Evidential Split-and-Merge: Application to Object-Based Image Analysis

Marie Lachaize[a,b], Sylvie Le Hégarat-Mascle[a], Emanuel Aldea[a], Aude Maitrot[b], Roger Reynaud[a]

[a] *SATIE laboratory, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay,France*
[b] *VEOLIA RECHERCHE & INNOVATION, 291 av. Dreyfous Ducas, Limay, France*

Abstract

This paper addresses the difficult problem of segmenting objects in a scene and simultaneously estimating their material class. Focusing on the case where, individually, no dataset can achieve such a task, multiple sensor datasets are considered, including some images for retrieving the spatial information. The proposed approach is based on mutual validation between class decision (using the most relevant dataset) and segmentation (derived from image data). The main originality relies in the ability to make these two modules (classification and segmentation) interactive. Specifically, our application focuses on object-level material labeling using classic RGB images, laser profilometer images and a NIR spectral sensor. Starting from a superpixel segmentation, the relevant data are introduced as constraints modifying the initial segmentation in a split-and-merge process, which interacts with the material labeling process. In this work, we use the belief function framework to model the information extracted from each kind of data and to transfer it from one processing module to another. In particular we show the relevance of evidential conflict measure to drive the split process and to control the merge one.
Experiments have been performed on actual scenes with stacked objects and difficult cases of material such as transparent polymers. They allow us to assess the performance of the proposed approach both in terms of material labeling and object segmentation as well as to illustrate some borderline cases.

Keywords: Information fusion, Belief Function Theory, Image segmentation, Object classification, Spectral data, RGB-D images

1. Introduction

In numerous computer vision applications dealing with scene analysis and understanding, object detection and labeling are key steps. In order to detect separate objects

*Sylvie Le Hégarat-Mascle
Email addresses: marie.lachaize@u-psud.fr (Marie Lachaize), sylvie.le-hegarat@u-psud.fr (Sylvie Le Hégarat-Mascle), emanuel.aldea@u-psud.fr (Emanuel Aldea), aude.maitrot@veolia.fr (Aude Maitrot), roger.reynaud@u-psud.fr (Roger Reynaud)

on a background, or to detect objects among other ones of a different kind or class, one may rely on their intrinsic features, e.g., material or color depending on the application. However, a much more difficult task is to detect individual instances of similar objects. In this case, the help of additional features such as object spatial features related to their relative location is required. Now, to acquire these additional features, there is generally a strong benefit from using several sensors, each of them being dedicated to a given feature. Indeed, one can expect that with dedicated sensors, each feature of interest is measured in a more accurate and robust way. For instance, by measuring the whole spectral response, spectral sensors (e.g., spectral camera or spectrometer) allow for a very accurate characterization of materials, whereas, by measuring directly the distance between sensor and object surface, 3D sensors (e.g, laser triangulation) provide precise point locations in the 3D space.

Then, our problem becomes a data fusion problem and defining an appropriate approach to combine heterogeneous types of information is all the more challenging that the data geometry and scale vary. Apart from the image co-registration problem, object-oriented approaches provide a solution to avoid the registration at pixel level by designing a segmentation, which yields to a new spatial support to derive and to handle meaningful pieces of information.

The benefits of this approach have been proven by a lot of research works and applications ([10], [26]). In particular, it has been invoked for the classification of remote sensing images based on acquisitions from different sensors at different resolutions and incidence angles as OBIA or GEOBIA for (Geographic) Object Based Image Analysis ([2], [9], [11] [3]).

The complementarity of the different data resolutions is exploited as follows: at the finest (pixel) level, the high spatial resolution allows for an exhaustive delineation of the objects or object sub-parts, while at region level higher-order statistical information derived from the high resolution sensors enriched by additional information provided by the coarser sensors allows for a more reliable region-based classification.

However, in cited works, the segmentation is done preliminarily and independently of the following processing steps. Any segmentation error impacts permanently the subsequent stages. In this work, we use the data complementarity to detect some imperfections in the segmentation.

Specifically, we consider an application of detection and classification of stacked-upon objects involving several sensors: a NIR (Near-InfraRed) spectral sensor, a 3D sensor and an RGB camera. The first one provides the spectral information, which is crucial for material classification. However, this sensor has a low framerate that involves a subsampling of the spectral data with regard to the RGB data. In the following, this kind of data is referred to as a pseudo-image. Meanwhile, the 3D sensor and the RGB camera provide high resolution images of the scene. The 3D sensor (a laser profilometer) also provides an image that can be linked to the intensity reflected by an object, which offers additional information. It will be denoted as the brightness image. Table 1 highlights the main features of the considered sensors, namely spatial resolution and ability for material labeling, but also ability to distinguish different objects as whole entities (the ideal case would correspond to *high* object separation ability associated with *high* intra-object homogeneity).

The complementarity of the information derived from each sensor as well as the

Table 1: Sensor features with regard to object detection and classification.

Sensors / Data type	**Intra-object homogeneity**	**Object separation ability**	**Material labeling ability**	**Spatial resolution**
NIR spectral	**High** except for multimaterial objects	**High** Except for objects of the same material	Yes	Low
RGB	Low	Medium	Low	High
3D	**High** Except for transparent and specular material	**High** except for same height objects, transparent and specular material	No	High
Brightness	Low Except for transparent material	Medium	Low	High

non-coincidence of the spatial supports naturally lead to an object-oriented approach. According to sensor features given in Table 1, it seems judicious to base the classification process on the spectral data, and the segmentation task on the 3D and RGB information. Then, following a cross-validation approach, in addition to its specific task, each processing module (segmentation and classification) will assess the reliability of the other one. The objective is to get a result under the form of material-labeled regions as close as possible to objects or object sub-parts in case of heterogeneous objects: (i) two different objects should not be in the same region, and (ii) the number of regions within a given object should be as small as possible (ideally equal to one).

In this work, we propose to handle the interactions between segmentation and classification modules using belief functions. This formalism has been proposed for numerous problems in various fields ([5, 24, 15, 19, 17]) and is now recognized for its ability to model partial or imprecise information as well as its ability to evaluate the conflict between pieces of information. In this work, the belief function framework will be firstly used to model the uncertainty of the classification step. Secondly, it will provide the two main tools for making segmentation and classification interact. Essentially, while segmentation will allow for robust classification, classification will allow for the evolution of the segmentation in a process akin to a Split-and-Merge strategy. During these interactions, the evidential combination will carry information throughout the evolution of the spatial support, and the evidential conflict measure will be used to evaluate the relevance of any potential action (such as labeling, splitting or merging).

The paper is organized as follows: Section 2 introduces the belief function tools and notations used in this work. Section 3 details the proposed approach and how we model interactions between the classification and segmentation modules. Section 4 discusses our experimental results. Section 5 summarizes the conclusions and perspectives of this work.

2. Preliminaries on Belief Function Theory (BFT)

In this section, we introduce the tools and notations used in this study. For a reader not familiar with BFT, we refer to the seminal book of [22].

Let Ω denote the **discernment frame**, i.e. the set of mutually exclusive hypotheses representing the solution possibilities and let 2^Ω denote the power set of Ω, i.e. the set of subsets of Ω elements. Denoting by $|A|$ the cardinality of a set A, $\left|2^\Omega\right| = 2^{|\Omega|}$. A bba (basic belief assignment) is defined through its **mass function** m^Ω such that: $m^\Omega : 2^\Omega \to [0,1]$, $\sum_{A\in 2^\Omega} m^\Omega(A) = 1$. The quantity $m^\Omega(A)$ represents the belief that the solution is in A without further specification. Let $\mathcal{F}_m = \left\{A \in 2^\Omega, m^\Omega(A) > 0\right\}$ denote the set of the focal elements of the bba m^Ω. Under the open-world assumption, $\emptyset$ may be a focal element with its mass representing the belief that Ω is at odds with the solution. If $\emptyset \in \mathcal{F}_m$, the bba is called *sub-normal*, otherwise (i.e. when $m^\Omega(\emptyset) = 0$), it is called *normal*.

The **conjunctive combination rule** is widely used because of its simplicity, its ability to specify the information and its convenient mathematical properties (in particular commutativity and associativity). It holds in the case of two independent bbas m_1^Ω and m_2^Ω defined on the same discernment frame:

$$\begin{aligned}\forall A \in 2^\Omega, m^\Omega_{1\textcircled{\cap}2}(A) =& m_1^\Omega \textcircled{\cap} m_2^\Omega(A)\\ =& \sum_{\substack{(B,C)\in \mathcal{F}_{m_1}\times\mathcal{F}_{m_2},\\ B\cap C = A}} m_1^\Omega(B)\, m_2^\Omega(C)\,. \end{aligned} \tag{1}$$

The mass on the empty set, $m^\Omega_{1\textcircled{\cap}2}(\emptyset)$, appearing after a bba conjunctive combination is often interpreted as a measure of disagreement or **conflict** between the combined bbas. In this work, it will allow us to quantify the disagreement between two beliefs. Now, since conflict can only increase following conjunctive combination, bba normalization is necessary to prevent conflict propagation from previous steps.

