SVM Classifier Fusion Using Belief Functions: Application to Hyperspectral Data Classification

Marie Lachaize[1,2], Sylvie Le Hégarat-Mascle[1(✉)], Emanuel Aldea[1],
Aude Maitrot[2], and Roger Reynaud[1]

[1] SATIE, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France
{marie.lachaize,sylvie.le-hegarat,emanuel.aldea,roger.reynaud}@u-psud.fr
[2] Veolia Recherche & Innovation, 291 Av. Dreyfous Ducas, Limay, France
{marie.lachaize,aude.maitrot}@veolia.com

Abstract. Hyperspectral imagery is a powerful source of information for recognition problems in a variety of fields. However, the resulting data volume is a challenge for classification methods especially considering industrial context requirements. Support Vector Machines (SVMs), commonly used classifiers for hyperspectral data, are originally suited for binary problems. Basing our study on [12] bbas allocation for binary classifiers, we investigate different strategies to combine two-class SVMs and tackle the multiclass problem. We evaluate the use of belief functions regarding the matter of SVM fusion with hyperspectral data for a waste sorting industrial application. We specifically highlight two possible ways of building a fast multi-class classifier using the belief functions framework that takes into account the process uncertainties and can use different information sources such as complementary spectra features.

Keywords: Hyperspectral classification · Belief Function Theory · Waste sorting

1 Introduction

Hyperspectral imaging acquires an almost continuous sampling of the spectral data in each pixel of a scene. The large number of narrow and contiguous bands allows us to detect even minor variations in a spectrum and provides a significant advantage for the distinction of different materials (natural or artificial). Hyperspectral imaging popularity has thus increased during the last decade, due both to the increase in computing power (that allows us to process these voluminous data) and to the need of separation of increasingly specific classes, e.g. different kinds of polymers for industrial application or different mineralogical compositions of surfaces for remote sensing applications. The acquisition of numerous wavebands enables the identification of fine classes belonging to various fields: atmosphere analysis, ecosystems monitoring, military applications and industrial applications (e.g. [5]).

Recently, hyperspectral imaging industrial applications appeared in waste sorting field. This latter needs new automated processes to improve the working

conditions and decrease the cost of recycled materials. This requires sensors able to discriminate among waste materials as close (in terms of spectral response) as different kinds of plastics or different fibrous materials. Then, some efficient classification methods should be defined to use the full potential of hyperspectral data in this industrial context: methods adapted to waste material classes, robust to significant object occlusion (objects stacked on one another), and fast enough to match the waste flow.

Hyperspectral imagery provides detailed spectral information for each pixel of the picture. This information richness comes with an obvious drawback: the huge amount of data to process. Indeed, it involves significant computing resources (time, memory...) that may be an issue in an industrial context. In addition to spectral information, several works also propose to include spatial information in their classification process, e.g. [13]. However, in this study, we focus on blind classification, i.e. classification of each pixel independently (not taking into account the neighbouring pixels that provide the spatial information). For our industrial application, it is a first step towards a classification process at object level which couples blind classification and spatial segmentation using others sensors.

Support vector machines (SVMs) are classically used to meet the challenge of hyperspectral classification, see [7,8,10] for instance. Basically, they project data into a higher dimensional space in which classes can be separated using (optimal) hyperplanes. SVMs are widely used in a probabilistic framework (e.g. through logistic regression). However, probabilities are not able to distinguish between uncertainty (in decision process) due to ambiguity, e.g. because of overlapping between classes, and uncertainty due to imprecision, e.g. because of low number of samples for some score values during the SVM training and calibration processes. Thus, [12] has extended several classical regressions used for SVM binary classifiers to the Belief Function Theory framework.

