# IMPSAC: Synthesis of Importance Sampling and Random Sample Consensus

P. H. S. Torr and C. Davidson
Microsoft Research Ltd
7 JJ Thompson Avenue, Cambridge,CB3 OFB, UK
philtorr@microsoft.com

October 2002

Technical Report
MSR-TR-2002-99

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
http://www.research.microsoft.com

[1]

# 1 Introduction

The goal of this work is to obtain accurate matches and epipolar geometry between two images of the same scene, where the motion is unlikely to be smooth or known *a priori*. Once the matches and two view image relation have been recovered, they can be used for image compression, for building 3D models [3, 33, 35, 48], for object recognition [19], for extraction of images from databases [31]

It transpires that the correspondence problem is one of the most difficult parts of structure recovery, especially when these images are far apart (the *wide baseline* problem) or when they undergo large rotations (the *large image deformation* problem). Whereas small baseline image matching technology have made large advances over the past decade [1, 2, 3, 7, 14, 22, 35, 42, 50], there has been comparatively little progress in wide baseline matching technology. Furthermore, the small baseline methods do not work on every image pair. For example, feature based cross correlation methods may fail if (1) there are insufficient features in the image pair, (2) there is too much repeated structure for features to get a good match, or (3) there is an image deformation that causes the cross correlation to fail.

There has been some work on solving these problems. Pritchett and Zisserman [27] present a set of recipes for special cases, but no unified theory of how to solve the problem in general. Cham and Cipolla [6] present a multi scale method for feature matching when making mosaics. The work is valid only if there is no parallax, i.e. if the image motion is governed by a homography. Furthermore the formulation is flawed as it propagates parameters using the estimate at the coarser level as a prior for the estimate at the finer, but since the images at fine and coarse resolution are not independent, the prior and likelihood are not independent. This leads to an erroneous posterior, which is then used (in their method) as a prior for the next level, compounding the error.

As suggested in Tao [33] low level segmentation may be combined with matching, recent work by Matas *et al* [20] suggest the use of extremal regions for wide baseline stereo, this approach seems to hold some promise.

Another style of approach is to first detect features in both images and then use a measure, invariant to large image deformations, to establish an initial set of correspondences between the images. Such invariant features have been suggested in [19, 24, 30, 31, 45]. Making a measure invariant to large image deformations means that many false matches (outliers) will be made, as a rule of thumb the more invariant the measure the more outliers will be produced. Thus a parametric motion model (e.g. homography or fundamental matrix) must be estimated by a robust estimation method to remove the outliers. However the sheer number of outliers renders most robust estimators useless (some work developing new estimators capable of dealing with a greater proportion of outliers is reported in [25]), unless the number of parameters in the model is small (for instance an affine transformation is used in [19]).

Rather than separate the process of matching and of estimating the epipolar geometry, we propose to fuse the two into the same algorithm. This means that the ambiguity inherent in matching using invariant measures can be avoided. A related work to this paper, that appeared at the same time as this work, is that of Dellaert *et al* [7], this paper too proposes simultaneous estimation of both matches and epipolar constraint in a Bayesian framework. The method of Dellaert *et al* is designed for the small baseline case, however it would be interesting to extend their methods (based on MCMC and EM) to the wide baseline case.

The method presented here is also a Bayesian solution to the image deformation and wide baseline matching problems. It has long been known that a coarse to fine approach can be beneficial in matching problems [4, 28]. In this approach information about the matching (and anything else of interest e.g epipolar geometry) is passed from the coarser levels to the finer. The coarse to fine strategy is beneficial for a number of reasons. It furnishes a solution to the wide baseline problem because the search window, and thus the number of potential false matches per corner, is reduced at the finer levels. Furthermore, at the coarser level, it is less computationally intensive to estimate the global image deformation, by testing different hypotheses for the deformation of the cross correlation between image patches.

Ideally the information to be transmitted would be the posterior distribution of the parameters at the coarser level. Encoding this posterior distribution and its relation to the finer level is an intricate task, not least because the normalization constant of the distribution is unknown. Four powerful statistical methods are enlisted to create

[1] Part of this work first appeared in [38].

a solution: (1) to represent the distributions as a set of particles, this overcomes the problem of representation. If a parametric form of the distribution is required a mixture of Gaussians can be fitted to these particles. (2) the use of importance sampling to generate the particles as unbiased draws from the posterior distribution, (3) RANSAC to generate the initial set of particles, (4) MCMC to redistribute the particles so that they better approximate the posterior. In this way the posterior distribution at the coarse level is used as an importance sampling function to draw samples from the posterior distribution at the finer level. This allows for a statistically unbiased estimate of the posterior at the next level down, which would not be the case if we simply used the posterior at the level above as a prior for the posterior at the level below. As a result, the epipolar geometry is estimated by using features at many different scales, solving the problem of having to select this scale manually. Furthermore the generation of importance sampling functions using RANSAC, dubbed IMPSAC may be of more general interest.

The paper is organized as follows. The types of two view relations that might arise are described in Section 2, and the likelihood of the matches given these relations in Section 2.1. It is shown how to derive a Bayesian likelihood in two cases, (a) where the matches are given, and then (b) where the matching is unknown. Existing geometry based matching methods are reviewed in Section 3, they comprise two stages: (a) estimate best cross correlation matches, (b) estimate epipolar geometry using a robust estimator. The Bayesian algorithm for optimal robust estimation of two view geometry, MAPSAC, is described in Section 3.1. However this approach breaks down for the image deformation and wide base line cases. The IMPSAC algorithm is described in Section 4 it consists of three stages at each level of the coarse to fine image pyramid, (1) first a set of particles are estimated which encode the posterior of the two view relation using, at the first level, random sampling or in finer levels SIR to propagate information from coarse to finer levels. (2) Then MCMC is used to adjust the particles to better represent the posterior. (3) Then a new importance sampling function is made from those particles to be used in generating particles at the next finer level. These steps are described in Sections 4.1, 4.2 and 4.3. The results are given in Section 5, where the algorithm is demonstrated on the wide baseline and the image deformation problems.

**Notation** A 3D scene point projects to homogeneous three vectors $\mathbf{x}^1$, and $\mathbf{x}^2$ in the first and second images. The inhomogeneous coordinate of an image point of the $i$th point in the $j$th image is $(x_i^j, y_i^j)$. The $k$th correspondence between two views will also be represented by the vector $\mathbf{m}_k = (x_i^1, y_i^1, x_j^2, y_j^2)^\top$, meaning that feature $i$ corresponds to feature $j$, if the matches are given then the features are ordered such that $\mathbf{m}_i = (x_i^1, y_i^1, x_i^2, y_i^2)^\top$. The set of all observed point matches between two views will be denoted by $\mathcal{D}$ ($\mathcal{D}$ standing for *data* as these are the observations). The probability density function (p.d.f.) of $x$ given $y$ is $p(x|y)$. $\mathcal{M}$ is a hypothesized *model* for the data, consisting of a noise model and two view relation $\mathcal{R}$ with parameters $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$, whereas, $\Re^D$, is $D$ dimensional Euclidean space, typically bounded. $\mathcal{V}$ is the algebraic variety for $\mathcal{R}$ with parameters $\boldsymbol{\alpha}$. An algebraic variety (or just variety), $\mathcal{V}$, is defined as the set of points in $\Re^D$ satisfying a polynomial set of equations $g_q(\mathbf{m}_q) = 0$, $q = 1 \ldots Q$ [47]. Furthermore within this paper $\mathcal{V}$ is also a manifold, of dimension $d$, codimension $r = D - d$. Note if the functional forms of $g_q()$ are not polynomial then $\mathcal{V}$ is a manifold but not a variety, this distinction is made for pedantry and is unimportant to the rest of the analysis.

