Belief Functions Clustering for Epipole Localization

Huiqin Chen, Sylvie Le Hégarat-Mascle*, Emanuel Aldea

SATIE Laboratory, UMR 8029, Paris-Saclay University

Abstract

This work deals with the clustering of information sources for epipole estimation in a multi-camera system. For this problem, each pair of matched visual features in the images can be considered as an elementary information source. The epipole is then estimated by combining these elementary sources taking into account their inadequacy, in particular large imprecision and presence of outliers, as well as the very large number of sources. We address the challenges introduced by a large number of sources with a strategy based on clustering and intra-cluster fusion using the Belief Functions framework. When evaluated on real data, the proposed algorithm exhibits more robustness in terms of accuracy and precision than the standard approaches which provide singular solutions.

Keywords: Clustering, Belief Function Theory, Localization, Multi-camera system

2010 MSC: 00-01, 99-00

1. Introduction

Multi-camera systems are more and more used since they can address complex tasks such as 3D reconstruction [1, 2, 3], localization [4] or navigation [5]. Now, when these cameras (or a subset of the cameras system) are moving (embedded on a pedestrian or vehicle), their localization is a key information of interpreting their images and processing them with respect to other data. To

*Corresponding author

Email address: sylvie.le-hegarat@u-psud.fr (Sylvie Le Hégarat-Mascle)

localize a given camera in the field of view of another one, one can try to directly detect its carrier. However, in practice, such a detection is ambiguous once there are several similar carriers in the scene, such as in the case of wearing a camera in a crowd, or within a fleet of drones or a vehicle network. Complementary localization evidences have thus to be considered in order to raise the ambiguities. Given a pair of cameras $\{C_1, C_2\}$, the *epipole* (of C_2 in C_1) is the 2D projection of the C_2 optical center in the image plane of C_1. The epipole location is such an evidence for carrier localization since it indicates the position of camera C_2 in the image provided by camera C_1.

Epipole localization is closely related to the relative pose estimation between two cameras, defined as the 3D rotation and 3D translation (six degrees of freedom) to relate the respective positions of the cameras. Despite the fact that this latter problem has been studied extensively for more than 30 years [6, 7, 8], there is still ongoing work in order to improve the achieved performance in adverse conditions introduced by wide baselines, large non-salient areas or repetitive structures specific to urban settings [9, 10, 11, 12]. The difficulty stems from the fact that the proposed solutions rely on the detection of keypoints in each view and their association to form pairs of *matched* keypoints. However, in practice, the derived set of matches contains a significant ratio of outlier matches which skew the solution. Despite the existence of robust estimation methods, such as those based on the very popular RANSAC [13] principle, one may still experience failures in the aforementioned adverse conditions where the outlier ratio raises generally above 50%. On the other hand, ensemble approaches have promoted the idea of considering several estimations in order to mitigate the impact of a few erroneous ones. In the case of epipole localization, such an idea has been developed in [14] using a voting strategy. However, in difficult settings, the correct location may be supported by only few estimations so that a more sophisticated modeling and combination within the ensemble is required.

In this work, we focus on the Belief Function Theory (BFT) framework. This formalism was made popular by various real-world applications [15, 16, 17, 18, 19, 20] for which it provides an efficient modelling of imprecise information,

allowing for fairer and more consistent decisions. However, for real applications, BFT scalability raises some challenges, 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 having cardinality $|\Omega|$, there are potentially $2^{|\Omega|}$ hypotheses to consider. For localization applications, DF corresponds to possible positions of the carrier, e.g., for epipole, typically $|\Omega| = 10^6$ pixels in C_1 image assuming C_2 is included in its field of view. First solutions [17] use some tricks (e.g. conditioning) to consider only a DF subset at once. Then, the authors in [21, 22, 23] propose to avoid the 2^{Ω} element enumeration by only considering the elements of the focal set (that is usually a small subset of 2^{Ω}) provided that we are able to handle them through their own description. Specifically, in [21, 22], the focal elements are described as sets of rectangles (tiles) similarly to the representation used in Interval Analysis [24], whereas [23] 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 been redefined in an efficient way.

Secondly, considering a large number of sources, their combination may become challenging. Indeed, using the very popular conjunctive rule proposed by Smets [25], the mass on the empty set ($m(\emptyset)$), usually called degree of conflict, is an increasing function with respect to the number of combined BF. Considering alternative rules would not solve the issue: Dempster's rule or the orthogonal sum [26] hides potential conflict between sources (e.g. as in the case of the Zadeh example), some hybrid rules (e.g., those proposed by Yager [27] or Dubois and Prade [28]) performing a dispatching of the conflict are only quasi-associative [29], which in turn may raise additional issues about the combination ordering in presence of very conflictual sources. Thus, instead of searching alternatives to the conjunctive combination rule, some authors proposed to discount the Basic Belief Assignments (BBAs) so that their degree of conflict remains under control [30, 31]. However, applying global or semi-global corrections to

the source BBAs may be irrelevant when source reliability is highly variable. Indeed, considering a large number of sources also raises the issue of the presence of unreliable ones: the higher the number of sources, the more likely it is that some of them be unreliable. Such sources are outliers for the combination since they are inconsistent with the remainder of the sources. Proposed algorithms to handle some sets of sources including outliers either extend the q-relaxation [32] proposed for the Interval Analysis to BFT [30], or extend RANSAC [13] to BF [22, 33]. 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 the set of inliers, the conjunctive rule may be used provided that the number of sources ranges in the tens, which nevertheless remains much beyond the number of sources we aim at considering for epipole localization.

As far as we know, the only work actually handling a large number of sources is [34]. It proposes a two-step combination based on BBAs clustering. Specifically, using the canonical decomposition, the clusters are 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, cluster SSFs are discounted with respect to the number of initial SSFs in the cluster. However, such an approach has very restrictive hypotheses, such as the fact that the canonical decomposition of initial BBAs involves only a small set of SSFs, which is clearly not the case when considering a large 2D discernment frame.

In summary, for our topical application, the main issue comes from the fact that we have both a large solution space (and thus discernment frame) and a large number of pieces of evidence including a high ratio of outliers. Even if some previous works have provided partial solutions, none of them handle both scalability issues together. In this work, we keep the general idea of BBA clustering that was already proposed by [35], but both the clustering criterion and the use of clustering results are tailored with respect to our application. BBA clusters are firstly 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. Secondly, BBAs are combined in a conjunctive way only within clusters to provide cluster-BBAs that are ranked so that the correct solution is expected to appear among the top ranked clusters. In order to illustrate the general concept introduced by our work, we will consider different sources of evidence which may arise in localization applications. The baseline scenario consists in a pair of images providing exclusively visual cues via keypoint association. Then, a more complex setting considers additional evidences provided by a pedestrian (i.e., carrier) detector and an exteroceptive sensor. Finally, a third scenario involves static cameras within a dynamic scene, in which the temporal dimension provides the means for the accumulation of evidences.

The remainder of this paper is as follows: in Section 2 we recall the basics (including belief function tools) used for this study, then Section 3 describes the proposed approach that provides a set of ordered solutions. In the next sections, we propose algorithms for the exploitation of the set of ordered solutions, in a multi-source fusion task (Section 4), and in a multi-temporal fusion task (Section 5) respectively. Section 6 analyzes the results obtained on a public dataset before Section 7 draws the main conclusions and perspectives of our work.

2. Related background

2.1. Basics on Belief Function Theory (BFT)

Let us denote by Ω the considered discernment frame, i.e. the set of mutually exclusive solutions of our problem and by 2^{Ω} the Ω power set, i.e. the set of Ω subsets. BFT allows us to handle imprecision along with uncertainty thanks to five main functions defined on 2^{Ω}. Since these functions are in one-to-one relationships, the knowledge of one is sufficient to derive any other of them: usually, the mass function m corresponds to the basic belief assignment (BBA) representing knowledge provided by a given source. It satisfies two constraints:

(i) $\forall A \in 2^\Omega, m(A) \in [0,1]$ and (ii) $\sum_{A\in 2^\Omega} m(A) = 1$. A hypothesis A having a non null mass value, $m(A) > 0$, is a Focal Element (FE) and the set of such hypotheses is called the focal set and denoted $\mathcal{F}(m)$. In this study, we will also refer to the disjunction of all focal elements, which will be denoted $\bigcup_m$ with $\bigcup_m = \bigcup_{A\in\mathcal{F}(m)} A$. 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 criteria have been proposed to compare two BBAs, m_1 and m_2. Firstly, several orderings between BBAs have been established: e.g., pl-ordering or q-ordering ($m_1 \sqsubseteq_f m_2 \Leftrightarrow \forall A \in 2^\Omega$, $f_1(A) \leq f_2(A)$, $f \in \{pl, q\}$), s-ordering and w-ordering [26]. However, whatever the considered ordering, it is only partial. Secondly, various distances or dissimilarity measures between BBAs have been proposed [36]. In this study, we will consider the Jousselme's one for its simplicity, interpretable results and well-established mathematical properties. It is based on the scalar product definition given by Eq. (1): denoting by $|H|$ the cardinality of any hypothesis H, m_1 and m_2 being two BBAs, $\forall (i,j) \in \{1,2\}^2$,

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

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

Now, if we have several BBAs defined on the same discernment frame and assume every source is reliable, we aim at corroborating themselves in order to decrease the imprecision and the uncertainties, which is achieved by combining them in a conjunctive way. Among the most popular conjunctive rules, let us cite the Smets' conjunctive rule [25] (cf. Eq.(2)), its normalized version [26], and Denœux's cautious rule [37]. The first two rules assume cognitive independence between sources whereas the last one can handle correlated sources.

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

As conjunctive combinations are performed, the belief becomes more fragmented across more FEs. To keep the number of FEs under control, mainly for numerical reasons, the BBA has to be simplified by approximating it. Now, in the perspective of further combination and following the Least Commitment Principle (LCP), the approaches providing a generalization of the initial BBA are favored, in particular those aggregating some FEs. Then, the iterative aggregation techniques [38] are based on a selection criterion involving a quantitative measure of the BF approximation: e.g., precision measure [39] in [38] or Jousselme's distance in [17]. This latter case boils down to choosing the two FEs in $\mathcal{F}(m)$ minimizing Eq. (3) (cf. Appendix),

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

Until the desired number of FEs is reached, the BBA approximation iterates: (i) the choice of the pair of FEs to merge and (ii) their aggregation in a single FE gathering their masses. Note that, if the BBA approximation is performed along with associative BBA combination, such a simplification process breaks unfortunately the associativity of the combination (in addition to breaking the strictly conjunctive nature of the whole process, in order to comply with LCP).

Finally, after having combined all sources through their BBAs, a decision can be taken. It is generally done in Ω, i.e. only considering singleton elements so that two widely used criteria are (i) the maximization of the contour function (that is given by the plausibility function restricted to Ω elements and normalized) and (ii) the maximization of the pignistic probability [40]:

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

2.2. The case of a 2D discernment frame

The open source[1] library 2CoBel [23] has been developed in the applicative context of pedestrian monitoring in dense crowds, i.e. with a requirement of

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

precise localization on the ground plane. It is a fully scalable library for 2D discernment frames, in which FEs are represented by polygons coded as sets of ordered vertices, allowing for FEs with multiple connected components and for FEs with holes. Using a hashing table allows then for fast identification of FEs already encountered when performing summations in combination rules for instance (cf. Eq. (2)). Also, distances between BBAs can be very easily derived thanks to clipping operators that compute areas or intersection or union between two polygons.

