

FUNDAMENTAL MATRIX ESTIMATION WITHOUT PRIOR MATCH

Nicolas Noury

Frédéric Sur

Marie-Odile Berger

INRIA Lorraine / INPL / LORIA
LORIA, Campus Scientifique - BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France
{noury, sur, berger}@loria.fr

ABSTRACT

This paper presents a probabilistic framework for computing correspondences and fundamental matrix in the structure from motion problem. Inspired by Moisan and Stival [1], we suggest using an *a contrario* model, which is a good answer to threshold problems in the robust filtering context. Contrary to most existing algorithms where perceptual correspondence setting and geometry evaluation are independent steps, the proposed algorithm is an all-in-one approach. We show that it is robust to repeated patterns which are usually difficult to unambiguously match and thus raise many problems in the fundamental matrix estimation.

Index Terms— Fundamental matrix, probabilistic model, repeated patterns.

1. INTRODUCTION

Structure from motion analysis in video streams aims at estimating the camera pose and the 3D structure of the scene. Evaluating the fundamental matrix between two views, as explained by Hartley and Zisserman [2], is often one of the very first steps. Most algorithms proceed with previously matched points to estimate the fundamental matrix. Unfortunately, these algorithms are very sensitive to incorrectly tracked features and robust filtering methods such as RANSAC [3] have to be used to select coherent matches.

1.1. RANSAC: limitations and solutions

RANSAC-like methods involve several thresholds to choose whether a match is correct with respect to the estimated model or not. Setting these thresholds is quite touchy, since they depend on many features such as scene contrast, image quality, matching accuracy...

Statistical approaches attempt to address these difficulties. For instance, instead of just classifying point matches between outliers and inliers by thresholding the size of the consensus set, Torr and Zisserman [4] maximise the likelihood of the consensus set in a probabilistic model that involves the precision of the inliers. To get rid of parameters, Moisan and Stival [1] have proposed an *a contrario* model (first introduced in a series of articles by Desolneux, Moisan and Morel, see e.g. [5]), dedicated to the estimation of the fundamental matrix, that outperforms classical RANSAC.

1.2. The challenge of ambiguous correspondences

In these algorithms, correspondences are generated by using local patches similarities. They are therefore unreliable (from the epipolar geometry point of view) when confronted with repeated patterns, and the fundamental matrix estimation is definitely spoiled by these wrong matches. Moreover, RANSAC-based algorithms do not pay

attention to the quality of the local patch similarity. Both geometric and similarity constraints should be used to select those correct matches on which the matrix estimation is performed. Domke and Aloimonos [6] present a solution which consists in establishing an *a priori* model for the distributions of correspondences, computed for every image pixels. The drawback of this approach is its heavy computational cost. Dellaert et al. [7] propose a promising EM algorithm to assign correspondences between points of interest, but do not take into account local patches.

1.3. Contributions

While Moisan and Stival's model requires a set of matched points between the views, we present here an *a contrario* model dedicated to the case where correspondences are unknown. It allows us to simultaneously match interest points and to compute the fundamental matrix while weighting the role of the points with respect to the quality of their matching. Our contribution consists in setting up a model which has been sketched out in [1], specifying its implementation and in validating it. Contrary to [6], we keep a reasonable computational burden. We obtain better results than state-of-the-art methods in scenes featuring repeated patterns such as textures. Experiments prove that the number of inliers noticeably increases, as well as the precision of the fundamental matrix estimation.

2. AN A CONTRARIO MODEL FOR FUNDAMENTAL MATRIX ESTIMATION WITHOUT PRIOR KNOWLEDGE

This section is a digest of the theory leading to the proposed *a contrario* model. We take our inspiration from [1].

2.1. Motivations and notations

The problem at hand is to estimate the fundamental matrix between two views, based on two lists of points of interest (POI, given with a local descriptor), one list for each of the two images. Let us recall that matches between POI are here unknown and of course have to fulfil the epipolar constraints.

Let us give some notations. A collection S_1 (resp. S_2) of POI is given in image 1 (resp. image 2), $des(m)$ is a local descriptor of a small region around POI m , and a match between POI is a pair (m_1, m_2) , where $m_1 \in S_1$ and $m_2 \in S_2$. In what follows, m_i denotes a POI as well as its homogeneous coordinates.

