Camera Localization Based on Belief Clustering

S. Le Hégarat-Mascle, Huiqin Chen, Emanuel Aldea

Camera Localization Based on Belief Clustering

Huiqin Chen
SATIE UMR 8029
Paris-Saclay University
huiqin.chen@u-psud.fr

Emanuel Aldea
SATIE UMR 8029
Paris-Saclay University
emanuel.aldea@u-psud.fr

Sylvie Le Hégarat-Mascle
SATIE UMR 8029
Paris-Saclay University
sylvie.le-hegarat@u-psud.fr

***Abstract*—This work deals with epipole estimation related to egocentric camera localization in surveillance and security applications. Matching visual features in the images provides some evidences for various solutions, so that epipole localization can be addressed as a fusion task with a large number of sources including outlier ones. In order to deal with source imprecision and uncertainty, we rely on the belief function theory and a 2D framework suited for our application. In this framework, we address the challenges introduced by a large number of sources with a strategy based on clustering and intra-cluster fusion. The proposed method exhibits more robustness in terms of accuracy and precision when compared on real data with the standard algorithms which provide single solution. Since we provide a Basic Belief Assignment as a result, our strategy is particularly adapted for the prospective combination with additional sources of information.**

***Index Terms*—Egocentric camera localization, epipole uncertainty, large number of sources, belief function theory 2D**

I. Introduction

The relative pose estimation between two cameras is a fundamental task in computer vision, and is an integral part of other algorithms addressing more complex tasks such as 3D reconstruction [1]–[3], localization [4] or navigation [5]. Although relative pose estimation has been studied extensively for more than 30 years, there is still ongoing work in order to improve its performance in adverse conditions introduced by wide baselines, large non-salient areas or repetitive structures specific to urban settings. In order to solve the pose estimation problem, the established approach is to rely on keypoint associations between the two views [6]. Among these matches, a minimal number of them denoted as a *minimal set*, typically seven [7] or eight [8], are required in order to solve for a pose. However, the entire set of observations contains in practice a significant ratio of outlier matches which skew the solution if included in the estimation. Since it is outright infeasible to analyze exhaustively the consensus among all the available minimal sets due to the combinatorial nature of the problem, stochastic sampling is widely used [9], and typically the solution is obtained from the most consensual sample set.

The relative pose is composed of a 3D rotation-translation pair exhibiting six degrees of freedom. A visual interpretation of a solution may be derived by projecting the second camera location in the first camera image plane. The projected point is called the epipole, and it is of particular interest in surveillance and localization applications. Therefore, a pose proposal is associated to an epipole along with an elliptic uncertainty area provided by the covariance matrix of its location. Now, the epipole location suffers from the same sensitivity to outliers which represents even more of an issue that, for some applications (surveillance, planning in drone or vehicular networks) which rely on the joint localization of static and mobile camera wearers, wrong epipole identification has a detrimental effect and additional techniques must be used in order to provide reliable solutions. In this work, we assume that imprecision is preferable to unreliability. It means that we relax the constraint of single localization in favor of a few solutions provided they include the true location.

In order to handle the uncertainty introduced by a significant ratio of spurious observations, we model the input of our problem and perform the main steps of our algorithm in the Belief Function Theory (BFT) framework. This formalism was made popular by various real-world applications [10] for which it provides an efficient modelling of imprecise information, allowing for fairer and more consistent decisions. However, for some applications, scalability remains a challenge, either in terms of the size of the discernment frame or in terms of the number of sources to be combined.

Firstly, regarding the size of the Discernment Frame (DF), the issue is that belief functions (mass, plausibility, etc.) are defined on the DF powerset, so that for a DF denoted Ω whose cardinality is $|\Omega|$, there are potentially $2^{|\Omega|}$ hypotheses to consider and to enumerate. Such an issue arises typically in localization applications in which DF corresponds to possible positions of the considered target, i.e typically $|\Omega| = 10^6$ if we require a spatial resolution equal to 10×10 cm^2 within an area of 100×100 m^2. First solutions, e.g. [11], use some tricks (e.g. conditioning) to consider only a subpart of DF at once. Then, [12]–[14] propose a solution where 2^Ω elements (singleton and compound hypotheses) are no longer labelled (as for enumeration) while each element of the focal set (that is usually a small subset of 2^Ω) is handled through its own description. Specifically, in [12], [13], the handled focal elements are described as sets of rectangles (tiles) similarly to the representation used in Interval Analysis [15], whereas [14] provides a more general representation of any 2D shapes using polygons. In both cases, belief function operators based on set relationships (intersection, union etc.) have to be redefined in an efficient way.

Secondly, regarding the number of sources, a first difficulty

This study was supported by the S^{2}UCRE project, co-funded by the German BMBF grant 13N14463 and by the French ANR grant ANR-16-SEBM-0001

is related to the used combination rule. In particular, using the very popular conjunctive rule proposed by Smets [16], the mass on the empty set ($m(\emptyset)$), usually called conflict, is an increasing function with respect to the number N of combined BF: $m(\emptyset) \rightarrow 1$ when $N \rightarrow +\infty$. Note that this rule was proposed to avoid evidence modelling issues (e.g. as in the case of the Zadeh example) hidden by mass normalization as performed in the original Dempster's rule or orthogonal sum [17]. Considering alternative rules would not solve the issue. Indeed, some hybrid rules (e.g., those proposed by Yager [18] or Dubois and Prade [19]) performing a dispatching of the conflict are not associative, which in turn may raise additional issues about the combination ordering of the sources. Instead of searching alternatives to the conjunctive combination rule, some authors proposed to discount the Basic Belief Assignments (BBAs) to combine so that the conflict remains under control [20], [21].