As SVMs are well suited for 2-class problems, several strategies have been proposed to address the multiclass case. Classical approaches are: (i) the *one-versus-one* strategy, where a classifier is trained on each class pair and (ii) the *one-versus-all* strategy, where a classifier is trained to distinguish one class against all the others. For each strategy, the merging of these 2-class classifiers is an important issue. In this study, we propose and evaluate three strategies to combine binary SVM outputs in the BFT framework. We also show that the use of belief functions allows us to combine the information contained into different input data derived from the whole spectrum data, and that it is an efficient way to deal with the complexity involved by the high dimensionality of the hyperspectral data.

2   Strategies from Binary to Multiclass Belief Functions

Addressing a multiclass problem from binary SVMs (*one-versus-one* or *one-versus-all*), we have to combine their binary decisions while taking into account the imprecision of each SVM (classes overlapping, number of samples, . . .).

We choose to perform this combination in BFT framework considering SVM outputs as logical sources. The way an observed score generates a basic belief assignment (bba) is defined by the calibration step proposed by [12]. It is briefly recalled in Appendix. Then, in the following, we assume that each cluster of SVM produces as many bbas as there are SVMs in the cluster, and that each bba represents our belief in each of the two classes handled by the considered SVM as well as the ignorance left by this classifier.

Notation are as follows. For any *one-versus-one* SVM dealing with the pair of classes ω_j and ω_k, $\Omega_{j,k}^b$ denotes the associated discernment frame and $m^{\Omega_{j,k}^b}$ the bba derived from calibration. For any *one-versus-all* SVM dealing with the pair of classes ω_j and its complementary $\bar{\omega}_j$, Ω_j^b denotes the associated discernment frame and $m^{\Omega_j^b}$ the bba derived from calibration. In both cases, the upperscript b recalls that SVM was binary and the cardinality of $\Omega_{j,k}^b$ or Ω_j^b is 2. Then, the transition from binary classification to a multiclass one implies a change of discernment frame from $\Omega_{j,k}^b$ or Ω_j^b to $\Omega = \{\omega_1, \ldots, \omega_n\}$, the multiclass set.

2.1 Case of *one-versus-one* SVMs

In the *one-versus-one* strategy, $\Omega_{j,k}^b = \{\omega_j, \omega_k\}$ and $m^{\Omega_{j,k}^b}$ is interpreted as the result of a conditioning of a multiclass bba m^Ω on $\Omega_{j,k}^b$, noted $m^\Omega \left[\Omega_{j,k}^b\right]$. Then, from these conditioned estimations of the bba, [9] proposes to derive m^Ω by solving the optimization problem:

$$\min_{m^\Omega} \sum_{k>j} \sum_{\emptyset \neq A \subseteq \Omega} \left(m^\Omega \left[\Omega_{j,k}^b\right](A) - m^{\Omega_{j,k}^b}(A) \left(1 - m^\Omega \left[\Omega_{j,k}^b\right](\emptyset)\right)\right)^2, \qquad (1)$$

under the constraints: (i) $m^\Omega(A) \geq 0, \forall A \in 2^\Omega$, (ii) $m^\Omega(\emptyset) = 0$ and (iii) $\sum_{A \in 2^\Omega} m^\Omega(A) = 1$. In Eq. (1), the factor $1 - m^\Omega \left[\Omega_{j,k}^b\right](\emptyset)$ is due to the fact that $m^{\Omega_{j,k}^b}$ is a normal bba which may not be the case for $m^\Omega \left[\Omega_{j,k}^b\right]$. To solve such a constrained system, we have noticed that it can also be written as a matrix system $\min_{\mathbf{X}} |\mathbf{AX-B}|^2$.

