Fusing appearance and geometric constraints for estimating the epipolar geometry

Miguel Lourenço and Nuno Gonçalves
Institute of Systems and Robotics,
University of Coimbra, Portugal
{miguel, nunogon}@isr.uc.pt

Abstract

Recovering the epipolar geometry of a stereo image pair is important for many computer and robotic vision systems, for performing motion recovering, 3D reconstruction and, more recently, image retrieval from large databases. Most state-of-the-art methods for estimating the fundamental matrix rely solely in putative image correspondences, and, therefore, heavily depend on the capability of the low-level image features to provide enough distinctiveness capabilities for establishing correct matches.

In this paper we present a robust method for estimating the fundamental matrix based on all image features, and not only matching points. This is done by selecting the best correspondent keypoints between views through a proper weighting function that fuses local appearance of keypoints and distance to the epipolar lines. Several distance weighting functions are compared, with an intuitive theoretical analysis of the role of each function parametrization being analyzed. Experimental evidence shows that our approach outperforms the current state-of-the-art methods in terms of error magnitude, number of correct matches provided and computational time.

1. Introduction

The recovering of the epipolar geometry of an image pair has been intensively studied due to its usefulness in many computer and robotic vision applications, such as 3D reconstruction [18, 21], self calibration [7], and image localization [17] and retrieval [1]. The interest in the estimation of the epipolar geometry dates back to [10] where Longuet-Higgins introduced the essential matrix. More recently, the application of classic projective geometry enabled the development of the fundamental matrix (F). This matrix describes the projective relation between two images, encapsulating all the information that is possible to recover with a pair of fully uncalibrated cameras [7, 18].

Most state-of-the-art methods for estimating F matrix take as input a set of putative point correspondences [12, 15] between a stereo image pair. A straightforward solution for the estimation of the F matrix would involve three steps: (i) stack the observation obtained from the matching step; (ii) build a suitable cost function, and (iii) find the correct solution using a minimization procedure. Due to the frequent presence of erroneous matches, this approach must be preceded by a robust estimation step that finds a suitable initialization for the fundamental matrix. The robust estimation of the fundamental matrix is typically achieved by running the 7- or 8-point algorithm within a Sample Consensus framework [2–4, 14, 19, 21, 22]. The hypotheses are ranked according to a certain criterion and the set of putative correspondences is divided into inliers and outliers based on how well they fit the epipolar geometry defined by the winning candidate.

In this paper we propose a novel cost function that fuses the geometric error with local feature appearance, and that can be directly plugged into the existing robust estimation frameworks. Our contributions are the following:

- A novel cost function for the estimation of the fundamental matrix that uses local image appearance [9, 16] and geometric error [7]. We provide an intuitive explanation about the role of both features appearance and distance weighting functions. We briefly explain how to used this new cost function in RANSAC [4, 5] and MAPSAC [21];
- A new algorithm for estimating the fundamental matrix is presented. We start by using the proposed cost function inside a MAPSAC framework to obtain an initial estimation of F. The estimation is then refined using a set of newly selected correspondences that verify appearance and geometric constraints.

1.1. Paper Outline and Notation

The paper outline is as follows: Section 2 reviews the background theory related with the fundamental matrix and

978-1-4673-5054-9/13/$31.00 ©2013 IEEE

robust estimation schemes. Section 3 introduces the new cost function, and studies the role of different distance weighting functions during the fundamental matrix estimation. Additionally, we show how to include the proposed cost function in a RANSAC/MAPSAC framework and propose a new refinement step that provides state-of-the-art performance for the estimation of the fundamental matrix. The new algorithm is evaluated in simulation experiments with ground truth. Section 4 presents the results in four stereo pairs without ground truth using the benchmark criteria used in [2] and the discussion of the obtained results. We also include a Structure-from-Motion (*SfM*) experiment to assess the utility of our algorithm in this type of applications. Finally, we draw conclusions and point towards new directions of work in section 5.

To improve the readability of the manuscript we define the following notation: Matrices are represented by symbols in sans serif font, e.g. G, and image signals are denoted by symbols in typewriter font , e.g. $\mathtt{I}$. Vectors and vector functions are typically represented by bold symbols, and scalars are indicated by plain letters, e.g $\mathbf{x} = (x, y)^\mathsf{T}$ and $\mathbf{f}(\mathbf{x}) = (f_x(\mathbf{x}), f_y(\mathbf{x}))^\mathsf{T}$.