# 2 The Two View Relations

This section describes the putative motion models that, under the assumption of rigidity, constrain the motion of points between two views. Each motion model, $\mathcal{M}$, consists of a noise model (including an outlier model), and a *relation*, $\mathcal{R}$, described by a set of *parameters*, $\boldsymbol{\alpha}$, that define $Q$ *implicit functional relationships* on the image coordinates between two views, i.e. $g_q(\underline{x}_i^1, \underline{y}_i^1, \underline{x}_i^2, \underline{y}_i^2, \underline{\boldsymbol{\alpha}}) = 0, i = 1 \ldots n, q = 1 \ldots Q$. Which may also be written $\mathbf{g}(\underline{\mathbf{m}}_i, \underline{\boldsymbol{\alpha}}) = \mathbf{0}$, where $\mathbf{0}$ is the zero vector.

Two examples of $\mathcal{R}$ are considered: (a) the Fundamental matrix [9, 15] and (b) the planar projective transformation (a homography), and Both of which can be estimated from image correspondences alone.

The Fundamental matrix encodes the epipolar constraint and applies for general motion and rigid structure with uncalibrated cameras. The set of homogeneous image points $\{\underline{\mathbf{x}}_i\}, i = 1, \ldots n$, in the first image is related to the set $\{\underline{\mathbf{x}}_i'\}$ in the second image by one implicit functional relationship $\underline{\mathbf{x}}_i'^\top \underline{\mathbf{F}}\underline{\mathbf{x}}_i = 0$, where $\underline{\mathbf{x}} = (\underline{x}, \underline{y}, 1)^\top$, is a homogeneous image coordinate, and $\mathbf{F}$ is the Fundamental Matrix. In the case where all the observed points lie on a plane, or the camera rotates about its optic axis and does not translate, then all the correspondences lie on a homography, which is encoded by two independent functional relationships, $\underline{\mathbf{x}}' = \underline{\mathbf{H}}\underline{\mathbf{x}}$.

## 2.1 Likelihood of a Match Given a Relation

In this section a Bayesian framework for the robust estimation of the matches and epipolar geometry is laid out, we shall mostly focus on the fundamental matrix, noting that the techniques are also valid for homographies. First the case when the matches are known is considered, then this formalism will be extended to include the case when the matches themselves are unknown and must be estimated. In both cases it will be demonstrated that, although the problem contains a large number of latent variables, the likelihood can be determined from the parameters of the epipolar geometry only.

The observed image coordinates of a match are noise corrupted versions of the projections of some real world point, i.e.

$$\begin{aligned} x_i^j &= \underline{x}_i^j + \epsilon_{ij} \\ y_i^j &= \underline{y}_i^j + \gamma_{ij} \qquad\qquad j = 1, 2, \text{ and } i = 1 \dots n \end{aligned}$$

where $\epsilon_{ij}, \gamma_{ij}$ are sampled from the noise distribution and the implicit functional relationship defined by $\mathcal{M}$ at each $\underline{\mathbf{m}}_i = (\underline{x}_i^1, \underline{y}_i^1, \underline{x}_i^2, \underline{y}_i^2)$ is zero, i.e.

$$g_q(\underline{x}_i^1, \underline{y}_i^1, \underline{x}_i^2, \underline{y}_i^2, \underline{\boldsymbol{\alpha}}) = 0, \qquad i = 1 \dots n, \text{ and } q = 1 \dots Q.$$

The important thing to observe in this model is that the the observations, $x_i^1, x_i^2$ etc, are only indirectly related to the parameters, $\boldsymbol{\alpha}$, of the two view relation via their true locations $\underline{x}_i^1, \underline{x}_i^2$ etc. Given a particular value for $\boldsymbol{\alpha}$, let the estimates of the true (uncorrupted) matches be $\{\hat{x}_i^j, \hat{y}_i^j | \;\; g_q(\hat{x}_i^1, \hat{y}_i^1, \hat{x}_i^2, \hat{y}_i^2, \hat{\boldsymbol{\alpha}}) = 0, \quad j = 1, 2, \quad i = 1 \dots n\}$ which are parametrized by $\boldsymbol{\beta}$. Let $\boldsymbol{\beta}_i$ denote the parameters of the $i$th match $\mathbf{m}_i$. It might be thought that in the two view case it would require four parameters to describe $\boldsymbol{\beta}_i$, however, once $\mathbf{F}$ is known $\boldsymbol{\beta}_i$ can be described in three, with some mapping: $\Psi(\boldsymbol{\beta}_i) = (x_i^1, y_i^1, x_i^2, y_i^2)$. These three parameters could be interpreted as the projective 3D location of the point in space, $\mathbf{X}_i$, with $\Psi$, $\mathbf{x}_i = \Psi(\mathbf{X}_i)$, the projection mapping from 3D to 2D, in this case $\boldsymbol{\beta}_i = \mathbf{X}_i$, however this is not the only way of parametrizing $\boldsymbol{\beta}$, e.g. we could use any three parameters of a coordinate system implicitly described on the three dimensional manifold described by $\mathbf{F}$. Similarly once $\mathbf{H}$ is known, $\boldsymbol{\beta}_i$, can be described by two parameters.

Then assuming a zero mean Normal distribution for $e_i^2$, with standard deviation $\sigma$ for $\epsilon_{ij}, \gamma_{ij}$, and that all the errors are independent then

$$p(\mathcal{D}|\boldsymbol{\beta}, \mathcal{M}, \mathcal{I}) = K^n \prod_{i=1\dots n} \exp -\frac{e_i^2}{2\sigma^2}, \tag{1}$$

where $n$ is the number of correspondences, the normalization constant is $K = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^D$, $D = 4$, as the observations are in $\Re^4$, and $e_i^2 = \sum_{j=1,2}((\hat{x}_i^j - x_i^j)^2 + (\hat{y}_i^j - y_i^j)^2)$. Given a particular estimate of the parameters, $\boldsymbol{\alpha}$, of a homography or epipolar geometry the maximum likelihood of (1) can be readily obtained by the well known technique of minimizing reprojection error. A computationally efficient first order approximation to this is given in Torr *et al.* [40]. Thus the maximum likelihood solution for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be obtained by gradient descent on the space of $\boldsymbol{\alpha}$, or some similar algorithm.

Next a fully Bayesian solution is laid out for the case where matches might also be outliers. First we introduce $\boldsymbol{\gamma} = \{\gamma_1 \dots \gamma_n\}$, a set of indicator variables, $\gamma_i$, such that $\gamma_i = 1$, if the $i$th correspondence is an inlier, and $\gamma_i = 0$, if the $i$th correspondence is an outlier. Generally the observations occur within a bounded region of $\Re^D$, so the distribution for outliers is assumed to be uniform $\frac{1}{v}$, where $v$ is the volume of this space (measured in the same units as $\sigma$). Then the likelihood factors as

$$p(\mathcal{D}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{M}, \mathcal{I}) = \prod_{i=1\dots n} \left(\gamma_i K \exp\left(-\frac{e_i^2}{2\sigma^2}\right) + (1 - \gamma_i)\frac{1}{v}\right). \tag{2}$$

In order to optimally estimate $\boldsymbol{\theta}$, a MAP estimate, $\hat{\boldsymbol{\theta}}$, is made such that