In this study, we propose an alternative to derive m^Ω that consists in simply performing a deconditioning of every bba $m^{\Omega_{j,k}^b}$ on the frame of discernment Ω and then combining the deconditioned bbas $m^{\Omega_{j,k}^b \Uparrow \Omega}$ using a conjunctive rule: denoting $B = \bar{A}$ the complementary in Ω of hypothesis A,

$$m^\Omega(A) = \bigcirc_{(j,k) \in [1,n]^2} m^{\Omega_{j,k}^b \Uparrow \Omega}(A), \ \forall A \in 2^\Omega, \qquad (2)$$

with

$$m^{\Omega_{j,k}^b \Uparrow \Omega}\left(A \cup \bar{\Omega}_{j,k}\right) = m^{\Omega_{j,k}^b}(A), \ \forall A \in \{\omega_j, \omega_k, \{\omega_j, \omega_k\}\}. \qquad (3)$$

This second way is much simpler than the first one and it is theoretically founded on the independence between the binary SVM outputs. Now, it could provide an interesting approximation of the solution and even, it outperforms the optimization proposed by [9] in our application case. Therefore, on this latter at least, there are only advantages in using the proposed alternative.

2.2 Case of *one-versus-all* SVMs

In the *one-versus-all* strategy, the binary discernment frames Ω_j^b are some coarsenings of discernment frame Ω, or equivalently Ω is a common refinement of the different $\Omega_j^b = \{\omega_j, \bar{\omega}_j\}$. Then, the bbas defined on Ω_j^b are refined on Ω using classical refinement operator, before combination using Dempster's rule:

$$m^\Omega (A) = \oplus_{j\in[1,n]} m^{\Omega_j^b \uparrow \Omega} (A), \ \forall A \in 2^\Omega, \tag{4}$$

with

$$m^{\Omega_j^b \uparrow \Omega} (A) = m^{\Omega_j^b} (A), \ \forall A \in \{\omega_j, \bar{\omega}_j, \Omega\}. \tag{5}$$

Here, we use the orthogonal sum instead of the conjunctive combination rule since, due to the existence of singleton focal elements and to the high number of combinations (as much as the number of classes), conflict becomes very important even in standard cases.

2.3 Case of Hybrid Strategy

Each of the two previous strategies handles a given kind of binary SVMs, either *one-versus-one* or *one-versus-all*. Each of these strategies has some advantages: better separability of the classes for the *one-versus-one* SVMs, simplicity and speed for the *one-versus-all* SVMs. Then, in order to benefit from both advantages, we propose a hybrid strategy that handles SVMs belonging to both kinds of binary SVMs.

The proposed solution is based on some metaknowledge on the classes used to choose the considered SVMs. Let us first present the particular case of a hierarchical strategy through a toy example with only 4 classes: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, and assume ω_3 and ω_4 classes are difficult to separate. According to the proposed strategy, one will focus on most performing SVMs, i.e. SVMs 'ω_1 against $\{\omega_2, \omega_3, \omega_4\}$', '$\omega_2$ against $\{\omega_1, \omega_3, \omega_4\}$', '$\{\omega_3, \omega_4\}$ against $\{\omega_1, \omega_2\}$' and 'ω_3 against ω_4'. In a more general way, the classes of Ω are 'wisely' grouped to form a new coarsened discernment frame on which the *one-versus-all* approach performs well ('good' enough class separability, such as for $\{\omega_1, \omega_2, \{\omega_3, \omega_4\}\}$ in the example). Then, the considered classifiers are: the *one-versus-all* SVMs for classes corresponding to singleton hypotheses of the coarsened discernment frame (i.e. some of them are compound classes, e.g. hypothesis $\{\omega_3, \omega_4\}$ in the example) and the *one-versus-one* SVMs for classes belonging to a compound class of the coarsened discernment frame (e.g. SVM ω_3 versus ω_4 in the example). The bbas derived from outputs of previously cited SVMs that have been either refined or deconditioned on Ω are combined using conjunctive rule (or Dempster's orthogonal rule).

This hierarchical strategy appears as a compromise in terms of number of used (and thus trained) classifiers: if $n = |\Omega|$, with the pure *one-versus-one* strategy, we have to consider $\binom{n}{2} = \frac{n(n-1)}{2}$ classifiers, with the pure *one-versus-all* strategy, we have to consider n classifiers and with the hierarchical strategy involving l groups of n_i indistinguishable classes, we have to consider

$n - \sum_{i=1}^{l} (n_i - 1))$ *one-versus-all* classifiers and $\sum_{i=1}^{l} \frac{n_i(n_i-1)}{2}$ *one-versus-one* classifiers.

