New MPEG-7 Scalable Color Descriptor Based on Polar Coordinates for Multispectral Earth Observation Image Analysis

Florin-Andrei Georgescu, *Member, IEEE*, Dan Răducanu, and Mihai Datcu, *Fellow, IEEE*

***Abstract*— Continuously expanding high-resolution and very high resolution multispectral image collections, provided by remote sensing satellites, require specific methods and techniques for data analysis and understanding. Even though there are several patch-based approaches for image classification and indexing, none of them are integrated within a standard. Having the goal to develop an MPEG-7 compliant descriptor for patch-based multispectral earth observation image classification and indexing, we propose a new feature extraction method able to extract maximum information from all the available spectral bands that Sentinel 2, the last generation of remote sensing satellites, provides. Using the polar coordinate transformation of the reflectance values, we obtain illumination invariant features, which can be used along with the scalable color descriptor present in MPEG-7 standard. Also, our method proves to enhance land cover classification of the areas affected by clouds and their shadows and provide similar classification results compared with the homogeneous texture descriptor (HTD), spectral histogram (SH), concatenated HTD with SH features, spectral indices (SIs), and bag-of-words-based descriptors, such as bag-of-SIs and bag-of-spectral-values on cloud-free areas.**

***Index Terms*— Feature extraction (FE), land cover classification, MPEG-7, polar coordinates, scalable color descriptor (SCD).**

I. Introduction

DUE to large earth observation (EO) image data collections, many applications for satellite image classification and indexing have emerged. No matter the scope of these applications, the main processing steps rely on FE, feature selection, and feature classification (FC).

Inspired from multimedia image processing, most of the available FE and FC algorithms for EO image analysis and understanding are specialized in detecting specific image properties such as texture, color, and shape. Despite the diversity of proposed methodologies for remote sensing image classification and indexing, there is no general approach that can be used independently on the input data.

Manuscript received June 22, 2016; revised December 5, 2016; accepted December 29, 2016. Date of publication May 12, 2017; date of current version June 22, 2017. *(Corresponding author: Florin-Andrei Georgescu.)*

F.-A. Georgescu is with the "Politehnica" University of Bucharest, Applied Electronics and Information Engineering, 060042 Bucharest, Romania, and also with the Military Technical Academy, Geomatics Department, GEOINT Center, 050141 Bucharest, Romania (e-mail: florin.andreig@gmail.com).

D. Răducanu is with the Military Technical Academy, Geomatics Department, GEOINT Center, 050141 Bucharest, Romania.

M. Datcu is with the German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), 82234 Wessling, Germany, and also with the "Politehnica" University of Bucharest, Applied Electronics and Information Engineering, 060042 Bucharest, Romania.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2017.2656822

Many authors have developed FE methods sensitive to the spatial dependency of the pixels that can be used in image interpretation, segmentation, classification, and change detection applications. Some of these texture analysis techniques used in EO image analysis are based on gray-level cooccurrence matrix [1], wavelet transforms [2], Gauss–Markov random fields [3], and Gabor filtering [4]. Other methods rely on statistical text modeling approaches, such as the author–topic model [5] and author–genre–topic model [6], in which a latent Dirichlet allocation is involved.

In other scenarios, a closer look to the spectral components of the analyzed image should be taken into account. Thus, new methods for EO image data analysis based on spectral values [7], spectral indices (SIs) [8], spectral angles [9], spectral histograms (SHs) [10], and color moments [11] have emerged. These image features rely on the pure spectral information of the analyzed data, are easy to compute, and can be used in a wide range of remote sensing applications for scene-classification- and content-based image retrieval.

For EO data analysis, especially for multispectral images, a good FE method is not enough. Due to the high radiometric resolutions of the last-generation satellite sensors, classical distance metrics used in the classification process may not return the desired results. Thus, in [12], it is explained how the Euclidean distance, including transformed and weighted forms such as Mahalanobis and likelihood distances, is inherently insensitive to the shapes of the spectral pattern and that the cosine of the angle θ (spectral angle) provides a better definition of "similarity" due to its invariant nature to the linearly scaled variations.

