences, estimating motion for a large number of such random selections, and finding the motion that best fits the entire set of

correspondences. In RANSAC, the best fit is found as per the cardinality of the support set of the motion. Other robust algorithms have other criterion of best fit. Many similar robust algorithms [12] such as LMedS, M-estimator, pbM-estimator, MINPRAN [17], NAPSAC [7], MLESAC [20], MAPSAC [19], Guided-MLESAC [18] have evolved over the years. PROSAC [3] is another such algorithm that randomly samples from progressively larger sets of correspondences ranked in order of higher to lower similarity scores between SIFT features. Domke et. al [6] proposes a probabilistic notion of correspondence.

It has been observed that RANSAC needs more iterations than theoretically expected. This leads us to LO-RANSAC [2] which is inspired by the fact that a set of correspondences uncontaminated by outliers might not lead to a correct epipolar geometry, thus increasing the number of iterations. This is because many of the correspondences in the uncontaminated set might lie on the same plane. Such planes are called degenerate planes or critical surfaces. Chum et. al [4] proposes one of the degeneracy detection techniques known. The epipolar geometry problem has also been looked at by using 3 correspondences [15] at a time by finding maximally stable extremal regions [14]. Weak motion models [10] and BEEM [9] algorithms have also taken similar approach. Recently, epipolar geometry was estimated using 2 correspondences [9, 16].

1.1. Related Prior Art

The approach adopted in this paper is similar in school of thought as the Maximum Likelihood Estimation SAMpling Consensus (MLESAC) [20] approach. MLESAC models the residual error distribution of correspondences in a putative set of correspondences, given a motion hypothesis, based on an assumed set of correspondences, as a mixture of Gaussian inlier error distribution and uniform outlier error distribution. These conditional probabilities due to individual residual errors are assumed to be independent. The product of all the conditional probabilities leads to a measure of likelihood of the correspondence set given the motion hypothesis. For each motion hypothesis, maximum log likelihood and the probability of validity of matches that maximizes the log likelihood are found by expectation maximization. MLESAC looks for the motion hypothesis that maximizes the likelihood of the putative correspondence set. MAPSAC [19] and Guided-MLESAC [18] are two popular variants of MLESAC. MAPSAC is the Bayesian version of MLESAC which improves upon it by maximizing the a posteriori probability instead of likelihood. Guided-MLESAC extends on MLESAC by using prior knowledge of validity of correspondences.

There are three aspects of MLESAC-school of approach that form the background against which we advance the state of the art. First, is related to the models used for inlier

and outlier correspondences. While the inlier error distribution can be quite confidently modeled as Gaussian, assuming that outlier errors exhibit uniform distribution is arguable. The nature of noise might be quite structured such as in the presence of repeated pattern. Second, MLESAC does not assume any prior knowledge of the probability of validity of a correspondence match and all matches are given the same probability of validity for a single hypothesis. Third, inliers are assumed to be mutually independent, but mutual independence of the inliers might be a not be a correct assumption. Our algorithm seeks to improve on these probable shortcomings of MLESAC and return a non-degenerate epipolar geometry.

1.2. Overview of our approach

Let us say that searches are broadly of two type : global and local. We seek to maximize a cost function by random global and local searches. The global search helps us arrive at different parameters and local searches are done to fine tune these parameters to see if a nearby solution is better. In our algorithm, global searches are done using a distribution of similarity based weights and local searches are done using Trigg's [21] Joint Feature Distribution which essentially imposes the epipolar constraint in a much unified way. We randomly choose in each iteration whether to find a similarity guided sample or a JFD guided sample. JFD guided samples are drawn from a distribution of conditional probabilities of putative correspondences conditioned on best known 'correspondel' so far. The minimal set of correspondences, e.g. 8 correspondences, needed for epipolar geometry estimation is referred to in this work as a 'correspondel'. Thus, our guidance strategy necessarily follows a Monte Carlo Markov Chain. We employ the Metropolis-Hastings MCMC method in our algorithm.