Along with this geometrical representation of 2D FEs, 2CoBel [23] provides an useful and compact representation of the interactions between FEs under the form of a Directed Acyclic Graph (DAG). This allows for the representation of any intersection between FEs as a path on the DAG. In our case, these intersections that are required for the computation of the decision criterion. Indeed, the FE representation as polygon (or set of polygons) brings scalability while requiring a new definition of the finest distinguishable hypotheses based on the concept of Maximal Intersection [23]. Specifically, a Maximal Intersection is a hypothesis that corresponds to a non-empty intersection between FEs such that its intersection with any different FE would lead to an empty intersection. Note that there is no pair of elements of type Maximal Intersection having an inclusion relationship. Then, the decisions are taken within the Maximal Intersection set. Using the DAG to represent a Maximal Intersection as a path connecting the involved FEs, both maximizations of the contour and BetP functions boil down to comparing mass accumulation on paths of maximal length: $\forall (P, P') \subseteq 2^{\mathcal{F}(m)} \times 2^{\mathcal{F}(m)}$ 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, optimal decisions correspond to Maximal Length Paths (MLP), i.e. paths representing non empty intersections and having maximal length on the DAG. Now, since the systematic exploration of the whole graph may be numerically expensive, [23] provides tools for efficient exploration, e.g. early avoiding of non maximal length paths (by detection of subpath features).

Finally, note that the justification for Maximal Intersections being the most precise hypotheses that we can consider given the BBA m is supported by the fact that they define the singleton hypotheses when considering an equivalent 1D discernment frame [23]. Therefore, decisions which do not favor larger compound hypotheses are consistent with the usual definition of decision functions on Ω instead of 2^{Ω}.

Likewise, decisions which do not refine the localization within MLP hypotheses are consistent with the BFT least commitment principle. However, if for a specific application the solution has to be represented as a point (or pixel), taking the barycenter of the decided MLP hypothesis appears as the least bad option.

2.3. Basics on epipole estimation

As explained in Section 1, the epipole localization and its uncertainty derives from the relative pose estimation. Using keypoint matches, a popular criterion (to minimize) is the sum of errors between predicted and observed matches. To this aim, the fundamental matrix $\mathbf{F}$ is a valuable tool, as it links any point $\mathbf{x}$ viewed in the first camera with its corresponding match $\mathbf{x}'$ from the second camera [41]. Denoting the transpose operator by upper-script ^{T}, the compact constraint

$$\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0 \tag{5}$$

is extremely useful as it bypasses the need for additional 3D scene geometry information. By construction, $\mathbf{F}$ is a rank 2 matrix due to being expressed as a product of matrices, among which one is also rank 2 (see [41], Sec. 9.2.2).

Firstly, we compute the fundamental matrix $\mathbf{F}$ from the 8-point algorithm [42]. Using the formulation [43], we denote $X = \{x_i, y_i, x'_i, y'_i\}_{1 \leq i \leq 8}$ a set of $n = 8$ point matches and we derive $\mathbf{F}$ by reordering the epipolar constraint specified in Eq.(5) as a linear system with respect to the coefficients of $\mathbf{F}$, which gives

$\mathbf{Af} = \mathbf{c}$ with

$$\mathbf{A} = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 \\ x'_2x_2 & x'_2y_2 & x'_2 & y'_2x_2 & y'_2y_2 & y'_2 & x_2 & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_8x_8 & x'_8y_8 & x'_8 & y'_8x_8 & y'_8y_8 & y'_8 & x_8 & y_8 \end{bmatrix}, \tag{6}$$

where $\mathbf{f} = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32})^T$, $\mathbf{c} = -(1,1,1,1,1,1,1,1)^T$, and F_{33} is set to be 1.

Then, in order to impose the constraint of rank 2 for $\mathbf{F}$, the smallest singular value derived by SVD of $\mathbf{F}$ is forced to be 0: if $\mathbf{F} = \mathbf{UDV}^T$, then the constrained fundamental matrix $\tilde{\mathbf{F}}$ is

$$\tilde{\mathbf{F}} = \mathbf{UD} \begin{bmatrix} \mathbf{I_{2\times 2}} & \mathbf{0_2} \\ 0 & 0 \end{bmatrix} \mathbf{V}^T, \tag{7}$$

where $\mathbf{I_{k\times k}}$ and $\mathbf{0_k}$ are the identity matrix and the null vector of dimensionality k, respectively. Finally, as the epipole satisfies $\mathbf{Fe} = \mathbf{0}$, it can simply be derived as the right singular vector $\mathbf{v_3}$ where $\mathbf{V} = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} \end{bmatrix}$.

In this study we also need epipole uncertainty evaluation. Once more, the standard pipeline relies on the fundamental matrix and on uncertainty propagation by linearization [44]. Denoting the F uncertainty as Σ_F, the epipole uncertainty can be derived with forward propagation based on the linear system formulation ($\mathbf{Af} = \mathbf{c}$) as $\Sigma_{\mathbf{f}} = \mathbf{J}_X \Sigma_X \mathbf{J}_X^T$, where Σ_X is a diagonal matrix representing the uncertainty on the coordinates for each of the 8 point matches in X (in our case for sake of simplicity $\sigma \times \mathbf{I_{32\times 32}}$, with $\sigma = 1$px) and $\mathbf{J}_X$ is the Jacobian matrix of $\mathbf{f}$ with respect to the set of point matches X, which can be automatically computed with the chain rule. Then, the uncertainty on $\tilde{\mathbf{F}}$ can be derived as

$$\Sigma_{\tilde{\mathbf{F}}} = \mathbf{J}_{\tilde{\mathbf{F}}/\mathbf{f}} \begin{bmatrix} \Sigma_{\mathbf{f}} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{J}_{\tilde{\mathbf{F}}/\mathbf{f}}^T, \tag{8}$$

where $\mathbf{J}_{\tilde{\mathbf{F}}/\mathbf{f}}$ is the Jacobian matrix of $\tilde{\mathbf{F}}$ with respect to $\mathbf{f}$, whose explicit computation can be found in [43]. Finally, the epipole uncertainty $\Sigma_{\mathbf{e}}$ is

$$\Sigma_{\mathbf{e}} = \mathbf{J}_{SVD} \Sigma_{\tilde{\mathbf{F}}} \mathbf{J}_{SVD}^T, \tag{9}$$

with $\mathbf{J}_{SVD}$ the Jacobian of the SVD [45]. Geometrically, the epipole location uncertainty has the shape of an ellipse whose axes are defined by the eigenvectors and eigenvalues of $\mathbf{\Sigma_e}$ such that the bound of the confidence area is defined by the points $\mathbf{x}$ satisfying

$$(\mathbf{x}-\mathbf{e})^T\mathbf{\Sigma_e^{-1}}(\mathbf{x}-\mathbf{e}) = \kappa^2, \tag{10}$$

with κ defined by the considered confidence level.

In the previous equations, we assume the 8 point matches and their coordinates are exact. However, due to the presence of a significant amount of outliers within the set of observed matches, robust estimation techniques such as RANSAC are required.

The principle of RANSAC is to randomly draw subsets of observations (matches in our case) in order to find a subset composed only of inliers, that is recognized as such. For this, each solution derived from a drawn subset of observations is scored by its consensus degree corresponding to the number of inliers, defined as the observations (among the whole set) presenting an error lower than a given threshold (that is a parameter of the algorithm) with respect to the prediction of the considered solution. Then, the solution selected by RANSAC is the one that is the most consensual, i.e. that generates the highest number of inliers. Applied to our problem, it means that given the set of putative matches $\mathcal{I}$, at iteration i, RANSAC will sample a 8-tuple from $\mathcal{I}$, derive the fundamental matrix $\tilde{\mathbf{F}}_\mathbf{i}$ (provided that the 8-tuple does not correspond to a degenerated system), and evaluate the consensus degree associated to the $\tilde{\mathbf{F}}_\mathbf{i}$ solution, before reiterating independently. The output of RANSAC includes thus (i) the inlier set having greatest cardinality along with (ii) the corresponding solution ($\hat{\mathbf{F}}$). Usually, $\hat{\mathbf{F}}$ is re-estimated from the whole inlier set. However, in some applications, it is preferable to keep the initial estimation (from which the consensus degree was evaluated), and this will be the case for the proposed algorithm.

Note also that, for any solution of the fundamental matrix $\mathbf{F}$ (the most consensual but also some others as explained further), we will derive its covariance

matrix $\mathbf{\Sigma_F}$, the epipole location $\mathbf{e}$ as well as its covariance matrix $\mathbf{\Sigma_e}$ and the ellipses corresponding to different pre-selected uncertainty levels κ (cf. Eq.(10)).

3. Proposed belief clustering

3.1. Problem formulation

Common outlier rejection techniques fail in difficult settings where outliers have a strong majority. The basic idea of our approach is then to introduce a mutual validation test for any potential solution, based on the consistency among several solutions obtained independently. Note that this idea is the very core of ensemble approaches that aim to increase estimation robustness and accuracy by combining different algorithm outputs.

We propose to obtain several pieces of evidence as candidate beliefs on the epipole location by considering various solutions provided by RANSAC. Depending on the availability of a video stream or of only one image pair, RANSAC solutions are generated slightly differently. Whereas for a video stream, provided that the cameras are static, we can use each temporal step to derive a new estimation of epipole location, for a single image pair we propose to retain several solutions among the most consensual ones explored by the RANSAC algorithm.

Specifically, let $\mathcal{S}_\mathbf{F} = \left\{\tilde{\mathbf{F}}_\mathbf{1}, \ldots, \tilde{\mathbf{F}}_\mathbf{n}\right\}$ be the set of the n tested solutions ranked in decreasing order according to their consensus value (the cardinality of their inlier set $\mathcal{I}_i$). These solutions have been obtained independently by random drawing of 8-tuples in $\mathcal{I}$ (the set of keypoint matches, cf. Section 2.3) We consider the p first ranked elements in $\mathcal{S}_\mathbf{F}$ with p derived with respect to threshold $\theta \in (0, 1)$ such that $|\mathcal{I}_p| \geq \theta \times |\mathcal{I}_1| > |\mathcal{I}_{p+1}|$. For instance, setting $\theta = 0.9$ boils down to considering the solutions having a number of inliers greater than 90% of the most consensual inlier set. Then, for each corresponding fundamental matrix $\tilde{\mathbf{F}}_\mathbf{i} \in \left\{\tilde{\mathbf{F}}_\mathbf{1}, \ldots, \tilde{\mathbf{F}}_\mathbf{p}\right\}$, we derive the associated epipole location along with its uncertainty ellipse equation (Eq. (10)). Figure 1 illustrates this process called Multiple Model Sampling. Let us underline that this process does not

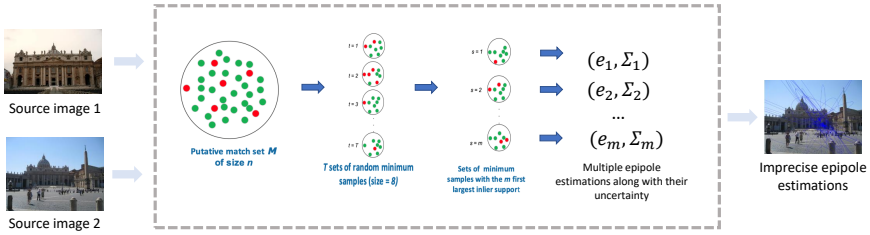


Figure 1: Multiple model sampling: From RANSAC solutions based on the observations, p models for epipole estimation are selected.