With respect to the hierarchical strategy, the hybrid strategy can in addition consider few other classifiers in order to increase the redundancy between classifiers and then provide more robust results. However, this number of additional classifiers should remain low to keep an interest in terms of complexity.

Finally, if the derivation of the metaknowledge is a subject beyond the scope of this article (here we simply assume that it can be either known *a priori* or learnt from training samples), let us underline that it is the crucial point for the proposed hybrid strategy since it allows us to add *prior* information (metaknowledge) that should be carefully chosen.

2.4   Decision

Independently of the strategy used, at the end, we have a bba resulting from the combination of the information pieces provided by different binary SVMs taking into account their own features (in particular the learning step conditions and results). From this bba, we can take a decision or, as seen in the experimental part, combine it with other bba(s) and then take the decision.

The belief function framework offers many possibilities for decision making [3]. Among them, pessimistic and optimistic strategies consist in maximizing, over the singleton hypotheses, the belief or plausibility function, respectively. A decision according to the pignistic probability provides intermediate results.

Here, specific to the *one-versus-one* strategy based optimization [9], we also test the following criterion: the decided singleton class ω_i maximizes the conflict generated by conditioning on binary subsets not including ω_i:

$$\omega_i = \arg\max_i \sum_{j,k \neq i} m^\Omega \left[\Omega_{j,k}^b \right] (\emptyset) . \tag{6}$$

3   Experiments

3.1   Industrial Context and Data Preprocessing

We applied the proposed classification strategies to waste sorting. Despite the use of some sorting components exploiting the mechanical properties of the waste materials in order to separate them, this industrial application has still several issues such as the automatic identification of some resembling materials or the detection of some 'intruders' in a set of similar wastes (e.g. paper, cardboard and plastic waste). For such purposes, the hyperspectral sensor appears relevant since it provides some information about the nature of the material itself that should help us to discriminate among different fibrous materials and polymers with high throughput and a reliability and robustness suitable for an industrial context.

As with any classification method, the performance of a SVM classifier strongly depends on the input data. Classical preprocessing on the spectrum

involves a filtering and derivation at different orders. Specifically, the Savitsky-Golay filter is widely used for hyperspectral data analysis [6,11]. This filter fits a low degree polynomial on data within a sliding window having fixed size. It allows us to smooth the data and to compute the derivatives from the fitted polynoms. The fact of considering different derivative orders (typically 0, 1 and 2) appears all the more justified since, for classification, not the whole spectrum is considered but only some selected features, in order to reduce both the data complexity and the correlation between the bands. Then, a classical way is to perform a Principal Component Analysis (PCA) on the filtered spectra, e.g. [1,2]. In our case, the number of selected components is set to represent 99 % of the information. It varies between 3 to about 20 whereas the whole spectrum dimensionality was about 275.

In summary, preprocessing involves the computation of different derivative orders (0, 1 and 2) of the spectrum by the Savitsky-Golay filter and then, for each of these derivatives, the computation of the PCA that provides the input data for the SVM classifiers. In the following, these input data are denoted S_0, S_1 and S_2 where subscript denotes the derivation order. The results obtained using these inputs will be compared in terms of classification performance. We also propose to use them as different logical sources so that, we combine the multiclass (defined on Ω) bbas derived for given input data (S_0, S_1 and S_2). Assuming that the PCA process makes input data cognitively independent, bba combination will be done through the conjunctive rule.