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}, \mathcal{I}) = \text{argmax}_{\boldsymbol{\theta}} \frac{p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}, \mathcal{I}) p(\boldsymbol{\theta}|\mathcal{M}, \mathcal{I})}{p(\mathcal{D}|\mathcal{M}, \mathcal{I})} \tag{3}$$

with $\hat{\boldsymbol{\theta}}$ corresponding to the estimate of the true $\boldsymbol{\theta}$. The two quantities $\mathcal{D}$ and $\mathcal{I}$ are defined as follows: $\mathcal{D}$ is the data, in this case the image correspondences, $\mathcal{I}$ is the given information, upon which all the probabilities are conditioned. The first term in the numerator is the likelihood term, the second term is the prior term, the denominator is called the evidence or normalizing constant which is constant for a fixed $\mathcal{M}$ and so can be discounted for MAP estimation, distributions for which this is unknown are referred to as *unnormalized*, a problem we shall return to in Section 4. Referring to (3) the likelihood term can be written

$$p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M},\mathcal{I}) = p(\mathcal{D}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{M},\mathcal{I}) = p(\mathcal{D}|\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{M},\mathcal{I}) \tag{4}$$

reflecting the fact that the likelihood component depends only on $\boldsymbol{\beta}$ the parameters of the matches and, $\boldsymbol{\gamma}$, which determines whether or not the matches are outliers or inliers; this is just equation (2). It will be assumed that, the prior term can be written

$$P(\boldsymbol{\theta}|\mathcal{M},\mathcal{I}) \;=\; p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|\mathcal{M},\mathcal{I}) \;=\; p(\boldsymbol{\beta}|\boldsymbol{\alpha},\mathcal{M},\mathcal{I})p(\boldsymbol{\gamma}|\boldsymbol{\alpha},\mathcal{M},\mathcal{I})p(\boldsymbol{\alpha}|\mathcal{M},\mathcal{I}). \tag{5}$$

within this paper the prior is assumed sufficiently diffuse as to be constant, however a more detailed discussion of the prior term can be found in [36], but see also Section 3.2 where a prior on *disparity* is defined. In this case the MAP solution for $\boldsymbol{\Theta}$ involves maximizing (2), (see Section 3.1 and equation (13)) over $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}$. At first this seems a very high space to search over, however really the search is just over all values of $\boldsymbol{\alpha}$. This is because, as has already been seen, the optimal value of $\boldsymbol{\beta}$ can be readily obtained given any $\boldsymbol{\alpha}$ [2]. Once $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are known $\boldsymbol{\gamma}$ can be determined with each $\gamma_i$ set to 0 or 1 so as to maximize (2), this is the basis of MLESAC [43], or MAPSAC [25].

## 2.2 Likelihood Formulation When the Matches are Unknown

In this section the likelihood for $\boldsymbol{\alpha}$ is formulated given a set of features in each image but an unknown matching between them. Let the set of $n$ features in image one be $x_i^1, y_i^1, i = 1 \ldots n$ and the set of $m$ features in image two be $x_j^2, y_j^2, j = 1 \ldots m$. Again it will be shown that even though the matching is unknown, this likelihood can be approximated only knowing $\boldsymbol{\alpha}$, which is important as when optimizing we need only search the space of $\boldsymbol{\alpha}$.

The analysis then proceeds as in equations (3)-(5) except that in place of the parameters $\boldsymbol{\gamma}$ a new set of indicator variables, $\boldsymbol{\delta}$, are introduced, such that $\delta_{ij} = 1$ indicates that the $i$th feature in image one is matched to the $j$th feature in image two. The fact that each feature can match at most one other (the *uniqueness* constraint) can be written $\sum_i \delta_{ij} = 0$ or 1, and $\sum_j \delta_{ij} = 0$ or 1.

The $\boldsymbol{\beta}$ parameter is kept, $\boldsymbol{\beta}_i$, representing the estimate of the true location of feature $i$ in image 1 and 2: $(\hat{x}_i^1\,, \hat{y}_i^1\,, \hat{x}_i^2\,, \hat{y}_i^2\,)$. In this new formulation (2) can be written as

$$p\left(\mathcal{D}|\boldsymbol{\beta},\boldsymbol{\delta},\mathcal{M},\mathcal{I}\right) \propto \prod_{i=1\ldots n}\left(\sum_{j=1\ldots m}\delta_{ij}K\exp\left(-\frac{e(\boldsymbol{\beta}_i,\delta_{ij})^2}{2\sigma^2}\right)+\left(1-\sum_j\delta_{ij}\right)\frac{1}{v}\right) \tag{6}$$

where

$$e(\boldsymbol{\beta}_i,\delta_{ij})^2 = (\hat{x}_i^1 - x_i^1)^2 + (\hat{y}_i^1 - y_i^1)^2 + (\hat{x}_i^2 - x_j^2)^2 + (\hat{y}_i^2 - y_j^2)^2. \tag{7}$$

The quantity $e(\boldsymbol{\beta}_i,\delta_{ij})^2$ is simply the reprojection error for matching the $i$th to $j$th feature under the parameters $\boldsymbol{\alpha}$. If the uniqueness constraint is dropped, which in general does not much affect the resulting likelihood, then $\boldsymbol{\delta}$ can be disposed of by either maximizing or marginalizing in (6). Marginalizing or maximizing over $\boldsymbol{\beta},\boldsymbol{\gamma}$ makes little difference if we are only interested in the MAP estimate of $\boldsymbol{\alpha}$, for $\boldsymbol{\delta}$, however, marginalization is preferred as this avoids an over early commitment to a particular matching, thus the $\delta_{ij}$ are replaced by their expectations in (6).

A likelihood for matching based purely on epipolar geometry would have many local minima if only the spatial position of the features is considered. In order to reduce this photometric information is incorporated into the likelihood. Suppose that the measure of photometric similarity between the image patches around the $i$th and $j$th feature is given by $\psi(i,j)$, such that

$$p\left(\mathcal{D}|\boldsymbol{\beta},\boldsymbol{\delta},\mathcal{M},\mathcal{I}\right) \propto \prod_{i=1\ldots n}\left(\sum_{j=1\ldots m}\delta_{ij}K\exp\left(-\frac{e(\boldsymbol{\beta}_i,\delta_{ij})^2}{2\sigma^2}-\psi(i,j)\right)+\left(1-\sum_j\delta_{ij}\right)\frac{1}{v}\right). \tag{8}$$

[2] For many cases the solution gained for $\boldsymbol{\alpha}$ is the same if *marginalization* is done over $\boldsymbol{\beta},\boldsymbol{\gamma}$ instead of maximization, for more details see [36].

The correlation measure $\psi(i,j)$ is based upon normalization cross correlation and is discussed further in Section 4.

# 3 Random Sampling Guided Matching

Within this section the state of the art in feature matching is described. This computation requires initial matching of points (e.g. corners detected to sub-pixel accuracy by the Harris corner detector [13]) between two images; the aim is then to compute the relation from these image correspondences. Given a corner at position $(x,y)$ in the first image, the search for a match considers all corners within a region centred on $(x,y)$ in the second image with a threshold on maximum disparity. The strength of candidate matches is measured by sum of squared differences in intensity. At this stage, the threshold for match acceptance is deliberately conservative in order to minimise incorrect matches. Nevertheless, many mismatches will occur because the matching process is based only on proximity and similarity. These mismatches (called outliers) are sufficient to render standard least squares estimators useless. Consequently robust methods must be adopted, which can provide a good estimate of the solution even if some of the data are outliers.

There are potentially a significant number of mismatches amongst the initial matches. Since correct matches will obey the epipolar geometry, the aim is to obtain a set of "inliers" consistent with the epipolar geometry using a robust technique. In this case "outliers" are putative matches which are inconsistent with the epipolar geometry. Robust estimation by random sampling (such as MLESAC, MAPSAC, LMS or RANSAC) have proven the most successful [10, 39, 49, 42]. These algorithms are well known and MAPSAC is summarized next.

