

FUSION OF HYPERSPECTRAL AND VHR MULTISPECTRAL IMAGE CLASSIFICATIONS IN URBAN AREAS

Alexandre Hervieu, Arnaud Le Bris, Clément Mallet

Université Paris-Est, IGN, MATIS laboratory, 73 avenue de Paris, 94160 Saint Mandé, FRANCE
arnaud.le-bris@ign.fr – clement.mallet@ign.fr

ICWG III/VII

KEY WORDS: multispectral imagery, hyperspectral imagery, urban land cover, regularization, fusion, classification, energy minimization, graphical model.

ABSTRACT:

An energetical approach is proposed for classification decision fusion in urban areas using multispectral and hyperspectral imagery at distinct spatial resolutions. Hyperspectral data provides a great ability to discriminate land-cover classes while multispectral data, usually at higher spatial resolution, makes possible a more accurate spatial delineation of the classes. Hence, the aim here is to achieve the most accurate classification maps by taking advantage of both data sources at the decision level: spectral properties of the hyperspectral data and the geometrical resolution of multispectral images. More specifically, the proposed method takes into account probability class membership maps in order to improve the classification fusion process. Such probability maps are available using standard classification techniques such as Random Forests or Support Vector Machines. Classification probability maps are integrated into an energy framework where minimization of a given energy leads to better classification maps. The energy is minimized using a graph-cut method called quadratic pseudo-boolean optimization (QPBO) with α -expansion. A first model is proposed that gives satisfactory results in terms of classification results and visual interpretation. This model is compared to a standard Potts models adapted to the considered problem. Finally, the model is enhanced by integrating the spatial contrast observed in the data source of higher spatial resolution (i.e., the multispectral image). Obtained results using the proposed energetical decision fusion process are shown on two urban multispectral/hyperspectral datasets. 2-3% improvement is noticed with respect to a Potts formulation and 3-8% compared to a single hyperspectral-based classification.

1. INTRODUCTION

In land-cover mapping, optical images with higher spatial resolution (<2 m) offers a slight increase in classification accuracy with respect to 2-10 m imagery (Khatami et al., 2016). It can be attributed to the significant amount of geometrical details present in the scenes. However it is also often accompanied with two main drawbacks. First, VHR data steadily increases the spectral variability of each land-cover class and decreases it between classes. Secondly, the spectral resolution of VHR data is often limited to three or four spectral bands (red, green, blue, infrared). Consequently, it appears relevant to merge such multispectral (MS) VHR images with hyperspectral (HS) data. The latter one gives a precise description of the spectral information but with a low geometric precision. Hence, HS data allows reliable land-cover classification results while MS image helps retrieving the geometric contours of such classes. Combining these two data sources may help reaching better classification results at the highest spatial resolution between both datasets.

Combining data with different dimensionalities, spectral and spatial resolutions is a standard remote sensing problem that has been extensively investigated in the literature (Chavez, 1991). Such issue has been exacerbated in the last years with the emergence of new optical sensors with various spatial and spectral configurations. Remote sensing is now inherently multi-modal. The possibility to acquire images of the same area by different sensors has resulted in scientific research focusing on fusing multi-sensor information as a means of combining the comparative advantages of each sensor. Complementary observations can thus be exploited for land-cover mapping purposes and combining existing observations can mitigate limitations of any one particular sensor in particular for land-cover issues (Gamba, 2014; Joshi et

al., 2016). None of them strictly outperforms all the others.

Fusion can be carried out at three different levels. First, it can be achieved at the observation level. For that purpose, pan-sharpening is a well known technique that integrates the geometric details of a high-resolution panchromatic image (PAN) and the color information of a low-resolution MS image to produce a high-resolution MS image. Pan-sharpening methods usually use PAN image to replace the high frequency part in the MS image (Carper, 1990). Other fusion algorithms have been proposed to merge MS and PAN images to combine complementary characteristics in terms of spatial and spectral resolutions (Loncan, 2016). Secondly, data sources can be merged at the feature level. Attributes are computed for each source separately but are fed into the same classifier through a unique feature set.

Thirdly, decision fusion can be performed (Benediktsson and Kanellopoulos, 1999) i.e., outputs of multiple independent classifiers are combined in order to provide a more reliable decision (Aitkenhead and Aalders, 2011; Huang and Zhang, 2012). Several types of fusion methods have been proposed, i.e., probabilistic, fuzzy and possibilistic fusion (Fauvel et al., 2006) as well as the evidence theory (Tupin, 2014). This latter type of fusion is the most popular nowadays and was developed by Dempster and Shafer (Shafer, 1976). This general framework for reasoning with uncertainty relies on the use of belief functions, and makes it possible to combine evidence from different observations to reach a certain degree of belief. Yet efficient in some cases, it is a theoretical complex framework that does not apply easily when dealing with heterogeneous and multiple data. Another effective solution is to propose as feature set to a classifier the probabilistic outputs of the several mono-source classifiers (Ceamanos et al., 2010).

In this paper, a fusion technique at the decision level is presented. Classification results are obtained from MS and HS images sep-

arately. The specific aim here is to use the class membership probability maps, an additional information given by standard classification algorithms. Such information is here used within a generic energetical framework, to be able to generalize the process to classifications obtained from several data sources at various spatial resolutions with complementary advantages (optical images, lidar and radar). At present, work has focused on spatial fusion and no advanced method (e.g. using evidence theory) was used to manage data uncertainty.

Hence, it was aimed at proposing a simple and adaptive method that will further be generalizable to integrate easily several types of data, with diversity in term of geometric, spectral and temporal resolutions. Thus, it can also be used when pansharpening methods are not adapted. The model presented in this paper is based upon graph-cut algorithms, that are well known and have been widely used in the computer vision and image processing community. These techniques rely on the definition of an energy composed with a data term and a regularization term, and the aim is to minimize the energy to obtain the desired result. Such techniques are popular for their simplicity of use and their flexibility while giving satisfying results in a wide range of application fields such as recognition, segmentation or 3D data reconstruction with a suitable computing time (Szeliski, 2010). Thus, such techniques were used to integrate an usually unexploited information (i.e., the probability maps of class membership) in order to propose an efficient fusion of classification results obtained on any pair of MS and HS images.