3.2   Experimental Results

The sample sets used for these experiments have been collected in the Veolia laboratories on a hyperspectral sensor (whose spectra contains about 275 wavelengths) via specimen boards with small material samples: four boards called *Paper*, *Plastic1*, *Plastic2a*, *Plastic2b*. Samples are divided in 9 classes, namely 7 polymers classes (not listed here for paper shortness) and 2 fibrous classes (paper and cardboard). From specimen boards, three different datasets were extracted. The first one, called training set, has 1000 samples per class and is used for SVM training. The second one, called calibration set, has 200 samples per class and is used for bba calibration. The last one has 1000 samples per class and is used for test and performance estimation. In addition to this test dataset, another board, exclusively used for testing and called *Superposition*, presents real objects stacked on top of each other to provide more realistic conditions.

Then, the training set allows for the estimation of each SVM classifier parameters, determined by cross validation and grid search, using Gaussian kernels. The calibration set allows for the estimation of sigmoid parameters and contour function defined for any score value (see Appendix for details). It also allows us to determine the classes to group for the hybrid strategy. Then, using the test set and the *Superposition* board, the first analysis (not presented here) puts forward some complementarity of classification performance for the input data S_1 and S_2, in particular for the 'difficult' pixels such as those present in the shadows or pixels corresponding to the superposition of two objects. The initial

Table 1. Correct classification rate (in %) for the 5 test sample boards (each one having 13500 pixels). Results are given for S_1 source, considering different strategies and decisions for the *one-versus-one* strategy.

| Strategy→	One-versus-one			One-versus-all	Hybrid
↓Boards	Score vote	Optimization	Deconditioning		
Paper	95.3	95.2	**95.5**	91.6	94.1
Plastic1	90.2	87.3	**91.4**	85.2	91.0
Plastic2a	79.1	83.7	**84.1**	77.5	79.5
Plastic2b	87.2	89.3	**89.7**	83.9	87.8
Superposition	82.8	88.0	**88.4**	79.5	80.1
Whole data	86.9	88.8	**89.8**	83.5	86.5

analysis also revealed that the S_0 input data provides results of little interest (low performance and lower complementarity) so that it has not been considered in the results presented further.

3.3   Comparison of the Different Strategies

The comparison of the different strategy is presented here in the case of S_1 data that prompts better results than S_0 or S_2. Considering S_1, the hybrid strategy was instantiated introducing two coarse classes: one grouping paper and cardboard and another grouping two classes of polymers (among the 7). A supplementary *one-versus-one* classifier is also considered to remove ambiguities between two other classes of polymers.

Classification results are analyzed in the perspective between comparison of (i) different multi-class strategies (*one-versus-one*, *one-versus-all*, hybrid), (ii) different decision making processes for the *one-vs-one* strategy. Quantitative results, computed on the whole datasets, are shown in Table 1. Our main conclusions are:

– In the case of the *one-versus-one* strategy, solving Eq. (1), the computation time increases dramatically with the number of classes (factor about 200 with the *one-vs-all* strategy computation time). Performing deconditioning (Eq. (2)) as proposed provides slightly better results for a much lower computation time.
– Comparison with classic decision rules, either score voting (shown in Table 1) or probabilistic decision (not shown) shows that the optimistic decision on bba obtained using either optimization or deconditioning (Eq. (2)) is better in every cases.
– The *one-vs-one* strategy always outperforms the *one-vs-all* strategy and the hybrid one, which is a standard result due to the better separability of the classes and much higher number of considered classifiers.
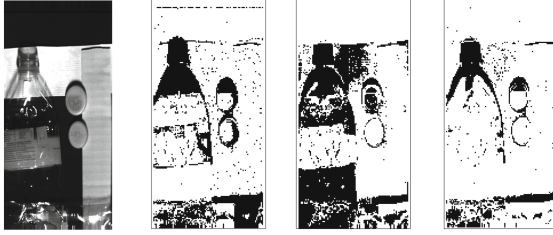
Fig. 1. Example of the fusion impact on classification rate (superposition board). From left to right: test board image, binary representations of the well classified pixels (in white) considering S_1, S_2 and their evidential fusion, respectively.

