

Robust matching in an uncertain world

Frédéric Sur

► **To cite this version:**

Frédéric Sur. Robust matching in an uncertain world. 20th International Conference on Pattern Recognition - ICPR 2010, Aug 2010, Istanbul, Turkey. IEEE, pp.2350 - 2353, 2010, ICPR 2010 - 20th International Conference on Pattern Recognition. <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5597003>. <10.1109/ICPR.2010.575>. <inria-00474083>

HAL Id: inria-00474083

<https://hal.inria.fr/inria-00474083>

Submitted on 20 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust matching in an uncertain world

Frédéric SUR

*École des Mines de Nancy (INPL) & LORIA (Magrit team, INRIA)
Campus Scientifique BP 239 - 54506 Vandœuvre-lès-Nancy cedex - France
frederic.sur@loria.fr*

Abstract

Finding point correspondences which are consistent with a geometric constraint is one of the cornerstones of many computer vision problems. This is a difficult task because of spurious measurements leading to ambiguously matched points and because of uncertainty in point location. In this article we address these problems and propose a new robust algorithm that explicitly takes account of location uncertainty. We propose applications to SIFT matching and 3D data fusion.

1. Motivation and related work

Point of interest (POI) matching is one of the very first steps of many computer vision tasks. The aim is to find point correspondences between several views (2D for a classic camera or possibly 3D for a depth-camera) of the same physical object or scene. 2D or 3D point matching is a prior step to registration, where a transformation between views is estimated from the matching step. A popular way to tackle the matching problem is to proceed in two separated steps: **1)** find out a list of putative correspondences, based on some features associated with POI (e.g. 3D shape descriptors or SIFT [7] in 2D), then **2)** extract from this list the largest possible subset of correspondences that are consistent with a common transformation.

Point matching is an intrinsically difficult problem. It is indeed impossible to match POI at step **1)** without introducing false correspondences (also called *outliers*, as opposed to *inliers*), i.e. correspondences between points that do not actually correspond to the same physical point. Moreover, POI location has limited accuracy. Even with a careful subpixel interpolation, the accuracy of image POI is limited to 0.2 / 0.5 pixel.

Robust algorithms (in the sense that they have to deal with outliers, and possibly to limited accuracy) are thus needed for step **2)**. The most popular choice is certainly RANSAC [5] and numerous methods derived from it. RANSAC is an iterative procedure where candidate transformations are tested over the set of putative corre-

spondences. Incorporating point location uncertainty in a robust matching algorithm is the main topic of several works: from 3-D registration [12] to point matching under projective transformation [2, 10, 14]. Uncertainty is somehow incorporated in MLESAC [16], where inliers location is assumed to be spoiled by an isotropic Gaussian perturbation. In [15], MLESAC is adapted to incorporate the uncertainty on the candidate transformations. In [14] we have incorporated the uncertainty of the fundamental matrix in the parameterless state-of-the-art RANSAC scheme of [9]. Following our work, [13] proposes to speed-up RANSAC by limiting the search for correspondences in the mapped error regions. All of these papers assume that POI location uncertainty is isotropic. Characterizing the uncertainty of POI location is a quite recent problem in computer vision. A recent paper [17] estimates the location uncertainty of the popular SIFT features [7] and demonstrates the soundness of the approach.

Our contribution is a general RANSAC-like robust algorithm, which explicitly takes into account the uncertainty of POI. It is inspired by the *a contrario* model from [9]. The proposed algorithm is general in the sense that it is well suited to 2D or 3D correspondences (contrary to existing methods), with broad potential applications. Still contrary to most existing methods, we do not a priori assume that point uncertainty is isotropic. In particular, the proposed algorithm can be applied for SIFT matching with the uncertainty derived from [17].

2. Gaussian error and Mahalanobis metric

We describe here the point matching problem whatever the dimensionality d and the transformation group (rotations, homographies, etc). We assume that a list of putative correspondences \mathcal{S} made of N couples of d -dimensional points $(x_i, y_i)_{1 \leq i \leq N}$ has been given by some algorithm exploiting feature similarity (step **1)**). The aim of a robust matching algorithm is to separate this list between inliers (for which there exists a mapping A such that $y_i \simeq Ax_i$) and outliers.

