A SATELLITE IMAGE CLASSIFICATION APPROACH BY USING ONE DIMENSIONAL DISCRIMINANT ANALYSIS

A. Saglam [1, *], N.A. Baykan [1]

[1] Selcuk University, Engineering Faculty, Computer Engineering 4210 Konya, Turkey - (saglam, nurdan)@selcuk.edu.tr

Commission VI, WG VI/4

KEY WORDS: Satellite Image, Image Classification, Linear Discriminant Analysis, Dimension Reduction

ABSTRACT:

The classification problem in the image processing field is an important challenge, so that in the process image pixels are separated into previously determined classes according to their features. This process provides a meaningful knowledge about an area thanks to the satellite images. Satellite images are digital images obtained from a satellite vehicle by the way scanning the interest areas with some specified sensors. These sensors provide the specific radiometric and spatial information about the surface of the object. This information allows the researchers to obtain reliable classification results to be used to solve some real life problems such as object extraction, mapping, recognition, navigation and disaster management. Linear Discriminant Analysis (LDA) is a supervised method that reduces the dimensions of data in respect to the maximum discrimination of the elements of the data. This method also transfers the data to a new coordinate space in which the discriminant features of the classes are highest using the objection data provided manually. In this work, we consider the classes as if the satellite images have two classes; one is foreground and the other is background. The true classes such as roofs, roads, buildings, spaces and trees are treated sequentially as the foreground. The area outside the foreground class is treated as the background. The one dimensional reduced feature values of pixels, such that each value is reduced according to the binary classification of each class, are considered as membership values to the classes. In this way, each pixel has membership values for each of the classes. Finally, the pixels are classified according to the membership values. We used the ISPRS WG III/4 2D Semantic Labeling Benchmark (Vaihingen) images includes the ground truths and give the accuracy result values for each class.

1. INTRODUCTION

Classification of image pixels which are the smallest elements of images is an important problem in the satellite image processing field as in most computer vision fields, because obtaining meaningful information from satellite images is essential process for many remote sensing applications. In the process of pixel classification, the pixels are assigned to previously defined classes (Lu and Weng, 2007; Vailaya et al., 1998). In image classification methods, the pixels are classified according to the some features of the pixels such as colour (radiometric), geometric, and pattern features (Wang et al., 2016).

Appointing the satellite image pixels to land-cover classes is also defined as "semantic segmentation" by the remote sensing community ('2D Semantic Labeling Contest', n.d.; Marmanis et al., 2018; Volpi and Tuia, 2017). Semantic segmentation has been an essential issue in satellite image processing, because this issue provides solutions for many ecological and socio-economic problems (Vailaya et al., 1998) such as landslide monitoring, inferring geographical information, guidance information for intelligent military systems, infrastructure design, and disaster management (Montoya, 2003; Paisitkriangkrai et al., 2015).

In remote sensing semantic segmentation, the pixels in the satellite images taken over an urban are generally labelled as road, building, tree, and vegetation. This task is a problem due to the fact that the pixels belonging to different classes may be similar to each other, and in the same way, some pixels belonging to the same class may be different from each other (Paisitkriangkrai et al., 2015). The main reasons of this case lie in that one class may contain too many objects, and there may be many redundant objects in the classes (Wang et al., 2017). Therefore, it is deduced from this case that if the pixels belonging to different classes have more distinguishing features, the classification would be better.

In the literature, there are some feature extraction methods that transform the data to another feature space providing more distinctive features for the data belonging to different classes. Linear Discriminant Analysis (LDA) is one of them (Fisher, 1936). LDA is a supervised method that extracts more distinctive features from available features. In image processing, the pixels belonging to two different classes can be distinguished more accurate using this method. Saglam and Baykan (2017) used LDA to define the roof pixels in the satellite images taken over from an urban (Saglam and Baykan, 2017). According to the study, the urban images are considered they consist of two classes; the roofs are handled as foreground, and the rest of the zone as background. They used only spectral (radiometric) features, not spatial or surface features. LDA transforms the multiple radiometric values to one distinctive feature for two classes.