In the context of angular metrics space, it is advisable to use polar coordinates instead of rectangular coordinates space. This approach can be helpful if we desire to obtain illumination invariant features.

In the MPEG-7 standard, also known as the multimedia content description interface, is described a suite of image descriptors for multimedia image processing [13] that has great potential in EO image analysis and understanding. Despite multimedia images, when dealing with EO data, the objects are represented using several spectral bands that equally influence the classification process. In order to compute MPEG-7 image descriptors on multispectral images, we need to apply a feature space transformation. Relying on [14], we convert the radiance values of each spectral band into polar coordinates equivalent, obtaining this way an illumination invariant feature space that

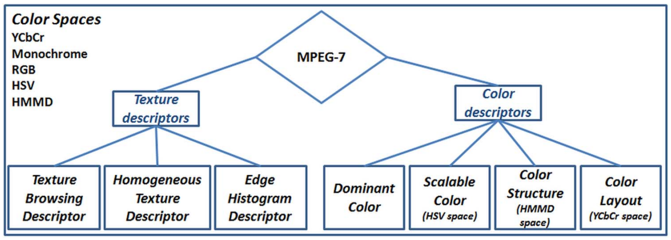

Fig. 1. MPEG-7 color and texture descriptors.

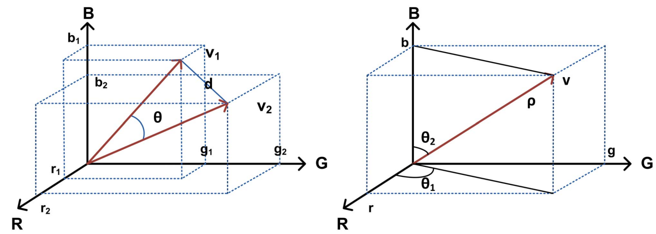

Fig. 2. (Left) Angular versus Euclidean distance in feature space. (Right) Polar versus Cartesian representation of R-G-B feature space.

can be used in the MPEG-7 scalable color descriptor (SCD) computation.

We propose an illumination invariant descriptor based on polar coordinate transformation of the spectral values. The polar feature space we create will provide the necessary support to compute in a high-dimensional spectral space the SCD described in the MPEG-7 standard.

Being motivated to adapt multimedia descriptors of the MPEG-7 standard to handle multispectral EO images, we assessed our polar-coordinates-based SCD (pSCD) for the EO land cover classification case. In order to provide relevant results, the suggested method is tested on different scenes acquired from the recently released Sentinel-2 satellite. Similar outcomes can be obtained with other multispectral satellite image data. Also the results are compared with other image descriptors such as the homogeneous texture descriptor (HTD), weber local descriptor, SH, SIs, and BoW-based descriptors, such as bag-of-SIs (BSI), and bag-of-spectral-values (BSV).

Section II presents the MPEG-7 standard descriptors and details the logic to perform the polar coordinate image transformation and integration with the SCD. The performance assessment of the proposed descriptor in various situations is presented in Section III, and Section IV is reserved for the main observations regarding the proposed pSCD FE method.

II. Earth Observation Image Understanding Using MPEG-7

It is known that the MPEG-7 standard, formally named the "multimedia content description interface," aims to provide methods and techniques of describing the multimedia content data and should support a high degree of interpretation of the information meaning in order to be accessed by a device or a computer code [13]. Although it was designed to support a wide range of multimedia applications, for both audio and visual data, this standard faces a great challenge when applied to remote sensing imagery.

In MPEG-7, the color descriptors provided are optimized for multimedia image processing and are usually applied on a perceptual color space and consist of a number of histogram descriptors, a dominant color descriptor, and a color layout descriptor [15]. On the other hand, the texture descriptors proposed in the standard, the texture browsing descriptor, HTD and local edge histogram descriptor, follow attributes such as directionality, regularity, coarseness, and homogeneity [15].

As shown in Fig. 1, MPEG-7 visual descriptors are designed to handle color and monochrome images. In the case of EO image information mining, where more than three spectral bands are involved, the default configuration of the standard MPEG-7 descriptors could not be used. Our interest is focused on the visual descriptors provided by the MPEG-7 standard, for both texture and color understanding and analysis, being highly motivated to adjust descriptors provided in this standard for multispectral EO data analysis and understanding.

