Clustering of Multispectral Images by Terrain Classes Using a Semantic Approach

Xi Zhou
Nanjing Research Institute of Electronics Engineering
Nanjing, China.
zhouxi_nju@126.com

Artsiom Nedzved
Belarusian State University
Minsk, Belarus
artiom.nedzved@gmail.com

Alexei Belotserkovsky
United Institute of Informatics Problems of the NAS of Belarus
Minsk, Belarus
alex.belot@gmail.com

***Abstract*—This article conducted a study on methods for clustering multispectral images by terrain classes using semantic technologies. Using Word2Vek technology, a semantic form of the image is constructed, which is used to determine the class of the image.**

***Keywords*—multispectral image, covariance, vector difference, Word2vec**

I. Introduction

Semantic scene segmentation based on images is one of the most important tasks in computer vision. Although we have made enormous progress in recent years using sophisticated image descriptors [1], [2] and more advanced machine learning techniques [3], [4], segmentation remains a challenging task. Although humans can easily interpret images semantically, this remains challenging for computer vision systems, primarily due to the ambiguity of the effect of light and surface reflections on a given pixel value. For example, dark pixels may be the result of reflections from dark surfaces under normal lighting conditions or reflections from light surfaces in shadows. Deciphering the contribution of light and reflection to images is a challenging task [5]. To solve this problem, we either need to make assumptions about the world or get more information.

In this work, we explore semantic segmentation of multispectral images using the latter approach. In particular, multispectral images from the Sentinel-2 satellite will be used.

The spectral imaging is imaging using multiple bands in the electromagnetic spectrum. While a conventional camera captures light in three wavelength ranges in the visible spectrum, RGB, spectral imaging involves a wide range of techniques beyond RGB. Spectral imaging can use infrared, visible light, ultraviolet, x-rays, or any combination of the above. This may involve acquiring image data simultaneously in the visible and non-visible ranges, illuminating beyond the visible range or using optical filters to capture a specific spectral range, and the ability to capture hundreds of wavelength ranges for each image pixel.

Multispectral remote sensing includes visible, near-infrared, and short-wave infrared imaging. These images were obtained over several broad wavelength ranges. Thus, multispectral imaging captures image data in a specific range of wavelengths across the entire electromagnetic spectrum. Different materials caught in the frame reflect and absorb rays at different wavelengths differently. In remote sensing, materials can be distinguished by their spectral reflectance signatures observed in remote sensing (Earth remote sensing) images. In this case, it is very difficult to make a direct identification, as described in [6].

A higher level of spectral detail in multispectral images provides a better ability to recognize subtle differences. For example, multispectral remote sensing makes it possible to distinguish three minerals due to the high spectral resolution. At the same time, the multispectral Landsat Thematic Mapper system cannot distinguish between these three minerals.

II. Multispectral satellite imaging

One of the priority areas for processing multispectral information and deciphering remote sensing data is theoretical and applied research aimed at increasing the efficiency of multispectral information processing. In theoretical and practical terms, the creation of systems that support the information processing process requires the development of new and improvement of existing methods and algorithms for information analysis, as well as the development of special mathematical, algorithmic and software for information processing and decision-making systems, which is explained by the following reasons. Firstly, the algorithms used to decipher remote sensing data do not provide the required accuracy and reliability of the results. Secondly, the use of clustering algorithms for multispecies data is not qualitatively satisfactory for expert assessment by specifying reference areas. Thirdly, the development of fundamentally new clustering algorithms is often not effective compared to improving existing algorithms, in terms of increasing processing speed and reducing the number of iterations.

In addition, measuring the degree of similarity of sensing objects is much simpler than forming feature descriptions.

One of the priority areas for processing multispectral information and deciphering remote sensing data is theoretical and applied research aimed at increasing the efficiency of multispectral information processing. In theoretical and practical terms, the creation of systems that support the information processing process requires the development of new and improvement of existing methods and algorithms for information analysis, as well as the development of special mathematical, algorithmic and software for information processing and decision-making systems, which is explained by the following reasons. Firstly, the algorithms used to decipher remote sensing data do not provide the required accuracy and reliability of the results. Secondly, the use of clustering algorithms for multispecies data is not qualitatively satisfactory for expert assessment by specifying reference areas. Thirdly, the development of fundamentally new clustering algorithms is often not effective compared to improving existing algorithms, in terms of increasing processing speed and reducing the number of iterations. In addition, measuring the degree of similarity of sensing objects is much simpler than forming feature descriptions.