As shown in [13], many combination rules presented as alternatives to Eq. (1) come down to the derivation of the *sub-normal* bba provided by Eq. (1) followed by a normalization step that is rule-specific. Among these possible normalizations, we focus on the one included in Yager's rule [27] that, in the absence of knowledge of the conflict origin, transfers it to the ignorance:

$$\begin{cases} \overline{m}^\Omega(\emptyset) &= 0,\\ \overline{m}^\Omega(\Omega) &= m^\Omega(\Omega) + m^\Omega(\emptyset),\\ \overline{m}^\Omega(A) &= m^\Omega(A), \qquad \forall A \in 2^\Omega \setminus \{\emptyset, \Omega\}\,. \end{cases} \tag{2}$$

Decision is generally made based on a function that supports a probabilistic interpretation. The three most used functions are the plausibility, *Pl*, the credibility, *Bel*, and the pignistic probability, *BetP*. The two first ones are in one-to-one relationship with m^Ω and may be interpreted as upper and lower bounds of an imprecise probability function, whereas *BetP* proposed in [23] is a probability measure defined on Ω

(provided that $m^{\Omega}(\emptyset) < 1$).

$$\forall A \in 2^{\Omega}, Pl^{\Omega}(A) = \sum_{B \in \mathcal{F}_m | A \cap B \neq \emptyset} m^{\Omega}(B), \tag{3}$$

$$\forall A \in 2^{\Omega}, Bel^{\Omega}(A) = \sum_{B \in \mathcal{F}_m | B \subseteq A} m^{\Omega}(B), \tag{4}$$

$$\forall \omega \in \Omega, BetP^{\Omega}(\omega) = \frac{1}{1 - m^{\Omega}(\emptyset)} \sum_{B \in \mathcal{F}_m | \omega \in B} \frac{m^{\Omega}(B)}{|B|}. \tag{5}$$

Finally, the imprecision measure quantifies the discrepancy between the pieces of evidence that support a hypothesis and the ones that are not in contradiction with the hypothesis, i.e., the plausibility and credibility values. The imprecision measure comes from interpreting BFT as a particular case of imprecise probability [25] so that, according to [6], the interval $[Bel(A), Pl(A)]$ receives a natural interpretation as an imprecise specification of some unknown probability $P(A)$. In this work, imprecision is evaluated considering the favorite hypothesis according to $BetP$ function, so that the lower the result of Eq. (6) is, the more reliable the decision making process is.

$$\iota = Pl^{\Omega}\left(\arg\max_{\omega \in \Omega} BetP^{\Omega}(\omega)\right) - Bel^{\Omega}\left(\arg\max_{\omega \in \Omega} BetP^{\Omega}(\omega)\right). \tag{6}$$

3. Proposed approach

3.1. Data characterization

Our problem is a joint problem of segmenting the objects of a scene and of labeling them according to their material class. If $\mathcal{S}$ denotes the desired segmentation, Ω the finite set of labels, and l_s the label of any given element s in $\mathcal{S}$, our objective is to infer both $\mathcal{S}$ and the associated labels $\{l_s\}_{s \in \mathcal{S}}$ in Ω.

In order to make the segmentation correspond (as much as possible) to the objects, the representation of the scene provided by the sensors should be homogeneous inside each object while highlighting object edges. However, in practice, we cannot avoid facing two types of limitations related to the observations: some data are incomplete or spatially imprecise while some others, that are spatially precise, fail to fully comply with the intra-object homogeneity constraint. In the following, we refer to data spatially incomplete (sparse, coarse) as *pseudo-images* as opposed to *true* images, which define the dense support of our setting. The word *pseudo-image* refers to the fact that any *pseudo-image* sample can be associated to a region in $\mathcal{S}$ (via a function of coarse co-registration). Then, denoting by $\mathcal{K}$ the set of indexes of the datasets, we distinguish within $\mathcal{K}$, the sets of *true* image indexes, $\mathcal{K}'$ with $\mathcal{K}' \subsetneq \mathcal{K}$, and of *pseudo-image* indexes, $\overline{\mathcal{K}'}$. The spatial information is thus encoded by $\mathcal{K}'$ data, and the homogeneity constraints for object segmentation will be derived from $\mathcal{K}''$ with $\overline{\mathcal{K}'} \subseteq \mathcal{K}'' \subsetneq \mathcal{K}$.

As an illustration, let us specify the data sets $\mathcal{K}'$, $\overline{\mathcal{K}'}$ and $\mathcal{K}''$ in the case of our application. $\mathcal{K}'$ corresponds to the data provided by the 3D sensor and by the RGB image, i.e. the data with the best spatial resolution in the three dimensions (i.e. including the height information). However, transparent materials are challenging for

both sensors and could lead to incorrect zone delimitations. $\overline{\mathcal{K}'}$ corresponds to spectral *pseudo-image*. Even if the sensor is quite efficient, the classification information can be ambiguous in the case of close spectral signatures of materials, or because of the possible presence of mixed pixels (at the borders of the objects for example). Finally, $\mathcal{K}''$ corresponds to the 3D data and to the spectral ones, which suits our application where most of the researched objects are homogeneous in terms of material and in terms of height localization.

With the previous notations, our problem comes down to the estimation of the labeled segmentation $(\mathcal{S}, \{l_s\}_{s\in\mathcal{S}})$, with $\mathcal{S}$ defined on $\mathcal{K}'$ spatial support and satisfying $\mathcal{K}''$ homogeneity constraints. However, this estimation is an *ill-posed* problem due to the *pseudo-images*, which do not allow for the accurate estimation of some edge parts.

3.2. Outline of the proposed approach

The proposed approach presents two original aspects. The first aspect derives directly from the previous formulation of our problem with respect to the available data: in order to overcome the ill-posed edge localization problem highlighted above, the set of *true* images denoted as $\overline{\mathcal{K}''}$ (which is necessarily included in $\mathcal{K}'$), is introduced despite its irrelevance with homogeneity constraints. These data allow us to compensate for the lack of edge information. For instance, RGB images will allow us to recover some edges corresponding to the violation of the homogeneity constraint on the material class, provided by the NIR spectral data (*pseudo-image*). However, since there are occasional mismatches between the corresponding representations for a same object, the derived set of edges exhibits both false negatives (in case of identical color for different materials, following our example of RGB supplementing NIR spectral data) and false positives (case of different colors for a mono-material object).

Practically, the proposed estimation starts by $\mathcal{K}'$ constraints (spatial information under the form of a first segmentation), then $\overline{\mathcal{K}'}$ constraints are added (refinement step) and finally, a subset of $\mathcal{K}'$ constraints (simplification step) is relaxed. The result of the refinement step is an intermediate segmentation denoted $\mathcal{S}^h$ such that every element s in this support is homogeneous with respect to the whole set of data while the result of the simplification step, that is denoted $\mathcal{S}^o$ in contrast to $\mathcal{S}^h$, corresponds to our estimation of $\mathcal{S}$. As represented in Figure 1, the refinement and simplification steps will be performed using segment split and segment merge processes, respectively.

The second novelty of this work is the use of the belief function framework to drive the split and merge processes. Firstly, belief functions (BF) are used to model the uncertainty inherent to the classification step. Then, they are also used to propagate the information from step to step by evaluating the consistency within any element of $\mathcal{S}$ and the consistency between adjacent elements of $\mathcal{S}$. Specifically, the conflict measure as well as the imprecision measure (Eq. (6)) are indicators of the consistency (either intra or inter $\mathcal{S}$ elements).

