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

Aveek Shankar Brahmachari

University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the American Studies Commons

Scholar Commons Citation

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
BLOGS: Balanced Local and Global Search for Non-Degenerate Two View Epipolar Geometry

by

Aveek Shankar Brahmachari

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science
Department of Computer Science and Engineering
College of Engineering
University of South Florida

Major Professor: Sudeep Sarkar, Ph.D.
Rangachar Kasturi, Ph.D.
Dmitry Goldgof, Ph.D.

Date of Approval:
June 12, 2009

Keywords: Similarity, Joint Feature Distributions, Jump Diffusion, Degeneracy, Epipolar Geometry

© Copyright 2009, Aveek Shankar Brahmachari
DEDICATION

To my loving family
ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Prof. Sudeep Sarkar for giving me an opportunity to work with him on such a topic of broad interest in computer vision. The courses he has taught and the research advice he has delivered would certainly have a strong impact on me for years to come. Over a period of two years I have always felt superbly guided and advised by him. His flows of thoughts have always triggered mine.

I am grateful to Prof. Rangachar Kasturi and Prof. Dmitry Goldgof for instilling an insight into computer vision related subjects through their wonderful courses. I am also thankful to them for being my committee members. Wide knowledge, down to earth and compassionate nature of Prof. Kasturi, Prof. Goldgof and Prof. Sarkar has always drawn my immense respect.

I want to thank Vasant Manohar for his technical help whenever needed and for alarming me with informations at the right time. I would also thank the technical support team of the department for their prompt actions. I am thankful to all my friends whose presence made life and work easier.

Finally, I would want to express my deepest gratitude to my parents, my sister and my fiance for both the cool shade and warmth of their love and affection. I thank them for standing firmly by me and encouraging me in my technical aspirations.
# TABLE OF CONTENTS

LIST OF TABLES iii

LIST OF FIGURES iv

ABSTRACT v

CHAPTER 1 INTRODUCTION 1
1.1 Correspondence and Epipolar Geometry 2
1.2 Motivation 4
1.3 Overview of Our Approach 5

CHAPTER 2 ROBUST STATISTICS AND EPIPOLAR GEOMETRY 7
2.1 RANSAC 7
2.2 Epipolar Geometry Model Generation 8
2.2.1 Normalized 8-point Algorithm 8
2.2.2 7-point Algorithm 9
2.3 Model Quality Estimation 10
2.3.1 Sampson’s Distance 10
2.4 Degeneracy 10
2.5 Stopping Criterion 10
2.6 Learning the Epipolar Geometry 11

CHAPTER 3 STATE OF THE ART 12
3.1 Least Median of Squares 12
3.2 M-estimator and pbM-estimator 12
3.3 MINPRAN 13
3.4 NAPSAC 13
3.5 MLESAC, Guided MLESAC, MAPSAC 14
3.6 LO-RANSAC 15
3.7 DEGENSAC 15
3.8 PROSAC 16
3.9 WMM based Guidance for Epipolar Geometry Estimation 16
3.10 BEEM 17

CHAPTER 4 BLOGS: BALANCED LOCAL AND GLOBAL SEARCH 19
4.1 The Problem Definition: Our Model 19
4.2 Joint Feature Distributions 22
4.3 The Algorithm 23
4.3.1 Quality Measure of the Correspondence 25
4.3.2 Degeneracy Measure 25
4.3.3 Jump Diffusion 27
4.3.3.1 Jump using Similarity-based Proposal Distribution 28
4.3.3.2 Diffusion using JFD-based Proposal Distribution 28
4.3.4 Acceptance of Sample 29

CHAPTER 5 EXTENSION 30
5.1 Extension to Multiple Putative Sets 30

CHAPTER 6 EXPERIMENTS 32
6.1 Test Data 32
6.2 Performance Measure 33
6.3 Results 33

CHAPTER 7 DISCUSSION AND CONCLUSION 39

REFERENCES 41
LIST OF TABLES

Table 6.1 Comparative performance analysis of LO-RANSAC, NAPSAC, MAP-SAC, BEEM and BLOGS (our method). 34
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Figure showing the 2D points 'u', 'v', 3D point 'X', optical centers 'O', 'O', epipolar lines 'l', 'l', epipoles 'e', 'e' and epipolar plane.</td>
<td>8</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Illustration of representation of correspondences as a vector.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Similarity matrix $S$ from which best mutual matches are found, brighter ones being better matches.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Images showing high probability correspondence search regions given by JFDs.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Flowchart of the sampling process used to find the best set of correspondences.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Computed epipolar line on Kmsm, Corridor and Bluna image pairs.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>Computed epipolar line on Flags, Steel Mesh and Pillar image pairs.</td>
<td>36</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>Computed epipolar lines on Building, University and Stones image pairs.</td>
<td>37</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>Computed epipolar lines on Parking, Cafeteria and Cars image pairs.</td>
<td>38</td>
</tr>
</tbody>
</table>
BLOGS: Balanced Local and Global Search for Non-Degenerate Two View Epipolar Geometry

Aveek Shankar Brahmachari

ABSTRACT

The problem of epipolar geometry estimation together with correspondence establishment in case of wide baseline and large scale changes and rotation has been addressed in this work. This work deals with cases that are heavily noised by outliers. The jump diffusion MCMC method has been employed to search for the non-degenerate epipolar geometry with the highest probabilistic support of putative correspondences. At the same time, inliers in the putative set are also identified. The jump steps involve large movements guided by a distribution of similarity based priors while diffusion steps are small movements guided by a distribution of likelihoods given by the Joint Feature Distribution (JFD) \[64\]. The 'best so far' samples are accepted in accordance to Metropolis-Hastings method. The diffusion steps are carried out by sampling conditioned on the 'best so far', making it local to the 'best so far' sample, while jump steps remain unconditioned and span across the correspondence and motion space according to a similarity based proposal distribution making large movements. We advance the theory in three novel ways. First, a similarity based prior proposal distribution which guide jump steps. Second, JFD based likelihoods which guide diffusion steps allowing more focused correspondence establishment while searching for epipolar geometry. Third, a measure of degeneracy that allows to rule out degenerate configurations. The jump diffusion framework thus defined allows handling over 90% outliers even in cases where the number of inliers is very few. Practically, the advancement lies in higher precision and accuracy that has been detailed in this work by comparisons. In this work, BLOGS is compared with LO-RANSAC \[10\], NAPSAC \[40\],
MAPSAC [59] and BEEM [21] algorithm, 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 state of the art. Performance of these algorithms and BLOGS are quantitatively benchmark for a comparison by estimating the error in the epipolar geometry given by root mean Sampson’s distance from manually specified corresponding point pairs which serve as a ground truth. Not just is BLOGS able to tolerate very high outlier rates, but also gives result of similar quality in 10 times lesser number of iterations than the most competitive among the compared algorithms.
CHAPTER 1