## 3.1 MAPSAC: An Algorithm for the MAP estimator

Within this section an algorithm is proposed to efficiently operationalize the finding of modes in the posterior. The RANSAC [10] algorithm has proven very successful for robust estimation, here I consider its extension to finding modes of the posterior in manifold fitting. Consideration of RANSAC shows that in effect it finds the minimum of a cost function defined as

$$C = \sum_i \rho\left(\frac{e_i^2}{\sigma^2}\right) \tag{9}$$

where $\rho()$ is

$$\rho\left(\frac{e^2}{\sigma^2}\right) = \begin{cases} 0 & \frac{e^2}{\sigma^2} < T \\ \text{positive constant} & \frac{e^2}{\sigma^2} \geq T \end{cases}. \tag{10}$$

One of the problems with RANSAC is that if the threshold, $T$, for considering inliers is set too high then the robust estimate can be very poor. In other words inliers score nothing and each outlier scores a constant penalty. Thus the higher $T$ is the more solutions there are with equal values of $C$ tending to poor estimation e.g. if $T$ were sufficiently large then all solutions would have the same cost as all the matches would be inliers. In Torr and Zisserman [42] it was shown that at no extra cost this undesirable situation can be remedied. Rather than minimizing $C$ a new cost function can be minimized which is simply the negative log of the posterior (3), which can be readily evaluated given $\boldsymbol{\alpha}$. Jointly maximizing $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ is achieved by maximizing

$$C_2 = \sum_i \rho_2\left(\frac{e_i^2}{\sigma^2}\right) \tag{11}$$

where $\rho_2()$ is

$$\rho_2\left(\frac{e^2}{\sigma^2}\right) = \begin{cases} \frac{e_i^2}{2\sigma^2} & \frac{e^2}{\sigma^2} < T \\ T & \frac{e^2}{\sigma^2} \geq T \end{cases}. \tag{12}$$

with $T$

$$T = -2\log\left(\frac{1}{vK}\right) \tag{13}$$

obtained by rearranging (2), and maximizing (2) over $\gamma_i$

1. Obtain data $\mathbf{m}_i, i = 1 \ldots n$ from some source, and determine the manifold to be fit.
2. Repeat (a)-(c) until a large number of samples have been taken or "jump out" occurs as described in [41].
   (a) Select a random sample of the minimum number of points (matches) $S_m = \{\mathbf{m}_i\}$ to estimate $\boldsymbol{\theta}$.
   (b) Estimate $\boldsymbol{\theta}$ consistent with this minimal set $S_m$.
   (c) Calculate the posterior for $\boldsymbol{\theta}$: $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}, \mathcal{I})$. This can either be approximated by maximizing $\boldsymbol{\gamma}$, or by marginalizing $\boldsymbol{\gamma}$ via EM
      i. *Either:* Approximate posterior by minimizing $C_2 = \sum_i \rho_2 \left(\frac{e^2}{2\sigma_i^2}\right)$, with $\rho_2()$ defined in (12). *(Lower computational cost).*
      ii. *Or:* use EM to marginalize $\boldsymbol{\gamma}$ and maximize the posterior $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}, \mathcal{I})$ directly. *(Higher computational cost).*
3. Select the best solution $\hat{\boldsymbol{\theta}}$ over all the samples i.e. that which maximized the posterior (or minimizes $C_2$).
4. Maximize the posterior (minimize $C_2$) using non-linear optimization as described in [41].

Table 1: *A brief summary of the stages of the Bayesian random sampling estimation algorithm* MAPSAC.

At a slight increase in computational cost and complexity of the algorithm the $\boldsymbol{\gamma}$ can be marginalized over too, which gives a slight increase in the accuracy of $\boldsymbol{\alpha}$ [42]. However the computational simple form of maximizing over $\boldsymbol{\gamma}$ is appealing and often it is as important to 3D reconstruction to identify outliers as well as get the motion correct.

It can be seen that outliers are still given a fixed penalty but now inliers are scored on how well they fit the data. The implementation of this new method yields a modest to hefty benefit to all robust estimations with absolutely no additional computational burden. *Once this is understood there is no reason to use* RANSAC *in preference to this method.* Similar schemes for robust estimation using random sampling and M-estimators were also proposed in [37] and [32]. After the random sampling the posterior is maximized using non-linear optimization as described in [41]. The generic method of using random samples to maximize the posterior (or an approximation of it) followed by gradient descent is dubbed MAPSAC *maximum a posteriori sample consensus* to distinguish it from the author's previous work MLESAC which was purely a maximum likelihood formulation. The MAPSAC algorithm is summarized in Table 1.

## 3.2 The Augmented Fundamental Matrix

In [44] it was shown that using $\mathbf{H}$ to guide matches throughout the sequence leads to fewer matches being extracted in the part of the sequence undergoing a general motion, as might be expected since the model underfits this part. However, when a loose threshold of 3 pixels was used (as opposed to a threshold of 1.25 pixels which is the two sigma window arising from interest point measurement noise) the homography is able to carry correct matches even when the planar assumption is broken. The explanation lies in the "plane plus parallax" model of image motion [17]: the estimated homography often behaves as if induced by a 'scene average' plane, or indeed is induced by a dominant scene plane; the homography map removes the effects of camera rotation and change in internal parameters, and is an exact map for points on the plane. The only residual image motion (which is *parallax* relative to this homography) arises from the scene relief relative to the plane. Often this parallax is less than the loose displacement threshold, so that all correspondences may still be obtained. Thus the homography provides

strong disambiguation for matching and the parallax effects do not exceed the loose threshold. Note that *disparity can only be computed with respect to some homography*, typically it is the one induced by camera rotation.

This suggests a new method for matching, in which one (or more) homographies *and* a fundamental matrix are estimated for the data. The homographies estimated at the coarser level are used to guide the search windows in order to detect matches for the features at the finer level. They can also be used to guide the cross correlation matching at the finer level in that the patches that are correlated can be corrected by transformation under the homography. This representation is referred to as the *augmented* fundamental $\mathbf{F}^+$. For the examples presented in this paper, one homography is sufficient to guide matching. This leads to a 10 parameter estimation problem for $\mathbf{F}^+$ (8 for the homography and 2 for the epipole, alternatively: 7 for the fundamental matrix and 3 for the plane of the homography). Given an epipole im image 2 $\mathbf{e}'$ and a homography from image 1 to 2, $\mathbf{H}$, then

$$\mathbf{F} = \begin{bmatrix} 0 & -e'_3 & e'_2 \\ e'_3 & 0 & -e'_1 \\ -e'_2 & e'_1 & 0 \end{bmatrix} \mathbf{H} \ . \tag{14}$$

Future work will consider the use of several planes to augment the fundamental matrix, but for many image sequences one seems to be sufficient to get good matches.

To account for disparity in (5) the term, $p(\boldsymbol{\beta}|\boldsymbol{\alpha}, \mathcal{M}, \mathcal{I})$, is adjusted to take into account the fact that lower disparities are favoured, following the "slow and smooth" prior of Weiss [46], the disparity, $\Delta_i$, of the $i$th feature, is defined as the Euclidean distance between $\hat{\mathbf{x}}_i{}^2$ and $\mathbf{H}\hat{\mathbf{x}}_i{}^1$, which is the distance moved along the epipolar line between the point predicted by $\mathbf{H}$ and that predicted by $\mathbf{F}$, which is assigned the Gaussian distribution:

$$p(\boldsymbol{\beta}_i|\boldsymbol{\alpha}, \mathcal{M}, \mathcal{I}) = \left(\frac{1}{\sqrt{2\pi}\sigma_\Delta}\right) \exp\left(-\frac{\Delta_i^2}{2{\sigma_\Delta}^2}\right), \tag{15}$$

assuming the $\boldsymbol{\beta}$ are independent then, $p(\boldsymbol{\beta}|\boldsymbol{\alpha}, \mathcal{M}, \mathcal{I}) = \prod p(\boldsymbol{\beta}_i|\boldsymbol{\alpha}, \mathcal{M}, \mathcal{I})$, this assumption is not entirely realistic, as ideally some sort of smoothness should be enforced on nearby disparities, however for features which are scattered across the image this condition can be relaxed.