In addition, remote sensing systems are currently being constantly improved, which makes it possible to obtain images of increasingly higher spectral and spatial resolution. There are systems that allow shooting in hundreds of spectral ranges. The spatial resolution of images is also constantly being improved. If in the 70s each pixel of a space image corresponded to 80 meters of the Earth, now there are images with a resolution of 1 meter, and sometimes better. Older techniques developed for lower-resolution imagery do not extract all the useful information contained in modern imagery. Therefore, there is a need for new methods for interpreting images of area objects that take into account the advantages of modern multispectral imaging.

As a rule, color satellite images are formed based on the absorption or reflection of certain radiation waves of the spectrum. As a result, in most cases, the color is formed based on filling the areas with different dyes with different concentrations. The color is formed by decomposing the absorption values of dye mixtures into the absorption values of individual spots, in this case, a simple decomposition according to the coordinates of color systems. Such decomposition does not allow to obtain a linear relationship between the dye concentration and absorption, which corresponds to the spectral line under monochromatic conditions. Most of the tasks of monitoring the Earth's surface are focused on polychromatic conditions, for which it is impossible to obtain accurate spectral values.

A. Landsat-8

One example of a multispectral sensor is Landsat-8. For example, Landsat-8 produces 11 images using the following bands:

1) Coastal aerosol (COASTAL AEROSOL) in range 1 (0.43-0.45 μm).
2) Blue (BLUE) in the range 2 (0.45-0.51 μm).
3) Green (GREEN) in the range 3 (0.53-0.59 μm).
4) Red (RED) in the range 4 (0.64-0.67 μm).
5) Near infrared (NIR) in the range 5 (0.85-0.88 microns).
6) Shortwave infrared 1 (SWIR 1) in the range 6 (1.57-1.65 μm).
7) Shortwave infrared 2 (SWIR 2) in the range 7 (2.11-2.29 μm).
8) Panchromatic (PANCHROMATIC) in the range 8 (0.50-0.68 microns).
9) Cirrus (CIRRUS) in the range 9 (1.36-1.38 μm).
10) Thermal Infrared 1 (TIRS 1) in the 10 range (10.60-11.19 μm).
11) Thermal Infrared 2 (TIRS 2) in the 11 range (11.50-12.51 μm).

Each band has a spatial resolution of 30 meters, with the exception of bands 8, 10 and 11. Band 8 has a spatial resolution of 15 meters, and bands 10 and 11 have a pixel size of 100 meters. In this case, there is no range between 0.88 and 1.36 μm because the atmosphere absorbs light at these wavelengths, as in Figure 1.

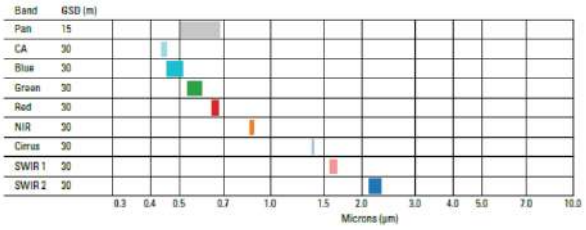

Figure 1. Ranges of spectral channels of Landsat satellites.

III. Description of the dataset

In [7] proposes a set of multispectral images captured by the Sentinel-2 satellite imagery. Sentinel-2 satellite imagery is publicly and freely available through the Copernicus Earth observation program. In [8], we present a new dataset based on Sentinel-2 satellite imagery, covering 13 spectral bands and consisting of 10 classes, containing a total of 27,000 tagged and georeferenced images. An example of one Industrial class image with the displayed spectral bands Red, Green, Blue can be seen in Figure 2. Of these, we have selected 8 of the most informative strips.

The dataset has the following classes:

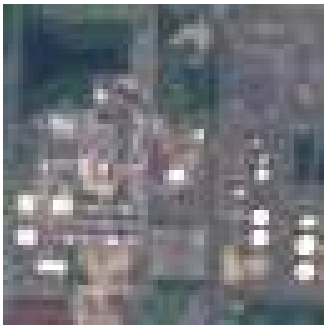

Figure 2. Industrial-1011 in RGB spectral bands.

