# Land Use Classification with Compressive Sensing Multifeature Fusion

Mohamed L. Mekhalfi, *Student Member, IEEE,* Farid Melgani, *Senior Member, IEEE,* Yakoub Bazi, *Senior Member, IEEE,* Naif Alajlan, *Senior Member, IEEE*

**Abstract—In this letter, we formulate the land use classification problem within a compressive sensing fusion framework. Compressive sensing (CS) aims at providing a compact representation form after a given query image has been processed with an opportune feature extraction type. In particular, residuals are generated from the image reconstruction with dictionaries associated with the available set of possible land uses and gathered to form a single-feature image pattern. The patterns obtained from different types of features are then fused to provide the final land use estimate. Two simple fusion strategies are adopted for such purpose. As demonstrated by experiments ran on the basis of a public benchmark database, the proposed method can achieve substantial classification accuracy gains over reference methods.**

***Index Terms*—Compressive sensing, co-occurrence of adjacent local binary patterns, data fusion, gradient local auto-correlations, histogram of oriented gradients, land-use classification.**

## I. INTRODUCTION

Land use (LU) classification from remote sensing images represents an important but challenging task. *Important* because it reflects the socio-economic activities of a given territory. *Challenging* since traditional low-level representations are not adapted for performing a high-level semantic description as required in LU classification.

As reported by the literature, several important contributions can be noted. In [1], for instance, a problem of 21 class LU image classification is considered, where three bag-of-visual-words-based (BOVW) schemes are proposed. The first one is a basic BOVW model, where a so-called codebook of visual words (of a limited size) is constructed by relying on the adopted set of gallery images, and then the occurrence of each visual word is observed in each target image, which turns out to produce an image signature consisting of frequencies of all the codebook elements. As the standard BOVW overlooks the spatial information, two other variants have also been considered, namely (i) spatial pyramid match kernel, whose underlying insight is to seek for approximate correspondences between set of points in high dimensional feature space by splitting this latter into progressively coarser grids, and (ii) spatial co-occurrence kernel, which derives benefits from the relative arrangement of image features. As for image features, the well-known Scale Invariant Feature Transform (SIFT) was opted for. Experiments conducted on the basis of a LU database have revealed that comparable or even improved performances can be envisioned while moving from the standard BOVW to the enhanced variants. Another extension, termed spatial pyramid co-occurrence, is suggested in [2], where the BOVW histogram involves both absolute and relative spatial arrangements of the visual words whose co-occurrence is counted over a spatial partitioning of the target image. This scheme has scored an improvement over the non-spatial BOVW. Once again, in order to overcome the fact that the standard BOVW neglects the spatial information, a model noted as pyramid-of-spatial-relatons (PSR) is presented in [3] for LU scene classification. For a considered image, the PSR performs over image sub-regions, which are obtained by successively partitioning the image into multiple regions, where each sub-region is represented by an orderless ensemble of spatial relaton histograms (i.e., a quantized local feature histogram). The PSR model captures both absolute and relative spatial relationships of local features by incorporating the BOVW signature and the spatial relaton pertaining to each image sub-region. Moreover, the PSR has shown robustness to rotation and translation. Validated on a LU database, notable ameliorations have been scored over the state-of-the-art. In [4], another variant of the BOVW is investigated under a LU classification perspective. The suggested model, named concentric circle-structured multiscale BOVW, is a six-pronged scheme. At first, multiple local features are extracted from multiple resolution images (constructed starting from the original image). Afterwards, different visual vocabularies are produced by means of the K-means algorithm for different resolutions. After that, each resolution image is divided into a number of annular subregions by an ensemble of concentric circles, where for each one a histogram of visual words is generated on the basis of the respective visual vocabulary. Subsequently, for each resolution image, all generated histograms are concatenated together. The final signature of the target image is then shaped by a further concatenation of the previously created resolution histograms. The latter image representations are finally fed into a Support Vector Machine (SVM) classifier as to finalize the decision making. Interesting results have been demonstrated on public land-use datasets. Another interesting remotely sensed image classification scheme is introduced in [5], where a multi-index learning method has been suggested as to infer low-dimensional representations derived out of complex scenes, and proved as efficient to the alternative common high-dimensional feature spaces. Multifeature fusion has also been a means as to take

Manuscript received xxx xxx, xxx.
This work was supported by the Deanship of Scientific Research of the King Saud University through the International Research Group under Project IRG14-20.