2. Background theory

2.1. Fundamental matrix estimation

The epipolar geometry representation is contained in the fundamental matrix. The fundamental matrix F is a 3×3 matrix, which given an image stereo pair $(\mathtt{I}, \mathtt{I}')$, relates two corresponding points $\mathbf{x} \in \mathtt{I}$ and $\mathbf{x}' \in \mathtt{I}'$ through the epipolar constraint:

$$\mathbf{x}'^\mathsf{T} \mathsf{F} \mathbf{x} = 0. \tag{1}$$

Geometrically this means that a point $\mathbf{x}$ lies on the epipolar line defined as $\mathbf{l} = \mathbf{x}'^\mathsf{T}\mathsf{F}$, and similarly for $\mathbf{x}'$: $\mathbf{l}' = \mathbf{x}^\mathsf{T}\mathsf{F}^\mathsf{T}$. A not necessary yet relevant step during the estimation of F is the normalization of the data. This pre-processing step permits to improve the conditioning of the problem and the stability of the estimated F matrix. We adopt the Hartley's normalization algorithm [6] that consists in finding a transformation T such that the centroid of $\hat{\mathbf{x}} = \mathsf{T}\mathbf{x}$ is coincident with the origin and the RMS distance of all points to it is $\sqrt{2}$ [6][1]. After a proper algebraic manipulation of Eq.1, it is possible to build an observation matrix A that verifies $\mathsf{A}\mathbf{f} = 0$, with

$$\mathbf{f} = (F_{(1,1)}, F_{(1,2)}, F_{(1,3)}, \cdots, F_{(3,1)}, F_{(3,2)}, F_{(3,3)})^\mathsf{T}$$

and $F_{(i,j)}$ denoting the F element in the i^{th} row and j^{th} column. A singular value decomposition is then applied on A (such that A is decomposed as $\mathsf{A} = \mathsf{U}\mathsf{S}\mathsf{V}^\mathsf{T}$) with the solution for $\mathbf{f}$ being the vector of V coincident with the minimum singular value of S.

For non perfect data, this fundamental matrix does not properly model the epipolar geometry since the epipolar lines do not intersect in a unique epipole. The rank 2 constraint is imposed by computing the SVD of $\mathsf{F}_{fullrank}$, setting the smallest eigenvalue to zero ($\mathsf{S} \rightarrow \hat{\mathsf{S}}$) and computing F using $\mathsf{F} = \mathsf{U}\hat{\mathsf{S}}\mathsf{V}^\mathsf{T}$. The fundamental matrix is finally *denormalized* by: $\mathsf{F} = \mathsf{T}'^\mathsf{T}\mathsf{F}\mathsf{T}$. This algorithm is called the 8-point normalized algorithm [6].

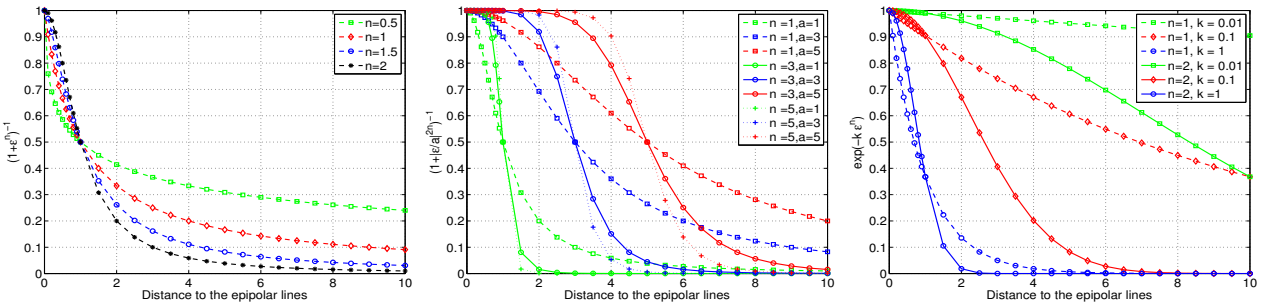
2.2. Robust estimation schemes

For robustness against erroneous point correspondences, the estimation of the fundamental matrix is typically performed by running the 8-point algorithm inside a Sample Consensus framework [3, 4, 14, 21, 22]. The RANSAC (RANdom SAmple Consensus) [3–5], or its variants, starts by sampling at random 8 correspondences for an initial estimation of the F matrix. The hypotheses are ranked according to a certain criterion [19] and the set of point correspondences is divided into inliers and outliers based on how well they fit the epipolar geometry defined by the winning candidate. The RANSAC cost function concerns maximizing the number of inliers, i.e. it outputs the F that best fits the matching points.

Recently some other methods based on RANSAC have been proposed. The MLESAC [23] (Maximum LikElihood SAmple Consensus) is a generalization of RANSAC. The MLESAC employs the same point selection strategy but the solution for F is selected as the one that maximizes a likelihood function, i.e. the shape of a normal distribution instead of the number of inliers. MAPSAC [21] (Maximum A Posteriori SAmple Consensus) extended the MLESAC framework by including Bayesian probabilities during the minimization step, which provides more robustness against noisy correspondences and outliers.