In this work, LDA is used for acquiring new multiple distinctive features for multi class discrimination from available pixel

---

* Corresponding author

features including radiometric and surface values. According to the method presented in this paper, LDA acts each class as foreground separately in the satellite images which consist of multiple classes such as roads, buildings, vegetation and trees. In this way, it extracts a new feature for each class, thence more than one feature (multiple LDA values) are obtained. Computing the LDA processing for each class, single dimensional data spaces are generated as much as the number of classes. Each LDA value represents a point in a single dimensional space. Finally, each data is assigned to a class in respect to the LDA values. In this work, we also developed a membership function which generates new membership values for each data using their LDA values to assign the elements of the data to classes.

In experiments, we used the Vaihingen dataset ('2D Semantic Labeling Contest', n.d.; Labeling and Vaihingen, 2016) to evaluate our method. The success of our method is demonstrated in this paper. In the implementation of the method, only the spectral features of the image pixels are taken into account, not spatial features. The results numerically and visually show that our method performs a respectable classification considering only the use of spectral properties of images and the simplicity of the method.

2. METHODOLOGY

2.1 Linear Discriminant Analysis (LDA)

LDA, also known Fisher's Linear Discriminant Analysis, is a feature extraction and dimension reduction method (Fisher, 1936). LDA enhances the inter-class variance while reducing the intra-class variance. This operation generates new attributes for the data, such that, these attributes provide the highest distinctive features for the data to be classified (Duda et al., 2000; Martis et al., 2013). LDA is a supervised method, because it needs a training data set which has to have goal ground-truth data.

According to the method, coefficient vector is generated using the ground-truth data, such that, the vector must have a number of values equals to the number of feature of the data (Lu and Weng, 2007). To obtain this vector, firstly, the intra-class covariance matrices are computed for each class, after that, the covariance matrices are also added together (1). In Eq. (1), $\mathbf{S}_w$ denotes the addition of the intra-covariance matrices of the classes.

$$\mathbf{S}_W = \sum_{c_k \in C} \sum_{\mathbf{x}_i \in c_k} (\mathbf{x}_i - \mathbf{\mu}_k)(\mathbf{x}_i - \mathbf{\mu}_k)^T \qquad (1)$$

where
C = set of classes
c_k = k th class
$\mathbf{x}_i$ = feature vector of i th element of c_k
$\mathbf{\mu}_k$ = mean vector of the feature vectors of c_k

After the matrix of the intra-class variances $\mathbf{S}_w$ is obtained, the matrix of the inter-class variances $\mathbf{S}_B$ is obtained as in Eq. (2).

$$\mathbf{S}_B = \sum_{k=1}^{m} (\mathbf{\mu}_k - \mathbf{\mu})(\mathbf{\mu}_k - \mathbf{\mu})^T \qquad (2)$$

where
m = the number of classes
μ = mean vector of the feature vectors of all the data

The next step is figuring out the maximization process which seems in Eq. (3). According to Eq. 3, the vector $\mathbf{w}$ which makes $j(\mathbf{w})$ at highest.

$$j(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \qquad (3)$$

In our method, LDA is calculated for each class separately as foreground. The other class in each LDA calculation is the class of background, which refers to the regions except the foreground. In the other words, we employee two classes as background and foreground for each LDA calculation, such that, a LDA calculation is needed one by one for all of the classes, for example four classes of foreground ("road", "building", "vegetation" and "tree" as the name of the classes) as taken into account in this study. If two classes are taken into account, Eq. (3) is in the direction of $(\mathbf{\mu}_1 - \mathbf{\mu}_2)$ linearly (Duda et al., 2000). For this reason, the calculation of the vector $\mathbf{w}$ can be directly calculated as in Eq. (4) instead of maximizing the equation in Eq. (3).

