

Guided Sampling via Weak Motion Models and Outlier Sample Generation for Epipolar Geometry Estimation

Liran Goshen · Ilan Shimshoni

Received: 12 December 2005 / Accepted: 22 January 2008 / Published online: 14 February 2008
© Springer Science+Business Media, LLC 2008

Abstract The problem of automatic robust estimation of the epipolar geometry in cases where the correspondences are contaminated with a high percentage of outliers is addressed. This situation often occurs when the images have undergone a significant deformation, either due to large rotation or wide baseline of the cameras. An accelerated algorithm for the identification of the false matches between the views is presented. The algorithm generates a set of weak motion models (WMMs). Each WMM roughly approximates the motion of correspondences from one image to the other. The algorithm represents the distribution of the median of the geometric distances of a correspondence to the WMMs as a mixture model of outlier correspondences and inlier correspondences. The algorithm generates a sample of outlier correspondences from the data. This sample is used to estimate the outlier rate and to estimate the outlier pdf. Using these two pdfs the probability that each correspondence is an inlier is estimated. These probabilities enable guided sampling. In the RANSAC process this guided sampling accelerates the search process. The resulting algorithm when tested on real images achieves a speedup of between one or two orders of magnitude.

Keywords Epipolar geometry estimation · Robust methods · Weak motion models

1 Introduction

Recovery of epipolar geometry is a fundamental problem in computer vision. It is an important building block in many applications such as ego-motion estimation, structure from motion, and visual servoing. Corresponding point pairs $p \leftrightarrow p'$ from two images of a scene, satisfy

$$p'^T F p = 0, \quad (1)$$

where F is the fundamental matrix. The two points can be concatenated yielding a point in \mathbb{R}^4 . The points satisfying (1) lie on a quadratic manifold in \mathbb{R}^4 .

The RANdom Sample Consensus algorithm (RANSAC) (Fischler and Bolles 1981) has been widely used in computer vision for solving a variety of robust estimation problems. One of its most important applications is for recovering the epipolar geometry from a set of putative matches of which a certain fraction is incorrect (outliers). The RANSAC algorithm is a simple but powerful tool. Repeatedly, random subsets are selected from the input data and the model parameters fitting the subset are calculated. The size of the random sample is the smallest sufficient for determining the model parameters. At each iteration the quality of the model is evaluated on the full data set. At the end, the model with the largest consensus set is returned. Two issues immediately arise:

1. How to check the quality of the model? Different cost functions may be used, the standard being the number of

This work was supported partly by grant 01-99-08430 of the Israeli Space Agency through the Ministry of Science Culture and Sports of Israel.

L. Goshen
Faculty of Industrial Engineering & Management,
Technion—Israel Institute of Technology, Haifa 32000, Israel

I. Shimshoni (✉)
Department of Management Information Systems, University of
Haifa, Haifa 31905, Israel
e-mail: ishimshoni@mis.haifa.ac.il

Table 1 The number of samples I required to ensure, with probability $p = 0.99$, that at least one sample has no outliers for a given sample size s and proportion of outliers ϵ

$s \backslash \epsilon$	25%	50%	60%	70%	80%	85%
3	8	34	70	168	573	1,362
7	33	588	2,808	21,055	3.5E05	2.6E06

data points consistent with the model. In Torr and Zisserman (2000) a method of maximum likelihood estimation by sampling consensus (MLESAC) is described. MLESAC evaluates the likelihood of the hypothesis, representing the error distribution as a mixture model in which the inlier error is Gaussian and the outlier error is uniform. In Chen and Meer (2003), Rozenfeld and Shimshoni (2005), Subbarao and Meer (2007) the need for user supplied threshold is eliminated by reformulating another robust method, the M-estimator, as a projection pursuit optimization problem. The projection based pbM-estimator automatically derives the threshold from univariate kernel density estimates. Similar methods were suggested (Wang and Suter 2004a, 2004b) which can also estimate models without the user having to give them a threshold. Other methods for automatically separating the inliers from the outliers were suggested in Stewart (1995), Yu et al. (1994), Lee et al. (1998) and Moisan and Stival (2004) which is a method which was designed to deal with the estimation of the epipolar geometry.

2. When to stop the algorithm? The number of iterations is chosen sufficiently high (Fischler and Bolles 1981; Torr 1995) to ensure with probability p that at least one of the random samples of s points is free from outliers. Usually p is chosen to be 0.99. Suppose ϵ is the probability that any selected feature is an outlier. Then at least I selections are required, where $(1 - (1 - \epsilon)^s)^I = 1 - p$, thus

$$I = \log(1 - p) / \log(1 - (1 - \epsilon)^s). \quad (2)$$

Table 1 gives the values of I for different values of s and ϵ .

Although ϵ is not generally known in advance, a lower bound can be estimated at runtime from the largest consistent set observed so far. It is widely appreciated that this stopping criterion is often wildly optimistic (Tordoff and Murray 2002; Chum et al. 2003; Torr and Davidson 2003) because with noisy data it is not enough to have a sample composed only of inliers, they must be inliers that lie on different regions of the manifold in order to give a good estimate of the fundamental matrix. This significantly reduces the number of sample sets that will accurately give a good estimate of the manifold.

In the case of narrow baseline images, the approximate position of the corresponding point is known and putative correspondences are only selected in that region. Therefore,

even if a wrong match is chosen an approximate fundamental matrix might be found. In the case of wide baseline stereo however, there is usually no prior knowledge where the corresponding point will lie in the second image and therefore putative matches are found all over the image. As a result, outliers usually yield gross errors which generate totally incorrect fundamental matrices. Our work focuses on this case for which the number of required iterations of the algorithm is well approximated by (2).

Several approaches have been suggested to speed-up RANSAC. ROR (Adam et al. 2001) can speed-up the sampling process by reducing the contamination level ϵ of the data points when the camera's internal parameters are known. ROR exploits the possibility of rotating one of the images to achieve some common behavior of the inliers. ROR may be run as a postprocessing step on output from any point matching algorithm. In R-RANSAC (Matas and Chum 2004) the computational savings are achieved by typically evaluating only a fraction of the data points for each of the hypothesized models. Hypothesized models which pass the initial test are evaluated with all of the data points. LO-RANSAC (Chum et al. 2003) exploits the fact that the model hypothesis from an uncontaminated minimal sample is often sufficiently near the optimal solution and a local optimization step is applied to selected models. The number of samples which the LO-RANSAC performs achieves a good agreement with the theoretical predictions of (2).

As can be seen in Table 1, the number of needed iterations increases when the percentage of outliers is over 50% and it dramatically increases after 70%. For ϵ values close to 1 the expression in (2) can be approximated by $I \approx \log(1 - p) / \log(1 - \epsilon)^s$. This is a polynomial in the variable $(1 - \epsilon)^{-1}$ of degree s . Thus, for large values of $(1 - \epsilon)^{-1}$ and for $s = 7$ the number of iterations is very large. Note that the iteration numbers in the table are the theoretic ones and in practice they can be much higher.

