# Descriptor Correlation Analysis for Remote Sensing Image Multi-Scale Classification

J. A. dos Santos[1,2], F. A. Faria[1], R. da S. Torres[1], A. Rocha[1], P-H. Gosselin[2], S. Philipp-Foliguet[2], A. Falcão[1]
[1]*RECOD Lab – Institute of Computing, University of Campinas, Brazil*
[2]*ETIS, CNRS, ENSEA, University of Cergy-Pontoise, France*
{jsantos,ffaria,rtorres,afalcao,anderson.rocha}@ic.unicamp.br, {gosselin,philipp}@ensea.fr

## Abstract

*This paper addresses the problem of remote sensing image multi-scale classification by: (i) showing that using multiple scales does improve classification results, but not all scales have the same importance; (ii) showing that image descriptors do not offer the same contribution at all scales, as commonly thought, and some of them are very correlated; (iii) introducing a simple approach to automatically select segmentation scales, descriptors, and classifiers based on correlation and accuracy analysis.*

## 1 Introduction

Region-based solutions for remote sensing image (RSI) classification normally rely on the use of segmentation techniques and robust pattern recognition methods [1]. Naturally, both segmentation and machine learning have unsolved problems on their own and often we need to cope with some of them when dealing with a specific scenario. For instance, in RSI, some researchers have addressed these problems in two ways: by exploiting regions with multiple sizes to improve segmentation [10], or by combining different features and classifiers [3, 4] to improve pattern recognition. In this sense, we proposed the Multi-Scale Classification (MSC) approach to combine different features at multiple segmentation scales [5]. The strategy is based on the boosting paradigm, whose principle is to combine weak classifiers to build an effective and efficient global one. MSC exploits a hierarchy of regions obtained by using the Guigues algorithm [7].

In reviewing RSI works that deal with combination of segmentation scales, features and classifiers, many questions arise related to how these combinations are done: (1) the ideal combination; (2) an intelligent way to combine and save time; (3) the best number of segmentation scales; and (4) the best number of descriptors. These are only some examples of questions and several others are possible. To the best of our acknowledge, there is no formal analysis concerning these issues and normally what we see is a compilation of several adhoc methods. In [5], we show that the combination of features at different scales improves the classification results, but these results still lack more explanation about how to select the best scales and descriptors.

In this context, the objective of this paper is to address such questions. First, we show that although using multiple scales does improve the results, not all scales have the same importance. Fusion of classifiers is also important, but not all combinations yield good results. Finally, in face of the observations made, we present a simple approach to automatically select segmentation scales, descriptors, and classifiers based on the correlation and the accuracy analysis, and all this need no a priori knowledge, no learning, everything is performed on line, during the system use.

## 2 Experimental Setup

To support our study, we have carried out experiments with the following components:

**RSI Data:** we used a high resolution SPOT scene ($3,000 \times 3,000$ pixels) which corresponds to the Monte Santo de Minas county, in the State of Minas Gerais, Brazil. It is a traditional place of coffee cultivation. The distortions in the relief and specific aspects of coffee crops increase the challenge of classifying this dataset. To evaluate the accuracy, we use a ground truth that indicates all coffee crops in the image. We divided the image into a $3 \times 3$ grid, generating nine $1,000 \times 1,000$ subimages. We carried out experiments with ten different combinations of the nine subimages used (three for training, three for validation, and three for testing). More details about the dataset can be found in [5].

**Segmentation Scales:** we consider five different scales to extract features: $\lambda_1, \ldots, \lambda_5$. We define the scales according to the principle of dichotomic cuts proposed in [7]. The higher the index, the coarser is the scale. The scale $\lambda_0$ is the pixel level.

**Features:** we extracted different features from the band composition IR-NIR-R (342) by using four color and three texture descriptors. The color descriptors are: Global Color Histogram (GCH), Color Coherence Vector (CCV), Color Autocorrelogram (ACC), and Border/Interior Pixel Classification (BIC). The texture descriptors are: Invariant Steerable Pyramid Decomposition (SID), Unser, and Quantized Compound Change Histogram (QCCH). These descriptors were selected based on previous results as reported in [6].

