



**HAL**  
open science

# Determining Epipole Location Integrity by Multimodal Sampling

Huiqin Chen, Emanuel Aldea, Sylvie Le Hégarat-Masclé

► **To cite this version:**

Huiqin Chen, Emanuel Aldea, Sylvie Le Hégarat-Masclé. Determining Epipole Location Integrity by Multimodal Sampling. 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Sep 2019, Taipei, Taiwan. pp.1-8, 10.1109/AVSS.2019.8909847 . hal-02472769

**HAL Id: hal-02472769**

**<https://hal.science/hal-02472769>**

Submitted on 10 Feb 2020

**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.

# Determining Epipole Location Integrity by Multimodal Sampling

Huiqin Chen, Emanuel Aldea and Sylvie Le Hégarat-Masclé  
SATIE - CNRS UMR 8029

Paris-Sud University, Paris-Saclay University, France

{huiqin.chen, emanuel.aldea, sylvie.le-hegarat}@u-psud.fr

## Abstract

*In urban cluttered scenes, a photo provided by a wearable camera may be used by a walking law-enforcement agent as an additional source of information for localizing themselves, or elements of interest related to public safety and security. In this work, we study the problem of locating the epipole, corresponding to the position of the moving camera, in the field of view of a reference camera. We show that the presence of outliers in the standard pipeline for camera relative pose estimation not only prevents the correct estimation of the epipole localization but also degrades the standard uncertainty propagation for the epipole position. We propose a robust method for constructing an epipole location map, and we evaluate its accuracy as well as its level of integrity with respect to standard approaches.*

## 1. Introduction

When a photo is taken within the field of view of a reference camera, it is possible to pinpoint the person taking the photo by locating the epipole in the reference view. Localizing the person wearing the camera has important applications in crowd surveillance for instance, when the wearer follows a specific event in the crowd and needs to localize accurately this event in an absolute reference system. Generally, the epipole localization problem is of significant interest for a wide range of tasks involving embedded cameras, such as locating law enforcement agents or pedestrian camera wearers in the field of view of static security cameras, or mutual localization inside swarms of coordinating UAVs. In these real applications, estimating the confidence region as a candidate search area is more beneficial than a single epipole position since a false epipole estimation may mislead completely the localization of the target. For this practical reason, the problem we address in this work is more general than the epipole estimation itself, and it regards its uncertainty estimation, especially how to improve its reliability.

The epipole of an image pair is defined as the point of

intersection of the baseline with the image plane. The most widely used approach for computing the epipole position is derived from the fundamental matrix estimation by singular value decomposition (SVD). Alternatively, the authors in [8] attempted to use a direct method based on the invariance of cross-ratio but fail to surpass the precision of the fundamental matrix method. Motion parallax methods [6, 5] find the epipole by determining the epipolar line with two points which coincide in the image. It requires to find firstly the homography for one or two planes inside the observed scene. In our work, we focus on the general method based on the fundamental matrix. For the epipole uncertainty estimation, the authors in [9] develop a closed-form solution for the Jacobian of the SVD and make it possible to estimate the epipole uncertainty with a first order propagation from the fundamental matrix, whose uncertainty is well developed in many existing works [1, 15, 12].

It is well known that the epipole estimation is unstable, depending on image content and interest point detection noise, but also due to the remaining outliers and to possible degenerate configurations. However, few work has been done for the reliability of its uncertainty estimation which is important in practice. Assuming that inliers can be identified by robust estimation such as RANSAC and the noise values follow a normal law with a small variance as typically  $\sigma \in [0, 1]$  px, the standard propagation pipeline proposed in [9] underestimates the epipole location uncertainty and yields a low level of integrity for covering the true epipole location due to the remaining false matches, as illustrated by Figure 1. The influence of outlier observations missed by RANSAC outlier rejection on the uncertainty has also been underlined in [11], when modeling the uncertainty of the simpler homography transform for image registration.

In order to set up a more reliable epipole location map, we develop a method based on multimodal sampling of the RANSAC process. In accordance with the considered application of wearable camera localization, we focus on the case when the epipole is visible in the view of the reference camera. The basic idea is to empirically evaluate the uncer-