In quest of adapting MPEG-7 descriptors to handle multispectral EO images, we propose a polar coordinate transformation of the color space instead of using the HSV color space. This will enable us to consider the full spectral resolution available in the EO data sets and obtain an illumination invariant image descriptor.

A. Polar Coordinate Image Transformation

Most of the techniques used in the multispectral image analysis and understanding are based on identifying the general characteristics of the pixels or image patches being analyzed. One of the conventional ways of representing multispectral data is to plot the image features in a multispectral vector space with as many dimensions as there are spectral components [14]. Considering the multispectral features space as a reference system, Okamura *et al.* [14] introduce an improved multispectral image analysis using polar coordinates.

The approach presented in [12] relies on the spectral angular distance for measuring distances in feature space and computing multispectral image classification and clustering. The angular distance is proved invariant to linearly scaled variations of the image features. If two spectral signatures defined by the feature vectors V_1 and V_2 are considered, the spectral distance (Euclidean distance) can be expressed using (1) and can represent the length of the segment $d = V_1V_2$ that connects the end points of the two vectors, while in the case of spectral angular distance, (1) is given by the angle θ between the two vectors as can be observed in Fig. 2

$$d = \sqrt{(r_2 - r_1)^2 + (g_2 - g_1)^2 + (b_2 - b_1)^2} \quad (1)$$

$$\theta = \arccos \frac{\overrightarrow{V_1} \cdot \overrightarrow{V_2}}{\| \overrightarrow{V_1} \| \cdot \| \overrightarrow{V_2} \|}. \quad (2)$$

Using the polar coordinate transformation of the image features (see the right of Fig. 2), Okamura *et al.* [14] have succeeded to obtain an illumination invariant descriptor. This is proved to be of high importance when classifying regions with powerful shadow and cloud coverages, a common phenomenon in multispectral EO images. Thus, the use of polar coordinate transformation will place patches of the same class that have different illuminations in the same sectors defined by the distance ρ and the spectral angles θ.

In our approach, we convert the radiances values of each spectral band from the multispectral image into polar coordinates and characterize the image through a distance ρ and $N-1$ angles θ as presented in the following equations:

$$\rho = \sqrt{x_N^2 + x_{N-1}^2 + \cdots + x_2^2 + x_1^2} \tag{3}$$

$$\theta_1 = \arctan \frac{\sqrt{x_N^2 + x_{N-1}^2 + \cdots + x_2^2}}{x_1} \tag{4}$$

$$\theta_2 = \arctan \frac{\sqrt{x_N^2 + x_{N-1}^2 + \cdots + x_3^2}}{x_2} \tag{5}$$

$$\cdots$$

$$\theta_{N-2} = \arctan \frac{\sqrt{x_N^2 + x_{N-1}^2}}{x_{N-2}} \tag{6}$$

$$\theta_{N-1} = 2 \arctan \frac{x_N}{x_{N-1} + \sqrt{x_N^2 + x_{N-1}^2}} \tag{7}$$

where N represents the number of the spectral bands present in the multispectral image, while the terms $x_1, x_2, \ldots, x_N$ are referring to the radiance values of the band i, where $i = \overline{1..N}$.

Using the conversion of the radiance values into polar coordinates, we aim to create a synthetic multispectral space that can be used in the SCD computation, as presented in the MPEG-7 standard.

B. Scalable Color Descriptor From Polar Coordinates

When dealing with multispectral images, with more than three spectral bands, the computation of HSV space is impossible. In order to compute SCD for multispectral images, we propose the polar coordinate transformation to obtain an HSV compliant color space on which we can extract SCD as presented in the MPEG-7 standard. As described in [15], we consider a uniform quantization of the polar coordinate color space to 256 bins. Furthermore, the Haar transform is applied.

Using polar coordinate transformation instead of HSV color space, we obtain an illumination invariant descriptor that can use full spectral resolution of the analyzed image. Also, our approach is aiming to create a new FE method that can be integrated within a standard. The proposed method can be computed on any type of multispectral remote sensing image.