The next sections specify how these evidential measures are used in the three main modules, *labeling* (Section 3.3.1), *Splitting* (Section 3.3.2), and *Merging* (Section 3.3.3), cf. Figure 1, such that:

Labeling This step involves the construction of a bba m_s^Ω for each element s in $\mathcal{S}$. The segmentation provides here a tool to select and propagate the classification

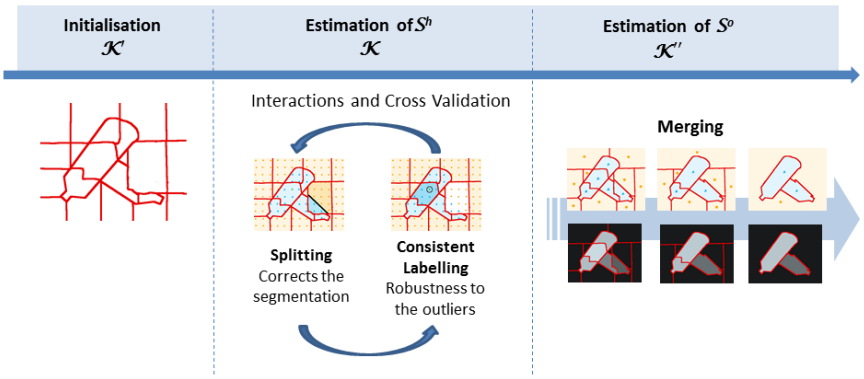


Figure 1: Layout of our approach: from $\mathcal{S}$ initialization towards $\mathcal{S}^h$ and then $\mathcal{S}^o$.

information from the pixel level to a higher level defined by the segments.

Splitting In this step we modify $\mathcal{S}$ based on the analysis of the bbas m_s^Ω, which allows for the detection of the inconsistent elements s in $\mathcal{S}$. After this step, $\mathcal{S}$ is an estimation of $\mathcal{S}^h$.

Merging This step aims to modify $\mathcal{S}$ in order to make the segmentation evolve towards the object level. Region merging is based on the detection of adjacent segments exhibiting consistent bbas that are defined on the frames of discernment associated with the features derived from $\mathcal{K}''$ data.

3.3. Evidential Split-and-Merge

In order to illustrate the different steps of our approach, we consider a running example. These steps are highlighted progressively in Figure 2 as we advance in presenting our algorithm. The first line shows an example of scene acquired by these different sensors.

Let us now detail the three main steps of our approach and their interactions.

3.3.1. Consistent labeling

Here, we aim at labeling every element s of $\mathcal{S}$. To do this, we draw some samples of the $\overline{\mathcal{K}'}$ data associated to s, expecting at least two benefits. First, drawing several samples and validating their consistency strengthens the labeling relative to the presence of some outlier samples. Secondly, it allows us to propagate the labeling information from the classification at the sample level to the segment level when faced with imprecise registration (e.g., using a margin with respect to segment borders). Finally, it may allow us to decrease the processing time by avoiding to label every data sample.

In the spirit of the RANSAC algorithm [4], for each subset of samples randomly drawn, the disagreement between its elements as well as its reliability are evaluated. In this work, consistency and reliability are both asserted by evidential measures. For any sample p, let m_p^Ω denote its associated bba on Ω (set of labels). In Section 4.1, we

specify the way m_p^Ω is derived in our application but alternative approaches could have been considered provided their output is a bba. Then, for any subset of samples, the bba combination is performed according to the conjunctive rule (Eq. (1)) and analyzed both in terms of consistency and in terms of commitment. On the one hand, if the combined bba exhibits a high conflict value, then either at least one of the samples was already tricky (having a high $\emptyset$ mass from the multilabel bba allocation) or the samples show a high disagreement among themselves. In both cases, the selected subset of samples is not consistent and should be rejected. On the other hand, very little committed bbas (i.e., highly imprecise or uninformative samples), while being much less conflicting, cannot be considered as reliable because of the lack of carried information. However, in this case, the selected subset of samples may be complemented by more committed samples also randomly drawn.

Algorithm 1 describes the proposed labeling of an $\mathcal{S}$ element by random sampling of the spectral pseudo-image. Besides the considered segment and the $\overline{\mathcal{K}'}$ data associated to s, the algorithm requires five parameters: the desired number of samples to combine, three threshold parameters and the maximal number of iterations (in the event that no consensus would be found). The first threshold parameter τ_0 is used in the prefiltering step of the drawn samples (noting that conflict may only increase with conjunction combination rule, we find useless to select samples already exhibiting a high conflict from the multilabel classification). The next step is the combination: combining the *normal* bbas (of the tested subset) using the conjunctive combination, the bba m_s^Ω is formed. Normalization (Eq. (2)) is applied to avoid bias when evaluating the consistency within a segment based on the samples. Then, the reliability of a decision according to m_s^Ω is evaluated. This evaluation involves the two threshold parameters τ_S and θ_S as follows: if the conflict measure is too high ($\kappa_s > \tau_S$), the subset is considered to be inconsistent and it is discarded. If the imprecision measure is high but the conflict is low ($i_s > \theta_S$), the subset is considered to require additional data, and bbas of extra samples are added to the combination. The output variables of the algorithm are a Boolean that indicates if m_s^Ω is sufficiently reliable to allow for a decision, m_s^Ω and the set $\mathcal{E}_s$ of the drawn samples (bba and location).

From the output of Algorithm 1, two actions can be performed:

- If m_s^Ω is reliable, a decision can be drawn from m_s^Ω to assign a label to the segment s. Besides, m_s^Ω is saved as a characteristic of the element s in the segment merging step (see Section 3.3.3).
- Otherwise, if m_s^Ω conflict is high, then the class information does not match the segmentation and the segment s shall be split to better suit the data (see Section 3.3.2).

3.3.2. Segmentation refinement

In this module, we assume that a first estimation of $\mathcal{S}$ has been performed based on the data in $\mathcal{K}'$. In Section 4.1, we will specify the way we derive the initial guess for $\mathcal{S}$ in our application, but alternative approaches could have been considered. At this stage, the elements in $\mathcal{S}$ do not take into account the homogeneity constraints derived from the data in $\overline{\mathcal{K}'}$, i.e., from the classification data.

Algorithm 1: Consensual labeling; inputs: s element of $\mathcal{S}$, dataset $\mathcal{D}_s$ associated to s; parameters: M number of samples to draw, τ_0 and τ_S conflict thresholds, θ_S imprecision threshold, n_{it} maximal number of iterations; outputs: m_s^Ω bba associated to segment s, $\mathcal{E}_s$ set of drawn samples, boolean ok indicating if the founded solution is consensual.

$\mathcal{E}_s \leftarrow \emptyset$; $\mathcal{E}'_s \leftarrow \emptyset$; $ok \leftarrow 0$;
Set integer $it \leftarrow 0$;
Initialize m'_s such that $m'_s(\emptyset) = 1$;
while $ok = 0$ *and* $it < n_{it}$ **do**
  $it \leftarrow it + 1$;
  while $|\mathcal{E}'_s| < M$ *and* $|\mathcal{E}_s| < |\mathcal{D}_s|$ **do**
    Randomly draw (without replacement) a sample p_i in $\mathcal{D}_s$;
    Derive its bba m_i^Ω;
    if $m_i^\Omega(\emptyset) < \tau_0$ **then**
      $\mathcal{E}'_s \leftarrow \mathcal{E}'_s \cup \left\{(p_i, m_i^\Omega)\right\}$;
    end
  end
  $\mathcal{E}_s \leftarrow \mathcal{E}_s \cup \mathcal{E}'_s$;
  Normalize the bbas of $\mathcal{E}'_s$: $m_i^\Omega \leftarrow \overline{m_i^\Omega}$ according to Eq. (2);
  Compute the combined bba $m_s^\Omega = \bigcirc\!\!\!\!\cap_{i/(p_i, m_i^\Omega) \in \mathcal{E}'_s} m_i^\Omega$;
  Deduce $\kappa_s = m_s^\Omega(\emptyset)$ and ι_s according to Eq. (6);
  if $\kappa_s < m'_s(\emptyset)$ **then**
    $m'_s \leftarrow m_s^\Omega$;
  end
  if $\kappa_s < \tau_S$ **then**
    if $\iota_s < \theta_S$ **then**
      $ok \leftarrow 1$;
    else
      $M \leftarrow M + 1$;
    end
  else
    $\mathcal{E}'_s \leftarrow \emptyset$;
  end
end
if $ok = 0$ **then**
  $m_s^\Omega \leftarrow m'_s$;
end

We focus on the segments pointed out by the consensual labeling step as "highly conflicting", and that are very likely to include several different materials. In other words, there is in this case significant evidence for a missing edge.

The efficiency of the derivation of the frontier inside the segment s depends on the

Algorithm 2: Segment splitting; inputs: segment s (set of pixels), $\mathcal{E}_s$ set of samples drawn inside s with each $\mathcal{E}_s$ element including the location and the bba on Ω of the sample; outputs: $\{s_j\}$ set of subsegments obtained by s splitting.

