Practical Belief Functions: Recent tools and examples of challenging applications

Sylvie Le Hégarat-Mascle

Université Paris-Saclay

sylvie.le-hegarat@universite-paris-saclay.fr

7 janvier 2026

Overview

1. Basics in belief functions
   - Representation of evidences
   - Combination of evidences
   - Decision
2. Belief functions for image processing
   - Spatial information as an information piece
   - Modeling ambiguities between classes
   - Consistency at region level
3. Belief functions for localization
   - Managing outlier issue
   - Handling 2D discernment frame
4. Belief functions for Deep Learning

Outline

1. Basics in belief functions
   - Representation of evidences
   - Combination of evidences
   - Decision
2. Belief functions for image processing
3. Belief functions for localization
4. Belief functions for Deep Learning

Information fusion

Kinds of tasks

- "lower-level tasks" (i.e., close to sensors) : detection, tracking, classification
- "higher-level tasks" (i.e., close to humans) : identification, threat/situation assessment, etc.

Expected benefits of fusion

- Source partial redundancy → capability to verify information from different (independent) sources, detect possible outliers (or deception) from conflicting information
- Source partial complementary → capability to reduce uncertainty (refine classes/positions/estimations, reduce ambiguities) of the information provided by individual sources
- Require (fusion challenge) correct consideration of the imperfections of the information and sources

Different facets of uncertainty

Klir and Yuan's typology of uncertainty (1995)

- *Fuzziness* = "lack of definite or sharp distinctions"
- *Ambiguity* = "one-to-many relationships"
- *Non-specificity* = "two or more alternatives [that] are unspecified"
- *Discord* = "disagreement in choosing among several alternatives"

Information quality

- *Specificity* = degree of precision/uniqueness
- *Consistency* = lack of internal conflict
- *Correctness* = agreement with Ground Truth
- *Credibility* = believable feature
- *Relevance* = usefulness

Source quality

- *Reputation* = general assessment
- *Confidence* = degree of belief in its own statements
- *Competency* = ability to provide specific information
- *Truthfulness* = honesty
- *Reliability* = degree to which information can be trusted

Historical aspects

Seminal works

1968 "A generalization of Bayesian inference", A. Dempster → statistical inference

1976 "A Mathematical Theory of Evidence", G. Shafer → epistemic uncertainty

⇒ Dempster-Shafer evidence theory

1994 "The transferable belief model", Ph. Smets

Increasing number of theoretical and applicative works

2016 : "40 years of Dempster-Shafer theory" (T. Denœux) ⇒

- Interpretation and relationships with alternative theories
- Reasoning tools : combination rules, conflict analysis, correction mechanisms, uncertainty propagation, graphical models, BBA approximation
- Applications : statistical inference, classification, regression, computer vision, filtering and tracking, multi-criteria decision making, etc.

Key ideas

- Ability to represent information even when they are vague, imprecise in addition to be possibly uncertain
- Higher levels of precision and certainty will come from the combination of several information pieces
- A belief function may be viewed as both a non additive measure (additiviy : $P(A) + P(\bar{A}) = 1$) and a generalized set
- Basic mechanisms for reasoning with belief functions extend both probabilistic operations (such as marginalization and conditioning) and set-theoretic operations (such as intersection and union)
- Rough links with
  - probability $\Leftrightarrow$ only singleton elements can be focal elements
  - possibility $\Leftrightarrow$ focal elements are all nested
  - interval analysis $\Leftrightarrow$ only one focal element per source

Outline

1. Basics in belief functions
   - Representation of evidences
   - Combination of evidences
   - Decision

Discernment frame

Discrete case

- Discernment frame $\Omega = \{\omega_0, \ldots \omega_N\}$ with ω_i hypotheses that are mutually exclusive and that can be solution of the considered problem
- *Open word* assumption $\Rightarrow \Omega$ may be non exhaustive
- 2^Ω = set of Ω subsets : $2^\Omega = \{A, A \subseteq \Omega\}$
  - *singleton* hypotheses = elements of 2^Ω having cardinality equal to 1 (Ω elements)
  - disjunctive or *compound hypotheses* = elements of 2^Ω with cardinality greater than 1

Continuous case

- $\Omega \subseteq \mathbb{R}^d$ with $d \geq 1$ (practically $d \leq 2$)
- Example : $\Omega = \mathbb{R}$ but only intervals will be 'believable' elements

Machine representation of hypotheses (Discrete case)

- Binary words : $N = |\Omega|, \forall k \in \left[0, 2^N - 1\right], k = \Sigma_{i \in [1,N]} 2^i \times k_i$ with $k_i = 1$ if and only if the kth singleton is included in hypothesis ω_k
- Use of bitwise operators for set operations
- Example
  $\Omega = \{\omega_1, \omega_2, \omega_3\} \Rightarrow \emptyset = 000,\ \omega_1 = 001,\ \omega_2 = 010,\ \{\omega_1, \omega_2\} = 011,$
  $\omega_3 = 100,\ \{\omega_1, \omega_3\} = 101,\ \{\omega_2, \omega_3\} = 110,\ \Omega = 111$
  $\{\omega_1, \omega_2\} \cap \{\omega_1, \omega_3\} \leftrightarrow 011\ \&\ 101 = 001 \leftrightarrow \omega_1$
  $\{\omega_1, \omega_2\} \cup \{\omega_1, \omega_3\} \leftrightarrow 011\ |\ 101 = 111 \leftrightarrow \Omega$

Basic Belief Assignment (BBA)

- A mass function m^Ω is such that :

$$m^\Omega : 2^\Omega \to [0,1] \text{ and } \sum_{A \in 2^\Omega} m^\Omega(A) = 1$$

- $m^\Omega(A)$ represents the belief allocated to the set of hypotheses included in A without possibility to distinguish them.
- *Focal Elements* (FE) : hypotheses $A \in 2^\Omega$ such that $m^\Omega(A) > 0$

Semantics of the mass function

- Meta-knowledge on a source → interpretation of its information

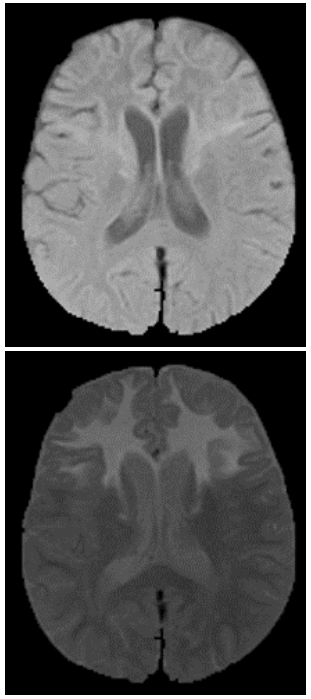
⇒ Random code formalism (Denœux, 2013) :

  - Decoder function (multi-valued mapping) $\Gamma : \mathscr{Y} \to 2^\Omega$
  - The source provides us information as an encoded message using a code in $\mathscr{Y}$ selected at random according to some known probability measure on $\mathbb{P}$
  - $\forall A \in 2^\Omega, m^\Omega(A) = \mathbb{P}(\{y \in \mathscr{Y} \,|\, \Gamma(y) = A\})$

Class ambiguities in multisource classification

Segmentation of brain tissues of patients suffering from adrenoleukodystrophy (Bloch, 1996)

- $\Omega = \{C_1, C_2, C_3\}$ with C_1 = Ventricles and cerebro-spinal fluid, C_2 = White Matter and Grey Matter, C_3 = ALD
- First MRI cannot distinguish C_2 from C_3, second MRI cannot distinguish C_1 from C_2
- $\rightarrow \{C_2, C_3\}$ FE in MRI 1, and $\{C_1, C_2\}$ FE in MRI 2



Segmentation of vegetation in SAR images

- C_1 = dense vegetation, C_2 = bare soil, C_3 = sparse vegetation, m_1 BBA band L, m_2 BBA band C,
- $\begin{cases} m_1(\Omega) = 1 - \alpha_1, \\ m_1(\{C_2, C_3\}) = \alpha_1 \times t, \\ m_1(\{C_1\}) = \alpha_1 \times (1-t), \end{cases}$ $\begin{cases} m_2(\Omega) = 1 - \alpha_2, \\ m_2(\{C_1, C_3\}) = \alpha_2 \times u, \\ m_2(\{C_2\}) = \alpha_2 \times (1-u), \end{cases}$ with $(t, u, \alpha_1, \alpha_2) \in [0,1]^4$, t and u depend on pixel

Correction of a mass function (Pichon, 2012)

Mass function can be modified according to meta-knowledge about source relevance and truthfulness :

- Mass function m^{Ω} defined under the hypothesis of reliable source
- Now, we learn that the source is relevant with probability $p_{re} = \alpha$ and truthful with probability $p_{th} = \beta$, $(\alpha, \beta) \in [0,1]^2$
- Function $\Gamma : 2^{\Omega} \rightarrow 2^{\Omega}$: $\forall A \in 2^{\Omega}, \Gamma(A|\text{re},\text{th}) = A$, $\Gamma(A|\overline{\text{re}}) = \Omega$, $\Gamma(A|\text{re},\overline{\text{th}}) = \overline{A}$
- $\forall A \in 2^{\Omega}, m^{\Omega}_{\Gamma}(A) = \alpha\beta m^{\Omega}(A) + \alpha(1-\beta) m^{\Omega}(\overline{A}) + [\![A = \Omega]\!](1-\alpha)$ where $[\![.]\!]$ denotes the Iverson brackets

Example : Thief investigation

- For a theft committed, there are three suspects : Arthur (A), Bastien (B) and Chloé (C).
- A first witness (T_1) claims that the thief was a man. A second witness (T_2) claims that Arthur was with him. A third witness (T_3) is 90% sure that the thief is not Bastien.
- A fourth witness stated that T_1 had only 80% reliable sight, and that T_2 was able to lie with a 50% probability.
- Belief function modeling encoding both the reliability (according to meta-knowledge) and meaning of a piece of evidence

$$\Rightarrow \begin{cases} m_1^{\Omega}(\{A,B\}) = 0.8 & // \text{ thief was a man according to } T_1 \\ m_1^{\Omega}(\Omega) = 0.2 & // \ T_1 \text{ sight defect} \\ m_2^{\Omega}(\{B,C\}) = 0.5 & // \text{ thief is not Arthur according to } T_2 \\ m_2^{\Omega}(A) = 0.5 & // \ T_2 \text{ lies} \\ m_3^{\Omega}(\{A,C\}) = 0.9 & // \text{ thief is not Bastien according to } T_3 \\ m_3^{\Omega}(\Omega) = 0.1 & // \ T_3 \text{ is not completely affirmative} \end{cases}$$

Examples of usual BBA

- *Normal* BBA : $m^{\Omega}(\emptyset) = 0$; otherwise, the BBA is called sub-normal
- *Categorical* (or *logical*) BBA : BBA having only one FE $A \in 2^{\Omega}$: $m^{\Omega}(A) = 1$ and $m^{\Omega}(B) = 0$ $\forall B \in 2^{\Omega}$, $B \neq A$; denoted m^{Ω}_{A}
- *Vacuous* BBA : Categorical BBA having Ω as unique FE : $m^{\Omega}(\Omega) = 1$ (represents complete ignorance)
- *Dogmatic* BBA : $m^{\Omega}(\Omega) = 0$
- *Simple* BBA : BBA having two and only two FEs among them Ω
- *Bayesian* BBA : BBA having only singleton hypotheses as FEs : $\forall A \in 2^{\Omega}$ such that $m^{\Omega}(A) > 0 \Rightarrow A \in \Omega$
- *Consonant* BBA : BBA having only nested FEs : $\forall (A, B) \in 2^{\Omega} \times 2^{\Omega}$ such that $m^{\Omega}(A) > 0$ and $m^{\Omega}(B) > 0 \Rightarrow A \subseteq B$ or $B \subseteq A$

Credibility, Implicability

Credibility

- Derived from the probability that an evidence implies A
- $\forall A \in 2^{\Omega}, bel^{\Omega}(A) = \sum_{B \subseteq A, B \neq \emptyset} m^{\Omega}(B)$
- $bel^{\Omega}(A)$ may be interpreted as the degree of support (belief) that is specific to A rather than $\overline{A}$, the complement of A in Ω

Implicability

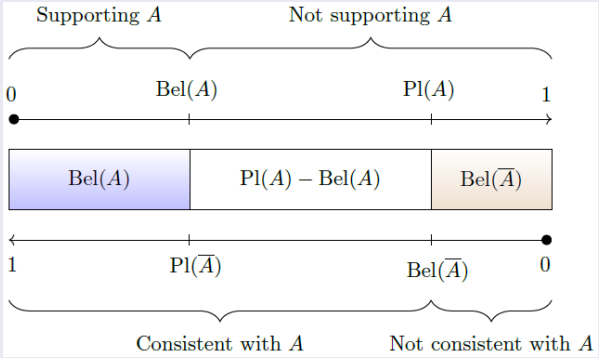
- $\forall A \in 2^{\Omega}, b^{\Omega}(A) = \sum_{B \subseteq A} m^{\Omega}(B)$
- $\forall A \in 2^{\Omega}, b^{\Omega}(A) = bel^{\Omega}(A) + m^{\Omega}(\emptyset)$
- Mainly used for computational conveniences

Plausibility, Commonality

Plausibility

- Derived from the probability that the evidence does not contradict A
- $\forall A \in 2^{\Omega}, pl^{\Omega}(A) = \sum_{B \cap A \neq \emptyset} m^{\Omega}(B)$
- $pl^{\Omega}(A)$ may be interpreted as the support degree that is compatible with A
- Normal BBA :

$$pl^{\Omega}(A) = 1 - bel^{\Omega}(\overline{A})$$



Commonality

- $\forall A \in 2^{\Omega}, q^{\Omega}(A) = \sum_{B \supseteq A} m^{\Omega}(B)$
- Ignorance degree after conditioning on A
- Mainly used for computational conveniences

Relationships between belief functions

- One-to-one relationships → equivalent representations of evidence

	$m(A)$	$pl(A)$	$bel(A)$	$q(A)$
$m(A)$	$=$	$\sum_{B\subseteq A}(-1)^{\lvert A\rvert-\lvert B\rvert+1}pl(\overline{B})$	$\sum_{\emptyset\neq B\subseteq A}(-1)^{\lvert A\rvert-\lvert B\rvert}bel(B)$	$\sum_{B\supseteq A}(-1)^{\lvert B\rvert-\lvert A\rvert}q(B)$
$pl(A)$	$\sum_{B\cap A\neq\emptyset}m(B)$	$=$	$bel(\Omega)-bel(\overline{A})$	
$bel(A)$	$\sum_{B\subseteq A,B\neq\emptyset}m(B)$		$=$	$\sum_{B\subseteq\overline{A}}(-1)^{\lvert\overline{B}\rvert}q(B)$
$b(A)$	$\sum_{B\subseteq A}m(B)$	$1-pl(\overline{A})$	$bel(A)+m(\emptyset)$	
$q(A)$	$\sum_{B\supseteq A}m(B)$			$=$

NB : All functions are defined in 2^{Ω} (even if omitted for notation shortness)

- Matrix calculus → derivation of a given belief function from another one and the adequate "transfer" matrix

(Dis)Simililarity between BBAs (1/2)

Three main ways of quantitatively comparing belief functions

- Minkowski family of distances :
  $\delta_M^{(\alpha)}(m_1, m_2) = \left(\sum_{A \subseteq \Omega} \left[\left(f_1 - f_2\right)(A)\right]^\alpha\right)^{\frac{1}{\alpha}}$, where $\alpha \in [1, +\infty)$ and f_1 and f_2 are transformations of m_1 and m_2
- Conflict (cf. § Conjunctive combination of BBAs)
- Inner product and cross-entropy

Examples of Minkowski distances

- Manhattan distances (L_1), e.g. $d_{bel}^{(1)}(m_1^\Omega, m_2^\Omega) \triangleq \sum_{A \in 2^\Omega} \left| bel_1^\Omega(A) - bel_2^\Omega(A) \right|$, $d_{pl}^{(1)}(m_1^\Omega, m_2^\Omega) \triangleq \sum_{A \in 2^\Omega} \left| pl_1^\Omega(A) - pl_2^\Omega(A) \right|$
- Chebyshev distances (L_∞), e.g. $d_T^{(\infty)}(m_1^\Omega, m_2^\Omega) \triangleq \max_{A \in 2^\Omega} \left| BetP_1^\Omega(A) - BetP_2^\Omega(A) \right|$ with $BetP^\Omega(A) \triangleq \sum_{B \in 2^\Omega \setminus \emptyset} \frac{|A \cap B|}{|B|} m^\Omega(B)$ → Tessem distance

(Dis)Simililarity between BBAs (2/2)

Case of Euclidian distances (L_2)

- General formulation based on inner product $\left\langle m_i^\Omega, m_j^\Omega \right\rangle_W \rightarrow$
$$d_W^{(2)}(m_1^\Omega, m_2^\Omega) \triangleq \left[\frac{1}{2}\left(\langle m_1^\Omega, m_1^\Omega\rangle_W + \langle m_2^\Omega, m_2^\Omega\rangle_W - 2\langle m_1^\Omega, m_2^\Omega\rangle_W\right)\right]^{\frac{1}{2}}$$
- Examples : Fixsen & Mahler (FM) distance, Jousselme (J) distance
  - $\left\langle m_i^\Omega, m_j^\Omega \right\rangle_{FM} \triangleq \sum_{A\in 2^\Omega\setminus\emptyset}\sum_{B\in 2^\Omega\setminus\emptyset} m_i^\Omega(A)m_j^\Omega(B)\frac{|A\cap B|}{|A||B|}$
  - $\left\langle m_i^\Omega, m_j^\Omega \right\rangle_{J} \triangleq \sum_{A\in 2^\Omega\setminus\emptyset}\sum_{B\in 2^\Omega\setminus\emptyset} m_i^\Omega(A)m_j^\Omega(B)\frac{|A\cap B|}{|A\cup B|}$

Inconsistency measures

- $\phi(A,B)$ a consistency index between hypotheses A and B, e.g. $\phi(A,B)=1$ if $A\cap B\neq\emptyset$ and 0 otherwise, $\phi(A,B)=\frac{|A\cap B|}{|A\cup B|}$ (Jaccard index), etc.
- Not consistency $\delta_2^2(m_1,m_2) = 1-\sum_{A,B\subseteq\Omega} m_1(A)m_2(B)\phi(A,B)$
- Cross-entropy $\delta_1^2(m_1,m_2) = \sum_{A\subseteq\Omega} m_1(A)\left(-log\sum_{B\subseteq\Omega} m_2(B)\phi(A,B)\right)$

BBA ordering and Least Commitment Principle (LCP)

Different BBA orderings, e.g.