Besides the choice of a tailored combination rule, a large number of sources raises the issue of the presence of unreliable ones. Indeed, the higher the number of sources, the more likely is the fact that one or some of them are unreliable. Such sources are called 'outliers' for the combination since they are incompatible/inconsistent with the remainder of the sources. Some authors have proposed algorithms to handle sets of sources including outliers, either by extending the q-relaxation [22] proposed for the Interval Analysis to BFT [20], or by extending RANSAC [9] to BF [13]. In the first case, the combination rule is modified to be robust to the presence of outliers, making it however intractable in the case of a large number of outliers (the q parameter being usually in the range of a few units). In the second case, having explicitly estimated a set of inliers (conversely to q-relaxation), the conjunctive rule is used however with a number of sources ranging in the tens.

Finally, [23] propose to handle a large number of sources by clustering BBAs and firstly combining the BBAs that belong to a same cluster. In their work, using the canonical decomposition, clusters are simply defined as sets of Simple Support Functions (SSF) having the same focal elements so that their combination is straightforward and also produces a SSF. Then, intra-cluster resulting SSFs are discounted with respect to the number of initial SSFs in the cluster. However, such an approach assumes that the canonical decomposition of initial BBAs involves a small set of SSFs, so that each element may appear several times when considering the different initial BBAs, which is clearly not the case when considering a large 2D discernment frame. Thus, even if in this work, we keep the general idea of BBA clustering that was already proposed by [24], both the clustering criterion and the use of clustering results are different. Indeed, in few words, firstly BBA clusters are derived using a hierarchical clustering based on Jousselme's distance that allows for taking into account focal element interactions. From clustering construction, these clusters correspond to possible but incompatible solutions for the epipole localization. Then, secondly, BBAs are combined in a conjunctive way only within clusters to provide cluster-BBAs that are ranked. We then show that the correct solution is among the top ranked clusters.

The remainder of this paper is as follows: Section II recalls the basics (including belief function tools) used for this study, Section III describes the proposed approach that provides a set of ordered solutions and Section IV analyzes the results obtained on a public dataset before Section V draws the main conclusions and perspectives of our work.

II. Background knowledge

In this section, we recall the BFT background notions which are required for our study.

A. Basics on BFT

Let Ω be the considered discernment frame, i.e. the set of mutually exclusive solutions of our problem. Belief Function Theory allows us to handle imprecision along with uncertainty thanks to functions defined on the Ω power set called 2^Ω, i.e. the set of Ω subsets. There are five main belief functions that are in one-to-one relationships so that the knowledge of one is sufficient to derive any other of them. Usually, the mass function m is considered as the BBA representing knowledge from a given source. It allows us to set a well-defined BBA provided that m satisfies two constraints: (i) $\forall A \in 2^\Omega, m(A) \in [0,1]$ and (ii) $\sum_{A\in 2^\Omega} m(A) = 1$. Note that the set of hypotheses having a non null mass value is called the focal set or set of focal elements. Apart from m, the plausibility (Pl) and commonality (q) functions are widely used, for decision and for computation respectively. They are related to m as follows: $\forall A \in 2^\Omega$, $Pl(A) = \sum_{B\in 2^\Omega, A\cap B\neq\emptyset} m(B)$, $q(A) = \sum_{B\in 2^\Omega, A\subseteq B} m(B)$.

Different kinds of criteria have been proposed to compare two BBAs, m_1 and m_2. Firstly, several orderings have been established between BBAs: pl-ordering ($m_1 \sqsubseteq_{pl} m_2 \Leftrightarrow \forall A \in 2^\Omega, Pl_1(A) \leq Pl_2(A)$), q-ordering ($m_1 \sqsubseteq_q m_2 \Leftrightarrow \forall A \in 2^\Omega, q_1(A) \leq q_2(A)$), s-ordering and w-ordering [17]. Secondly, various distances or dissimilarity measures between BBAs have been proposed [25]. In this study, we will consider the Jousselme's one for its simplicity, understandable behavior and well-established mathematical properties. It is based on the scalar product definition given by Eq.(1): denoting by $|H|$ the cardinality of any Ω subset H, $\forall (i,j) \in \{1,2\}^2$,

$$\langle m_i, m_j \rangle_J = \sum_{A\in 2^\Omega} \sum_{B\in 2^\Omega} \frac{|A\cap B|}{|A\cup B|} m_i(A)\, m_j(B), \quad (1)$$

such that Jousselme's distance $d_J(m_1,m_2)$ between m_1 and m_2 is equal to $\sqrt{\frac{1}{2}(\langle m_1,m_1\rangle_J + \langle m_2,m_2\rangle_J - 2\langle m_1,m_2\rangle_J)}$.

Whenever several sources of information are available, they are generally combined in a conjunctive way that boils down to assuming that every source is reliable, and to corroborating themselves in order to decrease the imprecision and the uncertainties. Among the most popular conjunctive rules, we can quote the Smets' conjunctive rule [16] (cf. Eq.(2)), its normalized version [17], and Denœux's cautious rule [26].

The first two rules assume cognitive independence between sources whereas the last one can handle correlated sources.

$$\forall A \in 2^{\Omega}, m_{1 \textcircled{\cap} 2}(A) = \sum_{B \in 2^{\Omega}} \sum_{\substack{C \in 2^{\Omega}, \\ B \cap C = A}} m_1(B)\, m_2(C). \quad (2)$$

As combinations are performed, the belief becomes more fragmented across more focal elements (FE). Then, mainly for numerical reasons, the BBA has to be approximated to keep a controlled number of FEs. This process is often called BBA simplification, and in the perspective of further combination, approaches providing a generalization of the initial BBA are favored, in particular those aggregating some FEs that follow the least commitment principle. Among the many methods proposed to choose the elements to be aggregated, iterative aggregation techniques [27] are based on a selection criterion involving a quantitative measure of the BF approximation: e.g., precision measure [28] is used in [27] whereas Jousselme's distance is used in [11]. This latter case boils down to choosing the two FEs of a given BBA minimizing Eq. (3)

$$\begin{aligned} d_J^2(A, B \mid m) = & \left(1 - \frac{|A|}{|A \cup B|}\right) m^2(A) + \\ & \left(1 - \frac{|B|}{|A \cup B|}\right) m^2(B). \end{aligned} \quad (3)$$

Then, for BBA approximation, a pair of hypotheses is chosen iteratively and their masses are aggregated until the desired number of FEs is reached. Note that, if performed along with BBA combination, such a simplification process breaks unfortunately the associativity (if it existed) of the combination.