In this paper we propose novel improvements to the robust estimation of epipolar geometry. The main goal is to assign probabilities to the putative correspondences and to use them in the RANSAC step. We use *weak motion models* (WMMs) to estimate these probabilities which are more informative than the correlation scores used for this purpose in Tordoff and Murray (2002). In Triggs (2001) a probabilistic motion model was presented for finding additional matches after an estimate for the fundamental matrix was found, limiting the motion to a segment of the epipolar line based on the motion of other corresponding pairs. In our case the WMMs are used to estimate the probabilities initially in order to speed up the fundamental matrix estimation process.

The algorithm generates a set of WMMs which are similar in spirit to weak classifiers in supervised learning cases. In the well known Adaboost (Freund and Schapire 1997)

algorithm a high quality classifier is generated from a combination of low quality (weak) classifiers. In our case each WMM roughly approximates the motion of a fraction of the correspondences from one image to the other. The algorithm represents the distribution of the median of the geometric distances of a correspondence to the set of WMMs as a mixture model of outlier correspondences and inlier correspondences. Typically the inlier correspondences are closer to the WMMs than the outlier correspondences. Therefore, a smaller distance will yield a higher probability that a match is an inlier.

The algorithm then generates a sample of outlier correspondences. This sample is used to estimate the outlier rate and to estimate the outlier pdf. When generating the sample of outliers we take into consideration the corner distribution on the images, the similarity between the corners and the corner matching technique. Using these two distributions, the probability that each correspondence is an inlier is estimated. These probabilities enable guided sampling, i.e. in each iteration a random subset of seven correspondences is chosen where each correspondence in this subset is chosen according to its probability.

The two main contributions of this paper are: (i) the use of WMMs to probe the correspondence set and (ii) the generation and the use of the sample of outlier correspondences. Our approach makes it possible to break the iteration number theoretical boundary given by (2). The acceleration is especially significant for outlier percentages above 70%. In such outlier percentages our algorithm achieves a speedup of between one or two orders of magnitude.

The paper is organized as follows. Section 2 describe the concept of the WMM. Section 3 presents the mixture density model and the generation of the sample of outlier correspondences. The generation of the WMMs is described in Sect. 4. Section 5 presents the details of the algorithm. The results are shown and discussed in Sect. 6. The paper is concluded in Sect. 7.

A short version of this paper was presented at the IEEE Conference on Computer Vision and Pattern Recognition (Goshen and Shimshoni 2005).

2 Weak Motion Model

Given a set of N putative correspondences $\{p_i \leftrightarrow p'_i\}$, we seek to find the set of inlier correspondences and we wish to compute the fundamental matrix between the two images. The algorithm that we describe in the next sections uses WMMs to solve these problems. A WMM is a motion model that roughly approximates the motion that a point underwent from the first image to the second image. In general each point in the first image moves to a point on its corresponding epipolar line, where the position on the line

depends of the depth of the 3D point. However, the motion of images of points which are close in 3D can be approximated by a more constrained motion. These probabilities enable guided sampling from the putative correspondences. Several possible transformations can be considered to be used as WMMs: a 2D homography, an affine transformation or a similarity transformation. An homography can be induced in a scene by a world plane, but in a general scene it is not easy to generate an homography since at least four point correspondences on a plane in the scene have to be found first. The other two transformations can be generated very easily: an affine transformation can be generated by any three correspondences and a similarity transformation by any two correspondences. These transformations have a geometric meaning as they are approximations of the homography, which is associated with the 3D plane on which the points lie, between the first and the second images. These approximations are not very accurate and the plane does not have to be part of the scene. This however is not critical, since the nature of the WMM is to give only rough approximations. We decided to use affine transformations as WMMs because they have more generalization ability than the similarity transformations and they are approximations of the corresponding homography.

The matrix representing the WMM is

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

where the matrix $A = [a_{ij}]$ is an invertible 2×2 matrix and (t_x, t_y) is a 2D translation. Note that three points in the joint image space (Triggs 1995) define a unique affine transformation and that an affine transformation is a linear transformation that defines a two dimensional manifold in the joint image space. We have found that very often an affine transformation that has been formed from three inlier correspondences can be used as a WMM.

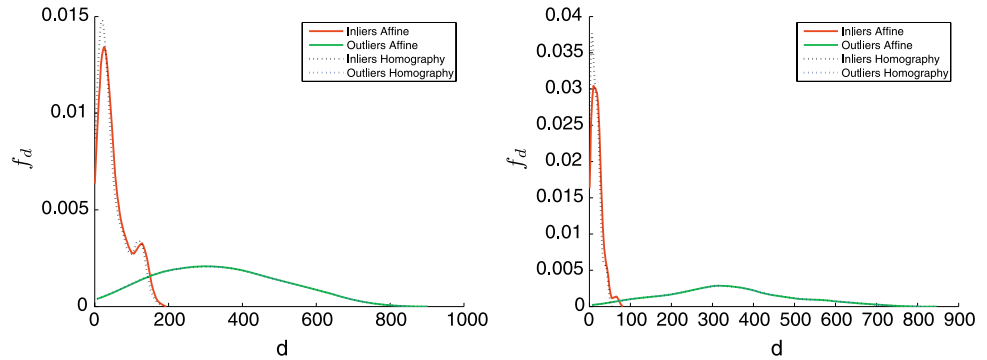
Let $\{w_j\}$ be a set of N_w WMMs. Let d_{ij} be the geometric distance from correspondence $p_i \leftrightarrow p'_i$ to the manifold defined by the WMM w_j , i.e.

$$\begin{aligned} d_{ij} &= \min_{\tilde{p}} \sqrt{\|p_i - \tilde{p}\|^2 + \|p'_i - H_{a_j} \tilde{p}\|^2} \\ &= \min_{\tilde{p}} \left\| \begin{bmatrix} H_{a_j} \\ I \end{bmatrix} \tilde{p} - \begin{pmatrix} p'_i \\ p_i \end{pmatrix} \right\|, \end{aligned}$$

where H_{a_j} is the affine transformation representing w_j . This geometric distance can be easily analytically calculated because the expression is a square of a linear transformation.

When an inlier pair is considered, the value of d_{ij} is the length of the sum of two error vectors. The first vector is the difference between position computed by the affine transformation and the true homography associated with the plane

Fig. 1 Simulation results comparing distances obtained using an homography WMM with distances obtained using an affine transformation WMM



spanned by the three projective 3D points created from the three pairs of points used to compute the affine transformation. The second vector is the result of the projective depth of the point pair relative to the plane. Usually the first component is quite small whereas the second component varies depending on the distance of the point from the plane. When an outlier pair is considered the distance is usually much larger as the point pairs usually do not satisfy the epipolar constraint and this usually results in large projective depth values.