- pl-ordering : $m_1^\Omega \sqsubseteq_{pl} m_2^\Omega \Leftrightarrow \forall A \in 2^\Omega, pl_1^\Omega(A) \leq pl_2^\Omega(A)$
- q-ordering : $m_1^\Omega \sqsubseteq_q m_2^\Omega \Leftrightarrow \forall A \in 2^\Omega, q_1^\Omega(A) \leq q_2^\Omega(A)$
- S-ordering :

$$m_1^\Omega \sqsubseteq_S m_2^\Omega \Leftrightarrow \exists S \text{ s.t. } \begin{cases} \forall (A,B) \in 2^\Omega \times 2^\Omega, S(A,B) > 0 \Rightarrow A \subseteq B \\ \forall B \in 2^\Omega, \sum_{A \in 2^\Omega} S(A,B) = 1 \\ \forall A \in 2^\Omega, m_1^\Omega(A) = \sum_{B \in 2^\Omega} S(A,B) m_2^\Omega(B) \end{cases}$$

- Some orderings are 'stronger', e.g. $m_1^\Omega \sqsubseteq_S m_2^\Omega \Rightarrow \begin{cases} m_1^\Omega \sqsubseteq_{pl} m_2^\Omega \\ m_1^\Omega \sqsubseteq_q m_2^\Omega \end{cases}$
- Vacuous BBA is the greatest element (less committed) for all these orderings

LCP

When 2 belief functions are compatible with all the constraints, the least committed must be favored

Approximation of BBAs (1/2)

Objective

- Constraint the number of focal elements (FE) of BBAs to be $\leq m \in \mathbb{N}_+ \rightarrow$ Given a BBA m^Ω, having n FEs, find $\tilde{m}^\Omega$ with less than m FEs and 'as close as possible' to m^Ω
- LCP $\Rightarrow \tilde{m}^\Omega$ less committed than m^Ω

Popular techniques

- Summarization
  - Basic idea : replace the $n-m+1$ FEs having lowest mass values by a single FE that is the union of removed FEs $\Rightarrow$ less committed BBA
  - Main drawback : increase of the plausibility of the removed elements !
- Iterative aggregation
  - Basic idea : transfer the mass of a pair of FEs to their union ; repeat until BBA has less than m FEs
  - Various criteria for FE pair selection

Approximation of BBAs (2/2)

Examples of criterion for FE pair selection in iterative aggregation

- $\mathscr{F}_m$ = set of focal elements
- Minimization of the credibility variation :
$$\underset{A,B\in\mathscr{F}_m^2, A\neq B}{\operatorname{argmin}} \quad 2^{|\Omega\setminus A|}m^{\Omega}(A)+2^{|\Omega\setminus B|}m^{\Omega}(B)-2^{|\Omega\setminus A\cup B|}(m^{\Omega}(A)+m^{\Omega}(B))$$
- Minimization of the contour function variation :
$$\underset{A,B\in\mathscr{F}_m^2, A\neq B}{\operatorname{argmin}} \quad m^{\Omega}(A)\times(|A\cup B|-|A|)+m^{\Omega}(B)\times(|A\cup B|-|B|)$$
- Minimum distance between BBAs
  - Jousselme's distance case →
$$\underset{A,B\in\mathscr{F}_m^2, A\neq B}{\operatorname{argmin}} \quad (m^{\Omega}(A))^2\times\left(1-\frac{|A|}{|A\cup B|}\right)+(m^{\Omega}(B))^2\times\left(1-\frac{|B|}{|A\cup B|}\right)$$

Outline

1. Basics in belief functions
   - Representation of evidences
   - Combination of evidences
   - Decision

Conditioning

- Let m^{Ω} be a BBA on Ω. Learning that the solution is not in $\overline{A}$, we derive $m^{\Omega}[A]$ such that :
  - $m^{\Omega}[A]$ is a specialization of m^{Ω}
  - $pl[A](\overline{A}) = 0$ (the complementary of A is no longer plausible)
  - $m^{\Omega}[A]$ is the less committed among solutions verifying 2 first items
- $m^{\Omega}[A](B) = \begin{cases} \sum_{C \subseteq \overline{A}} m^{\Omega}(B \cup C) & \forall B \subseteq A \\ 0 & \forall B \not\subseteq A \end{cases}$
- For all FEs of m^{Ω}, transfer of the mass toward the intersection with A (Transferable Belief Model, Smets)
- Deconditioning operation → less committed BBA whose conditioning on A is $m^{\Omega}[A]$: $\forall B \subseteq A, m^{\Omega}[\cancel{A}](\left\{B, \overline{A}\right\}) = m^{\Omega}[A](B)$

Usual combination rules

Formulation

- Smets' conjunctive combination :
  $\forall A \in 2^\Omega, m^\Omega_{1 \textcircled{\cap} 2}(A) = \sum_{\substack{(B,C) \in 2^\Omega \times 2^\Omega \\ B \cap C = A}} m^\Omega_1(B) m^\Omega_2(C)$
  $\forall A \in 2^\Omega \setminus \{\emptyset\}, q^\Omega_{1 \textcircled{\cap} 2}(A) = q^\Omega_1(A) q^\Omega_2(A) \Rightarrow$ decombination : $q^\Omega_{1 \cancel{\textcircled{\emptyset}} 2}(A) = q^\Omega_{1 \textcircled{\cap} 2}(A) / q^\Omega_2(A)$
- Dempster's rule (orthogonal sum) : normal BBA such that if $K \triangleq \sum_{\substack{(B,C) \in 2^\Omega \times 2^\Omega \\ B \cap C = \emptyset}} m^\Omega_1(B) m^\Omega_2(C) < 1$, then
  $\forall A \in 2^\Omega, m^\Omega_{1 \oplus 2}(A) = \frac{1}{1-K} m^\Omega_{1 \textcircled{\cap} 2}(A)$

Properties

- Assumes independence between pieces of evidence
- Commutativity, associativity, neutral element : m^Ω_Ω
- $K \in [0,1]$ is the degree of conflict between m^Ω_1 and m^Ω_2

Other combination rules

Disjunctive combination

- $\forall A \in 2^\Omega, m^\Omega_{1 \textcircled{\cup} 2}(A) = \sum_{\substack{(B,C) \in 2^\Omega \times 2^\Omega \\ B \cup C = A}} m^\Omega_1(B) m^\Omega_2(C)$

  $\forall A \in 2^\Omega \setminus \emptyset, b^\Omega_{1 \textcircled{\cup} 2}(A) = b^\Omega_1(A) b^\Omega_2(A)$
- Assumes at least one piece of evidence is reliable
- Commutativity, associativity, neutral element : $m^\Omega_\emptyset$

α-conjunctions and α-disjunctions $(\alpha \in [0,1])$

- α-conjunctions : $\forall X \in 2^\Omega$,

  $m^\Omega_{1 \textcircled{\cap}^\alpha 2}(X) = \sum_{\substack{(A,B,C) \in 2^\Omega \times 2^\Omega \times 2^\Omega \\ (A \cap B \cap \overline{C}) \cup (\overline{A} \cap \overline{B} \cap C) = X}} m^\Omega_1(A) m^\Omega_2(B) \alpha^{|\overline{C}|} (1-\alpha)^{|C|}$
- $\alpha = 1 \rightarrow$ non null term only for $C = \emptyset \rightarrow$ conjunctive rule,

  $\alpha = 0 \rightarrow$ non null term only for $C = \Omega \rightarrow$ logical equivalence

Hybrid combination rules : examples

Formulation

- Yager rule : normal BBA such that $\forall A \in 2^{\Omega} \setminus \{\emptyset\}$,
$$m^{\Omega}_{1 \textcircled{Y} 2}(A) = \sum_{\substack{(B,C)\in(2^{\Omega})^2,\\ B\cap C=A}} m_1^{\Omega}(B)\, m_2^{\Omega}(C) + [\![A=\Omega]\!] \sum_{\substack{(B,C)\in(2^{\Omega})^2,\\ B\cap C=\emptyset}} m_1^{\Omega}(B)\, m_2^{\Omega}(C)$$
- Dubois & Prade rule : normal BBA such that $\forall A \in 2^{\Omega}$,
$$m^{\Omega}_{1 \textcircled{DP} 2}(A) = \sum_{\substack{(B,C)\in(2^{\Omega})^2,\\ B\cap C=A}} m_1^{\Omega}(B)\, m_2^{\Omega}(C) + \sum_{\substack{(B,C)\in(2^{\Omega})^2,\\ B\cup C=A, B\cap C=\emptyset}} m_1^{\Omega}(B)\, m_2^{\Omega}(C)$$

Properties

- Conjunctive combination in the absence of conflict, otherwise redistribution of the conflict value
- Commutativity, neutral element : m^{Ω}_{Ω}, NON associativity

Example : Fruit recognition

- Recognize a fruit among banana, golden apple, red apple $\Rightarrow \Omega = \{B, A_1, A_2\}$
- Data : color sample $\Rightarrow \mathcal{Y}_c = \{yellow, red\}$ + shape $\Rightarrow \mathcal{Y}_s = \{round, long\}$
- Decoder functions : $\Gamma_c(yellow) = \{B, A_1\}$, $\Gamma_c(red) = \{A_2\}$;
  $\Gamma_s(round) = \{A_1, A_2\}$, $\Gamma_s(long) = \{B\}$
- However : in 20% cases, the color does not bring information on the fruit itself (illumination problem), in 30% cases, the shape is irrelevant (occlusion)
- If color is yellow : $m_c^\Omega(\{B, A_1\}) = 0.8, m_c^\Omega(\Omega) = 0.2$
  If shape is long : $m_s^\Omega(\{B\}) = 0.7, m_s^\Omega(\Omega) = 0.3$
- Case without conflict → conjunctive, Dempster's, hybrid rules → same result

	$\emptyset$	$\{B\}$	$\{A_1\}$	$\{B, A_1\}$	$\{A_2\}$	$\{B, A_2\}$	$\{A_1, A_2\}$	Ω
m_c^Ω	0	0	0	0.8	0	0	0	0.2
m_s^Ω	0	0	0	0	0	0	0.7	0.3
$m_{c\bigcap s}^\Omega$	0	0	0.56	0.24	0	0	0.14	0.06
$m_s'^\Omega$	0	0.7	0	0	0	0	0	0.3
$m_{c\bigcap s}'^\Omega$	0	0.7	0	0.24	0	0	0	0.06

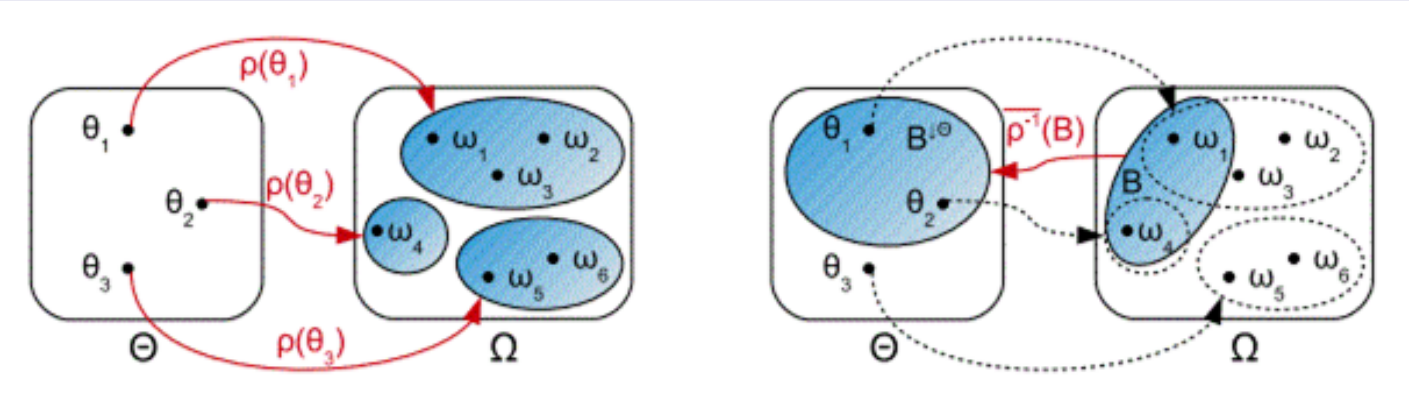
Zadeh example

- $\Omega = \{\omega_1, \omega_2, \omega_3\}$
- Non associative rules : in doubt, last prevails

	$\emptyset$	ω_1	ω_2	$\{\omega_1,\omega_2\}$	ω_3	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	Ω
m_1^Ω	0	0.9	0	0	0.1	0	0	0
m_2^Ω	0	0	0.9	0	0.1	0	0	0
$m_{1 \cap 2}^\Omega$	0.99	0	0	0	0.01	0	0	0
$m_{1 \oplus 2}^\Omega$	0	0	0	0	1	0	0	0
$m_{1 \textcircled{DP} 2}^\Omega$	0	0	0	0.81	0.01	0.09	0.09	0
$m_{1 \textcircled{Y} 2}^\Omega$	0	0	0	0	0.01	0	0	0.99
$m_{1 \cap 2 \cap \{1,2\}}^\Omega$	0.999	0	0	0	0.001	0	0	0
$m_{1 \oplus 2 \oplus \{1,2\}}^\Omega$	0	0	0	0	1	0	0	0
$m_{1 \textcircled{DP} 2 \textcircled{DP} 2}^\Omega$	0	0	0.81	0	0.019	0	0.009	0.162
$m_{1 \textcircled{DP} 2 \textcircled{DP} 1}^\Omega$	0	0.81	0	0	0.019	0.009	0	0.162
$m_{1 \textcircled{Y} 2 \textcircled{Y} 2}^\Omega$	0	0	0.891	0	0.1	0	0	0.009
$m_{1 \textcircled{Y} 2 \textcircled{Y} 1}^\Omega$	0	0.891	0	0	0.1	0	0	0.009

Refinement of a discernment frame

- A frame Ω is a refinement of a frame Θ if and only if there is a one-to-one mapping ρ between Θ and a partition of Ω (Ω elements can be obtained by splitting the elements of Θ, cf. next figure from Denœux)



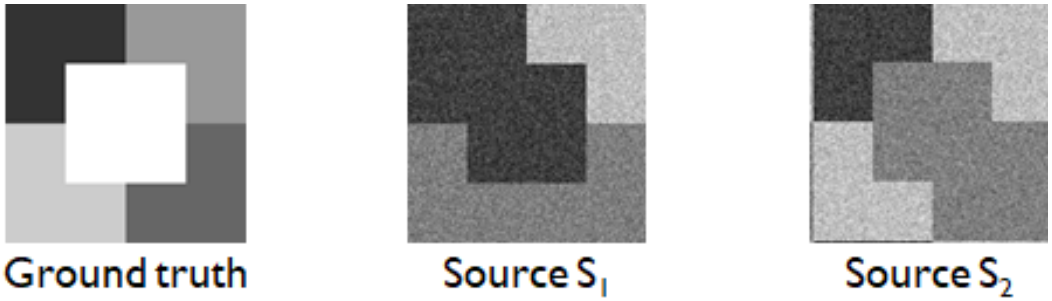
- Ω is a refinement of $\Theta \Leftrightarrow \Theta$ is a coarsening of Ω
- LCP $\Rightarrow \forall A \in 2^{\Theta}, m^{\Theta\uparrow\Omega}(\tilde{\rho}(A)) = m^{\Theta}(A)$, with $\tilde{\rho}(A) = \bigcup_{\theta_i \in A} \rho(\theta_i)$
- if ρ^{-1} not well-defined, LCP $\Rightarrow \forall A \in 2^{\Theta}$, $m^{\Omega\uparrow\Theta}(A) = \sum_{\substack{B \in 2^{\Omega}, \\ \exists \theta_i \in A | \rho(\theta_i) \cap B \neq \emptyset}} m^{\Omega}(B)$

Example : Minimum multisource discernment frame (1/2) (Le Hégarat-Mascle et al., 1997)

- Discernment frame Ω = set of distinguishable classes using multisource data → How learning this set ?
- Basic idea : $\forall i \in \{1,\ldots,n\}, \Omega_i \triangleq$ set of classes distinguishable by each source i is a coarsening of Ω
  - for any $(j,k) \in \{1,\ldots,n\}^2$, $j \neq k$, a source j can reveal subclasses in Ω_k elements and vice versa $\Rightarrow \Omega$ is a refinement of each Ω_i
  - Among the refinements common to all Ω_i, Ω is the chosen as the one with minimum cardinality

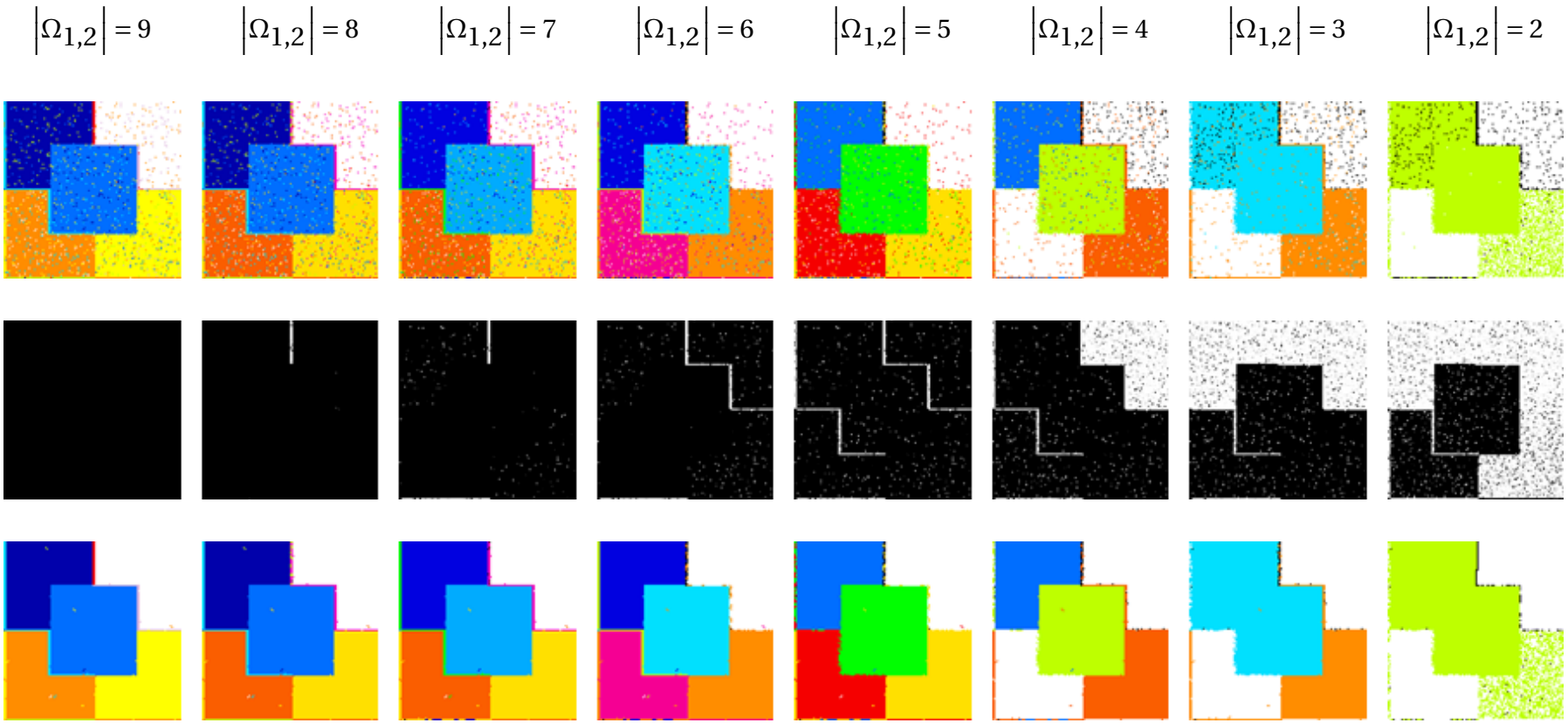
  $\Rightarrow$ Ω is the set of the non-empty intersections of every n-tuples of Ω_i
- Ex. $S_1 \rightarrow \{A_1, A_2\}$, $S_2 \rightarrow \{B_1, B_2\}$, the min. common discernment frame is $\left\{A_i \cap B_j \middle| A_i \cap B_j \neq \emptyset, (i,j) \in \{1,2\}^2\right\}$
- Labeling errors may produce fictitious intersections i.e. fake Ω elements → remove them based on a current multisource labeling → iterative algorithm

Example : Minimum multisource discernment frame (2/2)



2 kinds of imprecision :

- Class ambiguities
- Image registration error (1-2 pixels)



Supervised classification with partial classifiers

- $\Omega = \{\omega_1, \ldots, \omega_k\}$ known
- n supervised classifiers trained to distinguish a subset of classes against another, e.g.
  - *dense* dichotomizer : A versus $(\Omega \setminus A) = \overline{A}$ with $A \subsetneq \Omega \Rightarrow$ FE set $\subseteq \left\{A, \overline{A}, \Omega\right\}$
  - *sparse* dichotomizer : A versus B with $A \subsetneq \Omega$ and $B \subsetneq \overline{A} \Rightarrow$ FE set $\subseteq \{A \cup (\Omega \setminus (A \cup B)), B \cup (\Omega \setminus (A \cup B)), \Omega\} = \left\{\overline{B}, \overline{A}, \Omega\right\}$