2. METHODS

The core of the fusion method is based upon a specific data: the class membership probability for each pixel of the images. Thus, this section first shortly reminds some classification techniques providing such an output. The rest of the section is dedicated to the description of the fusion method.

2.1 Classification algorithms

Most classification approaches provide probability values for each class of interest as an output, namely the Random Forests (RF) method (Breiman, 2001) and the Support Vector Machines (SVM) one (Schölkopf, 2002). RF technique natively provides posterior class probabilities output whereas additional steps are required for SVMs (Platt, 2000). For classification purposes, SVMs usually give slightly better results than RFs, but with longer computation times. Since the processing time is not an issue here, our preference was given to SVMs, and more precisely to a Gaussian kernel SVM. Given a set of training pixels for each class, SVM learns a model and assigns new pixels into one of the considered class. As denoted before, SVM also provides posterior class probabilities, retrieved with the Platt's technique. $\forall u \in I, P(C_k(u)) = P(u \in C_k)$ where u is a pixel of the image I and C_k is one of the k classes of interest. SVM classification process is applied to both VHR MS and HS image, as depicted in Figure 1, which presents VHR MS and HS images and their corresponding classifications (in grey level). It must here be underline that in our experiments, VHR MS images were sometimes replaced by panchromatic ones so as to have a more challenging problem.

All SVM classifications considered in this article were obtained using sets of 100 randomly selected pixels for each class within the ground truth data (see Section 3.1 for more details).

2.2 Basic model

In this section, we present the basic model for MS and HS image classification fusion using posterior class probabilities. It is

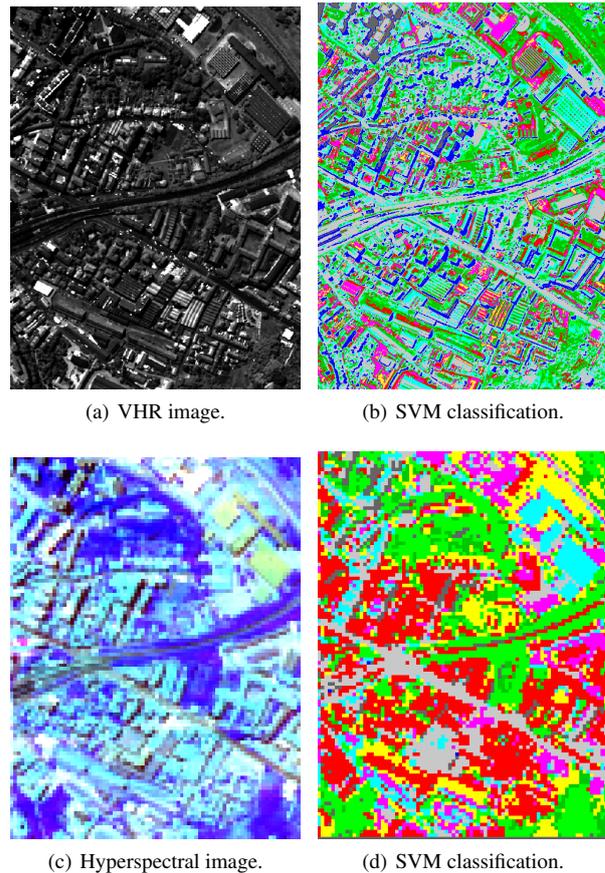


Figure 1: Inputs of our method. Two SVM classifications of one multispectral and one hyperspectral image. The label maps are accompanied with posterior class probabilities.

based upon the definition of an energy that will further be minimized. The energy is composed of two terms: a data attachment term E_{data} and a regularization term E_{regul} . This kind of formulation is well known in the image processing domain (Kolmogorov and Zabih, 2004) and has been successfully used for different applications. The energy model is a probabilistic graph taking into account the posterior class probabilities P_{MS} and P_{HS} from MS image I_{MS} and HS image I_{HS} , respectively, as well as the MS classification C_{MS} . In order to get probability maps with identical sizes, I_{HS} was first upsampled to the size of I_{MS} , meaning that each pixel of the VHR MS image has a corresponding pixel in the HS image. For a classification map C , our basic model defines the energy E as:

$$E(P_{MS}, P_{HS}, C_{HS}, C) = \sum_{u \in I_{MS}} E_{\text{data}}(C(u)) + \lambda \sum_{u, v \in N} E_{\text{regul}}(C(u), C(v)) \quad (1)$$

where:

$$\begin{aligned} E_{\text{data}}(C(u)) &= f(P_{HS}(C(u))), \\ E_{\text{regul}}(C(u) = C(v)) &= g(P_{MS}(C(u)), C_{MS}), \\ E_{\text{regul}}(C(u) \neq C(v)) &= h(P_{MS}(C(u)), C_{MS}), \end{aligned}$$

$\lambda \in [0, \infty[$ is the tradeoff parameter between both terms and N is the 8 connectivity neighborhood. E_{data} is a function of the probability map P_{HS} since I_{HS} is the data containing the most discriminative information for classification. If $P_{HS}(u)$ is high, the pixel u is prone to belong to class $C(u)$ and $E_{\text{data}}(C(u))$ will be small.

E_{regul} models the relationship between a pixel and its neighbors. The more a pixel u and its neighbors $v \in N$ correspond to the desired model, the smaller E_{regul} . The full model $E(C)$ expresses how the classification fits to the probability map P_{HS} and how neighboring pixels follow the model defined for E_{regul} .