Initialize to 0 the element of a table T [] of size $|\Omega|$;
forall $(m_i, p_i) \in \mathcal{E}_s$ **do**
    Compute its label l_i in Ω and save it in (m_i, p_i, l_i);
    Increment the l_i^{th} element of T: $T[l_i] \leftarrow T[l_i] + 1$;
end
From T deduce the Θ_s the restriction of Ω to the two most frequent labels l_i in s;
$\mathcal{B} \leftarrow \emptyset$;
forall $(m_i, p_i, l_i) \in \mathcal{E}_s$ **do**
    if $l_i \in \Theta_s$ **then**
        $\mathcal{B} \leftarrow \mathcal{B} \cup \{(x_i, y_i, l_i)\}$
    end
end
Train the linear SVM classifier with $\mathcal{B}$;
Derive the linear inner border **b** in s;
Label the pixels $\{p_k \in s\}$ of s in $\{-1, 1\}$ according to their position with respect to **b**;
The s subsegments $\{s_j\}$ are the connected components of the two subsets of $\{p_k \in s\}$ labeled either -1 or $+1$, respectively;

specific form of the data in $\overline{\mathcal{K}'}$. Specifically, to allow for the segment spatial refinement, the samples must be localized, at least roughly, within the segment s. Let $p_i = (x_i, y_i)$ represent the location of the sample i. In this work, we propose that the new frontier be simply inferred from the already drawn spectral samples for the labeling step, namely $\mathcal{E}_s$. Since this set is a subset of the samples associated to segment s, the s refinement based on $\mathcal{E}_s$ can be viewed as a trade-off between the final resolution of $\mathcal{S}$ elements and the algorithm complexity.

For the sake of simplicity, we assume that the considered segment s contains only two (main) classes. Besides being the case in practice in our application, such an assumption may be easily relaxed. Then, from the set of bbas saved in $\mathcal{E}_s$, the two classes are determined and their disjunction is denoted by Θ_s. Then, we only consider the set of triplets $\{(x_i, y_i, l_i)\}_{i \in \mathcal{E}_s, l_i \in \Theta_s}$ of the samples i of $\mathcal{E}_s$ whose label l_i (according to the $BetP_{s,i}$ decision) is in Θ_s. This triplet set provides the database to estimate the desired frontier inside s. Focusing on linear frontiers, the inner frontier is estimated using a linear SVM as presented in Algorithm 2.

Each new segment s' created is tagged with a new index. The splitting may result in more than two segments since the superpixels are not necessarily convex. Note also that this splitting process could be repeated if necessary (consensual labeling still not satisfied). Conversely, the splitting process can also be stopped if the segment size is too small.

The second line of Figure 2 illustrates the joint labeling and split processes. Fig-

ure 2d shows the segmentation initialization, Figure 2e the estimated labels on initial segmentation (with the RGB intensity image as the background with the colors modified according to the class label), Figure 2f the four splits, and Figure 2g the new label map. We note that the better object delineation (induced by the splits) also improves the spatial precision of the label map (at the pixel level).

3.3.3. Segmentation simplification

After the segmentation refinement step, $\mathcal{S}$ is a common spatial support with segments that are considered as homogeneous relatively to the data in $\mathcal{K}$. The $\mathcal{S}$ elements are labeled with $\{l_s\}_{s\in\mathcal{S}} \in \Omega^{|\mathcal{S}|}$ and the set of bbas associated to this labeling, namely $\left\{m_s^\Omega\right\}_{s\in\mathcal{S}}$, is stored.

To obtain object-level segmentation, $\mathcal{S}$ has now to be simplified by merging some adjacent segments. For this step, we propose to consider the conflict measure between the belief functions corresponding to adjacent segments as the merging criterion (in a similar way to [21]). The bbas representing material (class) information are already available in $\left\{m_s^\Omega\right\}_{s\in\mathcal{S}}$.

In order to consider height information as well, in this section we assume, for each segment, the availability of a bba summarizing the belief about its height. Denoting by $\mathcal{H}$ the discernment frame associated to height values, $\left\{m_s^\mathcal{H}\right\}_{s\in\mathcal{S}}$ denotes the set of height bbas. In Section 4.1, we specify the way $m_s^\mathcal{H}$ are derived in our application but alternative approaches could have been considered provided their output is a bba on $\mathcal{H}$.

Then, both kinds of bbas (defined on Ω and on $\mathcal{H}$) are used in the following iterative association algorithm.

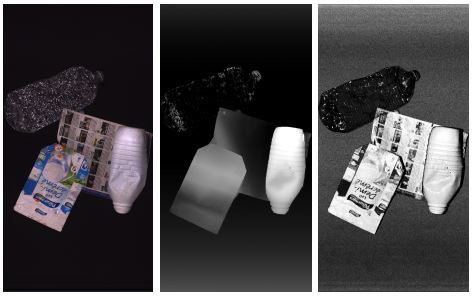
Mutual agreement algorithm. Let $\mathcal{N}_\mathcal{S}$ denote the set of neighboring pairs of $\mathcal{S}$ elements, i.e. that share a common border. The principle of the mutual agreement algorithm is to merge iteratively pairs of regions in $\mathcal{N}_\mathcal{S}$ if and only if they are their mutual best match. Practically, mutual agreement algorithm is instantiated using a valued graph $\mathcal{G}$ representing the region set, their interactions and the cost of each possible region merge. In our case, the nodes are the elements of the current segmentation $\mathcal{S}$, the graph edges are the $\mathcal{N}_\mathcal{S}$ elements and the edge cost is defined from the conflict value resulting from the combination of the bbas of the linked regions as follows.

Denoting by $m^\mathcal{H}_{i\textcircled{\cap}j}$ and $m^\Omega_{i\textcircled{\cap}j}$ the bbas resulting of the conjunctive combination of the bbas of the two adjacent regions i and j (region pair indexed by $\{i, j\} \in \mathcal{N}_\mathcal{S}$), height conflict value and class conflict value are, respectively, equal to $m^\mathcal{H}_{i\textcircled{\cap}j}(\emptyset)$ and $m^\Omega_{i\textcircled{\cap}j}(\emptyset)$, and $\{i, j\}$ cost, $v(\{i, j\})$, is set equal to

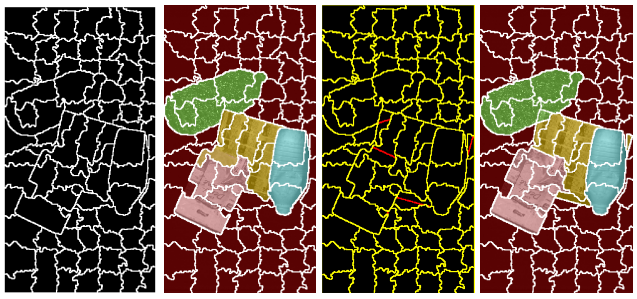
$$\begin{cases} v(\{i,j\}) = 1, & \text{if } m^\Omega_{i\textcircled{\cap}j}(\emptyset) > \tau^\Omega_\kappa, \text{ or } m^\mathcal{H}_{i\textcircled{\cap}j}(\emptyset) > \tau^\mathcal{H}_\kappa, \\ v(\{i,j\}) = m^\Omega_{i\textcircled{\cap}j}(\emptyset) \times m^\mathcal{H}_{i\textcircled{\cap}j}(\emptyset), & \text{otherwise,} \end{cases} \tag{7}$$

where τ^Ω_κ and $\tau^\mathcal{H}_\kappa$ are two threshold parameters.

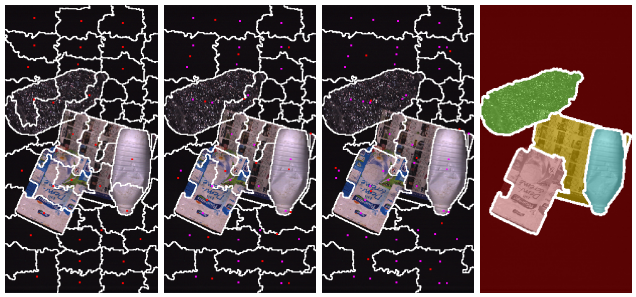
Algorithm 3 describes the region-merging step. Two conflict threshold parameters drive indirectly the stop criterion. Indeed, since conflict increases with conjunctive combination (when regions are merged), it can only exceed the conflict thresholds after



(a) RGB image (b) Height image (c) Brightness image



(d) Segmentation initialization (e) labeling before split step (f) Edges added by split step (g) labeling after split step



(h) Merge step, iteration 1 (i) Merge step, iteration 2 (j) Merge step, iteration 3 (k) Merge step, final result

Figure 2: Running example: $\mathcal{K}'$ images (1^{st} line); consensual labeling and splitting (2^{nd} line) with colored labels superimposed to gray-level intensity image (2e and 2g) and added splits in red in 2f; merge step iterations (3^{rd} line), involving 30, 18 and 10 segment merges for the three first iterations, respectively.

a bounded number of merge operations, making the region-merging stop by lack of satisfying candidates.