Finally, after having combined all sources through their BBAs, a decision can be taken. It is generally done in the discernment frame Ω, i.e. only considering singleton elements so that two widely used criteria are (i) the contour function (that is given by the plausibility function restricted to Ω elements and normalized) and (ii) the pignistic probability [29]:

$$\forall H \in \Omega, BetP(H) = \frac{1}{1 - m(\emptyset)} \sum_{A \in 2^{\Omega}, H \in A} \frac{m(A)}{|A|}. \quad (4)$$

B. The case of a 2D discernment frame

The open source[1] library 2CoBel [14] has been developed in the applicative context of crowd monitoring. Indeed, for such application, the localization has to be all the more precise that the crowd is dense. Therefore, a fully scalable library has been developed for 2D discernment frames, in which FEs are represented by polygons. Specifically, FEs are represented by sets of vertices, allowing both for FEs having multiple connected components and for FEs having holes (distinguishable by the ordering of the vertices). Besides, using a hashing table allows for fast identification of FEs already encountered when performing summations in combination rules or uncertainty propagation operators.

[1]Implementation available at: https://github.com/MOHICANS-project/2CoBel

Note also that distances between BBAs can be very easily derived thanks to clipping operators that compute areas or intersection or union between two polygons.

Besides this geometrical representation allowing for handling precise shapes of 2D hypotheses of interest, 2CoBel [14] provides an useful and compact representation of the interactions between FEs under the form of a Directed Acyclic Graph (DAG). In brief, any intersection between FEs can be represented by a path on the DAG. Now, there are two main cases in which the analysis of all the FE intersections is useful: (i) the definition of an equivalent 1D discernment frame for canonical decomposition (that we just mention since we do not use it), (ii) the computation of the decision criterion. Let us specify this last point. Both maximizations of the contour and BetP functions boil down to comparing mass accumulation on paths of maximal length. Indeed, denoting by $\mathcal{F}$ the set of FEs, $\forall (P, P') \subseteq 2^{\mathcal{F}} \times 2^{\mathcal{F}}$ such that $P \subset P'$ and $\cap_{A_i \in P} = \cap_{A_i \in P'}$, we have $Pl(P') = Pl(P) + \sum_{A_i \in P' \setminus P} m(A_i) > Pl(P)$ and $BetP(P') = BetP(P) + \frac{1}{1-m(\emptyset)} \sum_{A_i \in P' \setminus P} \frac{m(A_i)}{|A_i|} > BetP(P)$. Therefore, decisions have to be taken among paths representing non empty intersections and having maximal length on the DGA. Now, since the systematic exploration of the whole graph may be numerically expensive, [14] provides some tricks for efficient exploration, in particular early avoiding non maximal length paths (by detection of subpath features).

A keypoint is that the maximal length paths corresponding to non-empty intersections are the most precise hypotheses that we can consider given the BBA m. They are among the singleton hypotheses when considering the equivalent 1D discernment frame used for instance for canonical decomposition as proposed by [14]. Then, we are able to decide in favor of compound hypotheses of a 2D discernment frame, maximizing the $BetP$ criterion (or the contour function) in an equivalent 1D discernment frame.

C. Basics on epipole computation

Let us recall the usual way one computes the epipole localization and its uncertainty. The principle of RANSAC is to subsample the whole solution space (to reduce exploration complexity). In [9], the sampled solution correspond to exact solutions only considering a randomly drawn subset of observations. Then, at the end, the solution selected by RANSAC is the one that is the most consensual, i.e. that induces the highest number of inliers defined as measurements presenting noise level lower than a given threshold (that is a parameter of the algorithm). Applied to our problem, it means that providing the set of putative matches $\mathcal{I}$, at iteration i, RANSAC will sample a 8-tuple of pixel matches from $\mathcal{I}$, derive the fundamental matrix F_i (provided that the 8-tuple does not correspond to a degenerated system), and evaluate the consensus degree associated to this solution F_i, before reiterating independently. The output of RANSAC includes thus (i) the inlier set having greatest cardinality along with (ii) the corresponding solution ($\hat{F}$). Usually, this latter is re-estimated from the whole inlier set. However, in some

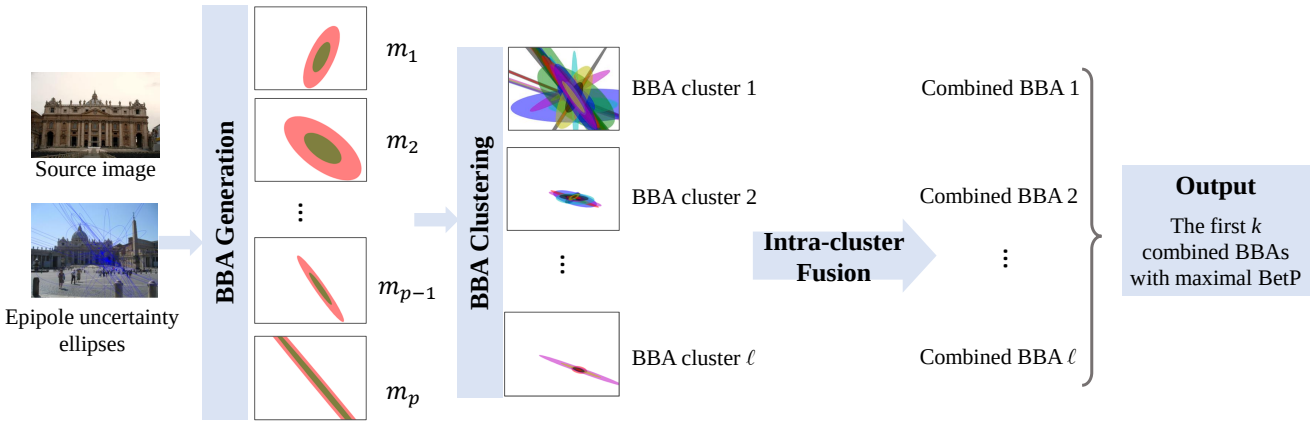


Fig. 1: Overview of the proposed method. First, p sources for epipole uncertainty estimation are generated based on the observations, then clustered by aggregation in l groups. Each group provides a solution as a combined BBA in a set which is ranked according to the pignistic probability BetP. In our application, we select the top k solutions.

applications (such as ours), it is preferable to keep the initial estimation (from which the consensus degree was evaluated).