Table 2. Correct classification rate (in %) for the 5 test sample boards. Results are given for the $S_1\&S_2$ fusion, considering different strategies.

Boards→	Paper	Plastic1	Plastic2a	Plastic2b	Superposition	Whole data
One-versus-one	93.3	**93.4**	**82.2**	**87.9**	**89.9**	**89.3**
One-versus-all	92.3	91.8	80.1	87.1	87.3	87.7
Hybrid	**95.9**	92.2	78.4	87.3	86.1	88.0

– The hybrid strategy shows intermediate results between the *one-vs-one* strategy and the *one-vs-all* one: depending on the considered test dataset, the improvement relatively to the *one-vs-all* strategy varies between 0.6 % and 5.8 %.

3.4   Combination of Sources S_1 and S_2

In this subpart, classification results are analyzed versus different input data, namely both S_1 and S_2 (combined by fusion) or only S_1 (shown in Table 1). Quantitative results are shown on Table 2. Our main conclusions are:

– The fusion of S_1 and S_2 sources provides no or low improvement compared to the *one-vs-one* strategy, however the complementarity is very beneficial to the *one-versus-all* strategy and narrows the gap between the classification results of the two strategies (from 6.3 % to 1.5 % on the whole dataset).
– Fig. 1 presents the *Superposition* results illustrating that the fusion improves particularly the classification of the difficult pixels, such as at the top of the bottle and the caps stacked on paper.
– The hybrid strategy still provides intermediate results between *one-versus-one* strategy and *one-versus-all* one, but the interest relatively to the *one-versus-all* is reduced (relatively to the case of the only-S_1 data) due to the high improvement of performance of *one-versus-all* strategy provided by fusion.

4   Conclusion and Perspectives

This study has investigated the possibility in BFT to build a multiclass classifier which would be fast and efficient enough to be considered in the industrial

context of waste sorting. We compare different ways to achieve it: using the deconditioning operator on bbas derived from *one-vs-one* classifiers, using the refinement operator on bbas derived from *one-vs-all* classifiers or a using hybrid strategy. According to our tests, using *one-vs-one* leads to better performance but requires more classifiers to train than using *one-vs-all* classifiers. Using optimization [9] to derive the multiclass bba from bbas derived from *one-vs-one* classifiers can be advantageously replaced by the proposed deconditioning and conjunctive combination in terms of computation time. Hybrid strategy seems a good compromise between pure *one-vs-one* and pure *one-vs-all* strategies, presenting both reasonable number of classifiers and interesting performance results. Combining multiclass bbas associated to different features of the hyperspectral spectra (different orders of derivative) enhances the classification results in a noticeable way.

A main perspective to our work is the automatic derivation of the meta-knowledge on data used to build the hybrid strategy. We saw that the results provided by hybrid strategy are encouraging. However, the impact of the meta-knowledge has to be investigated and this also indirectly raises the question of the calibration quality of the binary classifiers. Then, we also intend to investigate some evidential criteria that will allow us to analyze training calibration sets, for instance to group automatically some classes and make the most of the hybrid strategy.

Appendix: Evidential Calibration

Handling binary classifiers, the discernment frame is $\Omega^b = \{\{0\}, \{1\}\}$. Then, for a given SVM having its own features in terms of number of samples, learning step performance, we aim at defining a belief function for each score that reflects the confidence we may have in each class. Indeed, this belief function will be used for forecasting taking into account the whole training set specificities. Explicitly, for each score s, the mass function, denoted m, is derived from the contour function: $\omega \rightarrow pl_X(\omega, s)$, where $\omega = P(y = 1 | s)$ (note that, ω is not a class but a probability).