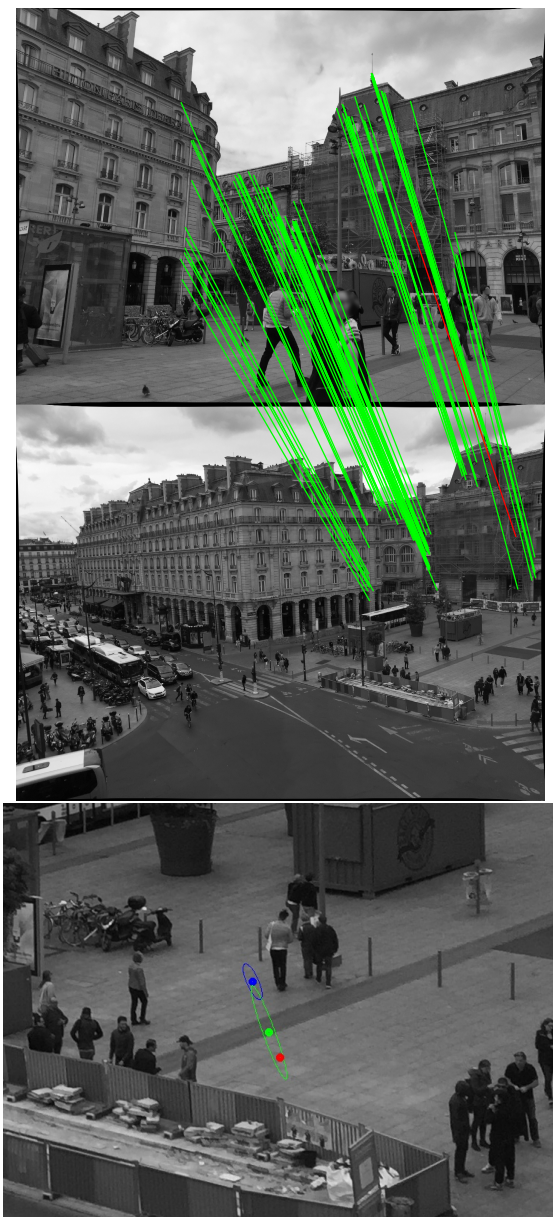


Figure 1: Illustration of an epipole uncertainty estimation. For a given pair of images, the matches selected as inliers by RANSAC (91 out of 208 initial matches) are shown with green and red lines. There is only one false positive match which is illustrated by the red line, all the other matches being true positives. In the figure below (close-up of the scene in the reference view), the estimated epipole uncertainty on all inliers (presented by blue ellipse) predicts a small standard deviation of the estimated epipole, and misses the true epipole (the red dot). When the single false positive is removed, the epipole ellipse (the green ellipse) predicts the uncertainty reliably.

tainty by simulating stochastic realizations of the matched point set. Our work focuses on a family of techniques which avoid relying explicitly on error models for outlier observations during uncertainty evaluation. Each of these realizations is processed following the standard pipeline to compute the epipole location and its uncertainty so that it provides an elementary geometric model. Global uncertainty is then derived from the fusion of all these elementary models using a tractable voting strategy.

The benefits of the proposed stochastic approach are twofold. First, we propose an epipole localization map which increases the accuracy in locating the true epipole while avoiding at the same time to underestimate the underlying uncertainty, contrarily to existing approaches. Second, our method has a low computational burden as the sampling process exploits intermediate results performed nevertheless during the robust estimation step which addresses outlier rejection.

## 2. Background

Let  $M = \{x_i, y_i, x'_i, y'_i\}_{1 \leq i \leq n}$  denote the set of  $n$  putative point matches between two views. A standard pipeline to compute the epipole consists in applying firstly RANSAC on  $M$  to filter out false matches, and in estimating the fundamental matrix  $F$  based on the inlier set  $I \subset M$ . The epipole  $e$  is derived from  $F$  by SVD. As the epipole estimation is based on the computation of the fundamental matrix, we review first the approaches for the estimation of the fundamental matrix uncertainty.

### 2.1. Fundamental matrix uncertainty

The existing methods for the estimation of the  $F$  covariance can be classified, depending on the estimation algorithm, into three categories.

**Uncertainty of the 8 point algorithm.** With  $n = 8$ ,  $F$  may be derived by solving the linear system  $Af = c$  where  $A$  is the stack of  $[x'_i x_i, x'_i y_i, x'_i y'_i, y'_i x_i, y'_i y_i, y'_i x_i, y_i, y_i]_{1 \leq i \leq 8}$  with  $f = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32})^T$  and  $c = -(1, 1, 1, 1, 1, 1, 1, 1)^T$ .  $F_{33}$  is set to be 1. Under this assumption, we get

$$f = A^{-1}c, \quad \Sigma_f = J_X \Sigma_X J_X^T, \quad (1)$$

where  $J_X$  is the Jacobian matrix of  $f$  with respect to the set of point matches  $X$ . To impose the constraint of rank 2 for the fundamental matrix, the smallest singular value derived by SVD of  $F$  is forced to be 0. We use  $F_{rank2}$  to denote the rank 2 fundamental matrix. Then the covariance matrix for  $F_{rank2}$  can be derived as

$$\Sigma_{F_{rank2}} = J_{F_{rank2}/f} \begin{bmatrix} \Sigma_f & 0 \\ 0 & 0 \end{bmatrix} J_{F_{rank2}/f}^T. \quad (2)$$

where  $J_{F_{rank2}/f}$  is the Jacobian matrix of  $F_{rank2}$  with respect to  $f$ . The explicit computation for  $J_X$  and  $J_{F_{rank2}/f}$  can be found in [12].