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{\mu}_1 - \mathbf{\mu}_2) \qquad (4)$$

After the weight vector $\mathbf{w}$ is obtained, each value in data $\mathbf{x}_i$ is multiplied by $\mathbf{w}$ as in Eq. (5).

$$x_i^{\text{LDA}} = \mathbf{w}^T \cdot \mathbf{x}_i \qquad (5)$$

As a result, each element in the data has one value $\mathbf{x}_i^{\text{LDA}}$ instead of multiple vector values. In this way, a dimension reduction process is also performed. However, in this study, each data has multiple values, because the LDA calculation is performed for each class as foreground. Thence, a weight set of $\mathbf{w}$ is obtained as $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$.

In the method used in this paper, it must be known that which class (foreground or background) represents minimum or maximum LDA values. Therefore, Eq. (4) must be used for the proposed method instead of Eq. (3). In Eq. (4), if $\mathbf{\mu}_1$ is the mean of the foreground, the elements which have highest LDA values belong to foreground, otherwise they belong to background. This case gives information about the direction of distribution between two classes (foreground and background).

2.2 Threshold value calculation for two-class labelling

As a result of the calculations of LDAs for each class, each data element has as a number of feature values equals to the number of classes. Each feature value obtained with LDA represents a point in a different single dimensional space, because each LDA value of an element obtained with the binary discrimination of the related class as foreground ("road", "building", "vegetation" or "tree" in this study) and remaining data as background. In each space, the distribution of all the data labelled as foreground or background is located as one of the four classes is foreground and the others are background. Namely, there would be four single dimensional spaces if four classes are handled as

foreground one by one. Therefore, if a threshold can be determined in a single dimensional space, the data would be separated into two classes. In the LDA calculations, if the foreground is specified as 2th class ($c_2 \rightarrow foreground$), meanly if μ_2 refers to the mean foreground in Eq. (4), the data belonging to the foreground would have smaller values in the space. Thus, separating the data into two classes in the space can be actualized as Eq. (6) and Eq. (7) using a threshold value t .

$$\text{if}(\mu_1 = \text{mean foreground}) \text{ in Eq. (4)}$$

$$x_i^{\text{label}} \leftarrow \begin{cases} \text{foreground} \Leftarrow x_i^k <= t^k \\ \text{background} \Leftarrow x_i^k > t^k \end{cases} \quad (6)$$

$$\text{if}(\mu_2 = \text{mean foreground}) \text{ in Eq. (4)}$$

$$x_i^{\text{label}} \leftarrow \begin{cases} \text{foreground} \Leftarrow x_i^k <= t^k \\ \text{background} \Leftarrow x_i^k > t^k \end{cases} \quad (7)$$

In the equations (6) and (7), x_i^{label} is the label value which class i th data is assigned to, for example "1" for background and "2" for foreground. If μ_1 in Eq. 4 refers the mean foreground, highest LDA values belong to the foreground. If μ_2 refers the mean foreground, smallest LDA values belong to the foreground. The value k refers the single space number, which represents the distribution of a real class, in which the data is separated into two classes as "foreground" and "background" for k th real class such as "road", "building", "vegetation", or tree".

In this case, the problem of specifying the threshold values for the spaces emerges. In this study, we first normalized the LDA values to integers in the range (0 to 255 in this study). After that, we test all of the integers in the range for the best classification success on the training data. The value which gives the best score is selected as threshold for that single dimensional space. For the score measurement for every integer, we use the evaluation of F1-score (8), which is also used for the success of the final classification results in this paper.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (8)$$

$$precision = \frac{tp}{tp + fp} \quad (9)$$

$$recall = \frac{tp}{tp + fn} \quad (10)$$

In the equations (9) and (10), tp defines the size of the true detected data for a class (true positives), fp defines the size of the false pair of the detected data for the class (false positives), and fn defines the size of the false pair of the outside the detected data for the class (false negatives). In the other words, it can be said that $tp + fp$ is the size of the detected data for a

class, and $tp + fn$ is the size of the true data for the class (Wang et al., 2017).