Then, for any solution of the fundamental matrix F, its covariance matrix Σ_F is derived as in [8]. The epipole location e as well as its covariance matrix Σ_e are then derived based on the constraint $Fe = 0$ and singular value decomposition (SVD) [30]. Geometrically, the epipole location uncertainty has the shape of an ellipse whose axes are defined by eigenvectors and eigenvalues such that $(x-e)^T\Sigma^{-1}(x-e) = k^2$, with k defined by the considered confidence level.

III. The Proposed Algorithm

In the proposed application, a major issue comes from the presence of outliers (in the keypoint matches) that prevents the correct estimation of the epipole. Various outlier rejection techniques fail in difficult configurations where, for instance, outliers have a strong majority. Then, the basic idea of our approach is to introduce a mutual validation test for any potential solution, based on the consistency among several solutions obtained independently, for instance using either different algorithms or, as in this case, different observations. Note that this idea is the very core of ensemble approaches.

In this work, we propose to obtain several evidences of the solution by considering various solutions provided by RANSAC (instead of retaining only the most consensual). Specifically, let $S_F = \{F_1, F_2, \ldots, F_i, \ldots, F_n\}$ be the set of the n tested solutions ranked in decreasing order according to their consensus value (namely the size of their inlier set $\mathcal{I}_i$). Recall that these solutions have been obtained independently by random drawing of 8-tuple in $\mathcal{I}$ (cf. Section II-C). Following [31], we will consider the p first elements in S_F with p derived with respect to threshold $\theta \in (0,1)$ such that $\frac{|\mathcal{I}_p|}{|\mathcal{I}_1|} \geq \theta > \frac{|\mathcal{I}_{p+1}|}{|\mathcal{I}_1|}$. Then, for each $F_i \in \{F_1, \ldots, F_p\}$, we derive the associated epipole location along with its uncertainty ellipse. Firstly, note that these additional steps do not increase the RANSAC complexity, since in basic RANSAC we also perform these draws and evaluate them; what differs here is the fact that instead of discarding them (except the best one) we save them for further processing. Due to the presence of erroneous keypoint matches, some of these ellipses do not include the true epipole location. In the following, we call outlier evidences such ellipses.

Then, our purpose is to provide as accurately as possible an estimation of the epipole location by combining epipole location evidences despite the presence of outliers. Among these p solutions, some of them exhibit consistency in terms of epipole location and others not. Furthermore, there may be different groups of consistent solutions. A direct accumulation of all these solutions does not allow to capture these characteristics. In this work, we propose to first cluster all considered solutions into different groups in terms of consistency and then explore the fusion strategy within each group. To respect the consistency inside the group and to highlight the inconsistency between different groups, the core idea is to keep individual solutions provided by each group instead of imposing a final fusion between different groups. The overview of the proposed method is illustrated in Fig. 1.

We use the 2CoBel library introduced in Section II-B for modelling epipole location evidences in terms of beliefs associated to 2D polygones (approximation of ellipses). Specifically in this work, each individual ellipse leads to a consonant BBA having two nested focal elements corresponding respectively to two uncertainty levels (95% and 50% in our case). Note that we chose these uncertainty levels in a consistent way with respect to previous works [32], [33]. Mass values associated to these focal elements will be discussed along with the experiments (Section IV). Each of these BBAs is very little committed so that the precise epipole location will comes from their combination.

A. Clustering

Given a set of BBAs $\mathcal{M} = \{m_1, m_2, ..., m_p\}$, we aim to cluster them without prior knowledge about the number of clusters. In this study, we focus on hierarchical clustering [34] for its agglomerating feature. This latter is based on distances

between samples (to cluster) or between samples and clusters so that samples/clusters are gathered according to increasing distance order. Note that, for the cluster distance, different criteria have been proposed: simple [35], average [36], complete [37].

Regarding BBA handling, we consider Jousselme's distance so that the pre-computed distance matrix $D_{p\times p}$, where $D_{ij} = d_J(m_i, m_j)$. Besides, in the perspective of conjunctive combination of BBAs belonging to a given cluster, we consider complete linkage which uses the $\max$ operator for computing the distance between two clusters from the distance values between samples. The theoretical threshold guarantying that two BBAs have at least one pair of focal elements intersecting (avoiding total conflict) can be computed as

$$\begin{aligned} d_{th} &= \sqrt{\frac{1}{2}(\langle m_i, m_i\rangle + \langle m_j, m_j\rangle)} \\ &= \sqrt{a^2 + (1-a)^2 + 2a(1-a)\frac{k_{50}^2}{k_{95}^2}}, \end{aligned} \tag{5}$$

where k_{95} corresponds to a confidence level equal to 95% in ellipse derivation (cf. Section II-C) and k_{50} to 50% confidence level, a and $(1-a)$ are the values of masses associated with the focal elements at 50% and 90% confidence levels, respectively. In our experiments, we set the maximum distance for clustering slightly below the theoretical value (namely, $d_{th} - 0.05$) in order to increase the consistency between BBAs in the same cluster. In addition, since even a complete linkage cannot guarantee that there is a common intersection for all BBAs in the same cluster (since only pairs of BBAs are considered), we set as a supplementary constraint that the intersection between all the largest focal elements of the BBAs within a given cluster is not empty (called "non empty BBA intersection" by language shortcut).