M. L. Mekhalfi and F. Melgani are with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: {mohamed.mekhalfi, melgani}@disi.unitn.it).
Y. Bazi and N. Alajlan are with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia (e-mail: {ybazi, najlan}@ksu.edu.sa).

advantage of all information latent within the remotely sensed images. In particular, in [6], both spectral and spatial features have been coupled into one single model, and demonstrated a good performance.

In this context, it can be made mention of the fact that, from the works cited above, the common denominator for image representation is the BOVW model, which indeed has proven effective in the LU context. On the other side, image representation and matching represents a research topic where there is still a plenty of room for further improvement. It is also a matter of fact that the interaction between different research disciplines has pointed out that substantial benefits can be achieved, take for instance the example of Compressive Sensing (CS) theory, which has emerged as a mainstream in information theory in recent years [7]-[8], and ever since has been adopted in various applications related mainly to computer vision [9], remote sensing [10], speech recognition [11], source separation [12], and assistive technologies [13]. CS appears particularly interesting for its capacity of representing a given signal/image in a remarkably compact manner.

In this letter, we formulate the LU image classification problem in a CS fusion framework. In greater detail, given a gallery of labeled (training) images, as many dictionaries as the number of classes are first set up. After an appropriate feature extraction phase which attributes a fixed length pattern (representation) to each image, the pattern of a given probe image is reconstructed on the basis of each dictionary, which turns out to generate a scalar reconstruction residual for each class. Hence a sequence concatenating all the residuals can be formed to represent the probe image. if *N* different types of features are adopted, *N* residual sequences are inferred and then normalized. The ultimate step is confined to fusing the *N* residual sequences for final decision as illustrated in the next section.

The remaining part of this letter is formatted as follows. Section 2 details the proposed approach, sheds light on the CS theory, and explains the adopted fusion mechanisms. Section 3 goes through experimental setup and results. Section 4 draws main conclusions, and elaborates future directions of the present work.

## II. METHOD DESCRIPTION

The implemented framework addresses the case of image classification contextualized within a LU perspective. The objective is to assign a class label to a given query image by making use of a labeled library beforehand prepared. On the one hand, such issue can be viewed as an image matching problem where the probe image is said to very likely hold the same label as the closest gallery image. In this context, as the images dealt with may convey different rotation, scale, and illumination acquisition conditions, comparing the images in their spectral form does not appear the best strategy to follow. This implies the need of adopting a kind of features that can deal, at least partially, with such issues. On the other hand, even a sophisticated feature type might not be sufficiently effective to cope with the given image classification task. In sum, we think that a salient kind of features incorporated with a sound classification scheme is ought to adequately deal with the LU case. For such purpose, in this letter, we propose a new framework which represents an image through an ensemble of compressive sensing encodings originating from different kinds of features and opportunely fused.

Let *I* be a probe image to be labeled. Let us assume available an already labeled gallery of images where those belonging to the same class are collected into a single group. Let *C* and *N* be the numbers of classes and feature types adopted for image representation, respectively. Each feature type (amongst the *N* adopted ones) is extracted from the probe image and then converted into a vector (likewise for the gallery images, for which the vectors are computed and stored offline in the form of a matrix). Next stage takes as input each probe feature vector and performs a CS-based reconstruction out of all the *C* gallery dictionaries (i.e., *C* reconstructions are performed). By performing a CS reconstruction of the probe pattern over all the dictionaries, as many residuals (scalars) as the number of classes will be generated, forming thereby a sequence of residuals denoted $R_{ij}$ $(i = 1 \dots N, j = 1 \dots C)$. The smallest quantity within $R_{ij}$ over *j* indicates the estimated class of the probe vector associated with the *i*th feature. For the *N* features, *N* residual sequences will thus be produced. In order to infer a global decision among the *N* feature types, a residual fusion layer is applied. The pipeline of the entire framework is illustrated in Fig. 1.