INTRODUCTION

Over many years, correspondence establishment and epipolar geometry estimation has been researched and now the problem is expanding its horizon to wider baseline, large scale changes and rotation. However, the problem still demands a lot of research. Current research is pushing the limit of minimal possible information shared by an image pair needed to come up with meaningful 3D structure estimate of the scene. Three things are important for epipolar geometry estimation: correspondence geometry, the camera geometry and the scene geometry. Indeed, the objective is to know the motion of the cameras and the 3D scene structure. Correspondences are required to estimate structure and motion, and knowledge of structure and motion can help establish correspondences. Thus, coupled update approaches form a natural solution to this problem. Consequently, starting without prior knowledge of the feature correspondences, feature distances or feature correlations are found to come up with a similarity or confidence measure based putative feature correspondences to bootstrap the search of correspondence and epipolar geometry. Among features, SIFT [31] and GLOH [37] point features have recently become very popular almost replacing Harris corners [25].

Epipolar constraint is the strongest constraint on the search for epipolar geometry, although weakness of the epipolar constraint lies in the fact that it does not discern between correspondences along the epipolar line. Other common constraints are uniqueness, similarity and proximity. Epipolar constraint together with appearance based similarity between features is thus most often used. Further, deviance from the epipolar constraint is the negative log likelihood of correctness of either the motion estimate or the correspondences in consideration [62]. 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 [12, 7, 34, 65]. 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.

From here on, the structure of the thesis is as follows. Section 1.1, 1.2 gives a brief overview of the research works pertaining to correspondence and epipolar geometry estimation and our motivation respectively. Section 1.3 gives a brief discourse of the novelty in our approach. Chapter 2 in general talks about the RANSAC robust estimation algorithm in section 2.1 and details issues related to epipolar geometry estimation algorithm, that are model generation, model quality estimation, degeneracy, stopping criterion and learning in section 2.2, 2.3, 2.4, 2.5, 2.6 respectively. Chapter 3 lays forth the central idea of various state of the art algorithms. Chapter 4 starts with describing the problem model in section 4.1, the Joint Feature Distribution in section 4.2, the flowchart of the BLOGS algorithm in section 4.3 followed by details of our quality measure in section 4.3.1, proposed degeneracy measure in section 4.3.2, the jump-diffusion method in section 4.3.3 and acceptance of the sample in 4.3.4. Chapter 5 talks about our extension to multiple putative sets. Chapter 6 talks about our test data, experiments and results in section 6.1, 6.2, 6.3. Chapter 7 ends up with discussion and conclusion.

1.1 Correspondence and Epipolar Geometry

There are significant amount of works on fundamental matrix estimation and structure from motion [27, 28, 43, 44, 32, 24]. Shapiro et al. [51] proposed a classic eigen space approach to the correspondence problem. 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 [30, 26] 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 [15] 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 putative 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 [27] such as LMedS, M-estimator [6], pbM-estimator [47], MINPRAN [52], NAPSAC [40], MLESAC [62], MAPSAC [59], Guided-MLESAC [56] have evolved over the years. PROSAC [8] 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. IMPSAC [60] proposed hierarchical matching. Few good algorithms reported that have also laid the epipolar geometry and correspondence problem in a probabilistic framework are [4, 13]. Epipolar geometry estimation algorithms that have claimed to solve the epipolar geometry problem at very high outlier rate nearly 90% are [39, 21, 20]. Tensor voting [55] applied to the epipolar geometry problem lead to resolving problems with as much as 60% outliers. Global optimality with feasible computation to solve the correspondence problem between sparse set of points was claimed in [33]. Optimally Randomized RANSAC [9] is another recent paper on the problem.

It has been observed that RANSAC needs more iterations than theoretically expected. This leads us to LO-RANSAC [10] 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. [11] came up with a good method of detecting planar degeneracy and estimating the epipolar geometry in presence of a dominant plane. Polleyfey et al. [17] also addressed the problem of degeneracy. Other research works addressing degeneracy are [63, 61, 46]. The epipolar geometry problem has also been looked at by using
3 correspondences \([36, 38]\) at a time by finding maximally stable extremal regions \([35]\). Weak motion models \([21]\) and BEEM \([20]\) algorithms have also taken similar approach. Recently, epipolar geometry was estimated using 2 LAF correspondences \([45]\). The BEEM algorithm paper \([20]\) also proposes estimation of epipolar geometry using 2 SIFT correspondences. It uses the dominant angle of SIFT features to produce 4 correspondences from 1. This results in 8 correspondences that are fed to the normalized eight points algorithm to get an epipolar geometry. The theoretically minimum points required are 5. The 5-point algorithm proposed by Nister \textit{et al.} \([42]\) has been very popular for needing minimal point correspondences. Other related works on structure and motion can be found in \([49, 58, 19, 53, 29, 57, 41, 18, 16, 3, 5]\).

1.2 Motivation

The approach adopted in this work is similar in school of thought as the Maximum Likelihood Estimation SAmpling Consensus (MLESAC) \([62]\) 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 \([59]\) and Guided-MLESAC \([56]\) are two popular variants of MLESAC. MAPSAC is the Bayesian version of MLESAC which improves upon it by maximizing the aposteriori 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 non-degenerate epipolar geometry.

1.3 Overview of Our Approach

Let us say that searches are broadly of two types: global and local. The global search is done by jump steps and local search is done by diffusion steps. We seek to maximize a cost function by random global and local searches. The global jump search helps us arrive at different parameters and local diffusion 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 Triggs’ [64] 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 Triggs’ 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 [40]. 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).