Finally, the clustering criterion boils down to the minimisation of both the cluster number and the intra-cluster distance under two constraints, namely the intra-cluster distance being lower than $d_{th} - 0.05$, and the non empty intersection between cluster elements. In the following, let l denote the number of obtained BBA clusters and $\{\mathcal{M}_1, ..., \mathcal{M}_i, ..., \mathcal{M}_l\}$ the set of clusters. This latter is a partition of $\mathcal{M}$, namely it satisfies (i) $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset, \forall (i,j) \in \{1,\ldots,l\}^2, i \neq j$ and (ii) $\cup_{i\in\{1,...,l\}}\mathcal{M}_i = \mathcal{M}$.

B. Fusion strategies

We now aim to combine the BBAs within each cluster using the conjunctive rule [16]. However, for clusters including more than a few ten BBAs, a step of BBA approximation has to be implemented (cf. Section II) to control computational complexity. Specifically, we introduce a step of BBA approximation each time the number of focal elements is larger than 20 in order to decrease it to 10. The used BBA approximation process is the same as in [11].

However, one issue introduced by the BBA approximation is the loss of the associativity of the combination. Let us then discuss the combination order since the result will depends on it. On the one hand, it may appear as more natural to gather first the closest BBAs (still according to Jousselme's distance) so that combination ordering follows distance one. On the other hand, one may remark that, since BBA approximation decreases BBA commitment, the BBAs combined in the end may have a greater impact in the final BBA. One can also wonder whether it is worth recomputing the distance with updated cluster BBA after each combination or if a preordering can be defined from the initial BBA distances.

In this study, we have experimented those different strategies and the two more efficient are presented in Section IV. They correspond to updated minimal (respectively maximal) distance ordering. Let $\mathcal{F}(m_i)$ denotes the set of focal elements of a BBA m_i and let $\mathcal{F}(m_i) \cap \mathcal{F}(m_j) = \{A \cap B\}_{(A,B)\in\mathcal{F}(m_i)\times\mathcal{F}(m_j)}$. Then, fusion ordering is performed as follows. The two first BBAs to combine are:

$$(i^*, j^*) = \underset{\substack{(i,j)\in\mathcal{F}(m_i)\times\mathcal{F}(m_j)\\ \mathcal{F}(m_i)\cap\mathcal{F}(m_j)\neq\emptyset}}{\arg\min} d_J(m_i, m_j); \tag{6}$$

whereas the next BBA to combine to the current BBA combination result $\tilde{m}$ is

$$j^* = \underset{\substack{(j\in\mathcal{F}(m_j)\\ \mathcal{F}(\tilde{m})\cap\mathcal{F}(m_j)\neq\emptyset}}{\arg\min} d_J(\tilde{m}, m_j); \tag{7}$$

Equations (6) and (7) correspond to min ordering. The max ordering is obtained replacing $\arg\min$ by $\arg\max$ in them.

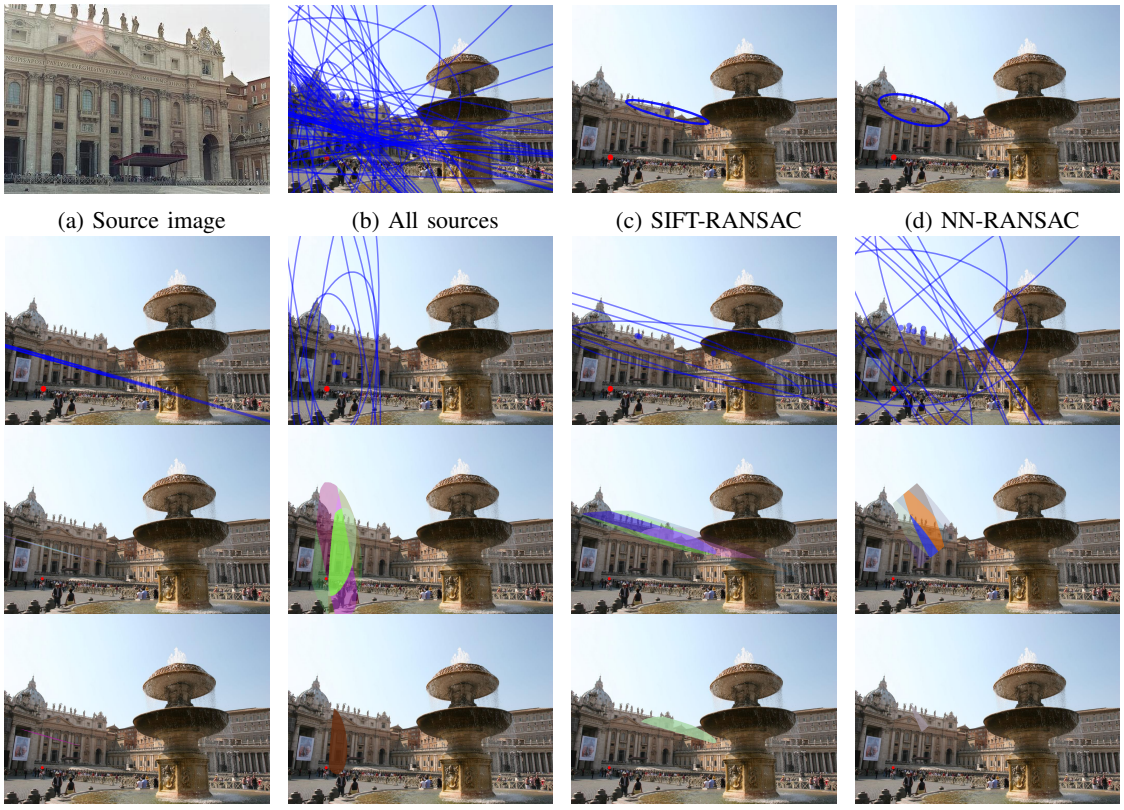
After the fusion step, we have derived l cluster BBAs $\{\tilde{m}_i\}_{i\in\{1,...,l\}}$ that can be ranked according to their maximum $BetP_i$ value ($max_{A\in\mathcal{F}(\tilde{m}_i)} BetP_i(A)$). Recalling that our aim in this study is to provide a set (as small as possible) of solutions (possibly under the form of BBA) including the ground truth, for result evaluation, we consider the k first cluster BBAs as the proposed solution set.