2.2.1 Data term: E_{data} is the data attachment term and is defined by a function f such as:

$$f(x) = 1 - x, \quad \text{with } x \in [0, 1].$$

The function f ensures that if the probability for a pixel u to belong to the class $C(u)$ is close to 1, E_{data} will be close to 0 and will not impact the total energy E . Conversely, if the pixel u is not likely to belong to class $C(u)$, $P_{HS}(C(u))$ will be close to 0 and the data attachment term will be close to its maximum for a pixel, i.e., 1.

2.2.2 Regularization term: The regularization term E_{regul} defines the interaction between a given pixel and its neighbors. For a pixel u and its neighbors in N (the 8 connectivity neighborhood), four cases have to be considered regarding the values $C(u)$, $C(v)$ and C_{MS} . The four cases considered are the next ones:

$$\begin{aligned} E_{\text{regul}}(C(u) = C(v) = C_{MS}(u)) &= 0, \\ E_{\text{regul}}(C(u) = C(v) \neq C_{MS}(u)) &= P_{MS}(C_{MS}(u))^\beta, \\ E_{\text{regul}}(C(u) = C_{MS}(u) \neq C(v)) &= 1 - P_{MS}(C_{MS}(u))^\beta, \\ E_{\text{regul}}(C(v) \neq C(u) \neq C_{MS}(u)) &= 1. \end{aligned}$$

Hence, the regularization term has two functions. The first one is tailored to smooth the results by favoring neighboring pixels to belong to a same class, i.e., when $C(u) = C(v)$. It is the basic idea of the Potts model (Schindler, 2012) where regularization term is simply defined by:

$$\begin{aligned} E_{\text{regul}}(C(u) = C(v)) &= 0, \\ E_{\text{regul}}(C(v) \neq C(u)) &= 1. \end{aligned}$$

The second role of the regularization term E_{regul} is to take into account the classification C_{MS} of the VHR MS image and, more specifically, the probability $P_{HR}(C_{MS}(u))$ associated to the most probable class $C_{MS}(u)$. Thus, if a pixel u and one of its neighbor v are assigned to a same class C and if it also corresponds to the most probable class $C_{MS}(u)$ given by the SVM in the MS image (i.e., if $C(u) = C(v) = C_{MS}(u)$), the regularization term is null. Indeed, in this case, it is the "ideal" configuration where the smoothing criterion is satisfied and classification $C(u)$ matches with the class $C_{MS}(u)$.

If $C(u) = C(v) \neq C_{MS}(u)$, the smoothing criterion is satisfied but the class $C(u)$ does not match with the most probable MS class C_{MS} . In this case, the regularization term is a function of $P_{HR}(C_{MS}(u))$. If $P_{HR}(C_{MS}(u))$ is high, E_{regul} is also high since the likelihood of pixel u to belong to class $C_{MS}(u)$ is strong.

In the third case, $C(u) = C_{MS}(u) \neq C(v)$, the classification $C(u)$ matches with the class $C_{MS}(u)$ but the smoothing criterion is not satisfied. If $P_{MS}(C_{MS}(u))$ is close to 1, which means that SVM on MS image is confident for pixel u to belong to class C_{MS} , then E_{regul} is small and the configuration is favored. Inversely, if $P_{MS}(C_{MS}(u))$ is close to 0, E_{regul} will be high (close to 1) and the configuration is more prone to be dismissed.

The last case is the case where the class $C(u)$ does not match $C_{MS}(u)$ while the smoothing criterion is not satisfied. It is the worst case, where $C(v) \neq C(u) \neq C_{MS}(u)$. In this case, the regularization term is set to its maximum value, i.e., 1.

The parameter $\beta \in [0, \infty]$ is a tradeoff parameter between the smoothing criterion and the importance of C_{MS} in the model. If β is high, the smoothing criterion is predominant and the model

comes close to a Potts model. On the opposite, if β is low, the model will tend to follow the classification given by C_{MS} . A property that will be used further for parameter selection is that when $\beta \rightarrow \infty$, the proposed model becomes a Potts model.

2.2.3 Energy minimization: The minimization of the energy is performed using a quadratic pseudo-boolean optimization method (QPBO)¹. This is a classical graph-cut method that builds a probabilistic graph where each pixel is a node. The minimization is computed by finding the minimal cut (Kolmogorov and Rother, 2007). QPBO performs regul classification, extension to multi-class problem is performed using an α -expansion routine (Kolmogorov and Zabih, 2004).

2.3 Adding contrast information

In this section, the proposed model is extended to a more general model that integrates an important visual property contained in the MS image, i.e., the contrast. The contrast is extracted from the VHR MS image since it is the data source with the highest spatial resolution. Indeed, the contrast in the VHR MS image helps retrieving the precise borders between classes and thus improves the classification details. Another smoothness solution could have been implicit e.g., by integrating larger regions of analysis such as superpixels that can be sharply detected with standard segmentation algorithms of MS images (Achanta et al., 2012).

Following (Rother et al., 2004) for contrast usage, $V_i(u, v) = \exp\left(\frac{-(I_i(u) - I_i(v))^2}{2(I_i(u) - I_i(v))^2}\right)$ is considered, where $I_i(u)$ is the intensity for pixel u in the MS image I_{MS} for dimension i . \bar{X} is the mean of X in the image. The function $V(u, v)$ computing the contrast value is then given by:

$$V(u, v, \epsilon) = \frac{1}{dim} \sum_{i \in [0, dim]} V_i(u, v)^\epsilon; \quad \epsilon \in [0, \infty[, \quad (2)$$

where dim is the number of dimensions of I_{MS} and ϵ is a parameter that modifies the standard deviation in the exponential terms.