Quality of a fundamental matrix is another important aspect 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.
CHAPTER 2

ROBUST STATISTICS AND EPIPOLAR GEOMETRY

In this chapter, we discuss the robust statistics technique RANSAC that has been the most rudimentary approach for epipolar geometry estimation. All other approaches are derived or similar to RANSAC. Moreover, all the important factors that influence the performance of state of the art epipolar geometry estimation algorithms are briefly introduced in this chapter. We begin with explaining RANSAC.

2.1 RANSAC

RANSAC stands for Random Sample Consensus. RANSAC is one of the most famous methods employed in robust statistics. The steps done in RANSAC are:

1. Randomly draw a sample of minimal size $s$ of data points needed to form a model hypothesis.
2. Generate a model hypothesis from the sample.
3. Verify the quality of the model using all the data points.
4. Store the 'best so far' model and its support set.
5. Repeat steps 1 through 4 for $i_n$ iterations.
6. 'Best so far' model and its support set are returned as result.

How many iterations does in RANSAC need to detect the model corresponding to the inlier data? Let us see how the number of iterations is derived by probabilistically. Let $p_o$ be the probability that a sampled correspondence is an outlier. Let $p_d$ be the probability of getting a de-noised minimal sample set of correspondences and $s$ be the number of correspondences in minimal sample set. Given $p_o$, $p_d$ and $s$, we can find $i_n$, the number of iterations required to attain a success rate of $p_d$ where the samples are noised by $p_o$ outlier.
rate. Probability of not getting $s$-tuple of inliers at least once in $i_n$ samples of $s$-tuples is 
\[(1 - (1 - p_o)^s)^{i_n} = 1 - p_d.\]

The number of iterations needed by RANSAC is thus given by:

\[i_n = \frac{\log(1 - p_d)}{\log(1 - (1 - p_o)^s)} \quad (2.1)\]

### 2.2 Epipolar Geometry Model Generation

#### 2.2.1 Normalized 8-point Algorithm

Write the $3 \times 3$ matrix $F$ in the form $f = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}]^T$. Let the homogeneous coordinates in $u = [u_1, u_2, 1]^T$ and $v = [v_1, v_2, 1]^T$.

\[x = v \otimes u = [v_1 u_1, v_1 u_2, v_1, v_2 u_1, v_2 u_2, v_2, u_1, u_2, 1] \quad (2.2)\]

\[A = [x_1, x_2, \cdots x_n]^T \quad (2.3)\]
A is a $2n \times 9$ matrix. Minimum value of $n$ is 8. The points need to be normalized before applying the algorithm. So, it is called a normalized eight point algorithm. The normalization is done by translation and scaling. Translation is done by translating the centroid of the points to the origin of the coordinates and scaling is done such that the RMS values of the distances of the points from the origin is $\sqrt{2}$.

We want $Af = 0$. So, we need to minimize $Af$. It is well known in linear algebra that the $f$ that minimizes $Af$ is given by the eigenvector corresponding to the least eigenvalue of $A^TA$.

One other property of the fundamental matrix is that it is singular and of rank 2. Thus, a rank 2 constraint needs to be imposed on $f$ vector. This is done by reshaping $f$ to $3 \times 3$ matrix $F$, followed by an eigen-decomposition and reconstruction using the two most significant eigen vectors with largest eigen values.

### 2.2.2 7-point Algorithm

Fundamental matrix has 7 degrees of freedom. Thus, 7 points are sufficient to get a fundamental matrix but the solution is non-linear. Note that the method is linear in using the Direct Linear Transform, but non-linear is solving for polynomial in degree three.

Let $F_1$ and $F_2$ be given by reshaping the eigen vectors corresponding to the two lowest eigen values in $A^TA$.

$$F = F_1 + \lambda F_2$$

(2.4)

The rank 2 constraint is imposed as

$$\det F = a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = 0$$

(2.5)

The above equation is a polynomial of degree 3, so 3 values of $\lambda$ give 3 fundamental matrices. Ignoring the complex solutions might lead to only 1 fundamental matrix. If a unique solution is not obtained, more points are needed to find the true fundamental matrix. All the solutions are in general used to generate model hypothesis.
2.3 Model Quality Estimation

The most important aspect of the epipolar geometry estimation algorithms is how the algorithm estimates the quality of the model hypothesis. Few methods are:

1. Maximum cardinality of support set out of all putative sets as in RANSAC, LMedS, MINPRAN.
2. Maximum likelihood of the putative correspondences as in MLESAC.
3. Maximum aposteriori of the putative correspondences as in MAPSAC.
4. Maximum weighted support of the entire putative set as in M-estimators.

2.3.1 Sampson’s Distance

The Sampson’s distance is a measure of the quality of the model as well as the correspondence whichever is judged. The Sampson’s distance of the point $x$ is given by:

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

(2.6)

2.4 Degeneracy

Degeneracy leads to loss of information leading to unstable estimates. If two points are required for estimating a line and they are almost coinciding, a minor error in precision and accuracy of the location of the points would make the estimate of the equation of the line joining the two points very unstable. Similarly a plane cannot be determined by three collinear or almost collinear points. Thus, the seven or eight points fed for model generation to a seven or eight point algorithm should carry distinct information about the structure of the scene.

2.5 Stopping Criterion

Number of iterations has been the most common stopping criterion in random or guided sampling algorithms. The paper on RANSAC [15] came up with an estimate of the number of iterations that would be needed to find an epipolar geometry supported by all inliers. It was found in [10] that the number of iterations needed to get good epipolar geometry is
not sufficient. The other common stopping criterion is number of iterations passed without updating the 'best so far'.

2.6 Learning the Epipolar Geometry

Dellaert et al. [12] proposed MCMC guidance for correspondence and epipolar geometry estimation. Guided-MLESAC [56] also proposes MCMC guidance. MCMC guidance is a method to learn from previously made hypothesis, which propels the algorithm in a guided fashion.
CHAPTER 3

STATE OF THE ART

In this chapter, ten of the most promising epipolar geometry algorithms have been briefly explained. An attempt has been made to order the algorithms both chronologically and by their central ideas. The core ideas of all the algorithms are briefly pointed out to help the reader understand the state of the art with little reading. Four of the algorithms explained here that are LO-RANSAC, NAPSAC, MAPSAC and BEEM have been compared in chapter 6. Let us start our discussion with LMedS.