III. Evaluation and Discussion

A. Experimental Setup

To demonstrate our pSCD, we used scenes from Sentinel-2 multispectral sensor acquired at different moments of time. The footprint of the analyzed images is over two different regions of Romania. Also, the evaluated scenes have different cloud coverage in order to assess how pSCD is handling in various situations. One of the tested scenes shown in Fig. 3 was acquired on December 23, 2015, covering a surface of 2500 km^2 over Bucharest, Romania, while the other was acquired on December 1, 2015, having a surface of 6400 km^2 with a cloud coverage of approximatively 39%, as can be seen in Fig. 6.

TABLE I
FE and FC Speeds With Associated Accuracies

	FE speed (patches/sec)	FC speed (patches/sec)		Accuracy	
		k-NN	***SVM***	***k-NN***	***SVM***
HTD	9.99	491.16	162.76	81.12%	87.42%
SH	1191.90	443.66	135.61	83.71%	84.76%
HTD_H	55.02	452.69	102.61	83.72%	85.56%
BSI	122.84	1261.03	376.93	83.97%	85.55%
BSV	234.47	1272.26	394.94	71.17%	69.58%
SI	1430.62	1562.50	848.90	79.83%	82.62%
pSCD	742.39	1602.56	1367.99	81.71%	82.84%

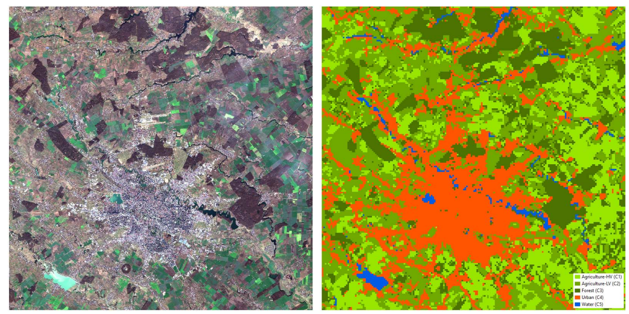

Fig. 3. Sentinel-2 scene and associated manual annotation.

In this letter, we propose a new MPEG-7 based descriptor, which is proved faster than classical descriptors such as HTD, SH, HTD_H, SI, BSI, and BSV descriptors and has comparable speeds to the FE methods based on SIs. Also, the classification speed of this new descriptor is higher because of the short feature vector obtained from Haar histograms.

The results of speed benchmarking for both FE and FC are presented in Table I and prove the computational performances of pSCD in contrast with those of other FE methods for multispectral EO image analysis presented in [17]. The values obtained for FE and FC assessment illustrated in Table I are computed for 1000 patches of 25×25 pixels each, with 13 spectral bands at a 10-m spatial resolution. The system architecture used in the processing has 16 GB of RAM and an eight-core 3.4-GHz CPU.

Using all of the 13 spectral bands, resampled at a 10-m resolution, we also evaluated the potential of pSCD features to classify the land cover of a small Sentinel-2 image, presented in the left of Fig. 3. For our assessment, we considered a few generic classes such as *Agriculture-HighVegetation* (C1), *Agriculture-LowVegetation* (C2), *Forests* (C3), *Urban* (C4), and *Water* (C5), which can be observed in Fig. 4. For the assessment we performed, the patches have 25×25 pixels in order to cover the relevant microtextures and color information present in all of the 13 spectral bands.

In the frame of qualitative and quantitative evaluation, we carefully created a manual annotation of the observed

TABLE II
k-NN AND SVM CLASSIFICATION ACCURACIES FOR SMALL SENTINEL-2 SCENE

	HTD	***SH***	***HTD_H***	***BSI***	***BSV***	***SI***	***pSCD***
knn	81.12%	83.71%	83.72%	83.97%	71.17%	79.83%	81.71%
SVM	87.42%	84.76%	85.56%	85.55%	69.58%	82.62%	82.84%

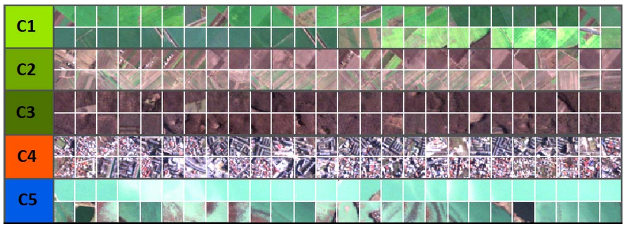

Fig. 4. Sentinel-2 database of thematic classes.