A. Annual Crop class

The Annual Crop class is a satellite image of an agricultural area represented by fields of annual crops. A feature of the class is the presence of both large areas of immature vegetation and large areas of mature vegetation, brightly illuminated by infrared radiation.

B. Forest class

The Forest class is a collection of satellite images of forested areas. A feature of this class is the overwhelming presence of vegetation, brightly illuminated by infrared radiation.

C. HerbaceousVegetation class

The HerbaceousVegetation class represents satellite imagery of hilly or steep terrain. A feature of the class is the rare presence of vegetation in the image.

D. Highway class

The Highway class is satellite imagery of the area in which a major highway passes. They are characterized by the fact that after image processing the highway should be clearly visible.

E. Industrial class

The Industrial class presents satellite images of industrial areas. A special feature of the class is a large number of buildings and a small area of terrain with vegetation. Also, after processing, the outlines of buildings are often lost, but they can be restored to construct an outline using the difference in the images.

F. Industrial class

The Pasture class contains satellite images of flat terrain. A feature of the class is the abundance of vegetation, brightly illuminated by infrared radiation.

G. PermanentCrop class

The PermanentCrop class is a satellite image of an agricultural area represented by fields of permanent crops. A special feature of the class is the presence of large areas of mature vegetation, brightly illuminated by infrared radiation.

H. Residential class

The Residential class represents satellite images of populated areas. A special feature of the class is the large number of buildings throughout the image.

I. River class

The River class is satellite imagery of an area where there is a river. They are characterized by the fact that after image processing the river should clearly stand out.

J. SeaLake class

The SeaLake class contains satellite images of the seascape. They are characterized by the overwhelming presence of water.

IV. Calculating the covariance matrix

To classify the image class represented by 8 spectral channels, it is proposed to use an algorithm to calculate the covariance matrix, where each cell i,j will denote the covariance of the i-th spectral layer and the j-th layer, as in Figure 3.

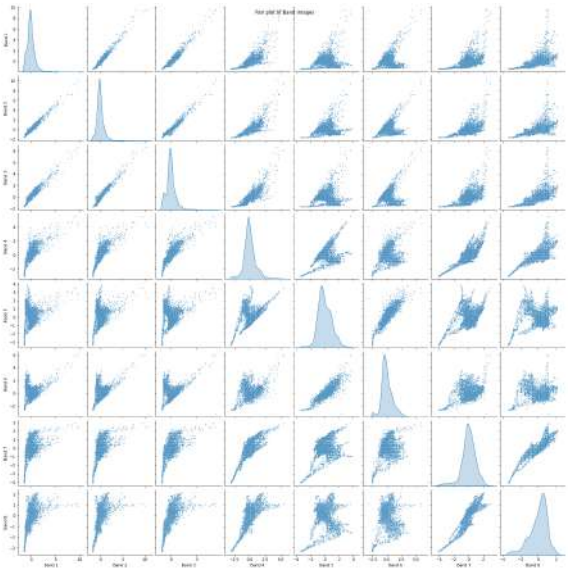

Figure 3. Visualization of the image covariance matrix with 8 spectral bands.

One such algorithm is the principle of principal components. He decides:

1) Approximate the data by linear manifolds of lower dimension: find a linear manifold of a given dimension $k < d$, the sum of squared distances to which is minimal.
2) Find a subspace of a given dimension, in the orthogonal projection onto which the spread (dispersion) (sample variance for $k = 1$) is maximum.
3) Find a subspace of a given dimension, in the orthogonal projection onto which the root-mean-square distance between each pair of points is maximum.

The principal component method consists of calculating eigenvectors and eigenvalues of the covariance matrix of the data space, then constructing projections in such a way that the direction of the maximum dispersion of the projection always coincides with the eigenvector having the maximum eigenvalue equal to the value of this dispersion. The covariance matrix after PCA processing can be seen in Figure 4.