Closely related work to ours can be found in [3, 20]. Instead of performing random sampling of the image correspondences, both methods guide the sampling procedure using the matching scores of the putative matches. While in [20], Tordoff and Murray replaced the random sampling in MLESAC with a sampling guided that uses the probability of correctness of individual correspondences, Chum and Matas [3] proposed to sample progressively larger subsets of top-ranked correspondences. In fact, they do not fuse geometric with appearance constraints since these constraints are used at different steps of the algorithm. Our contribution is different in the sense that we use appearance to weight the F hypothesis, and not to select the initial set of matches. In fact, our contribution is at most complementary to [3]. In this paper we are interested in observe if the proposed cost function is superior to the standard distance-based cost functions and, therefore, the methods of [3, 20] won't be

[1] A similar transformation is computed for $\mathbf{x}'$:$\hat{\mathbf{x}}' = \mathsf{T}'\mathbf{x}'$



(a) IWF (b) BSWF (c) EWF

Figure 1. Analysis of the parametrization of the selected distance weighting functions.

evaluated.

One typical post-processing step consists in refining the best hypothesis obtained from the RANSAC-like algorithms. This refinement step starts by using all inlying matches (and not only the initial eight matches of the sample) to re-compute a new estimation of F. Then, a search for new matching points is performed based on the distance of the points to the corresponding epipolar lines. Typically these two steps are iterated until the number of inliers converge [7]. In this paper, we propose an alternative solution to this refinement algorithm that considerably improves the accuracy of the F matrix.

3. Fusing distance and appearance metrics for estimating F

When estimating the fundamental matrix inside a Sample Consensus framework, a cost function must be used for performing the selection of the best estimation of F. Standard cost functions typically involve maximizing the number of inliers that fits in the estimated epipolar geometry [4, 5, 22], or minimizing the Sampson distance [21, 23], which is a first-order approximation of the geometric error [7]. For instance, the RANSAC algorithm can be seen as an optimization method that maximizes the following cost function [19]:

$$\mathcal{C}(\epsilon) = \begin{cases} 1 & \text{if } \epsilon < t \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where ϵ is the Sampson (or other proper) distance [7], and t for the inlier/outlier decision threshold. Despite the fact that different points might have different geometric errors, this is a binary decision that assumes that all points verifying the selection criteria contribute equally for maximizing $\mathcal{C}(\cdot)$.

3.1. Putting distance and appearance in $\mathcal{C}(\cdot)$

In this paper we argue that different distance magnitudes can provide different weights for the solution of F, and that using distance weighting functions with appearance metrics can be used for improving the estimation of the fundamental matrix, which we confirm later in the experimental evaluation. We start this section by analyzing several distance weighting functions, and by providing an intuitive explanation of how each function affects the estimation of F.

3.1.1 Analysis of different distance weighting functions

In this section we analyze three different distance weighting functions. For maximizing $\mathcal{C}(\epsilon)$ we select the following weighting functions for the Sampson distance:

- Inverse Weighting Func.: IWF $= {}^{1}/_{1+\epsilon^n}$ (3)
- Bell-shaped Weight. Func.: BSWF $= {}^{1}/_{1+({}^{|\epsilon|}/_{a})^{2n}}$ (4)
- Exponential Weight. Func.: EWF $= \exp(-k\epsilon^n)$ (5)

The IWF has only one parameter, which controls the decay of the weights with the distance. As we increase n, more weight is given to points with low geometric error. Points with lower localization precision will probably be rejected during the estimation of F. The BSWF can be seen as a generalization of the previous IWF. This weighting function has two parameters: a controls the aperture of the function (weight given to larger distance errors), while n controls the decay of the weighting. If we increase simultaneously (a, n) the BSWF allows to select inliers that do not accurately fit the estimated epipolar geometry, which can be relevant for applications that need a large number of matches instead of an accurate fundamental matrix estimation [1]. Finally, the EWF weights the distance using a decaying exponential with two parameters (k, n). The test of the two parameters allows to observe that k controls the aperture and n controls the decay of the weighting function. Apparently, this function seems to provide the best trade-off between an accurate estimation of F and the number of inliers provided.

Let us consider that two different matches have a similar geometric error. Is it possible to further distinguish the weight between the points? Appearance matching information can be included in the cost function to accomplish this task.

3.1.2 Including appearance in the cost function

Typically the matching information between two sets of features is disregarded during the estimation of the fundamental matrix, being only used for establishing the set of putative correspondences. However, it can provide a reliable metric to distinguish between higher and lower confidence matches. Including an appearance weighting term in the cost function can be accomplished as follows:

$$\mathcal{C}(\epsilon, \mathcal{A}) = \mathcal{C}_\epsilon \cdot \mathcal{C}_\mathcal{A}. \tag{6}$$