Finally, note that our approach differs from [23], in which the cardinality of a given cluster is used as an index of reliability (the higher the cardinality, the more reliable the cluster is assumed). For our application, such an assumption cannot be justified so that we will evaluate each cluster independently of their cardinality.

IV. Experiments and results

To evaluate the performance of the proposed method, we selected 128 pairs of images with various pose variations from the public dataset used in [38]. The number of inliers between each pair of image is at least larger than 15 and above 20% of putative matches. We obtain the ground truth epipole location from the calibration information provided with the dataset, which was computed using Structure From Motion. We compare the proposed method with the standard RANSAC method based on traditional features (**SIFT-RANSAC**) and based on the learned features (**NN-RANSAC** [38]).

We implement the proposed method in three steps. For each pair of images, we first follow the process in [31] and obtain a number of epipole uncertainty estimations. The number of iterations for sampling point matches during RANSAC is set to $n = 10^5$. In our experiments, the number of considered



(a) Source image (b) All sources (c) SIFT-RANSAC (d) NN-RANSAC

(e) BetP = 0.630 (rank 1) (f) BetP = 0.626 (rank 2) (g) BetP = 0.503 (low rank) (h) BetP = 0.299 (low rank)

Fig. 2: Qualitative illustration of our method. Upper row: the source image (a), set of epipole uncertainty ellipses (b) and the result of existing methods (c)-(d) (the ground truth is highlighted in red). (e): the top ranked cluster (ellipses, final BBA and BetP). (f): the second ranked cluster. (g)-(h): two clusters with a low rank/BetP due to the sources being less consistent.

sources p is set to be at most 100 (as long as their inlier support satisfies the condition depending on θ). Thus, note that p is not a sensitive parameter since it is determined mainly by θ ($p = f(\theta)$). The sensitive parameters are θ and k for which we provided some guideline on setting them according to our experiments. Then, we compute Jousselme's distance matrix for all the considered sources and feed it to the AgglomerativeClustering function of the module scikit-learn [39] with the complete linkage. Note that, since this function does not allow for applying an additional binary constraint, we introduce the "non-empty intersection" (cf. Section III) a posteriori, namely during the fusion step based on min distance ordering (in order to be consistent with the clustering criterion). Then, we apply the conjunctive combination for the evidences in each cluster by using the library 2CoBel following the two different fusion orders mentioned in Section III (**Fusion-min/max**). For SIFT-RANSAC and NN-RANSAC, the epipole uncertainty is derived as in [31] ("*Least squares SIFT*" and "*Least squares NN*").

Qualitative evaluation In Fig. 2 and 3, we illustrate some localization results, provided by existing methods, by top-ranked clusters and by low-ranked clusters respectively. In difficult settings, existing methods tend to be overconfident. Top ranked clusters exhibit a higher consistency among the BBAs which translates into a strong ellipse alignment, and even for challenging poses the true solution is present at the top. Conversely, the low rank clusters consist in uncertainty areas exhibiting less consistency.

Quantitative evaluation Since our output is in the form of a BBA (or an ordered set of BBAs) whereas RANSAC outputs are in the form of uncertainty ellipses (associated to 2D Gaussian distributions), for a fairer comparison, we convert these latter in a BBA. Specifically, we derive consonant BBAs having five equi-weighted focal elements, corresponding to the ellipses associated with respectively 95%, 75%, 50%, 25%, 10% confidence. We consider two different ways to compare different methods in terms of performance related to accuracy and precision.

The first evaluation counts the number of image pairs in which the ground truth epipole is included in at least one focal element of the considered BBA. Specifically, for the proposed method, if the ground truth epipole is included in at least one focal element of one of the k first evidences, we consider it as positive. The result is summarized in Table I for different values of k varying from 1 to 6 which ranks the variable number of clusters obtained for each image pair. According to this table, choosing higher values for the θ parameter allows for slightly better results. We interpret such

(a) Source image (b) All sources (c) SIFT-RANSAC (d) NN-RANSAC

(e) BetP = 0.791 (rank 1) (f) BetP = 0.724 (rank 2) (g) BetP = 0.567 (low rank) (h) BetP = 0.432 (low rank)

Fig. 3: Qualitative illustration of our method. Upper row: the source image (a), set of epipole uncertainty ellipses (b) and the result of existing methods (c)-(d) (the ground truth is highlighted in red). (e): the top ranked cluster (ellipses, final BBA and BetP). (f): the second ranked cluster. (g)-(h): two clusters with a low rank/BetP due to the sources being less consistent.

result as supporting the assertion that RANSAC filtering is beneficial, even if we relax the assumption that it is optimal. Secondly, we note the very low sensitivity of the results to the criterion min or max in the fusion. Thirdly, concerning the k parameter, the increasing and asymptotic behaviour of the number of image pairs including the ground truth is clearly visible. Note also that due to the difficulty of the geometry on some image pairs, the upper bound for performance is equal to 81 (i.e. we whecked that among the 128 image pairs, in 47 of them less than 2 ellipses among the p estimated from $F_i, i \in \{1, \ldots, p\}$ solutions include the ground truth). Finally, we specify that the results obtained using NN-RANSAC are biased by the fact that this latter has been trained on the same dataset and that much less performant results have been obtained considering other datasets.