3.1 Least Median of Squares

In the LMedS algorithm, the quality of the model is estimated by the median of squared errors for each data point. The model that minimizes the median of squared errors $\min\{\text{med}_i\{r_i^2\}\}$ where $r_i$ is the error, is sought and output. Evidently, LMedS does not handle more than 50% outliers.

3.2 M-estimator and pbM-estimator

M-estimator finds the weighted mean square error of all the data points for each model hypothesis. The model hypothesis for which the weighted mean square error is minimum is chosen as 'best so far' in each sampling iteration. Weighted error is negative log likelihood, and M-estimators look for maximum likelihood. Thus, M-estimators are Maximum likelihood estimators. Different weights give different M-estimators. So, in M-estimators we need to solve for $\min(\sum_{i=0}^{n}\omega(r_i^2))$ where $r_i$ is the error distance from the model of each point and $\omega$ is the weight function of the error.
pbM-estimator is projection based M-estimator where the likelihood maximization problem is given a form of kernel density estimator. pbM-estimator comes up with a weight based on the optimal Parzen window bandwidth for scaling the error estimate.

3.3 MINPRAN

MINPRAN stands for MINimum Probability of RANdomness. An important assumption in MINPRAN is that the outlier error distribution is uniform. Minimum residual error for each data point is found. The probability of finding at least $k$ points among total $n$ points such that they are within an error distance of $-r/ + r$ from the model hypothesis, given that all outliers are uniformly distributed between $-Z/ + Z$ is given by

$$F(r, k, n) = \sum_{i=k}^{n} \binom{n}{i} \left( \frac{r}{Z_0} ight)^i \left( 1 - \frac{r}{Z_0} \right)^{n-i}$$ (3.1)

It can be easily seen that the probability would increase if $r$ increases and decrease if $k$ increases. This is analogous to the motivation of standard RANSAC which looks for maximum inliers at a minimum threshold distance. If the probability is very less, the distribution is less likely to be uniform and the model hypothesis is better. $r$ and $k$ that minimizes the probability of randomness is found.

$$\mathbb{P} \left( \min_{1 \leq j \leq S, 1 \leq i \leq N} F(r_{\phi_j, i}, i, n) < F_0 \right) = p_0$$ (3.2)

In the above equation $r_{\phi_j, i}$ is the $i$th of residual errors of model $\phi_j$ that are arranged in ascending order. If the distance $r$ is $r_{\phi_j, i}$, number of inliers would obviously be $k = i$. For all $S$ models, minimum probability of randomness is found. The model with the ‘minimum probability of randomness’ is selected.

3.4 NAPSAC

NAPSAC stands for N-Adjacent Points SAmple Consensus. The motivation behind NAPSAC is that inliers tend to be closer to each other than outliers. NAPSAC uses this observation to guide its sampling strategy. NAPSAC is one of the algorithms that use the
distribution of data in high dimensional space for selection of the hypothesis model. This is how NAPSAC works to generate a hypothesis model.

1. Select an initial point \(x_0\) randomly from the set of all points.
2. Find all points within a hyper-sphere of radius \(r\) with \(x_0\) as the center.
3. If number of points found is less than the minimum required for model generation, failure.
4. Include \(x_0\) and uniformly sample rest of minimal number of points from all the points found.

NAPSAC is especially for problems in high dimensional space where even if the outlier rate is low, it is difficult to get an uncontaminated sample. This sampling strategy works fast in those cases as well. On the other hand, looking for proximity in points might lead to selection of a degenerate model.

### 3.5 MLESAC, Guided MLESAC, MAPSAC

MAPSAC is the Bayesian version of MLESAC. MLESAC stands for Maximum Likelihood Estimate for SAmping Consensus. MLESAC models the putative correspondence set as a mixture of outliers and inliers, where the error distribution of outliers is assumed uniform and that of inliers is Gaussian. Expectation Maximization is done to come up with a probability of validity of correspondence and also the maximum likelihood of the entire putative set given a model hypothesis. The hypothesis for which likelihood is maximum is accepted. MAPSAC looks for a maximum aposteriori estimate rather than maximum likelihood.

\[
p(r_i|M_h) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) e^{-\frac{r_i^2}{2\sigma^2}} + \frac{1}{w} (1 - p(v_i)) \quad (3.3)
\]

\[
\log(p(R|M_h)) = \sum_{i=1}^{n} \log \left\{ \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) e^{-\frac{r_i^2}{2\sigma^2}} + \frac{1}{w} (1 - p(v_i)) \right\} \quad (3.4)
\]

\[
Q_h = \max_{p(v)}(\log(p(R|M_h))) \quad (3.5)
\]
Equation 3.3 shows the probability of the residual error of a correspondence given a model hypothesis. Equation 3.4 assumes conditional independence of the correspondences multiplies probabilities of each correspondence and comes up with a probability of the entire putative set and a logarithm of the entire expression is done to arrive at likelihood. As in equation 3.5, EM algorithm is used to find the maximum likelihood and the probability of validity that maximizes the likelihood. The same process is repeated for all model hypotheses and the model that generates the maximum likelihood is returned.

Guided-MLESAC gets rid of the expectation maximization step by using feature correlation based priors replacing probability of validity in MLESAC’s formulation. Guided-MLESAC also extends to multiple putative sets. MAPSAC uses a prior which is not appearance based.

3.6 LO-RANSAC

LO in LO-RANSAC stands for Locally Optimized. LO-RANSAC is motivated from an observation that a sample uncontaminated by outliers might not lead to good model.

Local optimization works as follows:
1. Let the largest support set found so far be $S_u$.
2. Draw the next sample from $S_u$.
3. Let the sample size be $\min\left(\frac{S_u}{2}, 14\right)$ in case of epipolar geometry estimation. Estimation of epipolar geometry using larger sample size gives more stable estimates.
4. Use this estimate to find all inliers and repeat step 3 until no improvement is achieved.
5. Repeat step 2 through step 5 $I_k$ number of times.
6. Finally we get stable epipolar geometry and a stable support set.
7. The largest stable support set and corresponding epipolar geometry is accepted.