$\mathcal{C}_\epsilon$ can be one of the functions plotted in Fig. 1 and for $\mathcal{C}_\mathcal{A}$ we adopt the simple normalized cross-correlation with local pixel intensities:

$$\mathcal{C}_\mathcal{A} = \frac{\sum_\mathbf{x} \mathtt{T}(\mathbf{x})\mathtt{T}'(\mathbf{x}+\mathbf{u})}{\sqrt{\sum_\mathbf{x} \mathtt{T}(\mathbf{x})^2 \sum_\mathbf{x} \mathtt{T}'(\mathbf{x}+\mathbf{u})^2}}, \tag{7}$$

with $\mathtt{T}$ and $\mathtt{T}'$ denoting two local image patches. Note however, that our cost function can be used with more powerful descriptors, such as SIFT or GLOH, by using a simple decaying exponential (like the EWF) of the Euclidean distance between the descriptors. Including this cost function in an existing algorithm is a straightforward task, since the goal is always to compute the solution that maximizes Eq. 6. For RANSAC, we adapt the cost function of Eq. 2 as follows:

$$\mathcal{C}(\epsilon, \mathcal{A}) = \begin{cases} \mathcal{C}_\epsilon \cdot \mathcal{C}_\mathcal{A} & \text{if } \mathcal{C}_\epsilon \cdot \mathcal{C}_\mathcal{A} > t \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

while for a MAPSAC-based algorithm we simply substitute the traditional cost by:

$$\mathcal{C}(\epsilon, \mathcal{A}) = \max(\mathcal{C}_\epsilon \cdot \mathcal{C}_\mathcal{A}, t). \tag{9}$$

Henceforth, and to avoid confusing with the traditional implementation of this methods, we will refer to them as *RANSAC*$_{(\epsilon,\mathcal{A})}$ and *MAPSAC*$_{(\epsilon,\mathcal{A})}$, respectively. Note that Eq. 9 also weights the influence of the outliers during the estimation of F. In theory there is no need to impose such threshold t for the MAPSAC algorithm since we are already weighting the outliers by computing only $\mathcal{C}(\epsilon, \mathcal{A}) = \mathcal{C}_\epsilon \cdot \mathcal{C}_\mathcal{A}$. We only employ the threshold t for classifying the matching points as inliers or outliers. In practice, we do not observe any implication in thresholding the cost of Eq. 9 in terms of computational time, and accuracy of F.

3.2. Search of new matching points for refining F

Up to now, we discussed the new cost function, the aspects that should be taken into account when selecting the distance weighting function, and how to include Eq. 6 in the conventional Sample Consensus frameworks.

Now we present a refinement step for the estimation of F that searches for new correspondent points to improve the accuracy of the method. As input our algorithm takes: a set of candidate correspondences, automatically computed using appearance similarity; and the features detected with any keypoint detector [12, 16] in both images (the number can be different in each image). We use MAPSAC$_{(\epsilon,\mathcal{A})}$ with a few hundreds iterations (about 500) for performing a preliminary estimation of the fundamental matrix. After this fairly good estimation of the model, we only use the image features detected (and corresponding local brightness patches) in both images. A proper selection of new point correspondences is thus based using both local appearance and epipolar geometry constraints through Eq. 6.

One important observation is that it is impossible to find, for each point in one view, a single corresponding point in the other view, since there is no keypoint detector that is fully invariant to perspective changes due to camera motion. This means that nothing can assure that the features detected in one image are detected in the other view. Therefore, only points that verify $\mathcal{C}_\epsilon \cdot \mathcal{C}_\mathcal{A} > t$ (see line 6 of Algorithm 1) are selected as matching points. At each iteration of the method a new estimation of F is performed and a new search for inliers is performed. The algorithm is iterated until no improvement is obtained and convergence to the solution is obtained. The overall procedure for the estimation the fundamental matrix is shown in algorithm 1.