As mentioned above homographies represent accurately the distance from a plane. There is no way however to generate homographies from the data because it requires that four pairs of corresponding points whose 3D points lie on a plane be found. In the simulation results presented in Fig. 1 we generated a wide baseline image pair of a simulated 3D scene for which inlier and outlier matches were generated. We compared the distributions of distances obtained for one WMM using an affine transformation with the distances obtained for its corresponding homography. In the figure two sets of results are presented. When outliers are considered, the two distributions are practically identical. For inliers, the distances obtained for the homography are usually smaller than the ones obtained for the affine transformation and if homography-based WMMs could be generated the results would be slightly better, but in both cases the difference between the distribution of inliers and the outliers is considerable. The conclusion is that even if the affine transformation does not closely approximate the homography, using it should lead to satisfactory results.

Let d_i be the median distance of $p_i \leftrightarrow p'_i$ to the manifolds, i.e.

$$d_i = \text{med}\{d_{ij}\}_{j=1}^{N_w}.$$

Thus, for inliers only the $N_w/2$ affine models with smaller projective depths are used, resulting in a small value. Each of the affine transformations covers a certain part of the scene and therefore will produce a small distance for some of the inlier pairs. Small distances are also produced for some of the other transformations. Choosing the minimal distance

yielded worse results because quite a few outliers would have one low distance making them hard to be distinguished from the inliers.

The median distance can be considered as a random variable and is modeled as a mixture model:

$$f_d(d_i) = f_{in}(d_i)(1 - \epsilon) + f_{out}(d_i)\epsilon,$$

where $f_{in}(d_i) = f(d_i | p_i \leftrightarrow p'_i \text{ inlier})$, $f_{out}(d_i) = f(d_i | p_i \leftrightarrow p'_i \text{ outlier})$ and ϵ is the mixing parameter which is the probability that any selected correspondence is an outlier. Following the description given above the distribution of the distances associated with the inliers should be very different than the distances associated with the outliers.

3 Probability Estimation Using an Outlier Sample

The probability, $P_{in}(i)$, that correspondence $p_i \leftrightarrow p'_i$ is an inlier can be calculated using Bayes' rule. Due to the fact that d_i is a continuous variable the continuous version of Bayes' rule is applied in which probabilities are replaced by density functions (Duda et al. 2001, Chap. 2). The result is therefore:

$$\begin{aligned} P_{in}(i) &= P(p_i \leftrightarrow p'_i \text{ inlier} | d_i) \\ &= \frac{P(d_i | p_i \leftrightarrow p'_i \text{ inlier}) P(p_i \leftrightarrow p'_i \text{ inlier})}{P(d_i)} \\ &= \frac{f_{in}(d_i)(1 - \epsilon)}{f_d(d_i)}. \end{aligned}$$

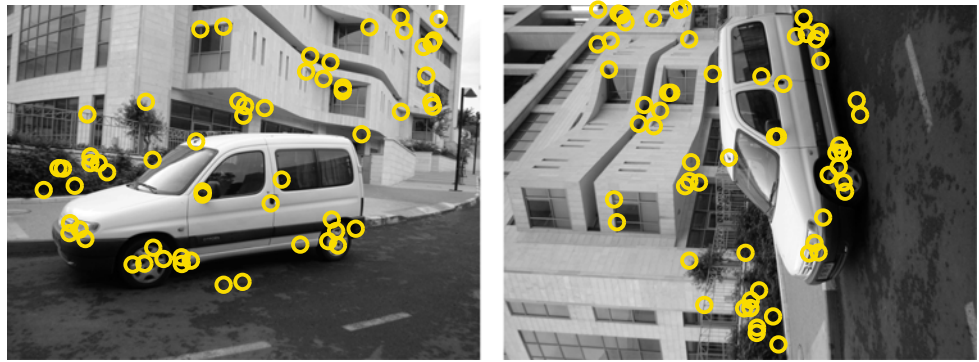
We estimate this probability in a non-parametric manner. All we assume is that the median distances of the inlier correspondences, d_i , are bounded by an unknown parameter D and that the outlier correspondences are not. Thus,

$$f_d(d_i) = \begin{cases} f_{in}(d_i)(1 - \epsilon) + f_{out}(d_i)\epsilon, & d_i \leq D; \\ f_{out}(d_i)\epsilon, & \text{otherwise.} \end{cases}$$

We obtain

$$P_{in}(i) = \begin{cases} \frac{f_d(d_i) - f_{out}(d_i)\epsilon}{f_d(d_i)}, & d_i \leq D; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Fig. 2 Car image pair, with correct matching points marked



The algorithm estimates $f_d()$ using a kernel density estimator. The kernel estimator with given kernel K is defined by

$$\widehat{pdf}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

where h is the kernel width, n is the number of sampled data points and X_i is the i^{th} observation of the random variable X . We use a Gaussian kernel function and the L-Stage Plug In method (Wand and Jones 1995) to estimate the kernel width. I.e.,

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5},$$

where $\hat{\sigma}$ is the sample standard deviation.

The estimation of $f_{out}()$, which is also needed to compute the probabilities, is more problematic. We usually do not have any prior knowledge about this pdf. Therefore we turn to generate a sample of outlier correspondences. Given such a sample $\{o_i \leftrightarrow o'_i\}, i = 1, \dots, N_o$ we can estimate $f_{out}()$ in the same way that we estimated $f_d()$.

We tried the following three methods to generate the outlier sample.

1. **Uniformly.** Each point of each correspondence is sampled uniformly from the image.
2. **Corner based.** Each point of each correspondence is sampled uniformly from the corner sets of the images.
3. **Algorithm guided.** Using the same algorithm that generates the putative correspondences $\{p_i \leftrightarrow p'_i\}$, only this time it excludes from the input of the algorithm the correspondences that already have been used to generate $\{p_i \leftrightarrow p'_i\}$. In our experiments only mutually best candidates were selected as putative correspondences. So the entries that were selected for the putative correspondences were removed and the outlier sample was generated using the same method.

We found that the algorithm guided method gives the best estimation to the outlier distribution. To demonstrate this, a

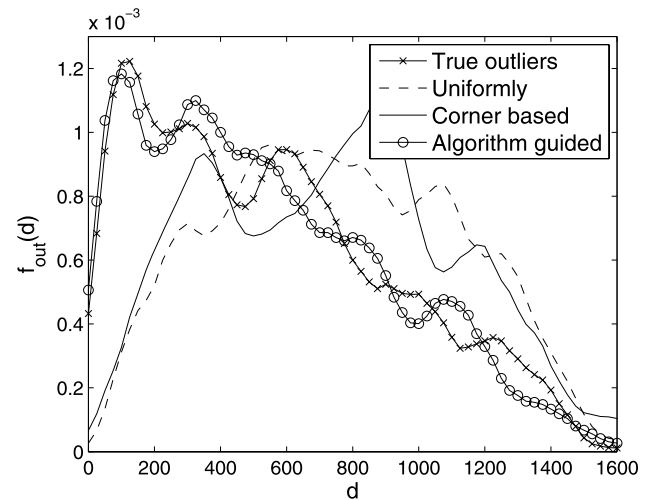


Fig. 3 The distributions for the car image pair of the actual outliers, uniformly method, corner based method and the algorithm guided method. The algorithm guided method gives the best estimation

set of ten WMMs was randomly generated for the image pair in Fig. 2 (using the method which will be described in Sect. 5). Figure 3 shows the distribution of the actual outliers and of the three aforementioned methods. The distributions in this figure and in rest of the figures in this paper are shown with Gaussian kernel smoothing with the L-Stage Plug In method for bandwidth estimation. In Fig. 3 the algorithm guided method gives a much better estimate of the $f_{out}()$ than the other two methods.