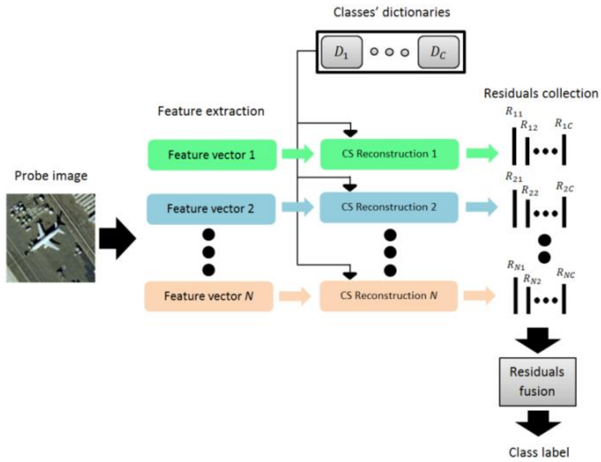


Figure 1. Global flowchart of the proposed classification scheme.

### *A. Compressive Sensing*

Throughout the explanations driven so far, as well as from Figure 1, it can be figured out that CS occupies a prime position in our work. The rationale standing behind such choice is that CS can be exploited to compactly represent a given pattern through the dictionary concept. That is, out of numerous images of a certain class, only one scalar (i.e., the residual) conveying the capacity of that particular class in terms of recovering the probe sample is obtained. Compressive sensing was recently introduced by Donoho [7], and Candès [8]. It aims at recovering an unknown sparse signal from a set of linear projections. Note that images hold a natural sparse representation with respect to a so-called dictionary (e.g., Fourier, wavelet) [14]. The fundamental aspect of the CS theory stems for its capability to sparsely represent any given measurement vector $V = D \cdot \alpha$ by solving the following L0-minimization problem:

$$\min \|\alpha\|_0 \quad \text{subject to } V = D \cdot \alpha \tag{1}$$

where *D* is a dictionary with a certain number of atoms (which in our case are image representations converted into vectors), $V$ is the input image representation (converted into vector) that can be represented as a sparse linear combination of these atoms, $\alpha$ is the set of coefficients intended as a compact representation of $V$.

In the literature, there exist several algorithms for solving the optimization problem expressed above. In the following, we briefly introduce an algorithm called stagewise orthogonal matching pursuit (StOMP) [15], which will be exploited in our work. By contrast to the basic orthogonal matching pursuit (OMP) algorithm, StOMP involves many coefficients at each stage (iteration) while in OMP only one coefficient can be involved. Additionally, StOMP runs over a fixed number of stages, whereas OMP may take numerous iterations. Hence, StOMP exhibits the advantage of a fast computation capability. The underlying StOMP routine is detailed below:

**Step 1:** Consider an initial solution $\alpha_0 = 0$, an initial residual $r_0 = V$, a stage counter $s$ set to 1, and an index sequence denoted as $T1 \ldots Ts$, which contains the locations of the non-zeros in $\alpha_0$.

**Step 2:** Compute the inner product between the current residual and the considered dictionary $D$:

$$C_s = D^t . r_{s-1} \tag{2}$$

$D$ collects all the training feature vectors (of the considered class), which are placed column-wise. $D^t$ is its transpose. $C_s$ refers to the CS coefficients. Its length is equal to the number of training vectors conveyed in $D$.