TABLE III
PRECISION–RECALL FOR SVM AND k-NN CLASSIFICATION OF THE MODIFIED MPEG-7 SCD

pSCD - SVM		pSCD-kNN		
Precision	***Recall***	***Precision***	***Recall***	
93.45%	87.39%	92.89%	88.97%	***C1***
91.71%	72.25%	92.61%	68.92%	***C2***
71.01%	88.28%	69.69%	89.00%	***C3***
63.04%	95.37%	59.06%	95.45%	***C4***
95.01%	56.71%	94.31%	67.80%	***C5***
82.84%	***80.00%***	***81.71%***	***82.03%***	***Mean***

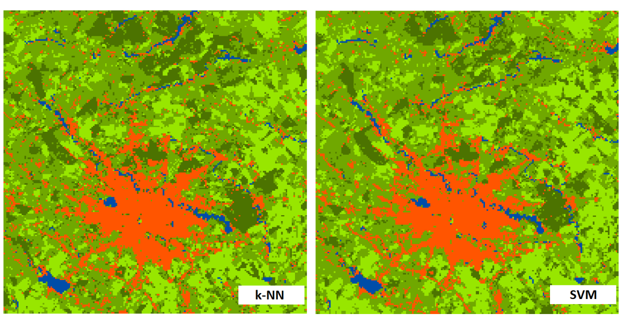

Fig. 5. k-NN versus SVM classification of pSCD image features.

scene, having into account all the spectral bands available. We used the manual annotation, present on the right of Fig. 3, to generate the confusion matrices for our classifications. In Table II, we present the mean accuracies of support vector machine (SVM) and k-nearest neighbors (k-NN) classifications of HTD, SH, HTD_H, SI, BSI, BSV, and pSCV features extracted from the scene illustrated in Fig. 3.

In Table III and Fig. 5, we present the detailed SVM and k-NN classification results of pSCD features computed for the Sentinel-2 scene illustrated in Fig. 3. Even though the results obtained for the pSCD descriptor are comparable to those obtained for other state-of-the-art descriptors, as shown in Tables I and II, we propose a method that can be integrated within the MPEG-7 standard without losing accuracy or computation speed. Taking into account what is stated in [18], we perform the quality and quantity assessment on a small region and then extrapolate for a larger surface where no ground truth is available.

Fig. 6. R-G-B bands of Sentinel-2 scene affected by clouds and shadows.

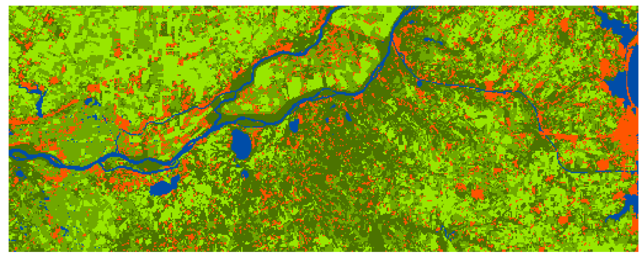

Fig. 7. Scene classification using SVM on pSCD features.

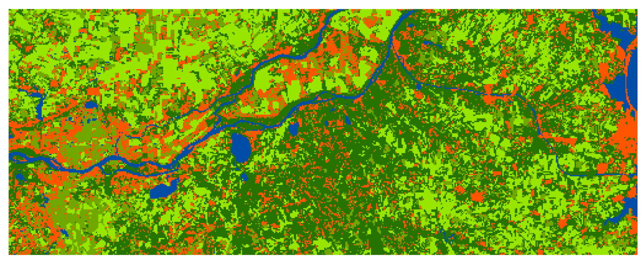

Fig. 8. Scene classification using k-NN on pSCD features.