3.7 DEGENSAC

DEGENSAC looks for non-degenerate epipolar geometry. The seven-point algorithm is used in DEGENSAC. 21 distinct triplets can be chosen from these 7 correspondences. Only 5 of them need to be tested for homography. They are \{1,2,3\}, \{4,5,6\}, \{1,4,7\},
\{2,3,7\}, \{5,6,7\}. No 4 points should lie on the same plane. The idea is simple and easy to implement.

3.8 PROSAC

PROSAC stands for Progressive Sampling Consensus. PROSAC uses the distance ratio between the closest and second closest SIFT features arranged in an order taking progressively larger set of putative correspondences starting from minimal sample size. PROSAC comes up with a strategy to increase the number of correspondences in descending order of similarity score one at a time and draws a number of samples from the progressively growing set such that it maintains the same sampling distribution as in simple RANSAC. Though PROSAC exploits the rank order of the putative correspondence, it does not exploit the similarity score.

3.9 WMM based Guidance for Epipolar Geometry Estimation

Weak Motion Models are used to roughly approximate the motion between two images. \(N_w\), WMMs are found and the geometric distance of points from these WMMs are modeled as a mixture of Gaussian inlier error distribution and uniform outlier error distribution. WMMs are basically affine transformations which require 3 correspondences to be defined. An outline of the algorithm is as follows:

1. Generate an outlier sample from the correspondences apart from the putative set.
2. Assume different outlier rates in set \(\epsilon\) gradually upto a user specified maximum \(\epsilon_j\) where \(j\) is the number of outlier rates specified.
3. Generate random WMMs for the outlier rate.
4. Fine tune the estimate of the outlier rate, as the assumed ones are not accurate.
5. Estimate inlier probabilities.
6. Estimate numbers of iterations \(N_s\) needed by LO-RANSAC step.
7. Guide the algorithm using the inlier probabilities if \(N_s\) is less than the maximum specified by user, otherwise go to the next outlier rate in \(\epsilon\).
8. If the stopping criterion is met, stop.
If the number of inliers indicated by the mixture model is approximately same as the number of inliers found by the guided LO-RANSAC step, then stop.

3.10 BEEM

Balanced Exploration and Exploitation Model (BEEM) search for efficient epipolar geometry estimation proposed few novel ideas. BEEM approach is closest to our approach in having a local as well as global search, but our approach is more unified and probabilistic and shares its motivation with MAPSAC. In the BEEM algorithm, there are two kinds of exploration and one kind of exploitation. While exploration is both local and global, exploitation is only local. BEEM uses the distance ratio between the closest and second closest SIFT feature to sample from. For ratios above 0.8 as suggested in [31], the distribution is uniform. Sampling from this distribution forms BEEM’s global search strategy. They call it global exploration. BEEM employs two local search strategies. BEEM algorithm estimates epipolar geometry using only 2 SIFT correspondences. It maintains a set of inliers and samples one correspondence from the inlier set and another correspondence from outside the inlier set or inside the inlier set with a given probability in order to escape degeneracy. This is called local exploration. Another local search is the local optimization step as in LO-RANSAC and it is called exploitation. In the exploration steps, if a model with larger support set is found, the next step is exploitation and thereafter the inlier set is maintained as the best so far. If a model with larger support set is not found, the quality of the model is estimated. The quality of the model is the probability that the best model found so far is not an outlier model. If the stopping criterion is not yet met in the quality estimation step, local exploration is chosen with the probability found, otherwise global search is chosen. This probability is also used to escape degeneracy in local exploration step. The stopping criterion is met if the ’best so far’ hypothesis is not updated for a number of iterations.

While we fix the probability of choosing local and global searches unbiased in BLOGS, BEEM prefers to dynamically allocate this probability by the quality of the model it is sampling at every iteration. The major novel approach in BEEM is the ability to compute
epipolar geometry using just 2 SIFT correspondences. BEEM exploits the dominant angle information in SIFT features to come up with 3 extra points per feature on the plane containing the SIFT feature. As mentioned in [27], 2 distinct homographies in a scene can yield an epipolar geometry. In BEEM, eight points thus found are supplied to eight points algorithm to get an estimate of the epipolar geometry. Sampling just two correspondences to get a model makes BEEM extremely fast compared to other algorithms.
4.1 The Problem Definition: Our Model

Given two images of a scene, without loss of generality, let us denote the image with smaller number of detected features to be $I_1$, containing $a$ features: $f_1 = [q_1, q_2, \ldots, q_a]$. To the feature set of the other images, $I_2$, add a NULL feature $r_0$ such that $f_2 = [r_0, r_1, \ldots, r_b]$. 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 $f_1$ can correspond to the NULL feature in $f_2$. This mapping, in general, is a NP-hard problem.

Figure 4.1. Illustration of representation of correspondences as a vector.
How many such distinct maps are possible? The number of NULL correspondences could be from 0 to $a$. We need to choose any one number given by $^aC_i$. The number of non-NULL correspondences would be number of permutations of the rest given by $^bP_{a-i}$. Thus, number of possible maps is a summation of the product of the two.

$$n_{map} = \sum_{i=0}^{a} ^aC_i \cdot ^bP_{a-i}$$  \hspace{1cm} (4.1)

Only one out of $n_{map}$ mappings is correct. For large values of $a$ and $b$, $n_{map}$ would be very high. Combinatorial explosion is very costly. Thus, approximation algorithms such as random sampling become a natural solution.

Each feature in $f_1$ and $f_2$ has a descriptor of length $l$ and a position of length 3 in homogeneous coordinate system. Let the descriptor set $D$ of a feature $f$ be $D(f) = [d_1(f), d_2(f), \ldots, d_l(f)]$. We can find the similarity between each feature of $f_1$ and $f_2$ using their descriptors. Let $S$ be the similarity matrix with $a$ rows and $b$ columns. For various possible similarity measures that can be used, the reader is referred to [48]. Similarity constraint is obtained from the descriptors and epipolar constraint is obtained from the positions.