**Step 3:** Perform a hard thresholding in order to find out the significant non-zeros in $C_s$ by searching for the locations corresponding to the 'large coordinates' $J_s$:

$$J_s = \{J : \alpha_s(J) > t_s \sigma_s\} \tag{3}$$

where $\sigma_s$ represents a formal noise level, and $t_s$ is a threshold parameter taking values in the range $2 \leq t_s \leq 3$.

**Step 4:** Merge the selected coordinates with the previous support.

$$T_s = T_{s-1} \cup J_s \tag{4}$$

**Step 5:** Project the vector $V$ on the columns of $D$ that correspond to the previously updated $T_s$. This yields a new approximation $\alpha_s$:

$$(\alpha_s)_{T_s} = (D^t_{T_s})^{-1} D^t_{T_s} V \tag{5}$$

**Step 6:** Update the residual according to: $r_s = V - D.\alpha_s$ (6)

**Step 7:** Check whether a stopping iterative condition (e.g., *smax*=10) is met. If so, $\alpha_s$ is considered as the final solution and as the residual left out of the recovery process. Otherwise, the stage counter s is incremented and the process is repeated from *Step 2*.

In our work, we rely on the amount defined as the sum of the elements of the residual vector, which is a scalar quantity highlighting how much the probe pattern $V$ is reconstructed over the dictionary $D$. Coming back to the generalized version of the problem (i.e., multiple dictionaries), then as many scalars (residuals) as the number of dictionaries are obtained. Subsequently, the class of the probe pattern refers to the dictionary that points out the smallest residual.

The previous CS classification routine is confined to the case of only one type of feature. In our contribution, however, we address the case of $N$ features in order to boost the classification process. Thus, $N$ residual vectors are generated and combined within a simple fusion process. Suggested fusion strategies are outlined below.

### *B. Sum Operator-based Strategy (SOS)*

From the previous section, it is to point out that in the case of $N$ features, $N$ residual sequences are generated. The linear residual fusion strategy starts by normalizing, for each feature type, all the $R_{ij}$ arrays so that the corresponding maximum peak would be at the unity. Such normalization step is important as the residual values pertaining to a certain kind of features might overweigh the remaining residuals in the fusion process. Afterwards, all the $R_{ij}$ vectors are linearly summed up to form a single sequence of the same length. Finally, the class of the probe pattern corresponds to the lowest residual of the final sequence.

$$R_j = \sum_{i=1}^{N} R_{ij} \quad \text{and} \quad class = \underset{j}{\operatorname{argmin}}(R_j) \tag{7}$$

### *C. Majority-based Operator Strategy (MOS)*

In this strategy, from each of the $N$ sequences of residuals, a decision is made by choosing the class corresponding to the smallest residual of the considered sequence. Hence, as many decisions (classes) as the number of sequences are produced. Next step is to pick up the class label that is the most frequent amongst the $N$ made decisions and assign it as a final class of the probe pattern. When a tie occurs, we chose the final class as that of the smallest residual amongst the classes in conflict.

## III. EXPERIMENTAL VALIDATION

In order to evaluate the proposed classification method, we exploited the UC Merced LU database, which was made available by Yang and Newsam [1]-[2]. The dataset was manually derived from another dataset of large aerial orthoimagery with about 30 cm of pixel resolution. It was downloaded from the USGS National Map of the following US regions: Birmingham, Boston, Buffalo, Columbus, Dallas, Harrisburg, Houston, Jacksonville, Las Vegas, Los Angeles, Miami, Napa, New York, Reno, San Diego, Santa Barbara, Seattle, Tampa, Tucson, and Ventura. It contains 2100 images, of 256×256 pixels each, categorized into 21 classes (100 images per class). The class labels are as follows: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis court. Samples of the LU database are depicted in Fig. 2.

The set of features adopted in our work consists of three types, namely:

(i) Histogram of Oriented Gradients (HOG): This kind of feature has gained a sound reputation in computer vision owing to its capacity of comprehensively describing an image through its local gradients [16]-[17].

