How Transferable Are Spatial Features for the Classification of Very High Resolution Remote Sensing Data?

Mathieu Fauvel *, Jocelyn Chanussot *  

*Department of Images and Signal Processing – GIPSA-Lab, Grenoble  
Grenoble Institute of Technology - INPG  
BP 46 - 38402 St Martin d’Heres - FRANCE  
E-mail: {mathieu.fauvel, jocelyn.chanussot}@lis.inpg.fr

Jon Atli Benediktsson °  

°Department of Electrical and Computer Engineering  
University of Iceland  
Hjardarhagi 2-6, 107 Reykjavik, ICELAND  
E-mail: benedikt@hi.is

Abstract — Knowledge transfer for the classification of very high resolution panchromatic data over urban area is investigated. Invariant feature are extracted with some morphological processing. The well-known spectral angle mapper (SAM) is proposed as a measure of transferability. Support vector machines (SVMs) are used to fit a separating hyperplane in a vector space defined by the extracted spatial features. The hyperplane is then used to classify other data set without any new training. Several experiments are presented. Results confirm the usefulness of spatial feature when the classification of two images from two separates data set is considered.

I. INTRODUCTION

With the full availability of very high resolution images (VHR), such as IKONOS or QUICKBIRD data, mapping of dense urban area is possible. Semi-automated or automated classification algorithms can be integrated into urban planning development, emergency response or Earth survey. Several approaches have been proposed in this last decades for the classification of urban area. Benediktsson et al. [1] uses mathematical morphology to extract spatial features, and a neural network to classify panchromatic IKONOS images. A Markov random field (MRF) based classifier is used in [2] and simultaneously exploits spatial and spectral information from the data. In [3], streets network is extracted on panchromatic IKONOS data. All these algorithms are based on the main characteristic of the VHR data: the high spatial resolution. As a matter of fact, the morphological processing extracts spatial relationship between pixel, the MRF uses a spatial energy term in the MRF model, and the detected streets network is regularized using an a priori spatial organization.

For the above-listed applications, speed and accuracy of the algorithms are critical issues. The main shortcoming of these methods is the need of a training step that has to be completed for each new data set, involving potentially tedious manual labelling. Recent works discuss the possibility of knowledge transfer for classification algorithms. In [4], the authors exploit existing classifiers to construct a new classifier, addressing a new problem. This task is very difficult: it is actually confronted to the variations of the class characteristics in the spectral domain. It is thus handled by using a Binary Hierarchical Classifier where an update step is added every time new data are processed.

For the analysis of urban area with VHR data, the features used for the classification are extracted from the structures of the image. In order to apply knowledge transfer for the classification of such data, it is necessary to extract features that do not change during time or over the space. For man made construction, shape, size, texture or orientation are features that should possibly remain almost invariant.

Using the Morphological Profile (MP) and its derivative (DMP) [1], information about the shape, the size and the local contrast of the structures present in the image can be extracted. In this paper, we propose to use the MP as an invariant features vector for the classification of VHR data of dense urban area.

The transferability of the features is first investigated theoretically. The spectral angle mapper between each features vector is computed for several real images. Then classification is performed using training parameters determined from one image and extended to the whole data set.

For the classification, a support vector machines (SVM) classifier with Gaussian kernel is used. This classifier has good performances for urban data analysis [5]. For one given sensor, differences of illumination result in images with different radiometric information. This point is also discussed and several histogram-based algorithms are used. Their influence on the knowledge transfer is assessed.