Suppose that image 1 and image 2 are linked through a fundamental matrix F . Constraints on matches (m_1, m_2) between POI are twofold: 1) the epipolar line Fm_1 is near m_2 and vice versa, and 2) descriptors $des(m_1)$ and $des(m_2)$ are alike (up to the geometry and the possible colorimetry change between the two views.) On the one hand, constraint 1 alone could enable a match between two POI

that satisfy the epipolar constraint “by chance” but actually do not correspond to the same 3D point. On the other hand, constraint 2 reinforces the possibility that m_1 and m_2 actually correspond to the same 3D point.

Let us introduce a probability law for each constraint:

1) $\frac{2D'}{A'} d(m_2, Fm_1)$ where A' is the area of image 2 domain, D' its diameter, and d the euclidean distance (this definition can be easily made symmetric.) Simple geometrical reasoning shows that this is the probability law of $d(m_2, Fm_1)$ when the m_2 's are supposed to be uniformly distributed.

2) $P(c) = \Pr(\delta(\text{des}(m_1), \text{des}(m_2)) < c)$, where δ is some metric between local descriptors and $P(c)$ is simply estimated by counting over the whole set of candidate matches (a subset of $S_1 \times S_2$).

2.2. A contrario model

Let us now consider a set $S \subset S_1 \times S_2$ (cardinality k) of pairs of POI (m_1^i, m_2^i) , and suppose that the fundamental matrix F has been estimated upon 7 pairs among S . Let us also denote

$$\alpha = \max_{i \in \{1, k\}} 2D'/A' \cdot d(m_2^i, Fm_1^i) \quad (1)$$

and

$$c = \max_{i \in \{1, k\}} \delta(\text{des}(m_1^i), \text{des}(m_2^i)). \quad (2)$$

Probabilities in the *a contrario* model \mathcal{M} are defined by the following assumption. The $k - 7$ points that are not involved in the computation of F are independent, as well as the k descriptors, and the distances between POI and descriptors are supposed to follow the probability laws above. This is called an *a contrario* model because if S is made of pairs of POI that are true matches (that is, matches between two POI corresponding to the same 3D point), these assumptions are surely not valid.

We then compute the probability of S under \mathcal{M} as

$$\begin{aligned} P_{\mathcal{M}}(S|F) &:= P_{\mathcal{M}}\left(\forall i \in \{1, \dots, k\}, \frac{2D'}{A'} d(m_2^i, Fm_1^i) \leq \alpha \right. \\ &\quad \left. \text{and } \delta(\text{des}(m_1^i), \text{des}(m_2^i)) \leq c\right) \\ &= \alpha^{k-7} P(c)^k. \end{aligned}$$

The lower $P_{\mathcal{M}}(S|F)$, the less likely it is that S is made of POI that match “just by chance”. A better explanation is that they obey the 3D geometry.

2.3. Meaningfulness of a group

Since the probability above does not permit to easily compare groups of matches with different cardinality, the following notion is introduced (inspired by Prop. 2 in [1]):

Definition 1. S is ε -meaningful if

$$\varepsilon_S(\alpha, c, k, n) := 3(n-7) \binom{n}{k} \binom{k}{7} \alpha^{k-7} P(c)^k \leq \varepsilon,$$

with α, c, k defined as above, and n the cardinality of the subset of $S_1 \times S_2$ in which reliable groups of matches are sought.

There are $\binom{n}{k}$ such S groups, $(n-7)$ choices for k (which values are between 8 and n), $\binom{k}{7}$ choices for the 7 pairs to estimate the fundamental matrix, and each 7-uplet gives (up to) 3 fundamental

matrices. There are in the end $3(n-7) \binom{n}{k} \binom{k}{7}$ fundamental matrices to test while enumerating the subsets S .

Therefore, ε_S is an upper bound to the expected number of groups like S which could be observed according to the *a contrario* model: if this expectation is lower than 1, we can be sure that the considered set is not observed by chance and assume having found a solution. ε_S makes a compromise between the quality of individual matches (measured by α and c) and the number of these matches that are involved in the group S .

We are thus interested in those S that minimise ε_S . Since a comprehensive scanning is out of the question, a heuristic search is called for.