### 3.3 Problems with Conventional Matching

There are two types of failure mode for the class of matching algorithms described in the last section. The first is the wide baseline case, see top row Figure 1, which shows two images taken at the same time instant [3] where the disparity is 160 pixels. In the conventional algorithm, described above, a search window must be set for putative matches. If this search window is too large (which it must be in this case to guarantee that the correct match lies within it), then there is a combinatorial explosion of putative matches. This leads to a catastrophic failure of correlation matching as there are too many potential false matches for each corner. The second failure mode is caused when the image is rotated (see bottom row Figure 1). In this case, standard correlation matching cannot be expected to succeed, because the correlation score is not rotationally invariant. Using a rotationally invariant correlation score does not correct this problem; instead it reduces the discriminating power of the score, increasing the number of mismatches even when the second image is not rotated. The answer to both these problems, presented here, is to adopt a coarse to fine strategy. The coarse to fine strategy has been used successfully for small baseline homography matching [4], but neglected for feature matching.

## 4 Coarse to Fine

In order to reduce the search region and amount of spurious matches a coarse to fine strategy is adopted. In the coarse to fine strategy, an image pyramid is formed by subsampling the image repeatedly by a factor of 2. The first stage is to generate the features at all levels. Then, at the coarsest scale, cross correlation scores are generated between all features, with each patch undergoing 16 evenly space rotations, the strongest matches over these rotations are stored and used as the basis for estimating $\boldsymbol{\alpha}$ using random sampling. The success of RANSAC-style methods proves that at least some of the generated samples lie in areas of high posterior probability. It would

[3] Kindly provided by Dayton Taylor.

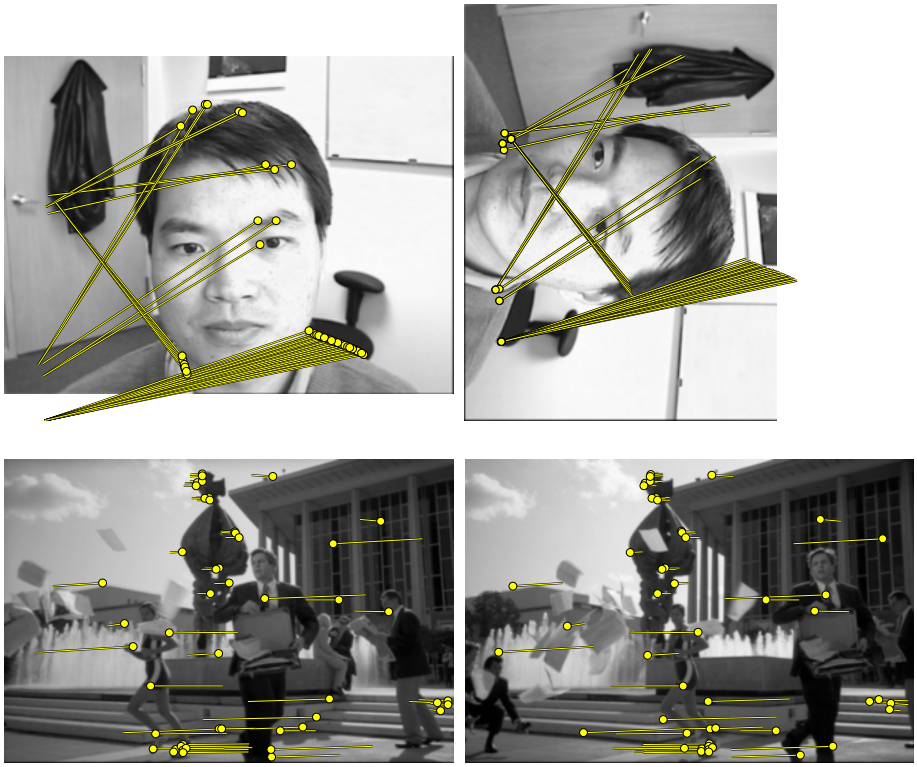

Figure 1: **Top Row Wide Baseline Failure of** MAPSAC/LMS/RANSAC**:** *50 matches from the first and last images of the Samsung sequence. The images are were imaged at the same time instance and are two of 50 taken from a 50 camera stereo rig. The features are shown in each image (circles) together with the line joining them to their correspondence in the other image, and are matched with a fundamental matrix. Although several of the features with small disparities have been correctly matched, features with large disparities are incorrectly matched. This is because, as the disparity increases, so does the number of potential mismatches.* **Bottom Row Catastrophic Failure of** MAPSAC/LMS/RANSAC **Due To Rotation:** *the second image in the Zhang sequence has been rotated by 90 degrees, in addition there is a slight change of pose of the head. The image correlation used is not invariant to rotation, so there are too many mismatches for* MAPSAC *to converge. Rotation-invariant correlation is not a solution to this problem, because it is less discriminating and thus results in too many mismatches even when the second image is not rotated.*

1. At each scale: Detect features.
2. Putative matching of corners over the coarsest two images using proximity and cross correlation under a variety of rotations.
3. At the **coarsest level**.
   (a) Generate initial particles at coarsest level using random sampling, as described in Section 3.1.
   (b) Apply MCMC to the particles to get a set which might be drawn from $p(\boldsymbol{\alpha}^0|\mathbf{D}_0)$, as described in Section 4.3.
   (c) Create an importance sampling function, $g_0(\boldsymbol{\alpha})$, from the particles as described in Section 4.1.
4. **levels l = 1 to l = finest level**.
   (a) Generate particles using the important sampling function of the level above, $g_{l-1}(\boldsymbol{\alpha})$, using SIR as described in Section 4.2.
   (b) Apply MCMC to the particles to get a set which might be drawn from $p(\boldsymbol{\alpha}^l|\mathbf{D}_l)$, as described in Section 4.3.
   (c) Create an importance sampling function, $g_l(\boldsymbol{\alpha})$, from the particles as described in Section 4.1.
5. The particle with the maximum posterior probability is taken as the MAP estimate. This can then be used as a starting point for a gradient descent algorithm.

Table 2: *Feature Matching Algorithm using* IMPSAC.

be nice to be able to harness the RANSAC mechanism in order to generate a good importance sampling function with which to propagate information from coarse to fine levels. Due to the ambiguity and inaccuracy in matching, at the coarsest level, in general it cannot be expected that a unique solution can be recovered using only the coarsest level. What is desired is a way to transmit the information on $\boldsymbol{\alpha}$ to the next level of fineness of the pyramid. One way to propagate information from one level to the next is to simply propagate down the mode of this distribution. However, at the coarsest levels this distribution is not expected to have a strong peak and often propagation of the mode does not convey sufficient information. Too soon a commitment to a single hypothesis may cause the algorithm to converge to the wrong solution. Rather, it is desirable to pass as much of the distribution as possible from one level to the next.

Unfortunately the parametric form of the distribution is not known. The problem is overcome by representing the distribution by a set of particles $\{\boldsymbol{\alpha}_1 \ldots \boldsymbol{\alpha}_m\}$ with weights $\{w_1 \ldots w_m\}$ in proportion to their likelihood. This sort of representation has been used with a good deal of success in the tracking literature [18]. Ideally the set of particles would be drawn from the posterior distribution. Note there is a duality between a sample and the distribution from which it is generated: clearly, the distribution can generate the sample; conversely, given the sample we can, at least approximately, re-create the distribution [5].

Define the distribution of the parameters $\boldsymbol{\alpha}$ of the relation $\mathcal{R}$ given the data, $\mathbf{D}_l$, at level, $l$, is $p(\boldsymbol{\alpha}^l|\mathbf{D}_l)$. The algorithm is summarized in Fig. 2. The aim of the algorithm is to generate the posterior distribution of $\boldsymbol{\alpha}^l$ at each level, which allows for a full exploration and description of the solution space, if one answer is required at the end the particles that maximizes the posterior can be selected.

## 4.1 Particles to Parametric Distribution

Within this paper the only reason we wish to convert particles to parametric distributions is so that we can create an importance sampling function, which is described in the next section. The approach used is to model the distribution of $\boldsymbol{\alpha}$ as a mixture of Gaussians and use EM to fit them [8, 21, 34]. Typically we have found 4 or 5 Gaussians to be sufficient to encode most of the ambiguity in many cases.