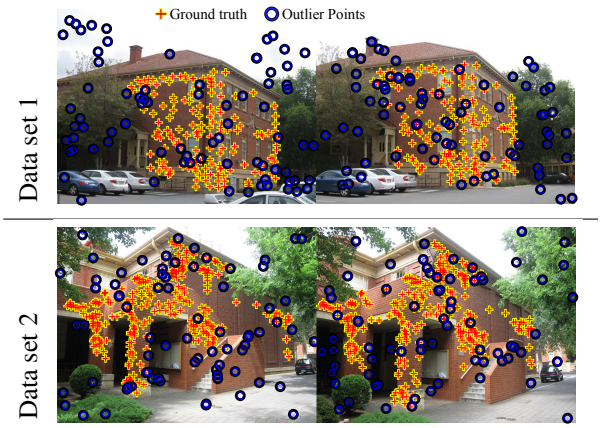
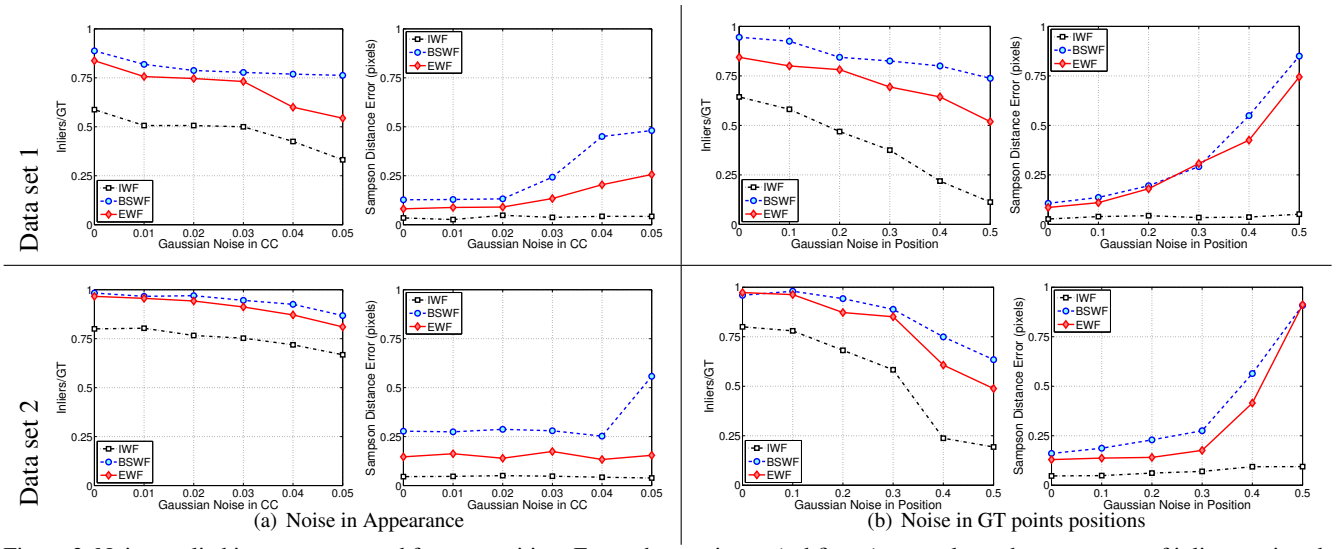


Figure 2. Data sets used under simulation.The cross represent points with a correct correspondence while the dark circles represent outlier points with no correspondence in the other view. The data sets have 33% and 21% of outliers, and 160 and 295 inliers, respectively.

3.3. Simulation Experiments

Herein we evaluate different weighting functions using two stereo image pairs (Fig. 2) with ground truth, using the data sets provided by [24]. We test the 3 distance weighting functions previously described using the following parametrization: (i) IWF - $n = 2$; (ii) BSWF - $(n, a) = (3, 3)$; and (iii) EWF - $(n, k) = (2, 0.1)$. We have experimentally observed that these are the parametrizations



(a) Noise in Appearance (b) Noise in GT points positions

Figure 3. Noise applied in appearance and feature position. For each experiment (subfigure) we evaluate the percentage of inliers retrieved (graphics on the left) and geometric error (graphics on the right).

Algorithm 1: Fundamental Matrix Estimation

Require: Stereo pairs of Images $\mathtt{I}_l$ and $\mathtt{I}_r$
Ensure: Fundamental matrix (F)

1: Compute images keypoints with Harris Corner detector and correspondences using $\mathcal{C}_{\mathcal{A}}(\mathbf{x}, \mathbf{x}')$
2: Store all appearance relations $\mathcal{C}_{\mathcal{A}}(\mathbf{x}, \mathbf{x}')$;
3: Use MAPSAC$_{(\epsilon,\mathcal{A})}$ to obtain F_{init};
4: Initialize $\mathsf{F}_f = \mathsf{F}_{init}$ and ϵ_f as the average Sampson distance;
5: **while** ϵ_f do not converge **do**
6: Search $\mathbf{x} \leftrightarrow \mathbf{x}'$ pairs based on $\mathcal{C}_\epsilon \cdot \mathcal{C}_{\mathcal{A}}$ constraints
7: Solve nonlinear least-squares for

$$\mathbf{x}'^{\mathsf{T}}\mathsf{F}\mathbf{x} \text{ s.t. } \max \mathcal{C}(\epsilon, \mathcal{A})$$

8: Evaluate solution using the Sampson distance ϵ_c
9: **if** $\mathsf{mean}(\epsilon_c) \leq \epsilon_f$ **then**
10: $\epsilon_f = \mathsf{mean}(\epsilon_c)$;
11: $\mathsf{F}_f = \mathsf{F}$;
12: **end if**
13: **end while**

that provide the best ratio of correct matches/geometric error for each weighting function.

Since our cost function encodes both local appearance and geometric error, two different experiments were conducted. In Fig. 3(a) we added Gaussian noise with an increasing standard deviation, up to $\sigma = 0.05$, to the local image patches using the Matlab *imnoise()* function. We selected this maximum value for σ since it was used in relevant image retrieval studies [8]. In the experiment of Fig. 3(b) we added noise to the keypoint GT position up to $\sigma = 0.5$. For each experiments, we used as evaluation metric the $^{Inliers}/_{GTpoints}$ ratio (graphics on the left) and Sampson distance (graphics on the right).