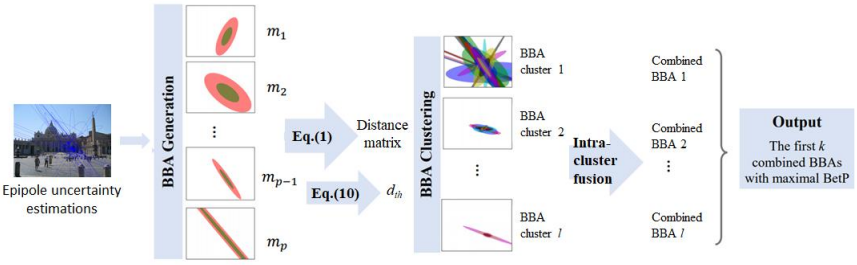


Figure 2: Overview of the proposed method. The p models interpreted as p uncertain and imprecise sources are formalized BBAs, then clustered by aggregation in l groups. Each group provides a solution as a combined BBA in a set which is ranked according to the pignistic probability BetP. In our application, we select the top k solutions.

increase the RANSAC complexity, since in standard 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.

From the set of the p putative imprecise solutions for the epipole location, we aim to derive a more accurate localization. However, due to the presence of erroneous keypoint matches, some of these ellipses do not include the true epipole location. Filtering them is all the more complicated that they nevertheless correspond to rather consensual solutions (among the p most consensual ones). However, we hope (and assume in the following) that there exists a subset of them including the true epipole location, and that this subset may be detected based on adequate criteria.

Figure 2 illustrates the proposed strategy that consists in clustering all con-

sidered solutions into different groups based on their consistency in order to derive solutions at cluster or group level before analysing them. Note that, since, by construction, the different groups are inconsistent between them (but consistent inside each group), they provide incompatible individual solutions.

We focus on the BF framework to model the uncertainty and the imprecision associated to the different solutions provided by Multiple Model Sampling. The considered discernment frame (Ω) is the set of the location solutions for the epipole, that is to say the plane containing the static camera image plane (which can be bounded to the image frame or not depending on the camera setting). Using the 2CoBel scalability, sub-pixel coordinates are handled.

Then, for each solution, a consonant BBA is derived whose nested FEs are 2D polygons approximating the ellipses corresponding to different values of confidence level κ in Eq. (10). Specifically, denoting by n_{FE}^0 the number of FEs, in our experiments, the default setting is $n_{FE}^0 = 2$, with uncertainty levels corresponding to 50% and 95% ($\kappa^2 = 1.386$ and $\kappa^2 = 5.991$, respectively) by referring to [46, 47], with equidistributed mass. Other settings will be discussed along with the experiments (Section 6). Note that these BBAs are dogmatic ($m(\Omega) = 0$) which means that we prefer discarding poor solutions (i.e., with the true epipole outside of the biggest focal element) rather than involving them in the fusion process. Indeed, these poor solutions will not bring relevant information for actual epipole location so that we propose to filter them (as presented in next subsection) in order to focus on smaller but relevant sets of sources.

3.2. BBAs clustering

Before explaining the proposed clustering solution, let us recall the specificity of the elements to cluster. Firstly, these elements are BBAs and not objects defined by their features (color, shape etc.) like in most clustering applications. This implies that usual distances such as Euclidean distances are not relevant to compare BBAs representing them for instance as a vector of focal elements (since the relationships between FE will not be taken into account). Secondly, the considered set of elements contains a large proportion of outliers (BBAs

corresponding to outlier solutions). Indeed for our localization application, only one cluster of BBAs will be relevant, the other ones corresponding either to individual outliers or to small grouping. Thus, either predicting the number of clusters or providing some cluster features or training examples is impossible from a practical standpoint. Then, we dismiss centroid-based algorithms such as c-means, or example-based ones such as k-nearest neighbors, despite the existence of evidential versions of such algorithms [48, 49].

Instead we focus on an approach which does not involve a cluster number parameter nor training or reference sample set. Then, the hierarchical clustering [50] appears rather self-evident. Indeed, it only requires the definition of a distance or similarity function, both of which have been widely explored in BFT [36], and of a threshold (maximum distance) parameter. Indeed, it 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, given a sample distance, for the cluster distance, different extensions of sample distance can be considered: simple [51], average [52], complete [53]. In [54], a Hierarchical Ascendant Clustering (HAC) was proposed for the clustering of objects having imprecise and uncertain features. Then the authors define BBAs representing the belief that two objects belong to a same or a different cluster. Such a problem is thus rather different from ours whose objective is to derive some subsets of compatible sources (non conflictual epipole solutions represented in terms of BBAs) for data fusion. We are leveraging the fact that BBAs are already defined for our location problem and cluster them in the perspective of their fusion, e.g., controlling the conflict degree in each cluster of BBAs. Let us detail how this can be accomplished based on a well-chosen BBA distance.

Among the proposed BBA distances, we focus on Jousselme's one [55] mainly for its simplicity and interpretable results (cf. Section 2.1). In this study, it also allows for consistency with the used BBA approximation. Based on it, let us compute the theoretical threshold guarantying that two BBAs have at least one pair of focal elements intersecting (avoiding total conflict). For two BBAs m_1

and m_2 having two nested focal elements of respective mass values a and $(1-a)$ and area ratio equal to $\frac{k_{50}^2}{k_{95}^2}$ (ellipses at 50% and 90% confidence levels), the Jouselme's distance between them in the case of conflict degree equal to 1, is

$$d_{th} = \sqrt{a^2 + (1-a)^2 + 2a(1-a)\frac{k_{50}^2}{k_{95}^2}}. \tag{11}$$

Proof. If m_1 and m_2 are totally conflictual, $\sum_{A \in 2^\Omega} \sum_{\substack{B \in 2^\Omega, \\ A \cap B = \emptyset}} m_1(A)\, m_2(B) = 1$. Under this hypothesis, it follows that

$$\forall (A,B) \in 2^\Omega \times 2^\Omega, m_1(A)m_2(B) > 0 \Rightarrow A \cap B = \emptyset.$$

Therefore, $\langle m_1, m_2 \rangle = 0$.

Besides, since m_1 has only two nested focal elements,

$$\forall (A,B) \in 2^\Omega \times 2^\Omega, m_1(A)m_1(B) > 0 \Rightarrow \frac{|A \cap B|}{|A \cup B|} = \begin{cases} 1 & \text{if } A = B, \\ \frac{k_{50}^2}{k_{95}^2} & \text{otherwise.} \end{cases}$$

Then, $\langle m_1, m_1 \rangle = a^2 + 2\frac{k_{50}^2}{k_{95}^2}a(1-a) + (1-a)^2$ and it is the same for $\langle m_2, m_2 \rangle$. Therefore,

$$\begin{aligned} \frac{1}{2}\left(\langle m_1, m_1 \rangle + \langle m_2, m_2 \rangle - 2\langle m_1, m_2 \rangle\right) &= \langle m_1, m_1 \rangle \\ &= a^2 + 2\frac{k_{50}^2}{k_{95}^2}a(1-a) + (1-a)^2 \end{aligned}$$

□

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 the perspective of conjunctive combination of all the BBAs belonging to the same given cluster, we consider complete linkage which uses the *max* operator for computing the distance between two clusters from the distance values between samples. It allows us to bound the distance between any pair of BBAs we will combine during the intra-cluster combination step. However, 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). Hence, 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. Indeed, for any set of j BBAs $\{m_i\}_{1\leq i\leq j}$, $\textcircled{\cap}_{m_i}(\emptyset) < 0 \iff \exists(A_1,\ldots,A_j) \in \mathcal{F}_{m_1}\times\cdots\times\mathcal{F}_{m_j}$ such that $\bigcap_{1\leq i\leq j} A_i \neq \emptyset$. If the BBAs are consonant, the m_i FEs are nested, and a non empty intersection with a FE implying a non-empty intersection with any FE including it, we have only to check the existence of at least one non-empty intersection between the largest FEs.

In summary, 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 $\bigcap_{m_i}$ for all m_i in the cluster. It allows us to derive consistent BBA clusters with respect to the conjunctive combination.

In the following, given a set of BBAs $\mathcal{M} = \{m_1, m_2, ..., m_p\}$ to cluster, let l denote the number of obtained BBA clusters and $\{\mathcal{M}_1, ..., \mathcal{M}_i, ..., \mathcal{M}_l\}$ the set of clusters with (i) $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset, \forall\,(i,j) \in \{1,\ldots,l\}^2, i \neq j$ and (ii) $\cup_{i\in\{1,\ldots,l\}}\mathcal{M}_i = \mathcal{M}$.

3.3. Intra-cluster fusion

From the partition $\{\mathcal{M}_1, ..., \mathcal{M}_i, ..., \mathcal{M}_l\}$ of the set of BBAs $\mathcal{M}$, the BBAs within each cluster shall be combined using the conjunctive rule [25]. This rule assumes cognitive independence between BBAs which here comes from the independence between RANSAC solutions (corresponding to different 8-tuples, i.e., different linear systems). Note that cognitive independence does not prevent BBAs to be similar, which is all the more expected for BBAs representing the ground truth epipole.

However, for clusters including more than a few tens of BBAs, a step of BBA approximation has to be implemented (cf. Section 2.1) to control computational complexity. Specifically, we perform BBA approximation each time the number of focal elements is larger than n_{FE}^{max} and decrease it to n_{FE}^{sum}, with typically $n_{FE}^{max} = 20$ and $n_{FE}^{sum} = 10$. The used BBA approximation process is the same as in [17].

Now, one issue introduced by the BBA approximation is the loss of the associativity of the combination which induces a dependency of the result with respect to the combination order. Different strategies have thus been explored. On the one hand, it may appear as more natural to gather first the closest BBAs (still according to Jousselme's distance) so that the combination ordering follows the 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 6. They correspond to updated minimal (respectively maximal) distance ordering. Let us define the intersection between two BBAs by $\mathcal{F}(m_i) \cap \mathcal{F}(m_j) = \{A \cap B\}_{(A,B)\in\mathcal{F}(m_i)\times\mathcal{F}(m_j)}$. Then, fusion is performed as follows. The two first BBAs to combine are:

$$(i^*, j^*) = \underset{\substack{(i,j)\in\mathcal{F}(m_i)\times\mathcal{F}(m_j)\\ \mathcal{F}(m_i)\cap\mathcal{F}(m_j)\neq\emptyset}}{\arg\min} d_J(m_i, m_j); \tag{12}$$

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

$$j^* = \underset{\substack{j\in\mathcal{F}(m_j)\\ \mathcal{F}(\tilde{m})\cap\mathcal{F}(m_j)\neq\emptyset}}{\arg\min} d_J(\tilde{m}, m_j); \tag{13}$$

Equations (12) and (13) correspond to *min* ordering. The *max* ordering is obtained replacing arg min by arg max in them.

After this intra-cluster fusion step, the l cluster BBAs $\{\tilde{m}_i\}_{i\in\{1,\dots,l\}}$ are ranked according to their maximum $BetP_i$ value or Pl_i value: $max_{A\in\mathcal{MI}(\tilde{m}_i)} f_i(A)$ with $f \in \{BetP, Pl\}$ and $\mathcal{MI}(m)$ the set of Maximal Intersections of BBA m, since 2CoBel [23] handles decisions on it. Recalling that our aim is to provide a set (as small as possible) of solutions (possibly under the form of BBAs) including the ground truth, we consider the k first cluster BBAs as the *proposed solution set*. Algorithm 1 describes the steps to derive this *proposed solution set*.