The x_i 's (resp. y_i 's) are modeled as random vari-

ables following a Gaussian distribution with mean \bar{x}_i and covariance matrix Σ_{x_i} (resp. \bar{y}_i and Σ_{y_i}). We note $x_i \sim \mathcal{N}(\bar{x}_i, \Sigma_{x_i})$. Let us remark that \bar{x}_i and \bar{y}_i are unknown, we only observe the uncertain POI x_i and y_i .

In the presented framework, A is a parametric mapping, which needs at least p correspondences for its estimation. For example, in the case of planar data ($d = 2$), estimating a homography requires 4 correspondences, and 5 for 3D data. In RANSAC, at each iteration A is estimated from exactly p correspondences. Since point location is uncertain, estimation of A is also uncertain.

The aim of this section is to define a distance between y_i and Ax_i taking account of the uncertainty. We will use the following classic theorem which is the key-stone of most papers handling uncertainty:

Proposition 1 (*propagation property*) *Let v be a random vector in \mathbb{R}^d , $v \sim \mathcal{N}(\bar{v}, \Sigma)$, and $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be an affine mapping such as $f(v) = f(\bar{v}) + A(v - \bar{v})$. Then $f(v) \sim \mathcal{N}(f(\bar{v}), A\Sigma A^T)$.*

When f is not linear, a first order Taylor series approximate gives an estimate of the covariance by replacing A by the Jacobian $J(\bar{v})$ of f at \bar{v} .

2.1. Uncertainty propagation

We estimate here the uncertainty of $A \cdot x$, when A is a random $d \times d$ Gaussian matrix with mean \bar{A} and covariance Σ_A (it is a $d^2 \times d^2$ matrix), while $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$.

Uncertainty of A . A is estimated from p uncorrelated correspondences $(x_1, y_1), \dots, (x_p, y_p)$. $A = (a_{i,j})_{1 \leq i, j \leq d}$ is seen here as the column vector $(a_{1,1}, \dots, a_{1,d}, \dots, a_{d,1}, a_{d,2}, \dots, a_{d,d})^T$.

By Prop. 1, Σ_A can thus be written as:

$$\Sigma_A = J \cdot \Delta(x_i, y_i)_{1 \leq i \leq p} \cdot J^T \quad (1)$$

where Δ is the $2dp \times 2dp$ block-diagonal matrix made of the Σ_{x_i} and Σ_{y_i} , and J is the $d^2 \times 2dp$ Jacobian matrix of the mapping: $(x_1, y_1, \dots, x_p, y_p) \mapsto A$ with respect to the d components of the p points x_i and y_i .

Let us remark that the computation of J has to be adapted to the actual algorithm estimating A from the p correspondences. With Direct Linear Transform [6], A is the solution of a Singular Value Decomposition (SVD), and J then comes from the Jacobian of the SVD [11]. In [14] we have estimated by a closed-form formula the uncertainty of the fundamental matrix obtained by the linear 8 point algorithm and SVD. In this paper, we estimate J by a finite-difference scheme.

Uncertainty of $A(x)$, which is a random variable as A and x . If A is linear, estimating the Jacobian of the bilinear mapping $(A, x) \mapsto A \cdot x$ is straightforward. It is the $d \times (d^2 + d)$ matrix: $\tilde{J} = (D_x \ A)$ where D_x is the $d \times d^2$ matrix $\text{diag}(x^T, x^T \dots x^T)$. Assuming

that A and x are uncorrelated (which holds if x is not used in estimating A), the covariance of $A \cdot x$ can be derived (using Prop. 1) as:

$$\Sigma_{A \cdot x} = D_x \cdot \Sigma_A \cdot D_x^T + A \cdot \Sigma_x \cdot A^T. \quad (2)$$

If A is not linear, the Jacobian matrix is explicitly computed to replace \tilde{J} , and then permits to propagate the covariance in Eq. (2). Calculus is here easily tractable.

2.2. Estimating the Mahalanobis distance

We now define a distance between $y \sim \mathcal{N}(\bar{y}, \Sigma_y)$ and $Ax \sim \mathcal{N}(\bar{A}\bar{x}, \Sigma_{Ax})$, when x and y are corresponding through A . Assuming $\bar{y} = \bar{A}\bar{x}$ and independence between the random variables, then $y - Ax$ has mean 0 and covariance matrix $\Sigma_y + \Sigma_{Ax}$. Assuming the random variables are Gaussian, a popular way to measure the similarity between y and Ax is to use the so-called (squared) Mahalanobis distance [8]:

$$d_M(y, Ax) = (y - Ax)^T (\Sigma_y + \Sigma_{Ax})^{-1} (y - Ax) \quad (3)$$

which follows a χ^2 law with d degrees of freedom (dof).