The process of obtaining the threshold t is performed for each class separately. In a result, a threshold set $\{t_1, t_2, \cdots, t_m\}$ is obtained. After obtaining the LDA values and the threshold value for each class, the data can be classified into two classes as foreground and background according to be separated class, using the related threshold values and the LDA values. But, the goal of the study is to separate all classes from each other in the data. In this paper, we classified all the data simultaneously using the LDA values and the threshold values.

2.3 Class selection functions for multi-class labelling

2.3.1 Min-Max selection function: Using only the obtained LDA values, firstly normalizing each in its related class in a range (e.g. 0 to 1), the data elements can be assigned to classes by the way selecting the index of the smallest LDA value (11) if μ_1=mean foreground in Eq. (4) or the highest value (12) if μ_2=mean foreground in Eq. (4) for each data. This way is called as "Min-Max function" in this paper. Each value in a LDA vector of an element was obtained for a different class (for one of the four classes in this study) in the LDA calculation in previous steps. In other words, for the LDA value $x_{i,"road"}^{\text{LDA}}$ was calculated handling the class "road" as foreground and the others as background in binary discrimination. The normalized LDA value $x_{i,"road"}^{\text{nLDA}}$ is calculated by normalizing all the $x_{i,"road"}^{\text{LDA}}$ in a certain range for each i among themselves

$$\text{if}(\mu_1 = \text{mean foreground}) \text{ in Eq. (4)}$$

$$x_i^{\text{label}} = \text{ind}(\max(x_{i,"road"}^{\text{nLDA}}, x_{i,"build."}^{\text{nLDA}}, x_{i,"veg."}^{\text{nLDA}}, x_{i,"tree"}^{\text{nLDA}})) \quad (11)$$

$$\text{if}(\mu_2 = \text{mean foreground}) \text{ in Eq. (4)}$$

$$x_i^{\text{label}} = \text{ind}(\min(x_{i,"road"}^{\text{nLDA}}, x_{i,"building"}^{\text{nLDA}}, x_{i,"veg."}^{\text{nLDA}}, x_{i,"tree"}^{\text{nLDA}})) \quad (12)$$

2.3.2 Max-membership selection function: The other way of selecting the class, is to applying a membership function. In this study, we assign membership values to pixels for each class. A membership value is the degree of an element in the data to assigning a class; such that it denotes the proximity of an element to the related class. A membership function transforms the current values to membership values. The membership function denoted for the class "road" (as an example) in the equations (13) and (14) is improved in the scope of this work. Eq. (13) defines the membership value $x_{i,"road"}^{\text{mem}}$ for the class "road" of i th element of the data if the foreground was considered as c_1 in the LDA calculation in previous steps (μ_1=mean foreground in Eq. (4)), and Eq. (14) if the foreground was considered as c_2 (μ_2=mean foreground in Eq. (4)). In the equations, $t_{"road"}$ refers the threshold value for the class "road" determined in previous steps.

if(μ_1=mean foreground) in Eq. (4)

$$x_{i,\text{"road"}}^{\text{mem}} = \begin{cases} \dfrac{x_{i,\text{"road"}}^{\text{LDA}} - t_{\text{"road"}}}{256 - t_{\text{"road"}}}, & x_{i,\text{"road"}}^{\text{LDA}} >= t_{\text{"road"}} \\[4mm] \dfrac{x_{i,\text{"road"}}^{\text{LDA}} - t_{\text{"road"}}}{t_{\text{"road"}} + 1}, & x_{i,\text{"road"}}^{\text{LDA}} < t_{\text{"road"}} \end{cases} \quad (13)$$

if(μ_2=mean foreground) in Eq. (4)

$$x_{i,\text{"road"}}^{\text{mem}} = \begin{cases} \dfrac{t_{\text{"road"}} - x_{i,\text{"road"}}^{\text{LDA}}}{t_{\text{"road"}} + 1}, & x_{i,\text{"road"}}^{\text{LDA}} <= t_{\text{"road"}} \\[4mm] \dfrac{t_{\text{"road"}} - x_{i,\text{"road"}}^{\text{LDA}}}{256 - t_{\text{"road"}}}, & x_{i,\text{"road"}}^{\text{LDA}} > t_{\text{"road"}} \end{cases} \quad (14)$$