The proposed criterion is a generalization of the one proposed by Moisan and Stival. Their algorithm, called ORSA (for Optimized Random Sampling Algorithm), selects reliable correspondences among already known matches by using only the geometric criterion α . They suggest taking into account the colorimetric distance c but intend to identify matches over all possible correspondences (that is to say over $S_1 \times S_2$), which is simply untractable.

3. IMPLEMENTATION CHOICES

We specify here several pending points: the local descriptors, the distance between descriptors, the probability distribution $P(c)$, and finally the proposed algorithm scheme aiming to find a correspondence set that minimises ε_S .

3.1. Feature extraction and similarity measurement

We use the SIFT extractor from D. Lowe [8] for finding interest points and computing local patch descriptors. These descriptors are just rotation-scale invariant, which we consider to be a good enough local approximation. The resemblance between descriptors is difficult to measure. Finding a distance between two descriptors consists here in computing a histogram distance. This subject has been studied in [9]. A compromise between speed and reliability can be achieved with the χ^2 distance, that we use.

We then evaluate the probability $P(c)$ as an empirical cumulative distribution function, estimated over the whole set of possible point matches. The study of the distance distribution between patches shows us that correct point matches have a rather small distance compared to others. This characteristic allows us to actually consider only a subset among the matching candidates by setting a rough threshold and discarding the unrelated patches. This cut only reduces the problem complexity and cannot really be considered as a supplementary threshold. The important selection phase has yet to be done.

3.2. Algorithm

Our algorithm follows the random sampling consensus scheme: we first select a random sample of minimal size in order to evaluate a fundamental matrix over it. Then, we build a meaningful set, containing correspondences which fit the model computed on the sample. Since we use the so-called 7-points algorithm to evaluate the fundamental matrices, random samples are made of 7 matched pairs and lead to 3 matrices. Thus for each sample, up to 3 meaningful sets have to be built. Let us explain how to select the random sample and build meaningful sets.

No matched couples are provided, we just have lists of points of interest (surrounded by local descriptors). Therefore, we build groups of putative matches as explained in 3.1. We then choose

a random sample by selecting 7 groups, and randomly one correspondence candidate for each group. We call this notion a *group of candidates*, that is, group linking an interest point in the first view with all possible correspondences in the second view: their number is limited by the rough cut made on descriptor distances.

Since we cannot consider an exhaustive search among all the possible subsets, which is the only way to find the most meaningful subset that we are interested in, we present a heuristic method. In order to compute the meaningful set of correspondences, we have to calculate α (defined in equation 1) for the points in correspondence outside the sample. At most one correspondence is valid in each group, so at most one correspondence should be used. We select the correspondence with the least epipolar distance, and thus the least α , among the group of candidates. The selected correspondences are ordered by increasing α . We thus obtain nested subsets of size k , with k between 8 and n , n being the number of *groups of candidates*. We associate to each one of them the maximum value $P(c)$. For each subset of size k , we are now able to compute the meaningfulness ε_S . The set with the minimum ε_S is the most meaningful one.

An advantage of the *a contrario* model is the meaning of ε_S : if $\varepsilon_S < 1$, we can be sure that the considered set is not observed by chance. It leads to an interesting optimization step. Once we obtain a set with $\varepsilon_S < 1$, the sample is chosen inside the set, and not among the whole groups. For each group chosen for the sample evaluation, we keep the last selected match instead of randomly sampling among the candidates in the group. The algorithm follows:

- extract interest points (SIFT)
- compute the distribution $P(c)$
- build the *groups of candidates*
- Let $\bar{\varepsilon} = +\infty$, repeat
 1. choose a sample: a set T of 7 groups, and one correspondence for each group (among all groups at first, and among previously found meaningful set during the optimization step)
 2. determine the most meaningful set U associated to each fundamental matrix F corresponding to T :
 - compute α for each group
 - sort groups with α increasing
 - choose the maximum $P(c)$ for each set with increasing α
 - compute ε_S and choose the subset minimising it.
 3. if $\varepsilon_S(U) < \bar{\varepsilon}$, then $\bar{\varepsilon} = \varepsilon_S(U)$ and $\bar{S} = U$
 4. if $\varepsilon_S(U) < 1$, enter the optimization step.

until the number of iteration is greater than N (N_{opt} in the optimization step, with $N_{opt} = \frac{N}{10}$)

- return \bar{S} and $\bar{\varepsilon}$.