Like MLESAC, MAPSAC and Guided-MLESAC, we randomly sample from a probability distribution, but unlike them, we do not characterize the outlier and inlier error distribution separately. We choose to do conditional characterization of the probability space of correspondence by using Trigg's Joint Feature Distribution and at the same time letting it both compete and benefit from similarity guided samples. The motivation of conditional characterization comes from the idea of conditional dependence of correspondences which also subsumes locality of valid correspondences as in NAPSAC [7]. NAPSAC (N-Adjacent Points random SAMpling Consensus) is inspired by the idea of locality of valid correspondences. Seeking locality in valid correspondences might lead us to correspondences on critical surfaces or degenerate planes. Instead, we used a measure of non-degeneracy of a 'correspondel' (8 correspondence pairs). Also, BEEM shares our motivation in having a local and a global search strategy.

Quality of a fundamental matrix is another important as-

pect that needs to be quantified. Cardinality of support set has been a popular measure for this. However, deciding upon the threshold of the error is a major problem. We measure the quality of a 'correspondel' using the probabilistic support given by the total sum of negative log likelihoods.

2. The Problem

Given two images of a scene, without loss of generality, let us denote the image with smaller number of detected features to be \mathbf{I}_1 , containing a features: $\mathbf{f}_1 = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_a]$. To the feature set of the other images, \mathbf{I}_2 , add a NULL feature \mathbf{r}_0 such that $\mathbf{f}_2 = [\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_b]$. The correspondence problem is to map a features to $b + 1$ features based on the image location and local photometric attributes. Any number of features in \mathbf{f}_1 can correspond to the NULL feature in \mathbf{f}_2 . This mapping, in general, is a NP-hard problem.

Each feature in \mathbf{f}_1 and \mathbf{f}_2 has a descriptor of length l and a position of length 3 in homogeneous coordinate system. Let the descriptor set \mathbf{D} of a feature \mathbf{f}_1 be $\mathbf{D}(\mathbf{f}_1) = [d_1(\mathbf{f}_1), d_2(\mathbf{f}_1), \dots, d_l(\mathbf{f}_1)]$. We can find the similarity between each feature of \mathbf{f}_1 and \mathbf{f}_2 using their descriptors. The similarity matrix \mathbf{S} has a rows and b columns. Similarity constraint is obtained from the descriptors and epipolar constraint is obtained from the positions.

A vector \mathbf{C} of size a is used to represent the mapping between the two feature sets. The null maps are denoted by 0 and non-null maps are values from 1 to b . This is illustrated in Fig. 1. Our objective is to correctly assign these values to \mathbf{C} and also estimate the correct epipolar geometry. Let \mathbf{S} be the similarity matrix with a rows and b columns.

We initialize \mathbf{C} in a greedy fashion with the index of the highest similarity values in each of the rows of \mathbf{S} . Of these maps, the match pairs $\{(\mathbf{u}_k, \mathbf{v}_k) | k = 1, \dots, n\}$ that exhibit highest similarity measure in \mathbf{S} both in its row and column,

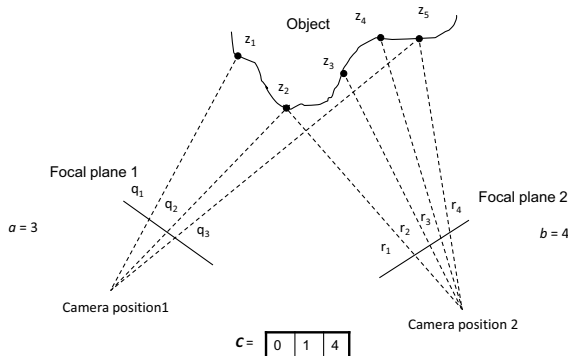


Figure 1. Illustration of representation of correspondences as a vector. \mathbf{C} represents the correspondences for features in camera 1. Zero represents null maps, for occluding or missing features

are selected to be the putative correspondences set \mathbf{X} . We represent each putative match pair as a 9 component tensor $\mathbf{x}_k = \mathbf{v}_k \otimes \mathbf{u}_k$. Thus, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$. These putative correspondences form the kernel from which rest of the correspondences will be built. Those that do not fit into the most consistent model will be mapped to NULL.

With each putative match pair, we associate a confidence measure which we refer to as similarity weights in our paper. Let the highest similarity in a row and column of a match \mathbf{x}_k be m_k . Let m_{kr} be the second highest similarity in its row and m_{kc} be the second highest in its column. We construct a weight t_k for the correspondence \mathbf{x}_k as

$$t_k = (1 - \exp^{-m_k})^2 (1 - \frac{m_{kr}}{m_k}) (1 - \frac{m_{kc}}{m_k}) \quad (1)$$

The distribution of prior weights in similarity weight vector \mathbf{t} is used to sample a correspondel. The Joint Feature Distributions, which are discussed next, are also used to sample correspondel and guide the MCMC as well. The use of the conditional JFD alleviates the need for assuming that correspondences are independent of each other, a common assumption in many random sampling methods.