In the second experiment, we performed pSCD FE on all 13 spectral bands of the Sentinel-2 image having almost 39% cloud coverage in order to prove the capabilities of the proposed descriptor to characterize the land cover classes. The analyzed scene can be observed in Fig. 6, while the qualitative results of classification with SVM and k-NN of the cloud-affected scene can be seen in Figs. 7 and 8.

The assessment we performed relies on supervised classification with k-NN that have $k = 10$ neighbors and SVMs, with a radial basis function kernel with $\gamma = 3.0518 \times 10^{-4}$ and the regression parameter $C = 5$. The parameters of the k-NN and SVM classifiers were chosen empirically in order to obtain best results for different remote sensing image types.

In the case of first scene shown in Fig. 3, we used 20 patch samples per class, which represent 0.0025% from the total of input image patches, while in the second scenario, for the image with cloud coverage presented in Fig. 6, we used 100 patch samples per class, representing almost 0.005% of the total patches in the scene.

B. Land Cover Classification of Cloud-Affected Areas

Having the purpose to extend the MPEG-7 standard for multispectral EO image analysis applications, we proposed to compute a polar coordinate transformation of the spectral

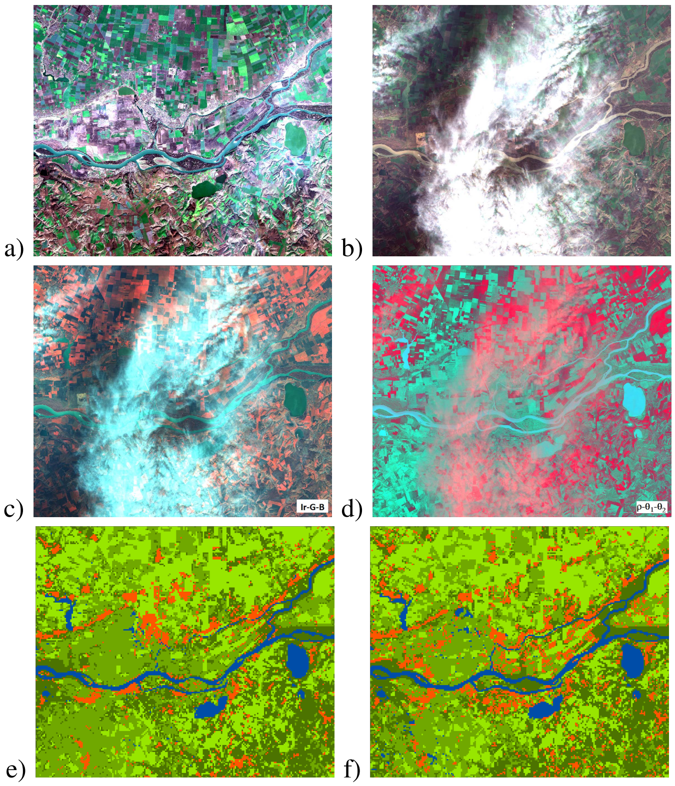

Fig. 9. Cloud-covered scene analysis. (a) Cloud free R-G-B bands, (b) Cloud affected R-G-B bands, (c) Ir-G-B bands, (d) pSCD transformation of Ir-G-B bands, (e) SVM classification with HTD, and (f) with pSCD.

values and then to compute the modified SCD. Doing so, we obtained an MPEG-7 compliant multispectral descriptor, which can perform faster than state-of-the-art FE methods and have comparable performances. Also, the descriptor we propose is proved to obtain good classification results even for images with dense cloud coverages and shadows.

The land cover classification for the cloud-affected scene presented in Fig. 6 is proved to provide better results when using pSCD features rather than HTD. A closer look of the cloud coverage can be observed in Fig. 9(a)–(d), where a comparison between different views is shown: no clouds in R-G-B bands [Fig. 9(a)], clouds in R-G-B bands [Fig. 9(b)], Ir-G-B bands [Fig. 9(c)], and polar coordinate transformation of the Ir-G-B bands [Fig. 9(d)].

Furthermore, in Fig. 9(e) and (f), the enhancement of SVM classification using pSCD [Fig. 9(f)] over HTD [Fig. 9(e)] can be observed.