(ii) Co-occurrence of Adjacent Local Binary Patterns (CoALBP): This is a variant to the popular LBP. The CoALBP covers spatial information of all the LBPs pertaining to the given image, and takes into account their co-occurrence. The CoALBP has shown richer representation and a robustness to illumination change [18].

(iii) Gradient Local Auto-Correlations (GLAC): This type of feature, additionally to the gradient information, also conveys the behaviour of a given image surface in terms of curvatures. Thereupon, it has demonstrated improvement over other reference features such as for instance the HOG [19].

These three different feature types are considered in our work due to their sound discrimination capability as motivated in [16]-[19].

For the sake of consistency with reference works, we articulate the evaluation upon a 5-fold cross validation, where the database is randomly split into 5 folds, each conveying 20 images per class. One held-out subset is used as probe (test), and the remaining four subsets are employed as gallery (training images). This process is performed five times, and their average is considered as the final classification accuracy, which is defined as the ratio of the correctly classified samples to the total number of test samples.

We first report the results obtained by using each type of feature individually, thus the fusion process is not involved at this point. Table 1 summarizes the results. The lowest rate was observed for the HOG features, scoring 68.67 %, followed by the GLAC pattern by yielding a raise of about 8 % over the first one. The best accuracy of 80.52 % was recorded for the CoALBP features. It can be noticed that the performances are different among the three features, which might be referred to the rationale that the images of the same class reveal various orientations and scale changes, not to mention the total number of classes (i.e., 21 categories), which is subject to raise a large within-class and a low inter-class variability.

With regards to the fusion of the previous features by means of the two simple strategies described above, we detail the results in Table 1. It appears that SOS exhibits a better accuracy (94.33 %) than MOS (87.95 %). First, as compared to relying on only one kind of feature, it is noteworthy that the fusion of multiple features by means of the CS representation has led to a drastic improvement. Second, the lower accuracy of MOS compared to SOS can be explained by the fact that the former disregards partial information (i.e., the residuals) used for the individual decisions on which it relies, and exploits only the final decisions (class labels) from the single classifiers. In other words, in our experiments, MOS combines 3 information (class labels) while SOS merges 3×21=63 information (residuals) to infer the fusion outcome. Exploiting all available information in the decision has thus emerged the best way to proceed with the fusion as performed by SOS.

As for comparing the proposed scheme with reference state-of-the-art methods on the same dataset, a number of interesting works is taken into account [1], [3], and [4]. They all opted for a similar 5-fold-based evaluation, except for [4], where a 2-fold validation was adopted. We therefore ran the experiments considering this latter validation too. The comparison is provided in Table 2. As shown, the proposed method resulted in sharply superior accuracies than all the considered works, where the lowest performance is observed for [1]. Our method exhibits two advantages over these reference methods: 1) the representation is very compact (reduced to 21, namely the number of classes) while the reference methods rely on the BOVW which produces representations with hundreds/thousands of bins; and 2) in our case, the reconstruction residuals represent already information from which decision can be inferred (normalized residuals can be viewed as one minus posterior probabilities, the smaller the residual the larger the posterior) while the other methods need to train a classifier (support vector machine) in a huge input space incurring in problems of curse of dimensionality because of the reduced number of training images per class. These two aspects mainly explain why our method outperforms the reference methods.

For additional comparison, we implemented two other fusion strategies. The first one is based on feeding a support vector machine (SVM) classifier with a vector concatenating the original three feature vectors. The second technique is still based on feature concatenation but, this time, transforms the concatenated feature vector through a single CS representation. The wining class is the one exhibiting the smallest residual. The classification results achieved with a 5-fold validation procedure are reported in Table 3, from which it emerges clearly that the proposed SOS strategy outperforms the two concatenation-based fusion techniques (Conc-SVM and Conc-CS).

