ORIGINAL RESEARCH PAPER

Computer Science

CLASSIFICATION AND SEGMENTATION IN SATELLITE IMAGE USING K-MEANS ALGORITHM

KEY WORDS: Classification, Segmentation, Image

Mr. P. Sedhupathy Research Scholar, Department of Computer Science, Gobi Arts & Science College, Karattatipalayam, Gobichettipalayam, Tamilnadu, India.

Dr. R. Prabahari Assistant Professor, Department of Computer Science, Gobi Arts & Science College, Karattatipalayam, Gobichettipalayam, Tamilnadu, India.

ABSTRACT

Image processing techniques are one of the fastest developing technologies in the current world. These techniques encompass generic enhancement operations on images that make images viable for extractions of required information (Maier et al., 2019; Fernández-Pacheco et al., 2014). They are core research areas within the disciplines of engineering and computer sciences. Computer algorithms essentially process digital images using various tasks including reconstructions, restorations, compressions, enhancements, estimations of spectrums in images. Executions of these tasks result in analyses or classifications or detections of objects in digital images. Though image processing techniques can contribute towards growth of agriculture and specifically in determining weeds in cultivations, classification of weeds from images is a huge challenge. This chapter details on classification of crops and weeds using image processing techniques while focusing on improving classification accuracies using preprocessing, feature extractions, optimizations of feature selections, classifications, and deep learning approaches

1. Supervised Classification Methods

1.1. Maximum-Likelihood Classifier

Unknown samples can be classified by the statistical properties of the samples, if the number of classes, the forms for the class-conditional density functions, the class parameters for the density functions, and the prior probabilities of the classes are known. However in practice, all these parameters and the density functions are not given. Instead, they can be learned from the training data. In Satellite image classification, the maximum number of classes is known since the number of images types is fixed. The parameters for the density functions can be estimated once the forms for the density functions are given. One can visually confirm the forms for the density functions by plotting the sample distributions or using the Parzen window density estimation. However as the dimensionality of the feature space increases, the required number of samples grows exponentially, and visualizations impossible when the dimensionality is more than three. In general, one can assume the form of the density functions based on knowledge of the data. These forms can be chosen from standard unimodal density functions that best describe the true underlying densities.

1.2 k Nearest Neighbor Classification

For the maximum-likelihood classifier, we have assumed the underlying density functions to be normal (unimodal). However, in many practical problems the density functions may not be unimodal. In such cases, two approaches can be possible: 1) the multimodal density functions can be modeled as having multiple sub-classes if the forms of the densities can be verified somehow, 2) incases where the dimensionality prohibits density estimation, a nonparametric method can be used with arbitrary distributions without assuming a form forth density functions. A popular nonparametric method is the nearest neighbor or the k nearest neighbor method. Given a set of training data, a test sample x is assigned to a class ωi when the nearest neighbor of x in the training data belongs to the class ωi. The error rate of the nearest neighbor method is greater than the Bayes rate, and never worse than twice the Bayes rate when an unlimited number of training samples is used. A simple extension of the nearest neighbor method is the k nearest neighbor method, which assign a test sample x to the most frequent class that its k nearest neighbors belong. As the value of k grows toward infinity, the error rate becomes the Bayes rate. However, in practice the aforementioned is not always true since the number of training samples is limited. In fact, the error rate can even increase as k increases. However it is a useful method when the number of training samples is so small that the class parameters cannot be estimated, or the underlying density functions do not fit a simple unimodal density function.

1.3 Unsupervised Classification Methods

Supervised classification methods, such as the Bayes classifier (parametric)and k-nearest neighbor clustering (nonparametric), require training data. If the number of classes and the form of the class-conditional probability density functions are known, the class parameters can be estimated from the training data, and a parametric classification method can be used. If the number of classes is known but the form of the class-conditional probability functions are unknown, then a nonparametric method such as k-nearest neighbor clustering can be used. In general, collecting and labeling a largest of samples can be extremely costly and even prohibitive for some cases. Fortunately we have a large collection of M-FISH images with ground truth. Thus the use of a supervised method is an adequate approach here. However, in an early stage of investigation regarding the structure of the data based on some features, an unsupervised method is desired since the samples are unlabeled. Then unsupervised methods can be used to generate the training data set and further to extract useful features. Popular unsupervised methods are k-means clustering and fuzzy k-means clustering, which group the samples into k clusters whether or not k classes actually exist in the data.

1.3.1 Classifier

In order to overcome the limitation of these unsupervised methods, introduce a simple but effective unsupervised and nonparametric classification method for M-FISH images. The concept arises from the fact that a set of samples bound to a particular probe set has an expected intensity pattern for each class. Those fundamental patterns can be used as templates or idea prototypes for the classes. If our normalization process is effective, then the distance between a normalized sample and its correct class mean (template) should be as small as possible. If that is true, then the minimum-distance classifier, without actually training the classifier, can be used to classify pixels, and the classification accuracy can be used to evaluate the effectiveness of the normalization.