Algorithm 1 : 2D belief clustering for epipole localization

Input: I_1 and I_2: A pair of images from two different views; m, θ: Multiple Model sampling parameters; n_{FE}^0: initial FE number; n_{FE}^{max}, n_{FE}^{sum}: summarization parameters; k: *proposed solution set* cardinality

Output: The *proposed solution set* $\tilde{\mathcal{M}} = \{\tilde{m}_1, \ldots, \tilde{m}_k\}$ of BBAs representing imprecise solutions for epipole location

Extract putative correspondence set $\mathcal{P} = \{p_1, \ldots, p_n\}$ between I_1 and I_2

Run RANSAC algorithm and, during it, select the m models with the largest inlier support, and rank them in $\mathcal{S}_{\mathbf{F}}^0 = \left\{\tilde{\mathbf{F}}_{\mathbf{1}}, \ldots, \tilde{\mathbf{F}}_{\mathbf{m}}\right\}$

$p = \arg\max_{l \in \{1,\ldots,m\}} l \times \mathbb{1}_{|\mathcal{I}_l| \geq \theta \times |\mathcal{I}_1|}$ ($\mathbb{1}_Z$ is the indicator function of set Z)

$\mathcal{S}_{\mathbf{F}} \leftarrow$ the p first elements of $\mathcal{S}_{\mathbf{F}}^0$

Initialize $\mathcal{M} = \emptyset$;

for $\tilde{\mathbf{F}}_{\mathbf{i}}$ in $\mathcal{S}_{\mathbf{F}}$ **do**

  Estimate epipole $\mathbf{e_i}$ (using SVD) and covariance matrix $\mathbf{\Sigma_{e_i}}$

  Build the 2D BBA m_i with n_{FE}^0 nested FEs corresponding to preset confidence levels κ in Eq. (10)

  $\mathcal{M} \leftarrow \mathcal{M} \cup \{m_i\}$

end for

Compute d_{th} from Eq. (11)

Set of clusters $\{\mathcal{M}_1, \ldots, \mathcal{M}_l\} \leftarrow$ output of HAC (complete linkage, Jousselme's distance, $d_{th} - 0.05$ threshold, input $\mathcal{M}$)

Initialize $\tilde{\mathcal{M}} = \emptyset$

for $i = 1$ to l **do**

  $\tilde{m}_i = \textcircled{\cap}_{m_j \in \mathcal{M}_i} m_j$

  $\tilde{\mathcal{M}} \leftarrow \tilde{\mathcal{M}} \cup \{\tilde{m}_i\}$

end for

Rank BBas in $\tilde{\mathcal{M}}$ according to max value of chosen decision criterion ($BetP$ or Pl) and only keep in $\tilde{\mathcal{M}}$ the k first top-ranked BBAs

Finally, note that for our application, we avoid any assumption about a possible relationship between cluster reliability and cardinality (as in [34]) so that we will evaluate each cluster independently of its cardinality.

In the two next sections, we present two application-dependent strategies for raising the ambiguity between the elements of *proposed solution set*. In the first strategy, we benefit from **additional sources** of localization whereas in the second one, we benefit from **multi-temporal** information.

4. Multi-source camera localization

The output of the belief clustering is the *proposed solution set* containing at most k BBAs representing the k most likely imprecise locations for the second camera. To raise the ambiguities among these different locations, additional sources shall be considered. Now, we investigate two different examples of such sources whose availability depends on the considered system. In both cases, we will define BBAs modeling the new evidence brought by each of these additional sources. These BBAs are defined on the same discernment frame as previously, namely Ω, the image plane of the static camera.

4.1. Belief from a pedestrian detector

When considering a wearable camera, the device is necessarily co-localized with the person carrying it. Thus, a computer vision pedestrian detector appears as a relevant source for localization in the view of the static camera. Among the many algorithms proposed to detect and localize pedestrians, Convolutional Neural Networks (CNN) have proven to be highly effective, e.g. [56, 57, 58] so that we naturally turn toward such approaches. Their output is a set of Bounding Boxes (BB) around each detected pedestrian. Now, in the perspective of BBA definition, we underline three main BB features: (i) as the camera is often held near the head or the shoulder, it is more likely located in the BB upper part than in the lower part; (ii) as a BB can imperfectly enclose the pedestrian (in particular in case of occlusion), the mobile camera may be

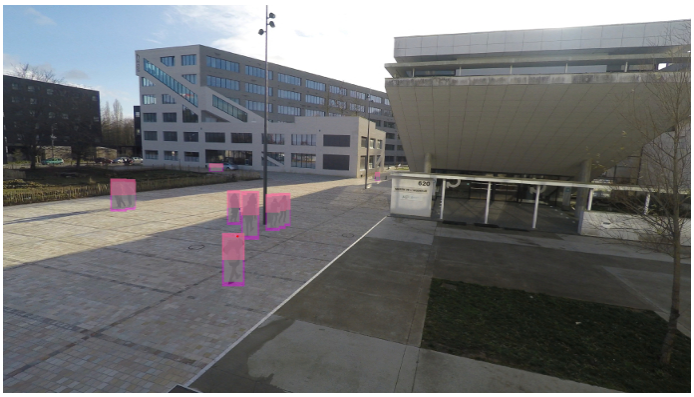


Figure 3: Focal elements of the BBA associated to the bounding boxes provided by the pedestrian detector.

outside BB although very close; (iii) in case of multiple pedestrians detected it is impossible to assume which one is more likely to wear the camera.

Based on the these BB features, we consider a set of BBs denoted by $\mathcal{B} = \{B_1, \ldots, B_i, \ldots, B_m\}$, followed with the dilation operation the dilation operation with disk structuring element of radius ρ denoted by δ_ρ, and the upper half of any box B_i denoted by B_i^{up}. We define a BBA associated to $\mathcal{B}$ with four focal elements: $B = \bigcup_{B_i \in \mathcal{B}} B_i$, $B^\delta = \bigcup_{B_i \in \mathcal{B}} \delta_\rho(B_i)$, $B^{up} = \bigcup_{B_i \in \mathcal{B}} B_i^{up}$ and $B^{\delta,up} = \bigcup_{B_i \in \mathcal{B}} \delta_\rho(B_i)^{up}$, illustrated in Figure 3. Among these four focal elements, the mass is about equally distributed even if we may refine the precise value during experiments. Note that as previously the BBA is dogmatic, but for a different reason. Here, we assume the pedestrian detector reliable enough to not miss the camera carrier. More specifically, we assume its ambiguities (between the different pedestrians present in the scene) are complementary to the ambiguities among BBA clusters. Then, we choose the BBA to be dogmatic in order to have a chance to detect, based on the degree of conflict, when it provides a completely erroneous solution.

4.2. BBA from GNSS data

In this setting, we assume that GNSS data are provided by a low-cost and light sensor embedded by the camera carrier. Thus the measured localization

is rather imprecise and has to be combined with our other localization pieces of evidence in order to increase spatial precision. The conversion of the rough 3D GNSS location into one or several focal elements in the image space requires the knowledge of the static camera pose parameters along with the theoretical 3D precision of GNSS. Besides, since the altitude (z coordinate) provided by the GNSS is far more imprecise than the horizontal plane location [22], we rather consider a prior (even if approximate) on the height h of the mobile camera (due to the fact it is held by a pedestrian).

Let P_m denote the plane parallel to plane $z = 0$ and having elevation equal to h and let (x, y) be the 2D GNSS coordinates representing its location on the Earth surface. Considering the GNSS intrinsic imprecision, we represent imprecise location areas as 2D nested disks in P_m centered on (x, y). Now, when the static camera elevation is large with respect to the interval (length) of the possible heights for mobile camera, the imprecision assuming a given height h is small with respect to GNSS intrinsic imprecision. Thus, in the following, we neglect the vertical uncertainty (along h) with respect to the horizontal one.

In order to derive the focal elements of the BBA associated to the GNSS data, we simply project the 2D nested disks in the ground plane P_m to the image plane of the static camera, which can be performed via a homography transformation for circle. Under the pinhole camera model, the homography between the plane P_m of equation $z = h$ and the image plane is represented by the matrix

$$\mathbf{H} = \mathbf{KR} \begin{bmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & h - c_z \end{bmatrix},$$

where $\mathbf{K}$ is the intrinsic matrix of camera, (c_x, c_y, c_z) is the 3D coordinates of camera center, and $\mathbf{R}$ the rotation matrix for camera's orientation. As a disk having center coordinates (x, y) with the radius r on P_m can be represented

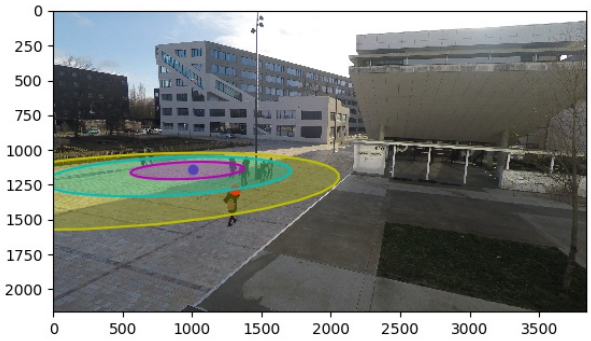


Figure 4: Illustration of the focal elements derived from GNSS localization. The three conics represent the uncertainty areas associated with GNSS noise level equal to σ, 2σ and 3σ, respectively.

with the general conic formulation as follows

$$\mathbf{C_r} = \begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \\ -x & -y & x^2 + y^2 - r^2 \end{bmatrix},$$

the projection on the image plane for the disk $\mathbf{C_r}$ on P_m is then represented by the conic matrix

$$\mathbf{C_e} = \mathbf{H}^{-T}\mathbf{C_r}\mathbf{H}^{-1}. \tag{14}$$

Note that, if in most cases, $\mathbf{C_e}$ is an ellipse, under rares circumstances, the projective projection can result in a hyperbole. In the application, the BBA representing the GNSS localisation information has three nested focal elements that are the inside areas of conics obtained by perspective projection of three P_m circles centered on (x, y), having radius equal to σ, 2σ and 3σ, respectively, with σ the theoretical uncertainty or error bar on (x, y). Using 2CoBel, these focal elements are approximated by 2D polygons, denoted by A_σ, $A_{2\sigma}$ and $A_{3\sigma}$. Figure 4 illustrates the conics (ellipses in this example) provided by the GNSS

data. As previously, the mass is about equally distributed between these three focal elements even if we may fine tune the precise values during experiment and the BBA is dogmatic, which means we assume it reliable enough to not miss $\mathbf{e}$, and otherwise the fusion process will be invalidated thanks to conflict detection.

4.3. Global fusion algorithm

From Section 3, we derive a set of l cluster BBAs $\tilde{\mathcal{M}} = \{\tilde{m}_i\}_{i\in\{1,\dots,l\}}$ and from Section 4 another set of BBAs brought by additional sources $\mathcal{M}^a = \{{m^a}_j\}_{j\in\{1,\dots,q\}}$. Now, if all $\mathcal{M}^a$ BBAs can be assumed reliable (i.e., there is at least one FE containing $\mathbf{e}$), we know that some BBAs of $\tilde{\mathcal{M}}$ correspond to outliers. Furthermore, by construction of the clusters, two different cluster BBAs are highly incompatible so that, for conjunctive fusion process, we can only retain one among the l cluster BBAs. Our fusion strategy is then to use the $\mathcal{M}^a$ BBAs to adjudicate.

Note that, to save some computational resources we use two tricks in Algorithm 2. Firstly, benefiting from the associativity property of conjunctive rule (Eq. (2)), $\mathcal{M}^a$ BBAs are combined only once before $\tilde{\mathcal{M}}$ BBAs inspection: $m^a_{\textcircled{\cap}} = \textcircled{\cap}_{m^a_i\in\mathcal{M}^a} m^a_i$, along with the disjunction of $m^a_{\textcircled{\cap}}$ focal elements, $\bigcup_{m^a_{\textcircled{\cap}}}$. Secondly, $\tilde{\mathcal{M}}$ BBAs are filtered by testing the intersection between $\bigcup_{m^a_{\textcircled{\cap}}}$ and the focal elements of $\tilde{m}_i \in \tilde{\mathcal{M}}$: if the intersection is empty, $\tilde{m}_i$ is withdrawn.