Fruit Example

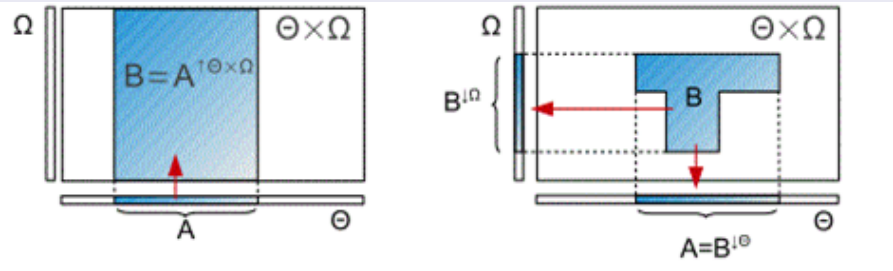
- $\Omega = \{B, A_1, A_2\}$ known
- 2 supervised dichotomizers : (1) $\{B\}$ vs $\{A_1, A_2\}$ and (2) $\{A_1\}$ vs $\{A_2\}$

	$\emptyset$	B	A_1	$\{B, A_1\}$	A_2	$\{B, A_2\}$	$\{A_1, A_2\}$	Ω
m_1^Ω	0	$\alpha_1 \times t$	0	0	0	0	$\alpha_1 \times (1-t)$	$1-\alpha_1$
m_2^Ω	0	0	0	$\alpha_2 \times (1-u)$	0	$\alpha_2 \times u$	0	$1-\alpha_2$
$m_{1 \textcircled{\cap} 2}^\Omega$	0	$\alpha_1 \times t$	$\alpha_1(1-t) \times$ $\alpha_2(1-u)$	0	$\alpha_1 \alpha_2 u$ $\times(1-t)$	0	$\alpha_1(1-t)$ $\times(1-\alpha_2)$	$(1-\alpha_1) \times$ $(1-\alpha_2)$

Cartesian product of discernment frames

- Cartesian product $\Omega \times \Theta$ refinement of both Ω and Θ, 2 discernment frames
- Independence → **Vacuous extension**

$$\begin{cases} m^{\Omega\uparrow\Omega\times\Theta}((A,\Theta)) = m^{\Omega}(A), & \forall A \in 2^{\Omega}, \\ m^{\Theta\uparrow\Omega\times\Theta}((\Omega,A)) = m^{\Theta}(A), & \forall A \in 2^{\Theta} \end{cases}$$



→ Allows for the combination of BBA initially defined on different discernment frames : e.g., $m_1^{\Omega} \textcircled{\cap} m_2^{\Theta} = m_1^{\Omega\uparrow\Omega\times\Theta} \textcircled{\cap} m_2^{\Theta\uparrow\Omega\times\Theta}$

- **Marginalization** → marginal mass function :

$$m^{\Omega\times\Theta\downarrow\Theta}(B) = \sum_{B\in 2^{\Omega\times\Theta}, A\downarrow\Theta=B} m^{\Omega\times\Theta}(A), \forall B \in 2^{\Theta}, \text{ where}$$

$$A \downarrow \Theta = \{\theta \in \Theta | \exists \omega \in \Omega, (\theta,\omega) \in A\}$$

→ Generalizes both set projection and probabilistic marginalization

⚠ $m^{\Omega} = m^{\Omega\uparrow(\Omega\times\Theta)\downarrow\Omega}$, but $m^{\Omega\times\Theta} \sqsubseteq m^{(\Omega\times\Theta)\downarrow\Omega\uparrow(\Omega\times\Theta)}$

Example : Fault diagnosis in an electrical system

- A system can exhibit 3 possible *fault causes* : sensor drift (D), control-loop issue (L), or perturbation/noise (N).
- Expert 1 "dominant issue is sensor drift (0.9 confidence) or a loop issue" ; Expert 2 "dominant issue is noise with (0.9 confidence) or a loop issue".
- If we assume mutual exclusivity with $\Omega = \{D, L, N\}$, combining these opinions yields a very high conflict and may lead to a paradoxical decision.
- **Key point :** in practice, these causes are *not* mutually exclusive (e.g., noise can coexist with poor tuning, and drift can coexist with both).
  Thus the appropriate discernment frame is $\Omega' = \{D, \overline{D}\} \times \{L, \overline{L}\} \times \{N, \overline{N}\}$.
- The induced basic belief assignments on Ω' are :

$$\left\{ \begin{array}{lclclcl} m_1^{\Omega'}\Big((D, \{L, \overline{L}\}, \{N, \overline{N}\})\Big) & = & 0.9, & & m_2^{\Omega'}\Big((\{D, \overline{D}\}, \{L, \overline{L}\}, N)\Big) & = & 0.9, \\ m_1^{\Omega'}\Big((\{D, \overline{D}\}, L, \{N, \overline{N}\})\Big) & = & 0.1, & & m_2^{\Omega'}\Big((\{D, \overline{D}\}, L, \{N, \overline{N}\})\Big) & = & 0.1. \end{array} \right.$$

- Therefore, the conjunctive combination yields :

$$\left\{ \begin{array}{ll} m_{1 \textcircled{\cap} 2}^{\Omega'}\Big((D, \{L, \overline{L}\}, N)\Big) = 0.81, & m_{1 \textcircled{\cap} 2}^{\Omega'}\Big((D, L, \{N, \overline{N}\})\Big) = 0.09, \\ m_{1 \textcircled{\cap} 2}^{\Omega'}\Big((\{D, \overline{D}\}, L, \{N, \overline{N}\})\Big) = 0.01, & m_{1 \textcircled{\cap} 2}^{\Omega'}\Big((\{D, \overline{D}\}, L, N)\Big) = 0.09. \end{array} \right.$$

Example of use of Cartesian product discernment frames for uncertain reasoning

- BBA m^{Ω} representing imprecise and/or partial knowledge on Ω
- BBA $m^{\mathscr{R}}$ representing source reliability, e.g. case of classic discounting : $\mathscr{R} = \{re, \overline{re}\}$ $m^{\Omega}[re] = m^{\Omega}$ and $m^{\Omega}\left[\overline{re}\right](\Omega) = 1$
- Combine all knowledge on $\Omega \times \mathscr{R}$:
  - m^{Ω} conditional extension : $\forall A \in 2^{\Omega}$, $m^{\Omega \Uparrow \Omega \times \mathscr{R}}(\{(A, re), (\Omega, \overline{re})\}) = m^{\Omega}(A)$
  - $m^{\mathscr{R}}$ vacuous extension : $\forall B \in 2^{\mathscr{R}}, m^{\mathscr{R} \uparrow \Omega \times \mathscr{R}}((\Omega, B)) = \begin{cases} \alpha & \text{if } B = re \\ 1-\alpha & \text{if } B = \overline{re} \end{cases}$
- $m^{\Omega \Uparrow \Omega \times \mathscr{R}} \textcircled{\cap} m^{\mathscr{R} \uparrow \Omega \times \mathscr{R}} = m^{\Omega \times \mathscr{R}}$ such that
  $\begin{cases} m^{\Omega \times \mathscr{R}}((A, re)) = \alpha m^{\Omega}(A) & \text{if } A \subseteq \Omega \\ m^{\Omega \times \mathscr{R}}((\Omega, \overline{re})) = 1 - \alpha & \backslash\backslash (1-\alpha) \times \sum_{A \in 2^{\Omega}} m^{\Omega}(A) \end{cases}$
- After reprojection on Ω : $m^{\Omega \times \mathscr{R} \downarrow \Omega}(A) = \begin{cases} \alpha m^{\Omega}(A) & \text{if } A \subsetneq \Omega \\ \alpha m^{\Omega}(\Omega) + (1-\alpha) & \text{if } A = \Omega \end{cases}$, i.e. usual formulation of classic discounting

Canonical decomposition (Smets)

- Generalized Simple BBAs : $\alpha \in [0,+\infty)$, $A^{\alpha}: \begin{cases} \mu^{\Omega}(A) = 1-\alpha \\ \mu^{\Omega}(\Omega) = \alpha \end{cases}$
- n-separable BBA : $\begin{cases} \exists \text{ a weight function } w: 2^{\Omega} \to [0,1] \text{ s.t.} \\ m^{\Omega} = \oplus_{\emptyset \neq A \subsetneq \Omega} A^{w(A)}; A^{w(A)} \text{ SBBAs only} \end{cases}$
- u-separable BBA : $\begin{cases} \exists \text{ a weight function } w: 2^{\Omega} \to [0,+\infty) \text{ s.t.} \\ m^{\Omega} = \bigcirc\!\!\!\!\cap_{A \subsetneq \Omega} A^{w(A)}; \ A^{w(A)} \text{ SBBAs and/or ISBBAs} \end{cases}$

→ Any non-dogmatic BBA is u-separable

→ Uniqueness of the decomposition if we restrict to the decompositions having only one A^{w} per subset A in 2^{Ω}

- $m_c^{\Omega} = \bigcirc\!\!\!\!\cap_{A \in 2^{\Omega}} A^{\min\{w(A),1\}}$, $m_d^{\Omega} = \bigcirc\!\!\!\!\cap_{A \in 2^{\Omega}} A^{\min\left\{\frac{1}{w(A)},1\right\}} \Rightarrow m^{\Omega} = m_c^{\Omega} \not\bigcirc\!\!\!\!\cap\, m_d^{\Omega}$
- $\forall A \in 2^{\Omega} \setminus \{\Omega\}$, $\ln(w(A)) = -\sum_{B \in 2^{\Omega}, B \supseteq A} (-1)^{|B|-|A|} \ln\left(q^{\Omega}(B)\right)$

Fruit example : $\Omega = \{B, A_1, A_2\}$

	$\emptyset$	$\{B\}$	$\{A_1\}$	$\{B, A_1\}$	$\{A_2\}$	$\{B, A_2\}$	$\{A_1, A_2\}$	Ω
$m^{\Omega}_{1 \textcircled{\cap} 2}$	0	0.7	0.24	0	0	0	0	0.06
$q^{\Omega}_{1 \textcircled{\cap} 2}$	1	0.76	0.3	0.06	0.06	0.06	0.06	0.06
$w^{\Omega}_{1 \textcircled{\cap} 2}$	3.8	0.079	0.2	1	1	1	1	
${}_{\{\emptyset\}}{}^{3.8}$	-2.8	0	0	0	0	0	0	3.8
${}_{\{B\}}{}^{0.079}$	0	0.921	0	0	0	0	0	0.079
${}_{\{A_1\}}{}^{0.2}$	0	0	0.8	0	0	0	0	0.2
${}_{\{B\}}{}^{0.079} \textcircled{\cap} {}_{\{A_1\}}{}^{0.2}$	0.737	0.184	0.063	0	0	0	0	0.016
${}_{\{B\}}{}^{0.079} \oplus {}_{\{A_1\}}{}^{0.2}$	0	0.7	0.24	0	0	0	0	0.06
${}_{\{B\}}{}^{0.079} \textcircled{\cap} {}_{\{A_1\}}{}^{0.2} \textcircled{\cap} {}_{\{\emptyset\}}{}^{3.8}$	0	0.7	0.24	0	0	0	0	0.06

- Remark : normalization by $k = \frac{1}{1-m^{\Omega}(\emptyset)}$ ($\frac{1}{1-0.737}$ in the example) $\Leftrightarrow$ combination with the ISBBA $\emptyset^{w(\emptyset)}$

Cautious conjunctive rule (Denœux, 2008)

w-ordering

- $m_1^\Omega \sqsubseteq_w m_2^\Omega \Leftrightarrow \forall A \subsetneq \Omega, w_1(A) \leq w_2(A)$
- $\mathscr{S}(m^\Omega)$: set of BBAs more committed than m according to $\sqsubseteq_w$

Denœux' cautious rule

- m_1^Ω, m_2^Ω being two non-dogmatic BBAs, the less committed BBA among $\mathscr{S}(m_1^\Omega) \cap \mathscr{S}(m_2^\Omega)$ exists and is defined by $m_1^\Omega \circledwedge m_2^\Omega = \bigcirc\!\!\!\!\cap_{A \subsetneq \Omega} A^{w_1(A) \wedge w_2(A)}$, with $\wedge$ the minimum T-norm
- Properties : commutativity, associativity, identity element : vacuous BBA, idempotence

Example with consonant BBAs

w function for consonant BBA

Consonant BBA $\rightarrow w$ derived from ordered q values : $q(\omega_{p(1)}) \geq \cdots \geq q(\omega_{p(k)})$

$$\Rightarrow w(\omega_{p(1)}) = \frac{q(\omega_{p(2)})}{q(\omega_{p(1)})}, \ldots w(\cup_{1 \leq i \leq k} \omega_{p(i)}) = \frac{q(\omega_{p(k)})}{q(\omega_{p(k-1)})}$$

$$\begin{cases} m_1^{\Omega}(A) = 0.2, m_1^{\Omega}(\{A,B\}) = 0.6 \\ m_1^{\Omega}(\Omega) = 0.2 \end{cases}$$

$$\begin{cases} m_2^{\Omega}(B) = 0.6, m_2^{\Omega}(\{A,B\}) = 0.2 \\ m_2^{\Omega}(\Omega) = 0.2 \end{cases}$$

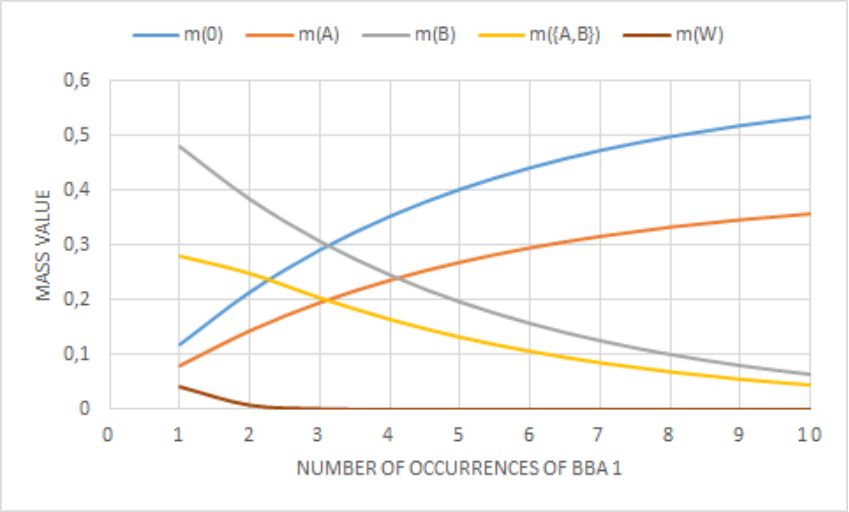
	$\emptyset$	$\{A\}$	$\{B\}$	$\{A,B\}$	$\{C\}$	$\{A,C\}$	$\{B,C\}$	Ω
q_1^{Ω}	1	1	0.8	0.8	0.2	0.2	0.2	0.2
w_1^{Ω}	1	0.8	1	0.25	1	1	1	
q_2^{Ω}	1	0.4	1	0.4	0.2	0.2	0.2	0.2
w_2^{Ω}	1	1	0.4	0.5	1	1	1	

BBA resulting of the combination of n times m_1 and 1 times m_2

Cautious rule

	$\emptyset$	$\{A\}$	$\{B\}$	$\{A,B\}$	Ω
$w_{1 \textcircled{\wedge} 2}^{\Omega}$	1	0.8	0.4	0.25	
$m_{1 \textcircled{\wedge} 2}^{\Omega}$	0.12	0.08	0.48	0.24	0.08

Conjunctive rule



Outline

1. Basics in belief functions
   - Representation of evidences
   - Combination of evidences
   - Decision

Decision in Ω

Bayes decision rule

- Decision among $\mathscr{A} = \{a_1, \ldots, a_n\}$;
- Reasoning with costs $\lambda(a_i, \omega_j)$ associated to action a_i under hypothesis $\omega_j \Rightarrow \hat{a} = \arg\min_{a_i \in \mathscr{A}} \sum_{\omega_j \in \Omega} \lambda(a_i, \omega_j) P(\omega_j)$
- Reasoning with utilities $u(a_i, \omega_j)$ associated to action a_i under hypothesis $\omega_j \Rightarrow \hat{a} = \arg\max_{a_i \in \mathscr{A}} \sum_{\omega_j \in \Omega} u(a_i, \omega_j) P(\omega_j)$

Constraints for BBA transformation into probability measure

- Transformation xP^{Ω} only depends on m^{Ω}, invariant w.r.t. Ω permutations
- $\forall \omega_j \in \Omega, xP^{\Omega}(\omega_j) \in [0,1], \sum_{\omega_j \in \Omega} xP(\omega_j) = 1$
- $\forall A \in 2^{\Omega}$ s.t. $pl^{\Omega}(A) = 0, \begin{cases} \forall \omega_j \in A, & xP^{\Omega}(\omega_j) = 0 \\ \forall \omega_j \in \overline{A}, & xP^{\Omega}(\omega_j) = xP^{\Omega}[\overline{A}](\omega_j) \end{cases}$

BBA transformation into probability measure

Pignistic transformation (Smets, 2005)

$$BetP^{\Omega}(\omega_j) = \frac{1}{1 - m^{\Omega}(\emptyset)} \sum_{\omega_j \in A \in 2^{\Omega}} \frac{m^{\Omega}(A)}{|A|}$$

$\Rightarrow$ uniform distribution of compound hypothesis masses toward their singleton hypotheses

Contour function (Cobb & Shenoy, 2006)

$$PlP^{\Omega}(\omega_j) = \frac{1}{\sum_{\omega_i \in \Omega} pl(\omega_i)} \times pl(\omega_j)$$

$\Rightarrow$ commute with the combination $\textcircled{\cap}$ and the decombination $\cancel{\textcircled{\cap}}$

$PlP^{\Omega}_{1\textcircled{\cap}2}$ computable from pl restriction to Ω : $PlP^{\Omega}_{1\textcircled{\cap}2}(\omega_j) = \frac{pl_1^{\Omega}(\omega_j) \times pl_2^{\Omega}(\omega_j)}{\sum_{\omega_i \in \Omega} pl_1^{\Omega}(\omega_i) \times pl_2^{\Omega}(\omega_i)}$

Decision w/o choosing a single probability (lower/upper)

From a belief function to a *set* of probabilities

Instead of selecting one xP (e.g., $BetP$, PlP), consider the **credal set** $\mathscr{P}(bel) = \{P \text{ on } \Omega : bel(A) \leq P(A) \leq pl(A) \ \forall A \subseteq \Omega\}$.

Expected utility becomes an interval : For any action $a \in \mathscr{A}$,

$$\underline{\mathbb{E}}(a) = \min_{P \in \mathscr{P}(bel)} \mathbb{E}_P[u(a,\omega)] = \sum_{A \subseteq \Omega} m(A) \min_{\omega \in A} u(a,\omega),$$

$$\overline{\mathbb{E}}(a) = \max_{P \in \mathscr{P}(bel)} \mathbb{E}_P[u(a,\omega)] = \sum_{A \subseteq \Omega} m(A) \max_{\omega \in A} u(a,\omega).$$

Decision rules (robust / optimistic / compromise)

- **Maximin (robust) :** $\hat{a} = \arg\max_{a \in \mathscr{A}} \underline{\mathbb{E}}(a)$ (guaranteed performance)
- **Maximax (optimistic) :** $\hat{a} = \arg\max_{a \in \mathscr{A}} \overline{\mathbb{E}}(a)$ (best-case)
- **Hurwicz :** $\hat{a} = \arg\max_{a \in \mathscr{A}} \mathbb{E}_\gamma(a)$ with $\mathbb{E}_\gamma(a) = \gamma \underline{\mathbb{E}}(a) + (1-\gamma) \overline{\mathbb{E}}(a)$, $\gamma \in [0,1]$

Special case : 0–1 utility ⇒ Belief vs Plausibility

0–1 utility (classification)

Action a_i means "decide ω_i" and $u(a_i, \omega) = \mathbb{1}[\omega = \omega_i]$.