From the analysis of Figs. 3(a) and 3(b), we can observe that, as we expected, IWF provides the estimation of the fundamental matrix with lower geometric error, at the expense of a considerable lost in the inlier points/GT points ratio. The BSWF clearly outperforms the two other distance cost functions in terms of the number of inliers retrieved, at the expense of providing the less accurate estimation of F. The EWF-version of our cost function seems to merge the strengths of IWF and BSWF, providing the best trade-off between number of inliers/geometric error.

In the sequel of the paper we evaluate our algorithm using both the BSWF- and EWF-versions of the proposed function. We aim to show that our algorithm it able to find more correspondent points than the standard methods for computing F while keeping the estimation very accurate.

4. Experimental Validation with Real Imagery

The previous section concerned the study of a new cost function to be used in a Sample Consensus framework for estimating the fundamental matrix. We also describe a new refinement step that relies on all the detected keypoints, and not only on the putative matches. The proposed cost function and refinement algorithm were evaluated under simulation, which allowed to draw preliminary conclusions about the advantages and drawbacks of each distance weighting function. Now, we focus on the evaluation of the proposed metric and algorithm in a set of real imagery experiments, where no ground truth correspondences are available.

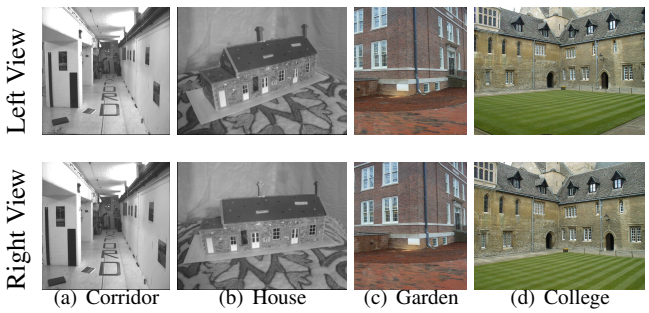


Figure 4. Data sets used in real experiments. Note that the data sets contain repetitive patterns, which difficult the task of matching using simply local image intensity values.

4.1. Benchmarked Experiments of [2]

In this section we evaluate the proposed RANSAC/MAPSAC$_{(\epsilon,\mathcal{A})}$ with the EWF- and BSWF-versions of the proposed cost function against the common implementations [7, 21]. We use the fundamental matrix toolbox used in [2][2]. For a fair comparison with our refinement algorithm described of section 3.2, we also include a standard guided-matching procedure [7] where further interest point correspondences are determined using the estimated epipolar geometry, with F being re-estimated using a non-linear procedure as decribed in [7]. We call this method *MAPSAC+Refine MIN DIST*.

For the image interest point detection we use our implementation of the Harris corner detector. Like in the simulation experiments, we just rely on pixel intensity values to encode the local appearance of the image features, with the similarity between feature patches being measured using normalized cross-correlation. As discussed earlier, using a scale invariant detector/descriptor [12, 15, 16] is quite straightforward with possible positive impact in more difficult situations, such as severe viewpoint.

The data set used for the evaluation can be seen in Fig. 4. It comprises a set of difficult situations, such as scale changes, wide-baseline and repetitive patterns. We adopt the benchmark evaluation criteria proposed in [2]: (i) the error in the fundamental matrix estimative using Sampson distance (Fig. 5(a)) [2, 7]; (ii) number of inliers/correct correspondences provided by each method (see Fig. 5(b)); and (iii) the computational time that each method requires to find a plausible solution (see Fig. 5(c)).

4.1.1 Results and their discussion

We will start the analysis by comparing the performance of the common RANSAC/MAPSAC implementations with the RANSAC/MAPSAC$_{(\epsilon,\mathcal{A})}$. The binary decision rule typically used in RANSAC does not provide a proper weighting for the final F solution. This is visible only by making a comparison between RANSAC and MAPSAC that employs different weights for different keypoint distances. Including our cost function in these algorithms clearly brings benefits in terms of the geometric error of the estimation and of the computational time required for estimating F. This corroborates our argument that different keypoint geometric error must be weighted differently, and that appearance cues should be used inside the cost function for better distinguish between high and lower confidence correspondences.

The inclusion of the refinement step *Refine MIN DIST* [7] as a complement to the MAPSAC framework brings clear benefits in terms of the geometric error obtained and in terms of the number of correspondences. We can also observe that our cost function outperforms this state-of-the-art algorithm in every evaluation criteria. In fact, our metric has one important advantage when compared to the distance-based cost function: **wrong** matches that **verify the epipolar geometry** are discarded if they present low correlation scores. Due this, the estimation of the fundamental matrix is very accurate at the first refinement steps, which permits to robustly incorporate new matching points without ruining the estimation in the subsequent steps.