Then, having evaluated to consistency of each cluster-BBA with the additional source BBA through their conjunctive combination, the output BBA is the one that will maximize the decision (epipole location) criterion (even if the algorithm itself does not provide such a crisp decision).

5. Multi-temporal epipole localization

In this application, we consider no longer a mobile camera within the field of view of a static camera, but a pair of static cameras. Assuming that these cameras capture synchronized video streams of a dynamic scene, we aim to

Algorithm 2 : BBA selection based on additional sources

Input: The set of cluster BBAs $\tilde{\mathcal{M}} = \{\tilde{m}_i\}_{i\in\{1,\ldots,l\}}$; The set of additional BBAs $\mathcal{M}^a = \left\{m_j^{add}\right\}_{j\in\{1,\ldots,q\}}$;

Output: BBA m_{out}

$m^a_{\textcircled{\cap}} = \textcircled{\cap}_{m_i^a\in\mathcal{M}^a} m_i^a$;

$\bigcup_{m^a_{\textcircled{\cap}}} = \bigcup_{B\subseteq\Omega, m^a_{\textcircled{\cap}}(B)>0} B$;

$\mathcal{M}_c = \emptyset$;

for $i = 1$ to l **do**

  if $\bigcup_{m^a_{\textcircled{\cap}}} \cap \bigcup_{\tilde{m}_i} \neq \emptyset$ **then**

    $m^{i,a} = \tilde{m}_i \,\textcircled{\cap}\, m^a_{\textcircled{\cap}}$

    Append $m^{i,a}$ to $\mathcal{M}_c$;

  end if

end for

$m_{out} = \underset{m^{i,a}\in\mathcal{M}_c}{\arg\max}\left\{max_{A\in\mathcal{MI}(m^{i,a})} Pl^{i,a}(A)\right\}$

exploit the temporal sequence for epipole localization. As the cameras are fixed and the scene is dynamic, each pair of frames can provide a new estimation for the fixed epipole location using a standard RANSAC process applied to image pairs. Note that, depending on the speed of the moving objects with respect to the frame acquisition frequency, we may have to sub-sample the sequence in order to ensure sufficient changes in the video content and cognitive independence of epipole estimations. In our case, the sequence frequency is 2 frames per second and we observe that, due to the fact that most matches occur on moving objects, the ratio of consecutive point matches (between each image pair) is about 12%, so that further sequence sub-sampling is not necessary.

Then, considering the reference camera image plane as discernment frame Ω, each instantaneous epipole solution is associated to a consonant BBA having $n^0_{FE} = 2$ nested focal elements corresponding to ellipses at 50% and 90% confidence levels using uncertainty propagation. This BBA construction is similar

to the one used in Section 3.2, except the derivation of the epipole solution performed here using standard RANSAC. Then, if the whole sequence contains N frames, we derive a set of N BBAs. Among them, many may be irrelevant (*outlier* BBAs) due to erroneous matches between keypoints.

Now, the basic idea of fusion is still to remove outliers based on cross-validation performed between multiple sources. Having a single temporal sequence, we create multiple sources by splitting it into T non-overlapping subparts with identical number of frames. To each subpart $t \in \{1, \ldots, T\}$ we associate a set $\mathcal{M}_t$ of $\lfloor \frac{N}{T} \rfloor$ BBAs. $\mathcal{M}_t$ is then processed according to the proposed belief clustering to derive *proposed solution sets* containing the k top-ranked cluster BBAs: $\tilde{\mathcal{M}}_t = \{\tilde{m}_{t,1}, \ldots, \tilde{m}_{t,k}\}$. Theoretically, we have to evaluate every T-tuples of BBAs formed by selecting one BBA in each set $\tilde{\mathcal{M}}_t$. However, as described in Algorithm 3, to save computational resources, we perform processing sequentially by detecting the BBAs with conflict degree equal to 1 as early as possible to avoid their combination. Then, like in Algorithm 2, we also base ranking of the combined BBAs on decision criterion. Now, the main difference is that we do not select the top BBA but the v first-ranked ones, in an ad-hoc and conservative spirit. These v first-ranked BBAs are then gathered in a single BBA using disjunctive combination since they are incompatible by construction (selection of incompatible clusters in at least one subpart of the sequence).

6. Experiments and results

6.1. Datasets, parameters and evaluation criterion

In order to evaluate the benefit of the proposed evidential epipole localization, we consider three datasets. Two of them are public datasets and one has been specifically acquired for this research. Since they have complementary features, they allow us to check the robustness of the belief clustering, and at the same time to evaluate epipole localization in different contexts:

- Firstly, to check the effectiveness of BBA clustering and k first-rank clusters selection on a public dataset, we selected 128 pairs of images in the

Algorithm 3 : Epipole estimation from video sequence

Input: The T sets of cluster BBAs $\tilde{\mathcal{M}}_t$, $t \in \{1, \ldots, T\}$;

Output: BBA m^{out};

1: $\mathcal{M}_{cur} = \tilde{\mathcal{M}}_1$;
2: $\mathcal{M}_{new} = \emptyset$;
3: **for** $i = 2$ to T **do**
4: **for all** $m \in \mathcal{M}_{cur}$ **do**
5: **for all** $\tilde{m}_{i,j} \in \tilde{\mathcal{M}}_i$ such that $\bigcup_m \cap \bigcup_{\tilde{m}_{i,j}} \neq \emptyset$ **do**
6: Add $m \textcircled{\cap} \tilde{m}_{i,j}$ to $\mathcal{M}_{new}$;
7: **end for**
8: **end for**
9: $\mathcal{M}_{cur} = \mathcal{M}_{new}$;
10: **end for**
11: Rank (according to chosen criterion, $BetP$ or Pl) BBAs in $\mathcal{M}_{new}$ and store the first-ranked v BBAs in $\mathcal{M}_{out}$
$m^{out} = \textcircled{\cap}_{m_s \in \mathcal{M}_{out}} m_s$;

public dataset used in [59]; these images are chosen to present various poses with the large view image containing the epipole of the mobile camera (even if acquisitions having been performed at different instants); the ground truth epipole location (estimated using Structure from Motion) is provided with the dataset.

- Secondly, to evaluate the benefit of the pedestrian detector for mobile camera localization, we use a dataset with simultaneous acquisition of static and mobile cameras; this latter has been specifically acquired with a camera wearer moving on the ground level in urban environment; it contains 196 synchronized image pairs, and the ground truth has been manually defined by a human annotator.

- Thirdly, to evaluate the benefit of multi-temporal acquisitions for epipole localization, we use two synchronized video streams; we consider the public

WildTrack dataset[2] [60] consisting of GoPro cameras mounted on tripods and recording the busy entrance of an university campus building. For this dataset the two cameras are static with large field of view so that the epipole no longer coincide with the position of one camera (that is no longer visible on the other camera images). The ground truth is provided by the dataset creators via an extrinsic calibration.

Let us recall the used parameters for BBA clustering. In the case of datasets 1 and 2, image pairs are processed independently. Then, for each pair of images, we derive multiple pieces of evidence (regarding the epipole location along with its uncertainty ellipses) by considering various solutions provided by RANSAC instead of retaining only the most consensual hypothesis (cf. Section 3.1). The number of iterations for sampling matches during RANSAC is set to $n = 10^5$. The number of retained solutions p is set to be at most 100 (as long as their inlier support satisfies the condition depending on θ). Under this bound, p exact value is determined by θ ($p = f(\theta)$). In our experiments, we study the result sensitivity to θ and k and infer some guidelines on setting them. For hierarchical clustering, we use the **AgglomerativeClustering** function of the module scikit-learn [61] with the complete linkage. Now, since this function does not allow for applying an additional binary constraint, we introduce the "non-empty intersection" constraint (between disjunctions of respective focal element sets, cf. Section 3) a posteriori, namely during the fusion step based on *min* distance ordering.

As quantitative evaluation criterion, we consider the modified metric [17]

$$\epsilon(\lambda) = \sum_{A \in 2^\Omega} d(\mathbf{e_{gd}}, A) m(A) + \lambda \sum_{A \in 2^\Omega} |A|\, m(A), \tag{15}$$

where $\lambda \in \mathbb{R}_{\geq 0}$ is a weighting parameter between terms, $\mathbf{e_{gd}}$ is the ground truth

[2] https://www.epfl.ch/labs/cvlab/data/data-wildtrack/

epipole location and $d(\mathbf{e_{gd}}, A)$ is defined as

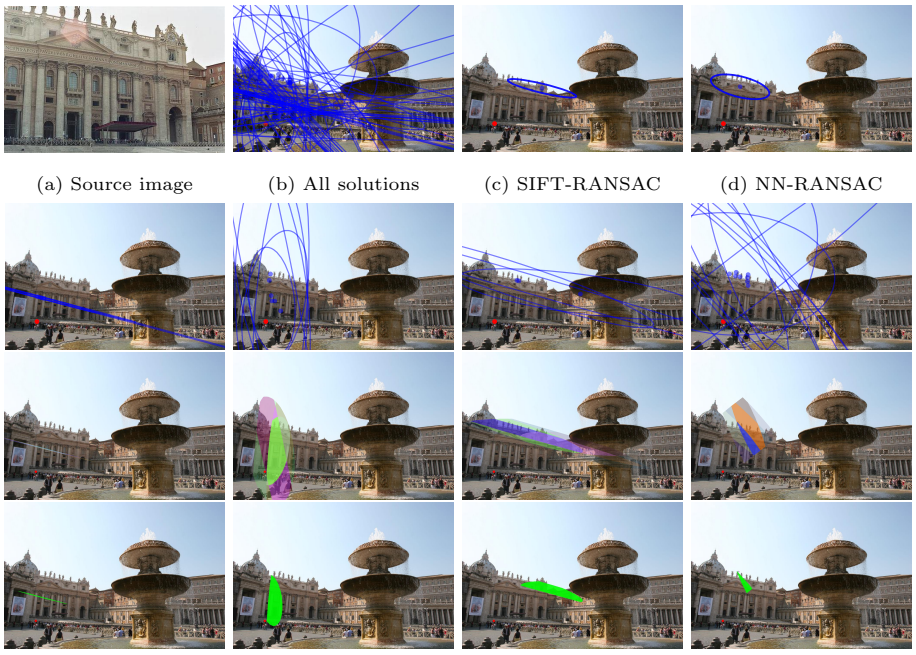
$$d(\mathbf{e_{gd}}, A) = \begin{cases} 0 & \text{if } \mathbf{e_{gd}} \text{ is included in } A, \\ \min\limits_{\mathbf{p} \in A} \| \mathbf{e_{gd}} - \mathbf{p} \|_2 & \text{otherwise,} \end{cases} \tag{16}$$

where $\| \, . \, \|_2$ is the Euclidean distance. 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 area of focal elements. BBAs with large focal elements (thus spatially imprecise) including $\mathbf{e_{gd}}$ exhibit low error values for λ close to 0, but higher error values when λ increases. Conversely, committed BBAs with small focal elements close to $\mathbf{e_{gd}}$ but not necessary including it are all the more badly scored that λ is low and better evaluated for λ having large positive values.