IV. Conclusion

In this letter, a new MPEG-7 compliant descriptor for multispectral EO image analysis is presented. We modified the SCD present in the MPEG-7 standard to compute a polar coordinate transformation of the spectral values in order to capture the full spectral information using all the spectral bands available.

The descriptor we propose is faster than classical state-of-the-art descriptors and provides similar results for land cover classification. Also, the pSCD FE method can prove itself very useful in the case of multispectral images that have clouds and shadows, the descriptor being able to perform classification of the terrain beneath the clouds and cloud shadows, due to its illumination invariant properties.

Our assessment is made using a patch-based approach that encapsulates the microtextures needed in the classification process. The patch size is chosen to cover continuous elements with a uniform distribution of each class.

The quality and quantity evaluation is made using confusion matrices computed on a manual annotation data set and each classification result. The database used in the quality and quantity evaluation was created to assess the extracted features, not the classification results.

References

[1] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[2] P. Scheunders, S. Livens, G. Van de Wouwer, P. Vautrot, and D. Van Dyck, “Wavelet-based texture analysis,” *Int. J. Comput. Sci. Inf. Manage.*, vol. 1, no. 2, pp. 22–34, 1998.

[3] R. D. Navarro, J. C. Magadia, and E. C. Paringit, “Estimating the Gauss–Markov random field parameters for remote sensing image textures,” in *Proc. IEEE Region Conf. TENCON*, Jan. 2009, pp. 1–6.

[4] Y. Hongyu, L. Bicheng, and C. Wen, “Remote sensing imagery retrieval based-on Gabor texture feature classification,” in *Proc. 7th Int. Conf. Signal Process. (ICSP)*, vol. 1. Aug. 2004, pp. 733–736.

[5] W. Luo, H. Li, and G. Liu, “Automatic annotation of multispectral Satellite images using author-topic model,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 634–638, Jul. 2012.

[6] W. Luo, H. Li, G. Liu, and L. Zeng, “Semantic annotation of Satellite images using author–genre–topic model,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1356–1368, Feb. 2014.

[7] T. Sikora, “The MPEG-7 visual standard for content description—An overview,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, Jun. 2001.

[8] G. Marchisio, F. Pacifici, and C. Padwick, “On the relative predictive value of the new spectral bands in the WorldWiew-2 sensor,” in *Proc. IEEE Symp. Geosci. Remote Sens.*, Jul. 2010, pp. 2723–2726.

[9] F. A. Kruse *et al.*, “The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data,” *Remote Sens. Environ.*, vol. 44, nos. 2–3, pp. 145–163, 1993.

[10] K. van de Sande, T. Gevers, and C. Snoek, “Evaluating color descriptors for object and scene recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2009.

[11] S. M. Singh and K. Hemachandran, “Content-based image retrieval using color moment and Gabor texture feature,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 1, pp. 299–309, 2012.

[12] Y. Sohn, E. Morán, and F. Gurri, “Deforestation in North-Central Yucatan(1985-1995)-mapping secondary succession of forest and agricultural land use in Sotuta using the cosine of the angle concept,” *Photogram. Eng. Remote Sens.*, vol. 65, no. 8, pp. 947–958, 1999.

[13] S.-F. Chang, T. Sikora, and A. Purl, “Overview of the MPEG-7 standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 688–695, Jun. 2001.

[14] T. Okamura, J.-I. Kudo, and K. Tanikawa, “Multispectral illumination invariant classification with three visible bands and a near-infrared band,” *Opt. Eng.*, vol. 42, no. 6, pp. 1665–1672, 2003.

[15] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, “Color and texture descriptors,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[16] F.-A. Georgescu, M. Datcu, and D. Răducanu, “Patch-based multispectral image classification assessment for SENTINEL-2 data analysis,” in *Proc. Conf. Big Data Space (BiDS)*, 2016, pp. 344–347.

[17] F.-A. Georgescu, C. Vaduva, D. Răducanu, and M. Datcu, “Feature extraction for patch-based classification of multispectral earth observation images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 865–869, Jun. 2016.

[18] A. Baraldi, L. Bruzzone, and P. Blonda, “Quality assessment of classification and cluster maps without ground truth knowledge,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 857–873, Apr. 2005.