4. EXPERIMENTAL RESULTS

In this section we show that the information brought by local patches enables a better selection among the point matches, especially when confronted with repeated patterns. Indeed, the classical matching methods simply threshold the distance between descriptors. It therefore furnishes lots of mismatched points in the case of repetitions, that spoil the fundamental matrix estimation.

Tests are conducted on three types of image pairs. The proposed algorithm is compared with Moisan and Stival’s ORSA (see section 2.3) feeded with correspondences that are brought by the

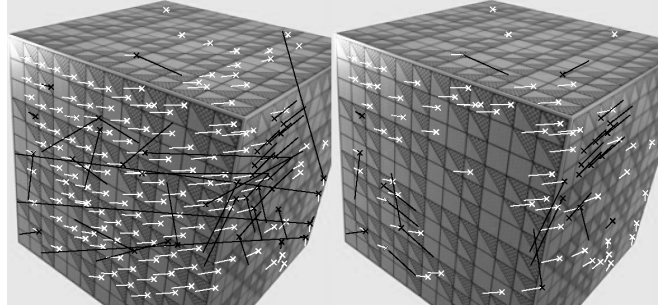


Fig. 1. Synthetic case: crosses show interest points, and straight lines their apparent displacement between two views. Detected inliers are in white, and outliers in black. In the proposed approach (on the left), we do not show every filtered outliers for visibility reasons. We get twice as much inlying correspondences with the proposed method as with ORSA. The estimated fundamental matrix is also more precise (the average epipolar distance is 0.2 pixels vs. 0.7), and computation time does not exceed one second.

	ORSA		proposed algo.	
	matches	accu.	matches	accu.
synthetic cube	97	0.73	193	0.19
real cube	247	0.19	371	0.15
Union Station	304	2.52	316	0.55
Loria	568	0.66	702	0.59

Table 1. Comparison of the two algorithms, showing the number of inlying correspondences and the average distance between inliers and their epipolar line (measured in pixels.)

popular SIFT matching procedure by Lowe. First, our algorithm is evaluated on computer generated views, then in a laboratory experiment, and finally on real-life pictures.

Figure 1 shows the synthetic case. We have generated 400×300 synthetic images with Blender software. See caption for comments.

Figure 2 and 3 show the “laboratory experiment” on a *real cube* pair from a 320×240 video acquired via a digital camera. Correspondences given by ORSA can be seen in figure 2. We get more correspondences using our algorithm (see figure 3), since we do not *a priori* throw away ambiguous feature points. The number of correspondences increases by about a factor 2 in areas containing repeated patterns.

Results of experiments led on pairs of real-life pictures and the two previous pairs are summarized in table 1. *Union Station* (see figure 4) and *Loria* are respectively 640×480 and 720×576 pairs of images. This confirm that both the accuracy (illustrated with *Union Station* details in figure 5) and the number of inliers increase.

5. CONCLUSION AND PERSPECTIVES

In this article, we have introduced an algorithm that allows us to simultaneously match local descriptors and estimate the fundamental matrix. The cornerstone of this approach is the *a contrario* model combining geometrical conditions and local similarities in images. We obtain a significant gain in accuracy, with generally most inlying correspondences, especially when confronted with repeated structures.

Nevertheless, the result quality depends on local descriptors invariance. As SIFT descriptors are only rotation-scale invariant, we