6.2. Evaluation for belief clustering

To evaluate the proposed belief clustering only, we consider the first dataset used in [59]. Let us recall that we propose evidential modeling and the customization of a well-known clustering algorithm for difficult settings where outliers exhibit a kind of consistency, so that there is no clear consensus. To show the benefit of such an approach with respect to approaches less conservative but deemed to be robust, we compare it with the standard RANSAC method based on traditional features (**SIFT-RANSAC**) and with the NN based outlier rejection (**NN-RANSAC** [59]). From their results, the epipole uncertainty is derived as in [14] ("Least squares SIFT" and "Least squares NN").

Figures 5 and 6 illustrate some localization results, provided by existing methods, by top-ranked clusters and by low-ranked clusters respectively. It illustrates that in difficult settings, the existing methods tend to be overconfident. Top ranked clusters exhibit a higher consistency among the BBAs which results in a strong ellipse alignment, and even for challenging poses the true solution is present at the top. However, we notice that the first-ranked BBA may fail in providing the right epipole location conversely to the second-ranked one that includes (or almost does) the ground truth while exhibiting a moderate level



(a) Source image (b) All solutions (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)

Figure 5: 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); Next three rows: for a given cluster, the corresponding original ellipses (first row), the final BBA (second row) and the maximum $BetP$ Maximal Intersection (third row, green area); cases of (e): the top ranked cluster, (f): the second ranked cluster, (g)-(h): two clusters with a low rank/BetP due to the sources being less consistent.

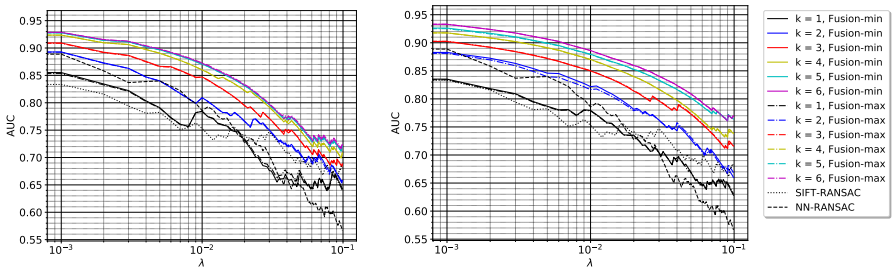
(a) Source image (b) All solutions (c) SIFT-RANSAC (d) NN-RANSAC

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

Figure 6: 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); Next three rows: for a given cluster, the corresponding original ellipses (first row), the final BBA (second row, random colors distinguishing the different focal elements) and the maximum $BetP$ Maximal Intersection (third row, green area); cases of (e): the top ranked cluster, (f): the second ranked cluster, (g)-(h): two clusters with a low rank/BetP due to the sources being less consistent.

of imprecision. Thus, these examples illustrate the benefit of keeping several cluster BBAs for fusion with addition sources. We also see that the low rank clusters consist in uncertainty areas exhibiting less consistency that can thus be harmlessly discarded. Finally, these examples also show the importance, for further fusion, to keep the whole BBA and not only the $BetP$ solution since a too early decision may miss the actual epipole location (not included in the $BetP$ selection in Fig. 5).

Then, to evaluate $\epsilon(\lambda)$ also in the case of uncertainty ellipses (as in the case of standard algorithms), for a fairer comparison, we convert these latter in a BBA. Specifically, we derive consonant BBAs having five equi-weighted focal elements, represented by the polygons approximation of the ellipses associated with respectively 95%, 75%, 50%, 25%, 10% confidence levels.



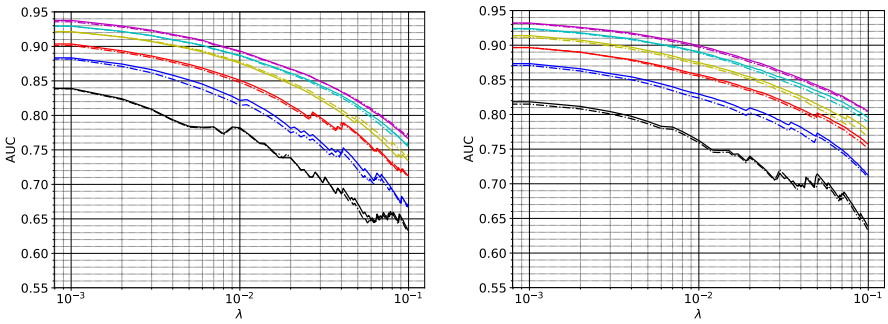
(a) Consensus threshold $\theta = 0.9$ (b) Consensus threshold $\theta = 0.5$

Figure 7: Curve of AUC for cumulative curve, $\epsilon(\lambda)$, versus λ.

Then, for each considered algorithm, we compute the error $\epsilon(\lambda)$ for each pair of images and then derive the empirical cdf (cumulative density function) of $\epsilon(\lambda)$. Since, as previously stated, the λ value strongly impacts the performance and then the ordering of the evaluated algorithms, we plot performance versus λ. For this, the whole cdf is summarized through its Area Under the Curve (AUC) value. The higher this value, the more efficient an approach is. Figure 7 shows, for the different algorithms we consider, 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 (as explored in next experiments) will allow for "good"

cluster selection. Nevertheless, we found interesting to evaluate, albeit in a preliminary way, the proposed approach on this public dataset that offers very various poses and scenes. The results underline that, as the value of k increases, the performance of the proposed BBA clustering improves as expected, and it outperforms other methods. In most cases, the desired estimation is within the 4 or 5 first-ranked clusters. The sub-figures (a) and (b) in Figure 7 allow for comparison between results achieved with $\theta = 0.9$ and $\theta = 0.5$. We notice that results are slightly better with $\theta = 0.5$ which means the consensus measure used by RANSAC is not optimal. We also note that results appear robust with respect to the fusion order introduced by different fusion strategies (*min* and *max* ordering discussed in Section 3.3).

Finally, let us evaluate the robustness of the approach with respect to the FE number parameters, either n_{FE}^0 during the allocation, or n_{FE}^{max} and n_{FE}^{sum} during BBA approximation. Figure 8 evaluates the impact of the approximation parameters in two cases of initial BBA allocations: two nested FEs ($n_{FE}^0 = 2$) corresponding to uncertainty levels 95% and 50% and five nested FEs ($n_{FE}^0 = 5$) corresponding to uncertainty levels 95%, 75%, 50%, 25% and 10%. The curves corresponding to different approximation parametrizations are represented with different line styles, so that we can check the very low impact of these parameters on AUC. Comparing the two subfigures, we also note the low impact of the number of initial FEs, even if more pronounced for $k = 1$. Figure 9 shows, versus the number of sources per cluster (i.e., the number of BBAs to combine), the average number of approximation and the average computation time in seconds, still distinguishing the two BBA allocation cases $n_{FE}^0 \in \{2, 5\}$. We notice that the approximation number curve mainly depends on n_{FE}^0 and, as expected, on the ratio $\frac{n_{FE}^{max}}{n_{FE}^{sum}}$ rather than on the absolute values of approximation parameters. Meanwhile, the average running time for cluster-BBA computation depends mainly on approximation parameters, in particular n_{FE}^{sum} which controls the complexity of the combination rule. Finally, noticing that the gain in AUC is either negligible when increasing $(n_{FE}^{max}, n_{FE}^{sum})$ or very low when increasing the initial number of FEs per BBAs, while the running time increases in a



(a) BBA allocation with $n_{FE}^0 = 2$

(b) BBA allocation with $n_{FE}^0 = 5$

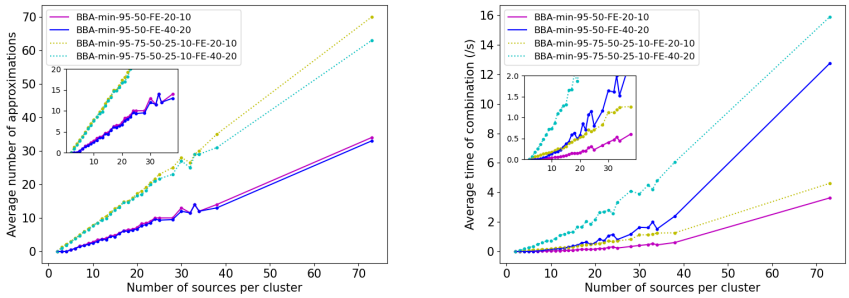
Figure 8: Impact of the approximation parameters on the AUC curves: different colors correspond to different numbers of kept clusters (from $k = 1$ in black to $k = 6$ in magenta), plain and dashed lines correspond to $(n_{FE}^{max}, n_{FE}^{sum})$ equal to $(20, 10)$ and $(40, 20)$, respectively.

significant way, leads us to set default parameters to $n_{FE}^0 = 2$, $n_{FE}^{max} = 20$, and $n_{FE}^{sum} = 10$.

6.3. Results for multi-source localization

In this experiments, we consider the second data set. For BBA clustering, accordingly to previous section, we set $\theta = 0.5$ and use *min* ordering for intra-cluster BBA combination. For the pedestrian detector, we focus on the widely used object detector **YOLO** [62]. As the GNSS source is not available in this dataset, we generate the simulation of GNSS position by adding a random normally distributed noise to the ground truth of epipole location. For each pair of images, we randomly sample a realization from the distribution $\mathcal{N}(\mathbf{e_{gd}}, \mathbf{\Sigma_{e_{gd}}})$ with $\mathbf{\Sigma_{e_{gd}}} = \sigma^2 \times \mathbf{I_{2\times 2}}$, where σ is the defined noise level.

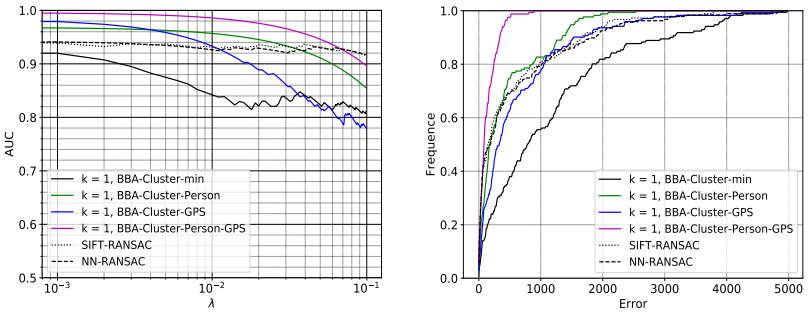
Since we investigate the benefit of multi-source fusion for our application of localization, the presented results contain a gradually increasing number of sources, from only considering **SIFT-RANSAC** or **NN-RANSAC** results or BBA clustering results with $k = 1$, to also considering either one additional source (the pedestrian detector or the GNSS simulation) or the two additional sources. As in the previous subsection, the evaluation is based on the AUC curve for localization error $\epsilon(\lambda)$ defined in Eq. (15).



(a) Average number of BBA approximations (b) Running time (in second) of cluster-BBA computation

Figure 9: Impact of BBA parameters n_{FE}^0, n_{FE}^{max}, and n_{FE}^{sum} ($n_{FE}^0 \in \{2, 5\}$ is called '95-50' or '95-75-50-25-10', respectively; $(n_{FE}^{max}, n_{FE}^{sum}) \in \{(20, 10), (40, 20)\}$ is called 'FE-20-10' or 'FE-40-20', respectively); subplots inside each subfigure are a zoom on $[0, 40]$ x-values.