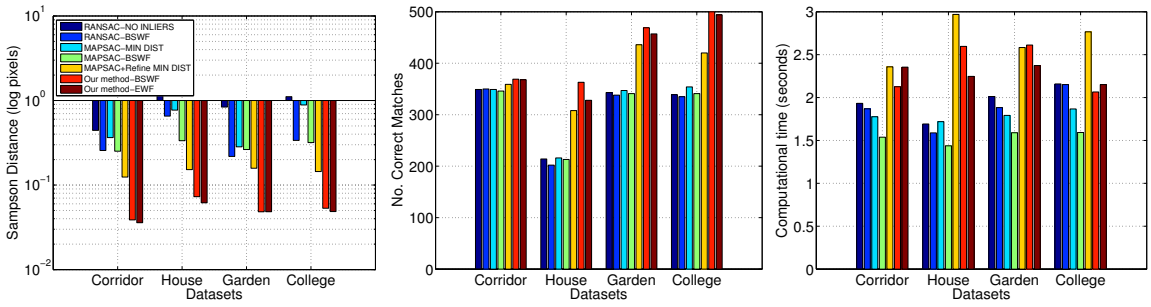
For all the cases tested, we have seen that the standard state-of-the-art methods that rely solely on point correspondences tend to provide a faster solution than the method herein presented (Fig. 5(c)). This was expected since our method requires an initial estimative of F for searching all possible correspondences using Eq. 6. Nevertheless, our method tends to be faster than its direct competitor (MAPSAC+*Refine MIN DIST*), while providing a higher number of inliers and a better estimation of the fundamental matrix. The faster convergence of our method happens mainly for two reasons: (i) the local matches scores are computed offline and stored into memory, i.e. they are not re-computed at every iteration of the method; (ii) since we provide more confidence to highly reliable correspondences based on geometric and appearance constraints the method tends to converge faster to the correct solution of the fundamental matrix.

4.2. Structure-from-Motion Experiments

In this section we perform a set of SfM experiments to access the usablity of our algorithm in this type of applications. We consider 2 data sets with 5 images each (two images of each sequence can be seen in Fig. 6). Every possible pair combination is performed, which totalizes 10 image pairs per sequence. We use the same algorithm as before for extracting and describe the local image keypoints.

The relative motion estimation is performed using the epipolar geometry. We start by estimating the fundamen-

[2]This toolbox provides the Matlab implementation of the methods supplied by the correspondent authors.



(a) Sampson distance (log pixels) (b) Number of correct matches (c) Computational time

Figure 5. Real Experiments. (a) compares the geometric error between the evaluated methods (note that the results are in logarithmic scale); (b) the number of inliers provided by each method, and (c) the computational time required by each method for selecting the best F. The results were average over 100 runs of each method.

Figure 6. Datasets used in SfM experiments. The average number of initial correct matches obtained with MAPSAC for each sequence are of 55.9 ± 7.8 for **DS 1** and 309 ± 12.7 for **DS 2**.

tal matrix F. Afterwards, we compute the essential matrix using $\mathsf{E} = \mathsf{K}^{\mathsf{T}}\mathsf{F}\mathsf{K}$, where K represent the camera intrinsics. Finally, E is factorized to obtain the relative displacement between frames [13].

In this set of experiments we only consider the EWF-version of our algorithm, which provide the best trade-off between inliers/geometric error, and the MAPSAC+*Refine MIN DIST*, with the results being averaged over 50 trials of each method. For the sake of a fair comparison, we initialize both methods using MAPSAC with 500 iterations for computing the initial estimate of the fundamental matrix F_{init}.

4.2.1 Results and their discussion

Table 1 shows the experiment comparison between the two refinement algorithms in SfM tasks. As is typical in these type of situations we evaluate the re-projection error as a measure of the quality of the estimated fundamental matrix. It can be seen that our re-projection error is significantly lower than that of *Refine MIN DIST*, which is a consequence of the highly quality estimation of the F matrix. Also important is the fact that our methods returns a higher number of inliers points, which is relevant if detailed 3D reconstructions are required.

It is also interesting to observe that for the **DS1** the standard *Refine MIN DIST* approach presents two times more variation in terms of inliers retrieved than our method. This means that using only distance in the refinement steps is more sensitive to a noisy initialization then our algorithm. Since our algorithm is more robust to a noisy initialization, it is capable of robustly incorporate new inliers without ruining the estimation of the refined F matrix.

5. Conclusions

In this paper we present a new cost function that fuses appearance and geometric error. An intuitive analysis of the role of each component is provided, with clear benefits arising when including this cost function in standard estimation frameworks, such as RANSAC or MAPSAC. We also propose and validate a new refinement algorithm, that takes as input a set of correspond-less keypoints and an initial estimation of the fundamental matrix. The proposed metric and algorithm are evaluated using simulation and real experiments, with clear benefits in terms of geometric error and the number of inliers retrieved.