The reason for this is that the outliers have a distribution that depends on several factors. It depends on the distribution of the corners in the images. Usually there are regions in the image that have a larger density of corners than others. For example in Fig. 2 the region of the white car has much less corners than the region of the bushes. The distribution also depends on the texture around the corners. For example a corner in the bushes of Fig. 2 has a better chance to be matched to another corner in the bushes, because the texture around the corners in the bushes is more similar to each other than to corners from other regions. The second

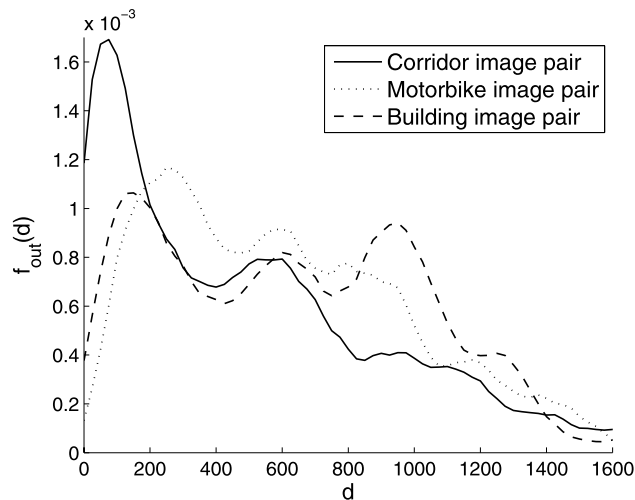


Fig. 4 Example of outlier distributions for different image pairs

method takes into consideration the distribution of the corners in the images. Figure 2 shows that the second method gives a slightly better estimation relative to the first method. The algorithm guided method gives the best estimation. It takes into consideration the corner distribution, the similarity between the corners and the matching technique.

One could think that the outlier distribution does not change from image pair to image pair, but this is not true. Figure 4 shows examples of three more outlier distributions of different image pairs. This figure demonstrates the variety of outlier distributions that can be obtained and that in order to achieve better results the outlier sample has to be generated for each image pair.

The median distances of the inlier correspondences are bounded by the unknown parameter D . At this point we already have the outlier sample and assuming that we know the outlier fraction ϵ the expected number of inliers is $(1 - \epsilon)N$. Therefore, for values less than D there should be $(1 - \epsilon)N$ inliers and the rest should be outliers. Since the artificially generated outliers are assumed to follow the same distribution as the outliers in the putative correspondences, we estimate that the same fraction of outliers of the N putative correspondences have values who are less than D , i.e.

$$D = \min \left\{ d \left| \left[\sum_{i=1}^N I(d_i < d) - \frac{N}{N_o} \sum_{i=1}^{N_o} I(d_i^o < d) \right] \right. \right. \\ \left. \left. = \lceil (1 - \epsilon)N \rceil \right\}, \quad (4)$$

where d_i^o is the median of the distances of $o_i \leftrightarrow o'_i$ to the WMM manifolds and $I()$ is the index function, i.e. $I(\text{True}) = 1$ and $I(\text{False}) = 0$. Figure 5 illustrates the D estimation method. As can be seen from the illustration all the inliers (whose number has already been estimated) have

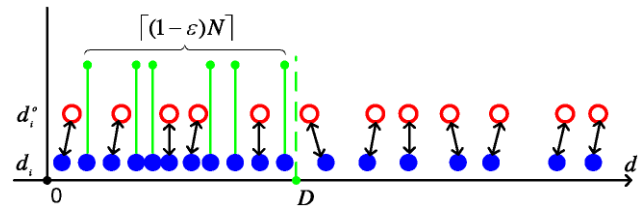


Fig. 5 Illustration of the estimation of D . The empty red bullets represent the outlier sample distances and the filled blue bullets represent the putative correspondences distances. D is the smallest distance at which the number of putative correspondences less the expected number of outliers according to the outlier sample is equal to the expected number of inliers, i.e. $\lceil (1 - \epsilon)N \rceil$

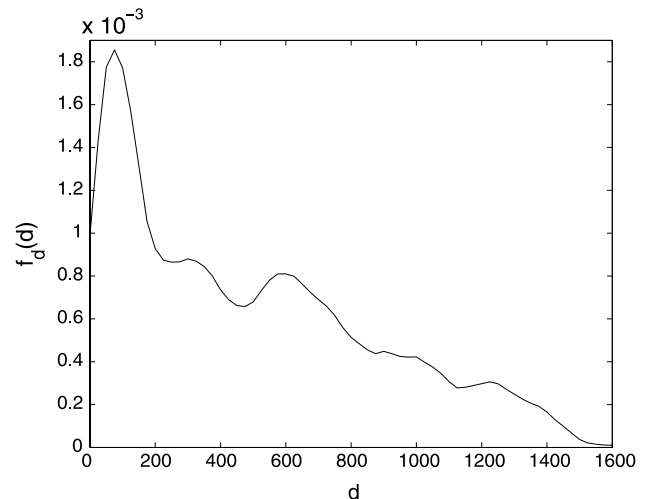


Fig. 6 $f_d(d)$ density function for the car image pair

values less than D . Thus, using the smallest value of D enables us to remove the largest number of outlier matches by assigning them probability zero.

In the method described above we have assumed the ϵ is known in advance. This is of course not true in general. A method for the estimation of ϵ is described in Sect. 5.

The following figures show results for the car image pair. Figure 6 shows the $f_d()$ density function, Fig. 7 shows the $f_{out}()$ and the $f_{in}()$ density functions. In practice the algorithm does not compute $f_{in}()$. The values shown in the graph were computed from the ground truth and are presented here for illustrative purposes. Figure 8 shows histograms of $P_{in}()$ for inlier and outlier correspondences computed using (3). Note that a large number of outlier correspondences get probability zero to be an inlier and a large number of the inlier correspondences get probability 0.6 to be an inlier while the original inlier rate in the putative correspondences is only 0.22. These probabilities are used to guide the RANSAC step of the algorithm and accelerate the search process. For illustration Fig. 9 shows some of the inlier correspondences with solid yellow lines, three outlier correspondences with probability greater than zero to be an inlier

with black solid lines and three outlier correspondences with probability zero to be an inlier with yellow dashed lines. Note that outlier correspondences with probabilities greater than zero have motions similar to the inliers and can not be discarded using the WMMs. However outlier correspondences with probability zero have a different motion relative to the inliers and have been successfully detected using the WMMs.