Algorithm 3: Segment merging; inputs: initial segmentation defined by region set $\mathcal{S}$ and adjacency relationships $\mathcal{N}_\mathcal{S}$, bba sets $\{m_s^\Omega\}_{s\in\mathcal{S}}$ and $\{m_s^\mathcal{H}\}_{s\in\mathcal{S}}$; parameters: threshold parameters τ_κ^Ω and $\tau_\kappa^\mathcal{H}$; outputs: new segmentation $\mathcal{S}'$, updated bba sets $\{m_s^\Omega\}_{s\in\mathcal{S}'}$ and $\{m_s^\mathcal{H}\}_{s\in\mathcal{S}'}$.

Define graph $\mathcal{G} = (\mathcal{S}, \mathcal{N}_\mathcal{S})$;
Associate to each $\mathcal{G}$ node s the two bbas m_s^Ω and $m_s^\mathcal{H}$;
Compute $\mathcal{G}$ arc values $v(\{i, j\})$ according to Eq. (7);
Set boolean $ok \leftarrow true$;
while ok **do**
  $ok \leftarrow false$;
  forall $(i, j) \in \mathcal{N}_\mathcal{S}$ *such that* $v(\{i, j\}) < 1$ **do**
    $i' = \arg\min_{k/(i,k)\in\mathcal{N}_\mathcal{S}} v(\{i, k\})$;
    $j' = \arg\min_{k/(j,k)\in\mathcal{N}_\mathcal{S}} v(\{j, k\})$;
    if $i' = j$ *and* $j' = i$ **then**
      $ok \leftarrow true$;
      Derive $m^\Omega_{i\textcircled{\cap}j}$ and $m^\mathcal{H}_{i\textcircled{\cap}j}$ from conjunctive combinations of i and j bbas;
      Update $\mathcal{G}$ nodes: one removal and one update with $\left\{m^\Omega_{i\textcircled{\cap}j}, m^\mathcal{H}_{i\textcircled{\cap}j}\right\}$;
      Update $\mathcal{G}$ links and cost values;
    end
  end
end
$\mathcal{S}' \leftarrow \mathcal{G}$ nodes;

The third line of Figure 2 illustrates the merging process. Figures 2h-2j show the three first iterations of the mutual agreement algorithm, with centroids of the pairs of merged segments at the current iteration in red and at previous iterations in pink. Figure 2k shows the final result with labeled segments.

3.4. Global algorithm

Algorithm 4 summarizes the proposed split and merge class-labeled segmentation. The evidential framework is used through the entire approach. Belief functions allow us to monitor information pieces and their uncertainty during the different steps and to transfer the already processed information to following processes. The influence of the different parameters of the algorithm is studied in next section.

4. Experimental Results

This section presents experimental results obtained on real scenes from our application, namely an automatic object classification for stacked manufactured objects. It

Algorithm 4: Global algorithm; inputs: data sets $\{I_k\}_{k\in\mathcal{K}}$, class set Ω, height interval $\mathcal{H}$; parameters: n_S number of superpixels, bba allocation parameters and conflict threshold parameters; outputs: object-level segmentation $\mathcal{S}^o$ with segments labeled in Ω: $\{l_s\}_{s\in\mathcal{S}}$.

Initial segmentation $\mathcal{S}^h \leftarrow$ output of the used superpixel algorithm having as inputs $\mathcal{K}'$ data and n_S;
forall $s \in \mathcal{S}^h$ **do**
    Consensual labeling of s according to Alg. 1 having s and associated spectral data as inputs and providing output $(m_s^\Omega, \mathcal{E}_s, ok)$;
    if $ok = false$ **then**
        Split s according to Alg. 2 having s and $\mathcal{E}_s$ as inputs and providing output s subregion set $\left\{s_j\right\}$;
        $\mathcal{S}^h \leftarrow \mathcal{S}^h \cup \left\{s_j\right\} \setminus \{s\}$;
    else
        Compute $m_s^\mathcal{H}$ the bba of s on $\mathcal{H}$;
    end
end
$\mathcal{S}^o$ and $\left\{m_s^\Omega\right\}_{s\in\mathcal{S}^o} \leftarrow$ output of Alg. 3 having $\mathcal{S}^h$ (and its adjacency graph) and bba sets $\{m_s^\Omega\}_{s\in\mathcal{S}^h}$ and $\{m_s^\mathcal{H}\}_{s\in\mathcal{S}^h}$ as inputs;
forall $s \in \mathcal{S}^o$ **do**
    Derive pignistic probability from m_s^Ω (according to Eq. (5));
    $l_s \leftarrow \arg\min_{l\in\Omega} BetP_s^\Omega(l)$;
end

allows us to evaluate the ability of our approach to address the analysis of a mixture of overlapping objects. Faced with the diversity of the objects and the materials, a multi-sensor device and sensor fusion seem necessary. For instance, in [18], the authors stress the necessity to introduce intelligence in the system to overcome the sensor limitations, e.g., using deep learning or imprecision modeling such as with belief functions.

The considered data have been obtained in Veolia laboratories using the three mentioned sensors: a NIR spectrometer, a 3D sensor (laser profilometer that provides both brightness and height images) and a color camera. The sensors have different resolutions, frame rates and fields of view so that spatial and temporal matching of the different data can only be imprecise. However, a rough registration has been performed among the four types of data (namely color, brightness, height and spectral data) in order to calibrate the functions providing, for each sample, the approximate position in a common frame of reference.

4.1. Experimental settings

The initialization of the segmentation and the bba allocations (on material class and height discernment frames) are prerequisite functionalities of the proposed approach. In this section, we specify them for our experiments.

Superpixel segmentation. Given that our split step based on *pseudo-image* samples can only add linear borders (i.e., spatial rough approximation of a missing edge), it seems preferable to limit the number of false negative edges in the initial segmentation while controlling the number of false positive edges. Then, from $\mathcal{K}'$ data, as $\mathcal{S}$ initialization, we carry out an over-segmentation. Matching this goal, several superpixel algorithms (e.g. [14, 1, 16]) allow us to control the mean size of the segments as well as the adherence to the object boundaries versus the contour regularity.

In this work, we considered the waterpixel algorithm [16] applied to a weighted combination of gradient images from $\mathcal{K}'$ data set. To take into account the fact that among the $\mathcal{K}'$ images, only the height images also belong to $\mathcal{K}''$, i.e. they do not generate false positive edges, their default weight value was set greater than those of brightness and color image weights. Finally, the driving parameter for this initial segmentation is the number of superpixels (that determines the approximate size of the superpixels, i.e. initial regions).

Labeling bba allocation. Labeling was performed using spectral data. Their inherent high dimensionality (spectra of several hundreds of wavelengths) justifies that they are commonly processed with learning-based classifiers such as Support Vector Machines (SVM). We focused on SVM as the best compromise between high efficiency in processing high dimension data and low effort required for their construction [8]. Faced with a multiclass problem (the set of classes Ω such that $|\Omega| > 2$) whereas SVM are primarily binary classifiers, we considered *Error Correcting Output Codes* (ECOC), which consist in splitting the multiclass problem in several simpler ones corresponding to simple classifiers whose individual responses are combined to estimate the class solution in Ω.

In order to handle both the uncertainty and the imprecision during the classification, we recently proposed an evidential ECOC approach, [12]. Here we are interested in the ECOC decoding part of this work. In brief, this part works as follows. An ECOC can be viewed as a set of l dichotomizers (i.e., binary classifiers, SVM in our case) for which the two considered hypotheses are non-overlapping subsets of Ω. Denoting these two hypotheses $A_{i,0}$ and $A_{i,1}$ for the i^{th} SVM, its output score for a given sample p is converted (see [12] for details) into an elementary bba m_i^b having $A_{i,0}$, $A_{i,1}$ and $A_{i,0} \cup A_{i,1}$ as focal elements. For non-dense ECOC, a deconditioning on $\overline{A_{i,0} \cup A_{i,1}}$ allows us to model the absence of information on $\overline{A_{i,0} \cup A_{i,1}}$ provided by this i^{th} dichotomizer. Then, the sample bba m_p^Ω is the conjunctive combination (Eq. (1)) of elementary bbas: $m_p^\Omega = \cap_{i=1}^{l} m_i^b$. The value $m_p^\Omega(\emptyset)$ represents how much the elementary classifiers agreed with each other. This justifies its use in the prefiltering step in Algorithm 1, which discards anomalous samples, such as mixed samples or outliers (e.g., due to shadow effect near object edges).

For experimental tests, the used ECOC was the one-vs-all. It is one of the simplest ECOC, relying on a limited number of dichotomizers ($|\Omega|$), which makes it well adapted for high number of classes, even if the learning step has to be carefully done [20] in presence of unbalanced hypotheses.