**Classifiers:** we use support vector machines (SVMs) with no kernels for each descriptor at scale $\lambda_i$. In the experiments with the MSC, we used "weakened" SVMs as weak learners. More details about the implementation of SVMs as weak learners can be found in [5].

**Effectiveness Measure:** We analyze the results by computing the overall accuracy, kappa index, and tau index for the classified images [9].

## 3 Results and Discussion

This section presents the correlation analysis and the proposed approach for selecting classifiers.

### 3.1 Correlation Analysis

The first study is concerned with the analysis of the accuracy of classifiers at different segmentation scales. The second study is the correlation analysis of each pair of classifiers. In our experiments, a classifier is defined for a descriptor and a segmentation scale. We use $Cor$ [8] to assess the correlation of two classifiers $c_i$ and $c_j$:

$$COR(c_i, c_j) = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad (1)$$

where $a$ is the percentage of images that both classifiers $c_i$ and $c_j$ classified correctly in the training set, $b$ and $c$ are the percentage of images that $c_j$ hit and $c_i$ missed and vice versa, and $d$ is the percentage of images that both classifiers missed.

#### 3.1.1 Classifier Accuracy for Different Segmentation Scales

Figure 1 shows the overall accuracy, the kappa index, and the tau index for each SVM classifier implemented using each descriptor/scale. We observe a large difference between the accuracy results (Figure 1(a)) of the classifier implemented with color and texture descriptors for almost all scales. Among the color descriptor accuracies, we have no significant difference, although BIC presents the higher values at all scales. Among the texture ones, they present almost the same accuracies at all scales except for QCCH that presents its best results at the coarser scales.

Regarding the tau indexes (Figure 1(b)), which is more discriminative than overall accuracy, we observe that BIC achieves the best results for all scales. GCH also yields the best result at the coarser scale $\lambda_5$. Among the texture descriptors, all of them are almost random at the finest scales ($\lambda_1$ and $\lambda_2$). QCCH presents the best results at the intermediate scale $\lambda_3$. The texture descriptors present their best results at the coarsest scales $\lambda_4$ and $\lambda_5$. At the coarsest scales, QCCH and Unser present better results than SID.

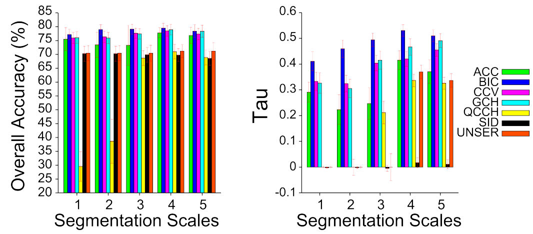


**Figure 1. Classifier effectiveness for each descriptor at various segmentation scales: Overall accuracy (on the left), and Tau index (on the right).**

The main conclusion of this experiment is that color descriptors are very important at all scales while texture features can contribute only at the coarsest ones.

#### 3.1.2 Classifier Correlation for Different Segmentation Scales

In this section, we analyze the correlation of each pair of classifiers at the segmentation scales.

Figure 2 shows the correlation scores considering the different descriptors and scales. We have observed that the correlation among the descriptors presents minor differences depending on the training set. We report in this paper the most common patterns observed in the experiments. Note that the correlation among the finer scales is large (scales $\lambda_1$ and $\lambda_2$), while the correlation among the coarser scales ($\lambda_4$ and $\lambda_5$) is small. As expected, the overall correlation between scales with regions of different sizes is low. This suggests that the use of different scales improves the classification of RSI according to what have been reported in the literature.