A vector $C$ 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. 4.1. Our objective is to correctly assign these values to $C$ and also estimate the correct epipolar geometry.

We initialize $C$ in a greedy fashion with the index of the highest similarity values in each of the rows of $S$. Of these maps, the match pairs $\{(u_k, v_k)|k = 1, \ldots, n\}$ that exhibit highest similarity measure in $S$ both in its row and column, are selected to be the putative correspondences set $X$. We represent each putative match pair as a 9 component tensor $x_k = v_k \otimes u_k$. Thus, $X = [x_1, x_2, \ldots, x_n]$. These putative correspondences form the kernel from which rest of the correspondences will be built. Random models are drawn from the putative set. 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 $x_k$ 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 $x_k$ as

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

(4.2)

The figure 4.2 is from the famous seminal paper by Scott and Longuet-Higgins [50] on spectral correspondences. We use the figure to show that putative correspondences are the best mutual match.

The Joint Feature Distributions, which are discussed next, are used to sample correspondences 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.

4.2 Joint Feature Distributions

Triggs [64] 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, JFDs are the joint probability distributions over the parameters of the corresponding 2D to 2D features. 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, $u$, into the corresponding $v$ by the probability $p(v|u)$. The form for this conditional probability will be centered around the underlying, deterministic, 2D to 2D epipolar constraint where $F_{3 \times 3}$ is the 3x3 fundamental matrix.

$$v^T F_{3 \times 3} u = 0 \quad (4.3)$$

This equation can be linearized by considering the tensor product of the corresponding points, $x = v \otimes u$, with dimension 9 by 1 and expressed in the form $A F_{9 \times 1} = 0$, where $F_{9 \times 1}$ is reshaped form of $F_{3 \times 3}$. This linear form implies that the JFD models are Gaussian in the tensor space,

$$p(x_k) \propto \exp \left(- \frac{L_k}{2} \right) \quad (4.4)$$

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

$$L_k = x_k^T W x_k \quad (4.5)$$

where $k$ varies from 1 to $n$. Thus, the JFD is parameterized by the homogeneous information tensor, $W$, 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 $\theta_s$ of size $s$ from $X$. Let the indices of the sampled match pairs be $h$. Then we build their $9 \times 9$ homogeneous scatter matrix

$$V = \frac{1}{s} \sum_s x_{hi} x_{hi}^T = \frac{1}{s} \theta_s \theta_s^T,$$

where $\theta_s = [x_{h1}, \ldots, x_{hs}]$ is the $9 \times s$ measurement matrix and $i$ varies from 1 to $n$. This measurement matrix also appears in linear matching tensor estimation. Triggs [64] has found that the inverse of this matrix is a good estimate of the information tensor $W \approx V^{-1}$. In practice, we have to compute $W \approx (V + \text{diag}(\epsilon, \ldots, \epsilon, 0))^{-1}$ to regularize the inversion.

The conditional probability of any match pair $(u_k,v_k)$ in $X$, given a set of correspondences $\theta_s$ which is a sample of size $s$ drawn from $X$ is given by the multivariate Gaussian distribution function as follows

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

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

In figure 4.3, first an image is shown with a point. Next, high probability correspondence region over second image found using JFD based on entire putative set is shown, followed by another image showing high probability correspondence region over second image found using JFD based on the best correspondel found. Note that the high probability region ellipse in 'b' does not cover the corresponding point while that in 'c' does. Also note that JFD for accurate set of correspondence in narrower.

4.3 The Algorithm

We start with the flowchart of our algorithm. All the necessary steps in the flowchart are detailed in the sections of this chapter. The overall approach of BLOGS shown in Fig. 4.4. 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-degeneracy, discussed later. After this, depending on a parameter $\alpha$, either a JFD guided sample or a similarity guided sample is drawn at each iteration. $\alpha$ is a control parameter used to strike a balance
Figure 4.3. Images showing high probability correspondence search regions given by JFDs.

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

### 4.3.1 Quality Measure of the Correspondence

Sampson’s error distance for a point correspondence can be taken as (negative log) likelihood of the point correspondence given a correspondence. We find the likelihood of each of the putative correspondences given a correspondence. We construct a measure of the quality of the correspondence 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 $\delta_k$. The quality of the correspondence or the associated motion is given by

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

(4.7)

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

### 4.3.2 Degeneracy Measure

Let $\theta_s$ be a sampled correspondence 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
Figure 4.4. Flowchart of the sampling process used to find the best set of correspondences.
be informative. In practice, the ratio of the smaller eigenvalue with respect to the larger eigenvalue should be above a certain threshold value $\beta$. 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.

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

(4.8)

where $\lambda_{1,ij}$ and $\lambda_{2,ij}$ 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 $\beta$. In our experiments, we choose the $\beta = 0.25$.

### 4.3.3 Jump Diffusion

MCMC sampling has been popular in correspondence and epipolar geometry estimation [12]. Jump diffusion process by introduced by Grenander and Miller [23]. Green’s [22] paper more specifically deals with image processing and vision related research. Han et al. [14] has used jump-diffusion framework for range image segmentation more recently. In the literature [54], many hybrid samplers have been reported. Jump-diffusion is one such hybrid sampler. A wide discourse on jump diffusion can be found in [2]. While jump is a global move, diffusion is a local move. We use appearance based prior distribution to draw jump move samples and epipolar constraint based JFD for a distribution of likelihoods is used to draw diffusion move samples. We see priors as a cue to global jump based search and likelihoods as a cue to local diffusion based search. The 'best so far' found sample is improved by both jump and diffusion moves under several MCMC iterations.

1. Jump: Jump moves explore the space of correspondence and motion unbiased on any previous learning. This makes jump a global search process. Jump is a global exploration move.

2. Diffusion: Diffusion moves explore the space of correspondence and motion biased on the best learnt motion so far. This makes diffusion a search process local to the best found motion so far. Diffusion is a local exploration move and exploitation move as well.