The general model for classification results fusion is now:

$$\begin{aligned} E(P_{MS}, P_{HS}, C_{HS}, C) &= \sum_{u \in I_{MS}} E_{\text{data}}(C(u)) \\ &+ \lambda \sum_{u, v \in N} E_{\text{regul}}(C(u), C(v)), \end{aligned}$$

where:

$$\begin{aligned} E_{\text{data}}(C(u)) &= 1 - P_{HS}(C(u)), \\ E_{\text{regul}}(C(u) = C(v) = C_{MS}(u)) &= 0, \\ E_{\text{regul}}(C(u) = C(v) \neq C_{MS}(u)) &= (1 - \gamma)P_{MS}(C_{MS}(u))^\beta, \\ E_{\text{regul}}(C(u) = C_{MS}(u) \neq C(v)) &= \\ &= (1 - \gamma)(1 - P_{MS}(C_{MS}(u))^\beta) + \gamma V(u, v, \epsilon), \\ E_{\text{regul}}(C(v) \neq C(u) \neq C_{MS}(u)) &= (1 - \gamma) + \gamma V(u, v, \epsilon). \end{aligned}$$

The parameter γ is a tradeoff parameter between the basic model (led by the MS classification C_{MS}) and the newly integrated contrast-based terms.

This model integrates the idea that if the contrast between two neighboring pixels u and v is important, two pixels are less prone to belong to a same class. Hence, for the condition, $C(u) = C_{MS}(u) \neq C(v)$, if the value $V(u, v, \epsilon)$ is high, the regularizing term E_{regul} will be high and the configuration will more likely be rejected. Inversely, for the last condition $C(v) \neq C(u) \neq$

¹An implementation is available [here](#).

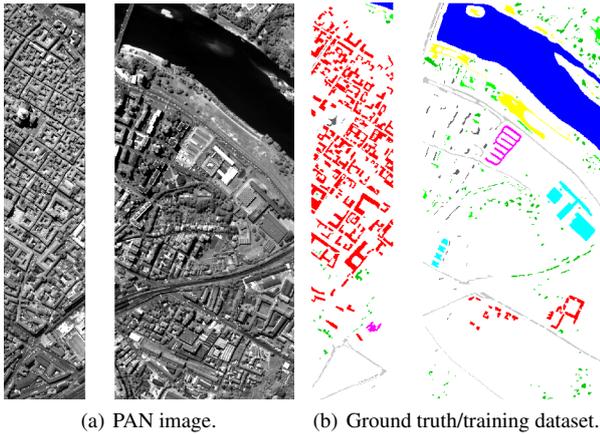


Figure 2: Pavia Centre dataset. Classes : Water - Meadows - Trees - Bare soil - Tiles - Bitumen - Asphalt - Self blocking bricks - Shadows.

$C_{MS}(u)$ where the class assigned to pixels u and v are different, a high contrast will lead to a small regularizing term. If $\gamma = 0$ and $\epsilon = 0$, the proposed model becomes a Potts model (Schindler, 2012) defined by:

$$\begin{aligned} E_{\text{regul}}(C(u) = C(v)) &= 0, \\ E_{\text{regul}}(C(v) \neq C(u)) &= 1. \end{aligned}$$

This property is used to further address the parameter selection step described in the following section.

2.4 Parameter selection

Parameter selection is always an issue when dealing with energies composed of several terms. The general fusion model depends on four parameters, i.e., λ , β , ϵ , and γ . Parameter selection is performed by cross-validation where a limited part (half of the data here) of the data is used to select the best parameter set (the highest percentage of correct classification). This parameter set is then used to process the rest of the data. The issue here is that choosing the best parameter set among the possible ones might be very costly in computing time.

In order to reduce the computing time of this step, we use the following properties of the general model proposed in Section 2.3:

- if $\gamma = 0$ and $\beta \rightarrow +\infty$, the general model is a Potts model;
- if $\gamma = 1$ and $\epsilon = 0$, the general model is a Potts model.

The hypothesis that the value λ_{\max} maximizing the classification result for a Potts model is the same value than the one maximizing the fusion model classification result is formulated here. Hence, λ_{\max} is computed using a simple Potts model. Then, with λ_{\max} and $\gamma = 0$, β_{\max} is found. Similarly, using λ_{\max} and $\gamma = 1$, the value ϵ_{\max} is computed. Lastly, the tradeoff parameter γ_{\max} maximizing results of the model is chosen in the $[0, 1]$ interval. The process can also then be iterated, optimizing parameters in the same order at each iteration.

3. EXPERIMENTS

This section provides first a description of the considered datasets and then presents a discussion on obtained results.

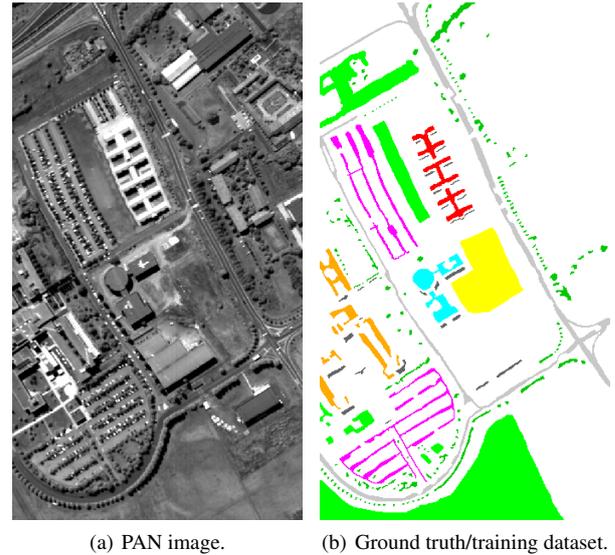


Figure 3: Pavia University dataset. Classes : Meadows - Gravel - Trees - Bare soil - Painted metal sheets - Bitumen - Asphalt - Self blocking bricks - Shadows.