Region **A** is related to the low correlation among QCCH-based classifiers and classifiers created using other descriptors. Region **B** refers to the low correlation of ACC-based classifiers with other ones. That suggests that ACC-based classifiers are good candidates to be combined. Region **C** refers to the high correlation observed among the classifiers created with texture descriptors, mainly when finer scales (small regions) are considered. Finally, the region labeled with **D** refers to the high correlation score observed for CCV and GCH descriptors. Classifiers based on those descriptors are *not* good candidates to be combined.

Figure 3 presents the correlation coefficient (see Equation 1) of each pair of descriptors at the segmentation scales $\lambda_1, \ldots, \lambda_5$. Note that the smaller the segmentation scale, the higher the correlation between the

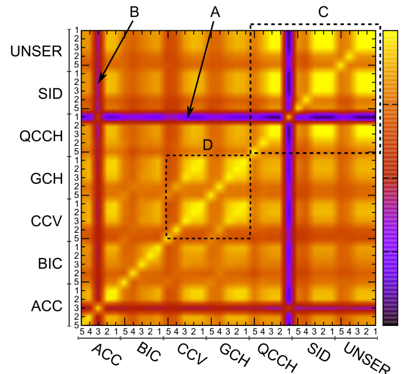


**Figure 2. Complete correlation coefficients for each descriptors at the segmentation scales $\lambda_1, \ldots, \lambda_5$.**

descriptors. The finer scales are composed by more homogeneous and smaller regions. In such scenario, global descriptors as those used in our experiments have less visual patterns to encode. This may be one of the reasons why region-based methods have presented better results than traditional pixel-based classification in the literature when high-resolution RSIs are considered.

In face of the results above, most promising combination would involve the classifiers implemented with color descriptors, at all scales. Some examples are ACC and BIC at $\lambda_4$, and BIC and GCH at $\lambda_5$. With regard to texture descriptors, one should consider only the created classifiers considering scales with large regions.

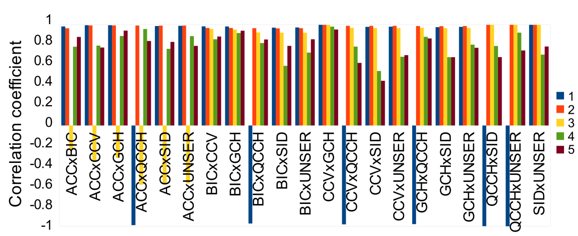


**Figure 3. Correlation of pairs of classifiers for different segmentation scales.**

Finally, with this experiment we can conclude that combining descriptors improves the classification results, but some descriptors contribute more than others and that depends on the scale. Furthermore, we assume that low correlated classifiers are good candidates to be combined as proposed in [2].

### 3.2 Selection of Descriptors

As we observed that not all scales and descriptors have the same contributions and that some classifiers might be very correlated to others, we need to devise a method to select the most promising combination pair (descriptor, classifier).

The simplest idea is to select the most accurate classifiers/descriptors for combination. However, by using only the overall accuracy as the majority of works in the literature, we can have a wrong notion about the results, mainly in binary classification problems. Therefore, we design a simple strategy using two other variables to select classifiers. The first is the tau index, which can be interpreted as a measure of difference to the classification randomly obtained. We used tau because it is more discriminative than the overall accuracy. The other one is the correlation between pairs of classifiers. Correlation gives a notion of diversity that can be used to select classifiers specialized in different kinds of features or subclasses and captures the ones more appropriate to be combined.

Consider a plane where the *x axis* and *y axis* represent the tau index and the correlation of a pair of classifiers, respectively. Let $\mathcal{C}$ be the set of pairs of classifiers in a given scale. The position $P_{(c_i,c_j)}$ of a pair of classifiers $c_i \in \mathcal{C}$ and $c_j \in \mathcal{C}$ on this plane is defined by the ordered pair $P_{(c_i,c_j)} = (Cor(c_i,c_j), \frac{\tau_{c_i}+\tau_{c_j}}{2})$, where $Cor(c_i,c_j)$ is the correlation of classifiers $c_i$ and $c_j$, given by Equation 1, and $\tau_{c_i}$ and $\tau_{c_j}$ are the classification effectiveness measured using the tau index for classifiers $c_i$ and $c_j$, respectively. Both the correlation and the tau index are computed on the validation set. Figure 4 shows the distribution of pairs of classifiers considering the $\lambda_5$ scale for one of the validation sets. Similar distributions are computed for all scales.