4 Weak Motion Model Generation

In order to obtain a set of N_w WMMs A RANSAC style algorithm is used. At each iteration a subset of three correspondences is randomly chosen from the putative correspondences. Each subset uniquely defines an affine transformation. After $N_w/(1-\epsilon)^3$ iterations on average N_w subsets containing only inlier correspondences have been generated. Each subset gets a score. The score is the geometric distance of the $(1-\epsilon)N$ th closest correspondence to the manifold of the affine transformation of the subset. A subset that fits an

affine transformation that can be used as a WMM gets a low score because at least $(1-\epsilon)N$ inliers are relatively close to the manifold and therefore the geometric distance of the $(1-\epsilon)N$ th closest correspondence is relatively small. An affine transformation based on a subset that includes outliers usually can not be used as a WMM and does not have a set of $(1-\epsilon)N$ correspondences that are relatively close to its manifold and because of that will get a relatively high score. We take the set of N_w models with the lowest scores to be the set of the WMMs.

5 The Algorithm

In this section we describe the flow of the algorithm. The generation of the WMMs and the computation of the probability of each correspondence to be an inlier depend on ϵ . In general ϵ is not known in advance. The algorithm searches for the correct ϵ starting from the lowest rate level up to the highest rate level from a rate level set, $\{\epsilon_j\}$, given as input by the user. The algorithm is summarized in Algorithm 1. The details of the algorithm are as follows:

Algorithm 1 Guided Sampling via WMMs and Outlier Sample Generation

1. Generate an outlier sample
 2. For each ϵ_j starting from the smallest up to the highest:
 - (a) Generate random WMMs for the current rate level ϵ_j
 - (b) Fine tune the estimate of the outlier rate
 - (c) Estimate inlier probabilities
 - (d) Estimate the number of iterations, \hat{N}_s , for the guided LO-RANSAC
 - (e) If $\hat{N}_s < N_t$ then execute guided LO-RANSAC
 - (f) If the stopping criteria are validated then terminate
-

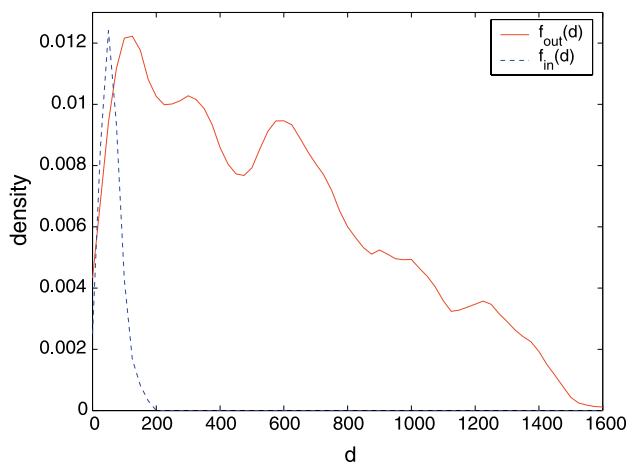


Fig. 7 $f_{out}(d)$ and $f_{in}(d)$ density functions for the car image pair

Fig. 8 Histograms for $P_{in}()$ of
(a) inlier correspondences
(b) outlier correspondences for the car image pair

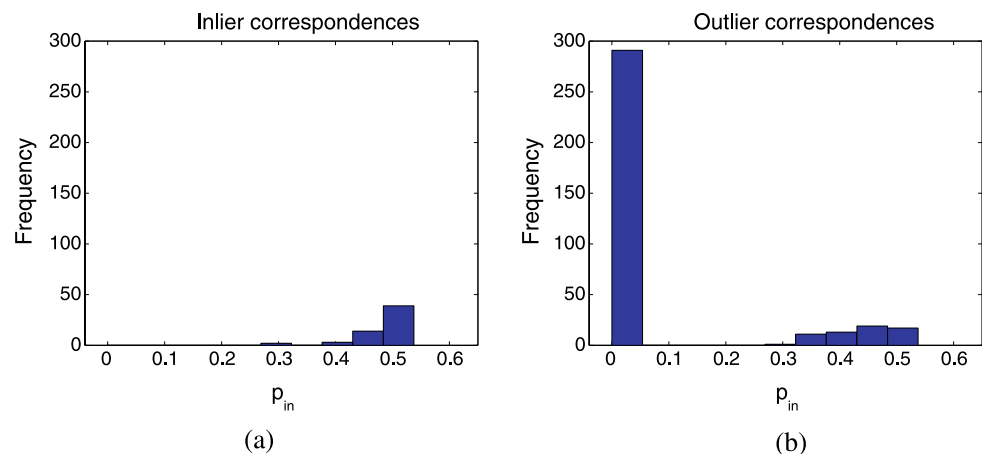




Fig. 9 (Color online) Several inlier correspondences are shown with *solid yellow lines*, three outlier correspondences with a probability greater than zero to be an inlier are shown with *black solid lines* and three outlier correspondences with probability zero to be an inlier are shown with *dashed white lines*

Generation of Outlier Sample

The algorithm generates an outlier sample. This step is done only once at the beginning of the algorithm using the algorithm guided method.

Generation of Random Weak Motion Models

The generation of the random WMMs is done using the method described in Sect. 4. However in Sect. 4 we assumed that ϵ is known in advance. As in practice this is not the case we work on the whole rate level set $\{\epsilon_j\}$ in parallel. The data structure used for the generation of the WMMs is illustrated in Fig. 10. Each rate level, ϵ_j , is associated with a number of iterations $N_{I_j} = \lceil N_w / (1 - \epsilon_j)^3 \rceil$. Each rate level has a set, B_j , of N_w suspected WMMs. These sets are updated during the generation of the random affine transformations.

Each set B_j keeps the N_w best affine transformations found until now that are suspected to be WMMs. In each iteration a new subset of three correspondences is randomly chosen and the appropriate affine transformation is formed. The geometric distance of each of the correspondences to the manifold of the affine transformation is calculated. Each rate level ϵ_j gives a different score to this affine transformation. The score is the distance of the $(1 - \epsilon_j)N$ th closest correspondence to the manifold of the affine transformation. Each of the rate levels that the algorithm did not pass until now, keeps the current affine transformation if it has one of the N_w lowest scores obtained till now. In our implementation each of the WMM sets, B_j , is kept in a different heap. Thus, the sets can be updated very efficiently. The random generation of the affine transformations continues until the number of iterations reaches N_{I_j} . Then the algorithm can be applied to ϵ_j .