The membership function is operated for each class using the related LDA value. In conclusion of the membership calculation process, each value has membership values instead of the LDA values. Each membership value is in the range [-1,1]. The higher the membership value is the greater the membership degree for the related class. Finally, the index of the highest membership value of an element is selected, and the element is assigned to the class labelling with this index (15). This way is called as "Max-membership function" in this paper

$$x_i^{\text{label}} = \text{ind}(\max(x_{i,\text{"road"}}^{mem}, x_{i,\text{"build."}}^{mem}, x_{i,\text{"veg."}}^{mem}, x_{i,\text{"tree"}}^{mem})) \quad (15)$$

3. EXPERIMENTAL RESULTS

We applied the developed method on the Vaihingen dataset provided by the ISPRS Commission III ('2D Semantic Labeling Contest', n.d.; Karakoyun et al., 2017; Labeling and Vaihingen, 2016). The dataset includes 33 satellite images; the size of each is about 2500×200 pixels. The ground-truth of 16 of them is shared for training and validation to researcher, and that of 17 of them is saved for centrally benchmarking by the commission. In this paper, we used the 16 images whose ground truths are

Figure 1. The sample image "area30" in the dataset (a) and the projection of its normalized DSM (b)

shared. We used 12 of them for obtaining the weight vector set $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$ (for LDA values) and the threshold set $\{t_1, t_2, \cdots, t_m\}$ as training data. We used 4 of them (area5, area7, area23, and area30) (Wang et al., 2017) as validation data to present the accuracy results of the method.

The images in the dataset consist of NIR (Near Infrared), R (Red), and G (Green) channels. The DSMs (Digital Surface Maps) and nDSMs (Normalized DSMs) are also provided by the commission. In Fig. 1, the image "area30" in the data set and its nDSM projection is showed as an example. The MATLAB program is used in this study for the method and image presentations. We use the values of NIR-R-G_DSM-nDSM as the feature vectors of the data.

The ground truth includes 6 class; those are "road", "building", "vegetation", "tree", "car", and "clutter". The classes "car" and "clutter" occupy a small area in the dataset; so we ignored these classes because of similarity of their radiometric features to the other large classes and misdirecting the general classification.

At first, we calculated the LDA values and obtained the threshold values for each class on the training data. For computing the LDA values and determining the threshold values, the binary ground-truths are needed for each class. For example, for the class "road", the binary ground-truth of the class is set as "road" as foreground and "not road" as background. But, in the data set, the truths of all the classes are in the same data together. Therefore, we firstly extracted the truth class separately for LDA calculation and threshold specifying. The threshold values obtained and their F1-scores for the related classes are given in the Table 1. In Fig. 2, the binary classification ground-truth and the results of the binary classification of the sample image are presented according to the obtained threshold values from training data.

It should be noted that the binary classification in the training images for only obtaining the threshold values, and the binary classification in the sample image which is one of the validation images is only for the presentation here. The binary classification results are not used for the actual classification.

After the LDA weight vectors and the threshold values for every class are obtained, the LDA values of the validation data are computed using the LDA vectors, and the LDA values are normalized to LDA integers. As the next step, the membership values are computed using the LDA integers of the validation and the threshold values obtained using the training data previously.

Class	Threshold {0,1,2, ..., 255}	Best Binary Classification Score
"Road"	127	0.7984
"Building"	152	0.7447
"Vegetation"	107	0.6438
"Tree"	88	0.8009

Table 1. The threshold values and their binary classification results on the training data

Here, we tested two labelling methods that are "Min-Max function" and "Max-membership function" mentioned in previous section. "Min-max functions" only uses the LDA

"road"   "building"   "vegetation"   "tree"

(a)

(b)

Figure 2. The truth foregrounds obtained from the actual ground-truth (a) and the binary classification of them using the LDA
weight vectors and the threshold values resulted from the training images (b)

values, and "Max-membership function" uses the membership values to assign the pixels to the classes. In the Table 2, the two labelling methods are compared on the validation data. In Fig. 3, the classification results of the validation images are also presented visually.