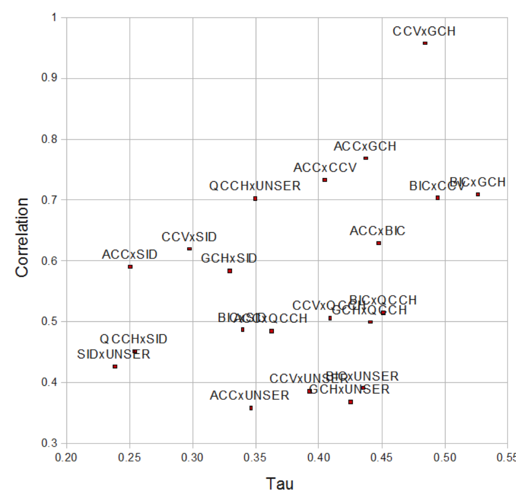


**Figure 4. Distribution of pairs of classifiers considering the $\lambda_5$ scale for one of the validation sets.**

An ideal pair of classifiers should have low correlation and high tau index. Let $\mathcal{P}$ be the position of the ideal pair of classifiers. In our approach, $\mathcal{P} = (1.0, 0.0)$. The set $\mathcal{R}$ of selected pairs of classifiers for a given scale is defined by the K-nearest neighbours of $\mathcal{P}$:

$$K-NN(\mathcal{P}) = \{\mathcal{R} \subseteq \mathcal{C}, |\mathcal{R}| = K \\ \wedge \forall x \in \mathcal{R}, y \in \mathcal{C} - \mathcal{R} : \rho(\mathcal{P},x) \leqslant \rho(\mathcal{P},y)\} \quad (2)$$

where $\rho$ is the distance between two points. In our case, we use the Euclidean distance. We use this strategy with $K = 1$ to select the nearest pair of classifiers to the ideal position for each scale. Since we consider five scales, 10 classifiers are selected for combination.

We perform experiments using the MSC approach to assess the effectiveness of our selection strategy. However, any other method could be used without loss of generalization. The objective is to show that the effectiveness of MSC is the same, when it uses the small set of relevant classifiers selected by our approach. The MSC, which is based on boosting of weak classifiers, defines a weight for each selected classifier along $T$ rounds. The "strong" final classifier is a linear combination of these weak classifiers [5].

Table 1 presents the average overall accuracy (O. A.), Kappa, and Tau measures of the MSC considering the 10 selected classifiers by our strategy ($MSC_{10}$) and using all available classifiers ($MSC_{35}$, five classifiers per scale). One can see that the accuracies are almost the same. The time spent for training MSC, however, are very different. $MSC_{10}$ takes around $9h$, while $MSC_{35}$ takes $16h$. Table 2 shows the weight computed by $MSC$ for weak classifiers selected across the training rounds, considering all classifiers and those 10 found by our selection strategy. As it can be observed, the set of weak classifiers and their weights are almost the same for both configurations.

**Table 1. Classification results using $10$ and $35$ classifiers.**

| Comb. approach | O.A. (%) | Kappa ($\kappa$) | Tau ($\tau$) |
|---|---|---|---|
| $MSC_{10}$ | $82.01 \pm 1.11$ | $0.5475 \pm 0.02$ | $0.6203 \pm 0.02$ |
| $MSC_{35}$ | $82.28 \pm 0.99$ | $0.5587 \pm 0.02$ | $0.6321 \pm 0.01$ |