3.1 Data

The proposed model was tested on three distinct urban land cover datasets, namely *Pavia Centre*, *Pavia University* (Italy) and *Toulouse Centre* (France). *Pavia University* and *Pavia Centre* are well known datasets used for years in the hyperspectral community².

Pavia Centre and Pavia University. *Pavia Centre* and *University* hyperspectral images have respectively 102 and 103 spectral bands ranging from 430 to 860 nm. *Pavia Centre* is composed of two images of sizes 228×1096 and 569×1096 pixels. *Pavia University* is a 335×610 pixels image. Both have a ground sample distance (GSD) of 1.3 m. Both image ground truths are composed with different sets of 9 classes. The land cover classes associated to *Pavia Centre* scene are: *trees*, *asphalt*, *self-blocking bricks*, *bitumen*, *tiles*, *shadows*, *meadows*, and *bare soil*. The set of classes for *Pavia University* is composed of: *meadows*, *gravel*, *trees*, *painted metal sheets*, *bare soil*, *bitumen*, *self-blocking bricks*, *shadows*. For *Pavia Centre* and *Pavia University* datasets, a panchromatic image (PAN) with the initial geometric resolution (1.3 m) was created. In the following, these PAN images were used instead of MS images (considering PAN is an extreme case of MS image with only one dimension). This more challenging problem was a way to highlight the properties of the proposed model by using a complex configuration where the geometric precise MS data has very low spectral discriminative properties. The considered HS images were a resampled version at a lower resolution of 7.8 m of the initial HS images. Hence, we have a set of images where the MS image (i.e., here, PAN images) has a 1.3 m geometric resolution and HS image has a 7.8 m geometric resolution. Figures 2 and 3 present the PAN/MS image and the ground truth for both *Pavia Centre* and *University* datasets.

Toulouse. *Toulouse Centre* dataset was simulated from HS images acquired at a 1.6 m GSD and composed with 405 spectral bands ranging from 400 to 2500 nm (Adeline et al., 2013). MS images composed of 5 bands (in the visible and near infra-red wavelengths) were then simulated at a 1.6 m GSD using the spectral configuration of *Pliades* satellites. HS images were down-sampled to a 8 m GSD. 13 land cover classes were considered:

²Available [here](#).

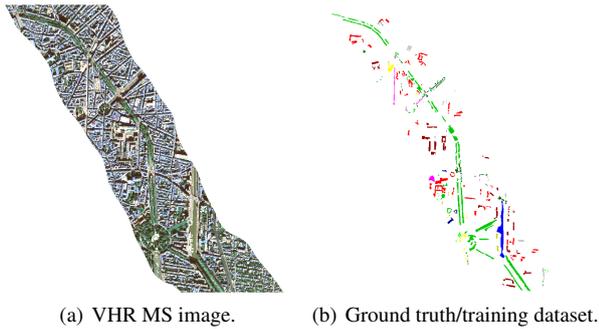


Figure 4: Toulouse Centre. Classes : Tiles - Roofing metal 1 - Roofing metal 2 - Gravel roof - Slates - Cement - Bare soil - Asphalt - Stone pavement bricks - Water - Meadows - Trees - Shadows.

water, high vegetation (trees) and low vegetation (bush), asphalt, tiles, bare soil, metal roof 1 and 2, gravel roof, train track, pavement, cement, slates. Figure 4 shows the MS image of Toulouse Centre dataset and the corresponding annotated ground truth.

3.2 Results

Results obtained on the three datasets are now presented. Figure 6 shows a table containing the classification results obtained with several variants of the proposed energy. The results shown here were obtained using a cross-validation technique where half of the data was used for parameter selection and the rest is the testing data. Parameter selection for those results followed the method described in Section 2.4. The columns VHR MS and HS correspond to a simple SVM classification applied to the VHR MS (or PAN) and HS data, without any fusion step. These are the input of the basic and general fusion algorithms described in the previous sections. In this figure, the results obtained using a Potts model and the general model proposed in Section 2.3 (FUSION column) are also presented.

The obtained results show the contribution of the fusion of VHR MS and HS images for urban land cover processing. Hence, on the three datasets, classification results were highly increased between SVM classifications on HS/VHR MS data and classification using the proposed fusion model. Moreover, comparison with a simple Potts model shows that the fusion models presented in this article not only perform smoothing as designed but also correctly use the probability maps to retrieve details available from the VHR MS image (while keeping to some extent the good classification properties of the HS data).

One should note the ground truth corresponds to limited parts of the areas. Thus, even if the classification results are better, qualitative assessment through visual evaluation on all images remains necessary. This is illustrated in Figure 5 that shows results of the classification on a restricted area (for visual convenience). The difference on the details of the general shape between the results of the Potts model and the fusion models results is obvious. Indeed, we can distinguish the shape of buildings and roads while it was not possible with the best Potts results. Hence, more than the classification results, the method visually efficiently gets the best of the MS and HS information, providing an efficient classification (from HS data) while retrieving details provided by the highest geometric resolution available (from VHR MS data). These properties of the proposed models are observable on Figures 7, 8 and 9. However, it can also be noted that the proposed model is too much focused on MS classification, resulting sometimes in error propagation (e.g., shadows are classified as water in Figure 5).