Finally, it is important to discuss how to select the best distance weighting function concerning the type of high-level vision application. For image retrieval/matching applications [1], the BSWF is clearly the best choice since it provides a higher number of correspondences. On the other hand, in a motion recovery scenario [11] an accurate estimation of the epipolar geometry is preferred, with the best choice for this case being the IWF. For 3D reconstruction, where the accuracy of the fundamental matrix estimation is as important as the number of inliers retrieved, we advise to use the EWF-version of our cost function.

Table 1. Evaluation in Structure from Motion scenario. The last three columns compare the two refinement algorithms in terms of number of 3D points reconstructed, re-projection error (in pixels) and computation time (in seconds).

	Method	Inliers F	Rep. Err.	Time
DS 1	*Refine MIN DIST*	335 ± 34.2	1.9 ± 1.2	3.5 ± 1.1
	Ours	402 ± 16.9	0.8 ± 0.6	2.5 ± 0.8
DS 2	*Refine MIN DIST*	575 ± 15.4	1.5 ± 1.1	2.7 ± 0.6
	Ours	648 ± 12.1	0.62 ± 0.45	2.1 ± 0.9

As future work, we will investigate the integration of more powerful image description schemes in the proposed cost function, and its practical implications in image sequences with large viewpoint variations.

Acknowledgments

The authors gratefully acknowledge the Portuguese Science Foundation (FCT) that generously funded this work through grants PTDC/EIA-CCO/109120/2008 and SFRH/BD/63118/2009.

References

[1] R. Arandjelović and A. Zisserman. Efficient image retrieval for 3D structures. In *BMVC*, pages 30.1–30.11, 2010. 1, 3, 7

[2] X. Armangue and J. Salvi. Overall view regarding fundamental matrix estimation. *Image and Vision Computing*, 21:205–220, 2003. 1, 2, 6

[3] O. Chum and J. Matas. Matching with PROSAC - Progressive Sample Consensus. In *IEEE-CVPR*, pages 220–226, 2005. 1, 2

[4] R. Den Hollander and A. Hanjalic. A Combined RANSAC-Hough Transform Algorithm for Fundamental Matrix Estimation. In *BMVC*, pages 41.1–41.10, 2007. 1, 2, 3

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981. 1, 2, 3

[6] R. Hartley. In defense of the eight-point algorithm. *IEEE-TPAMI*, 19(6):580 –593, 1997. 2

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2nd edition, 2004. 1, 3, 6

[8] Y. Ke and R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *IEEE-CVPR*, pages 506–513, 2004. 5

[9] J. P. Lewis. Fast normalized cross-correlation. In *Vision Interface*, pages 120–123, 1995. 1

[10] Longuet. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981. 1

[11] M. Lourenco, J. P. Barreto, and F. Vasconcelos. sRD-SIFT Keypoint Detection and Matching in Images with Radial Distortion. *IEEE-TRO*, 2012. 7

[12] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2):91–110, 2004. 1, 4, 6

[13] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3D Vision: From Images to Geometric Models*. Interdisciplinary Applied Mathematics. Springer-Verlag, 2003. 7

[14] A. Méler, M. Decrouez, and J. Crowley. BetaSAC: A New Conditional Sampling For RANSAC. In *BMVC*, 2010. 1, 2

[15] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE-TPAMI*, 27(10):1615–1630, 2005. 1, 6

[16] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A Comparison of Affine Region Detectors. *IJCV*, 65:43–72, 2005. 1, 4, 6

[17] P. Newman, D. Cole, and K. Ho. Outdoor slam using visual appearance and laser ranging. In *IEEE-ICRA*, pages 1180 –1187, 2006. 1

[18] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the World from Internet Photo Collections. *IJCV*, 80(2):189–210, Nov. 2008. 1

[19] D. Tegolo and F. Bellavia. noRANSAC for fundamental matrix estimation. In *BMVC*, pages 1–11, 2011. 1, 2, 3

[20] B. Tordoff and D. W. Murray. Guided sampling and consensus for motion estimation. In *ECCV*. Springer-Verlag, 2002. 2

[21] P. H. S. Torr. Bayesian Model Estimation and Selection for Epipolar Geometry and Generic Manifold Fitting. Technical report, Microsoft Research, 2002. 1, 2, 3, 6

[22] P. H. S. Torr and D. W. Murray. The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *IJCV*, 24(3):271–300, 1997. 1, 2, 3

[23] P. H. S. Torr and A. Zisserman. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry . *Comput. Vis. Image. Unders.*, 78(1):138–156, 2000. 2, 3

[24] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter. Dynamic and hierarchical multi-structure geometric model fitting. In *IEEE-ICCV*, pages 1044–1051, 2011. 4