On a focal set A

$$\min_{\omega \in A} u(a_i, \omega) = \begin{cases} 1 & \text{iff } A \subseteq \{\omega_i\} \text{ (i.e. } A = \{\omega_i\}) \\ 0 & \text{otherwise} \end{cases} \qquad \max_{\omega \in A} u(a_i, \omega) = \begin{cases} 1 & \text{iff } \omega_i \in A \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\underline{\mathbb{E}}(a_i) = \sum_A m(A) \min_{\omega \in A} u(a_i, \omega) = \sum_{A \subseteq \{\omega_i\}} m(A) = bel(\{\omega_i\}),$$

$$\overline{\mathbb{E}}(a_i) = \sum_A m(A) \max_{\omega \in A} u(a_i, \omega) = \sum_{A : \omega_i \in A} m(A) = pl(\{\omega_i\}).$$

Interpretation

- Maximize $\underline{\mathbb{E}}$ ⇒ maximize bel (pessimistic / robust).
- Maximize $\overline{\mathbb{E}}$ ⇒ maximize pl (optimistic / contour-based).
- Hurwicz interpolates between both strategies.

Example : Ellsberg paradox

One-urn Ellsberg paradox

1 urn containing 90 balls : 30 balls are red, while the remaining 60 balls are either black or yellow in unknown proportions 2 gamble scenarii

- gamble A : win if draw a *red* ball versus gamble B : win if draw a *black* ball → participants the majority of participants choose gamble A
- gamble C : win if draw a *red or yellow* ball versus gamble D : win if draw a *black or yellow* ball → participants the majority of participants choose gamble D

Probabilistic formulation

- Sample space : $\Omega = \{R, Y, B\}$, $P(R) = 1/3$, $P(Y) + P(B) = 2/3$
- gamble A choice $\Rightarrow P(R) > P(B) \Rightarrow P(B) < 1/3$
- gamble D choice $\Rightarrow P(R \cup Y) < P(B \cup Y) \Rightarrow P(R) + P(Y) < P(B) + P(Y)$ $\Rightarrow 1/3 < P(B)$

Ellsberg paradox explained

Belief formulation

- Sample space : $\Omega = \{R, Y, B\} \Rightarrow 2^{\Omega} = \{\emptyset, R, Y, \{R, Y\}, B, \{R, B\}, \{Y, B\}, \Omega\}$
- $m^{\Omega}(R) = 1/3,\ m^{\Omega}(\{Y, B\}) = 2/3$

	$\emptyset$	$\{R\}$	$\{Y\}$	$\{R,Y\}$	$\{B\}$	$\{R,B\}$	$\{Y,B\}$	Ω
bel^{Ω}	0	1/3	0	1/3	0	1/3	2/3	1
pl^{Ω}	0	1/3	2/3	1	2/3	1	2/3	1

- decision based on bel^{Ω} function :
  - gamble A choice $\Rightarrow bel^{\Omega}(R) > bel^{\Omega}(B)$
  - gamble D choice $\Rightarrow bel^{\Omega}(\{R, Y\}) < bel^{\Omega}(\{B, Y\})$

$\Rightarrow$ completely consistent with participant majority choice

Wine-water paradox revisited (A. Jousselme)

Probabilistic formulation

- Assume an area with a mixture of commercial vessels and warships. We assume the ratio of warships over commercial vessels $r_{W/V} \in \left[\frac{1}{3},3\right]$. What is the probability that $r_{W/V} \leq 2$?
- Principle of Indifference $\rightarrow r_{W/V}$ uniformally distributed in $\left[\frac{1}{3},3\right] \Rightarrow P(r_{W/V} \leq 2) = \frac{2-\frac{1}{3}}{3-\frac{1}{3}} = \frac{5}{8}$
- Paradox : Having the same reasoning with random variable $r_{V/W} \in \left[\frac{1}{3},3\right]$, one finds $P(r_{V/W} \geq \frac{1}{2}) = \frac{3-\frac{1}{2}}{3-\frac{1}{3}} = \frac{15}{16} \neq P(r_{W/V} \leq 2)$

Belief formulation

- Discernment space : $\Omega = \mathbb{R}^+ \Rightarrow 2^\Omega = 2^{\mathbb{R}^+}$, i.e. $\mathbb{R}^+$ intervals, for both $r_{W/V}$ and $r_{V/W}$
- In both cases, categorical BBAs : equal to $m^\Omega_{W/V}(\left[\frac{1}{3},3\right]) = 1$ and $m^\Omega_{V/W}(\left[\frac{1}{3},3\right]) = 1$ resp.
- Deduced bel and pl functions
  - considering $r_{W/V}$: $bel^\Omega_{W/V}(\left[\frac{1}{3},2\right]) = 0$ and $pl^\Omega_{W/V}(\left[\frac{1}{3},2\right]) = 1$
  - considering $r_{V/W}$: $bel^\Omega_{V/W}(\left[\frac{1}{2},3\right]) = 0$ and $pl^\Omega_{V/W}(\left[\frac{1}{2},3\right]) = 1$

$\Rightarrow$ completely consistent whatever the considere random variable

Outline

1. Basics in belief functions
2. Belief functions for image processing
   - Spatial information as an information piece
   - Modeling ambiguities between classes
   - Consistency at region level
3. Belief functions for localization
4. Belief functions for Deep Learning

Field of BBAs

Examples of problems

- Detection of specific image features (contours, line segments, objects, etc.) → binary map of detections
- Image segmentation → map of region labels
- Image semantic segmentation → map of classes

Pixel level

- In each pixel, decision in Ω ($\{0,1\}$, set of regions, set of classes)

⇒ one BBA m^Ω per pixel p → one BBA field per image

- Case of several sources : as many BBAs $m_1^\Omega,\ldots,m_n^\Omega$ as information sources → m^Ω derived from $m_1^\Omega,\ldots,m_n^\Omega$ combination

Outline

2. Belief functions for image processing
   - Spatial information as an information piece
   - Modeling ambiguities between classes
   - Consistency at region level

Contextual semantic segmentation

- Basic idea : neighbor label configuration is an information source
- Modeling through a BBA m_N^Ω to combine with the other BBAs

Example

- $\Omega = \{R, G, Y\}$, 8-connectivity neighborhood :

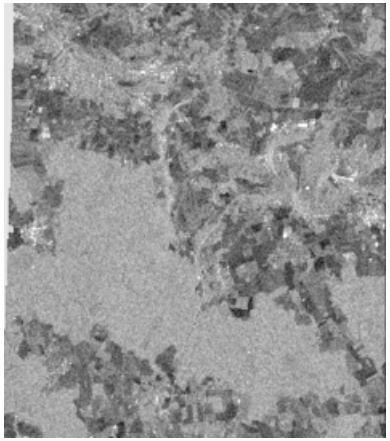
- Forbidden configurations : e.g., no lonely pixel → dogmatic BBA (→ conditioning on the not-forbidden solutions) ; example : $m_N^\Omega(\{R, Y\}) = 1$
- Likely configurations → consonant BBA s.t. $BetP(\omega_i)$ given by local configuration ; example :

$BetP(\{R\}) = 5/9, BetP(\{Y\}) = 3/9, BetP(\{G\}) = 1/9 \Rightarrow$ FEs $= \{R\}, \{R, Y\}, \Omega$

$\Rightarrow m_N^\Omega(\{R\}) + \frac{1}{2} m_N^\Omega(\{R, Y\}) + \frac{1}{3} m_N^\Omega(\Omega) = 5/9, \frac{1}{2} m_N^\Omega(\{R, Y\}) + \frac{1}{3} m_N^\Omega(\Omega) = 3/9, \frac{m_N^\Omega(\Omega)}{3} = 1/9$

$\Rightarrow m_N^\Omega(\{R\}) = \frac{2}{9}, m_N^\Omega(\{R, Y\}) = \frac{4}{9}, m_N^\Omega(\Omega) = \frac{1}{3}$

Application to SAR-Optical data fusion

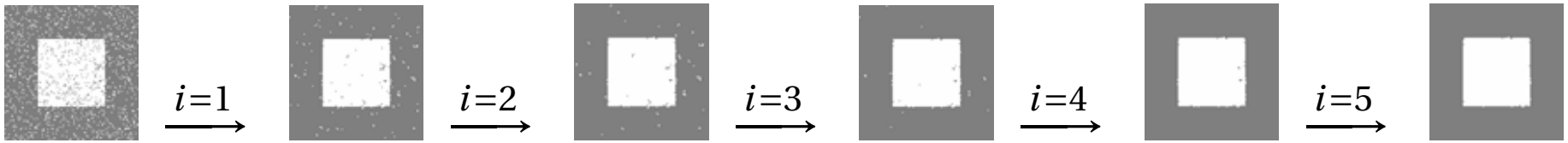


Objective : forest area detection $\rightarrow \Omega = \left\{F, \overline{F}\right\}$;
Sources : radar and optical images
$\Rightarrow m_R^\Omega$ and m_O^Ω such that

$$\begin{cases} \text{clouds} \rightarrow m_O^\Omega = \text{vacuous BBA} \\ \text{cloud borders} \rightarrow m_O^\Omega(\Omega) \searrow \text{ with dist. to cloud} \\ \text{speckle} \rightarrow \text{ global discounting of } m_R^\Omega \end{cases}$$

Neighborhood estimation : Iterative process

1. initializations : $m_D^\Omega = m_R^\Omega \textcircled{\cap} m_O^\Omega$; $i = 0$; $\left(m_N^\Omega\right)^0 =$ vacuous BBA
2. Perform multisource classification $\left(m^\Omega\right)^i = m_D^\Omega \textcircled{\cap} \left(m_N^\Omega\right)^i$
3. Update $\left(m_N^\Omega\right)^{i+1}$ with respect to neighbor labels derived from $\left(m^\Omega\right)^i$
4. If stopping criterion is not verified : $i \leftarrow i+1$; goto step 2.

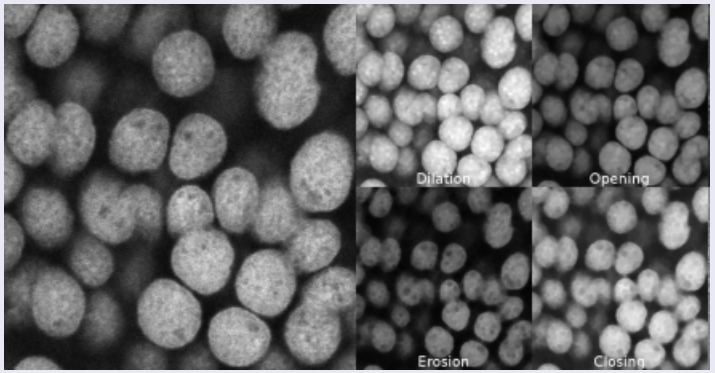


Mathematical Morphology operators *in-1-slide*

- Math. Morphology (MM) operators defined on complete lattice $\mathscr{L}$ (partial ordering $\leq$) ; for gray level images, $\mathscr{L}$ is the lattice of functions f on the image domain $\mathscr{S}$, $\leq$ is the partial order on numerical functions
- Translation invariance $\rightarrow$ structuring element ν ; case of flat structuring element : defined by its support $\mathscr{N}$

$$\begin{cases} \text{Erosion :} & \forall s \in \mathscr{S}, \quad \epsilon_\nu(f)(s) = \min_{\forall t \in \mathscr{N}_s} f(t), \\ \text{Dilation :} & \delta_\nu(f)(s) = \max_{\forall t \in \mathscr{N}_s} f(t), \\ \text{Opening : } \gamma_\nu(f) = \delta_{\check{\nu}} \circ \epsilon_\nu(f) ; & \text{Closing : } \phi_\nu(f) = \epsilon_{\check{\nu}} \circ \delta_\nu(f) \end{cases}$$

- γ_ν and ϕ_ν are MM filters
- Duality between ϵ_ν and δ_ν, and between γ_ν and ϕ_ν



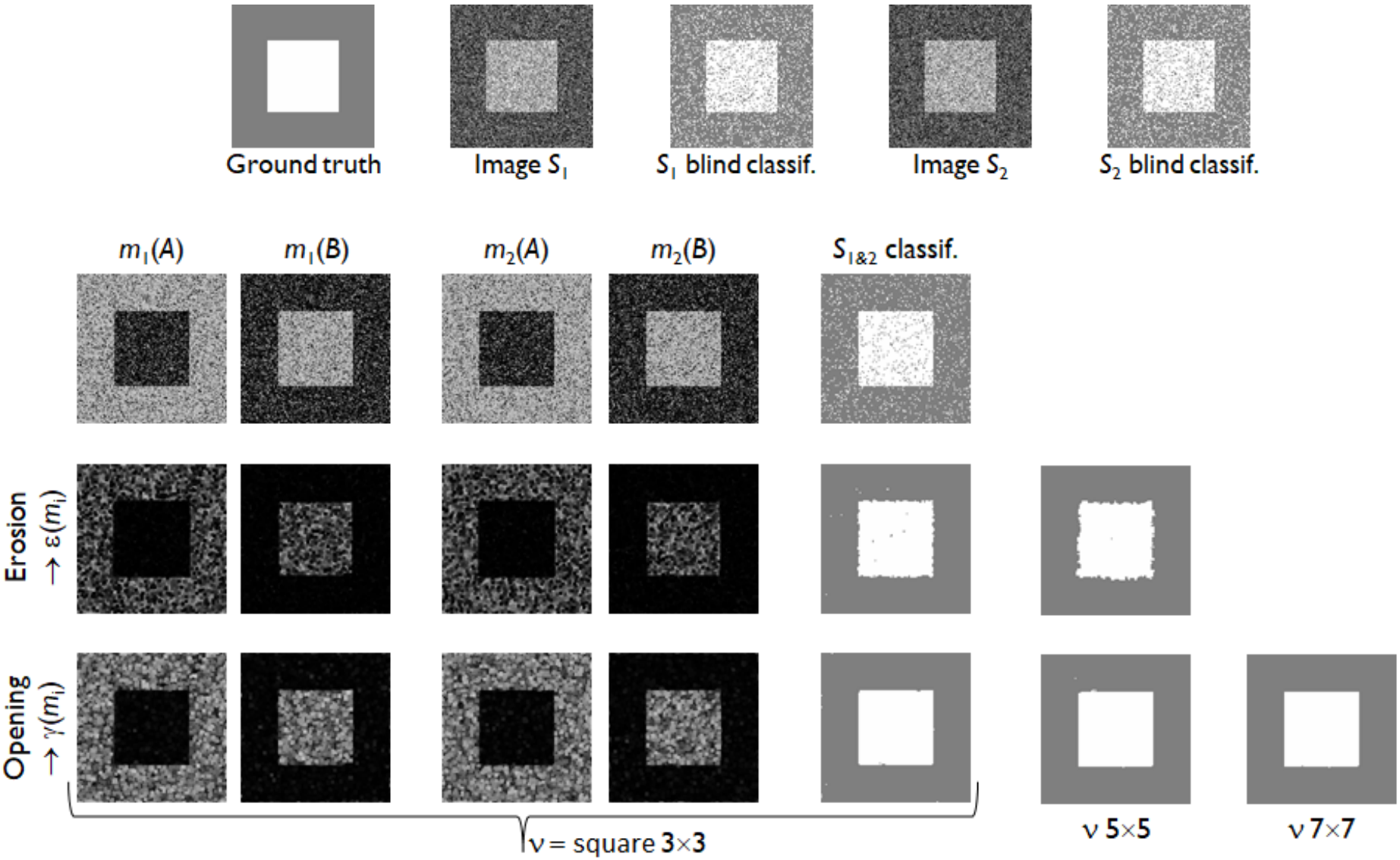
(Bloch, 2008)

- Basic idea : use the Mathematical Morphology (MM) operators to determine the beliefs on disjunctions
- Use a structuring element to evaluate spatial imprecision
- Duality between pl^{Ω} and bel^{Ω} corresponds to duality between MM erosion and dilation or between MM opening and closing
- Case of initial BBA m_0^{Ω} normal and having 'at most' 2 disjoint FEs A, B : $A \cap B = \emptyset \Rightarrow pl^{\Omega}(A) = bel^{\Omega}(A) + m^{\Omega}(\{A,B\}) = 1 - bel^{\Omega}(B)$

  $\Rightarrow m^{\Omega}(\{A,B\}) = 1 - bel^{\Omega}(A) - bel^{\Omega}(B) = m_0^{\Omega}(A) - m^{\Omega}(A) + m_0^{\Omega}(B) - m^{\Omega}(B)$
- $$\begin{cases} \forall C \in \{A,B\}, m^{\Omega}(C) \begin{cases} = \epsilon_{\nu}\left(m_0^{\Omega}(C)\right) \\ \text{or} \\ = \gamma_{\nu}\left(m_0^{\Omega}(C)\right) \end{cases} \\ m^{\Omega}(\{A,B\}) = m_0^{\Omega}(\{A,B\}) + \sum_{C \in \{A,B\}} \left(m_0^{\Omega}(C) - m^{\Omega}(C)\right) \end{cases}$$

Example on simulated data

- 2 sources $(S_1,\ S_2)$, 2 classes : $\Omega = \{A, B\} \Rightarrow B = \overline{A}$



Pedestrian detection (Vandoni et al., 2019) (1/2)

- Detection problem → 2 classes : $\Omega = \left\{H, \overline{H}\right\}$
- calibration of SVM classifier scores → sigmoid function

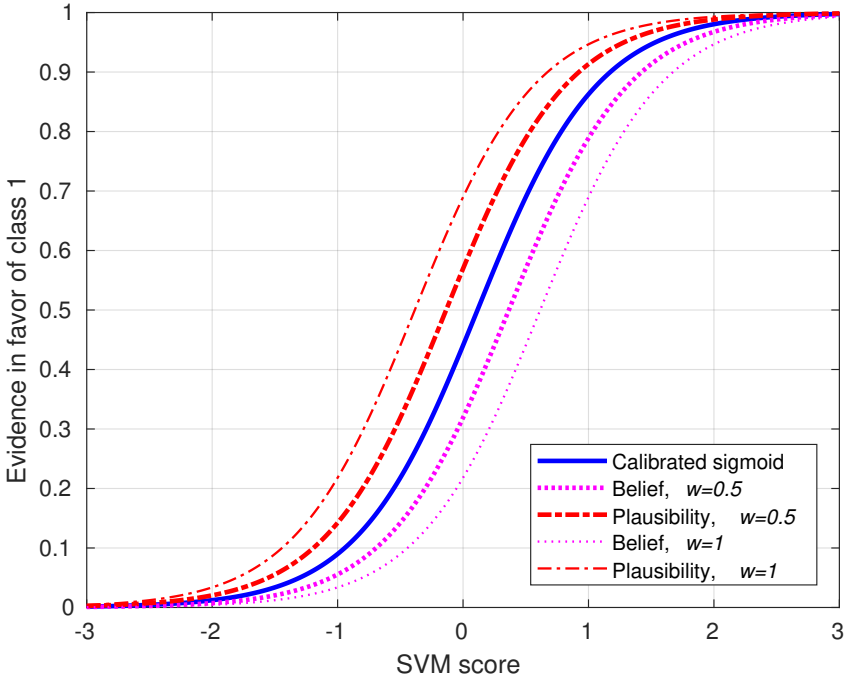


Figure – Bayesian m_0^Ω provided by sigmoid function (SVM score calibration) and derived bel^Ω and pl^Ω bounds at different structuring element w sizes ; class 1 $= H$.