The considered set of classes (Ω) included 11 classes among which different sorts of plastic and paper along with metal and the background. These classes have been

defined according to the materials we aimed to separate in our application. Introducing a new class can be simply achieved by modifying the ECOC and training the additional necessary dichotomizers.

Height bba allocation. Bba allocation for the height feature (at region level) should represent both the absolute value of the height of the region and the diversity of height values. The discernment frame is the real number interval of any possible height values: $\mathcal{H} \subset \mathbb{R}$. On $\mathcal{H}$, we represented the height belief through a consonant bba: for each region s, the focal elements of the bba $m_s^{\mathcal{H}}$ are nested intervals defined from the statistical distribution of the height values on s. Specifically, interval bounds correspond to some percentiles computed on the s histogram of height values: denoting by $n_{\mathcal{F}^{\mathcal{H}}}$ the fixed number of focal elements ($n_{\mathcal{F}^{\mathcal{H}}} \in \{2, 3\}$), and h_{α_i} the α_i-percentiles, with $100 \geq \alpha_i \geq \alpha_{i+1} > 0, \forall i \in \{1, \ldots, n_{\mathcal{F}^{\mathcal{H}}} - 1\}$, the initial focal elements are the real intervals $[h_{\alpha_i}, h_{100-\alpha_i}]$.

In our experiments, $n_{\mathcal{F}^{\mathcal{H}}} \in \{2, 3\}$, $\alpha_1 = 30$, and $\alpha_2 \in \{0, 10, 20\}$ corresponding respectively, to a 'large' interval (minimum and maximum on the considered region), to a 'medium' one and to a 'narrow' one. Initial masses of those two focal elements were equal to 0.3 and 0.7, respectively.

To take into account the local reliability of the 3D sensor, $m_s^{\mathcal{H}}$ may be discounted. Specifically, brightness information allows us to detect some of the *transparent* objects, for which the 3D information is totally unreliable. In such cases, $m_s^{\mathcal{H}}(\mathcal{H}) = 1$ so that the region height will not be considered further. In a more general way, for a discounting of parameter α ($\alpha \in [0, 1]$), $m_s^{\mathcal{H}}([h_{30}^s, h_{70}^s]) = (1-\alpha)\times 0.3$, $m_s^{\mathcal{H}}([h_{\alpha_2}^s, h_{1-\alpha_2}^s]) = (1-\alpha)\times 0.7$ and $m_s^{\mathcal{H}}(\mathcal{H}) = \alpha$.

4.2. Results

4.2.1. Qualitative analysis

In this section, we show some typical result examples obtained on scenes acquired on VEOLIA experimental device bench. Figure 3 presents a first set of results. Each line corresponds to a different scene. On this figure, the obtained results are globally satisfying. Specifically, per scene, the main features are:

- Line 1: Scene with several bottles made up of the same material; their separation is made possible due to their height feature.
- Line 2: The top bottle among the milk bottles is not made up of the same material as the ones below, and only spectral analysis allows for correct material labeling. Note also that the cap of the top bottle being of the same material as the bottles below, the algorithm generated a right split, even if it failed to estimate the spatial shape of the obtained segment.
- Line 3: Scene with four objects, among them a transparent bottle that is correctly detected due to the brightness information; however, brightness measures being noisy, this entails spatial imprecision, including for the fibrous rectangular box.

- Line 4: Fibrous material and can on top of transparent bottles are correctly classified and detected due to the simultaneous use of brightness and spectral data for transparent bottles.

- Line 5: Scene with different parts of deformed objects: e.g., the plastic label has a non-standard shape that does not correspond to classic object databases; note that the merge step allows for the correct recovering of the object global shape despite the color inhomogeneity.

- Line 6: Complex scene with two transparent bottles (same material but presenting different appearance) and color-mixed fibrous objects that are separated due to their height feature; note however a small segmentation error due to the fact that no split can correct a missed border between two objects of the same material and height.

Figure 4 presents a second set of results in order to show some limits of our approach even if results are still globally satisfying:

- Line 1: Scene with overlapping flat objects (among others): the two paper sheets made of the same material and roughly at the same height cannot be distinguished.

- Line 2: object heterogeneousness in terms of height; conversely to the previous scene, here we present a case of an object (the bottle partially crushed) with subparts presenting different heights that cannot be merged.

- Lines 3: Transparent bottle on fibrous materials. The transparent material is correctly detected and classified using the hyperspectral sensor coupled with the brightness information. However, the bottle creates a shadow that prevents the correct classification of a subpart of the fibrous object below and consequently the object segmentation; a possible solution would be to learn such spectra.

- Line 4: One can notice that on this second set of results, the obtained segments are more imprecise than in the case of Figure 3; this is particularly true with this scene even if there is no classification error nor glaring segmentation error.

- Line 5: A rather complex scene with one transparent bottle, color-mixed fibrous objects that are separated due to their height feature, and a dark plastic sheet (top, middle-left of the scene); the dark plastic is missed by the proposed approach (almost invisible for every data); besides, we also can notice a typical case of labeling error between different fibrous classes (which are generally rather difficult to distinguish).

Figure 5 presents two very complex scenes with 8 and 14 objects, respectively. The flat fibrous objects underline the necessary compromise when fitting the algorithm parameters, which results in an inability to distinguish some fibrous objects located at the same height in the first scene at the same time as a fragmentation of the paper sheet in the second scene. This second scene also illustrates the problem of heterogeneous

Figure 3: First set of results: each line corresponds to a different scene, with RGB image (left column) and labeled object-segmentation result with colored labels overlaid on the gray-level intensity image (right column).

Figure 4: Examples of results underlining some limits of the proposed approach and presenting, for each scene (line): the RGB image (left column) and the labeled object-segmentation result with colored labels overlaid on the gray-level intensity image (right column).

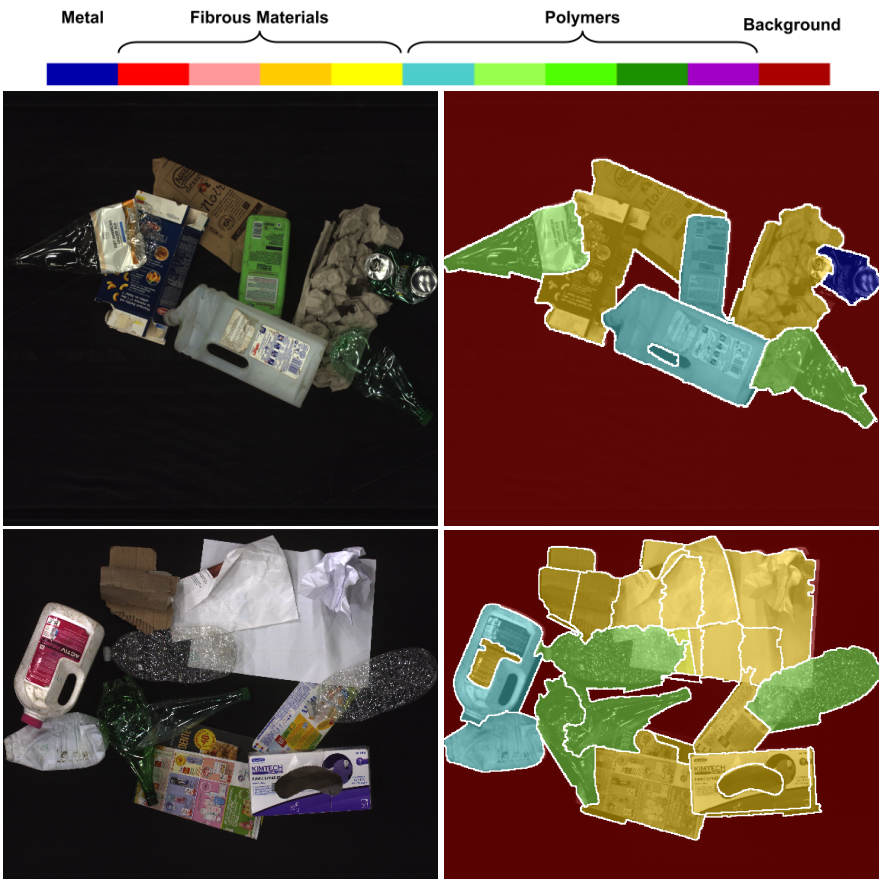


Figure 5: Example of results on two complex scenes: RGB image (left column) and labeled object-segmentation result with colored labels overlaid on the gray-level intensity image (right column).