Figure 10a shows the AUC curves versus the λ parameter whereas Figure 10b shows the cdf for $\lambda = 0.01$ (knowing that the first term of $\epsilon(\lambda)$ ranges in $\left[0, 5.10^3\right]$ and the second term in $\left[0, 10^5\right]$). We notice the very high performance achieved considering the three sources (BBA clustering on RANSAC solutions, pedestrian detector and GNSS data). Specifically, we note that the single use of BBA clustering method with $k = 1$ is less competitive than the standard RANSAC approaches. This is mainly due to the fact that the standard approaches provide a larger ellipse than the BBA clustering, that nevertheless was defined to provide several rather committed solutions for further fusion purposes. Note also that, if on the previous dataset NN-RANSAC provided better results than SIFT-RANSAC, its results on this dataset are rather disappointing. This is due to the fact that, on the first dataset, NN-RANSAC results were good as this latter has been trained on the same dataset. Now, combining the output of BBA clustering with one additional source (either pedestrian detector or GNSS data) allows us to overcome the limitations of these traditional approaches, and using all sources results in a rather significant leap in performance. Such results both highlight the effectiveness of the filtering of the set of initial solutions based on BBA clustering and the ability of an additional source to select the correct BBA



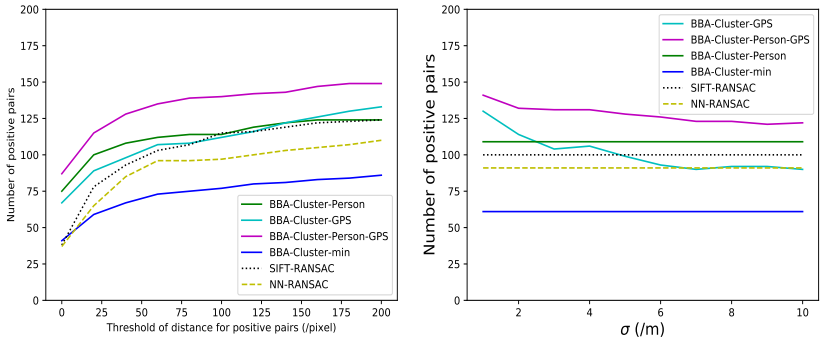
(a) AUC versus λ (b) Cumulative density function for $\lambda = 0.01$

Figure 10: Results in terms of AUC and CDF of the error $\epsilon(\lambda)$ achieved by the multi-source localization of mobile camera.

cluster.

In order to provide an alternative evaluation of our results, we look at the evolution of the number of positive image pairs with respect to the distance threshold to the ground truth epipole. Now, the distance will be computed as in Eq. (16) with $A = \bigcup_m$ (which boils down to simplifying or summarizing our resulting BBA in a categorical BBA). Figure 11a provides the number of such positive pairs versus the used distance threshold. Note that, in our application, we consider it is still meaningful to localize the target camera wearer even when the distance threshold to the ground truth epipole is increased to 100 pixels compared to the large resolution of the reference image (4K). Figure 11a confirms the conclusions previously drawn from Fig. 10a.

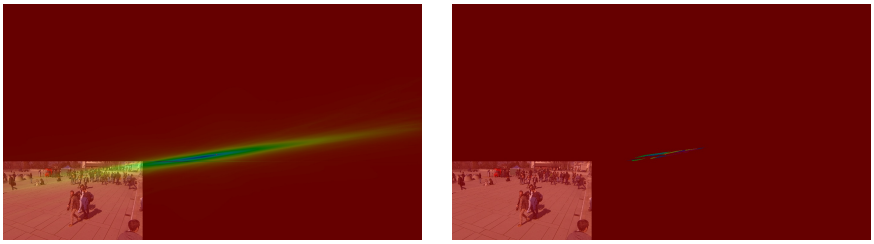
Finally, for the distance threshold equal to 50 pixels, we look at the number of positive pairs versus GNSS noise. We note that, as expected, as the GNSS noise increases, the number of positive pairs obtained by fusion involving GNSS data decreases. Nevertheless, when used in addition to the BBA clustering output and pedestrian detector, the GNSS data appear useful even with rather high noise levels (up to 10 m) since performance overcomes the results obtained without it.



(a) Number of positive pairs versus distance threshold; GNSS noise $\sigma = 3m$.

(b) Number of positive pairs versus GNSS noise level; distance threshold = 50px.

Figure 11: Number of positive pairs for different sources used for mobile camera localization.



(a) Accumulated Vote

(b) BBA clustering

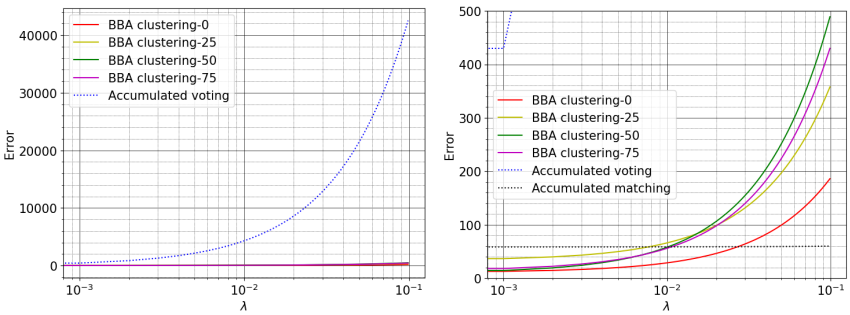
Figure 12: Qualitative comparison between accumulated voting and the proposed multi-temporal BBA fusion.

6.4. Results for video localization

In this section, we aim at evaluating the epipole localization using a pair of static cameras. The localization algorithm is the one described in Section 5, Algorithm 3 with $T = 4$ and $v = 10$. Since the whole sequence contains 400 frames, each of the 4 image subsets contains 100 consecutive frames, from which BBA clustering allows us to retain the 5 first-ranked cluster BBAs ($\forall t \in \{1, \ldots, 4\}$, each $\tilde{\mathcal{M}}_t$ contains 5 BBAs). For comparison, we also consider the accumulated voting proposed in [14] which has proved to be more conservative/cautious than RANSAC in case of difficult settings.

Figure 12 illustrates qualitatively the results of the accumulated voting and of the proposed Algorithm 3. It clearly appears that our result is much more precise than the voting strategy proposed earlier [14]. Now, SIFT-RANSAC considering the whole sequence of 400 images provides a very confident result (ellipse with axes of a few pixels; not shown). Even if RANSAC result may appear rather good since it is actually close to the actual epipole location (58 pixels), it completely fails in estimating the actual uncertainty of its solution. Indeed, with so many data (keypoint matches accumulated through the temporal sequence), RANSAC algorithm (whatever the considered variant) will be overconfident in the obtained result missing its actual reliability. The proposed method appears then as a good compromise between perhaps too cautious results obtained using accumulated voting and too committed ones obtained using RANSAC sampling.

Quantitative evaluation is presented on Figure 13 which shows Eq. (15) with respect to λ values. In contrast to the previous experiments, in this setting we have only one result sample (obtained considering the whole sequence) so that we cannot plot error statistics (pdf, AUC). In order to check the sensitivity of our approach to the considered subsets in the sequence, we introduce different offsets (called as '-0', '-25', '-50' and '-75') in the sequence split. From Figure 13 left, we see that, due to the very large size of the obtained uncertainties, the accumulated voting completely fails in providing an interesting result. Then, zooming on low error values (Figure 13 right), we notice that, for low λ values,



(a) Whole error dynamic range

(b) Zoom with bounded error dynamic range $[0, 500]$

Figure 13: $\epsilon(\lambda)$ (Eq. (15)) versus λ; the four curves 'BBA-clustering-XX' refer to different offset values in the proposed approach, 'accumulated voting' refers to the method proposed in [14], and 'Accumulated matching' to SIFT-RANSAC considering the keypoint matches accumulated during the entire sequence.

our approach slightly outperforms RANSAC result. Indeed the distance term in Eq. (15) is about $10px$ against about $60px$. However, since uncertainty is much higher, once $\lambda > 10^{-2}$ the RANSAC solution error becomes lower. Nevertheless, as stated previously, the evidential approach result seems a good compromise between result precision and actual uncertainty. Finally, let us underline that the subset offset does not seems to impact results. More specifically, the epipole location is influenced very marginally but the focal element size does vary a little bit more (which explains the fact that error curves separate when λ increases).

7. Conclusion

In this work, our objective is to propose a fusion strategy suited for contexts in which a large number of sources, including a significant ratio of outliers, need to be combined. The adopted approach for mitigating the impact of the presence of outliers is to perform 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 af-

terwards 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. For an algorithm performing multi-modal fusion, our approach is intended to rank favourably the BBAs in the perspective of further fusion with additional sources, and indeed the experiments highlight that in this setting the promoted evidences improve the global result.

Our strategy exploits the fact that our algorithm is less committed than the standard vision-based solutions and thus more favorable to the use of additional sources. The closest applications to our work are related to pedestrian or vehicular transportation, but the underlying strategy of intra-cluster fusion may be helpful in a wider range of problems which benefit from large amounts of conflicting data sources.

Acknowledgment

This study was supported by the S^{2}UCRE^3 project (*Safety & Security of Urban Crowded Environments*), co-funded by the German BMBF grant 13N14463 and by the French ANR grant ANR-16-SEBM-0001.

References

[1] N. Snavely, S. M. Seitz, R. Szeliski, Modeling the world from internet photo collections, International journal of computer vision 80 (2) (2008) 189–210.

[2] P. Moulon, P. Monasse, 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] `https://www.s2ucre.eu/`

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

[4] B. Williams, G. Klein, I. Reid, Automatic relocalization and loop closing for real-time monocular slam, IEEE transactions on pattern analysis and machine intelligence 33 (9) (2011) 1699–1712.

[5] F. Fraundorfer, D. Scaramuzza, Visual odometry: Part ii: Matching, robustness, optimization, and applications, IEEE Robotics & Automation Magazine 19 (2) (2012) 78–90.

[6] R. Mohr, E. Arbogast, It can be done without camera calibration, Pattern recognition letters 12 (1) (1991) 39–43.

[7] Q.-T. Luong, O. D. Faugeras, On the direct determination of epipoles: A case study in algebraic methods for geometric problems, in: Proceedings of 12th International Conference on Pattern Recognition, Vol. 1, IEEE, 1994, pp. 243–247.

[8] A. Verri, E. Trucco, Finding the epipole from uncalibrated optical flow, Image and vision computing 17 (8) (1999) 605–609.

[9] L. Puig, J. J. Guerrero, Self-location from monocular uncalibrated vision using reference omniviews, in: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2009, pp. 5216–5221.

[10] J. Bentolila, J. M. Francos, Conic epipolar constraints from affine correspondences, Computer Vision and Image Understanding 122 (2014) 105–114.

[11] A. Ramirez, E. Ohn-Bar, M. M. Trivedi, Go with the flow: Improving multi-view vehicle detection with motion cues, in: 2014 22nd International Conference on Pattern Recognition, IEEE, 2014, pp. 4140–4145.

[12] S. Ardeshir, A. Borji, Egocentric meets top-view, IEEE transactions on pattern analysis and machine intelligence 41 (6) (2018) 1353–1366.

[13] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381–395.

[14] H. Chen, E. Aldea, 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.

[15] M. Lachaize, S. Le Hégarat-Mascle, E. Aldea, A. Maitrot, R. Reynaud, Evidential split-and-merge: Application to object-based image analysis, International Journal of Approximate Reasoning 103 (2018) 303–319.

[16] N. Helal, F. Pichon, D. Porumbel, D. Mercier, É. Lefèvre, The capacitated vehicle routing problem with evidential demands, International Journal of Approximate Reasoning 95 (2018) 124–151.

[17] C. André, S. Le Hégarat-Mascle, R. Reynaud, Evidential framework for data fusion in a multi-sensor surveillance system, Engineering Applications of Artificial Intelligence 43 (2015) 166–180.