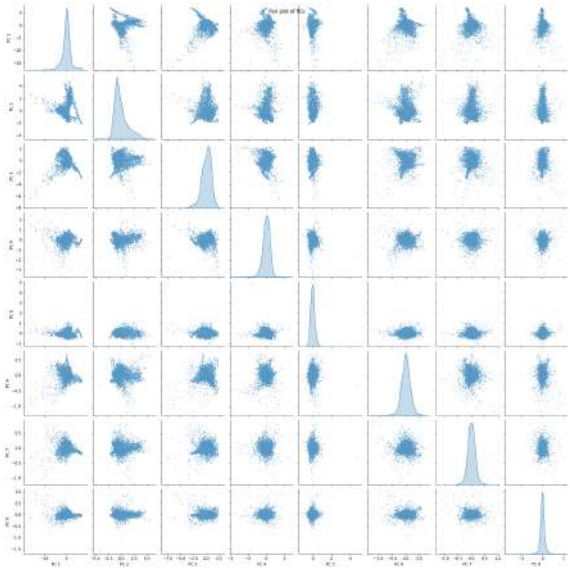

Figure 4. Visualization of the image covariance matrix with 8 spectral bands.

The next step of the algorithm is to reduce the dimension of the data space, but in the case of multispectral images this is not a necessary step: each projection is a new image layer that stores the necessary data. Instead, in [9] it is proposed to work only with the most informative image from the resulting projections.

Different images of the same class will have almost identical spectral data ratios. Indeed, for Forest class images the frequencies of 560 nm will prevail. and 842 nm., when for the SeaLake class 490 nm. and 945 nm. As a consequence, since each cell of the covariance matrix denotes either the variance of some layer of the multispectral image if that cell lies on the diagonal, or the covariance of two specific layers if it does not lie on the diagonal, the covariance matrices for each class will have the same patterns of dominant relationships.

V. Construction of a semantic form of the area

If we consider the matrix as a vector, then using the Word2vec technology from [10], from the previous statements we can conclude that the image is converted into a word that has metric characteristics, as shown in Figure 5.

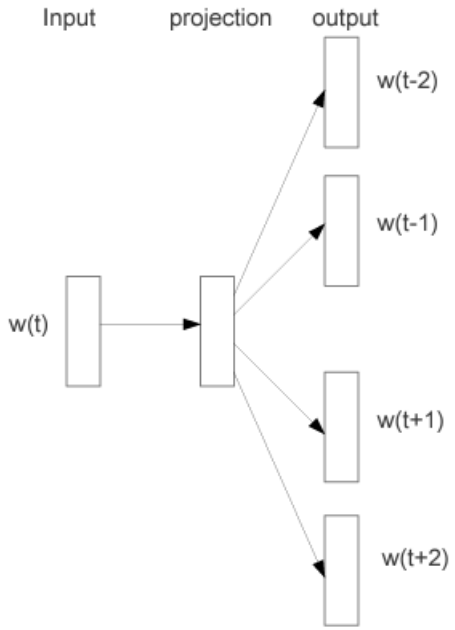

Figure 5. Illustration of how Word2vec technology works.

Following the Word2vec principle, we can build a semantic model to determine the semantic image of an image. Thus, the image is converted into a covariance ratio vector, where each cell uniquely defines the covariance of the two spectral layers. Consequently, if defining vectors are selected for each class, then by the difference between the vectors of the vector space, the dimension of which is n*n, where n is the number of spectral bands in the image, it will be possible to predict which class the image belongs to by the semantic difference of the vectors.

Using this semantic approach, we convert the image into a covariance ratio vector. Each element in this vector encapsulates the covariance between the two spectral layers, encoding not only spectral information but also semantic nuances. By selecting definition vectors for each terrain category, we create a semantic reference frame. Subsequently, using the semantic differences between these vectors within a vector space (which is expanded according to the number of spectral bands in the image), images can be accurately classified based on their semantic properties. By leveraging semantic techniques, we overcome the limitations of traditional pixel-based analysis and gain a deeper understanding of the semantic landscape of multispectral images.

VI. Demonstration of the method

To demonstrate how the method works, consider matrices for three different classes: Industrial, Forest, SeaLake. All three classes have different spectral characteristics that will uniquely determine the covariance matrix for the images representing the class. By calculating the average value of the covariance matrix for classes whose

sample included 1000 images, as well as the vector of eigenvalues of these matrices.

A. Calculation of class matrices

Matrix for class Industrial:
[[1 0.981 0.971 0.839 0.328 0.33 0.583 0.676]
[0.981 1. 0.976 0.876 0.414 0.426 0.64 0.696]
[0.971 0.976 1. 0.877 0.347 0.348 0.653 0.738]
[0.839 0.876 0.877 1. 0.549 0.45 0.823 0.823]
[0.328 0.414 0.347 0.549 1. 0.886 0.59 0.323]
[0.33 0.426 0.348 0.45 0.886 1. 0.504 0.255]
[0.583 0.64 0.653 0.823 0.59 0.504 1. 0.903]
[0.676 0.696 0.738 0.823 0.323 0.255 0.903 1.]]