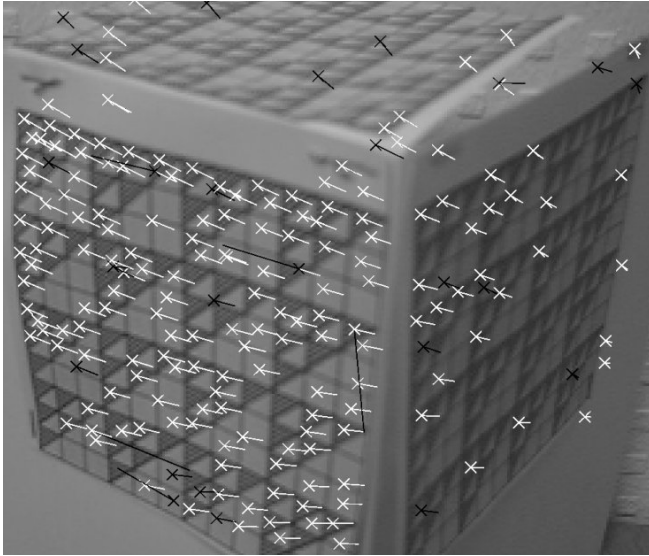


Fig. 2. ORSA case: correspondences in presence of repeated patterns. The front side has not as many matches as one could expect.

intend to study other descriptors. Let us note that our algorithm only needs a metric between descriptors, and is independent from their building process. We also envisage using preemptive tests from Nistér [10] to improve the real-time response and use our technique in tracking problems.

6. REFERENCES

- [1] L. Moisan and B. Stival, "A Probabilistic Criterion to Detect Rigid Point Matches between Two Images and Estimate the Fundamental Matrix," *IJCV*, vol. 57, no. 3, pp. 201–218, 2004.
- [2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.
- [3] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [4] P. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *CVIU*, vol. 78, pp. 138–156, 2000.
- [5] A. Desolneux, L. Moisan, and J.-M. Morel, "Maximal Meaningful Events and Applications to Image Analysis," *Annals of Statistics*, vol. 31, no. 6, 2003.
- [6] J. Domke and Y. Aloimonos, "A Probabilistic Notion of Correspondence and the Epipolar Constraint," in *Proc. of 3DPVT*, 2006.
- [7] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun, "Structure from motion without correspondence," in *IEEE CVPR*, 2000.
- [8] D. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.
- [9] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *IEEE CVPR*, 2006, vol. 1, pp. 246–253.
- [10] D. Nistér, "Preemptive RANSAC for Live Structure and Motion Estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.

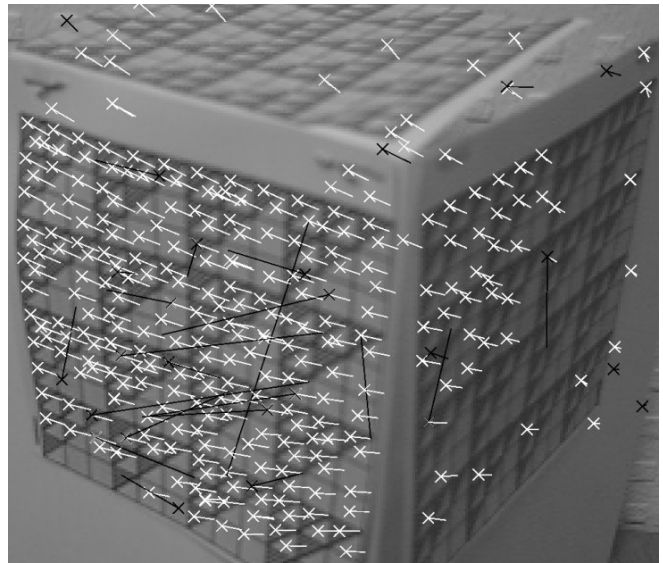


Fig. 3. When adding local patches similarity to the *a contrario* model: twice as much reliable matches are generated. A better fundamental matrix estimation is also obtained.



Fig. 4. The *Union Station* image pair. Notice the repeated structures.

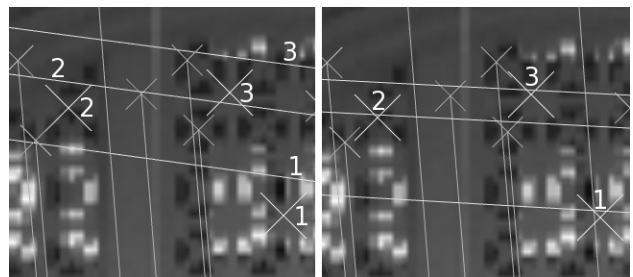


Fig. 5. *Union Station* details. Numbers show the interest points and the associated epipolar lines, on which they should lay. This illustrates the gain in accuracy with the proposed algorithm (on the right) towards ORSA (on the left).