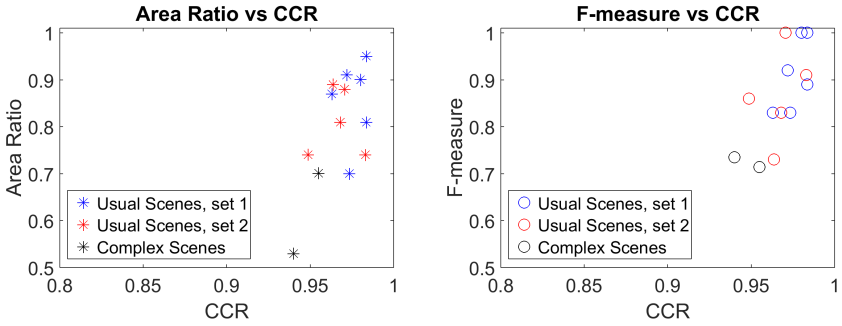


Figure 6: Quantitative evaluation: For each of the three scene sets with increasing complexity, Area ratio criterion versus CCR in % (left) and F-measure criterion versus CCR (right).

objects such as the polymer bottle having a fibrous sticker. However, despite these segmentation errors on few objects, once more we notice that segment labeling is excellent and that most of the objects are at least roughly detected.

4.2.2. Quantitative analysis

Since our problem was double (object segmentation and labeling), different specific criteria are used for quantitative evaluation. Firstly, the labeling results were evaluated at pixel level in terms of correct classification rate *CCR* that is the sum of correctly classified pixels relative to the total number of pixels (in the common frame of reference). Secondly, the segmentation was evaluated at object level in terms of precision and recall measures from which their harmonic average (F-measure) was derived. For this, each segment was associated with the object with which it had the largest intersection (in terms of pixel number), and then: for any object associated with i segments ($i \geq 0$), we counted $min\{i, 1\}$ true positive, $max\{i-1, 0\}$ false positive(s) and $1 - min\{i, 1\}$ (i.e. $max\{1-i, 0\}$) false negative. Specifically, denoting by n_{TP}, n_{FP} and n_{FN} the total numbers of true positives, false positives and false negatives on the scene, precision $Prec = \frac{n_{TP}}{n_{TP}+n_{FP}}$, recall $Rec = \frac{n_{TP}}{n_{TP}+n_{FN}}$ and F-measure is equal to $2 \times \frac{Prec \times Rec}{Prec+Rec}$. Thirdly, in order to evaluate the shape of the chosen segments we also computed the area ratio (*AR*) that is the ratio of the areas in numbers of pixels between the intersection and the union of every pair of associated segment and object. Note that *CCR* performance at pixel level also provides an indirect evaluation of segmentation accuracy since it is penalized by coarse segments.

Figure 6 shows the *AR* and F-measure criteria versus the *CCR* one for the three previous example sets. The numerical values corroborate the previous qualitative comments about slight decrease of performance with increasing scene complexity. We also notice that the *CCR* values are very high (greater than 94% for all scenes). This is due to the efficiency of hyperspectral data for material classification, even if fibrous classes remain the most difficult to distinguish. Then, *CCR* is mainly impacted by the segment spatial imprecision versus the pixel level classification evaluation. Concerning object detection, on the one hand, recall values are very good (average value over all scenes equal to 0.95), which means that there is almost no improper merge (but the

mentioned case of same material/same height objects, typically flat fibrous). Precision values, on the other hand, are impacted by object fragmentation that is more frequent, if only because of the cases of partial occlusion inducing multiple components for a single object. In the same way, the overlapping area criterion is penalized both by false positives (since, when an object is divided, only the main segment is associated to the object, and the intersecting area with the ground truth object is decreased) and false negatives (since when a segment overlaps two objects, the union area is all the more important). In particular, the second complex scene has an approximate 50% *AR* value (lowest point), which means that object shape and surface are only half recovered. However, as visible on Figure 5 second line, such a value derives from the average between uneven *AR* values, namely quite good values and quite low values such as for the white paper sheet or two among the three transparent bottles in the center left (that are in fact almost distinguished except for a thin junction).

4.2.3. Discussion on algorithm parameters

The proposed algorithm involves different parameters, among them some threshold parameters and the parameters of initial segmentation, namely number of regions and used dataset.

Threshold parameters appear in Algorithm 1 (conflict and imprecision thresholds, maximal number of iterations) and in Algorithm 3 (conflict thresholds). They drive the split process and the merge one, respectively. The parameter sensitivity study showed however that they have less impact than the parameters driving the initial segmentation. For instance, varying conflict threshold τ_S and number of samples M in Algorithm 1, (τ_S, M) (Alg. 1) in $\{0.1, 0.2, 0.4, 0.6\} \times \{5, 8, 10\}$, CCR values varied in $[94.1, 95.1]$ for the first complex scene (cf. Figure 5). However, their effect is not null: when increasing τ_S, the probability to accept (as reliable) a drawn subset of samples increases too so that less split operations will be required. Conversely, the number of performed splits increases with the number of samples to draw M. This comes from two phenomena: firstly, since the conflict increases with each conjunctive combination between samples, it can only increase with M; secondly, the chance to draw an actually conflicting sample (e.g., outlier, noisy sample) increases with M.

Regarding merging step parameter thresholds, the situation is even more favorable (in terms of robustness), since the mutual agreement mechanism is dominant in the choice of the segments to merge. This allows for improved robustness relatively to $\tau_\kappa^{\mathcal{H}}$ that in practice is chosen close to 1. Finally, for τ_κ^{Ω}, 0.2 has been set as a default value. Note however that it depends on the intrinsic conflict of the material class bba, that may vary with the used ECOC or more generally with the bba allocation (cf. Section 4.1) and with the consensual labeling step parameters.

Initial segmentation number of regions drives the initial object fragmentation. Using waterpixels, it directly controls the rough size of the superpixels that has a great influence on the final result. Table 2 sums up the impact of the growing size of the superpixels on several phenomena that impact the quality of the result.

The phenomena in favor of big superpixels are: (i) the robustness to outliers that, for a fixed number of outliers, increases with the number of total samples, (ii) the meaningfulness of the height allocation since bigger segments are more representative of the objects, and (iii) the number of required merges needed to recover the whole

Table 2: Overview of the impact of the growing superpixel size on different algorithm phenomena.

Phenomena	**Growing size of the superpixels** ↗
Robustness to outliers	☹ ↗ ☺
Detection of mixed segments	☺ ↘ ☹
Meaningfulness of a linear inner border	☺ ↘ ☹
Meaningfulness of the height information with respect to the object	☹ ↗ ☺
Number of straddling segments	☺ ↗ ☹
Number of merges needed	☹ ↗ ☺

object. Indeed, since the conflict grows with the number of merges, if this number is too large, then the natural increase of the conflict may make the algorithm stop prematurely.

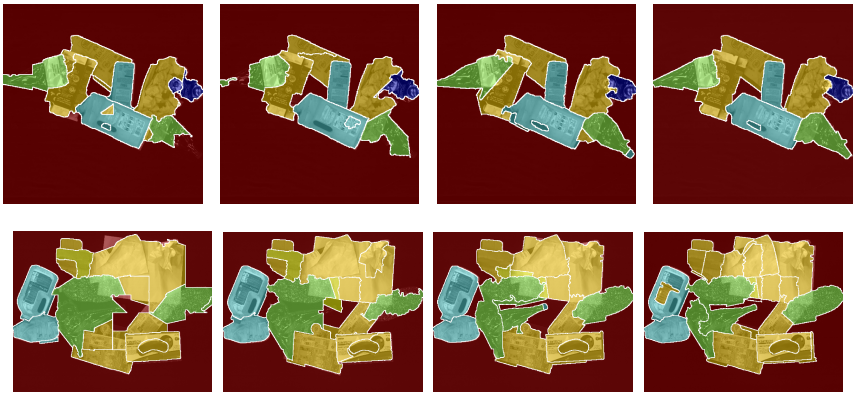
The phenomena in favor of small superpixels are: (i) the absence of straddling segments, (ii) the approximation of a missing border by a linear frontier (that is all the more relevant that this frontier is not too long), and (iii) the fact that mixed segments are easier to detect. Indeed, in case of two classes A and B present in a given segment, we expect that the pixels of class B be detected as outliers. Thus, in the same way that big segments are more robust to outliers, they also are less prone to detect the presence of class B.

Dataset $\mathcal{K}'$ can also be an adjustment variable whose *optimal* value depends on the object discriminant features and available sensors. To illustrate its impact, we considered initial segmentations (superpixels) obtained from only the 3D data, or also the brightness and/or the color information for both scenes. Figure 7 presents some result examples obtained on the two complex scenes (cf. Figure 5).

First of all, our evidential split-and-merge algorithm provided expected results regardless of the segmentation. In particular, the separations within the stack of objects are performed, even between objects of the same material.