Eigenvalues:
[5.566 1.449 0.706 0.145 0.059 0.012 0.024 0.041]

Matrix for class Forest:
[[1. 0.878 0.907 0.853 0.724 0.75 0.841 0.854]
[0.878 1. 0.907 0.888 0.805 0.856 0.861 0.863]
[0.907 0.907 1. 0.917 0.738 0.731 0.882 0.907]
[0.853 0.888 0.917 1. 0.888 0.793 0.972 0.976]
[0.724 0.805 0.738 0.888 1. 0.898 0.901 0.857]
[0.75 0.856 0.731 0.793 0.898 1. 0.794 0.755]
[0.841 0.861 0.882 0.972 0.901 0.794 1. 0.992]
[0.854 0.863 0.907 0.976 0.857 0.755 0.992 1.]]

Eigenvalues:
[5.568 1.259 0.627 0.218 0.015 0.066 0.136 0.112]

Matrix for class SeaLake:
[[1. 0.492 0.446 0.366 0.27 0.23 0.295 0.289]
[0.492 1. 0.477 0.408 0.32 0.26 0.329 0.296]
[0.446 0.477 1. 0.532 0.418 0.394 0.402 0.373]
[0.366 0.408 0.532 1. 0.583 0.552 0.537 0.501]
[0.27 0.32 0.418 0.583 1. 0.582 0.585 0.519]
[0.23 0.26 0.394 0.552 0.582 1. 0.54 0.482]
[0.295 0.329 0.402 0.537 0.585 0.54 1. 0.561]
[0.289 0.296 0.373 0.501 0.519 0.482 0.561 1.]]

Eigenvalues:
[4.047 1.147 0.589 0.506 0.492 0.383 0.413 0.425]

And also consider two images from the dataset, Industrial-1011, shown in Figure 2, and SeaLake-1016, shown in Figure 6.

B. Calculation of image matrices

The first proposed method for determining the class of an image will be the difference in the metrics of the matrix space L0. This method allows you to quickly and even visually determine whether an image belongs to a class. The main problem of this method is reducing the dimension of space from eight stripes to one number, as a result of which collisions arise when semantically different vectors return a measure that is close in value. Because of this, a significant number of errors arise when determining the class of an image.

Figure 6. SeaLake-1016 in RGB spectral bands.

The second method for determining the class membership of an image is the nearest neighbor search algorithm. This algorithm consists of three steps:

1) The distance between each image and the eigenvectors of each class is calculated. The distance is taken to be the quadratic difference of vectors.
2) Find the minimum distance for each image.
3) The class to which the image belongs is determined by comparing the minimum distances.

This method requires a little more calculations, but it takes into account the ratio of the spectral bands of the matrices in a certain order, as well as the difference between the corresponding bands.

The image Industrial-1011 obtained the following values of the covariance matrices and eigenvectors:

Matrix for class Industrial-1011:
[[1. 0.982 0.957 0.841 0.136 0.195 0.564 0.774]
[0.982 1. 0.975 0.873 0.214 0.275 0.604 0.778]
[0.957 0.975 1. 0.899 0.18 0.216 0.63 0.818]
[0.841 0.873 0.899 1. 0.383 0.321 0.816 0.929]
[0.136 0.214 0.18 0.383 1. 0.925 0.69 0.36]
[0.195 0.275 0.216 0.321 0.925 1. 0.608 0.302]
[0.564 0.604 0.63 0.816 0.69 0.608 1. 0.894]
[0.774 0.778 0.818 0.929 0.36 0.302 0.894 1.]]

Eigenvalues:
[5.47 1.862 0.487 0.089 0.042 0.032 0.006 0.014]

Based on the calculation results, image Industrial-1011, the quadratic distance between the image vector and the Industrial class vector is 0.482, the Forest class vector is 0.662, and the SeaLake class vector is 1.839. For clarity, distances are rounded to the third decimal place. The probability of an image belonging to the Industrial class is 84%, to the Forest class is 78%, and to the SeaLake class is 38%. Thus, we can conclude that the image belongs to the Industrial class, but it is worth noting the presence of local vegetation in the image.