3. Joint Feature Distributions

Triggs [21] proposed the concept of Joint Feature Distributions (JFDs) to provide a flexible and robust alternative to the strict and deterministic geometric constraints used for projective matching. In our context, we are particularly interested in two-camera 2D to 2D epipolar constraint. Simply put, they summarize the statistics of a given set of correspondences and does not rigidly constrain them to a deterministic geometry. That is why they are an ideal formalism to account for small non-rigid distortions and errors that will inevitably be present in any camera.

We can model the noisy mapping of the 2D features, \mathbf{u} , into the corresponding \mathbf{v} by the probability $p(\mathbf{v}|\mathbf{u})$. The form for this conditional probability will be centered around the underlying, deterministic, 2D to 2D epipolar constraint where \mathbf{F} is the 3×3 fundamental matrix.

$$\mathbf{v}^T \mathbf{F} \mathbf{u} = 0 \quad (2)$$

This equation can be linearized by considering the tensor product of the corresponding points, $\mathbf{x} = \mathbf{v} \otimes \mathbf{u}$, with dimension 9 by 1 and expressed in the form $\mathbf{A} \mathbf{x} = 0$, where \mathbf{A} is the constraint matrix formed by correspondences \mathbf{x} and \mathbf{f} is a 9×1 estimate of \mathbf{F} . This linear form implies that the JFD models are Gaussian in the tensor space,

$$p(\mathbf{x}_k) \propto \exp - \left(\frac{L_k}{2} \right) \quad (3)$$

where the negative log-likelihood function, L_k , is given by

$$L_k = \mathbf{x}_k^T \mathbf{W} \mathbf{x}_k \quad (4)$$

where k varies from 1 to n . Thus, the JFD is parameterized by the homogeneous information tensor, \mathbf{W} , which is symmetric positive definite 9 by 9 matrix generalizing the homogeneous information. We can estimate this from sample correspondences as follows.

Let us draw a random sample of correspondences θ_s of size s from \mathbf{X} . Let the indices of the sampled match pairs be \mathbf{h} . Then we build their 9 by 9 homogeneous scatter matrix $\mathbf{V} = \frac{1}{s} \sum_s \mathbf{x}_{h_i} \mathbf{x}_{h_i}^T = \frac{1}{s} \theta_s \theta_s^T$, where $\theta_s = [\mathbf{x}_{h_1}, \dots, \mathbf{x}_{h_s}]$ is the 9 by s measurement matrix and i varies from 1 to n . This measurement matrix also appears in linear matching tensor estimation. Triggs [21] has found that the inverse of this matrix is a good estimate of the information tensor $\mathbf{W} \approx \mathbf{V}^{-1}$. In practice, we have to compute $\mathbf{W} \approx (\mathbf{V} + \text{diag}(\epsilon, \dots, \epsilon, 0))^{-1}$ to regularize the inversion.

The *conditional probability* of any match pair $(\mathbf{u}_k, \mathbf{v}_k)$ in \mathbf{X} , given a set of correspondences θ_s which is a sample of size s drawn from \mathbf{X} is given by the multivariate Gaussian distribution function as follows

$$p(\mathbf{x}_k | \theta_s) = \left| \frac{\mathbf{V}(\theta_s) + \epsilon}{(2\pi)^{4.5}} \right|^{-1} \exp - \left(\frac{\mathbf{x}_k^T (\mathbf{V}(\theta_s) + \epsilon)^{-1} \mathbf{x}_k}{2} \right) \quad (5)$$

where $\mathbf{V}(\theta_s)$ is a 9 by 9 matrix constructed from θ_s , as specified earlier. We will use this conditional probability function to sample from the correspondence space.