The paper is organized as follows. In the next section, morphological processing is detailed. The third section is dedicated to the theoretical study of the transferability. In the fourth section, classification using SVM is briefly recalled and experiments are presented.
II. MORPHOLOGICAL PROFILE

A. Morphological profile

The Morphological Profile is introduced in remote sensing by Pesaresi and Benediktsson in [6]. It is a morphological granulometry [7]. A MP is composed of the opening profile (OP) and the closing profile (CP), respectively. The OP at the pixel \( x \) of image \( I \) is defined as an \( n \)-dimensional vector:

\[
OP_i(x) = \gamma_R^{(i)}(x), \forall i \in [0,n]
\]

where \( \gamma_R^{(i)} \) is the opening by reconstruction [8] with a structuring element (SE) of size \( i \), and \( n \) is the total number of openings. Similarly, the CP at the pixel \( x \) of image \( I \) is defined as an \( n \)-dimensional vector:

\[
CP_i(x) = \phi_R^{(i)}(x), \forall i \in [0,n]
\]

where \( \phi_R^{(i)} \) is the closing by reconstruction with an SE of size \( i \). Clearly we have \( CP_0(x) = OP_0(x) = I(x) \). By concatenating the \( OP \) and the \( CP \), the MP of image \( I \) is defined as the following \( 2n + 1 \)-dimensional vector:

\[
MP(x) = \{CP_0(x), \ldots, I(x), \ldots, OP_n(x)\}
\]

Local spatial information is contained in the MP for each pixel as the response of the considered pixel to opening/closing by reconstruction with a structuring element of increasing size. Fig. 1 shows an example of such an MP.

B. Tuning the size of the MP

To build the \( MP \), one has to properly choose the shape and size of the SE. Classically, disks are used. It has the property of being isotropic, i.e. the property of being independent to changes of orientation. The number of opening/closing and the step size of the SE have to be chosen to cover all the structures of interest in the image. This is chosen in accordance with the resolution of the data and the range of possible variations in the size of the structures of interest. In our application, we chose the following parameters:

- SE: disk,
- Initial size (radius): \( R=2 \) pixels,
- Number of opening/closing: 15,
- Step: 2.

It results in an \( MP \) of size 31, for each pixel.

III. STUDY OF THE TRANSFERABILITY

A. Typical profile

First of all, we have to estimate a typical profile for each class. The \( MP \) has been previously computed for all the test images. Using some manually labelled pixels, several referenced profiles are available. A typical profile is estimated for each class by averaging all the corresponding referenced profiles. Considering the high number of samples, this is a reasonable estimate. Other statistical values have been considered to compute this estimate, such as the median, but the mean value gave the best results. It is thus the only one reported in this paper.

Fig. 2 presents the mean \( MP \) for three images and for four classes, namely: road, building, shadow and open area. Images 1 and 2 were extracted from the same panchromatic image, at two different locations. Image 3 is extracted from another panchromatic image but was acquired with the same sensor.

In the following we compare the profiles obtained for the same class, but from different images.

B. Comparison

To analyze the transferability, we compare the different profiles extracted with the \( MP \). We use the spectral angle mapper (SAM) defined as follows:

\[
\alpha_{kl} = \frac{\langle P(k), P(l) \rangle}{\|P(k)\| \|P(l)\|}
\]

where \( P(k) \) is the mean profile of the class \( k \). The SAM is a scale invariant metric.

The SAM measured between two images for each class is reported in Table I. Road and building profiles appear to be stable from one image to another. It fits intuition, since structures with typical shape in one image still have their typical shape in other images. For example, buildings have in general an approximately squaured shape, while roads are thin and elongated structures. On the contrary, the profile obtained for shadows seems to be less stable. As a matter of fact, corresponding shapes depend on the other structures (buildings, trees ...), but also on the sun orientation, e.g. when the sun is high in the sky, the resulting shadows are very small.

The \( MP \) was computed on the original data set. In the next subsection, different histogram-based transformations are presented to correct radiometric differences.
C. Scaling

Before the construction of the MP, three different scaling were tested, and their influence on the SAM analyzed. Considering an image $I$, $I(x)$ a pixel, $I_{\text{min}}$ (respectively $I_{\text{max}}$) the lowest (highest) value of $I$, $\mu_I$ and $\sigma_I^2$ the mean and variance of $I$, respectively, and $H_I$ the histogram of $I$, the scaling algorithms were:

1) Histogram stretching: The pixel values are linearly stretched into $[0, 1]$,

$$I'(x) = \frac{I(x) - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}}$$  \hspace{1cm} (6)

2) Standardization of the histogram: The transformed data have a zero-mean and an unitary variance,

$$I''(x) = \frac{I(x) - \mu_I}{\sigma_I}.$$  \hspace{1cm} (7)

3) Histogram equalization: The transformed data have a flat histogram,

$$I'''(x) = I(x) \int_0^x H_I(z)dz.$$  \hspace{1cm} (8)

The process previously described is applied to the transformed data and the SAM is computed in a similar way. Results are presented on Table II, III and IV, showing an increase of the SAM for the different scalings.