## 4.2 Sampling-Importance-Resampling

Initially importance sampling was developed to evaluate integrals-and hence calculate posterior normalizing constraints and moments-however we can also exploit the idea to produce simulated samples from any function that represents a distribution. Furthermore the normalization constant, $p(\mathcal{D}|\mathcal{M},\mathcal{I})$ in (3), need not be known for the distribution.

Sampling-importance-resampling [11], or SIR, is a method for drawing samples from complicated high dimensional posterior distributions for which the normalization factor is unknown. In this case we are going to use it to generate an initial set of particles that represent the posterior distribution of $\boldsymbol{\alpha}$ at a given resolution, *using* the information from the higher (coarser) level. By this process information about the distribution of $\boldsymbol{\alpha}$ is transmitted from coarser to finer levels in a statistically consistent way. Suppose it is of interest to draw samples from such a distribution $q(\boldsymbol{\alpha})$, and there exists a normalized positive density function, the importance sampling function, $g(\boldsymbol{\alpha})$ from which it is possible to draw samples (in our case we have chosen $g(\boldsymbol{\alpha})$ to have the form of a mixture of Gaussians particularly because this is easy to sample from e.g. see [29]). The algorithm proceeds as follows:

1. Generate a set of $M$ draws $S^t = \{\boldsymbol{\alpha}_1, \ldots \boldsymbol{\alpha}_M\}$ from $g(\boldsymbol{\alpha})$.
2. Evaluate $q(\boldsymbol{\alpha})$ for each element of $S^t$.
3. Calculate importance weights $w_i = \frac{q(\boldsymbol{\alpha}_i)}{g(\boldsymbol{\alpha}_i)}$ for each element of $S^t$.
4. Sample a new draw from $S^{t+1}$ from $S^t$ where the probability of taking a new $\boldsymbol{\alpha}_i$ is proportional to its weight $w_i$.

Iterating this procedure from step 2 is called *sampling importance resampling* (SIR). This process, in the limit of $M$, produces a fair sample from the distribution $q(\boldsymbol{\alpha})$ [11].

In our example $g(\boldsymbol{\alpha})$ is the approximation to the posterior, $p(\boldsymbol{\alpha}^{l-1}|\mathbf{D}_{l-1})$, at the coarser level obtained by fitting a mixture of Gaussians (it is necessary to fit the Gaussians in order to be able to generate samples from $g(\boldsymbol{\alpha})$), and $q(\boldsymbol{\alpha})$ is the posterior at the level below, $p(\boldsymbol{\alpha}^{l}|\mathbf{D}_{l})$, for which we want to generate a representative set of particles.

The rate of convergence is determined by the suitability of the importance function $g(\boldsymbol{\alpha})$. The worst possible scenario occurs when the importance weights, $w_i = \frac{q(\boldsymbol{\alpha}_i)}{g(\boldsymbol{\alpha}_i)}$, are small with high probability and large with low probability, i.e. we cant $g(\boldsymbol{\alpha})$ to resemble $q(\boldsymbol{\alpha})$ for the particles. There is no general purpose method for choosing a good importance sampling function, however in this paper the importance sampling function $g(\boldsymbol{\alpha})$ is either chosen via RANSAC and MCMC at the coarsest level, level $l = 0$, or as an approximation to the posterior at levels below, in either case this importance sampling function, $g(\boldsymbol{\alpha})$, should have similar shape to the, $q(\boldsymbol{\alpha})$, which is the posterior of $\boldsymbol{\alpha}$ at the level below, and hence good convergence properties.

## 4.3 An MCMC Step

Within this section the algorithm will first be described in general terms and then for our specific example. Markov Chain Monte Carlo (MCMC) is an alternate method for simulating the drawing of samples $\boldsymbol{\alpha}$ [26]. Markov chain simulation, like importance sampling, is a general method based on drawing values of $\boldsymbol{\alpha}$ from an approximate distribution and then correcting those draws to approximate better the target distribution. The key difference is that the samples are drawn sequentially, with the distribution of the sampled draws depending on the last value drawn.

There are two uses for MCMC in our algorithm. At initialization the purpose of the MCMC step is to alter the particles generated by RANSAC to make them better approximate the posterior distribution of $p(\boldsymbol{\alpha}^0|\mathbf{D}_0)$. For each subsequent level an additional aim is to also ensure that all the probability does not become concentrated into

one particle, thus if the same particle is repeatedly sampled in stage 4 of SIR, the MCMC will randomly perturb each particle whilst still making the particles representative of the posterior.

There are a number of ways to generate a Markov process which converges to a particular distribution. These include the Metropolis-Hastings [23, 16] class of algorithms, in which transitions or jumps are drawn from a user-defined jumping distribution $J_t(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1})$ and the updated model is accepted or rejected based on a scoring function $f(\boldsymbol{\alpha})$. The Metropolis algorithm, creates a sequence of random points $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \ldots$ whose distribution converges to the target distribution, in our case we shall apply the MCMC algorithm to *every* particle generated.

1. Draw an initial point $\boldsymbol{\alpha}_0$ from a starting distribution $p_0(\boldsymbol{\alpha})$.
2. For $t = 1..T$
   (a) Draw candidate point $\boldsymbol{\alpha}_*$ from the jumping distribution $J_t(\boldsymbol{\alpha}_*|\boldsymbol{\alpha}_{t-1})$.
   (b) Calculate the ratio
   $$r = \frac{f(\boldsymbol{\alpha}_*)}{f(\boldsymbol{\alpha}_{t-1})} \tag{16}$$
   (c) Set $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_*$ with probability min($r$,1), otherwise set $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1}$
   (d) Retain history of all points $\boldsymbol{\alpha}_t$ visited after a "burn-in" period during which the process converges.

In this case the jumping distribution is chosen to be a Gaussian.

A common pitfall of Markov chain simulation when applied to a multi modal distribution (i.e. one with multiple peaks) is that it may converge to a local maximum and thus fail to fully explore the distribution. One way to guard against this is to seed multiple Markov chains at random points in parameter space. This is akin to Monte Carlo sampling of the parameter space, and algorithms which operate on this principle are known as Markov Chain Monte Carlo (MCMC) algorithms [12, 26]. By running MCMC on the whole set of particles generated by importance sampling or RANSAC we are effectively starting from many different starting points, and thus hopefully exploring the whole space.

How many iterations should the MCMC algorithm be run for? The answer to this question depends on the dimension and multi modality of the distribution and can be tens of thousands of iterations, however by using RANSAC and SIR to seed the starting points for MCMC we have found that sometimes as little as one or two iterations can be made (although one would expect better results for many iterations), this is because the particles are (hopefully) already well positioned in areas of high probability by RANSAC and SIR.

# 5 Results

The final stage of the algorithm in Table 2 is to select the most likely particle at the finest level as the most likely hypothesis. This is the particle $\boldsymbol{\alpha}_{imax}$ which maximizes $p(\boldsymbol{\alpha}_i|\mathbf{D})$. The $i^{th}$ feature in the first image is matched to the feature $j$ in the second image which maximizes $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, |\mathbf{D})$. The top row of Figure 2 shows the successful matching of two images with up to 160 pixels disparity, demonstrating the capacity of IMPSAC for wide baseline matching. The bottom row Figure 2 shows how IMPSAC is robust to large rotations of the image. In figure 3, mismatches of MAPSAC are corrected by rematching with the augmented likelihood, doubling the number of matched features.

When does the algorithm break? Typically when there are too few particles to explore the parameter space, in the examples considered 2000 particles were used; if the number drops to 500 then 2 times in 3 the algorithm fails. It is difficult to make a general rule for how many particles are required, as this depends to a large extent on the multi modality of the posterior distribution, but in most cases the 2000 is the minimum.