The independence assumption holds if (x, y) does not intervene in the estimation of A . To the best of our knowledge, such a use of the Mahalanobis distance has been first studied in [12]. The value of this distance is never used in robust matching: for example in [2, 10, 13] it is just used to define a search region once a fixed confidence level is set a priori.

The metric is symmetrized with:

$$d_M^s(x, y) = d_M(y, Ax) + d_M(x, A^{-1}y) \quad (4)$$

which follows a χ^2 law with $2d$ d.o.f. under independence assumption.

How to actually compute d_M^s ? Eq. (2) would give $\Sigma_{A^{-1}y}$, provided $\Sigma_{A^{-1}}$ is known. This latter covariance can be derived from Σ_A . For example, if A is a rotation matrix, then $A^{-1} = A^T$ and $\Sigma_{A^{-1}}$ is just a reordering of the coefficients of Σ_A . If A is a (linear) affine transformation, then A^{-1} is the inverse of A and $\Sigma_{A^{-1}}$ can be estimated from A via the non-linear propagation property. In this case, one just needs the Jacobian matrix of the mapping from A to A^{-1} . Since $d(A^{-1}) = -A^{-1}dAA^{-1}$: (differentiate $AA^{-1} = I$), one has: $\frac{\partial(A^{-1})}{\partial E_{i,j}} = -A^{-1}E_{i,j}A^{-1}$ where $E_{i,j}$ is the canonical base matrix with entries equal to 0 except in position (i, j) which is 1. If A is a homography, A^{-1} is still the inverse of A in homogeneous coordinates, but the normalisation step (here $\|A\|_2 = 1$) must be taken into account in the propagation.

3. A *contrario* RANSAC for uncertain points

One of our contribution is to propose a robust algorithm for uncertain point matching via a method based

on a so-called *a contrario* model. Books [3] and [4] give a survey of *a contrario* models in many computer vision problems. The idea behind *a contrario* models is that independent, structureless random features can produce structured groups only with a very small probability. However, though it is rarely pointed out, it is also the underlying idea of every method using the χ^2 law to set a threshold over the Mahalanobis distance based on the 5% significance: found correspondences are unlikely under independence assumption.

Let us be more specific. As in every RANSAC-like algorithm, a candidate transformation A is first estimated from a minimal set s made of p correspondences. A fitness measure for a subset S of \mathcal{S} containing s is:

$$d_{\max}(A, S, s) = \max_{(x_i, y_i) \in S} d_M^s(x_i, y_i) \quad (5)$$

$$\text{or } d_{\text{sum}}(A, S, s) = \sum_{(x_i, y_i) \in S} d_M^s(x_i, y_i). \quad (6)$$

We now define the null hypothesis \mathcal{H}_0 .

Definition 1 \mathcal{H}_0 : $(x_i, y_i)_{i \in S \setminus s}$ are independent Gaussian random variables, $x_i \sim \mathcal{N}(\bar{x}_i, \Sigma_{x_i})$, $y_i \sim \mathcal{N}(\bar{y}_i, \Sigma_{y_i})$, also independent of the random mapping $A \sim \mathcal{N}(\bar{A}, \Sigma_A)$ such that $\forall i$, $\bar{y}_i = \bar{A}\bar{x}_i$.

Let δ be some positive number. Thanks to independence assumption among points in $S \setminus s$ and between A (estimated over s) and points of $S \setminus s$, one computes, if k denotes the cardinality of S :

$$\Pr(d_{\max}(A, S, s) \leq \delta | \mathcal{H}_0) = \left(f_{\chi_{2d}^2}(\delta) \right)^{k-p} \quad (7)$$

$$\text{and } \Pr(d_{\text{sum}}(A, S, s) \leq \delta | \mathcal{H}_0) = f_{\chi_{2(k-p)d}^2}(\delta) \quad (8)$$

where $f_{\chi_{\alpha}^2}$ is the cumulative distribution function (c.d.f.) of the χ^2 law with α d.o.f. Indeed, the c.d.f. of the max of $(k-p)$ i.i.d. variables is the common c.d.f. to the power $(k-p)$, and the sum of $k-p$ i.i.d. variables following a χ^2 law with $2d$ d.o.f. follows a χ^2 law with $2(k-p)d$ d.o.f. We will use d_{sum} in the sequel; d_{\max} gives similar results.