**Uncertainty for least squares estimation ( $n > 8$ ).** With  $n > 8$ , it is possible to apply the Jacobian of the SVD [9] to least squares estimation for fundamental matrix uncertainty as mentioned in [12].  $F$  can be computed as in [4] by SVD from the matrix  $B$  which is the stack of  $[x'_i x_i, x'_i y_i, x'_i, y'_i x_i, y'_i y_i, y'_i, x_i, y_i, 1]_{1 \leq i \leq n}$ . By setting  $f = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^T$  with  $\|f\| = 1$ , The covariance matrix of  $f$  is

$$\Sigma_f = J_{SVD} \Sigma_X J_{SVD}^T \quad (3)$$

One can use the same strategy as with the 8 point algorithm to impose the rank-2 constraint.

**Uncertainty for nonlinear minimization estimation.** When  $F$  is estimated iteratively, for example by minimizing the following criterion:

$$\min_F \sum_{i=1}^n (d^2(\mathbf{x}'_i, F\mathbf{x}_i) + d^2(\mathbf{x}_i, F^T \mathbf{x}'_i)), \quad (4)$$

the covariance of  $F$  may be computed with the formula presented in [15]:

$$\Sigma_{F_p} = \frac{2R}{n - d_p} H^{-T}, \quad \Sigma_F = \frac{\partial F}{\partial F_p} \Sigma_{F_p} \frac{\partial F}{\partial F_p}^T, \quad (5)$$

where  $F_p$  is the  $F$  parameterization vector for the estimation of the fundamental matrix and  $R$  is the residual error for the achieved minimum,  $d_p$  is the number of independent parameters of the problem, and  $n$  is the number of used point matches.  $H$  is the Hessian matrix of the cost function related to parameters for the achieved minimum.

The first method is specific for 8 point matches. The second and third method can be used in general case. Let us notice that the fundamental matrix uncertainty estimated by the first and the second method are related to the uncertainty in match localization (due to detector noise), as well as to their geometrical structure (specified by their coordinates). Differently, the uncertainty of  $F$  computed by the third method depends mainly on the residual error and on the number of point matches.

## 2.2. Epipole uncertainty

Building upon the strategies mentioned above, there are two fundamental approaches to characterizing the uncertainty of the epipole: the analytical solution based on the introduced covariance matrix estimation for the fundamental matrix, and a statistical method based on Monte Carlo simulation.

**Analytical pipeline.** Given the estimation of the fundamental matrix  $F$  and its covariance  $\Sigma_F$ , the epipole  $e$  and

the corresponding  $e'$  are computed by performing the SVD of  $F$ , as the epipoles satisfy the constraints  $Fe = 0$  and  $e'F^T = 0$  respectively. For the epipole covariance, the authors in [9] derive the analytical solution by computing the Jacobian of the SVD:

$$\Sigma_e = J_{SVD} \Sigma_F J_{SVD}^T. \quad (6)$$

Thus, the analytical pipeline consists in firstly estimating  $\Sigma_F$  with the previous presented methods on the inlier set  $I$ , by assuming that the noise of point matches in  $I$  follows an independent Gaussian distribution with a small variance. Then  $\Sigma_e$  is obtained with Equation (6). Specifically, the epipole and its covariance matrix can be directly extracted from the parameters  $F_p$  and its covariance  $\Sigma_{F_p}$  in Equation (5) as the parameterization of  $F_p$  is based on epipoles (we refer the reader to [15] for more details).

To visualize the epipole uncertainty related to the estimates  $e$  and  $\Sigma_e$ , we use the k-hyper-ellipsoid based illustration [10] defined by the following equation:

$$(x - e)^T \Sigma_e^{-1} (x - e) = k^2. \quad (7)$$

By choosing  $k^2 = 5.991$ , the ellipse represents the 95% confidence region according to the probability  $P_{\chi^2}(k, 2)$  for the epipole to lie inside.

**Monte Carlo Simulation.** An alternative solution to estimate uncertainty is a statistical method based on a standard Monte-Carlo (MC) simulation [4]. Considering the input point matches contaminated with Gaussian noise, one can draw multiple point realizations by adding to the input points some noise following the given noise distribution, and estimate the epipole with a given estimation method. The epipole uncertainty is characterized by the distribution of all the computed epipole realizations. Compared to the analytical solution, the MC approach avoids the computation of derivatives and the assumption of linearity with the price of more expensive computations.

We argue that both the analytical solution and the MC techniques based on simulating the measurement noise will not avoid the issue highlighted in Figure 1, since the final error is not due to the covariance matrix approximations, but to the incorrect modelling of the observation noise. In fact, the assumption of Gaussian noise with small variance is not realistic at all for the outliers. Therefore, we propose in the following section a solution which overcomes the above limitations by avoiding to rely on explicit assumptions about outlier noise.

## 3. Method

When the nature of observation noise is unknown, bootstrapping has been proposed [3] for evaluating the robustness of a solution. We firstly justify the interest of a bootstrap resampling technique and then present how to integrate it into epipole uncertainty estimation in our work.