4. BLOGS :Balanced Local and Global Search

The overall approach is shown in Fig. 2. The algorithm begins by randomly drawing a sample from the distribution of similarity weights of the putative correspondences. This is continued till we get a non-degenerate correspondel as per our threshold on the measure of non-degeneracies, discussed later. After this, depending on a parameter α , either a JFD guided sample or a similarity guided sample is drawn at each iteration. α is a control parameter used to strike a balance between the two kind of guidance for search. Similarity guided samples are independent and explore more global regions whereas JFD guided samples are conditionally dependent on the best known sample so far and thus effectively search local regions. α should be such that it allows global regions so that the MCMC does not converge to a local optimal solution and should also allow local searches near the optimal before it is discarded by a better solution via global search. Getting a better solution anyways is acceptable, but we want to have a balanced local and global search so that we are less likely to miss samples that are potentially optimal. In our experiments, we used a fixed $\alpha = 0.5$ without claiming it to be the best choice. Like in Metropolis Hastings MCMC method, we decide whether to accept the newly drawn sample, given the previous best known sample drawn so far. The decision to accept is made using a correspondence hypothesis quality measure, a de-

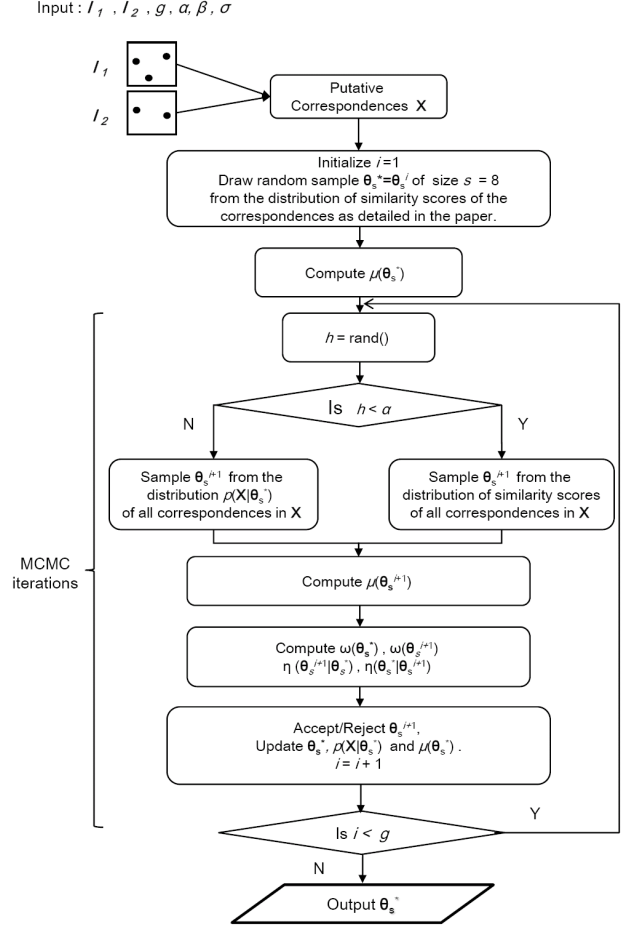


Figure 2. Flowchart of the sampling process used to find the best set of correspondences

generacy measure, and effective proposal probability. The algorithm ends after g iterations. In the following subsections, we outline the various aspects of the algorithm.

4.1. Quality Measure of the Correspondel

Sampson's error distance for a point correspondence can be taken as (negative log) likelihood of the point correspondence given a correspondel. We find the likelihood of each of the putative correspondences given a correspondel. We construct a measure of the quality of the correspondel by the sum of these likelihoods. Summation is more robust to the presence of outliers than product. Let the error, i.e. distance from epipolar line, of the k th putative correspondence be δ_k . The quality of the correspondel or the associated motion is given by

$$\mu(\theta_g) = \sum_{k=1}^n \exp \left(\frac{-\delta_k \cdot \sigma}{2} \right) \quad (6)$$

We set $\sigma = 10^4$ in our experiments.

4.2. Degeneracy Measure

Let θ_s be a sampled correspondel of size s which is 8 in our case. 8-point algorithm fails when points lie on the critical surface or degenerate planes. We can identify if the correspondences that lie on a degenerate plane.

For each putative correspondence tensor, we do an eigenvalue decomposition to get the eigenvalues. We should have two non-zero eigenvalues for both the correspondences to be informative. In practice, the ratio of the smaller eigenvalue with respect to the larger eigenvalue should be above a certain threshold value β . If all the pairs of correspondences in the correspondel are above the threshold, the correspondel is accepted as non-degenerate. $\omega(\theta_8)$ is either 0 or 1 implying degenerate and non-degenerate respectively.

$$\omega(\theta_8) = \prod_{i=1}^7 \prod_{j>i}^8 \left(\frac{\lambda_{2ij}}{\lambda_{1ij}} > \beta \right) \quad (7)$$

where λ_{1ij} and λ_{2ij} are the eigenvalues obtained by the eigen decomposition of the i th and j th correspondence tensors in the sampled correspondel. The expression in bracket is either 1 or 0 for all pairs of correspondences depending on the threshold β . In our experiments, we choose the $\beta = 0.25$.