$\tilde{m}(H)$ neighborhood of x_1

	0.8	
0.01	$\boldsymbol{x_1}$	0.2
	0.7	

$\tilde{m}(H)$ neighborhood of x_2

	0.95	
0.95	$\boldsymbol{x_2}$	0.94
	0.98	

$\tilde{m}(\overline{H})$ neighborhood of x_1

	0.19	
0.8	$\boldsymbol{x_1}$	0.4
	0.2	

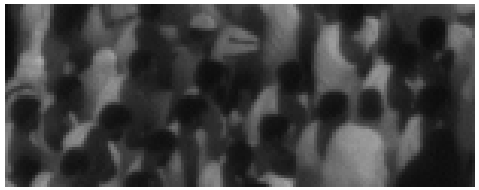
$\tilde{m}(\overline{H})$ neighborhood of x_2

	0.04	
0.03	$\boldsymbol{x_2}$	0.03
	0.01	

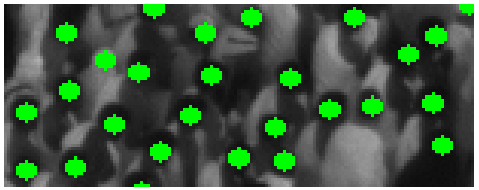
Table – Example of BBA allocation for samples x_1 and x_2, with flat 4-connectivity structuring element (simplified case).

Sample	score-based BBA	Final BBA
x_1	$\tilde{m}(H) = 0.12$ $\tilde{m}(\bar{H}) = 0.49$ $\tilde{m}(\Omega) = 0.39$	$m(H) = 0.01$ $m(\bar{H}) = 0.19$ $m(\Omega) = 0.8$
x_2	$\tilde{m}(H) = 0.95$ $\tilde{m}(\bar{H}) = 0.01$ $\tilde{m}(\Omega) = 0.04$	$m(H) = 0.94$ $m(\bar{H}) = 0.01$ $m(\Omega) = 0.05$

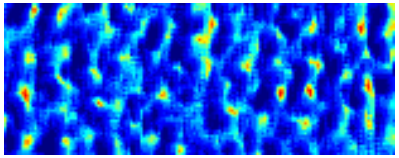
Real application : Pedestrian detection (2/2)



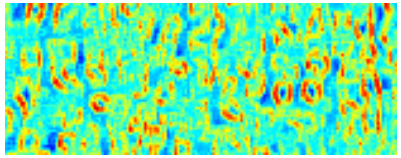
(a) Image patch



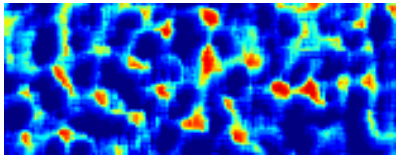
(b) Ground-truth



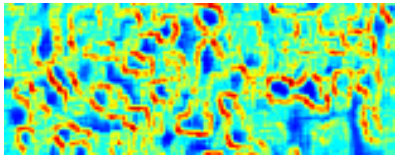
(c) $m_{HOG}(H)$



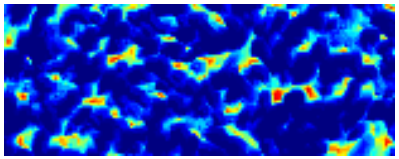
(d) $m_{HOG}(\Omega)$



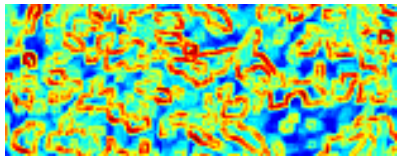
(e) $m_{LBP}(H)$



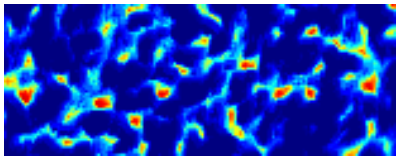
(f) $m_{LBP}(\Omega)$



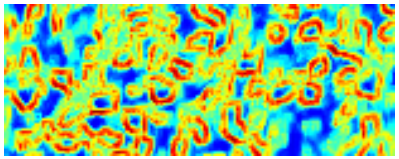
(g) $m_{GABOR}(H)$



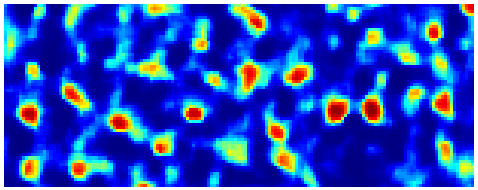
(h) $m_{GABOR}(\Omega)$



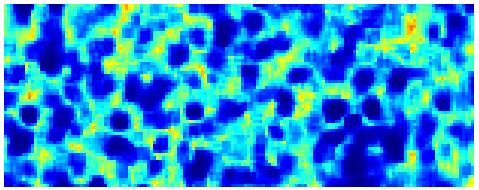
(i) $m_{DAISY}(H)$



(j) $m_{DAISY}(\Omega)$



(k) $m(H)$



(l) $m(\Omega)$

Outline

2. Belief functions for image processing
   - Spatial information as an information piece
   - Modeling ambiguities between classes
   - Consistency at region level

Classifier Ensemble

Basic idea

- Combining the decisions of different classifiers to get better performance generalization, i.e. overcome the Bias-Variance tradeoff
- 2 main strategies for classifier selection :
  - few strong (rather generic and accurate) classifiers
  - many weak (or very specialized) classifiers, e.g. AdaBoost
- Many ways to introduce diversity between classifiers, e.g.
  - Bagging (Bootstrap Aggregating) : diversity $\leftarrow$ different training datasets
  - Random subspaces : diversity $\leftarrow$ different feature vectors
  - Error-correcting output codes (ECOC) : diversity $\leftarrow\neq$ classifiers

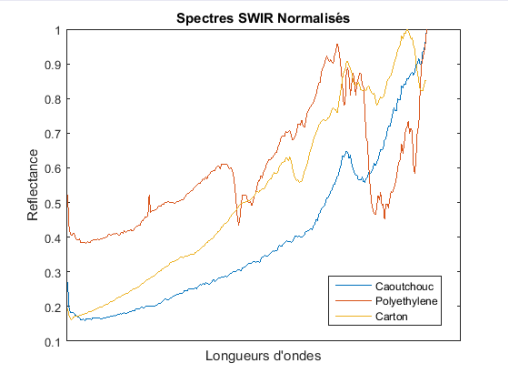
Divide & Conquer strategy

Split a *difficult* problem into several *simpler sub*problems $\rightarrow$ for multiclass classification,

- How choosing the subsets of classes considered by the different classifiers ?
- How combining the classifier outputs dealing with $\neq$ classes or class subsets ?

Hyperspectral data classification

Hyperspectral data



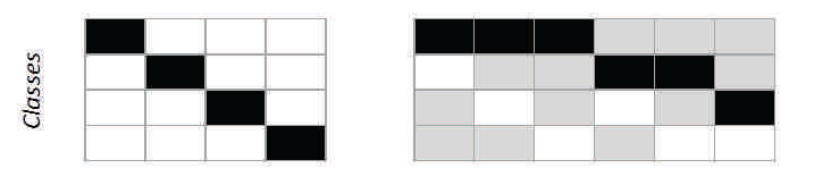
High dimensionality data ➡ Preprocessing steps :

- Basic spectrum + ≠ derivation orders (e.g., from Savitzky-Golay filter) ;
- Dimensionality reduction ('only' 99% information → # channels $\approx 20 \ll 400$)

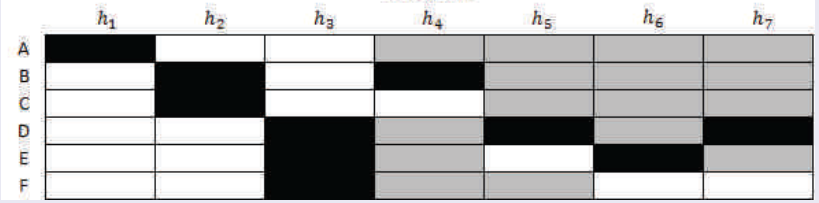
Support Vector Machines (SVM) with multiclasses → ECOC classifier

Use a set of binary classifiers (dichotomizers) to derive a global decision

Examples :



(a) One-Versus-All (left) and One-Versus-One (right) strategies ; 4-class case.

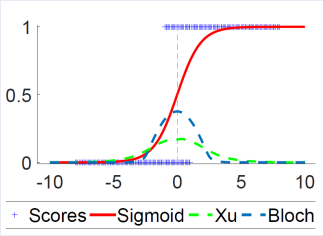


(b) Confusion Matrix Superclass ECOC strategy ; 6 classes, 3 superclasses.

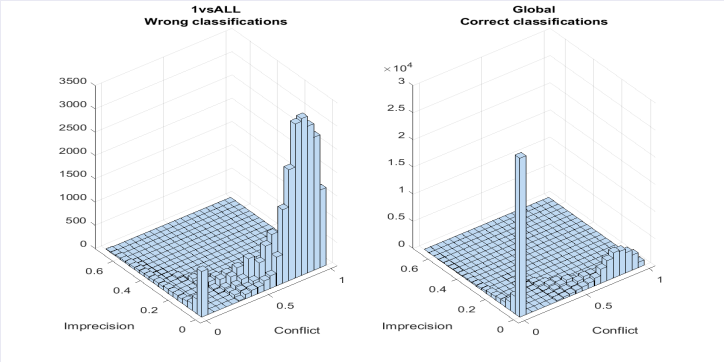
Benefit of BF for dynamic ECOC definition

Decoding part

- BBA defined from SVM (Support Vector Machine) scores →
- Refinement ρ_i of dichotomizer classes $\{-1,+1\}$ to Ω classes → FE elements are $\rho_i(\{-1\})$, $\rho_i(\{+1\}$ and Ω (dense dichotomizer) or $\overline{\rho_i(\{+1\})}$, $\overline{\rho_i(\{-1\})}$ and Ω (sparse dichotomizer)



Coding part



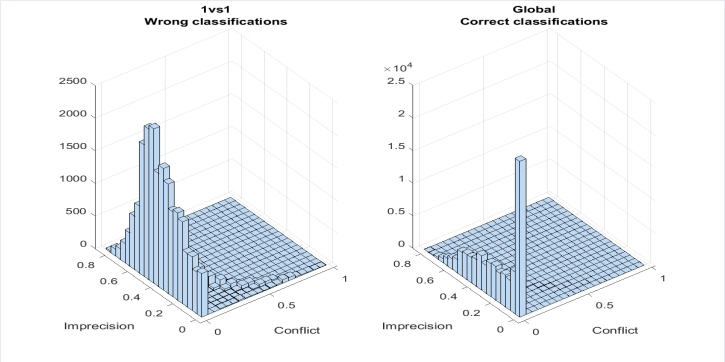
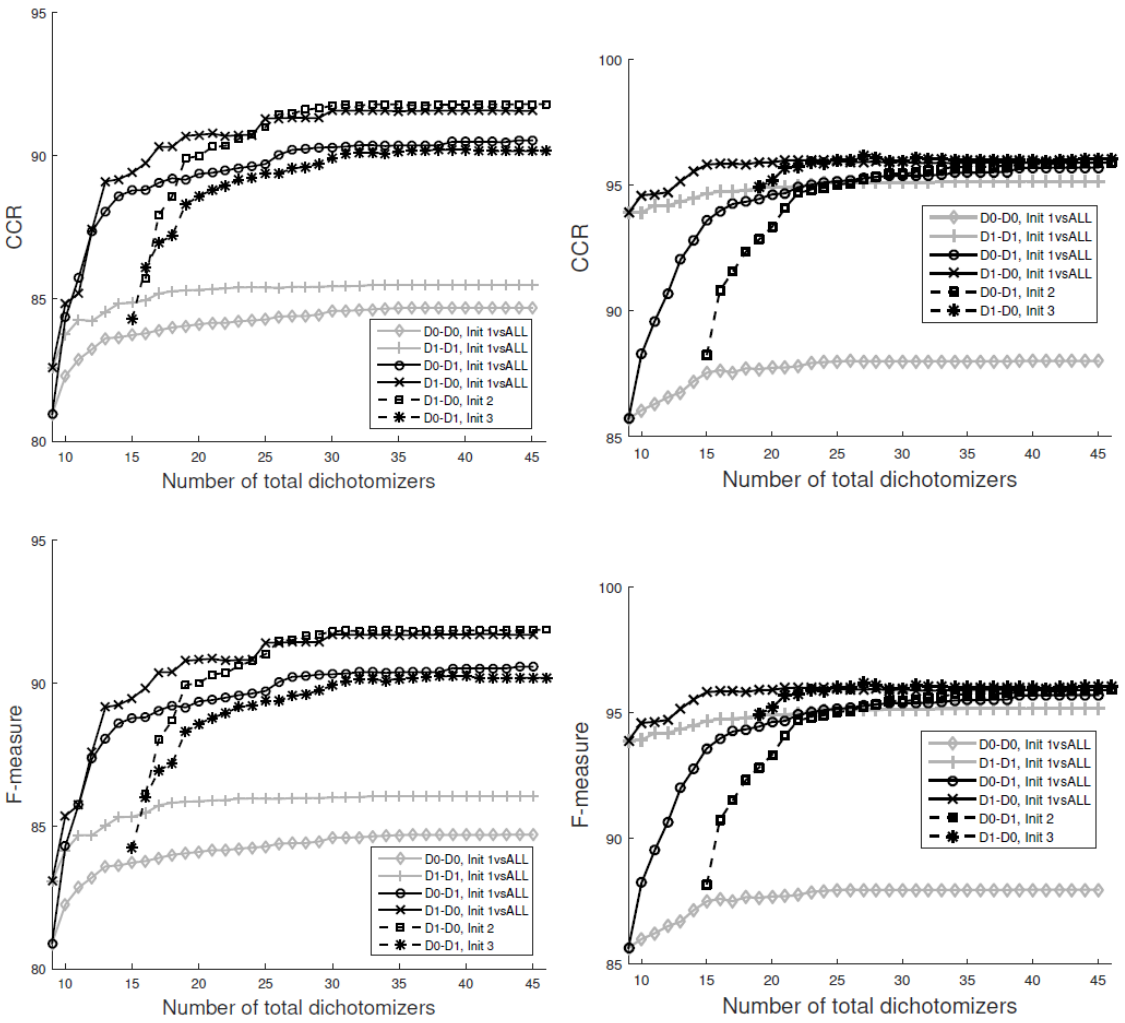


Figure – $(m(\emptyset), m(\Omega))$ 2D-histograms of wrong label samples and correct label samples with OVA (One-Versus-All) strategy (left) and OVO (One-Versus-One) strategy (right).

⇒ data-driven approach : selection of tricky pixels and estimation of the ambiguous classes for each of these pixels (based on partial ECOC result), + performance-driven approach : dichomizer selection among those separating the ambiguous classes

Benefit of combining decisions derived from different derivatives (Lachaize et al., 2018a)



(a) HSI_1 data (b) HSI_2 data

- Performance (Correct Classification Rate, CCR, and F1-score) vs number of dichotomizers for different combinations of 0 and 1st derivative of the data spectrum
- 1st and 2nd columns → 2 different hyperspectral sensors, HSI_1 and HSI_2

Outline

2. Belief functions for image processing
   - Spatial information as an information piece
   - Modeling ambiguities between classes
   - Consistency at region level

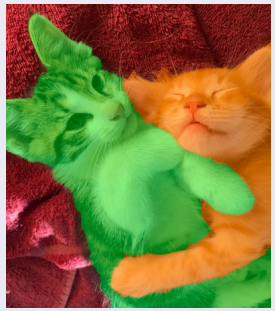
Semantic segmentation problem

Basic image processing problem

- Objective : define a partition of the image lattice so that each region has homogeneous features
- Spatial information not only to confirm or infirm decisions at pixel level but also to retrieve object/structure subparts
- Toward scene interpretation

Challenges

- Find a representation such that the data appear homogeneous inside the objects and heterogeneous with respect to the background or the other objects (i.e.,instance segmentation)



- Case of multisensor data : manage data complementary and redundancy in terms of region homogeneity

Semantic segmentation based on heterogeneous sensors

Data constraints

Complementary data both in terms of features and spatial sampling :

- Some data are spatially dense (*true* images) but do not verify the intra-object homogeneity constraint
- Some data are spatially subsampled or imprecise (low resolution) → *pseudo-images*

Ill-posed semantic segmentation problem

$\mathscr{S}$: region set (segmentation), Ω : label set, l_s label of one element $s \in \mathscr{S}$

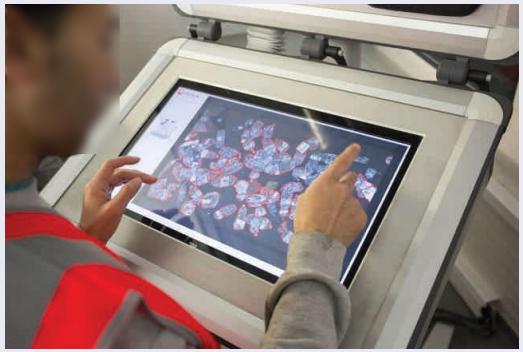
$\Rightarrow$ Estimate $\mathscr{S}$ AND the associated labels $\{l_s\}_{s \in \mathscr{S}}$

Pseudo-images (i.e., missing borders) $\Rightarrow$ ill-posed problem

Real application case : Characterization of deformed objects using hyperspectral imaging (1/2)

Application context : Waste separation and recognition

- First sorting steps : based on mechanical properties
- Final sorting step : either manually by a human operator or using a robotic arm and human interface



Computer vision dedicated to the conveyor belt

- Deformed objects ⇒ cannot rely on shape to recognize the category of waste
- Dirty objects ⇒ can neither rely on color
- Many occlusions ⇒ need to separate objects

Real application case : Characterization of deformed objects using hyperspectral imaging (2/2)

- Hyperspectral (HS) camera → very good at distinguishing ≠ materials but expensive and with high processing costs
- Depth sensor → mainly to separate stacked objects
- Color (RGB) camera → usually in complement to other sensors
- Dedicated sensors : stiffness, brightness, sound sensor etc. → useful to recognize specific materials

Complementary of the data

- Assumption : objects consistent in terms of *material* and *height*
- *True* images (RGB, 3D) provide spatial information *but* not necessary relevant w.r.t. objects : RGB may induce irrelevant sub-regions (multi-color objects), 3D data do not distinguish objects on a same plane
- *Pseudo*-images (HS data) provide material information *but* only in sub-sampled points

Proposed approach

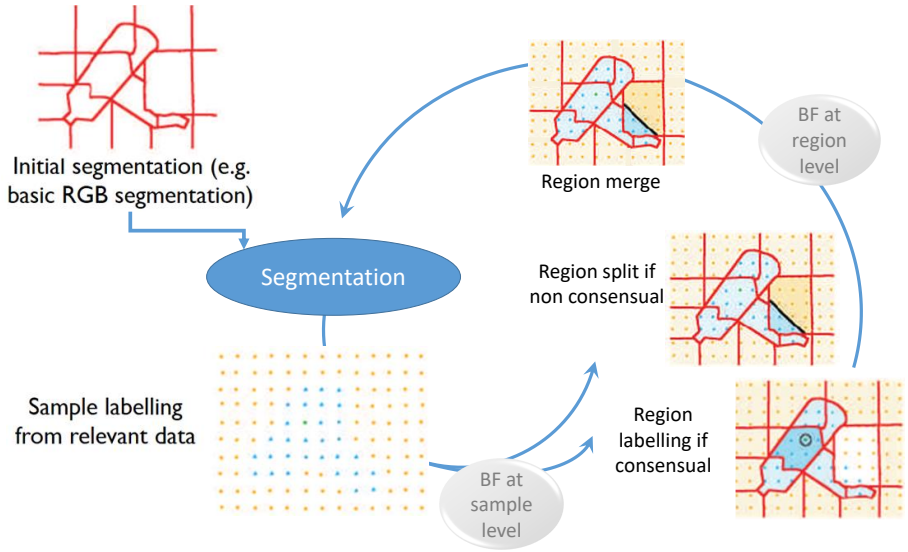


Figure – Semantic segmentation scheme using heterogeneous data.