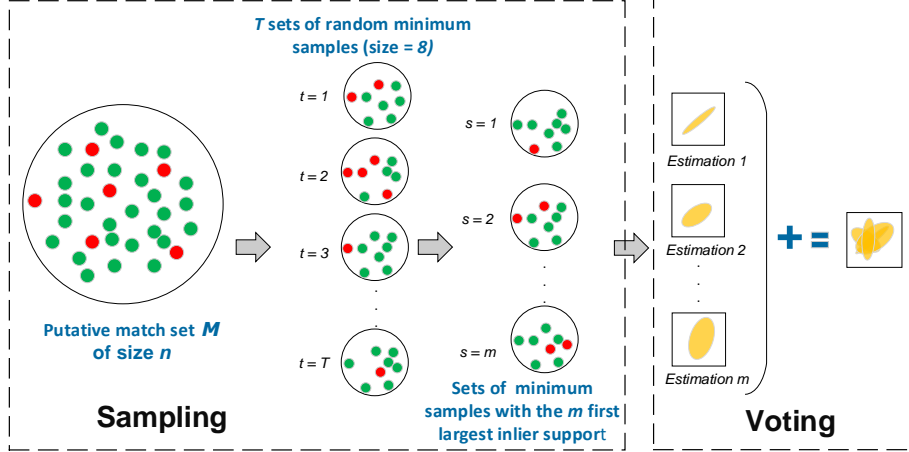


Figure 2: Method overview. The proposed algorithm is divided into two stages: Sampling and Voting. Given the input of putative match set  $M$ , containing true point matches (green dots) and false point matches (red dots), we randomly sample  $T$  sets of minimum number of point matches and then select the sets with the  $m$  first largest inlier support during the stage of sampling. The second stage constructs the final map by cumulative voting of the estimations from these  $m$  sets.

Assuming the observations represent a set of independent and identically distributed data points  $\mathcal{D}$ , the bootstrap process consists in generating different sets of data points  $\mathcal{D}_i$  by resampling the same number of points from  $\mathcal{D}$  with replacement. The same parameter estimation procedure is then performed on each generated  $\mathcal{D}_i$ . Although it belongs to the stochastic approach category as well, bootstrapping differs from MC approaches, since the solution diversity does not derive from any noise assumption but from dataset intrinsic variability. Instead of sampling data points by following a probably wrong noise distribution, the bootstrap method samples points exclusively among the actual observation dataset  $\mathcal{D}$ . As mentioned in [10], the basic idea is that the dataset  $\mathcal{D}$  itself is considered as the best and at the same time the only available estimator for the underlying probability distribution when the noise is unknown.

In order to deal with the unknown error distribution introduced by outliers, we adopt an approach for epipole uncertainty estimation inspired by the bootstrap method. A key point is that it manages to exploit part of the computation performed for outlier rejection.

### 3.1. Proposed algorithm

The core idea is to draw different realizations by sampling point matches inside the given dataset itself. A straightforward way is to generate sample models by sampling different sets of minimal number of point matches directly from  $M$  and draw the distribution of epipoles estimated from each realization. Although this mechanism is closely related to RANSAC, the difference is that the latter process considers all the models instead of only the best one. For evaluating the uncertainty of the solution, it is not

reliable to rely only on the most consensual model as we illustrated in Figure 1. However, considering any models is not a wise choice either, since many of them do not provide useful information because they capture few inliers, especially in difficult visual contexts.

In our work, we propose a solution which performs an accelerated model set simulation by considering models with a preference based on the number of captured inliers, illustrated by Figure 2. Let  $S = \{S_1, S_2, \dots, S_k, \dots, S_T\}$  be the set of all sets of minimal number of point matches randomly sampled from  $M$  during RANSAC, ranked in decreasing order with respect to the size of their inlier support, denoted by  $|I_{S_k}|$ . The epipole uncertainty estimation is based on these  $m$  models estimated from  $m$  first sets of point matches in  $S$ :  $S^m = \{S_1, S_2, \dots, S_k, \dots, S_m\} \subset S$ . In order to devise a continuous probability map for the epipole location over the image domain, we present each model with the epipole  $e_k$  and the covariance matrix  $\Sigma_{e_k}$ , where  $\Sigma_{e_k}$  is computed with the propagation formula related to 8 point algorithm (see Equation (2), (6)). For each model, the uncertainty map of epipole localization is computed by

$$P_k(p) \propto \exp \left\{ -\frac{1}{2} (p - e_k)^T \Sigma_{e_k}^{-1} (p - e_k) \right\}, \quad (8)$$

where  $p$  is the pixel position in image domain  $\mathcal{P}$  of dimension  $[0, w] \times [0, h]$ .  $P_k(p)$  denotes the probability of  $p$  being the true epipole coordinate given the epipole and its covariance matrix estimated on model  $S_k$ . Finally, we compute the global uncertainty map as the cumulative vote of the uncertainty map based on all  $m$  models, followed by a normalization by dividing with the maximum value over the

map:

$$P(p) = \frac{\sum_{k=1}^m P_k(p)}{\max_{p \in \mathcal{P}} \sum_{k=1}^m P_k(p)}. \quad (9)$$

This uncertainty map  $P$  exhibits the coherence with the distribution of  $\{e_k\}_{1 \leq k \leq m}$  as the maximum value is still located in the position of  $e_k$  for each single uncertainty map  $P(k)$ . The cumulative voting guarantees to follow the global consistency among all considered models instead of only one single estimation. Although more complex fusion strategies exist, voting is widely used due to its tractability for large numbers of hypotheses.