# 6 Future Work

Due to space constraints, model selection is not the topic of this paper. However it will be briefly illustrated how importance sampling can be used to evaluate the marginal likelihoods required for model comparison. Given a

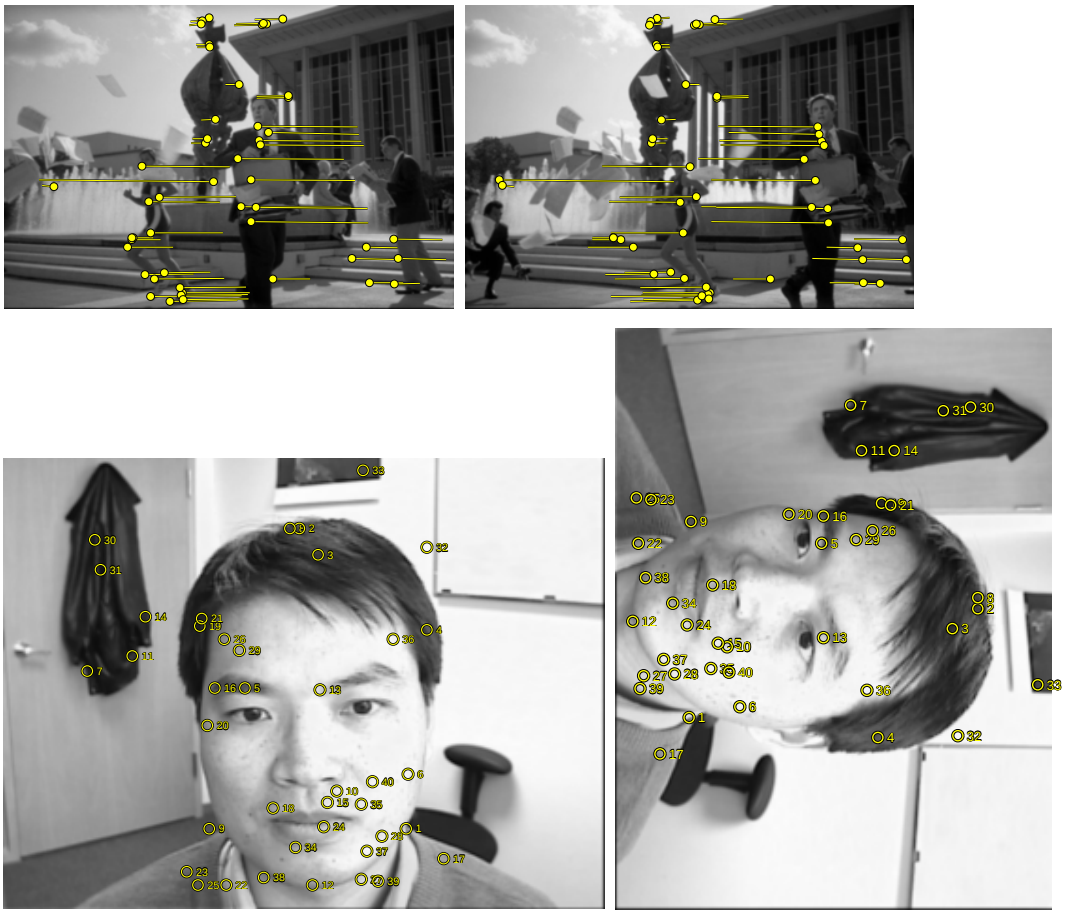

Figure 2: **Top Row Wide Baseline Success of** IMPSAC**:** *the first and last images from the Samsung sequence, captured at the same time but from different positions. The disparity between the images is up to 160 pixels, yet only 3 or 4 of the 50 example matches shown are mismatched.* **Bottom Row Rotation Success of** IMPSAC**:** *Despite the combination of a rotation of 90 degrees and the change in pose of the face, the features are correctly matched. Although just* 40 *features are shown for clarity, over* 1000 *were matched.*

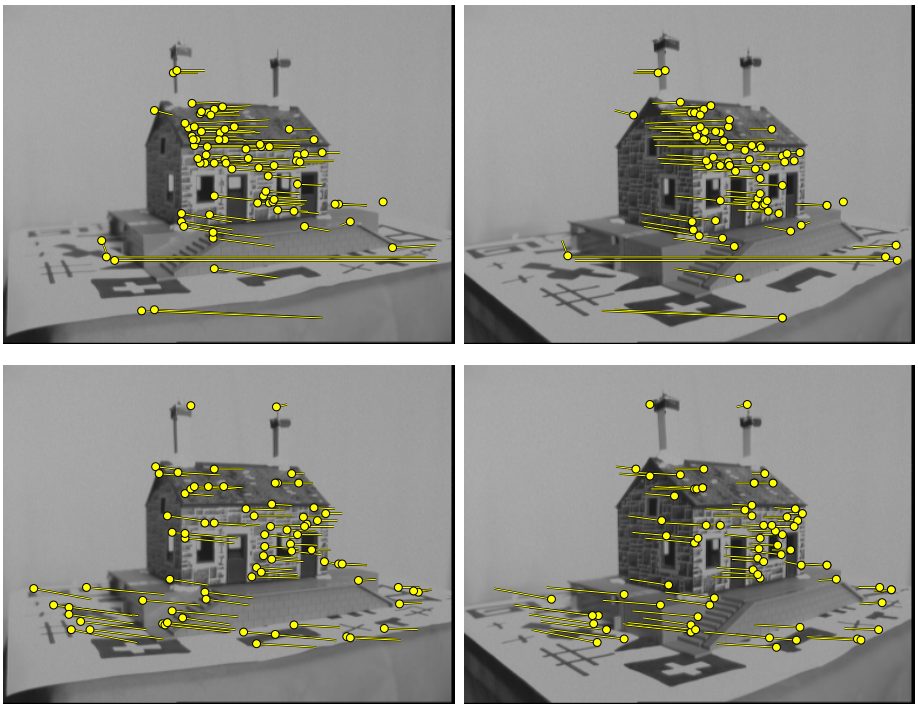

Figure 3: MAPSAC **Mismatches Corrected by Augmented Likelihood:** *(Above)* MAPSAC *matches with fundamental matrix include numerous mismatches. (Below) From the same* MAPSAC *hypothesis, rematching with augmented likelihood increases number of matches from 509 to 1274, also reducing mismatches.*

set of $k$ models $\mathbf{M}_1 \ldots \mathbf{M}_k$ that can explain the data $\mathbf{D}$ (here the models are fundamental matrix, homography, augmented fundamental matrix etc.) then Bayes rule leads us to

$$p(\mathbf{M}_i|\mathbf{D}\mathbf{I}) = \frac{p(\mathbf{D}|\mathbf{M}_i\mathbf{I})p(\mathbf{M}_i|\mathbf{I})}{p(\mathbf{D}|\mathbf{I})}\,, \tag{17}$$

where $\mathbf{I}$ is the prior information assumed about the world. Note $p(\mathbf{D}|\mathbf{I})$ is the same for all models. Assuming that all the models are equally likely *a priori* i.e. $\mathbf{M}_i = \frac{1}{k}$, the key posterior likelihood of each model is the evaluation of $p(\mathbf{D}|\mathbf{M}_j\mathbf{I})$, which is called the evidence. This is the integral of the likelihood over all possible values of the model's parameters:

$$p(\mathbf{D}|\mathbf{M}_j\mathbf{I}) = \int p(\mathbf{D}|\mathbf{M}_j\boldsymbol{\alpha}\mathbf{I})p(\boldsymbol{\alpha}|\mathbf{M}_j\mathbf{I})\partial\boldsymbol{\alpha} \tag{18}$$

where $\boldsymbol{\alpha}$ are the $j$th model's parameters, and $p(\boldsymbol{\alpha}|\mathbf{M}_j\mathbf{I})$ is the prior distribution of parameters of the model. One method for numerically evaluating this integral would be to uniformly sample the parameter space and sum the posteriors of the samples. Unfortunately the high dimensionality of the parameter space precludes this. One could draw samples from the prior and sum the posterior of these samples, but typically the prior is too diffuse to yield samples around the peak of the distribution. Importance sampling furnishes a Monte Carlo method for performing this integration [11], the advantage of which is that samples can be taken more densely around the expected peak of the posterior and less densely in areas of little interest. If the importance sampling function is $g(\boldsymbol{\alpha})$ ($g(\boldsymbol{\alpha})$ is a normalized density), then given a set of $M$ particles drawn from $g(\boldsymbol{\alpha})$

$$p(\mathbf{D}|\mathbf{M}_j\mathbf{I}) \rightarrow \sum_{i=1}^{i=M} \frac{p(\mathbf{D}|\mathbf{M}_j\boldsymbol{\alpha}\mathbf{I})p(\boldsymbol{\alpha}|\mathbf{M}_j\mathbf{I})}{g(\boldsymbol{\alpha})} \text{ as } M \rightarrow \infty \tag{19}$$

Evaluation of this leads to the selection of an augmented fundamental matrix model for the Samsung sequence shown in Figure 2, a homography model for the Zhang sequence shown in Figure 2, and an augmented fundamental matrix for Figure 3.