IV. SUPPORT VECTOR MACHINES

A. Generalities

Given a set of known samples $S = \{x^i, y^i\}$, $i \in [1, N]$, with $(x, y) \in \mathbb{R}^n \times \{-1, 1\}$, the support vector machine (SVM) separates the data into two class $(-1, 1)$ by an Optimal Hyperplane (OH) (see Fig. 3). The OH is found through a convex optimization problem [9]:

$$\min \left(\frac{(w, w)_{\mathbb{R}^n}}{2} + C \sum_{i=1}^N \xi_i\right)$$  \hspace{1cm} (9)

subject to

$$y_i((w, x^i)_{\mathbb{R}^n} + b) \geq 1 - \xi_i, \xi_i \geq 0 \ \forall i \in [1, N]$$

where $w$ and $b$ are the OH’s parameters, $\xi$ are slack variables which are introduced to deal with misclassified samples and $C$ is a constant that controls the amount of penalty.

The optimal parameter is found using Lagrangian:

$$w = \sum_{i=1}^N y_i \alpha_i x^i.$$  \hspace{1cm} (10)

where $\alpha_i$ are Lagrangian coefficient.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>SPECTRAL ANGLE, IN RADIANS, FOR EACH CLASS BETWEEN THE DIFFERENT IMAGES.</th>
<th>TABLE II</th>
<th>SPECTRAL ANGLE, IN RADIANS, FOR EACH CLASS BETWEEN THE DIFFERENT IMAGES LINEARLY SCALED.</th>
<th>TABLE III</th>
<th>SPECTRAL ANGLE, IN RADIANS, FOR EACH CLASS BETWEEN THE DIFFERENT IMAGES STANDARDIZED.</th>
<th>TABLE IV</th>
<th>SPECTRAL ANGLE, IN RADIANS, FOR EACH CLASS BETWEEN THE DIFFERENT IMAGES EQUALIZED.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MP$</td>
<td>1 &amp; 2</td>
<td>1 &amp; 3</td>
<td>2 &amp; 3</td>
<td>$MP$</td>
<td>1 &amp; 2</td>
<td>1 &amp; 3</td>
<td>2 &amp; 3</td>
</tr>
<tr>
<td>Road</td>
<td>0.0469</td>
<td>0.0354</td>
<td>0.0488</td>
<td>Building</td>
<td>0.0661</td>
<td>0.0812</td>
<td>0.0926</td>
</tr>
<tr>
<td>Building</td>
<td>0.0336</td>
<td>0.0396</td>
<td>0.0620</td>
<td>Building</td>
<td>0.0423</td>
<td>0.1012</td>
<td>0.1150</td>
</tr>
<tr>
<td>Shadow</td>
<td>0.0740</td>
<td>0.2466</td>
<td>0.2145</td>
<td>Shadow</td>
<td>0.0656</td>
<td>0.2122</td>
<td>0.2477</td>
</tr>
<tr>
<td>Open Area</td>
<td>0.0423</td>
<td>0.1591</td>
<td>0.1447</td>
<td>Open Area</td>
<td>0.0574</td>
<td>0.0427</td>
<td>0.0837</td>
</tr>
</tbody>
</table>

Fig. 2. Mean Morphological Profiles extracted from the panchromatic images: “+” corresponds to the class road, “−” to the class building, “.” to the class open area.
The solution vector is a linear combination of some samples of the training set, whose $\alpha_i$ are non-zero. These samples are called the Support Vectors. The hyperplane decision function can thus be written as:

$$y_u = sgn \left( \sum_{i=1}^{N} y_i \alpha_i \langle \mathbf{x}^u, \mathbf{x}^i \rangle + b \right)$$

(11)

where $\mathbf{x}^u$ is an unseen sample.