### 3.2. Discussion on parameters

The proposed approach has one key parameter, namely  $m$  the number of considered models. The value of  $m$  is expected to relate to the distribution of the support sizes  $|I_{S_i}|$  for the most consensual models, as well as to the largest number of inliers  $|I_{S_1}|$  captured during random sampling process. Otherwise stated,  $m$  is set to make sure that all the chosen models capture a sufficiently large number of inliers with respect to the best model. Following our validation in difficult visual contexts, we set a fixed, reasonably large value  $m = 1000$  during the RANSAC sampling stage. This choice is well adapted in case significant, frequent ambiguities among models which capture similar numbers of inliers which are also close to  $|I_{S_1}|$ . In less ambiguous contexts, when the process is most certain about the best model and the values  $|I_{S_i}|$  decay faster, the choice of a fixed value  $m$  may introduce more unreliable, low quality models. This problem may be easily avoided by setting a threshold  $\tau$  for the number of inliers during the voting stage. If  $|I_{S_i}| < \tau \times |I_{S_1}|$ , then the model estimated from the set of point matches  $S_i$  will be discarded during the cumulative voting (See Equation (9)). The threshold  $\tau$  is suggested to be set in the range 0.7 to 0.9.

### 3.3. Computational complexity

A significant computational advantage of our approach is that it is tailored to benefit from the intermediate evaluations performed by RANSAC, that are generally wasted by pose estimation algorithms. Therefore, we make a clear difference below between the sampling cost which is contained within the RANSAC algorithm, and the additional incurred cost with respect to outlier rejection.

As it is well known, the number of iterations  $T$  performed during RANSAC depends on the approximate ratio of true inliers among the observations, irrespective of our method. The sorting algorithm required for selecting the  $m$  most consensual models may be performed by insertion operations in a sorted vector, with an overall underlying complexity of  $\mathcal{O}(T \log(m))$ . This additional cost is negli-

Algorithm	Input	Estimation Method
Least Squares SIFT	$I_{SIFT}$	LS (Eq. (3)), Eq.(6)
Least Squares NN	$I_{NN}$	LS (Eq. (3)), Eq.(6)
Minimization SIFT	$I_{SIFT}$	Min (Eq. (5)), Eq.(6)
Minimization NN	$I_{NN}$	Min (Eq. (5)), Eq.(6)
Ours SIFT	$M_{SIFT}$	Sampling and Voting

Table 1: Summary of compared algorithms.

gible with respect to the  $F$  estimation and evaluation of its support performed at each RANSAC iteration.

For the epipole location voting step (the vote domain being the image domain as well), the naive complexity will be  $\mathcal{O}(m \times w \times h)$  which may be significant for large images. However, when applying Equation (8) in order to compute the uncertainty map for each model, one may constrain the domain of pixels to be updated  $p$  from  $\mathcal{P}$  to the confidence region  $\mathcal{C}$ , where  $\mathcal{C}$  is the  $k$ -hyper-ellipsoid defined in Equation (7). With this acceptable approximation, the complexity may be significantly reduced to  $\mathcal{O}(m \times a \times b)$ , with  $a$  and  $b$  being the main axes of the uncertainty ellipse. As for most voting approaches, this second step which is specific for our algorithm may be trivially parallelized if real time constraints are critical.

## 4. Experiments and results

In the experiments part, we test the performance of the proposed method for epipolar uncertainty estimation with the public urban dataset used in [14]. Four variants based on the existing literature are compared with the proposed method in terms of accuracy and precision for qualitative and quantitative evaluation.

### 4.1. Dataset

We firstly choose images from the given dataset with various change of view and generate different pairs of images. We keep the ones whose epipole is visible in the reference view. In order to guarantee sufficient overlap between two views with a fair estimation of the epipolar geometry, we remove images pairs whose number of inliers found by RANSAC is less than 10 or below 20% of putative matches. The final number of image pairs is 1118. The ground truth epipole location is computed from the provided calibration information, which was acquired using Structure From Motion.