4.3. Proposal Distribution

Putative correspondences are given by $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$. Let θ_8 be a sample of 8-tuple of correspondences drawn from \mathbf{X} . We define the importance function as follows.

$$\gamma(\theta_8) = \frac{\omega(\theta_8)\mu(\theta_8)}{\int (\omega(\theta_8)\mu(\theta_9))} \quad (8)$$

where $\mu(\theta_8)$ is the quality of the sample of 8-tuple θ_8 and $\omega(\theta_8)$ is the degeneracy measure.

JFD based proposal distribution used for local search is given by

$$e_k = \frac{p(\mathbf{x}_k|\theta_8)}{\sum_{k=1}^n p(\mathbf{x}_k|\theta_8)} \quad (9)$$

The similarity weights vector \mathbf{t} can be obtained as mentioned in section 2. The similarity based proposal distribution used for global search is given by

$$e_k = \frac{t_k}{\sum_{k=1}^n t_k} \quad (10)$$

4.4. Acceptance of Sample

We sample from \mathbf{X} based on $\mathbf{E} = \{e_1, \dots, e_n\}$, constructed either based on JFD or similarity. Let $\{j_1, \dots, j_n\}$ be the indexes of the random samples drawn from \mathbf{X} . We define the effective proposal probability of a sample $\eta(\theta_s|\theta_8^i)$ as the

product of effective probabilities from \mathbf{E} without replacement of the drawn samples.

We draw the next sample by using the information from the best known sample. Let us define the 8-tuple sample drawn at iteration i as θ_8^i . Similarly, the 8-tuple sample drawn at iteration $i+1$ would be θ_8^{i+1} . We also define the best sample known so far as θ_8^* . To start off the process, θ_8^* is initialized to θ_8^1 . Effective proposal probability at iteration $i+1$ is thus given by

$$\eta(\theta_8^{i+1}|\theta_8^*) = \frac{e_{j_1} e_{j_2} \dots e_{j_8}}{(1-e_{j_1})(1-e_{j_1}-e_{j_2}) \dots (1-e_{j_1} \dots -e_{j_7})} \quad (11)$$

The Metropolis-Hastings sampling step is now given by

$$w(\theta_8^{i+1}) = \frac{\eta(\theta_8^{i+1}|\theta_8^*)}{\eta(\theta_8^*|\theta_8^{i+1})} \quad y(\theta_8^{i+1}) = \frac{\gamma(\theta_8^{i+1})}{\gamma(\theta_8^*)} \quad (12)$$

We can also write $y(\theta_8^{i+1})$ as

$$y(\theta_8^{i+1}) = \frac{\omega(\theta_8^{i+1})\mu(\theta_8^{i+1})}{\omega(\theta_8^*)\mu(\theta_8^*)} \quad (13)$$

If $w(\theta_8^{i+1}) > 1$ and $y(\theta_8^{i+1}) > 1$

$$\theta_8^* = \theta_8^{i+1}$$

θ_8^{i+1} is accepted as optimal if both $w(\theta_8^{i+1})$ and $y(\theta_8^{i+1})$ are greater than 1. θ_8^* is the optimal 8-tuple of correspondences found so far. The first 8-tuple is sampled using \mathbf{t} and thereafter MCMC sampling is triggered. This is repeated to over g number of iterations.

5. Experiments

We compare our algorithm with LO-RANSAC [2], NAPSAC [7], MAPSAC [19] and BEEM [9], which are the current state of the art competing methods. We have used LO-RANSAC implementation as in the WBS Image Matcher (executable only), available on the internet. NAPSAC and MAPSAC implementations have been used from Torr's Toolkit for Structure from Motion. We have used SIFT [13] features. Generating a putative correspondence set is a part of our algorithm. Our putative correspondence set is a set of less than 200 correspondences with highest similarity. For each image, our putative correspondence set is also used as input to NAPSAC, MAPSAC and BEEM, to remove variability due to feature choice. However, for LO-RANSAC we had to use the set of putative correspondences computed by it because we had access to just the executable. For NAPSAC and MAPSAC algorithms we re-computed the fundamental matrices using the normalized 8-point algorithm on all the computed inliers. We did this to make sure that we get an unscaled fundamental matrix from these algorithms. We used the fundamental matrix given by the LO-RANSAC's implementation directly for comparison after finding that it is unscaled.