During testing, the hypothesis tailored for parameter selection

| Methods \ Datasets | VHR MS | HS | POTTS | FUSION |
|--------------------|--------|------|-------|--------|
| Pavia Centre | 62.7 | 94.1 | 96.5 | 97.1 |
| Pavia University | 26.7 | 83.1 | 88.2 | 91.3 |
| Toulouse Centre | 85.2 | 89.3 | 94.2 | 96.2 |

Figure 6: Classification results (in %).

proved to be empirically efficient. Indeed, we computed the classification results using much more exhaustive sets of parameters. We found out that the best classification results were very close from the ones obtained using the procedure described in Section 2.4. Hence the considered hypothesis seems to be valid while significantly reducing the computation costs for parameter selection. For example, if the choice is carried out among $N_\epsilon = N_\lambda = N_\beta = N_\gamma = 100$ values for the parameter set $\{\epsilon, \lambda, \beta, \gamma\}$, 10^8 classification processes need to be computed. Such a procedure enables to reduce hugely the computation time for the parameter selection step. Parameters ϵ , λ , β and γ are selected within grids of sizes $N_\epsilon = N_\lambda = N_\beta = N_\gamma = 100$, the number of classifications to be computed falls from 10^8 to 400, which is still quite important

4. CONCLUSION

In this paper, we presented a method for fusion of VHR multi-spectral (MS) and hyperspectral (HS) data in urban areas at the decision level. The idea was to combine the best of those two types of data, i.e., the high spatial resolution of MS images and the discriminative properties of the HS images. The proposed fusion model relied on posterior class probabilities, available for most of existing classification techniques. Here, a SVM method was adopted. The probabilities were integrated through an energy minimization process, enabling to improve classification results of a single source and to retrieve details observable in the VHR MS data while keeping the good classification properties obtained with the HS data. The model is generic and intends to be applied to other configuration involving VHR image and richer remote sensing data.

The perspectives are threefold. First, to some extent, the energy should be modified to less rely on posterior class probabilities in order to better manage initial misclassifications or spatial inconsistencies. Secondly, we will extend our model to a larger set of images (instead of only two here) of various spectral dimensions and geometric resolutions. The idea is to be able to process and fusion any information of a given scene to get the best classification at the most precise spatial resolution available. This can be performed using a multi-layer process where each layer is the model described in this paper for fusion of two images. Another final interesting perspective is to process MS/HS images acquired at different epochs. This opens the field for change detection in urban areas, and will require more advanced knowledge fusion model.

ACKNOWLEDGMENTS

This work was supported by the French National Research Agency (ANR) through the HYP project on Hyperspectral imagery for Environmental urban Planning (ANR-14-CE22-0016).

BIBLIOGRAPHY

Achanta, R., Shaji, A., Smith, K. Lucchi, A., Fua, P., Süsstrunk, S. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(11), pp.2274–2282.

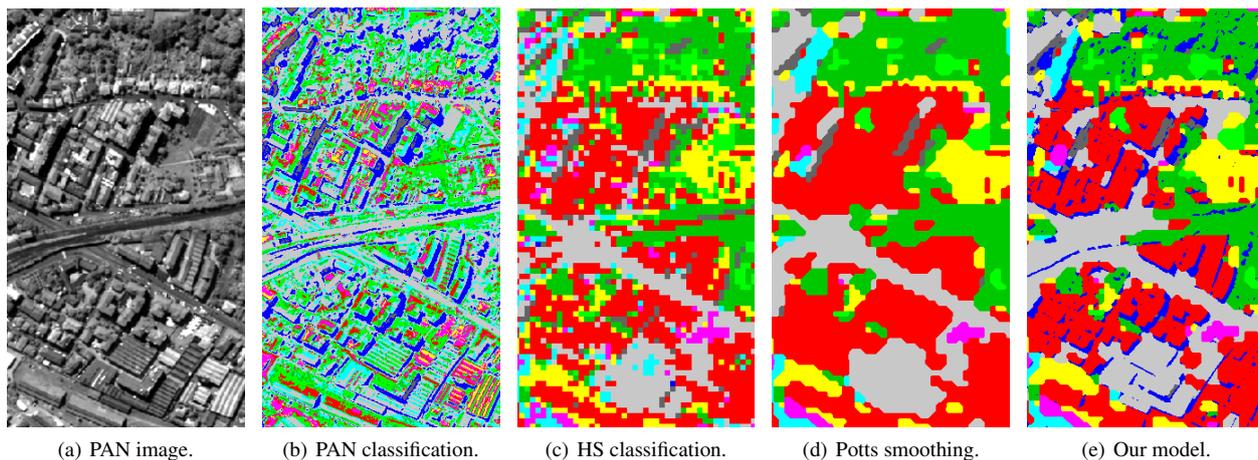


Figure 5: Results on a part of Pavia Centre dataset.

Adeline, K., Le Bris A., Coubarde, F., Briottet, X., Paparoditis, N., Viallefont, F., Rivire, N., Papelard, J.-P., David, N., Déliot, P., Duffaut, J., Poutier, L., Foucher, P.-Y., Achard, V., Souchon, J.-P., Thom, C., Airault, S. and Maillet, G., 2013. Description de la campagne aéroportée UMBRA : étude de l'impact anthropique sur les écosystèmes urbains et naturels avec des images THR multispectrales et hyperspectrales. *Revue Française de Photogrammétrie et de Télédétection*, 202, pp.79–92.

Aitkenhead, M. and Aalders, I., 2011. Automating land cover mapping of Scotland using expert system and knowledge integration methods. *Remote Sensing of Environment*, 115(5), pp.1285–1295.

Benediktsson, J.A. and Kanellopoulos, I., 1999. Decision fusion methods in classification of multisource and hyperdimensional data. *IEEE Trans. on Geoscience and Remote Sensing*, 37(3), pp.1367–1377.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), pp.5–32.

Carper, W.J., Lillesand, T.M. and Kiefer, R.W., 1990. The use of Intensity-Hue-Saturation transform for merging SPOT panchromatic and multispectral Image Data. *Photogrammetric Engineering & Remote Sensing*, 56 (4), pp.459–467.

Ceamanos, X., Waske, B., Benediktsson, J.A., Chanussot, J., Fauvel, M. and Sveinsson, J.R., 2010. A classifier ensemble based on fusion of Support Vector Machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4), pp.293–307.

Chavez, S., Stuart, J., Sides, C. and Anderson, J.A., 1991. Comparison of three different methods to merge multiresolution, multispectral data Landsat TM and SPOT panchromatic. *Photogrammetric Engineering & Remote Sensing*, 57(3), pp.259–303.