### 4.2. Implementation

Let  $M_{SIFT}$  denote the set of putative point matches between two views, computed by SIFT with standard ratio test 0.75 [7] and  $I_{SIFT}$  the inlier set selected by RANSAC from



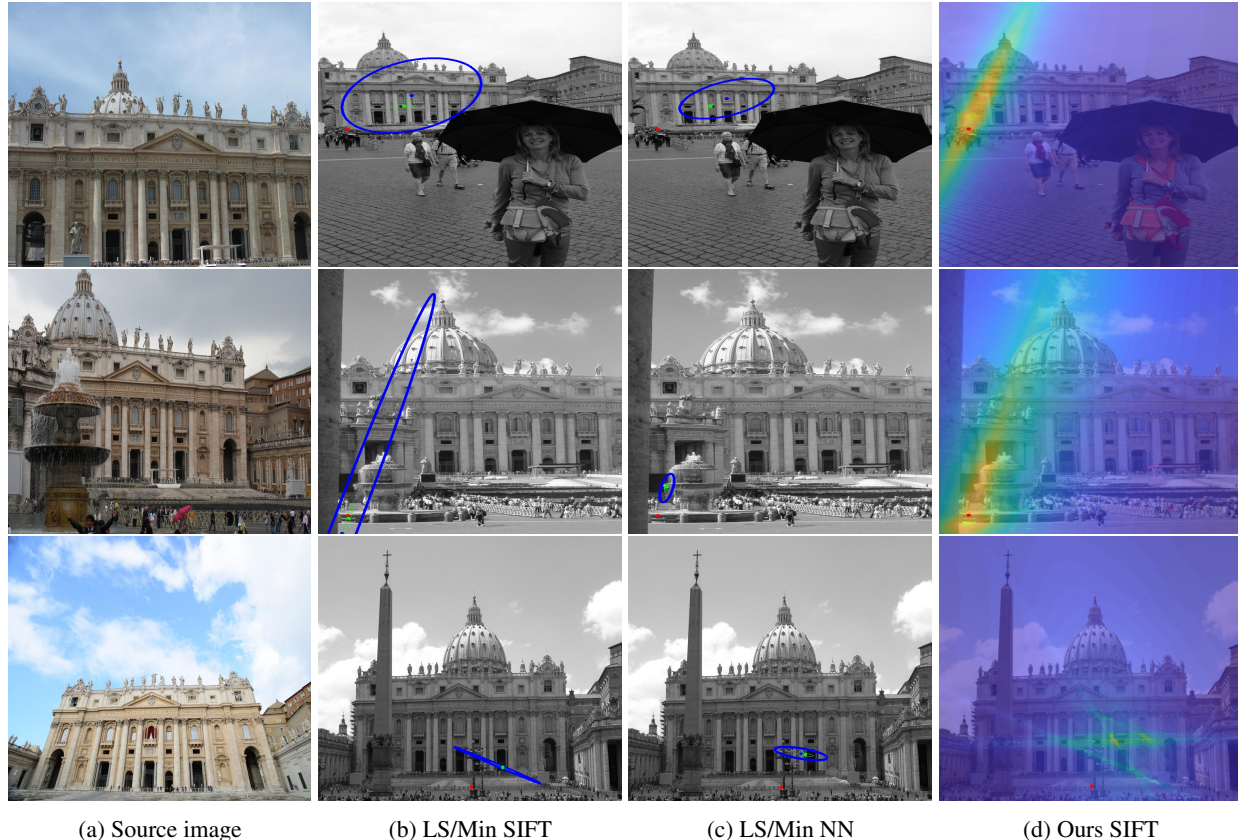


Figure 3: Qualitative performance. (a) presents the source image taken in the reference view. (b), (c) and (d) show the epipole uncertainty estimation in the reference view with different algorithms as mentioned in Table 1. The red dot is the ground truth epipole location. In (b) and (c), the blue ellipse presents 95% confidence region for the epipole location based on least squares estimation and the green ellipse for minimization. The proposed estimation is illustrated in (d) by a heat color map (the yellow color corresponds to the highest probability).

$M_{SIFT}$ . In order to compare with the method in [14], we apply their geometric neural network on  $M_{SIFT}$  to filter out false matches and generate the inlier set  $I_{NN}$ .

The proposed method is applied directly on  $M_{SIFT}$ . During sampling stage, we set the number of RANSAC iterations  $T = 100000$  based on the expected ratio of outliers. The number of considered models  $m = 1000$ . For voting,  $\tau$  is set to be 0.9. The output is an uncertainty map for epipole location computed by Equation (9). Under this configuration, the standard RANSAC process takes on average 21.3s, with an addition 2.2s during our algorithm voting stage on the domain of size  $768 \times 1024$ . To compare the proposed uncertainty estimation of epipole with the standard pipeline based on the analytical solution (See Section 2.2), we develop four variants depending on the input point matches and uncertainty estimation method, detailed in Table 1. We compute a simple uncertainty map with estimated epipole  $e$  and its covariance matrix  $\Sigma_e$  using Equation (8) for these variants.

### 4.3. Evaluation Metric

Given the estimated epipole uncertainty map  $P$  and the ground truth epipole location  $e_{gd}$ , the evaluation is based on accuracy as well as precision, defined as follows.