[18] Q. Gao, D.-L. Xu, An empirical study on the application of the evidential reasoning rule to decision making in financial investment, Knowledge-Based Systems 164 (2019) 226–234.

[19] Z.-G. Liu, Y. Liu, J. Dezert, F. Cuzzolin, Evidence combination based on credal belief redistribution for pattern classification, IEEE Transactions on Fuzzy Systems 28 (4) (2019) 618–631.

[20] T. Denœux, 40 years of dempster-shafer theory, International Journal of Approximate Reasoning 79 (2016) 1–6.

[21] W. Rekik, S. Le Hégarat-Mascle, R. Reynaud, A. Kallel, A. B. Hamida, Dynamic object construction using belief function theory, Information Sciences 345 (2016) 129–142.

[22] S. Zair, S. Le Hégarat-Mascle, Evidential framework for robust localization using raw gnss data, Engineering Applications of Artificial Intelligence 61 (2017) 126–135.

[23] N. Pellicanò, S. Le Hégarat-Mascle, E. Aldea, 2cobel: A scalable belief function representation for 2d discernment frames, International Journal of Approximate Reasoning 103 (2018) 320–342.

[24] L. Jaulin, M. Kieffer, O. Didrit, E. Walter, Interval analysis, in: Applied interval analysis, Springer, 2001, pp. 11–43.

[25] P. Smets, The combination of evidence in the transferable belief model, IEEE Transactions on pattern analysis and machine intelligence 12 (5) (1990) 447–458.

[26] G. Shafer, A mathematical theory of evidence, Vol. 42, Princeton university press, 1976.

[27] R. R. Yager, On the dempster-shafer framework and new combination rules, Information sciences 41 (2) (1987) 93–137.

[28] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational intelligence 4 (3) (1988) 244–264.

[29] R. R. Yager, Quasi-associative operations in the combination of evidence, Kybernetes.

[30] F. Pichon, S. Destercke, T. Burger, A consistency-specificity trade-off to select source behavior in information fusion, IEEE transactions on cybernetics 45 (4) (2014) 598–609.

[31] Y. Zhao, R. Jia, P. Shi, A novel combination method for conflicting evidence based on inconsistent measurements, Information Sciences 367 (2016) 125–142.

[32] V. Drevelle, P. Bonnifait, A set-membership approach for high integrity height-aided satellite positioning, GPS solutions 15 (4) (2011) 357–368.

[33] T. Denoeux, Distributed combination of belief functions, Information Fusion 65 (2021) 179–191.

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

[35] J. Schubert, Clustering decomposed belief functions using generalized weights of conflict, International Journal of Approximate Reasonning 48 (2) (2008) 466–480.

[36] A.-L. Jousselme, P. Maupin, Distances in evidence theory: Comprehensive survey and generalizations, International Journal of Approximate Reasoning 53 (2) (2012) 118–145.

[37] T. Denœux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artificial Intelligence 172 (2-3) (2008) 234–264.

[38] T. Denœux, Inner and outer approximation of belief structures using a hierarchical clustering approach, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9 (04) (2001) 437–460.

[39] D. Dubois, H. Prade, Consonant approximations of belief functions, International Journal of Approximate Reasoning 4 (5-6) (1990) 419–449.

[40] P. Smets, R. Kennes, The transferable belief model, Artificial intelligence 66 (2) (1994) 191–234.

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

[42] R. I. Hartley, In defense of the eight-point algorithm, IEEE Transactions on pattern analysis and machine intelligence 19 (6) (1997) 580–593.

[43] F. Sur, N. Noury, M.-O. Berger, Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation, in: BMVC 2008, 2008, p. 10.

[44] S. Julier, J. Uhlmann, H. F. Durrant-Whyte, A new method for the nonlinear transformation of means and covariances in filters and estimators, IEEE Transactions on automatic control 45 (3) (2000) 477–482.

[45] T. Papadopoulo, M. I. Lourakis, Estimating the jacobian of the singular value decomposition: Theory and applications, in: ECCV, Springer, 2000, pp. 554–570.

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

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

[48] T. Denœux, O. Kanjanatarakul, S. Sriboonchitta, Ek-nnclus: a clustering procedure based on the evidential k-nearest neighbor rule, Knowledge-Based Systems 88 (2015) 57–69.

[49] M.-H. Masson, T. Denoeux, Ecm: An evidential version of the fuzzy c-means algorithm, Pattern Recognition 41 (4) (2008) 1384–1397.

[50] W. H. Day, H. Edelsbrunner, Efficient algorithms for agglomerative hierarchical clustering methods, Journal of classification 1 (1) (1984) 7–24.

[51] R. Sibson, Slink: an optimally efficient algorithm for the single-link cluster method, The computer journal 16 (1) (1973) 30–34.

[52] H. K. Seifoddini, Single linkage versus average linkage clustering in machine cells formation applications, Computers & Industrial Engineering 16 (3) (1989) 419–426.

[53] D. Defays, An efficient algorithm for a complete link method, The Computer Journal 20 (4) (1977) 364–366.

[54] W. Maalel, K. Zhou, A. Martin, Z. Elouedi, Belief hierarchical clustering, in: International Conference on Belief Functions, Springer, 2014, pp. 68–76.

[55] A.-L. Jousselme, D. Grenier, É. Bossé, A new distance between two bodies of evidence, Information fusion 2 (2) (2001) 91–101.

[56] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, S. Tubaro, Deep convolutional neural networks for pedestrian detection, Signal processing: image communication 47 (2016) 482–489.

[57] C. Li, X. Wang, W. Liu, Neural features for pedestrian detection, Neurocomputing 238 (2017) 420–432.

[58] T. Zou, S. Yang, Y. Zhang, M. Ye, Attention guided neural network models for occluded pedestrian detection, Pattern Recognition Letters 131 (2020) 91–97.

[59] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, P. Fua, Learning to find good correspondences, CVPR (2018) 2666–2674.

[60] T. Chavdarova, P. Baqué, S. Bouquet, A. Maksai, C. Jose, T. Bagautdinov, L. Lettry, P. Fua, L. Van Gool, F. Fleuret, Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5030–5039.

[61] 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, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[62] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767.

Appendix

Proof. Jousselme's distance $d_J(m, \tilde{m})$ between BBAs m and $\tilde{m}$ is equal to $\sqrt{\frac{1}{2}\left(\langle m, m\rangle_J + \langle \tilde{m}, \tilde{m}\rangle_J - 2\langle m, \tilde{m}\rangle_J\right)}$, with

$$\langle m_i, m_j\rangle_J = \sum_{H_i \in \mathcal{F}(m_i)} \sum_{H_j \in \mathcal{F}(m_j)} \frac{|H_i \cap H_j|}{|H_i \cup H_j|} m_i(H_i)\, m_j(H_j),$$

where $\mathcal{F}(m_i)$ and $\mathcal{F}(m_j)$ are the sets of focal elements of m_i and m_j, respectively. If m and $\tilde{m}$ are such that $\tilde{m}$ corresponds to an aggregation of two focal elements of m, i.e. $\exists (A, B) \in \mathcal{F}(m) \times \mathcal{F}(m), A \neq B$ s.t.

$$\begin{cases} \tilde{m}(A) = \tilde{m}(B) = 0, \\ \tilde{m}(A \cup B) = m(A \cup B) + m(A) + m(B), \\ \forall H \in \Omega \setminus \{A, B, A \cup B\}, \tilde{m}(H) = m(H). \end{cases}$$

Using the linearity of the scalar product, we have

$$\begin{aligned} \langle m, m\rangle_J + \langle \tilde{m}, \tilde{m}\rangle_J - 2\langle m, \tilde{m}\rangle_J &= \langle m, m - \tilde{m}\rangle_J + \langle \tilde{m}, \tilde{m} - m\rangle_J, \\ &= \langle m, \delta m\rangle_J - \langle \tilde{m}, \delta m\rangle_J, \end{aligned}$$

with $\delta m = m - \tilde{m}$, non null only for $H \in \{A, B, A \cup B\}$ with $\delta m(A) = m(A)$, $\delta m(B) = m(B)$ and $\delta m(A \cup B) = -m(A) - m(B)$.
Introducing the notation $k(H_1, H_2) = \frac{|H_1 \cap H_2|}{|H_1 \cup H_2|}$,

$$\begin{aligned} \langle m, \delta m\rangle_J &= \textstyle\sum_{(H) \in \mathcal{F}(m)} \quad k(A, H) m(H) m(A) + k(B, H) m(H) m(B) + \\ &\qquad\qquad k(A \cup B, H) m(H)(-m(A) - m(B)), \\ \langle \tilde{m}, \delta m\rangle_J &= \textstyle\sum_{(H) \in \mathcal{F}(\tilde{m})} \quad k(A, H) \tilde{m}(H) m(A) + k(B, H) m(H) \tilde{m}(B) + \\ &\qquad\qquad k(A \cup B, H) \tilde{m}(H)(-m(A) - m(B)) \end{aligned}$$

Thus, with $\mathcal{F}^\star(m) = \mathcal{F}(m) \setminus \{A, B, A \cup B\}$ and $\mathcal{F}^\star(\tilde{m}) = \mathcal{F}(\tilde{m}) \setminus \{A \cup B\}$ (A

and $B \notin \mathcal{F}(\tilde{m})$),

$$\begin{aligned}
\langle m, \delta m\rangle_J - \langle \tilde{m}, \delta m\rangle_J &= \sum_{H\in\mathcal{F}^\star(m)}[\quad k(A,H)m(A)(m(H)-\tilde{m}(H))+ \\
&\qquad\qquad k(B,H)m(B)(m(H)-\tilde{m}(H))+ \\
&\qquad\qquad k(A\cup B,H)(m(A)m(B))(\tilde{m}(H)-m(H))] \\
&+ \sum_{H\in\{A,B,A\cup B\}}[\quad k(A,H)m(A)(m(H)-\tilde{m}(H))+ \\
&\qquad\qquad k(B,H)m(B)(m(H)-\tilde{m}(H))+ \\
&\qquad\qquad k(A\cup B,H)(m(A)m(B))(\tilde{m}(H)-m(H))]
\end{aligned}$$

Since $\mathcal{F}^\star(m) = \mathcal{F}^\star(\tilde{m})$ and $\forall H \in \mathcal{F}^\star(m), m(H) = \tilde{m}(H)$, then the first sum in previous equation is equal to 0, so that, remarking that $k(X,X) = 1, \forall X \in 2^\Omega$, developing and gathering terms, we obtain

$$\begin{aligned}
\langle m, \delta m\rangle_J - \langle \tilde{m}, \delta m\rangle_J &= m^2(A) + 2k(A,B)m(A)m(B) - 2k(A,A\cup B)m(A)(m(A)+m(B)) \\
&\quad +m^2(B) - 2k(B,A\cup B)m(B)(m(A)+m(B)) + (m(A)+m(B))^2, \\
&= 2m^2(A)\left[1-\tfrac{|A|}{|A\cup B|}\right] + 2m^2(B)\left[1-\tfrac{|B|}{|A\cup B|}\right] + \\
&\quad 2m(A)m(B)\left[1+\tfrac{|A\cap B|}{|A\cup B|}-\tfrac{|A|}{|A\cup B|}-\tfrac{|B|}{|A\cup B|}\right].
\end{aligned}$$

Finally, since $|A\cup B| = |A| + |B| - |A\cap B|$, we find

$$\tfrac{1}{2}\left(d_J\left(m,\tilde{m}\right)\right)^2 = m^2(A)\left[1-\tfrac{|A|}{|A\cup B|}\right] + m^2(B)\left[1-\tfrac{|B|}{|A\cup B|}\right].$$

925 □