**Table 2. Weak classifiers chosen by the MSC for each round $t$ considering 10 automatically selected classifiers and all 35 classifiers.**

| | $MSC_{10}$ | | $MSC_{35}$ | |
|---|---|---|---|---|
| | Classifier | Weight | Classifier | Weight |
| 0 | BIC,$\lambda_3$ | 0.73 | BIC,$\lambda_3$ | 0.73 |
| 1 | BIC,$\lambda_5$ | 0.21 | BIC,$\lambda_5$ | 0.21 |
| 2 | Unser,$\lambda_4$ | 0.10 | Unser,$\lambda_4$ | 0.10 |
| 3 | Unser,$\lambda_5$ | 0.02 | GCH,$\lambda_4$ | 0.10 |
| 4 | BIC,$\lambda_5$ | 0.16 | BIC,$\lambda_5$ | 0.16 |
| 5 | ACC,$\lambda_2$ | 0.25 | GCH,$\lambda_5$ | 0.18 |
| 6 | Unser,$\lambda_5$ | 0.08 | ACC,$\lambda_3$ | 0.20 |
| 7 | ACC,$\lambda_1$ | 0.07 | CCV,$\lambda_2$ | 0.15 |
| 8 | BIC,$\lambda_1$ | 0.21 | ACC,$\lambda_5$ | 0.14 |
| 9 | BIC,$\lambda_5$ | 0.12 | GCH,$\lambda_5$ | 0.08 |

# 4 Conclusions

Recent works in the literature have showed that combining different scales can improve RSI classification results. In this paper, we carried out experiments that confirm this, but we showed that not all scales contribute in the same way. Coarser scales offer great power of description while the finer ones can improve the classification by detailing the segmentation. Another branch of studies showed that the use of different descriptors is important. However, the descriptors do not contribute equally at all scales. Finally, in face of the observations we presented a simple approach to choose suitable descriptors/scales for RSI classification by using the correlation measure and the tau index, achieving promising results while spending about half of the time of a fusion approach without a selection policy. Future work includes the analysis of the correlation per accuracy plane with more powerful techniques (e.g., clustering algorithms).

# 5 Acknowledgements

We thank FAPESP (2008/58528-2 and 2010/14910-0), CAPES, and CNPq for the financial support, and Cooxupé for the dataset.

# References

[1] T. Blaschke. Object based image analysis for remote sensing. *Journal of Photogrammetry and Remote Sensing*, 65(1):2 – 16, 2010.

[2] G. Brown and L. Kuncheva. "Good" and "Bad" Diversity in Majority Vote Ensembles. In *Multiple Class. Systems*, volume 5997, chapter 13, pages 124–133. 2010.

[3] J. A. dos Santos, F. A. Faria, R. T. Calumby, R. da S. Torres, and R. A. C. Lamparelli. A genetic programming approach for coffee crop recognition. In *IGARSS*, pages 3418–3421, 2010.

[4] J. A. dos Santos, C. D. Ferreira, R. da S.Torres, M. A. Gonçalves, and R. A. C. Lamparelli. A relevance feedback method based on genetic programming for classification of remote sensing images. *Information Sciences*, 181(13):2671 – 2684, 2011.

[5] J. A. dos Santos, P. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. X. Falcão. Multiscale classification of remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions on*, PP(99):1–12, 2012.

[6] J. A. dos Santos, O. A. B. Penatti, and R. da S. Torres. Evaluating the potential of texture and color descriptors for remote sensing image retrieval and classification. In *VISAPP*, pages 203–208, 2010.

[7] L. Guigues, J. Cocquerez, and H. Le Men. Scale-sets image analysis. *IJCV*, 68:289–317, 2006.

[8] L. I. Kuncheva and C. J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *ML*, 51(2):181–207, 2003.

[9] Z. Ma and R. Redmond. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 61(4):439–453, 1995.

[10] A. Tzotsos, K. Karantzalos, and D. Argialas. Object-based image analysis through nonlinear scale-space filtering. *ISPRS*, 66(1):2–16, 2011.