**Accuracy.** The accuracy represents the predicted likelihood for the true epipole and the precision underlines how tightly the candidate area is delineated inside the image space. The accuracy is measured with the score

$$s = P(e_{gd}), \quad (10)$$

A large score is expected for a prediction with high confidence level. We compute first the score for each pair of images and then compute the success ratio by counting the percentage of image pairs whose score is higher than a chosen threshold  $\tau_s$ . By setting different values for  $\tau_s$ , we obtain the curve of success ratio for each method.

**Precision.** The precision is measured in terms of similarity between the estimated uncertainty map  $P$  and ground truth map  $P_{gd}$ , which is defined by setting a value

$P_{gd}(e_{gd}) = 1$ , and null values elsewhere over the domain of image. This measure is complementary to accuracy, since high scores across large areas of the image space increase the accuracy. In the following, we present two measures considered for assessing the similarity between the two maps. For both these measures, we first normalize the estimated uncertainty map across the image space in order to convert it to a proper probability distribution function.

- **KL divergence based distance.** Given distributions  $G$  and  $Q$ , the KL divergence is defined as:

$$D_{KL}(G \parallel Q) = \sum_{x \in \mathcal{X}} G(x) \log \frac{G(x)}{Q(x)}. \quad (11)$$

In order to get a symmetric measure, one may also compute the reverse  $D_{KL}(Q \parallel G)$  and take the average of  $D_{KL}(G \parallel Q)$  and  $D_{KL}(Q \parallel G)$ . The Kullback-Leibler (KL) divergence is widely used for measuring to what extent two probability distributions are related (a divergence of zero indicating actually identical distributions). However, in the case of an indicator ground truth distribution,  $D_{KL}(P_{gd} \parallel P)$  is null  $\forall p \in \mathcal{P} \setminus \{e_{gd}\}$ , as  $P_{gd}(p \neq e_{gd}) = 0$ . Thus, it only depends on the value of  $P(p = e_{gd})$ . Conversely,  $D_{KL}(P \parallel P_{gd})$  is not sensitive to the values of  $P(p)$  for the same reason. Therefore, in our setting the KL divergence provides a good estimate for accuracy (as the score  $s$ ), but fails to take into account the overall spatial geometry of the uncertainty map in order to provide helpful information about the localization precision (i.e. the desirable property according to which the inferred likelihood map  $P$  outputs significant values around the true epipole location).

- **Optimal transport based distance.** Optimal transport (OT) has emerged as a powerful tool to evaluate the similarity of two distributions based on their spatial layout [2, 13], by computing the minimal cost in order to transform a distribution into the other one. In spite of its suitable behaviour, the use of OT has been limited by its computational cost in multidimensional spaces (more details about the underlying formalism may be found in [2]). However, in our case the optimal transport based distance also has a simplified form due to the indicator value in  $P_{gd}$  (viewed as a degenerated point distribution), and it may be conveniently computed in an exact form as follows:

$$D_{OT}(P, P_{gd}) = \sum_{p \in \mathcal{P}} P(p) \|p - e_{gd}\|_2, \quad (12)$$

when considering the Euclidean norm on  $\mathbb{R}^2$ , denoted by  $\|\cdot\|_2$ . Unlike the KL divergence, the OT distance is able to discriminate among multiple distributions which, for the same score  $s$ , place more likelihood in the proximity of the ground truth epipole location.

Based on the characteristics of the two distance measures, we choose  $D_{OT}$  as the measure of precision. A small value presents a high level of precision. We compute  $D_{OT}$  between  $P$  and  $P_{gd}$  for each pair of image and draw the normalized histogram of  $D_{OT}$  over all image pairs.

#### 4.4. Results

**Qualitative Evaluation.** We present qualitative evaluation results using three examples of image pairs in Figure 3. The above row presents a successful scenario in which the proposed algorithm performs very well in terms of accuracy and precision with respect to existing algorithms. In this case, the most consensual model (selected by RANSAC) is invalid, therefore standard pipelines provide a false uncertainty estimation. In the middle row, we present a case where the methods we considered for comparison exhibit very variable levels of performance, despite the fundamentally similar pipeline used for uncertainty propagation. In this case, our method still exhibits a very good accuracy. Finally, the last row shows a failure case, for which all methods provide wrong evaluations due to the degeneracy of the fundamental matrix evaluation, since the inlier set is located exclusively on the same facade.

Overall, the integrity of our method is excellent. In difficult scenes, it tends to enlarge the uncertainty area which is a desirable behavior, and will fail only in the presence of degenerate geometric configurations (very limited overlap or in the presence of a single planar dominant structure).

**Quantitative Evaluation.** The quantitative evaluation is based at the same time on the accuracy as well as on the precision of the localization.