The second evaluation consists in applying the modified metric proposed in [11] with the following definition:

$$\epsilon(\lambda) = \sum_{A \in 2^\Omega} d(e_{gd}, A) m(A) + \lambda \sum_{A \in 2^\Omega} |A|\, m(A), \qquad (8)$$

where e_{gd} is the ground truth epipole location and $d(e_{gd}, A)$ is defined as

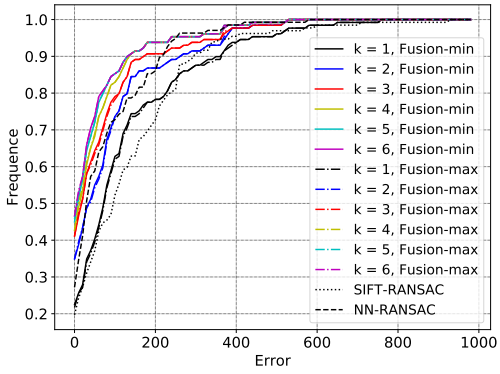
$$d(e_{gd}, A) = \begin{cases} 0 & \text{if } e_{gd} \text{ is included in } A; \\ \min\limits_{x \in \mathcal{C}_A} \| e_{gd} - x \| & \text{otherwise,} \end{cases} \qquad (9)$$

where $\mathcal{C}_A$ is the set of contour points for the focal element A. This measure allows one to control the compromise between the guarantee for the ground truth epipole belonging to the set of focal elements in the considered solution, and the imprecision related to the size of focal elements. λ is the weighting parameter between them.

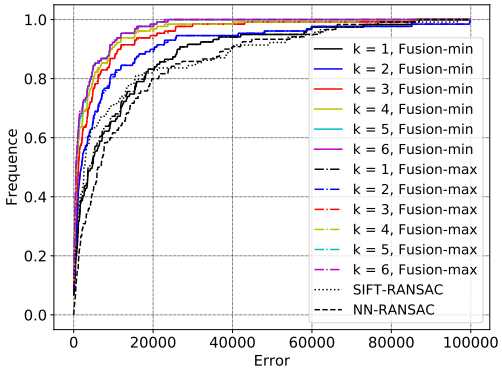
Figure 4 illustrates the different subparts of this error. From

TABLE I: Number of image pairs on which the respective method (left column) which contains the ground truth epipole. For the proposed fusion, we present results with consensus threshold values $\theta = 0.9$ and $\theta = 0.5$.

Method	#Image pairs including the ground truth					
	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
SIFT-RANSAC	39	-	-	-	-	-
NN-RANSAC	63	-	-	-	-	-
Fusion-min ($\theta = 0.9$)	31	44	55	58	63	65
Fusion-max ($\theta = 0.9$)	30	44	55	58	64	65
Fusion-min ($\theta = 0.5$)	23	40	47	49	56	60
Fusion-max ($\theta = 0.5$)	23	39	48	50	56	60

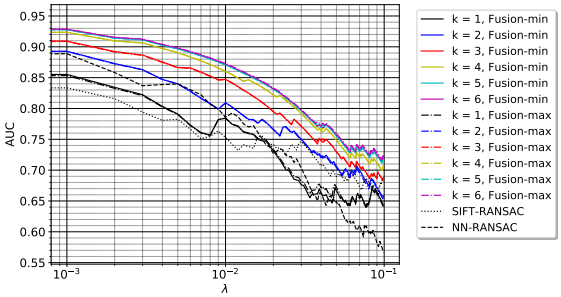


(a) Cumulative histogram of error related to $\sum_{A \in 2^{\Omega}} d(e_{gd}, A) m(A)$: $\theta = 0.9$

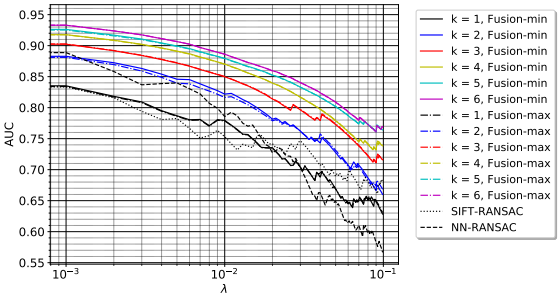


(b) Cumulative histogram of error related to $\sum_{A \in 2^{\Omega}} |A| \, m(A)$: $\theta = 0.9$

Fig. 4: Cumulative curves of error subparts versus $\epsilon(\lambda)$.



(a) Consensus threshold $\theta = 0.9$



(b) Consensus threshold $\theta = 0.5$

Fig. 5: Curve of AUC for cumulative curve, versus $\epsilon(\lambda)$.

these cumulative curves and by setting the λ value, we can estimate empirically the $\epsilon(\lambda)$ cdf (cumulative density function) and compute its Area Under the Curve (AUC) value. The higher this value, the more efficient an approach is. Now, since this value depends on the λ, in Fig. 5, we plot the AUC versus λ. Specifically, for the proposed method, we consider the solution with the smallest value of $\epsilon(\lambda)$ among the proposed k solutions. It corresponds to an optimistic assumption that an additional source (GPS, person detector for synchronized data acquisitions) will allow for "good" cluster selection. The results for different methods reported in Fig. 5 underline that, as the value of k increases, the performance of the proposed fusion method improves and outperforms others methods in terms of (i) the guarantee to include the ground truth epipole as well as (ii) the localization precision based on the evaluation methods introduced above. The proposed method also exhibits a robust behavior to the fusion order according to the similar performance achieved by different fusion strategies (min and max).

V. Conclusion

Our work explores a strategy for the fusion of a large number of sources including outliers. We propose to mitigate the impact of the outlier presence by introducing a preliminary clustering process, which organizes the sources in coherent groups. This step allows for intra-cluster fusion to be performed without increasing the mass on the empty set or requiring the user to dispatch it. The resulting BBA across the source clusters may be used afterwards for fusion with additional sources of information. In our application, namely the epipole localization which is closely related to the relative pose estimation problem in computer vision, we show that the pignistic probability related to each source cluster is a good indicator of the estimation quality, and that the evidence we obtain is competitive with respect to the state of the art.