Classes	Min-Max F1-score	Max-membership F1-core
"Road"	0.7074	0.6913
"Building"	0.2907	0.6635
"Vegetation"	0.5252	0.5659
"Tree"	0.5382	0.5486
Overall	**0.5154**	**0.6173**

Table 2. F1-score validation results of the methods with the
functions of "Min-Max" and "Max-membership"

4. CONCLUSION

In this paper, the classifications of some satellite images obtained from a benchmark dataset are intended using one dimensional LDAs. For this purpose, the binary distributions of LDA is used, and applied the binary classification for each class as if they are foreground for obtaining a threshold value for each. Two labelling functions which assign the pixels to the classes is presented; one uses the LDA values and the other uses the membership values generated from the LDA values and the threshold values. In result, the two methods are compared to each other. The results show that the function that uses

membership values is better than the other. Max-membership method is practical to put into practice for supervising classification for large size images. We offer a new approach in this paper using the spectral features of images, not using the spatial features. For further studies, this method can be developed incorporating the spatial features.

REFERENCES

2D Semantic Labeling Contest [WWW Document], n.d. URL http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification. New York John Wiley, Sect. https://doi.org/10.1038/npp.2011.9

Fisher, R.A., 1936. The use of multiple measures in taxonomic problems. Ann. Eugen. 7, 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Karakoyun, M., Saglam, A., Baykan, N.A., Altun, A.A., 2017. Non-locally color image segmentation for remote sensing
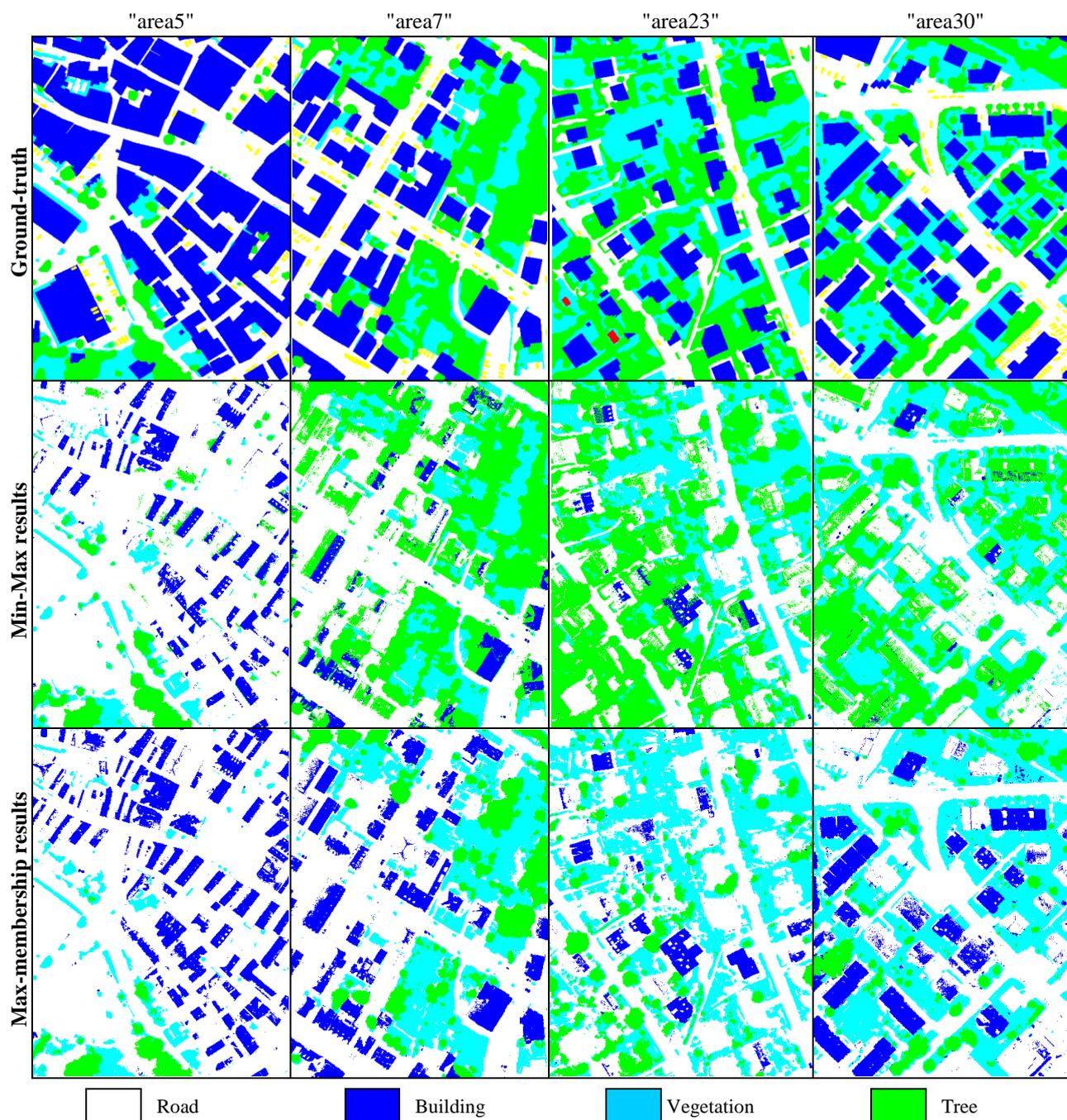
Figure 3. Visual classification results of the validation images with "Min-Max function" and "Max-membership function"