1.4 Post processing Methods

1.4.1 Majority and Plurality Filtering

Since many misclassified pixels are surrounded by correctly classified pixels, small local pixel misclassifications can be corrected using neighborhood information. A kernel of a proper size is applied, pixel by pixel, to the initial classification result. In majority filtering, a pixel value is replaced with a majority of the pixel values under the kernel, if the majority exists. If a majority is not found, then the pixel values remains unchanged. Given anN × N kernel, a majority is the value that occurs more than N2/2 times. In plurality filtering, a pixel value is replaced with the most common value under the kernel. If there is a tie, the pixel value remains unchanged. When the kernel is placed near the boundaries, the background pixels are ignored for the counting. Caution needs to be used when selecting the proper kernel size.

1.4.2 Accuracies of Classification Methods Before and After Normalization

The pixel classifications were performed with three different conditions: no preprocessing, background correction, and EM normalization. Both unsupervised-nonparametric (the minimum-distance classifier) and supervised parametric(the maximum-likelihood classifier) methods were used for classification. Since the maximum-likelihood classifier requires training, a set of images of normal male specimens was selected as training samples for each probe set as shown. A total of 26 out of 185 images were used for training: 9 out of 85 images for Vysis images, 8 out of 71 images for satellite images, and 9 out of 29 images for PSI images. All 185 images were tested using both classification methods. The maximum-likelihood classifier assuming the distributions were normal and used for the minimum-distance classifier to classify pixels. As table 4.2 shows, the overall classification accuracy without any normalization was about 50%, which increased significantly after background correctionto about 60%, and further improved with EM normalization to about70% for both classification methods. EM normalization increased the classification accuracy from 50% to 70%, which is a 40% increase in accuracy. The classification accuracies of the commonly cited images in previous papers regarding M-FISH pixel classification. Note that the results shown in this paper are the initial pixel classification accuracies without any post-processing to correct obvious misclassifications using such methods as majority filtering, and also note the rates are regarding the satellite pixels only.

1.5 Experimental Results

Table 5.1: PSNR

Existing 1	Existing 2	proposed
27	44	55
55	67	67
67	87	89
145	176	107
167	189	178

Table 6.1 represents PSNR values in this table. Proposed values are compared with Existing 1 and Existing 2 values. Their proposed values are higher than compare with other existing values.

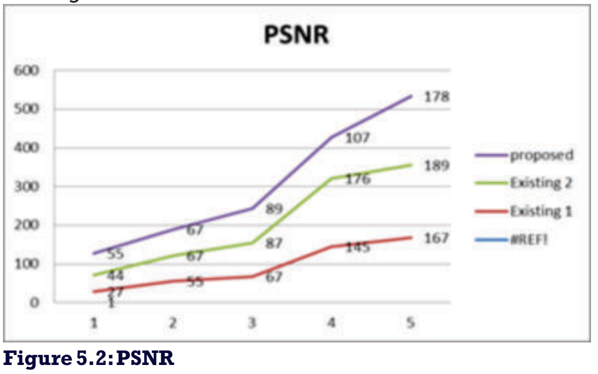

Figure 5.2: PSNR

Figure 6.2 represents PSNR values are compare with them. All values are only positive. The proposed values are higher than in this diagram. Existing 1 is a lower than compare with existing 2 and proposed values.

REFERENCES

1. Ahmed, A.M, Rashid, T.A and Saeed, S.A.M, (2020). Cat swarm optimization algorithm: a survey and performance evaluation. Computational intelligence and neuroscience, vol. 2020 pp. 1-20
2. Ahmed, F, Bari, A.H, Hossain, E, Al-Mamun, H.A and Kwan, P, (2011). Performance analysis of support vector machine and bayesian classifier for crop and weed classification from digital images. World Applied Sciences Journal, vol. 12, no. 4, pp. 432-440.
3. Alam, M, Alam, M.S, Roman, M, Tufail, M, Khan, M.U and Khan, M.T, (2020), Real-time machine-learning based crop/weed detection and classification for variable-rate spraying in precision agriculture. In 7th International Conference on Electrical and Electronics Engineering (ICEEE), pp. 273-280.
4. Alamdar, F and Keyvanpour, M, (2011). A new color feature extraction method based on QuadHistogram. Procedia Environmental Sciences, vol. 10, pp. 777-783.
5. Albawi, S, Mohammed, T.A and Al-Zawi, S, (2017), Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET), pp. 1-6.
6. Ali, H, Lali, M.I, Nawaz, M.Z, Sharif, M and Saleem, B.A, (2017). Symptom based automated detection of citrus diseases using color histogram and textural descriptors. Computers and Electronics in agriculture, vol. 138, pp. 92-104.
7. Ali, M, Guru, D.S and Suhil, M, (2018). Classifying Arabic Farmers Complaints Based on Crops and Diseases Using Machine Learning Approaches. In International Conference on Recent Trends in Image Processing and Pattern Recognition, pp. 416-428.
8. Alippi, C, Disabato, S and Roveri, M, (2018). Moving convolutional neural networks to embedded systems: the alexnet and VGG-16 case. In 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp. 212-223.
9. Arkeman, Y, Buono, A and Hermadi, I, (2017). Satellite image processing for precision agriculture and agroindustry using convolutional neural network and genetic algorithm. IOP Conference Series: Earth and Environmental Science, vol. 54, no. 1, pp. 1-7.
10. Bačanin Džakula, N, (2019). Convolutional Neural Network Layers and Architectures. In Sinteza-International Scientific Conference on Information Technology and Data Related Research, pp. 445-451.