# 7 Conclusion

Within this paper how to use the posterior distribution of a two view relation at a coarse level to generate the posterior at a fine level is demonstrated. This has been achieved through the synthesis of powerful statistical techniques. The concept of using a random sampling estimator to seed MCMC and to generate the importance sampling function, IMPSAC, is a general mechanism that can be used in a wide variety of statistical problems beyond this. It provides a solution to the general problem of how to create importance sampling functions for outlier corrupted data. The coarse to fine strategy helps overcome the wide baseline problem. The resultant is a general purpose and powerful image matching algorithm that can be used for 3D reconstruction or compression.

# References


[1] N. Ayache. *Artificial vision for mobile robots*. MIT Press, Cambridge, 1991.

[2] P. Beardsley, A. Zisserman, and D. W. Murray. Navigation using affine structure and motion. LNCS 800/801, pages 85–96. Signal Processing, 1994.

[3] Beardsley P., Torr P., and Zisserman A. 3D model acquisition from extended image sequences. LNCS 1064/1065, pages 683–695. Signal Processing, 1996.

[4] J. R. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. 2nd European Conference on Computer Vision, LNCS 588, Santa Margherita Ligure*, pages 237–252, 1992.

[5] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory*. Wiley, New York, 1994.

[6] T. Cham and R. Cipolla. A statistical framework for long range matching in uncalibrated image mosaicing. In *Conference on Computer Vision and Pattern Recognition*, pages 442–447, 1998.

[7] F. Dellaert, S.M. Seitz, C.E. Thorpe, and S. Thrun. Structure from motion without correspondence. pages II:557–564, 2000.

[8] N.M. Dempster, A.P. Laird and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39:185–197, 1977.

[9] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision, LNCS 588, Santa Margherita Ligure*, pages 563–578. Springer–Verlag, 1992.

[10] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, vol. 24:381–95, 1981.

[11] A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis*. Chapman and Hall, 1995.

[12] W. Gilks, S. Richardson, and D. Spiegelhalter (editors). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London., 1996.

[13] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Alvey Conf.*, pages 189–192, 1987.

[14] C. J. Harris. Determination of ego-motion from matched points. pages 189–192, 1987.

[15] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd European Conference on Computer Vision, LNCS 588, Santa Margherita Ligure*, pages 579–587. Springer-Verlag, 1992.

[16] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[17] M. Irani and P. Anandan. Parallax geometry of pairs of points for 3D scene analysis. In B. Buxton and R. Cipolla, editors, *Proc. 4th European Conference on Computer Vision, LNCS 1064, Cambridge*, pages 17–30. Springer, 1996.

[18] M. Isard and A. Blake. Condensation — conditional density propagation for visual tracking. *International Journal of Computer Vision*, 28(1):5–28, 1998.

[19] D.G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, 1999.

[20] J. Matas, O. Chum, M. Urban, and T. Pajdl. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, pages 384–393, 2002.

[21] G.I. McLachlan and K. Basford. *Mixture models: inference and applications to clustering*. Marcel Dekker. New York, 1988.

[22] P. F. McLauchlan and D. W. Murray. A unifying framework for structure from motion recovery from image sequences. In *International Conf. on Computer Vision*, pages 314–320, 1995.

[23] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44:335–341, 1949.

[24] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. volume 1, pages 128–142, 2002.

[25] D. Myatt, J. M. Bishop, R. Craddock, S. Nasuto, and P.H.S. Torr. NAPSAC: High noise, high dimensional robust estimation - it's in the bag. In *BMVC*, pages 458–467, 2002.

[26] R. M. Neal. Probabilistic inference using monte carlo markov chains. Technical Report CRG-TR-93-1, University of Toronto, 1993.

[27] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. 6th Int'l Conf. on Computer Vision, Bombay*, pages 754–760, January 1998.

[28] L.H. Quam. Hierarchical warp stereo. In Lee S. Baumann, editor, *Image Understanding Workshop (New Orleans, LA, October 3-4, 1984)*, pages 149–155. Defense Advanced Research Projects Agency, Science Applications International Corp., 1984.

[29] B. D. Ripley. *Stochastic Simulation*. New York, 1987.

[30] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *Proc. 8th International Conference on Computer Vision, Vancouver, Canada*, 2001.

[31] C. Schmid and A. Zisserman. Automatic line matching across views. In *Conference on Computer Vision and Pattern Recognition*, pages 666–671, 1997.

[32] C. V. Stewart. Bias in robust estimation caused by discontinuities and multiple structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.PAMI-19,no.8:818–833, 1997.

[33] Hai Tao, Harpreet S. Sawhney, and Rakesh Kumar. A global matching framework for stereo computation. In *Int. Conf. Computer Vision*, pages 532–539. IEEE Computer Society, 2001.

[34] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, Inc., 1985.

[35] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorisation approach. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[36] P. H. S. Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting, msr-tr-2002-29. Technical report, Microsoft Research, 2002. to appear IJCV.

[37] P. H. S. Torr, P. A. Beardsley, and D. W. Murray. Robust vision. In J. Illingworth, editor, *Proc. 5th British Machine Vision Conference, York*, pages 145–155. BMVA Press, 1994.

[38] P. H. S. Torr and C. Davidson. Impsac: A synthesis of importance sampling and random sample consensus to effect multi-scale image matching for small and wide baselines. In *ECCV2000*, volume 1, pages 819–833, 2000.

[39] P. H. S. Torr and D. W. Murray. Outlier detection and motion segmentation. In P. S. Schenker, editor, *Sensor Fusion VI*, pages 432–443. SPIE volume 2059, 1993. Boston.

[40] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *Int Journal of Computer Vision*, 24(3):271–300, 1997.

[41] P. H. S. Torr and A Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–607, 1997.

[42] P. H. S. Torr and A. Zisserman. Robust computation and parametrization of multiple view relations. In U Desai, editor, *ICCV6*, pages 727–732. Narosa Publishing House, 1998.

[43] P. H. S. Torr and A Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *CVIU*, 78(1):138–156, 2000.

[44] P.H.S. Torr, A. Fitzgibbon, and A. Zisserman. The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. *IJCV*, 32(1):27–45, 1999. Marr Prize Paper ICCV 1999.

[45] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *British Machine Vision Conference (BMVC2000)*, pages 412–422, 2000.

[46] Y. Weiss and E.H. Adelson. Slow and smooth: a bayesian theory for the combination of local motion signals in human vision. Technical Report MIT AI Memo 1624 (CBCL Paper 158), MIT, 1998.

[47] Eic Weisstein. Maths encyclopedia. http://mathworld.wolfram.com/, 2002.

[48] C. Zeller. *Projective, Affine and Euclidean Calibration in Compute Vision and the Application of Three Dimensional Perception*. PhD thesis, RobotVis Group, INRIA Sophia-Antipolis, 1996.

[49] Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *AI Journal*, vol.78:87–119, 1994.

[50] Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis*. Springer-Verlag, 1992.