Using kernel methods, it is possible to perform a non linear classification [10]. The inner product is substituted by a kernel function that allows an implicit formulation of the algorithm in some feature space [9]. One classical kernel function is the radial basis function:

$$k_\gamma (\mathbf{x}, \mathbf{z}) = \exp \left( -\frac{\| \mathbf{x} - \mathbf{z} \|^2}{2\gamma^2} \right)$$

(12)

where the norm is the Euclidean-norm and $\gamma \in \mathbb{R}^+$ tunes the flexibility of the kernel. A short comparison of kernels for remotely sensed image classification can be found in [5].

Finally, Eq. (11) can be rewritten as follows:

$$y_u = sgn \left( \sum_{i=1}^{N} y_i \alpha_i k_\gamma (\mathbf{x}^u, \mathbf{x}^i) + b \right).$$

(13)

B. Knowledge transfer

A feature vector of size 31 is associated to each pixel. OH’s parameters (i.e. $w$, $b$, $\alpha$ and $\gamma$) are found based on these features.

For one single data set, a hyperplane that separates the data in a feature space is found using standard SVM training process. Then, this decision function is used to classify data from another data set, without any new training process.

V. EXPERIMENTS

Experiments were conducted on various panchromatic images. For each image, the $MP$ is computed, see section II-B, and a classification is performed using the SVM. Two series of experiments were conducted, with and without scaling, respectively. Results are presented in the two following sections.

The data set consists of three panchromatic images extracted from simulated Pleiades images provided by CNES (satellite to be launched in 2008). The spatial resolution is 0.75 meter by pixel. All images are urban areas. Uncorrelated train and test set were built for each image, see Table V.

A. $MP$ versus Grayscale information

In the first experiments, we compare the knowledge transfer with and without $MP$. Results are reported in Tables VI, VII and VIII.

From the Tables, we can see that for two images extracted from the same data set (Im. 1 and 2), the grayscale information leads to a better classification results. Moreover, the results are not so different, e.g. $88.83\% \rightarrow 87.75\%$ in Table VII, whereas they are clearly worst when the $MP$ is used.

When considering images extracted from two different data sets (Im. 1 and 3, Im. 2 and 3), the results are better with the $MP$. In that case, the radiometric information has changed and the spatial information included in the $MP$ is helpful for the knowledge transfer.

B. Influence of scaling

As explained in section III-C, different histogram based transformations were tested, prior to the construction of the

TABLE V

<table>
<thead>
<tr>
<th></th>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>train</td>
<td>test</td>
<td>train</td>
<td>test</td>
<td>train</td>
<td>test</td>
</tr>
<tr>
<td>Class 1</td>
<td>90.18</td>
<td>77.98</td>
<td>25.16</td>
<td>90.18</td>
<td>77.98</td>
<td>25.16</td>
</tr>
<tr>
<td>Class 2</td>
<td>83.81</td>
<td>70.89</td>
<td>72.06</td>
<td>83.81</td>
<td>70.89</td>
<td>72.06</td>
</tr>
<tr>
<td>Class 3</td>
<td>92.80</td>
<td>69.30</td>
<td>69.30</td>
<td>92.80</td>
<td>69.30</td>
<td>69.30</td>
</tr>
<tr>
<td>Class 4</td>
<td>90.33</td>
<td>75.49</td>
<td>62.37</td>
<td>90.33</td>
<td>75.49</td>
<td>62.37</td>
</tr>
<tr>
<td>Average</td>
<td>88.83</td>
<td>87.75</td>
<td>54.39</td>
<td>88.83</td>
<td>87.75</td>
<td>54.39</td>
</tr>
</tbody>
</table>