Table 1. Comparative performance analysis of LO-RANSAC, NAPSAC, MAPSAC, BEEM and BLOGS (our method). The first column lists the images. The second column notes the number of inlier correspondences and outlier rates for each image. This was manually determined. It captures the “hardness” of each image pair. The images are sorted in the table based on this. For each algorithm we list the Root Mean Sampson’s error for the 16 ground truth correspondences for different number of iterations. For LO-RANSAC, we did not have the flexibility to change the number of iterations. For BEEM, we find the average pixel error and average iterations if it converges within 200000 iterations. The best result for each image is shown in bold.

Image Pair	Putative Correspondence Quality (Inliers, Outlier Rate)	Typical Root Sampson’s Distance (pixel errors) from 16 hand marked correspondences on executing LO-RANSAC , and samples needed to converge	Root Mean Square Sampson’s Distance (pixel errors) from 16 hand marked correspondence in 100 executions of each algorithm										
			NAPSAC iterations			MAPSAC iterations			BEEM		BLOGS iterations		
			500	1000	5000	500	1000	5000	(pixel, error	average iterations)	500	1000	5000
KMsm	171, 1.16	1.01, 69	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.59, 164	0.54	0.54	0.54
Corridor	44, 47.6	59.98, 119	18.29	13.40	6.78	18.47	14.97	5.59	22.89, 57	13.30	9.25	3.27	
Bluna	66, 63.3	1.07 , 947	62.67	50.03	38.88	42.13	26.98	5.74	59.83, 510	7.48	6.56	3.17	
Flags	47, 75.5	85.47, 200000	46.50	31.20	9.56	22.34	17.32	6.78	16.83, 2805	8.65	5.27	1.72	
Steel mesh	31, 84.0	3.08, 45068	8.96	5.92	4.80	5.41	4.82	4.57	3.24 , 360	3.90	3.67	2.64	
Pillars	27, 85.4	74.77, 200000	40.48	40.44	36.02	39.98	35.42	33.20	28.91, 377	30.89	25.53	15.34	
Building	22, 88.3	1.11 , 23597	41.98	35.41	14.10	42.22	44.10	38.12	3.69 , 448	19.29	16.82	6.72	
University	19, 89.9	123.32, 1033	19.89	10.11	6.32	25.46	19.95	10.82	13.14, 813	6.66	6.35	6.03	
Stones	15, 92.3	65.11, 16158	65.06	58.60	55.62	53.61	52.46	50.63	did not converge	43.29	42.36	24.85	
Parking	14, 92.7	123.60, 128538	30.95	31.45	31.39	37.14	34.89	29.81	27.73, 619	26.81	22.82	13.34	
Cafeteria	11, 94.3	80.47, 200000	87.59	86.01	85.93	76.02	74.38	67.04	did not converge	69.42	65.68	59.17	
Cars	11, 94.4	74.65, 58177	84.53	84.02	79.75	81.00	75.58	75.39	did not converge	76.44	74.29	60.61	

5.1. Test Data

We have benchmarked performance on 12 image pairs including two images from another work. The Bluna image pair and KMsm image pair were taken from dataset along the WBS Image Matcher’s website. The Bluna image pair and KMsm image pair are of size 480x360, whereas the rest of the images have a size of 712x534. The test data contains image pairs that have a very wide baseline, scale changes, rotation and occlusion. Such image pairs are not sufficiently addressed in the literature. We have manually quantified the number of inliers and the outlier rate in the putative correspondence set for each image pair to give an idea about the hardness of the epipolar geometry search required for each method. These are listed, sorted according to difficulty level, in Table. 1. Note that the last two pairs in the table are particularly challenging image pairs. They test the limits of current approaches, including ours, and help motivate future research to solve such problems.

5.2. Performance Measure

For each image pair, we manually marked 16 correspondences, different from the SIFT features used to estimate the epipolar geometry. These serve as the ground truth. The root mean Sampson’s distance of these 16 hand marked correspondences serve as the quantitative performance measure. For the sake of proper comparison, we averaged the error over 100 executions of the algorithm. LO-RANSAC

returns the same epipolar geometry each time, so we did not need to do this for it.