images in different color spaces by using data-clustering methods. 5th Int. Conf. Adv. Technol. Sci. 6–12.

Labeling, S., Vaihingen, B., 2016. Use of the Stair Vision Library within the ISPRS Use of the Stair Vision Library within the ISPRS 2D. https://doi.org/10.13140/2.1.5015.9683

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28, 823–870. https://doi.org/10.1080/01431160600746456

Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: Improving semantic image segmentation with boundary detection. ISPRS J. Photogramm. Remote Sens. 135, 158–172. https://doi.org/10.1016/j.isprsjprs.2017.11.009

Martis, R.J., Acharya, U.R., Min, L.C., 2013. ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform. Biomed. Signal Process. Control 8, 437–448. https://doi.org/10.1016/j.bspc.2013.01.005

Montoya, L., 2003. Geo-data acquisition through mobile GIS and digital video: An urban disaster management perspective, in: Environmental Modelling and Software. pp. 869–876. https://doi.org/10.1016/S1364-8152(03)00105-1

Paisitkriangkrai, S., Sherrah, J., Janney, P., Van-Den Hengel, A., 2015. Effective semantic pixel labelling with convolutional networks and Conditional Random Fields, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp. 36–43. https://doi.org/10.1109/CVPRW.2015.7301381

Saglam, A., Baykan, N.A., 2017. Roof detection on satellite images, in: 2017 25th Signal Processing and Communications Applications Conference (SIU). IEEE, pp. 1–4. https://doi.org/10.1109/SIU.2017.7960413

Vailaya, A., Jain, A., Zhang, H.J., 1998. On image classification: City images vs. landscapes. Pattern Recognit. 31, 1921–1935. https://doi.org/10.1016/S0031-3203(98)00079-X

Volpi, M., Tuia, D., 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55, 881–893. https://doi.org/10.1109/TGRS.2016.2616585

Wang, H., Wang, Y., Zhang, Q., Xiang, S., Pan, C., 2017. Gated convolutional neural network for semantic segmentation in high-resolution images. Remote Sens. 9. https://doi.org/10.3390/rs9050446

Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T., Eklund, P., 2016. A review of road extraction from remote sensing images. J. Traffic Transp. Eng. (English Ed. 3, 271–282. https://doi.org/10.1016/j.jtte.2016.05.005