TABLE VI

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>83.81</td>
<td>82.67</td>
<td>74.73</td>
<td>91.41</td>
<td>85.86</td>
</tr>
<tr>
<td>91.71</td>
<td>89.43</td>
<td>79.50</td>
<td>79.55</td>
<td>86.43</td>
</tr>
<tr>
<td>93.00</td>
<td>85.30</td>
<td>76.92</td>
<td>85.94</td>
<td>77.74</td>
</tr>
<tr>
<td>84.63</td>
<td>77.74</td>
<td>75.84</td>
<td>87.93</td>
<td>65.35</td>
</tr>
<tr>
<td>89.20</td>
<td>83.79</td>
<td>62.50</td>
<td>92.98</td>
<td>60.66</td>
</tr>
<tr>
<td>75.49</td>
<td>70.89</td>
<td>72.06</td>
<td>90.33</td>
<td>62.37</td>
</tr>
<tr>
<td>90.18</td>
<td>77.98</td>
<td>25.16</td>
<td>90.33</td>
<td>75.49</td>
</tr>
<tr>
<td>83.81</td>
<td>70.89</td>
<td>72.06</td>
<td>83.81</td>
<td>70.89</td>
</tr>
<tr>
<td>92.80</td>
<td>69.30</td>
<td>69.30</td>
<td>92.80</td>
<td>69.30</td>
</tr>
<tr>
<td>90.33</td>
<td>75.49</td>
<td>62.37</td>
<td>90.33</td>
<td>75.49</td>
</tr>
</tbody>
</table>

TABLE VII

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>89.20</td>
<td>83.79</td>
<td>62.50</td>
<td>92.98</td>
<td>60.66</td>
</tr>
<tr>
<td>89.20</td>
<td>83.79</td>
<td>62.50</td>
<td>92.98</td>
<td>60.66</td>
</tr>
<tr>
<td>89.20</td>
<td>83.79</td>
<td>62.50</td>
<td>92.98</td>
<td>60.66</td>
</tr>
<tr>
<td>89.20</td>
<td>83.79</td>
<td>62.50</td>
<td>92.98</td>
<td>60.66</td>
</tr>
</tbody>
</table>
For the sake of clarity, we only report results of knowledge transfer for images extracted from two different data sets (as described in the previous section, radiometric information is sufficient for images extracted from one single data set). Results are presented in Tables IX, X and XI. Im. a ← Im. b means that image a is classified by the OH found with the training set of image b. Results in brackets are the classification accuracy obtained with the original training set.

From the Tables, every radiometric correction lead to better classification accuracies, linear scaling and standardization providing the best results. The equalization artificially stretches the data, according to the cumulative histogram. However, the cumulative histogram is too dependent of the image, and may thus not be appropriate for knowledge transfer.

The best results were obtained with linear scaling: when the image 3 is classified with the OH found with image 1 (respectively 2), the final classification accuracy is 86.47% (resp. 74.74%) against 71.15% (resp 64.98%) without scaling. Is it interesting to note that the best knowledge transfer is done with the image which have the largest training set. The SVM’s training algorithm found the samples that support the separating hyperplane, and, as a consequence, with a larger training set, more discriminative samples can be extracted. An algorithm has been proposed by Bruzzone et al. [11] based on this principle. An OH is found and using some unlabeled samples the OH is updated by adding/removing some support vectors found with the unlabeled samples.

Finally, one should note that despite the SAM between the MP actually increased with the radiometric normalization, the classification accuracy is improved.

VI. CONCLUSION

Transferability of spatial features for the classification of urban area has been discussed. Morphological processing was used to extract invariant features and support vector machines were used to classify the data.

The spectral angle mapper does not seem to be a good measure of the features’ transferability, since its variation does not follow the classification results.

For two images extracted from the same data set, radiometric information performs well, leading to good classification performances. However, the classes were defined at a coarse level: building, road ... If finer definition is desired, spatial definition should help for knowledge transfer.

For two images extracted from two different data sets, MP with linear scaling of the data gave promising results. More advanced SVM algorithm should help for the classification.

Finally, we conclude that spatial features extracted with the MP are almost invariant. Our current research is now focused on invariant texture features extraction.

REFERENCES