Segmentation task

- Start from a segmentation based *true* images
- Correct it based on object features derived from HS and 3D data → split inconsistent regions and merge neighbour regions that are consistent
- Intra-region and inter-regions *consistency* measured in terms of conflict ($m(\emptyset)$) between region samples

➡ Classification and segmentation tasks interacting at superpixel level → allowing us to cope with pseudo-images

➡ BF framework → modelling missing data and class ambiguities, controlling consistency based on *conflict* degree

Evidential Split&Merge (Lachaize et al., 2018b)

Segmentation evolution by splitting regions and/or merging regions

Split step : criterion = conflict between intra-region BBAs

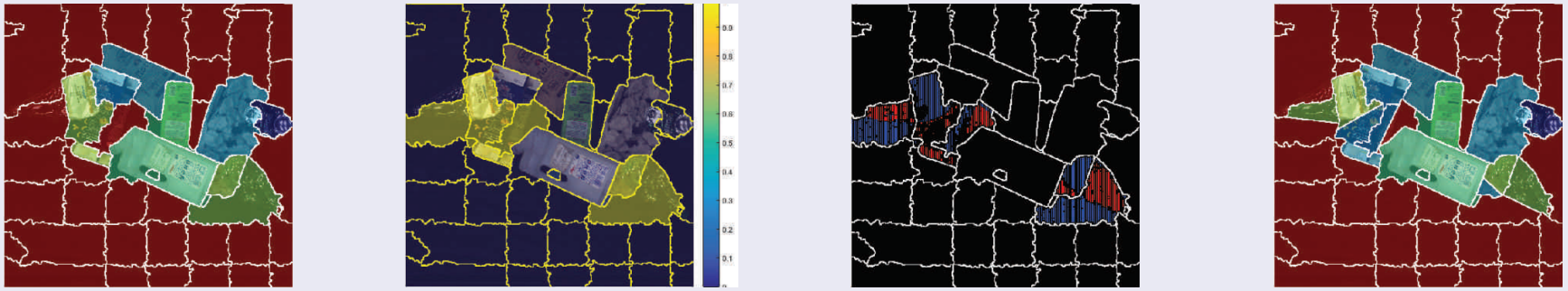


Figure – Example : (a) initial labeled-segmentation, (b) conflict per region, (c) linear border intra-region, (d) split step final labeled-segmentation.

Merge step : criterion = conflict between inter-region BBAs

Figure – Example : (a) initial (post split) labeled-segmentation, (b-c) iterations 1 and 3, (d) final labeled-segmentation.

Results for the characterization of deformed objects using hyperspectral imaging

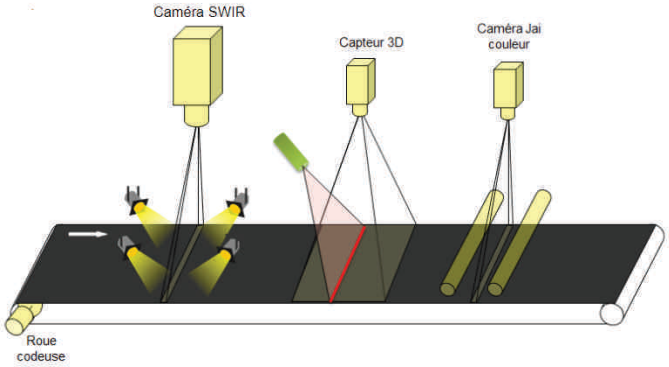


Figure – Scheme and photo of the VEOLIA conveyor for waste sorting.

Figure – Examples of results with (a) stack of white plastic bottles, (b) different white plastics, (c) multicolored and transparent objects, (d) transparent plastic bottles, (e) flat and multicolored sheets, (f) mixed items.

Conclusion on BF tools to integrate spatial information

Motivating application	**Key information**	**Coding**	**Tool**
Semantic segmentation	Neighborhood labels	Spatial regularization (BBA)	Iterative neighborhood BBA estimation
Contextual pixel labeling	Spatial ignorance	Discounting of BF	Erosion (or opening) of mass function
Object semantic segmentation	Spatial conflict vs consistency	Current segmentation revision	Evidential Split&Merge

Outline

1. Basics in belief functions
2. Belief functions for image processing
3. Belief functions for localization
   - Managing outlier issue
   - Handling 2D discernment frame
4. Belief functions for Deep Learning

Localization problem

Sensor dependent problem

- Embedded sensors : either providing absolute localization (e.g., GNSS receiver) or relative displacement (e.g., IMU, odometers)
- External sensors having different perception fields (from PIR sensors to multi-camera system) → detection in the perception field, tracking in the chosen referential (e.g., 2D ground plane)

Uncertainty issue

- Uncertainty sources : sensors, measurements, model uncertainty, etc.

➡ probabilistic modeling (aleatory uncertainty) : uncertainty ellipsoids, Multiple Hypothesis Tracking, etc. ; set approaches (imprecision) : interval analysis...

Outlier issue

- Different definitions : physics, statistics (OOD), data fusion

➡ robust estimation, data relaxation, outliers removal vs correction

Outline

3. Belief functions for localization
   - Managing outlier issue
   - Handling 2D discernment frame

Interval Analysis q-relaxation (Drevelle and Bonnifait, 2011)

- No estimation of the set of inliers
- Under hypothesis that only r sources have to be considered out of the N sources, for any N-tuple $\boldsymbol{A} = \{A_1, A_2, \ldots, A_N\} \in \left(2^{\Omega}\right)^N$ the relaxation of $q = N - r$ results in $\Gamma_r(\boldsymbol{A}) \in 2^{\Omega}$, the disjunction of all the intersections between elements of r-tuples

$$\Gamma_r(\boldsymbol{A}) = \bigcup_{\mathscr{A} \subseteq \{A_1, \cdots, A_N\}, |\mathscr{A}| = r} \left(\bigcap_{A_i \in \mathscr{A}} A_i \right)$$

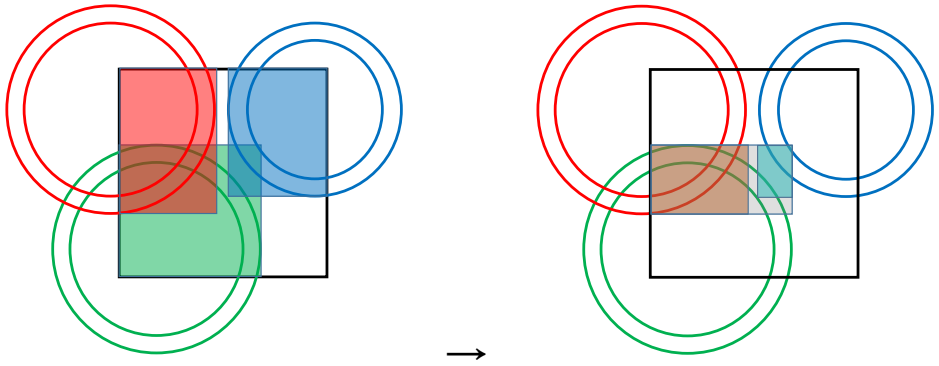


Figure – Visual example for q-relaxation in Interval Analysis.

Evidential q-relaxation (Pichon et al., 2016)

Mathematical definition

$$\forall B \in 2^{\Omega}, m\left[H_r^N\right](B) = \sum_{\substack{\boldsymbol{A} \subseteq \Omega^N; \\ \Gamma_r(\boldsymbol{A}) = B}} \left[\prod_{i=1}^{N} m_i^{\Omega}(A_i)\right] \text{ with } \Gamma_r(\boldsymbol{A}) \text{ defined as in IA q-relaxation}$$

A	$\emptyset$	$\{\omega_1\}$	$\{\omega_1,\omega_2\}$	$\{\omega_3\}$	$\{\omega_1,\omega_3\}$	Ω
m_1^{Ω}	0	0.5	0	0	0	0.5
m_2^{Ω}	0	0	0	0	0.2	0.8
m_3^{Ω}	0	0	0	0.6	0	0.4
$m^{\Omega}[q=2]$	0	0	0	0	0.06	0.94
$m^{\Omega}[q=1]$	0	0.06	0.04	0	0.24	0.66
$m^{\Omega}[q=0]$	0.36	0.2	0.04	0.24	0	0.16

For instance, mass product $m_1^{\Omega}(\{\omega_1\}) \times m_2^{\Omega}(\{\omega_1,\omega_3\}) \times m_3^{\Omega}(\{\omega_3\})$ transferred to $(\{\omega_1\} \cap \{\omega_1,\omega_3\}) \cup (\{\omega_1\} \cap \{\omega_3\}) \cup (\{\omega_1,\omega_3\} \cap \{\omega_3\})$

- Generalizes the classic BFT combination rules, e.g. conjunctive combination and the disjunction one → in the example, $m^{\Omega}[q=2] = m_1^{\Omega} \textcircled{\cup} m_2^{\Omega} \textcircled{\cup} m_3^{\Omega}$ and $m^{\Omega}[q=0] = m_1^{\Omega} \textcircled{\cap} m_2^{\Omega} \textcircled{\cap} m_3^{\Omega}$
- Categorical BBA ⇒ evidential q-relaxation consistent with the IA q-relaxation

RANSAC (Fischler and Bolles, 1981) and variants

RANdom Sample Consensus

⇒ estimation of the set of inliers

- Sampling the solution space
- For every solution, evaluate its consensus degree as the number of inlier samples, i.e. samples presenting an error lower than a given threshold
- For the most consensual solution, $\mathscr{I}$ = inlier set

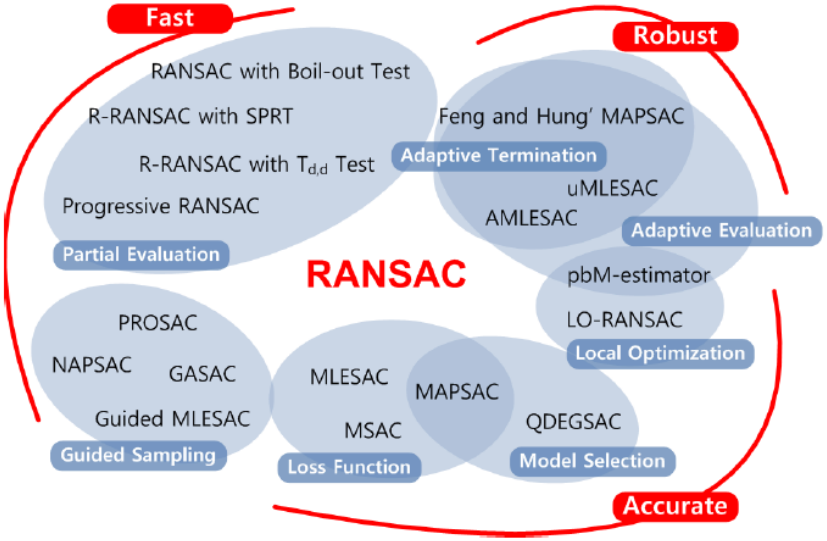


Figure – Choi et al., "Performance Evaluation of RANSAC Family", BMVC 2009

Evidential RANSAC (Zair and Le Hégarat-Mascle, 2017)

Take into account the imprecision of the observations in order to mitigate their weight
⇒ represent the observations through BBAs to combine :

- Use a measure of consistency (maximum of the contour function) or inconsistency (conflict) and an a priori threshold to estimate the solution
- Graph traversal in a graph to sub-sample the solution space

Table – BBAs of the toy example presented on the figure ; $\Omega = \{\omega_1, \omega_2, \omega_3\}$.

A	m_1	m_2	m_3	m_4	m_5
$\{\omega_1\}$	0	0	0	0.7	0.3
$\{\omega_2\}$	0.05	0	0.8	0	0
$\{\omega_1, \omega_2\}$	0.75	0.8	0	0.3	0
$\{\omega_2, \omega_3\}$	0	0	0.2	0	0
Ω	0.2	0.2	0	0	0.65

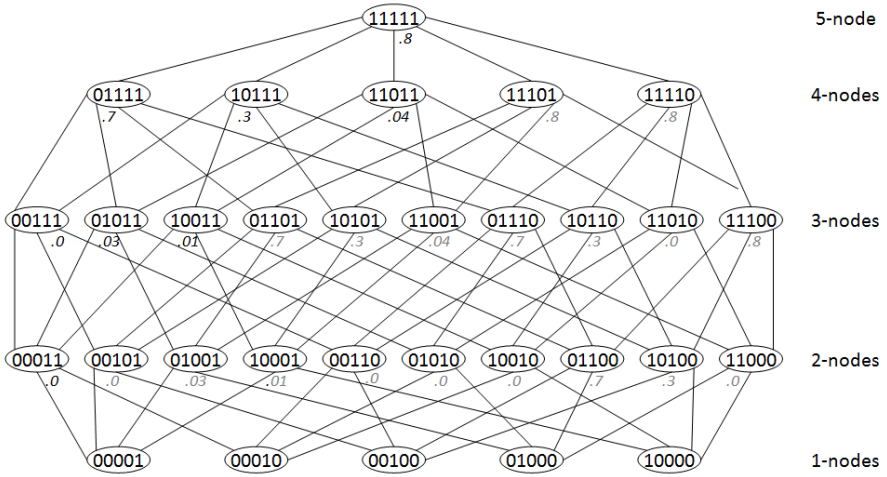


Figure – Graph associated to the toy example ; en italic : conflict values ; only those in black color are actually computed when starting from vertex 00011 and setting conflict threshold $\kappa_{th} = 0.1$

Example of real application : GNSS localization in constraint environment

GNSS signals

- In-sight signal vs multi-path signal
- Obstacles degrade the measurement of time travel between satellite and receiver
- Non-line-of-sight and multi-path signals are considered as outliers

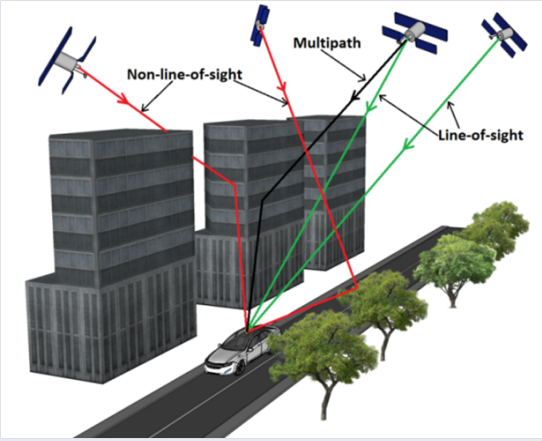


Figure – Constrained environment.

Localization

- Data : 3D coordinates of satellites i ; receiver-satellite pseudo-ranges $\rho_i, \forall i \in \mathscr{S}$ with $\mathscr{S}$ the set of satellites
- 4-parameter observation model : $\tilde{\rho}_i = f(\underbrace{x, y, z, \delta t}_{\mathbf{X_r}})$,

  Residues $r_i(t) = \left[\tilde{\rho}_i(\mathbf{X_r}(\mathbf{t})) - \rho_i(t)\right]$

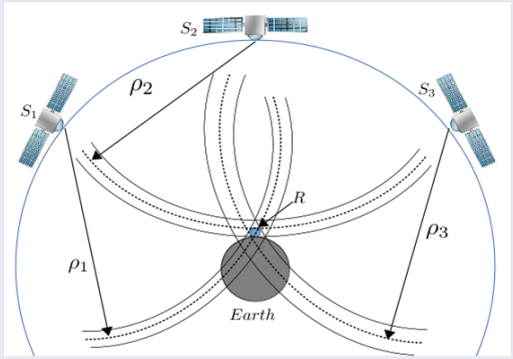
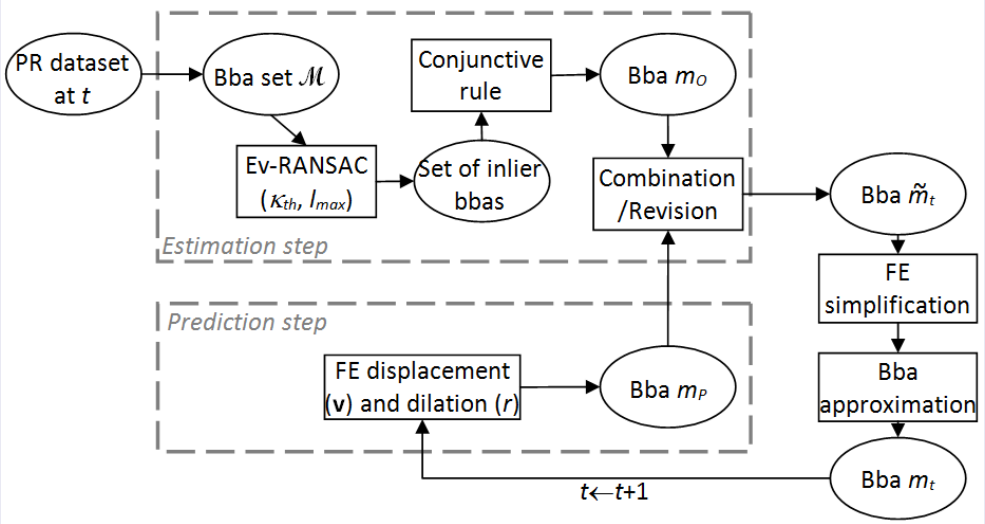


Figure – Trilateration :
$\hat{\mathbf{X}}_\mathbf{r}(\mathbf{t}) = \arg\min_{\mathscr{I}, \mathbf{X_r}(\mathbf{t})} \sum_{i \in \mathscr{I}} g\left(r_i(t)\right)$ with $\mathscr{I} \subseteq \mathscr{S}$.

Evidential GNSS localization

Algorithm



Involved belief function tools : usual operators +

- BBA allocation : from PR measure to 2D ring sector focal elements (FEs)
- Evidential RANSAC
- BBA summarization : controls number of FEs using BBA approximation (André et al., 2015)
- BBA prediction : 2D-translation of FEs w.r.t. displacement model and 2D-dilation

Result example

- Error measure : $\mathcal{E}(\lambda) = \sum_{A \in 2^{\Omega}} (|A| + \lambda [\![\hat{\mathbf{x}}_{\mathbf{r}} \notin A]\!]) \times \frac{m(A)}{1-m(\emptyset)}$
- Comparison between some variants of EV-RANSAC w.r.t. the used combination or revision rule and with interval analysis

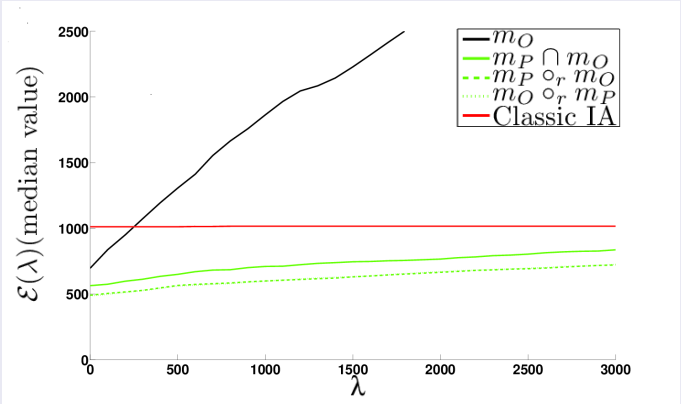


Figure – Performance according to $\mathcal{E}(\lambda)$ versus λ.