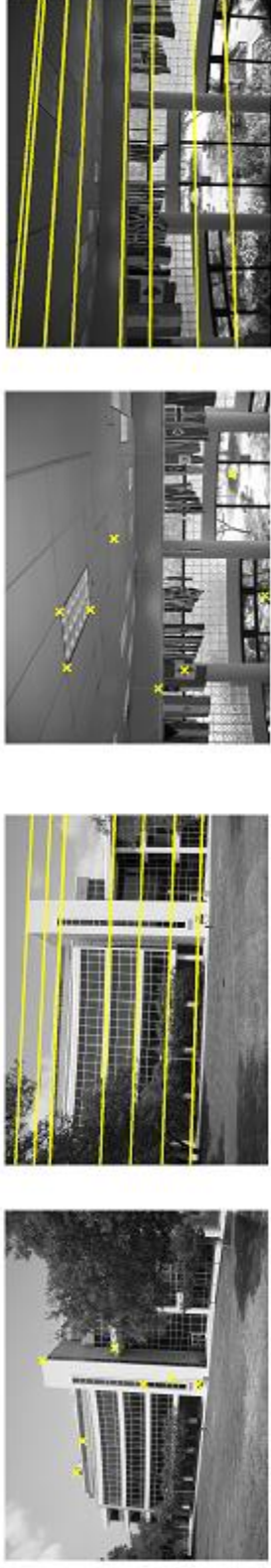
The Sampson’s distance is given by :

$$d(\mathbf{x} = \mathbf{v} \otimes \mathbf{u}, \mathbf{F}) = \sum_{i=1}^n \frac{(v_i^T \mathbf{F} \mathbf{u}_i)^2}{(\mathbf{F} \mathbf{u}_i)_1^2 + (\mathbf{F} \mathbf{u}_i)_2^2 + (v_i^T \mathbf{F})_1^2 + (v_i^T \mathbf{F})_2^2} \tag{14}$$

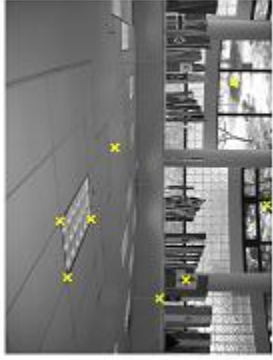
We test NAPSAC, MAPSAC and BLOGS for 500, 1000 and 5000 iterations or samples, while LO-RANSAC and BEEM execute upto their convergence. Note that our use of manually specified ground truth correspondences is a movement of a current evaluation methodology toward a more rigorous one.

5.3. Results

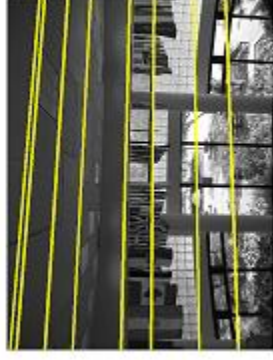
Fig. 3 showcases results on some image pairs in terms of the computed epipolar lines. Note the significant change in scale, rotation, and baseline that is represented by these images. There is significant amount of viewpoint change and occlusions. The image include both man-made and natural objects. There is significant amount of repeated structure in some of the images. For some of the ground truth points on the left images, we can shown the corresponding epipolar lines on the other image. We can see that the computed epipolar lines are of high quality.



(a) Building image pair



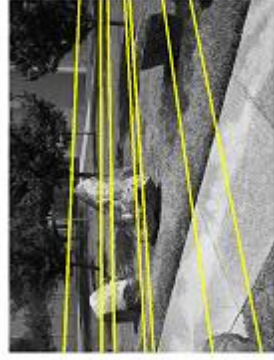
(b) Flags image pair



(c) Corridor image pair



(d) Stones image pair



(e) Parking image pair



(f) University image pair

Figure 3. Computed epipolar lines on few of the image pairs on which the proposed algorithm was tested. On the left image, we show the ground truth feature points with 'x'. The corresponding epipolar lines are drawn on the other image. These were created with `epipolar_viewer.m`, courtesy of David Leibowitz.