Fauvel, M., Chanussot, J. and Benediktsson, J.A., 2006. Decision fusion for the classification of urban remote sensing images. *IEEE Trans. on Geoscience and Remote Sensing*, 44(10), pp.2828–2838.

Gamba, P., 2014. Image and data fusion in remote sensing of urban areas: status issues and research trends. *International Journal of Image and Data Fusion*, 5(1), pp.2–12.

Huang, X. and Zhang, L., 2012. A multilevel decision fusion approach for urban mapping using very high-resolution multi-hyper-spectral imagery. *International Journal of Remote Sensing*, 33(11), pp.3354–3372.

Joshi, N. et al., 2016. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1):70.

Khatami, R., Mountrakis, G. and Stehman, S., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. *Remote Sensing of Environment*, 177, pp.89–100.

Kolmogorov, V. and Zabih, R., 2004. What energy functions can be minimized via Graph Cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(2), pp.65–81.

Kolmogorov, V. and Rother, C., 2007. Minimizing non-submodular functions with Graph Cuts - a review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(7), pp.1274–1279.

Loncan, L., Almeida, L. B., Bioucas-Dias, J. M., Liao, W., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G., Simoes, M., Tourneret, J.-Y., Veganzones, M., Vivone, G., Wei, Q. and Yokoya, N., 2015. Hyperspectral pansharpening: a review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 27–46.

Platt, J., 2000. Probabilistic outputs for Support Vector Machines and comparison to regularized likelihood methods. In: A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans (eds.): *Advances in Large Margin Classifiers*. MIT Press, Cambridge, MA, USA.

Rother, C., Kolmogorov, V. and Blake, A., 2004. "Grabcut" - Interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics*, 23(3), pp.309–314.

Schölkopf, B. and Smola, A.J., 2002. *Learning with kernels: Support Vector Machines, regularization, optimization and beyond*. MIT press, Cambridge, MA, USA.

Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Trans. on Geoscience and Remote Sensing*, 50(11), pp.4534–4545.

Szeliski, R., 2010. *Computer Vision: algorithms and applications*. Springer-Verlag, London, UK.

Shafer, G., 1976. *A mathematical theory of evidence*. Princeton University Press, Princeton, NJ, USA.

Tupin, F., Inglada, J. and Nicolas, J.-M., 2014. *Remote sensing imagery*. Wiley, Hoboken, NJ, USA.

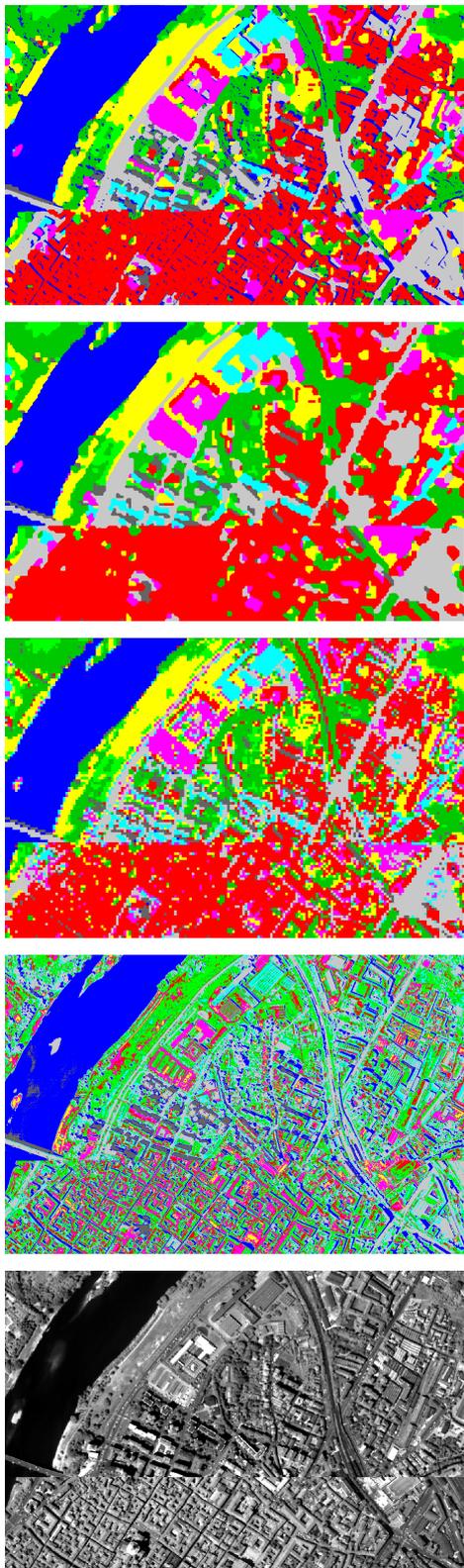


Figure 7: Pavia Centre results.