Figure 3 depicts an example of image classification based on the SOS strategy. In this example, the test image belongs to the first class ‘agricultural’. It can be noticed that, for all the three adopted features, the lowest residual points correctly to the first class, which is likewise observed in the final SOS-based residual. Another point to stress in this example is that the GLAC feature yields a residual vector of undesirable shape as compared to the other two features. Indeed, the ideal case would be a residual vector where the minimum pertains to the correct class while the remaining part is much larger and uniformly distributed. The SOS fusion permits to compensate this issue by generating a global residual vector close to the desired shape leading to a correct but also more confident decision.

Table 1. Classification accuracies achieved by the single feature-CS representations and by the CS fusion strategies.

| | Feature Typology | | | Fusion | |
|---|---|---|---|---|---|
| | HOG | CoALBP | GLAC | SOS | MOS |
| Acc(%) | 68.67 | 80.52 | 77.1 | 94.33 | 87.95 |

## IV. CONCLUSION

This letter puts forward a novel classification scheme within the context of LU image classification. The underlying idea is to represent compactly images within a CS and a multifeature framework. The CS reconstruction residuals originating from different kinds of features are fused based on two simple but efficient strategies. Indeed, as the results point out, promising outcomes have been obtained. Furthermore, the proposed method exhibits very substantial classification accuracy gains over reference methods.

Nevertheless, we believe that the current version of the algorithm can undergo further sophistication. For instance, we recommend the investigation of more elaborated fusion methods such as the induced ordered weighted averaging (IOWA) operators [20]. The second element is related to the size of the input feature vectors, which can be beforehand reduced while maintaining or even improving the classification performance. Fisher discriminant analysis could represent a good candidate for such purpose [21]. Thirdly, raising the number of feature types in the ensemble (in this work we considered just three kinds) might be another ingredient of amelioration. Finally, reformulating the proposed method under an active learning perspective could also be another interesting way to explore [22].

## ACKNOWLEDGMENT

The Authors would like to thank D. Donoho and Y. Tsaig for providing the SparseLab toolbox [23].


## V. REFERENCES


[1] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in Proc. 18th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst., San Jose, CA, USA, pp. 270–279, 2010.

[2] Y. Yang and S. Newsam, "Spatial pyramid co-occurrence for image classification," in Proc. IEEE Int. Conf. Comput. Vis., pp. 1465–1472, 2011.

[3] C. Shizhi, and Y. Tian. "Pyramid of Spatial Relatons for Scene-Level Land Use Classification," IEEE Trans. Geosci. Remote Sens., vol. 53, no. 4, pp. 1947-1957, 2015.

[4] L. Zhao, P. Tang, and L. Huo. "Land-Use Scene Classification Using a Concentric Circle-Structured Multiscale Bag-of-Visual-Words Model," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, in press, 2015.

[5] X. Huang, Q. Lu, L. Zhang, "A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas," ISPRS Jour. of Photogr. Remote Sens., vol. 90, pp. 36-48, 2014.

[6] X. Huang, L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 1, pp. 257-272, 2013.

[7] D. L. Donoho, "Compressed Sensing," IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289-1306, 2006.

[8] E. J. Candès, J. Romberg, T. Tao, "Robust Uncertainty Principle-Exact Signal Reconstruction from Highly Incomplete Frequency Information,"IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489-509, 2006.

[9] J. Wright, A. Yang, S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, 2009.

[10] L. Lorenzi, F. Melgani and G. Mercier, "Missing Area Reconstruction in Multispectral Images Under a Compressive Sensing Perspective", IEEE Trans. Geosci. Remote Sens., vol. 51, pp. 3998 - 4008, 2013.

[11] J. Gemmeke, T. Virtanen, and A. Hurmalainen, "Exemplar-based sparse representations for noise robust automatic speech recognition," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 7, pp. 2067–2080, 2011.