In the *a contrario* methodology, one does not directly deal with the probabilities but rather with the so-called *Number of False Alarms* (NFA).

Definition 2 A set S of correspondences is ε -meaningful if there exists a mapping A such that:

$$\text{NFA}(A, S) := (N-p) \binom{k}{p} \binom{N}{k} f_{\chi_{2(k-p)d}^2}(\delta) \leq \varepsilon \quad (9)$$

where δ is the observed value of $d_{\text{sum}}(A, S, s)$.

This quantity automatically balances accuracy and number of points in S (ε -meaningful sets with larger k must have smaller δ), while the standard method consisting in thresholding eq. 5 or 6 via a fixed significance rate over-favours large k . Def 2 is motivated by

Proposition 2 The expected number of ε -meaningful set under \mathcal{H}_0 is lower than ε . (see [3, 4])

Remark. Now, from Eq. 2, 3, and 4, one remarks that the larger the uncertainty of A (or A^{-1}), the smaller the Mahalanobis distance. In order to prevent to fall into degenerate cases (for example homographies with points in s aligned), one cannot just minimize the NFA. Our aim is actually twofold: minimizing both the NFA and the uncertainty of A and A^{-1} . We thus decide to bound from above the uncertainty of the tested transformation A when searching for a most meaningful set. This was also remarked in [13].

Algorithm. Iteratively sample p correspondences from \mathcal{S} and derive the corresponding A . If the largest eigenvalue of Σ_A or $\Sigma_{A^{-1}}$ is above some threshold (10.0 in Sec. 4), then go to the next iteration. Otherwise build the subset $S \subset \mathcal{S}$ with the lowest NFA by remarking (as in [9]) that when k is fixed, the subset of cardinality k with the lowest NFA is made of the k correspondences among \mathcal{S} with the smallest d_M^s distances. In the end, return the subset with the lowest NFA.

4. Proof-of-concept experiments

SIFT matching under homography. SIFT uncertainty comes from [17]. Fig. 1 and 2 (top) shows two views with POI putative correspondences superimposed. Feature matching (step 1) in Sec. 1) is voluntarily set to produce many outliers. Results are shown with the covariances $\Sigma_y + \Sigma_{A_x}$ and $\Sigma_x + \Sigma_{A^{-1}y}$ (90% ellipsoid, yellow) superimposed to POI. One can see that taking into account the Mahalanobis distance automatically adapts the inlier/outlier threshold: it is gentler for points with higher uncertainty (e.g. because they are distant from the 4 points estimating A - in red). For example, bottom-left POI in Fig. 2 would not have been matched with classic RANSAC. In Fig. 2 the algorithm is run twice: once the set with the lowest NFA has been found, the algorithm is run on the whole set of correspondences minus the first group. Both groups (consistent with a homography) correspond to aligned planes. (The cathedral and the building frontage are actually aligned, which explains the matches in the 2nd group).

3D point cloud merging. We propose here an experiment about 3D data fusion based on synthetic data. The aim is to identify the 3D homography between two clouds of matched points, which contain outliers. We plan to test our algorithm on 3D reconstructions of a scene obtained by Structure From Motion algorithms or by depth-cameras. Propagating the uncertainty of SIFT POI or of the depth-camera measurements to the 3D points could indeed provide valuable information for partially overlapping reconstructions [1].