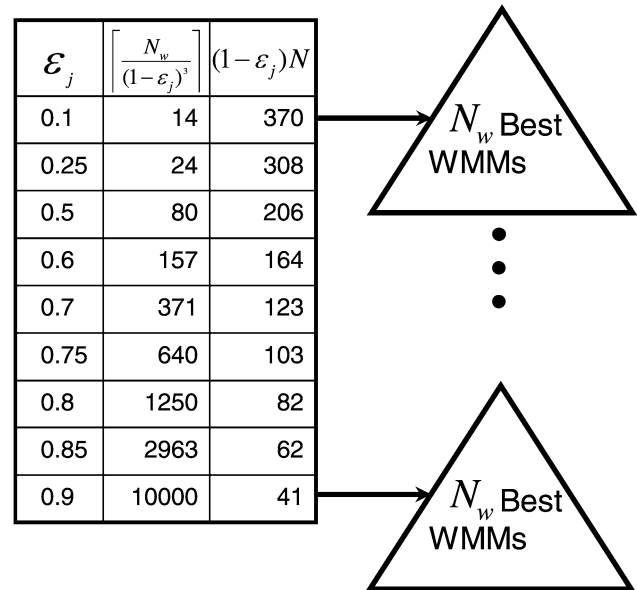


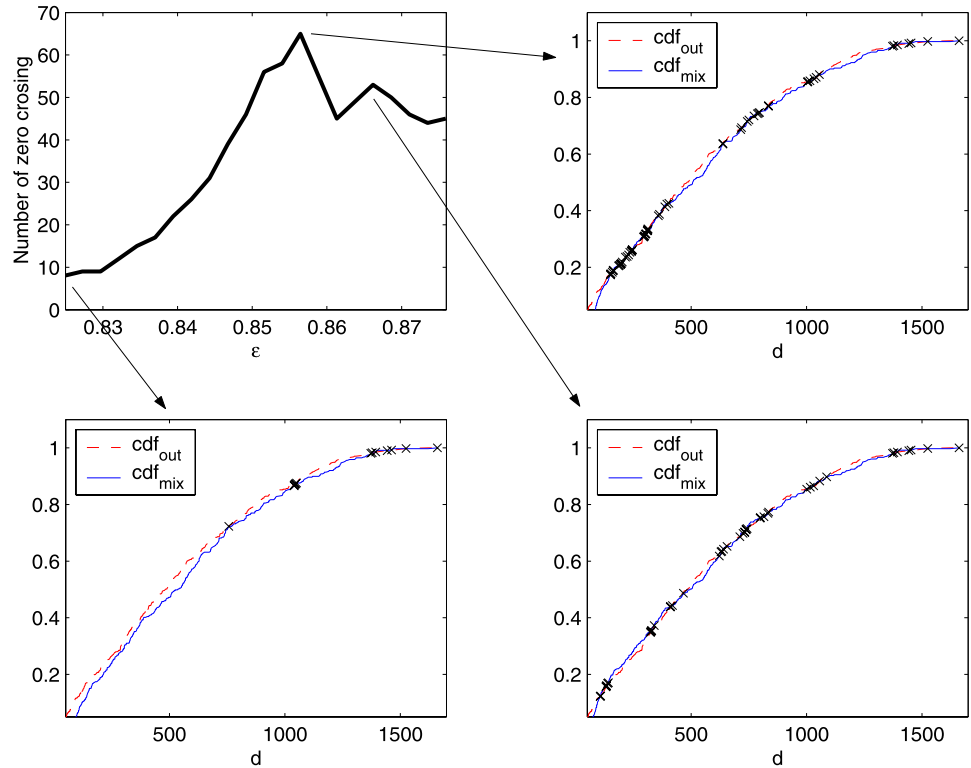
Fig. 10 Illustration of the generation of WMMs stage. Each rate level ϵ_j is associated with the number of iterations $\lceil N_w / (1 - \epsilon_j)^3 \rceil$ and a score that is the distance of the $(1 - \epsilon_j)N$ th closest correspondence to the manifold of the affine transformation. $N = 411$ is the number of putative correspondences of the car image pair. Each rate level keeps in a heap the N_w affine transformations with the lowest scores obtained until now

Fine Tune the Estimate of The Outlier Rate

The outlier rate level set $\{\epsilon_j\}$ is given as an input from the user. So no ϵ_j is expected to be the accurate outlier rate of the putative correspondences. In this step the algorithm fine tunes the estimation of the outlier rate. First the median distance sets of the putative correspondences $\{d_i\}$ and of the outlier sample $\{d_i^o\}$ are calculated according to the current set of WMMs, B_j . The search for the accurate outlier rate is carried out around ϵ_j in the region $A = [\frac{\epsilon_j + \epsilon_{j-1}}{2}, \frac{\epsilon_j + \epsilon_{j+1}}{2}]$. Note that for each $\epsilon \in A$ there is an appropriate D_ϵ according to (4).

Any correspondence whose median distance is larger than D is considered an outlier. Therefore, for distances larger than D we have two distributions of outliers: median distance distribution from the putative correspondences and the median distance distribution from the outlier sample. Therefore, the algorithm seeks to find $\hat{\epsilon}_j$ that gives the best fit between these two distributions for distances greater than D . Let $cdf_{out}()$ be the empirical cumulative distribution function of the outlier sample, i.e. $cdf_{out}(d) = \frac{\sum_{i=1}^{N_o} I(d_i^o < d)}{N_o}$ and let $cdf_{mix}()$ be the empirical distribution function of the outlier pdf given by the mixture model, i.e. $cdf_{mix}(d) = (\frac{\sum_{i=1}^N I(d_i < d)}{N} - (1 - \epsilon)) / \epsilon$ for $d > D$. These two distributions are noisy and when they fit each other we expect that their curves will cross each other in many places. Therefore, the quality of the fit between these cdfs is measured

Fig. 11 An example of the fine tuning process of $\hat{\epsilon}$ for the car image pair. The figure shows the number of zero crossing with respect to ϵ and several corresponding cdf functions. The zero crossing locations are marked by \times 's



by the number of zero crossings of $cdf_{out}(d) - cdf_{mix}(d)$, for $d > D$, where a large number represents a good fit (see Fig. 11). To ensure that this score will not be biased toward lower values of ϵ the parameter D is set to be $D_{(\epsilon_j + \epsilon_{j-1})/2}$ for all of the checked values of ϵ . Other methods for cdf comparison can also be used in this step, such as the Kolmogorov-Smirnov test. In that test $\max_d |cdf_{out}(d) - cdf_{mix}(d)|$ is computed. As the starting position in the cdf from where the test is applied D is only approximated, the maximal value obtained might be noisy and might yield high values. Therefore we chose the other method which yields a high value for two approximations of the true outlier cdf.

Consider the special case when the algorithm is given a pair of non-overlapping images. In this case the matches given to the algorithm and the outlier sample of matches come from the same distribution, the outlier distribution. Therefore this step of the algorithm will detect a value of ϵ close to one and decide to terminate.

Inlier Probability Estimation

The probability, $P_{in}(i)$, that correspondence $p_i \leftrightarrow p'_i$ is an inlier is calculated for all the correspondences by

$$P_{in}(i) = \begin{cases} \frac{f_d(d_i) - f_{out}(d_i)\hat{\epsilon}_j}{f_d(d_i)}, & d_i \leq D\hat{\epsilon}_j; \\ 0, & \text{otherwise.} \end{cases}$$

The estimate of the density function at each point is done using the kernel density estimator described in Sect. 3.