The SeaLake-1016 image obtained the following values of covariance matrices and matrix space norms:

Matrix for class SeaLake-1016 :
[[1. 0.465 0.383 0.433 0.409 0.306 0.316 0.124]
[0.465 1. 0.56 0.621 0.605 0.459 0.477 0.137]
[0.383 0.56 1. 0.546 0.512 0.439 0.386 0.172]
[0.433 0.621 0.546 1. 0.647 0.51 0.481 0.202]
[0.409 0.605 0.512 0.647 1. 0.553 0.562 0.238]
[0.306 0.459 0.439 0.51 0.553 1. 0.42 0.174]
[0.316 0.477 0.386 0.481 0.562 0.42 1. 0.095]
[0.124 0.137 0.172 0.202 0.238 0.174 0.095 1.]]

Eigenvalues:
[3.981 0.955 0.746 0.611 0.553 0.453 0.371 0.331]

Based on the calculation results, image Industrial-1011, the quadratic distance between the image vector and the Industrial class vector is 1.902, the Forest class vector is 1.823, and the SeaLake class vector is 0.310. For clarity, distances are rounded to the third decimal place. The probability of an image belonging to the Industrial class is 47%, to the Forest class is 49%, and to the SeaLake class is 92%. Thus, we can conclude that the image belongs to the SeaLake class, and unambiguously.

Diagram of the algorithm

The general diagram of the algorithm is presented in Figure 7.

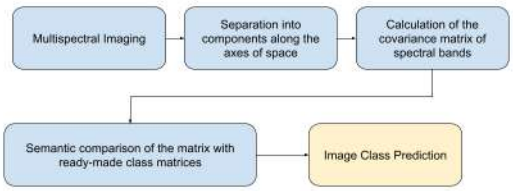

Figure 7. General diagram of the algorithm.

Acknowledgment

Financial support for the project “Agreement on the development of technology for developing algorithms for processing images of remote sensing of the Earth”, agreement number: 22CETC19-ICN1785.

References

[1] J. Shotton, J.M. Winn, C. Rother, A. Criminisi, “TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation,” A. Leonardis, H. Bischof, A. Pinz, ECCV 2006, Part I. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg, 2006.
[2] G. Csurka, F. Perronnin, “An efficient approach to semantic segmentation,” IJCV 95, 2011.
[3] J. Verbeek, B. Triggs, “Region classification with markov field aspects models,” CVPR, 2007.
[4] L. Ladicky, C. Russell, P. Kohli, P.H.S. Torr, “Graph Cut Based Inference with Co-occurrence Statistics,” K. Daniilidis, P. Maragos, N. Paragios: ECCV 2010, Part V. LNCS, vol. 6315, pp. 239–253. Springer, Heidelberg, 2010.
[5] G.D. Finlayson, M.S. Drew, B.V. Funt, “Color constancy: generalized diagonal transforms suffice,” Journal of the Optical Society of America 11, 3011–3019, 1994.
[6] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[7] Patrick Helber, Benjamin Bischke, Andreas Dengel, Damian Borth. “Eurosat: a new dataset and deep learning benchmark for land use and land cover classification.” IEEE Journal on Selected Topics in Applied Earth Observation and Remote Sensing, 2019.
[8] Patrick Helber, Benjamin Bischke, Andreas Dengel. “Introducing EuroSAT: a new dataset and deep learning benchmark for land use and land cover classification.” 2018 IEEE International Symposium on Geosciences and Remote Sensing.
[9] Santosh Kumar R., “Principal Component Analysis: Drilling Insight through Image Visualization,” 2020.
[10] Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, 2013.

ОПРЕДЕЛЕНИЕ КЛАССА МУЛЬТИСПЕКТРАЛЬНОГО ИЗОБРАЖЕНИЯ ПО СЕМАНТИЧЕСКОЙ РАЗНОСТИ КОВАРИАЦИОННЫХ МАТРИЦ

Бу Цин, Недзьведь А. А., Белоцерковский А.

Аннотация: В данной статье проведено исследование спутниковых мультиспектральных изображений, спектральных данных местности, а также представлен метод определения принадлежности изображений к классам местности с использованием семантического анализа вектора собственных значений ковариационной матрицы спутникового мультиспектрального изображения.

Received 27.03.2024