Outline

3. Belief functions for localization
   - Managing outlier issue
   - Handling 2D discernment frame

Handling 2D focal elements

Examples of applications using belief functions

- GNSS localization (BFT → modeling pseudo-range imprecision, managing conflict) ⇒ FEs = ring sectors ;
- Object reconstruction from fragmentary detections (BFT → object-detection association, spatio-temporal conditioning, e.g. Rekik et al., 2016)
- Tracking in surveillance applications (BFT → modeling PIR measurements, sound sensors, video detection, e.g. André et al., 2015) ⇒ FEs = 2D geometric shapes

Solutions avoiding exhaustive coding of every subset of the 2D grid

- Conditioning on sub-parts of the discernment frame, multi-resolution ⇒ complexity, loss in spatial resolution
- Geometric approximation by sets of 2D rectangles ⇒ non uniqueness
- Discretizing using cloud points (N-D case) ⇒ numerical approximation of Demspter rule (using Monte Carlo simulation), Sui et al., 2018

Focal element representation

- FEs = polygons (or sets of polygons)
- 1 FE = set of closed paths ; a closed path = a table of ordered vertices (sense → hole versus inside region)

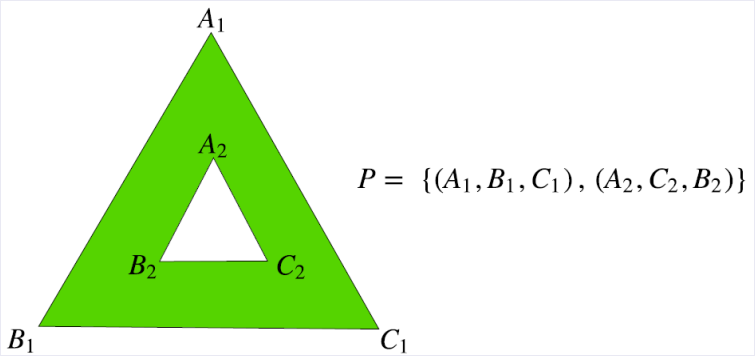


Figure – Coding of a polygon with hole.

- ∃ efficient methods to perform intersection, union, difference, XOR... (extension of Vatti's algorithm for clipping, used in image synthesis)
- Use of 1 hash function for early identification of FEs → particularly useful for the sums in the combination rule(s) and for uncertainty propagation

Graph representing the interactions between focal elements

- Directed Acyclic Graph (DAG) $\mathscr{G}$ such that :
  - vertices represent FEs ordered by decreasing cardinality →
  - arcs represent the non-null intersections (or inclusion relationships) between FEs

⇒ Any intersection between FEs is represented by a path within $\mathscr{G}$

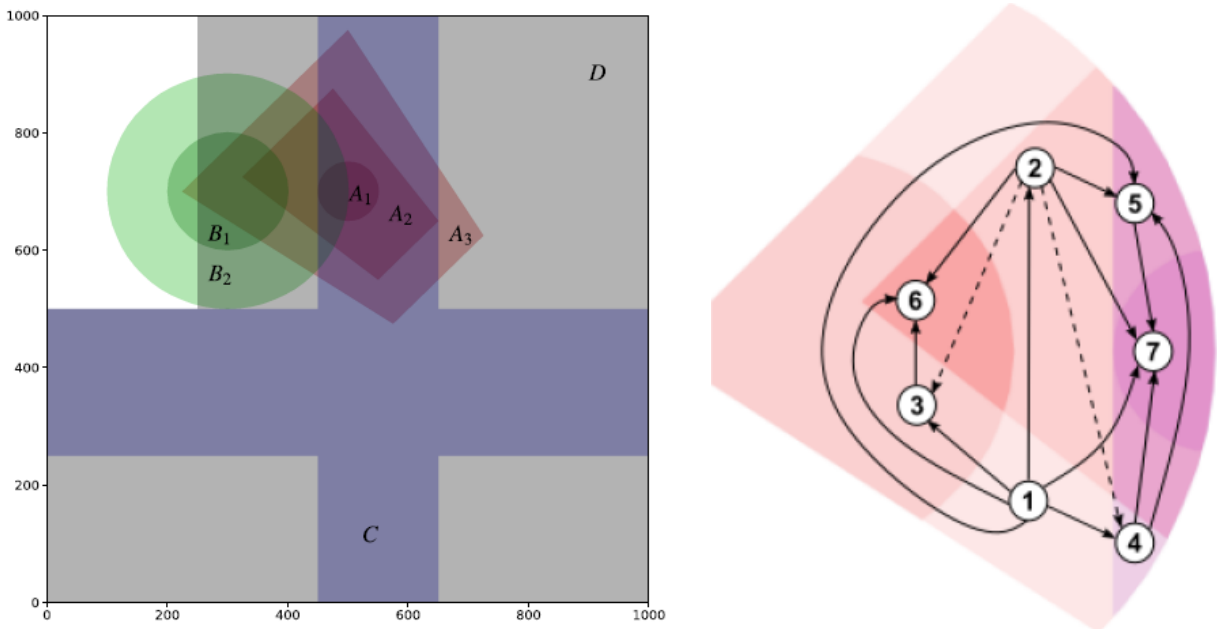


Figure – (a) FEs associated to : camera detection m_1 (in red), track at $t-1$ m_2 (in green), road prior m_3 (in blue), building-exclusion prior m_4 (in gray) ; (b) intersection-inclusion graph after conjunctive combination (→ = inclusion, ⇢ = intersection).

Graph simplification with a view to graph traversal

Definitions

- For any pair of vertices (v_1, v_2), $v_1 \subset v_2$ iff the FE represented by v_1 is included in the FE represented by v_2
- For any path P, 'P intersection' = intersection between the FEs represented by the vertices of P
- *dead path* = path with empty intersection, *superpath*, *subpath*, *redundant path* $\rightarrow \exists$ 1 *superpath* of same intersection ;

$\Rightarrow$ focusing on *non-redundant, non-dead* paths

1. *Root suppression* : Consider as root only the vertices not included in some others
2. *Early stopping* : Stop the path once a vertex is included in a vertex of higher cardinality than the root
3. *Graph simplification* : If a vertex is included in several other vertices, keep only the incoming inclusion connection from the vertex with the lowest cardinality

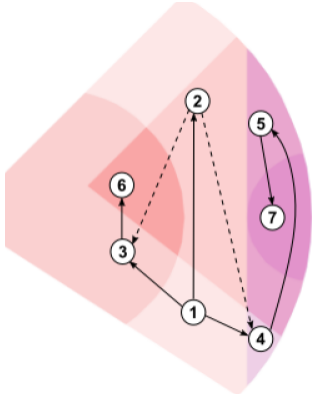
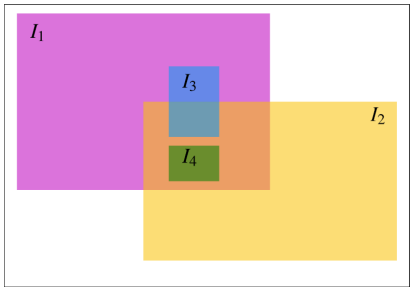


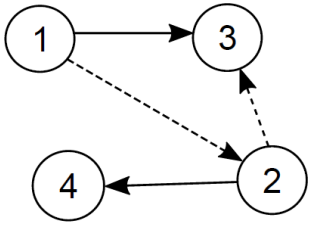
Figure – Simplified graph. → : inclusion, ⇢ : intersection.

Computation of an equivalent BBA 1D

Set of disjoint sets $\mathscr{D}$ = partition of the union of the FEs such that no element straddles 2 FEs and no element is "unnecessarily" subdivided (maximal cover)

- ⇒ Any 2D FE can be written as the disjunction of $\mathscr{D}$ elements → equivalent BBA on $\mathscr{D}$ (1D)
- ⇒ Possibility of using any usual BFT tools, e.g. Fast Mobius Transform to perform the canonical decomposition of the BBA on $2^{\mathscr{D}}$





	$\tilde{D} = \bigcap_i I_i$				D
	I_1	I_2	I_3	I_4	
A	✓	✓	✓		$D_A = \tilde{D}_A$
B	✓	✓		✓	$D_B = \tilde{D}_B$
C	✓	✓			$D_C = \tilde{D}_C \setminus \{D_A, D_B\}$
D	✓		✓		$D_D = \tilde{D}_D \setminus \{D_A\}$
E	✓				$D_E = \tilde{D}_E \setminus \{D_A, D_B, D_C, D_D\}$
F		✓			$D_F = \tilde{D}_F \setminus \{D_A, D_B, D_C\}$

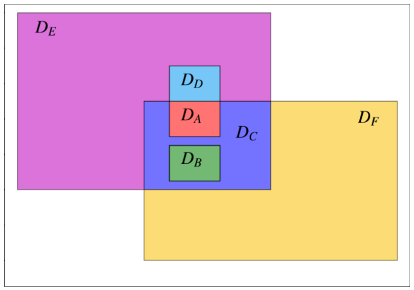


Figure – Illustrative example of coarse discernment frame extraction for canonical decomposition. (a) initial 2D BBA (FE labeled $I_j, j \in \{1,\ldots,4\}$) ; (b) simplified intersection-inclusion DAG : I_4 is included in I_1 and I_2 (but only edge toward v_1 is removed) ; (c) details of the derivation of the equivalent 1D representation ; (d) obtained *set of disjoint sets* $\mathscr{D} = \{D_A, D_B, D_C, D_D, D_E, D_F\}$

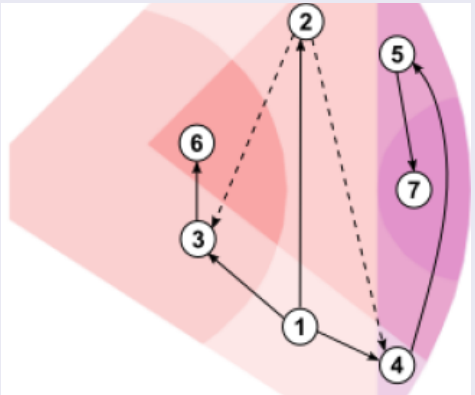
Decision-making

Issue : decision usually taken on singleton hypothesis : how to guaranty scalability ?

Maximal intersections

- Definition : Path on the graph s.t. any supplementary vertex $\rightarrow \emptyset$
- Property : any *maximal intersection* can be represented as a non-redundant path that is not a subpath (of any other non-redundant path)
- Decision $(BetP,\ pl)$ taken on the set $\mathscr{I}_{\mathscr{M}}$ of *maximal intersections*

Example



Path	Maximal intersection
$\langle v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_6 \rangle$	yes
$\langle v_1 \rightarrow v_2 \rightarrow v_4 \rightarrow v_5 \rightarrow v_7 \rangle$	yes
$\langle v_1 \rightarrow v_3 \rightarrow v_6 \rangle$	no
$\langle v_1 \rightarrow v_4 \rightarrow v_5 \rightarrow v_7 \rangle$	no

Figure – (a) Simplified graph ($\rightarrow$: inclusion, $\dashrightarrow$: intersection). (b) Derivation of maximal intersections ($\rightarrow \mathscr{I}_{\mathscr{M}}$). Maximum $BetP$ obtained for $I_1 \cap I_2 \cap I_4 \cap I_5 \cap I_7$.

Multi-camera pedestrian detection (Pellicanò et al., 2017)

Figure – Multi-camera system with single camera head detector

Algorithm using variable-height homographies

- hyperpixel = a pixel and its homographic projections in the other views
- hysteresis thresholding, head seg. & top-head detection
- fusion of top-head detections → heat detection map

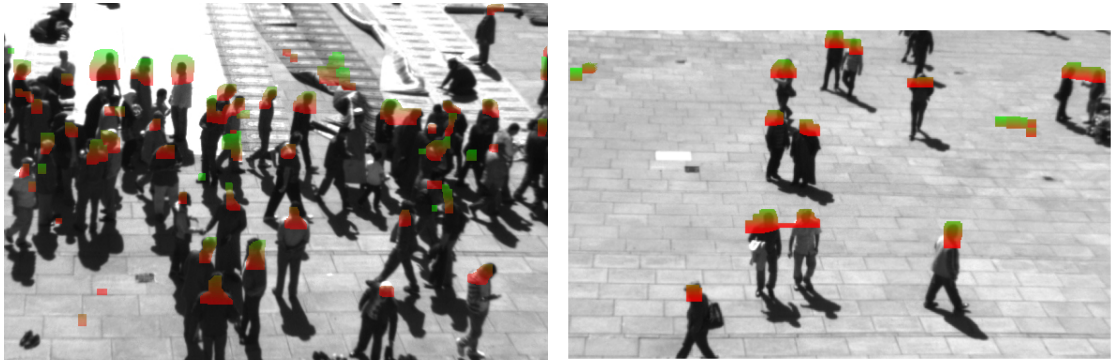


Figure – Pedestrian detector → detection map along with pedestrian height estimations $\in [1.4m, 2m]$

Application to pedestrian tracking (Pellicanò et al., 2018)

Pedestrian tracking using geometric constraints : detections in image ($t=0$), height between $1.4m$ (rouge) et $2m$ (vert).

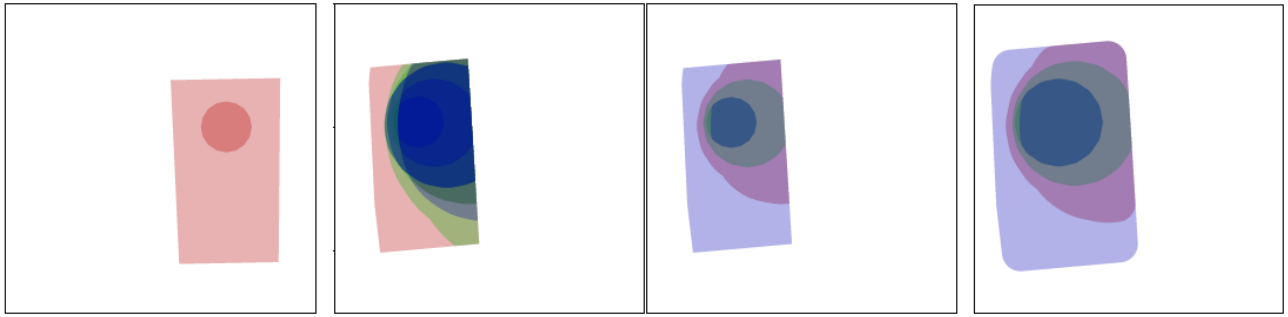
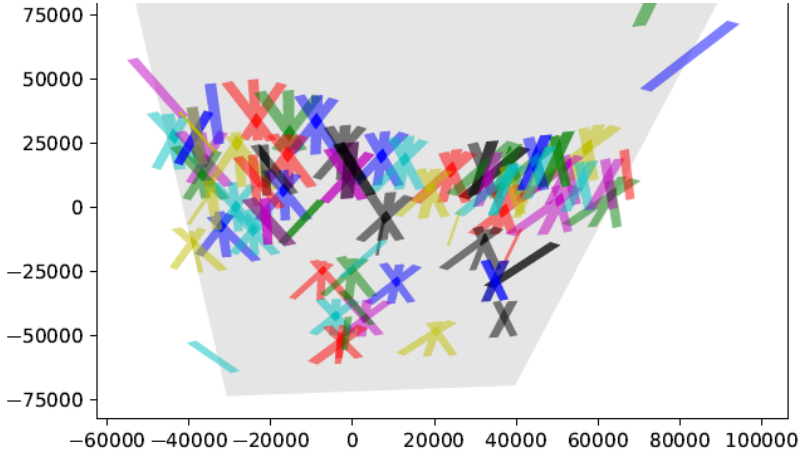
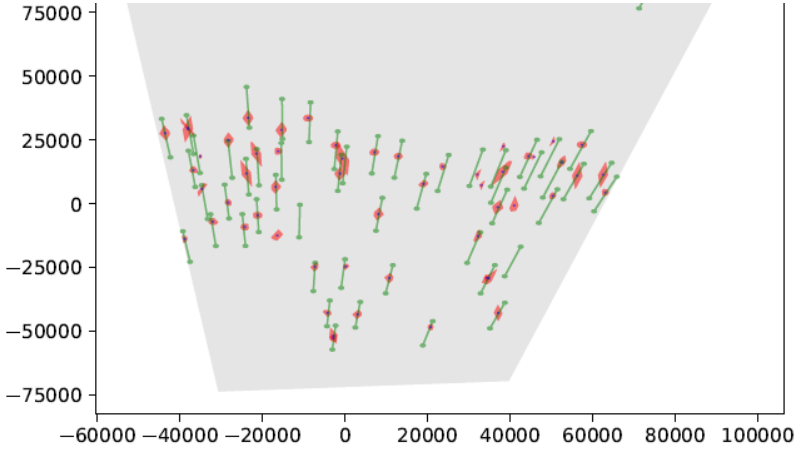


Figure – Examples of BBAs on ground plane (2D) discernment frame : (a) FEs of BBA $m_{d_{0,0}}$ (derived from pedestrian blob at $t=0$) : size of largest FE $\approx 1\times 2$ m^2), (b) 16 FEs of BBA $\tilde{m}_{t_{0,7}}$ at $t=7$ after conjunctive combination between detection BBA and track BBA, (c) FEs of summarized BBA $\tilde{m}_{t_{0,7}}$ obtained after simplification (5 FEs), FEs after dilation (polygon offsetting), i.e. corresponding to track BBA $m_{t_{0,8}}$.

Results combining appearance detector (CNN) and geometric detector (Pellicanò, 2019)



(a) Example of FE of appearance detector BBAs ; area of interest of the central camera.



(b) Example of result : GT (green segments), results (red polygons) and their barycenters (blue).

Benefits of the combination with appearance detector

- Solve the localization ambiguities
- ➡ Confirm the detections → almost no FN, only few FP ($Rec = 97.9\%$ & $Prec = 87.3\%$ versus $Rec = 74.4\%$ & $Prec = 66.8\%$ only using geometric detector)
- ➡ Increase the localization precision (average loc. error $13.1cm$ vs $19.9cm$)

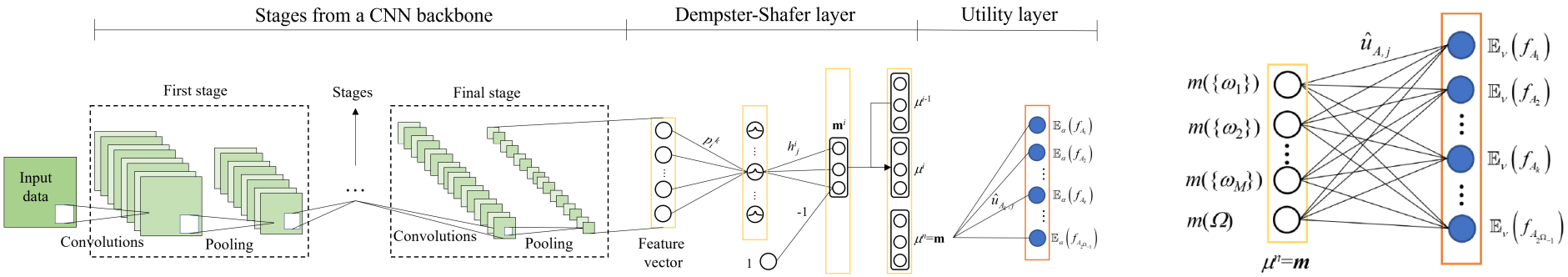
Conclusion on BF recent tools for localization

Motivating application	**Key information**	**Coding**	**Tool**
GNSS localization	Pseudo-range signals	Distinction inliers vs outliers	Evidential RANSAC
Pedestrian tracking	Imprecision on head height, imprecision on projected detections	2D FEs and their interactions	library 2CoBel https ://github.com/MOHICANS-project/2CoBel

Outline

1. Basics in belief functions
2. Belief functions for image processing
3. Belief functions for localization
4. Belief functions for Deep Learning

Evidential NN (Denœux, 2000 ; Tong et al., 2021)



(a) (b)

Figure – Architectures of (a) an evidential CNN classifier and (b) the utility layer ; (Tong et al., 2021)

Utility layer

- *Imprecise* decision performed using a utility matrix $U \in [0,1]^{c \times (2^c - 1)}$ that quantifies the utility to move from singletons to set-valued labels
- Basic idea : better predict an imprecise label than a wrong one ; e.g., $\{\text{cat}, \text{dog}\}$ rather than $\{\text{dog}\}$ if $\omega^* =$ 'cat'

Noisy label issue

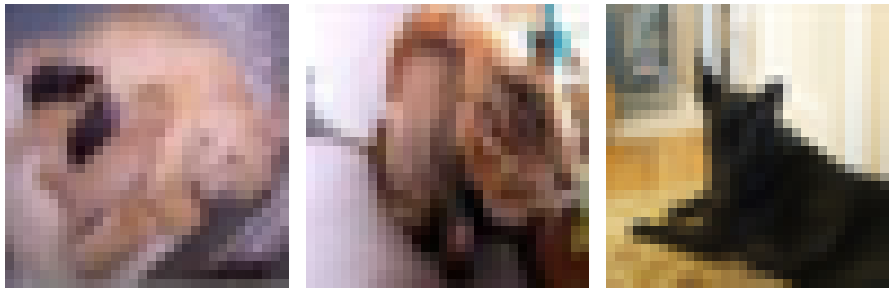
Problem statement

- Deep learning requires large well annotated datasets
- Label noise ratio in real world datasets $\in [8\%, 38\%]$ (Xiao et al., 2015)

➡ How to train DNNs under label noise in a classification setting ?