However, the initial superpixels have a noticeable impact on the quality of the final object borders as well as on the number of segment(s) representing them at the end. Indeed, according to Table 3, we notice that the CCR values are very stable and high. This is due to the fact that all the segments are correctly labeled and that errors are only due to some border imprecision or missed object subpart. The F-measure is also rather stable but achieves much lower values. This is due to the presence of a few false negatives (indistinct objects in $\mathcal{S}^o$) such as with the three transparent bottles adjacent in the second scene or to the presence of a few false positives (objects divided in several regions in $\mathcal{S}^o$) such as with the paper sheet. For this last case, we note that the merge process stops due to the flatness of the object that induces narrow focal elements for the height bbas. The overlapping area criterion exhibits the greatest variation versus $\mathcal{K}'$ in complex scene 1 case.

Analyzing the evolution of the performance versus the $\mathcal{K}'$ dataset, we note that:



(a) 3D (b) 3D-RGB (c) 3D-Brightness (d) 3D-Brightness-RGB

Figure 7: Result examples using different datasets for the initial segmentation; case of the two complex scenes.

- Even with a segmentation based only on the 3D data, performance measures are relatively correct thanks notably to the splitting step that can retrieve some of the missing borders.
- Using additional data in order to estimate a better initial segmentation can be more efficient in terms of segmentation and classification evaluation.
- Rather unexpectedly but fortunately, the couple (3D, brightness data) appears very interesting since it allows for results competitive or possibly even better than using the three images (3D, RGB, Brightness).
- However, the most helpful data varies with the considered scene since these additional sources of information have their own strengths (detection of transparent of flat materials) and flaws (creation of useless borders in multi-material objects). Then, most of the limitations of the algorithm underlined in Figure 7 are related to specific configurations for which the sensors we considered did not provide discriminative enough features. Irrespective of the split-and-merge algorithm, each set $\mathcal{K}'$ being considered for a range of input objects is bound to exhibit difficulties with certain configurations in a similar way.

5. Conclusion

In this work, we have proposed and studied the benefits of the belief function framework for an object oriented classification. The proposed algorithm was applied to the double problem of segmentation and material classification.

Using belief functions allows for an easy modeling of the interactions between the classification and the segmentation modules involved in object-oriented methods. In this work, those interactions consist in labeling, splitting and merging and involve several levels of fusion. The information propagation from one level to another is ensured

Table 3: Performances of the algorithm with regard to the used data to initialize the segmentation. The results are given in terms of F-measure (F), pixel level CCR and Area Ratio (AR).

$\mathcal{K}'$	**Only 3D**			**3D and RGB**			**3D and Bright**			**3D-RGB-Brightness**		
	F	CCR	AR	F	CCR	AR	F	CCR	AR	F	CCR	AR
Complex scene 1	$76,9$	$94,8$	$68,8$	$76,9$	$95,1$	$54,3$	$74,1$	$95,6$	$74,6$	$71,4$	$95,5$	$70,1$
Complex scene 2	$72,0$	$92,1$	$51,8$	$76,6$	$93,3$	$51,0$	$75,0$	$93,4$	$56,1$	$73,5$	$94,0$	$53,0$

by the evidential combination. The conflict measure is used to detect disagreements, firstly between the classification samples, and secondly between the segments. In this sense, it drives the interactions between the modules. Eventually, we show that the full approach allows for a relevant segmentation in the perspective of object identification, even in the case of missing data (subsampling).

Indeed, a main feature of our approach is that it considers heterogeneous data, namely images and pseudo-images, and get them involved at the right step(s) of the processing: 3D sensor data and RGB image in the initial segmentation, spectral data in the classification, both spectral data and 3D ones in the final segmentation. Then, inner limits include the fact that objects of similar height and similar material have little chance to be separated using the current method without adding new information. Multi-label objects as well as objects composed of parts of different heights cannot be assembled. Finally, the proposed approach contains several parameters that allow for the customization of the algorithm to the considered data and application.

In the following work, we may investigate the use of supplementary information such as the hue or the brightness to steer the merging process and induce more relevant fusions. In a more general way, we aim to include association rules in the approach derived from the application in order to tackle the problem of multi-material objects for example. Those rules will be formulated in the belief framework using prior bba and contextual discounting operator. In the same way, partial occlusions of an object that make it appear split in two distant subparts (case of stacked objects), cannot be solved without the consideration of supplementary prior information (for instance about the expected shape of the object). We will then formulate this problem as an association problem in the belief function framework, which will allow us to benefit from evidential association algorithms already proposed, e.g. [7].

References

[1] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282.

[2] Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16.

[3] Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R. Q., van der Meer, F., van der Werff, H., van Coillie, F., et al. (2014). Geographic object-based image analysis–towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:180–191.

[4] Bolles, R. C. and Fischler, M. A. (1981). A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In *IJCAI*, volume 1981, pages 637–643.

[5] Capelle, A.-S., Colot, O., and Fernandez-Maloigne, C. (2004). Evidential segmentation scheme of multi-echo mr images for the detection of brain tumors using neighborhood information. *Information Fusion*, 5(3):203–216.

[6] Denœux, T. (1999). Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning*, 20(1):79–111.

[7] Denoeux, T., El Zoghby, N., Cherfaoui, V., and Jouglet, A. (2014). Optimal object association in the Dempster–Shafer framework. *IEEE transactions on cybernetics*, 44(12):2521–2531.

[8] Gu, Y., Chanussot, J., Jia, X., and Benediktsson, J. A. (2017). Multiple kernel learning for hyperspectral image classification: A review. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6547–6565.

[9] Hay, G. J. and Castilla, G. (2008). Geographic object-based image analysis (geobia): A new name for a new discipline. In *Object-based image analysis*, pages 75–89. Springer.

[10] Hoiem, D., Efros, A. A., and Hebert, M. (2007). Recovering surface layout from an image. *International Journal of Computer Vision*, 75(1):151–172.

[11] Kim, M., Warner, T. A., Madden, M., and Atkinson, D. S. (2011). Multi-scale geobia with very high spatial resolution digital aerial imagery: scale, texture and image objects. *International Journal of Remote Sensing*, 32(10):2825–2850.

[12] Lachaize, M., Le Hégarat-Mascle, S., Aldea, E., Maitrot, A., and Reynaud, R. (2018). Evidential framework for the error correcting codes method. *Engineering Applications of Artificial Intellligence*, 73:10–21.

[13] Lefevre, E., Colot, O., and Vannoorenberghe, P. (2002). Belief function combination and conflict management. *Information Fusion*, 3(2):149–162.

[14] Levinshtein, A., Stere, A., Kutulakos, K. N., Fleet, D. J., Dickinson, S. J., and Siddiqi, K. (2009). Turbopixels: Fast superpixels using geometric flows. *IEEE transactions on pattern analysis and machine intelligence*, 31(12):2290–2297.

[15] Liu, Z.-G., Pan, Q., and Dezert, J. (2014). A belief classification rule for imprecise data. *Applied Intelligence*, 40(2):214–228.

[16] Machairas, V., Faessel, M., Cárdenas-Peña, D., Chabardes, T., Walter, T., and Decencière, E. (2015). Waterpixels. *IEEE Transactions on Image Processing*, 24(11):3707–3716.

[17] Minary, P., Pichon, F., Mercier, D., Lefevre, E., and Droit, B. (2017). Face pixel detection using evidential calibration and fusion. *International Journal of Approximate Reasoning*, 91:202–215.

[18] Rahman, M., Hussain, A., and Basri, H. (2014). A critical review on waste paper sorting techniques. *International Journal of Environmental Science and Technology*, 11(2):551–564.

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

[20] Rifkin, R. and Klautau, A. (2004). In defense of one-vs-all classification. *Journal of Machine Learning Research*, 5(Jan):101–141.

[21] Ristic, B. and Smets, P. (2006). The TBM global distance measure for the association of uncertain combat id declarations. *Information fusion*, 7(3):276–284.

[22] Shafer, G. (1976). *A Mathematical Theory of Evidence*, volume 1. Princeton university press Princeton.

[23] Smets, P. and Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, 66(2):191–234.

[24] Tabassian, M., Ghaderi, R., and Ebrahimpour, R. (2012). Combining complementary information sources in the Dempster–Shafer framework for solving classification problems with imperfect labels. *Knowledge-Based Systems*, 27:92–102.

[25] Walley, P. (1991). Statistical reasoning with imprecise probabilities.

[26] Xu, P., Davoine, F., Bordes, J.-B., Zhao, H., and Denœux, T. (2016). Multimodal information fusion for urban scene understanding. *Machine Vision and Applications*, 27(3):331–349.

[27] Yager, R. R. (1987). On the Dempster–Shafer framework and new combination rules. *Information sciences*, 41(2):93–137.