27
3. Balance between jump and diffusion: Coordination can be done using a parameter supplied to the algorithm or can be dynamically set as the algorithm proceeds. We prefer to fix it as a parameter. As the quality of the motion increases, both jump moves and diffusion moves gradually are unable to improve the quality of motion any further. While diffusion moves are more likely to get stuck in a maxima, jump steps over time attempt to upstage the quality of motion with which diffusion moves can compete and benefit, providing better results. Diffusion steps always keep on trying to improve the 'best so far' by using it to draw more samples, even when jump steps do not lead to improvements. Since, we see that jump and diffusion are equally important; we keep the parameter unbiased for both jump and diffusion.

Putative correspondences are given by $X = \{x_1, x_2, ..., x_n\}$. Let $\theta_8$ be a sample of 8-tuple of correspondences drawn from $X$. We define the importance function as follows.

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

(4.9)

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

4.3.3.1 Jump using Similarity-based Proposal Distribution

We sample correspondels from the distribution of similarity weight vector $t$ 4.2. The proposal distribution in this case is given by

$$e_k = \frac{t_k}{\sum_{k=1}^{n} t_k}$$

(4.10)

4.3.3.2 Diffusion using JFD-based Proposal Distribution

Our proposal distribution is given by

$$e_k = \frac{p(x_k|\theta_8)}{\sum_{k=1}^{n} p(x_k|\theta_8)}$$

(4.11)

The $p(x_k|\theta_8)$ is obtained from equation 4.6
4.3.4 Acceptance of Sample

We sample from $X$ based on $E = \{e_1, \ldots, e_n\}$, constructed either based on JFD or similarity. Let $\{j_1, \ldots, j_n\}$ be the indexes of the random samples drawn from $X$. We define the effective proposal probability of a sample $\eta(\theta_s | \theta^*_s)$ as the product of effective probabilities from $E$ 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 $\theta^i_s$. Similarly, the $8$-tuple sample drawn at iteration $i + 1$ would be $\theta^{i+1}_s$. We also define the best sample known so far as $\theta^*_s$. To start off the process, $\theta^*_s$ is initialized to $\theta^0_s$. Effective proposal probability at iteration $i + 1$ is thus given by

$$\eta(\theta^{i+1}_s | \theta^*_s) = \frac{e_{j_1} e_{j_2} \cdots e_{j_8}}{(1 - e_{j_1})(1 - e_{j_1} - e_{j_2}) \cdots (1 - e_{j_1} \cdots - e_{j_8})}$$  \hspace{1cm} (4.12)

The Metropolis-Hastings sampling step is now given by

$$w(\theta^{i+1}_s) = \frac{\eta(\theta^{i+1}_s | \theta^*_s)}{\eta(\theta^*_s | \theta^{i+1}_s)} \quad y(\theta^{i+1}_s) = \frac{\gamma(\theta^{i+1}_s)}{\gamma(\theta^*_s)}$$  \hspace{1cm} (4.13)

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

$$y(\theta^{i+1}_s) = \frac{\omega(\theta^{i+1}_s) \mu(\theta^{i+1}_s)}{\omega(\theta^*_s) \mu(\theta^*_s)}$$  \hspace{1cm} (4.14)

If $w(\theta^{i+1}_s) > 1$ and $y(\theta^{i+1}_s) > 1$, $\theta^*_s = \theta^{i+1}_s$  \hspace{1cm} (4.15)

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

In the previous section, we saw that the two important things that make a correspondence more probable of being correct are its similarity value and its distinctness in its row and column. In this section, we describe how we apply these criteria without fixing the putative set. We start with a set of correspondence that exhibit high similarities above a certain threshold. All non-zero rows and columns remaining in the similarity matrix after thresholding are considered without applying the uniqueness constraint that allows only diagonal permutations. For each non-zero similarity value in the matrix, a penalty based on the sum of the number of other non-zero similarity values in its row and column is imposed, to take care of the distinctness criterion. A proportional sampling probability is derived and we arrive at a distribution of similarity weights. While sampling is done from the distribution of similarity weights in the jump steps, diffusion steps are guided by JFD based distribution of probabilities for all correspondences that have non-zero similarity values in the similarity matrix after thresholding. Initially, we threshold the similarity values using the standard deviation at each row and column in the similarity matrix and a specified constant assumed to be same for all images. We suggest this constant to be \( \tau = 1 - 1/e = 0.63 \) according to the form of our similarity measure, although it is an arbitrary choice.

Let the similarity matrix be \( S \) with size \((a, b)\). Let \( S_{ij} \) be an element of \( S \).

\[
S_{ij} = (1 - \exp \left( -\frac{1}{D_{ij}} \right))
\]  

(5.1)
where $D_{ij}$ is the distance between the descriptors of $i$th and $j$th features. $S_{ij}$ is subject to a threshold to meet the high similarity criterion. To apply the distinctness criterion we come up with a matrix $G$ whose elements $G_{ij}$ are defined as follows.

$$S_{ij} = 0 \quad \text{if} \quad S_{ij} < 3\sigma(R_i) + \tau \quad \text{or} \quad S_{ij} < 3\sigma(C_j) + \tau \quad (5.2)$$

$$G_{ij} = \frac{S_{ij}}{-\log(N_i + N_j)} \quad (5.3)$$

where $N_i$ and $N_j$ are the number of non-zero similarity values in $i$th row and $j$th column respectively.

The jump step uses the distribution of weights according to $G$ and for diffusion step conditional probability values based on JFDs for the correspondences with non-zero values in $G$ are estimated. The distribution of these conditional probability values is used for the diffusion steps. These conditional probability values are stored in $G'$.

Samples are drawn from $G$ and $G'$. While $G$ does not change, $G'$ keeps on changing as per the sample correspondel that generated the best model so far. A quality measure is found for each non-zero similarity feature in $G$ or $G'$ whichever is used in the jump or diffusion step respectively. A sum of negative log likelihood of correspondences given the model is found as a quality measure. One to one mapping or uniqueness is later imposed by applying Hungarian algorithm on $G$, or by summing in greedy fashion while taking care of mutual exclusion of rows and columns, or by considering the sum of only those values which are maximum in their row and column. Summing in greedy fashion has been a choice in our case, due to its simplicity. Moreover, complexity of the Hungarian algorithm is far too high and our need to distinguish inliers from outliers rather than maximizing the sum makes greedy strategy a better choice.
CHAPTER 6

EXPERIMENTS

We compare our algorithm with LO-RANSAC [10], NAPSAC [40], MAPSAC [59] and BEEM [21] 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. BEEM [21] implementation is also available on the Shimshoni’s website. We have used SIFT [31] 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.