Iteration Number Estimation

The inlier probability estimation is used to accelerate the search for the fundamental matrix using a guided RANSAC algorithm. The correspondences are sampled according to their inlier probability. In each guided RANSAC iteration a subset sample of seven correspondences is chosen. Correspondence i has probability $\frac{P_{in}(i)}{\sum_{i=1}^N P_{in}(i)}$ to be chosen. The number of samples N_s is chosen sufficiently high to ensure with probability p that at least one of the samples of seven correspondences is free from outliers. Thus (2) has to be generalized. Let S be the set of all possible subsets of seven correspondences. The probability of subset $s_i \in S$ to be chosen in the guided sampling step is

$$P_c(s_i) = \prod_{j=1}^7 \frac{P_{in}(c_{ij})}{\sum_{k=1}^N P_{in}(k)},$$

where c_{ij} is the j^{th} correspondence in subset s_i . The probability that this seven tuple consists of only inliers is:

$$P_t(s_i) = \prod_{j=1}^7 P_{in}(c_{ij}).$$

The calculation of N_s has to take into consideration all the possible combination series of subsets s_i where each subset is chosen according the probability function $P_c()$. This type

of calculation is infeasible. We can however estimate N_s using a Monte Carlo method. We generate a series M of length N_M of subsets s_i . Each member in the series is chosen in the same manner as in the guided sampling step, i.e. according to the probability function $P_c()$. The series M represents an “average” subseries of the guided sampling step. We calculate how many such subseries should be taken in order to ensure with probability p that at least one of the samples of seven correspondences is free from outliers. We have

$$\left(\prod_{i=1}^{N_M} (1 - P_t(m_i)) \right)^K = 1 - p,$$

where $m_i \in M$ is a subset of seven correspondences and K is the number of subseries needed to be taken in order to ensure the probability p . We obtain

$$K = \log(1 - p) / \log \left(\prod_{i=1}^{N_M} (1 - P_t(m_i)) \right).$$

Now we can estimate N_s as

$$\hat{N}_s = K * N_M.$$

The length of series M should be sufficiently long to ensure that M represents an “average” subseries of the guided sampling step. We chose $p = 0.99$ and $N_m = 2000$. As a result the estimated N_s will yield a sample of seven inlier correspondences with probability of at least $p = 0.985$ for widely diverse probability distributions.

When the WMMs did not provide enough probabilistic information to reduce the number of iterations \hat{N}_s below a user defined threshold N_t the next step is not performed and the next ϵ_j is tested.

Guided LO-RANSAC

In this step a slightly different version of LO-RANSAC algorithm is applied. The difference is in the sampling technique. Instead of random sampling, the samples are chosen according to the probabilities $P_{in}(i)$, i.e. correspondence i has a probability

$$\frac{P_{in}(i)}{\sum_{i=1}^N P_{in}(i)}$$

to be chosen. In each global iteration a subset of seven correspondences is sampled and the fundamental matrix satisfying these matches is computed. The set of putative matches whose Sampson distance (see its definition in Hartley and Zisserman (2000)) is less than a certain preset threshold is found. If the size of this set, which we will call the support set of the candidate fundamental set, is the largest so far the LO-RANSAC step is applied several times. In this step

a larger set of matches from the support set is chosen and from it the fundamental matrix is estimated. Due to the set's size, the estimate of the fundamental matrix is usually more accurate, yielding an even larger support set. The number of global iterations is limited to \hat{N}_s . At the end of this step, the fundamental matrix F with the largest support set N_{in_j} is kept.

Note that the algorithm has two RANSAC style steps, the WMM generation step and the guided LO-RANSAC step. Each iteration of the second type is more costly because in each iteration a fundamental matrix has to be computed based on seven correspondences. This involves computing the roots of a polynomial of degree three which may have up to three solutions for the fundamental matrix and the distances of the putative correspondences to each of the non linear manifolds of the fundamental matrices are calculated. On the other hand, in each iteration of the random WMM generation step a unique affine transformation is formed using a linear algorithm and the distances of the putative correspondences to the linear manifold of the affine transformation have to be calculated. In our implementation the running time of an average iteration of the LO-RANSAC is 4.5 times more costly than an average iteration of the random WMM generation step.

Checking the Stopping Criteria of the Algorithm

In this step the algorithm checks the stopping criteria. The algorithm has two estimates of the number of inliers: $N(1 - \hat{\epsilon}_j)$ from the mixture model and N_{in_j} from the guided LO-RANSAC. If the two estimations are approximately the same then the algorithm terminates and returns the fundamental matrix F and the set of N_{in_j} inlier correspondences. The algorithm also terminates in cases that the current estimation of the number of inliers is lower than the largest inlier set found until this stage. In such cases the algorithm already has found a solution with a number of inliers that is greater than the expected number of inliers in the current stage. Therefore, it is not worthwhile to continue the search. This situation is checked after the estimation of $\hat{\epsilon}_j$ in the fine tuning step of the algorithm. In this case the algorithm returns the largest found inlier set and its fundamental matrix.

6 Experiments

The proposed algorithm has been tested on many image pairs of indoor and outdoor scenes several of which are presented here. The proposed algorithm can use putative correspondences generated by any technique. In the experiments we have tested two methods. In the first method, the putative correspondences were detected and matched automatically using a very simple method. Corners were detected by the Harris corner detector. Cross correlation scores

Table 2 Experiment characteristics and results. For image pairs marked with asterisk the putative correspondences were generated using the SIFT keypoint detector. The putative correspondences of the rest of the image pairs were generated using Harris corner detector with rotated cross correlation scores. N is the number of putative correspondences, “In.” is the number of inlier correspondences and ϵ is the outlier rate level. For each algorithm “75%” is the average number of samples until at least 75% of the true inliers are found, “End” is the average number of iterations until the termination of the algorithm and “In.” is the average number of inliers that were found until the termination of the algorithm. The numbers in parentheses are the average numbers of the WMMs that have been generated. The last two columns show the average time speedup rate until at least 75% of the possible inliers are found and until the termination of the algorithm. Several numbers in the table are shown with the “greater than” symbol that indicates that there have been cases in which the LO-RANSAC algorithm terminated when it exceeded one million iterations