To build the contour function on ω, [12] uses the idea behind the logistic regression: approximating the probability $P(y = 1 | s)$ by a sigmoid $sig_s(\theta) = [1 + exp(\theta_0 + \theta_1 s)]^{-1}$, where the parameter $\hat{\theta} = (\theta_0, \theta_1)$ is determined by maximizing the likelihood function $L_X(\theta)$ over the training set $X = \{(s_1, y_1), \ldots, (s_N, y_N)\}$ where, for each sample number i, $s_i \in \mathbb{R}$ is the score given by the considered classifier and $y_i \in \{0, 1\}$ is its true label. Then, the contour function of interest is drawn for a given value of score. It derives from the 2D function plotting the relative value of the likelihood function $\frac{L_X(\theta)}{L_X(\hat{\theta})}$ versus $\theta = (\theta_0, \theta_1)$. Then, for any given pair (s, ω), the set of θ (i.e. $sig_s^{-1}(\omega)$) values is a straight line in $\mathbb{R}^2$. Then, the contour function value can be determined as the maximum value over this straight line:

$$pl_X^{\Omega_{j,k}^b}(\omega) = \begin{cases} 0 & \text{if } \omega \in \{0, 1\}, \\ \sup_{sig_s^{-1}(\omega)} \frac{L_X(\theta)}{L_X(\hat{\theta})} & \text{otherwise,} \end{cases} \tag{7}$$

with $L_X(\theta) = \prod_{i=1}^{N} p_i^{y_i}(1 - p_i)^{1-y_i}$ where $p_i = \frac{1}{1+exp(\theta_0+\theta_1 s_i)}$. Finally, from each $pl_X^{\Omega_{j,k}^b}$, the corresponding mass function $m^{\Omega_{j,k}^b}$ on binary discernment frame $\Omega_{j,k}^b$ is derived using the 'likelihood based' belief function for statistical inference approach proposed by Shafer and further justified by Denœux [4].

References

1. Cavalli, R., Licciardi, G., Chanussot, J.: Archaeological structures using nonlinear principal component analysis applied to airborne hyperspectral image. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **6**(2), 659–669 (2013)
2. Chen, G., Qian, S.-E.: Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. IEEE Trans. Geosci. Remote Sens. **49**(3), 973–980 (2011)
3. Denoeux, T.: Analysis of evidence-theoretic decision rules for pattern classification. Pattern Recogn. **30**(7), 1095–1107 (1997)
4. Denœux, T.: Likelihood-based belief function: justification and some extensions to low-quality data. Int. J. Approx. Reason. **55**(7), 1535–1547 (2014)
5. Grahn, H., Geladi, P.: Techniques and Applications of Hyperspectral Image Analysis. Wiley, New York (2007)
6. King, R.L., Ruffin, C., LaMastus, F., Shaw, D.: The analysis of hyperspectral data using Savitzky-Golay filtering-practical issues. 2. In: Proceedings of IGARSS 1999, vol. 1, pp. 398–400. IEEE (1999)
7. Kuo, B.-C., Ho, H.-H., Li, C.-H., Hung, C.-C., Taur, J.-S.: A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7**(1), 317–326 (2014)
8. Melgani, F., Bruzzone, L.: Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. **42**(8), 1778–1790 (2004)
9. Quost, B., Denœux, T., Masson, M.-H.: Pairwise classifier combination using belief functions. Pattern Recogn. Lett. **28**(5), 644–653 (2007)
10. Samiappan, S., Prasad, S., Bruce, L.M.: Non-uniform random feature selection, kernel density scoring with SVM based ensemble classification for hyperspectral image analysis. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **6**(2), 792–800 (2013)
11. Vaiphasa, C.: Consideration of smoothing techniques for hyperspectral remote sensing. ISPRS J. Photogram. Remote Sens. **60**(2), 91–99 (2006)
12. Xu, P., Davoine, F., Zha, H., Denoeux, T.: Evidential calibration of binary SVM classifiers. Int. J. Approx. Reason. **72**, 55–70 (2016)
13. Yang, J.-H., Wang, L.-G., Qian, J.-X.: Hyperspectral image classification based on spatial and spectral features and sparse representation. Appl. Geophys. **11**(4), 489–499 (2014)