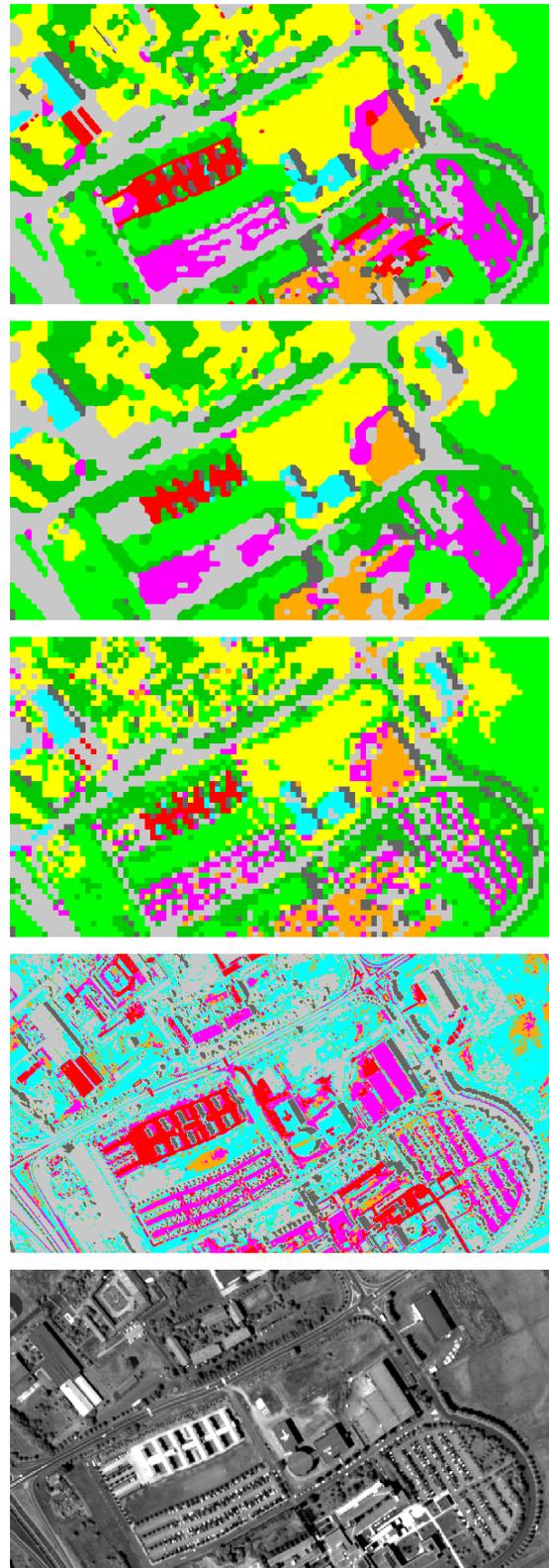


Figure 8: Pavia University results.

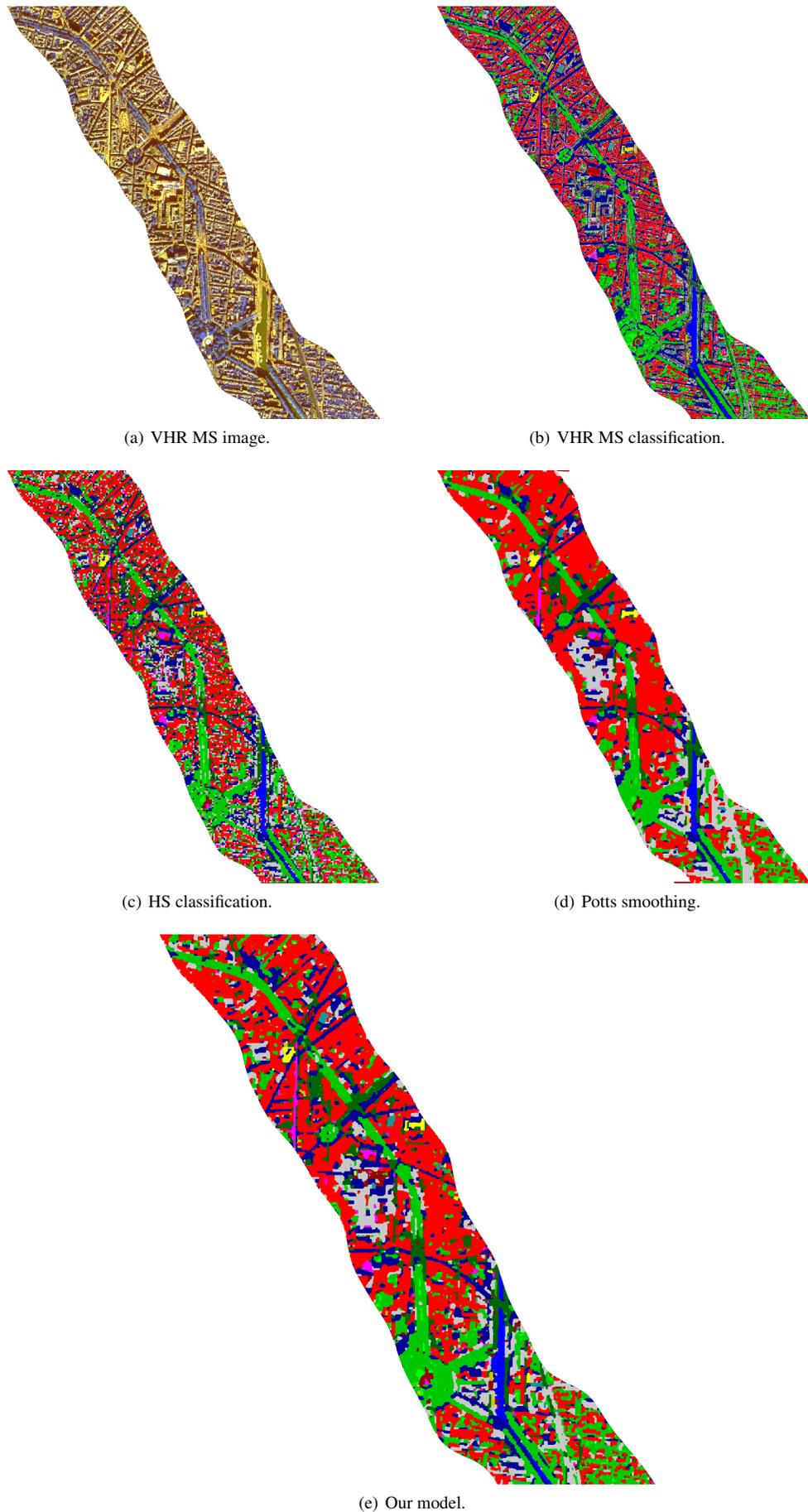


Figure 9: Toulouse Centre results.