[12] Z. S. He, S. L. Xie, S. X. Ding, and A. Cichocki, "Convolutive blind source separation in the frequency domain based on sparse representation", IEEE Trans. Audio, Speech, and Language Process, vol. 15, no. 5, pp. 1551-1563, 2007.

[13] M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan, "A Compressive Sensing Approach to Describe Indoor Scenes for Blind People", IEEE Trans. Circuits Systems for Video Tech., vol. 25, in press, 2015.

[14] E. J. Candès, T. Tao, "Decoding by Linear Programming", IEEE Trans. Inform. Theory, vol. 51, no. 12, pp. 4203-4215, 2005.

[15] D. L. Donoho, Y. Tsaig, , I. Drori, , J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," IEEE Trans. Information Theory, vol. 58, no. 2, pp. 1094-1121, 2012.

[16] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 886-893, 2005.

[17] T. Moranduzzo, and F. Melgani, "Detecting Cars in UAV Images With a Catalog-Based Approach," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 10, pp. 6353-6367, 2013.

[18] R. Nosaka, Y. Ohkawa, and K. Fukui. "Feature extraction based on co-occurrence of adjacent local binary patterns," In Advances in image & video technology, pp. 82-91, 2012.

[19] T. Kobayashi, N. Otsu, "Image Feature Extraction using Gradient Local Auto-Correlations," in Proc. 10th European Conf. on Computer Vision, Part I. LNCS, vol. 5302, pp. 346-358, 2008.

[20] Y. Bazi, N. Alajlan, F. Melgani, H. AlHichri, and R. Yager. "Robust Estimation of Water Chlorophyll Concentrations With Gaussian Process Regression and IOWA Aggregation Operators, " IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 7, pp. 3019-3028, 2014.

[21] H. Huang, H. Feng, C. Peng, "Complete local Fisher discriminant analysis with Laplacian score ranking for face recognition", Neurocomputing, vol. 89 , pp. 64-77, 2012.

[22] B. Demir and L. Bruzzone, "A Novel Active Learning Method in Relevance Feedback for Content-Based Remote Sensing Image Retrieval," IEEE Trans. Geosci. Remote Sens., vol. 53, in press, 2015.

[23] Available at: http://sparselab.stanford.edu/



Figure 2. Image samples from the UC Merced LU dataset.

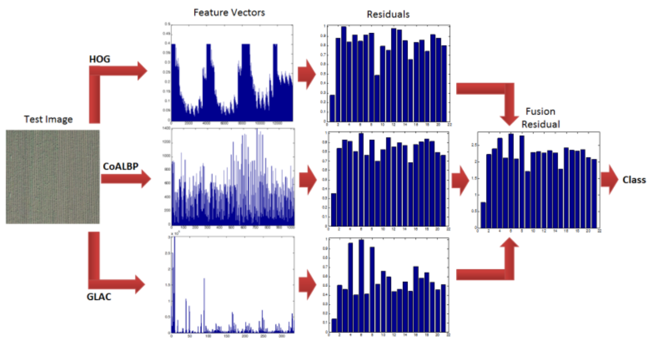


Figure 3. Image classification example showing the input feature vectors, the corresponding CS residual vectors and the fusion-based residual vector.

Table 2. Comparison against state-of-the-art methods.

| Work | Description | Acc (%) | Validation |
|---|---|---|---|
| Yang and Shawn [1] | Spatial BOVW | 81.19 | 5-Fold |
| Chen and Tian [3] | Pyramid of Spatial Relatons | 89.1 | 5-Fold |
| Proposed Method | CS Multifeature Fusion | **94.33** | 5-Fold |
| Zhao, Tang, and Hou [4] | Concentric Multiscale BOVW | 86.64 | 2-Fold |
| Proposed Method | CS Multifeature Fusion | **91.1** | 2-Fold |

Table 3. Comparison with two concatenation-based fusion strategies.

| | Conc-SVM | Conc-CS | SOS |
|---|---|---|---|
| Acc (%) | 77.52 | 81.18 | 94.33 |