6.1 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. 6.1. 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.

6.2 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 serves 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. BEEM converges on meeting its stopping criterion. Average error and average iterations were used for to evaluate the performance of BEEM.

The Sampson’s distance is given by:

$$d(\mathbf{x} = \mathbf{v} \otimes \mathbf{u}, \mathbf{F}) = \sum_{i=1}^{n} \frac{(v_i^T \mathbf{F} u_i)^2}{(\mathbf{F} u_i)_1^2 + (\mathbf{F} u_i)_2^2 + (v_i^T \mathbf{F})_1^2 + (v_i^T \mathbf{F})_2^2}$$  \hspace{1cm} (6.1)

6.3 Results

We test NAPSAC, MAPSAC and BLOGS for 500, 1000 and 5000 iterations or samples, while LO-RANSAC and BEEM execute up to their convergence. While LO-RANSAC always converges at the same number of iterations, BEEM converges after different number of iterations each time. Note that our use of manually specified ground truth correspondences is a movement of a current evaluation methodology toward a more rigorous one.

In the table 6.1, 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
Table 6.1. Comparative performance analysis of LO-RANSAC, NAPSAC, MAPSAC, BEEM and BLOGS (our method).

<table>
<thead>
<tr>
<th>Image Pair</th>
<th>Putative Correspondence Quality (Inliers, Outlier Rate)</th>
<th>Typical Sampson’s Distance (pixel errors) from 16 hand marked correspondences</th>
<th>Root Mean Square Sampson’s Distance (pixel errors) from 16 hand marked correspondence in 100 executions of each algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NAPSAC iterations</td>
<td>MAPSAC iterations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>KMsm</td>
<td>171, 1.16</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>Corridor</td>
<td>44, 47.6</td>
<td>59.98</td>
<td>119</td>
</tr>
<tr>
<td>Bluna</td>
<td>66, 63.3</td>
<td>1.07</td>
<td>947</td>
</tr>
<tr>
<td>Flags</td>
<td>47, 75.5</td>
<td>85.47</td>
<td>20000</td>
</tr>
<tr>
<td>Steel mesh</td>
<td>31, 84.0</td>
<td>3.08</td>
<td>45068</td>
</tr>
<tr>
<td>Pillars</td>
<td>27, 85.4</td>
<td>74.77</td>
<td>200000</td>
</tr>
<tr>
<td>Building</td>
<td>22, 88.3</td>
<td>1.11</td>
<td>23597</td>
</tr>
<tr>
<td>University</td>
<td>19, 89.9</td>
<td>123.32</td>
<td>1033</td>
</tr>
<tr>
<td>Stones</td>
<td>15, 92.3</td>
<td>65.11</td>
<td>16158</td>
</tr>
<tr>
<td>Parking</td>
<td>14, 92.7</td>
<td>123.60</td>
<td>128538</td>
</tr>
<tr>
<td>Cafeteria</td>
<td>11, 94.3</td>
<td>80.47</td>
<td>200000</td>
</tr>
<tr>
<td>Cars</td>
<td>11, 94.4</td>
<td>74.65</td>
<td>58177</td>
</tr>
</tbody>
</table>

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. The best result for each image is shown in bold. For BEEM, we find the average of the pixel error and average iterations if it converges within 200000 iterations.

The quantified performances are shown in Table 6.1. We observe that LO-RANSAC performs well in few cases and does not do so in rest of the cases. 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. BEEM converges fast on meeting its own stopping criterion, but the result was improved only once and was almost the same on one other image pair. It did not converge even after 200000 iterations on 3 image pairs. Very little differences might be ignored due to possible inaccuracy in hand-marked points, although points were marked with utmost care. Even on the hard images (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
Figure 6.1. Computed epipolar line on Kmsm, Corridor and Bluna image pairs.

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.
Figure 6.3. Computed epipolar lines on Building, University and Stones image pairs.
Figure 6.4. Computed epipolar lines on Parking, Cafeteria and Cars image pairs.
The problem of feature correspondence and motion estimation still remains a challenge. Despite many years of research, the accuracy (correctness) and precision (convergence to the same solution on repetitions) of the state of the art algorithms for feature correspondence and motion estimation in case of wide baseline, highly scaled and rotated image pairs is far from being practically exploitable. Inlier correspondences may not yield correct epipolar geometry if the inlier correspondences are degenerate. On the other hand, correct epipolar geometry might not be supported by only inlier correspondences. Outlier correspondences along the epipolar lines are accepted as inliers. The irony is that outlier might be used to come up with a good motion model but inlier might not. Triggs’ Joint Feature Distribution comes in picture here as it gives a high probability narrow search elliptical region with its major axis aligned along the epipolar line. And we also weed out degenerate correspondences by a dimensionality check such that no pair of correspondences in the sampled correspondences has a dimensionality of 1 rather than 2. JFD based guidance would tend to sample more of correct correspondences than the conventional approaches. We used this strength of JFDs and address the weaknesses by not using it as a measure of quality and also weeding out the degeneracy. JFD based guidance is only as correct as the sampled correspondences used to estimate the JFD and it shares its strength and weaknesses with the 8-point algorithm except for being focused on a region rather than a line and not requiring the rank 2 constraint being explicitly imposed.

The success of the BLOGS approach can be attributed to three aspects: the similarity weights-guided global search, 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. In this paper, we proposed and used a novel strategy to rule out correspondels with correspondences from a degenerate plane. This takes us a step further theoretically and experimentally. The proposed algorithm takes care of all known aspects of epipolar geometry estimation in a simple and unified way. We compared our algorithm to BEEM, MAPSAC, NAPSAC and LO-RANSAC on many image pairs with heavily noised correspondences and found that our algorithm performed better in almost all the cases.
REFERENCES