Usual approaches

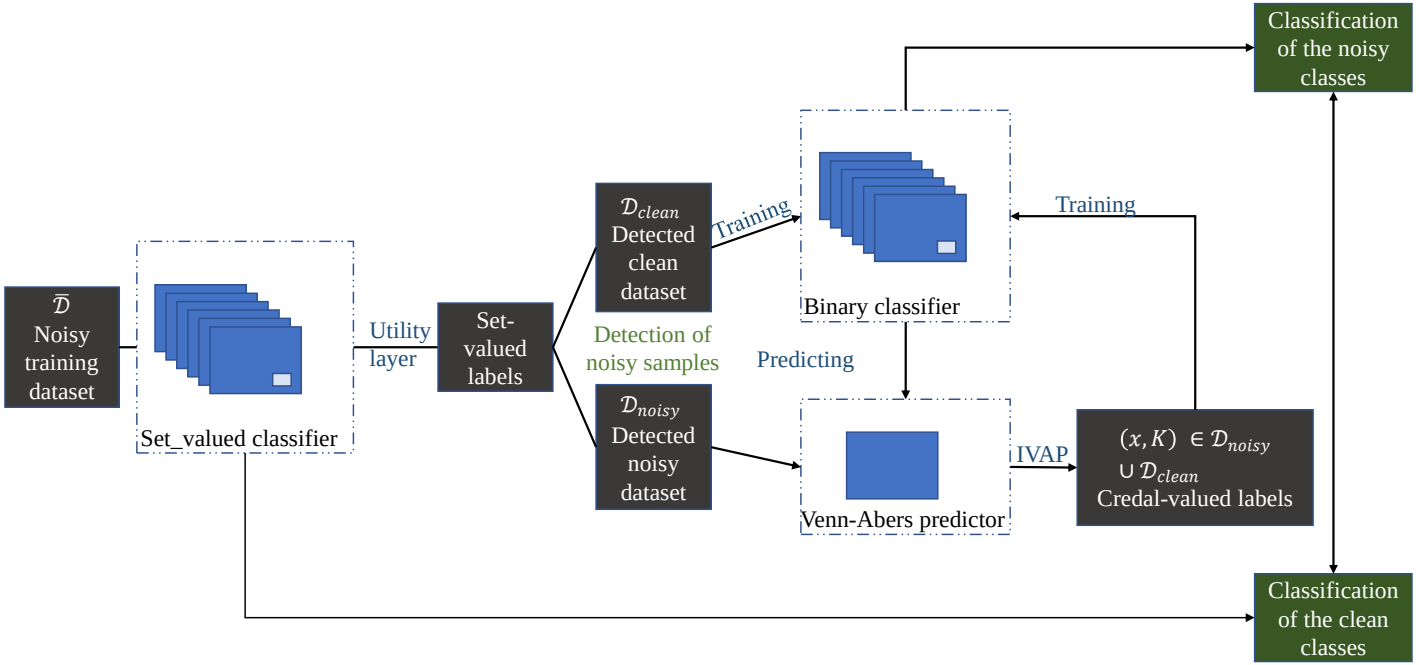
- Robust architectures
- Robust regularization
- Loss adjustment
- Sample evaluation/selection



(a) Cat (b) Cat (c) Dog

Figure – Examples of confusing samples CIFAR-10.

Robust classification with noisy labels using Venn-Abers predictors (Lemghari et al., 2024)

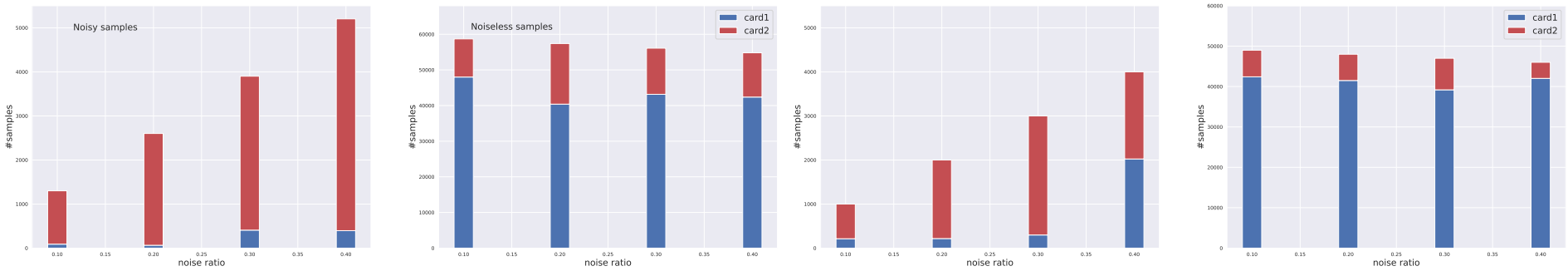


Predicted score : $f_\Theta(x) \Rightarrow 2$ isotonic regressions g_j from $\tilde{\mathscr{P}}_{cal} \cup \left\{\left(f_\Theta(x), \omega_j\right)\right\}, j \in \{0,1\}$ $\Rightarrow$ interval bounds $\underline{p}_x = g_0\left(f_\Theta(x)\right)$ and $\overline{p}_x = g_1\left(f_\Theta(x)\right)$

Credal loss adapted from Kullback Leibler divergence :

$$\mathscr{L}\left(\underline{p}_x, \overline{p}_x, f_\Theta(x)\right) = \begin{cases} 0 & \text{if } \underline{p}_x \leq f_\Theta(x) \leq \overline{p}_x, \\ \mathscr{L}^{KL}(\underline{p}_x, f_\Theta(x)) & \text{if } f_\Theta(x) \leq \underline{p}_x, \\ \mathscr{L}^{KL}(\overline{p}_x, f_\Theta(x)) & \text{if } f_\Theta(x) \geq \overline{p}_x. \end{cases}$$

Examples of results



(a) MNIST dataset (b) CIFAR-10 dataset

Figure – Numbers of set-valued labels (red) and single-valued labels (blue) versus noise level ($\in \{10\%, 20\%, 30\%, 40\%\}$) distinguishing the cases of the noisy samples (left) and not noisy ones (right)

Method/Dataset	MNIST				CIFAR-10			
noise level (%)	10	20	30	40	10	20	30	40
CNN model	0.972	0.958	0.939	0.923	0.924	0.914	0.892	0.870
T-forward	0.974	0.960	0.952	0.948	0.930	0.920	0.905	0.878
Our method	**0.990**	**0.983**	**0.973**	**0.964**	**0.941**	**0.935**	**0.913**	**0.885**

Table – Accuracy values for different percentages of symmetric noise for MNIST and CIFAR-10 datasets.

Domain-Informed and Neural-Optimized Belief Assignments (Daimellah et al., 2025)

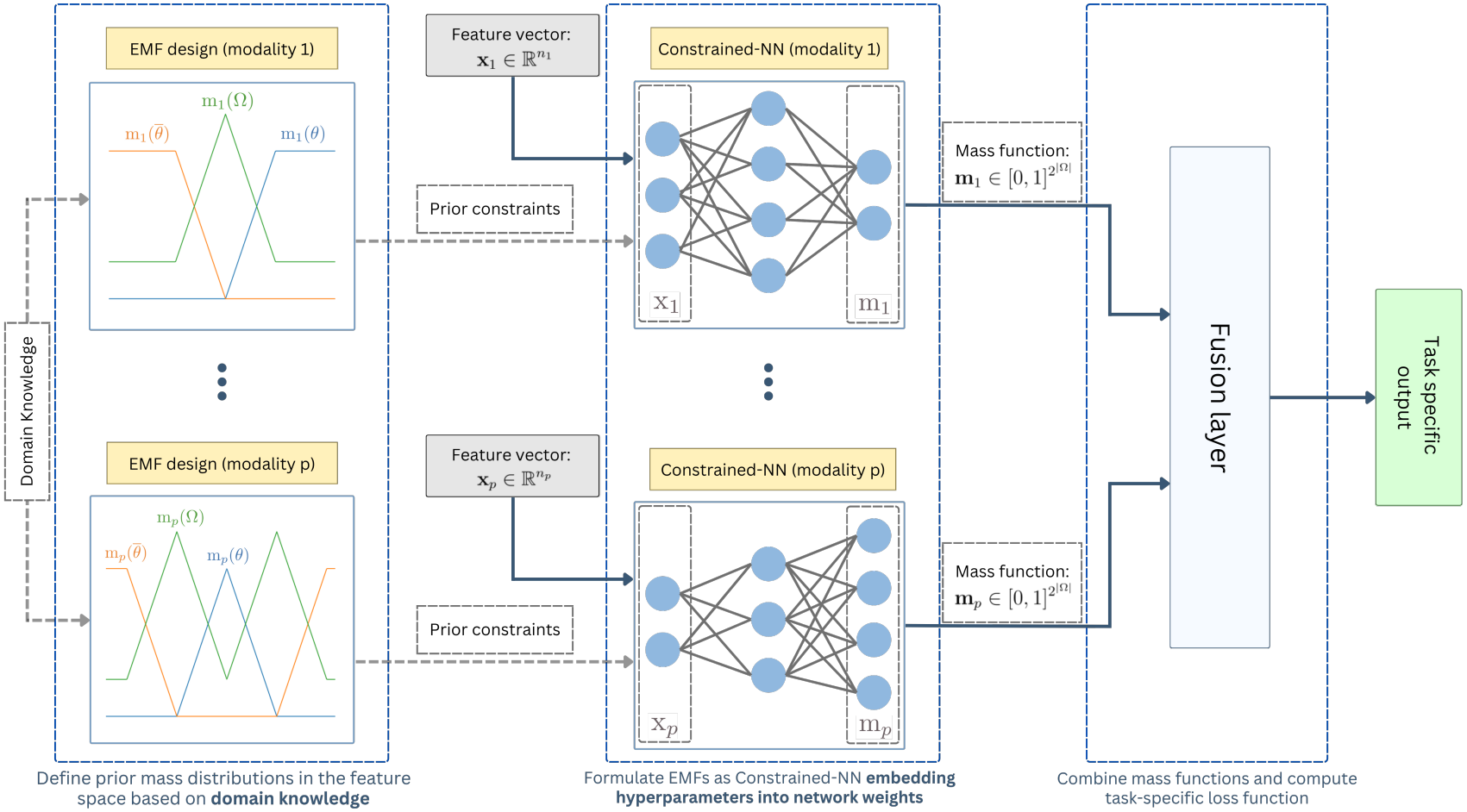


Figure – Multimodal framework for mass function assignment ; $i \in [\![1, p]\!]$ is the modality index and n_i is the dimension of the feature vector of modality i.

Domain-Informed and Neural-Optimized Belief Assignments (Daimellah et al., 2025)

Loss choice

with i data sample index, j class index, $\hat{y}_{ij} = BetP^{\Omega}(\omega_j)$,

$$\mathscr{L}_{\text{sup-1}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{|\Omega|} y_{ij}\log(\hat{y}_{ij}) + \lambda\cdot\frac{1}{N}\sum_{i=1}^{N}\left(m_i^{\Omega}(\emptyset) + \sum_{A\subseteq\Omega} m_i^{\Omega}(A)|A|\right),$$

⇒ sum of CE term, inconsistency measure and non-specificity measure

$$\mathscr{L}_{\text{sup-2}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{|\Omega|} y_{ij}\log\left(\sum_{\substack{A\subseteq\Omega\\ A\ni\omega_j}}\frac{m_i^{\Omega}(A)}{|A|}\right).$$

⇒ simplified form with implicit penalization of high conflict values (that make the resulting output vectors no longer sum to one).

Examples of results

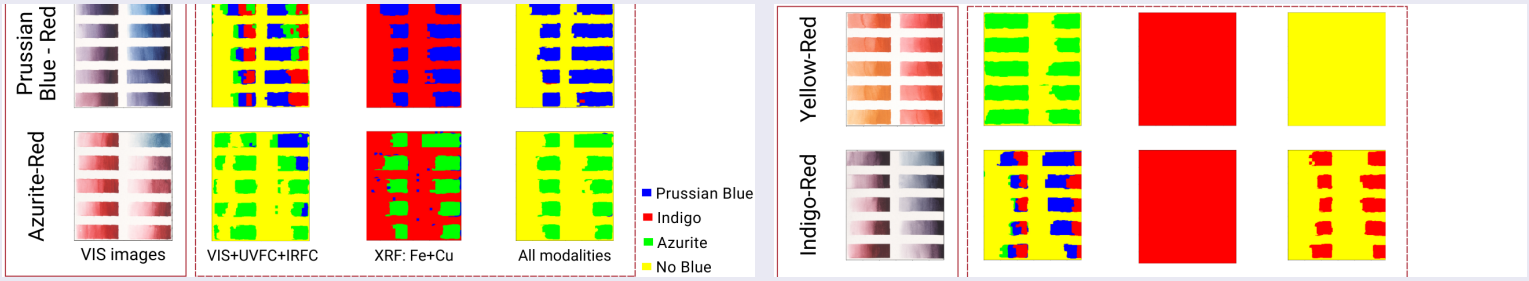


Figure – Qualitative comparison of pigment classification results using optical modalities alone, XRF modalities alone, and multimodal fusion combining all modalities.

Perspectives

- Toward more automatic definition of BBAs using binary measures of *conflict*, *cross-entropy* to compare/confront BBAs... $\rightsquigarrow$ make some correction mechanisms for BBAs automatic

- Toward explainability $\rightsquigarrow$ use sensitivity spaces to evaluate the impact
  - of (i) each BBAs (and or its hyperparameters) or (ii) piece of information (prior, observation, etc.)
  - on (i) final decision or (ii) the final BBA and its information content,

- Toward interpretability $\rightsquigarrow$ use unarity measures of *internal inconsistency*, *non-specificity* to evaluate BBA information content provided
  - that BBAs are calibrated, i.e. there are known relationships between self-evaluation metrics and decision quality,
  - experts can provide explicit 'costs' of delayed decisions

Thank you for your attention

Bibliography (1/3)

- André, C., Le Hégarat-Mascle, S., & Reynaud, R. (2015). Evidential framework for data fusion in a multi-sensor surveillance system. *Engineering Applications of Artificial Intelligence*, 43, 166-180.
- Bendjebbour, A., Delignon, Y., Fouque, L., Samson, V., & Pieczynski, W. (2001). Multisensor image segmentation using Dempster-Shafer fusion in Markov fields context. *IEEE Trans. on Geoscience and Remote Sensing*, 39(8), 1789-1798.
- Bloch, I. (1996). Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letter*, 17(8), 905-919.
- Bloch, I. (2008). Defining belief functions using mathematical morphology-application to image fusion under imprecision. *Int. J. of Approximate Reasoning*, 48(2), 437-465.
- Choi, S., Kim, T., & Yu, W. (1997). Performance evaluation of RANSAC family. *J. of Computer Vision*, 24(3), 271-300.
- Cobb, B. R., & Shenoy, P. P. (2006). On the plausibility transformation method for translating belief function models to probability models. *Int. J. of Approximate Reasoning*, 41(3), 314-330.
- Drevelle, V., & Bonnifait, P. (2011). A set-membership approach for high integrity height-aided satellite positioning. *GPS solutions*, 15(4), 357-368.
- Dempster, A. P. (1968). A generalization of Bayesian inference. *J. of the Royal Statistical Society : Series B (Methodological)*, 30(2), 205-232.
- Denœux, T. (2008). Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2-3), 234-264.

Bibliography (2/3)

- Denœux, T., El Zoghby, N., Cherfaoui, V., & Jouglet, A. (2014). Optimal object association in the Dempster-Shafer framework. *IEEE Trans. on Cybernetics*, 44(12), 2521-2531.
- Denœux, T. (2016). 40 years of Dempster-Shafer theory. *Int. J. of Approximate Reasoning*, 79(C), 1-6.
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
- Jousselme, A. L., & Maupin, P. (2012). Distances in evidence theory : Comprehensive survey and generalizations. *Int. J. of Approximate Reasoning*, 53(2), 118-145.
- Lachaize, M., Le Hégarat-Mascle, S., Aldea, E., Maitrot, A., & Reynaud, R. (2018). Evidential framework for error correcting output code classification. *Engineering Applications of Artificial Intelligence*, 73, 10-21.
- Lachaize, M., Le Hégarat-Mascle, S., Aldea, E., Maitrot, A., & Reynaud, R. (2018). Evidential split-and-merge : Application to object-based image analysis. *Int. J. of Approximate Reasoning*, 103, 303-319.
- Le Hégarat-Mascle, S., Bloch, I., & Vidal-Madjar, D. (1997). Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Trans. on Geoscience and Remote Sensing*, 35(4), 1018-1031.
- Le Hégarat-Mascle, S., Bloch, I., & Vidal-Madjar, D. (1998). Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover. *Pattern Recognition*, 31(11), 1811-1823.
- Le Hégarat-Mascle, S., & Seltz, R. (2004). Automatic change detection by evidential fusion of change indices. *Remote Sensing of Environment*, 91(3-4), 390-404.

Bibliography (3/3)

- Mercier, D., Lefevre, É., & Jolly, D. (2011). Object association with belief functions, an application with vehicles. *Information Sciences*, 181(24), 5485-5500.
- Pellicanò, N., Le Hégarat-Mascle, S., & Aldea, E. (2018). 2CoBel : A scalable belief function representation for 2D discernment frames. *Int. J. of Approximate Reasoning*, 103, 320-342.
- Pichon, F., Dubois, D., & Denœux, T. (2012). Relevance and truthfulness in information correction and fusion. *Int. J. of Approximate Reasoning*, 53(2), 159-175.
- Pichon, F., Mercier, D., Lefevre, É., & Delmotte, F. (2016). Proposition and learning of some belief function contextual correction mechanisms. *Int. J. of Approximate Reasoning*, 72, 4-42.
- Ristic, B., & Smets, P. (2006). The TBM global distance measure for the association of uncertain combat ID declarations. *Information Fusion*, 7(3), 276-284. Smets, P., & Ristic, B. (2007). Kalman filter and joint tracking and classification based on belief functions in the TBM framework. *Information Fusion*, 8(1), 16-27.
- Shafer, G. (1976). *A mathematical Theory of Evidence*. Princeton university press.
- Smets, P., & Kennes, R. (1994). The transferable belief model. *Artificial intelligence*, 66(2), 191-234.
- Smets, P. (2005). Decision making in the TBM : the necessity of the pignistic transformation. *Int. J. of Approximate Reasoning*, 38(2), 133-147.
- Vandoni, J., Aldea, E., & Le Hégarat-Mascle, S. (2019). Evidential query-by-committee active learning for pedestrian detection in high-density crowds. *Int. J. of Approximate Reasoning*, 104, 166-184.
- Zair, S., & Le Hégarat-Mascle, S. (2017). Evidential framework for robust localization using raw GNSS data. *Engineering Applications of Artificial Intelligence*, 61, 126-135.