The quantified performances are shown in Table 1. We observe that LO-RANSAC performs well in few cases and does not do so in rest of the cases. BEEM either converges fast or does not at all. BEEM is also found to converge far from optimality at times. MAPSAC performs better than both NAPSAC and LO-RANSAC in most of the cases. We found that MAPSAC can handle high outlier rates as well. MAPSAC handling such high outlier rates is unreported to our knowledge. Our algorithm consistently performed better than NAPSAC and MAPSAC, while producing a little more pixel errors on one occasion. Very little differences might be ignored due to possible inaccuracy in hand-marked points, although points were marked with utmost care. Even on the challenge image pairs (the last two), our performance is better than the others. The results show that we attain almost the same accuracy in 500 iterations as MAPSAC attains in 5000 iterations and that BLOGS is capable of gaining more from barely sufficient number of inliers that might include degenerate inliers as well. Our algorithm is able to push to the limit of over 90 percent outliers with acceptable pixel errors in 5000 iterations in few cases. However, what value of pixel error is acceptable depends on the application. The two challenge image pairs remain unsolved, motivating us to come up with even better solutions.

6. Conclusion

The success of the BLOGS approach can be attributed to three aspects: the similarity weights-guided sampling, the JFD guided sampling, and the degeneracy criterion to weed out degenerate correspondences. The similarity-based and JFD-based sampling strategies complement each other, the former induces search over larger global regions and the later is a local search. One major problem which is still extensively researched is that an all inlier correspondel can have a very high support set even if they lie on a degenerate plane that leads to wrong epipolar geometry estimation. In this paper, we proposed and used a novel strategy to rule out correspondels with correspondences from a degenerate plane. The proposed algorithm takes care of all known aspects of epipolar geometry estimation in a simple and unified way. We compared our algorithm to MAPSAC, NAPSAC and LO-RANSAC and BEEM on many image pairs with heavily noised correspondences and found that our algorithm performed significantly better in almost all the cases.

References

- [1] X. Armangue and J. Salvi. Overall view regarding fundamental matrix estimation. *Image and Vision Computing*, 21(2):205–220, 2003.
- [2] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. In *Proceedings of DAGM*. Springer-Verlag, 2003.
- [3] O. Chum and J. Matas. Matching with prosac - progressive sample consensus. In *Proceedings of CVPR*, volume 1, pages 220–226, 2005.
- [4] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *Proceedings of CVPR*, pages I: 772–779, 2005.
- [5] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun. Structure from motion without correspondence. In *Proceedings of CVPR*, volume 2, pages 557–564, 2000.
- [6] J. Domke and Y. Aloimonos. A probabilistic framework for correspondence and egomotion. In *Proceedings of WDV06*, pages 232–242, 2006.
- [7] D.R. Myatt, P. Torr, S. Nasuto, J. Bishop, and R. Craddock. Napsac: High noise, high dimensional robust estimation - its in the bag. In *Proceedings of BMVC*, volume 2, pages 458–467, 2002.
- [8] M. Fischler and R. Boles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [9] L. Goshen and I. Shimshoni. Balanced exploration and exploitation model search for efficient epipolar geometry estimation. *IEEE Transactions on PAMI*, 30(7):1230–1242, 2008.
- [10] L. Goshen and I. Shimshoni. Guided sampling via weak motion models and outlier sample generation for epipolar geometry estimation. *IJCV*, 80(2):275–288, 2008.
- [11] R. Hartley. In defence of the 8-point algorithm. *IEEE Transactions on PAMI*, 19(6):580–593, 1997.
- [12] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, New York, NY, USA, 2000.
- [13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.
- [14] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.
- [15] J. Matas, S. Obdrzalek, and O. Chum. Local affine frames for wide-baseline stereo. In *Proceedings of ICPR*, pages 363–366, 2002.
- [16] M. Perdoch, J. Matas, and O. Chum. Epipolar geometry from two correspondences. In *Proceedings of ICPR*, pages 215–219, 2006.
- [17] C. Stewart. Minpran: A new robust operator for computer vision. *IEEE Transactions on PAMI*, 17(10):925–938, 1995.
- [18] B. Tordoff. Guided-mlesac: Faster image transform estimation by using matching priors. *IEEE Transactions on PAMI*, 27(10):1523–1535, 2005. Member-David W. Murray.
- [19] P. Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *IJCV*, 50(1):35–61, 2002.
- [20] P. Torr and A. Zisserman. Mlesac: a new robust estimator with application to estimating image geometry. *CVIU*, 78(1):138–156, 2000.
- [21] B. Triggs. Joint feature distributions for image correspondence. In *Proceedings of ICCV*, pages 201–208, 2001.