Experimental setting. 1,000 points are uniformly drawn in a $100 \times 100 \times 100$ block, and separated between inliers (I) and outliers (O). Inliers x_i are trans-



Figure 1. Top: 44 SIFT correspondences. Many outliers can be seen. Bottom: 10 inliers are retrieved with our algorithm. Please zoom in on the pdf file.

formed by a 3D homography into the y_i , and outliers are associated with uniformly drawn points into the 3D area limited by the inliers. Each point is given a random covariance which may have standard deviation up to 6 in a direction, and its position is changed with a standard deviation of 3 (compare to the size of the block: the problem is quite challenging). The table gives statistics for our algorithm (average over 10 runs) in 6 situations.

| # I | # O | # retrieved | # O among retrieved |
|-------|-----|-------------|---------------------|
| 1,000 | 0 | 154.8 | 0 |
| 900 | 100 | 122.7 | 1.4 |
| 700 | 300 | 89.3 | 1.6 |
| 500 | 500 | 61.8 | 4.5 |
| 300 | 700 | 45.5 | 6.9 |
| 100 | 900 | 27.5 | 17.5 |

While the algorithm misses a large amount of inliers, the number of outliers among the retrieved correspondences is excellent, even with large rate of outliers in the dataset. Interestingly, if one stops the algorithm as soon as a 1-meaningful group is retrieved (this is sound because of prop. 2), the retrieved group has about 20-30 correspondences and is returned on average after less than 4 iterations (up to 50% outlier rate). This makes us expect possible densification strategies as in [13].

Acknowledgements: thanks are due to Gilles Simon and Nicolas Noury for providing the images, and to Marie-Odile Berger for numerous valuable discussions.

References

- [1] G. Boström, J. Gonalves, and V. Sequeira. Controlled 3D data fusion using error-bounds. *JPRS* 2008.
- [2] S. Brandt. On the probabilistic epipolar geometry. *IVC* 2006.
- [3] F. Cao, J. Lisani, J.-M. Morel, P. Musé, and F. Sur. *A theory of shape identification*. Springer 2008.
- [4] A. Desolneux, L. Moisan, and J.-M. Morel. *From Gestalt theory to image analysis*. Springer 2008.

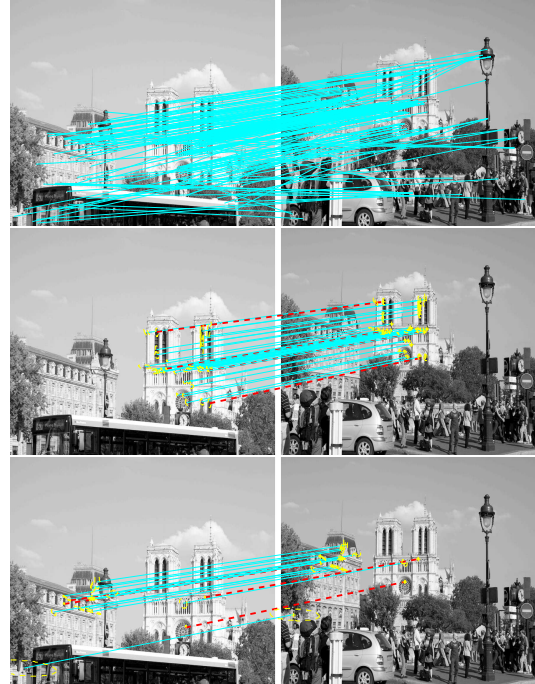


Figure 2. Top: 160 initial SIFT matches. Middle: 1st group (41 matches). Bottom: 2nd group (20 matches).

- [5] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting. *Comm. of the ACM* 1981.
- [6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press 2000.
- [7] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV* 2004.
- [8] P. Mahalanobis. On the generalised distance in statistics. *Proc. of the Nat. Inst. of Sciences of India* 1936.
- [9] L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *IJCV* 2004.
- [10] B. Ochoa and S. Belongie. Covariance propagation for guided matching. *SMVP* 2006.
- [11] T. Papadopoulos and M. Lourakis. Estimating the Jacobian of the Singular Value Decomposition: theory and applications. In *ECCV* 2000.
- [12] X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *IJCV* 1997.
- [13] R. Raguram, J.-M. Frahm, and M. Pollefeys. Exploiting uncertainty in random sample consensus. *ICCV* 2009.
- [14] F. Sur, N. Noury, and M.-O. Berger. Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation. *BMVC* 2008.
- [15] B. Tordoff and R. Cipolla. Uncertain RanSaC. *MVA* 2005.
- [16] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *CVIU* 2000.
- [17] B. Zeisl, P. Georgel, F. Schweiger, E. Steinbach, and N. Navab. Estimation of location uncertainty for scale invariant feature points. *BMVC* 2009.