	Characteristics			LO-RANSAC			Proposed algorithm			Speedup rate	
	N	In.	ϵ	75%	End	In.	75%	End	In.	75%	End
Slide pr.	565	122	0.78	70,878	384,171	114.8	326 (427)	1,388 (1,274)	118.9	18.2	59.2
Desk	303	63	0.79	80,841	341,958	61.7	168 (411)	495 (908)	62.1	42.1	78.3
Lab	730	131	0.82	>232,596	>860,590	124.3	109 (2,216)	1,214 (2534)	129.0	>90	>114
Corridor	401	70	0.83	>228,675	1,000,000	59.7	1,038 (1,327)	1,645 (2,227)	66.4	>137	>178
Motorbike	592	86	0.85	>302,867	1,000,000	80.5	160 (2,963)	737 (2,963)	85.5	>95	>197
Car	411	59	0.86	>248,987	1,000,000	53.5	42 (2,877)	1,016 (2,877)	58.4	>114	>238
Building	363	50	0.86	>602,181	1,000,000	40.3	614 (3,067)	2,023 (3,667)	49.7	>208	>251
Road	401	50	0.88	>547,547	1,000,000	37.6	1,532 (1,937)	5,168 (10,000)	49.2	>270	>181
Yard	508	53	0.90	>508,519	1,000,000	43.5	91 (10,000)	2,843 (10,000)	50.9	>265	>185
Statue*	328	58	0.82	>485,417	1,000,000	43.4	130.2 (1,945)	575 (2,630)	54.7	>103	>141
Van*	323	59	0.82	>327,871	1,000,000	42.1	159.7 (2,231)	492 (2,827)	56.2	>88	>123
Parking*	301	48	0.84	>221,793	1,000,000	37.9	302 (3,015)	459 (4,732)	45.9	>91	>137

were generated between all corners with each patch undergoing 36 evenly spaced rotations, the strongest matches over these rotations are stored. Only mutually best candidates were selected as putative correspondences. In the second method, the putative correspondences were generated using the SIFT keypoint detector (Lowe 2004). We have compared our algorithm with the LO-RANSAC algorithm which gives near perfect agreement with the theoretical performance that is given by (2). The termination criterion based on (2) was set to $p = 0.99$. In cases where the number of iterations exceeded one million the algorithm also terminated. The parameters for the proposed algorithm were as follows. The outlier rate level set, $\{\epsilon_j\}$, was set to $\{0.1, 0.25, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.9, 0.925, 0.95\}$, the number of WMMs generated for each of the outlier rate levels was set to $N_w = 10$, the length of the series M was set to $N_M = 1000$ and the threshold for the guided LO-RANSAC step was $N_t = 3000$. Twelve experiments are presented in this section. Some of the image pairs are shown in Figs. 12 and 13. The cases that are presented here are difficult cases in which the outlier rate is between 0.78–0.9. The characterization of the scenes, the number of putative correspondences, the number of inlier correspondences and the outlier rate are summarized in Table 2. Each algorithm has been applied to each image pair twenty times. For each algorithm the following statistics are presented: The number of samples until at least 75% of the true inliers are found (as was suggested in Tordoff and Murray (2002)), the number of iterations until the termination of the algorithm, and

the number of inliers that have been found until the termination of the algorithm. The number of generated WMMs is also given. The average speedup rate time achieved in the experiments until at least 75% of the possible inliers have been found is between 18.2–270 and the average speedup rate time until the termination of the algorithm is between 59.2–251. Note that several numbers in the table are shown with the “greater than” symbol that indicates that at least in one case the LO-RANSAC algorithm terminated when it exceeded one million iterations. Removal of this early termination criterion yields much higher results. In addition, the number of inlier correspondences found by the proposed algorithm is larger than in the ones achieved by LO-RANSAC and the results are near perfect.

7 Summary and Conclusions

In this work, we presented a novel algorithm for automatic robust estimation of the epipolar geometry in cases where the correspondences are contaminated with a high percentage of outliers. The algorithm uses WMMs that roughly approximate the motion of correspondences from one image to the other. The algorithm generates a set of WMMs and represents the distribution of the median of geometric distances of a correspondence to the WMMs as a mixture model of outlier correspondences and inlier correspondences. An outlier correspondences sample is then generated using the algorithm guided method. This method takes into consider-



(a) Corridor image pair



(b) Building image pair



(c) Motorbike image pair

Fig. 12 Some of the real image pairs on which the algorithms were tested



(a) Van image pair



(b) Parking image pair



(c) Statue image pair

Fig. 13 Another set of real image pairs on which the algorithms were tested

ation the corner distribution in the images, the similarity between the corner neighborhoods and the matching technique. The outlier sample is used to estimate the outlier rate and to estimate the outlier pdf. These pdfs are then used to estimate the probability that each correspondence is an inlier. These probabilities guide the sampling which accelerates the search process. The performance of the algorithm was compared to the performance of the LO-RANSAC algorithm that gives near perfect agreement with the number of iterations given by (2). The resulting algorithm when tested on real image pairs achieves speedup rates of 59.2–251.

References

- Adam, A., Rivlin, E., & Shimshoni, I. (2001). ROR: Rejection of outliers by rotations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(1), 78–84.
- Chen, H., & Meer, P. (2003). Robust regression with projection based m-estimators. In *International conference on computer vision* (pp. 878–885).
- Chum, O., Matas, J., & Kittler, J. V. (2003). Locally optimized RANSAC. In *German pattern recognition symposium* (pp. 236–243).
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.
- Freund, Y., & Schapire, R. (1997). A decision-theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.
- Goshen, L., & Shimshoni, I. (2005). Guided sampling via weak motion models and outlier sample generation for epipolar geometry estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. I: 1105–1112).
- Hartley, R. I., & Zisserman, A. (2000). *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press.
- Lee, K. M., Meer, P., & Park, R. H. (1998). Robust adaptive segmentation of range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2), 200–205.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
- Matas, J., & Chum, O. (2004). Randomized RANSAC with $T_{d,d}$ test. *Image and Vision Computing*, 22(10), 837–842.
- Moisan, L., & Stival, B. (2004). A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 57(3), 201–218.
- Rozenfeld, S., & Shimshoni, I. (2005). The modified pbM-estimator method and a runtime analysis technique for the ransac family. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. I: 1113–1120).
- Stewart, C. V. (1995). Minpran: a new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10), 925–938.
- Subbarao, R., & Meer, P. (2007). Discontinuity preserving filtering over analytic manifolds. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–6).
- Tordoff, B., & Murray, D. W. (2002). Guided sampling and consensus for motion estimation. In *European conference on computer vision* (pp. I: 82–96).
- Torr, P. H. S. (1995). *Motion segmentation and outlier detection*. Ph.D. thesis, Department of Engineering Science, University of Oxford.
- Torr, P. H. S., & Davidson, C. (2003). IMPSAC: Synthesis of importance sampling and random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3), 354–364.
- Torr, P. H. S., & Zisserman, A. (2000). MLESAC: a new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1), 138–156.
- Triggs, B. (1995). Matching constraints and the joint image. In *International conference on computer vision* (pp. 338–343).
- Triggs, B. (2001). Joint feature distributions for image correspondence. In *International conference on computer vision* (pp. II: 201–208).
- Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman & Hall.
- Wang, H., & Suter, D. (2004a). Mdpe: a very robust estimator for model fitting and range image segmentation. *International Journal of Computer Vision*, 59(2), 139–166.
- Wang, H., & Suter, D. (2004b). Robust adaptive-scale parametric model estimation for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11), 1459–1474.
- Yu, X. M., Bui, T. D., & Krzyzak, A. (1994). Robust estimation for range image segmentation and reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5), 530–538.