- **Accuracy** The accuracy evaluation is presented in Figure 4(a). The proposed method achieves the best performance on accuracy. For other methods based on inliers of standard RANSAC, even with a very small threshold of score, their success rate is still below 40%. On the contrary, the proposed method achieves 40% success rate even with a large threshold such as 0.6. In agreement with the behaviour predicted in Section 4.3, Figure 4(d) shows the KL divergence based distance for the different methods, which correlates with the score performance.
- **Precision** The evaluation on precision is shown in Figure 4(b). Due to a more conservative estimation of the localization, the proposed method exhibits the largest distance histogram due to small value predictions over the image domain, for which the OT distance is very sensitive. However, Figure 4(c) shows the evolution of the average  $D_{OT}$  when the estimated probability maps are set to zero below a threshold (specified on the horizontal axis). Once the smallest values are gradually removed, the average distance of the proposed method approaches the distance of others methods, a fact which underlines



that the modes of the pdf are closely located to the real location.

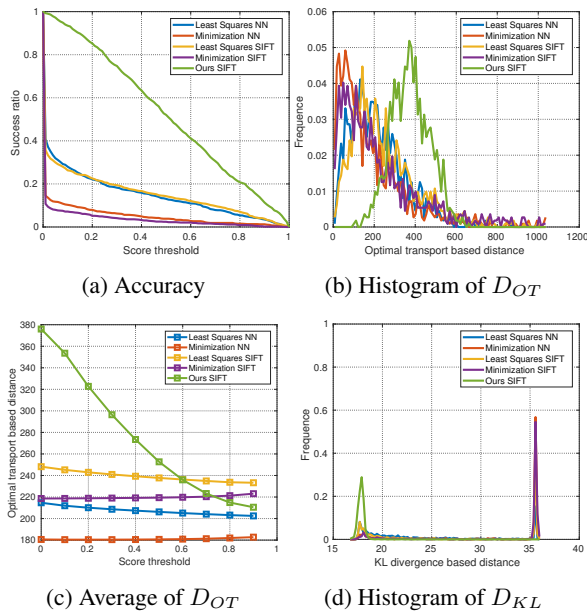


Figure 4: Quantitative Evaluation. For all evaluated algorithms, (a) compares the percentage of image pairs whose score is larger than the correspondent threshold. (b) illustrates the normalized histogram of the optimal transport based distance for the precision evaluation. (c) is the average of the optimal transport based distance over the tested image pairs with different score threshold. (d) compares the histogram of the KL divergence based distance.

Both qualitative and quantitative evaluations reveal that the proposed method improves the integrity of the estimated epipole. Compared to the standard pipeline based on a single estimation of inliers found by RANSAC, which yields a low integrity of the estimated epipole due to the presence of outliers, the proposed multimodal sampling method explores different model estimations on potentially valid sets of point matches. The estimation on such sets of matches helps to decrease the chance to include outliers. The voting strategy by multiple models follows the global tendency and finally outputs a reliable prediction for epipole localization.

## 5. Conclusion

To improve the epipole uncertainty estimation, we proposed a majority voting strategy by sampling multiple models based on the standard RANSAC and uncertainty propagation pipeline. The constructed probability map provides at the same time, and contrary to baseline methods, a high accuracy and a good precision for high probability areas. Our work opens a promising avenue for coupling the algo-

rithm with a pedestrian detector in order to further constrain the location of the person acquiring the photo in the reference view.

Also for future work, we intend to integrate our cumulative sampling approach into a neural network by learning the voting strategy. Following our study, the KL-divergence based and OT based distances are good candidates for the loss function supposed to guide the learning process.

## Acknowledgment

This study was supported by the S<sup>2</sup>UCRE project, co-funded by the BMBF grant 13N14463 and by the ANR grant ANR-16-SEBM-0001.

## References

- [1] G. Csurka, C. Zeller, Z. Zhang, and O. D. Faugeras. Characterizing the uncertainty of the fundamental matrix. *CVIU*, 68(1):18–36, 1997.
- [2] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NIPS*, 2013.
- [3] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.
- [4] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [5] J. M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion parallax. In *European Conference on Computer Vision*, pages 205–210. Springer, 1994.
- [6] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. of the Royal Soc. of London. Series B. Biological Sciences*, 208(1173):385–397, 1980.
- [7] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [8] Q.-T. Luong and O. D. Faugeras. On the determination of epipoles using cross-ratios. *CVIU*, 71:1–18, 1998.
- [9] T. Papadopoulos and M. I. Lourakis. Estimating the jacobian of the singular value decomposition: Theory and applications. In *ECCV*, pages 554–570. Springer, 2000.
- [10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.
- [11] P. O. Sundlie, C. N. Taylor, and J. A. Fernando. Sources of uncertainty in feature-based image registration algorithms. In *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VI*, volume 9464, page 94640Z. Int. Soc. for Optics and Photonics, 2015.
- [12] F. Sur, N. Noury, and M.-O. Berger. Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation. In *BMVC 2008*, page 10, 2008.
- [13] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [14] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua. Learning to find good correspondences. *CVPR*, pages 2666–2674, 2018.
- [15] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *IJCV*, 27(2):161–195, 1998.