In the future, we will exploit the fact that our algorithm is less committed than the standard vision-based solutions, and thus more favorable to the use of additional sources. We intend to perform fusion in synchronized video streams based on the results provided by our algorithm and on other localization sensors (GPS or inertial data).

References

[1] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *International journal of computer vision*, vol. 80, no. 2, pp. 189–210, 2008.

[2] P. Moulon, P. Monasse, and R. Marlet, "Global fusion of relative motions for robust, accurate and scalable structure from motion," in *Proceedings of the IEEE ICCV*, 2013, pp. 3248–3255.

[3] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *CVPR*, 2016.

[4] B. Williams, G. Klein, and I. Reid, "Automatic relocalization and loop closing for real-time monocular slam," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 9, pp. 1699–1712, 2011.

[5] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part ii: Matching, robustness, optimization, and applications," *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 78–90, 2012.

[6] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[7] A. Bartoli and P. Sturm, "Nonlinear estimation of the fundamental matrix with minimal parameters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 3, pp. 426–432, 2004.

[8] R. M. Haralick, H. Joo, C.-N. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Pose estimation from corresponding point data," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1426–1446, 1989.

[9] M. A. Fischler and R. C. 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.

[10] T. Denœux, "40 years of dempster-shafer theory," *International Journal of Approximate Reasoning*, vol. 79, pp. 1–6, 2016.

[11] C. André, S. Le Hégarat-Mascle, and R. Reynaud, "Evidential framework for data fusion in a multi-sensor surveillance system," *Engineering Applications of Artificial Intelligence*, vol. 43, pp. 166–180, 2015.

[12] W. Rekik, S. Le Hégarat-Mascle, R. Reynaud, A. Kallel, and A. B. Hamida, "Dynamic object construction using belief function theory," *Information Sciences*, vol. 345, pp. 129–142, 2016.

[13] S. Zair and S. Le Hégarat-Mascle, "Evidential framework for robust localization using raw gnss data," *Engineering Applications of Artificial Intelligence*, vol. 61, pp. 126–135, 2017.

[14] N. Pellicanò, S. Le Hégarat-Mascle, and E. Aldea, "2cobel: A scalable belief function representation for 2d discernment frames," *International Journal of Approximate Reasoning*, vol. 103, pp. 320–342, 2018.

[15] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, "Interval analysis," in *Applied interval analysis*. Springer, 2001, pp. 11–43.

[16] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 5, pp. 447–458, 1990.

[17] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42.

[18] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information sciences*, vol. 41, no. 2, pp. 93–137, 1987.

[19] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Computational intelligence*, vol. 4, no. 3, pp. 244–264, 1988.

[20] F. Pichon, S. Destercke, and T. Burger, "A consistency-specificity trade-off to select source behavior in information fusion," *IEEE transactions on cybernetics*, vol. 45, no. 4, pp. 598–609, 2014.

[21] Y. Zhao, R. Jia, and P. Shi, "A novel combination method for conflicting evidence based on inconsistent measurements," *Information Sciences*, vol. 367, pp. 125–142, 2016.

[22] V. Drevelle and P. Bonnifait, "A set-membership approach for high integrity height-aided satellite positioning," *GPS solutions*, vol. 15, no. 4, pp. 357–368, 2011.

[23] K. Zhou, A. Martin, and Q. Pan, *A belief combination rule for a large number of sources*. Infinite Study, 2018.

[24] J. Schubert, "Clustering decomposed belief functions using generalized weights of conflict," *International Journal of Approximate Reasonning*, vol. 48, no. 2, pp. 466–480, 2008.

[25] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, no. 2, pp. 118–145, 2012.

[26] T. Denœux, "Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence," *Artificial Intelligence*, vol. 172, no. 2-3, pp. 234–264, 2008.

[27] ——, "Inner and outer approximation of belief structures using a hierarchical clustering approach," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 04, pp. 437–460, 2001.

[28] D. Dubois and H. Prade, "Consonant approximations of belief functions," *International Journal of Approximate Reasoning*, vol. 4, no. 5-6, pp. 419–449, 1990.

[29] P. Smets and R. Kennes, "The transferable belief model," *Artificial intelligence*, vol. 66, no. 2, pp. 191–234, 1994.

[30] T. Papadopoulo and M. I. Lourakis, "Estimating the jacobian of the singular value decomposition: Theory and applications," in *European Conference on Computer Vision*. Springer, 2000, pp. 554–570.

[31] H. Chen, E. Aldea, and S. Le Hégarat-Mascle, "Determining epipole location integrity by multimodal sampling," in *Proceedings of the 16th IEEE International Conference on AVSS, The 3th International Workshop on Traffic and Street Surveillance for Safety and Security (IWT4S)*, 2019.

[32] R. Raguram, J.-M. Frahm, and M. Pollefeys, "Exploiting uncertainty in random sample consensus," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2074–2081.

[33] J. Lawn and R. Cipolla, "Reliable extraction of the camera motion using constraints on the epipole," in *European Conference on Computer Vision*. Springer, 1996, pp. 161–173.

[34] W. H. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of classification*, vol. 1, no. 1, pp. 7–24, 1984.

[35] R. Sibson, "Slink: an optimally efficient algorithm for the single-link cluster method," *The computer journal*, vol. 16, no. 1, pp. 30–34, 1973.

[36] H. K. Seifoddini, "Single linkage versus average linkage clustering in machine cells formation applications," *Computers & Industrial Engineering*, vol. 16, no. 3, pp. 419–426, 1989.

[37] D. Defays, "An efficient algorithm for a complete link method," *The Computer Journal*, vol. 20, no. 4, pp. 364–366, 1977.

[38] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," *CVPR*, pp. 2666–2674, 2018.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.