Red Tide Image Recognition using Roundness and Entropy based on Semantic Feature

Sun Park[1], Yoo Kang Ji[2], ByungRae Cha[3]

[1] Institute Research of Information Science and Engineering,
Mokpo National University, South Korea
[2]Department of Information and Communication Engineering, DongShin University, South Korea
[3]SCENT Center, GIST, South Korea
[1]sunpark@mokpo.ac.kr, [2]neobacje@gmail.com, [3]brcha@nm.gist.ac.kr

Abstract. Red tide is a natural phenomenon that changes sea color by harmful algal bloom. There has been a lot of study on red tide due to increasing of red tide damage. However, to automatically recognize the red tide algae images present in previous studies are not enough. It is difficult to recognize red tide algae images because they do not have matching features for recognizing various species of algae. Besides, the accuracy of algae image recognition is low since previous red tide recognition method mostly uses a few species of red tide harmful algae images for training. In order to resolve the above limitation, this paper proposes the red tide algae recognition method using roundness and entropy based on semantic features of LSA (latent semantic analysis). The experimental results demonstrate that the proposed method achieves better performance than other red tide recognition methods.

Keywords: red tide algae, image recognition, LSA (latent semantic analysis), roundness, entropy.

1 Introduction

Red tide is a temporary natural phenomenon involving harmful algal blooms (HABs) in company with a changing sea color from normal to red or reddish brown, and which has a bad influence on coast environments and sea ecosystems [1]. The characteristic of red tide blooms in South Korea spreads fast by currents. Since the red tide has a broad occurrence range (e.g., a varied events area as the southern, western, and eastern part of sea) and spreads quickly, the direct detection of red tide blooms in real time for mitigation activity is difficult. Seven of the 63 species of harmful red tide algae damage fisheries in coastal areas of South Korea. The HABs have inflicted massive mortality on fin fish and shellfish, damaging the economies of fisheries for almost every year from 1993 to 2007. Particularly, there were huge harmful algal blooms of cochlodinium polykrikoides in 1995, resulting in about 95 million dollars in fishery damage [1]. In recent years, there has been no serious damage to fisheries from red tide events of harmful algae, but much money is spent for the mitigation activity of red tides blooms every year.

There have been many studies on red tide because the damage to sea farming by red tide blooms of harmful algae occurs every year. The research on red tide events of harmful algae in the coastal area of South Korea has mostly focused chemical properties and investigation of biological causes and is generally used for mitigation activity [2]. However, research on a variety of automatic methods for monitoring and recognizing red tide images of harmful algae on the coast of South Korea is still insufficient [1].

Recently, studies of red tide recognition mostly use classification techniques as follows. Jiang et al. proposed the red tide algae classification method using Bayesian classifier and SVM (support vector machine). Their method removes unknown and polluted algae images using Bayesian classifier. Red tide algae images are classified using SVM [3]. However, there is no good performance to remove unknown and polluted images. Besides, they use only 8 species of red tide algae for training images classification. Jiang et al. proposed red tide algae recognition method using two step classifiers. Their method uses first step classifier in which SVDD (support vector data description) removes unknown and polluted algae images. Algae images are classified into the predefined class label using second step SVM classifier. Also this method uses only 8 species of red tide algae for learning image classification [4]. Lili et al. use SVM and FCM (fuzzy c-means) clustering method for classifying red tide images. Their method uses SVM to identify class of test sample. The method identifies the misclassified data using probability based method, and then the misclassified data is reclassified using FCM. Their method also uses 5 species of red tide algae to train image classification [5].

Their methods [3, 4, 5] are difficult to classify red tide algae images except training algae images. In addition, they do not consider matching point between training images and query image, since the training images have different forms regarding various species of red tide algae (i.e., there are blooming red tides with relation to 63 species of algae in the coastal area of South Korea).

In order to resolve the above limitations of few training algae images and without matching point, this paper proposes a new red tide image recognition method using roundness and entropy based on semantic features of LSA. The proposed method has following advantages. First, training of the method uses the semantic features of 63 species of red tide algae image by means of LSA which can recognize various species of red tide algae images. Second, the method uses roundness of shape of algae images to apply matching point of the images for the images recognition. Final, our method uses entropy of red tide images which improves the recognition accuracy of algae images.

2 Latent semantic analysis

We define the matrix notation as follows: Let X_{*j} be the j'th column matrix vector X, X_{i*} be the i'th row vector, and X_{ij} be the element of the i'th row and the j'th column.

Latent semantic analysis (LSA) is a decomposition method using singular value decomposition (SVD). This method decomposes matrix A into three matrices, U, D, and V^T [6, 7].

$$A = UDV^T, \quad A \approx \tilde{U}\tilde{D}\tilde{V}^T \tag{1}$$

where U is an $m \times n$ orthonormal matrix of eigenvectors of AA^T (left singular vectors), and V is an $n \times n$ orthonormal matrix of eigenvectors of A^TA (right singular vectors). $D = diag(\sigma_1, \sigma_2, ..., \sigma_n)$ is an $n \times n$ diagonal matrix whose diagonal elements are non-negative eigenvalues sorted in descending order. $\tilde{U}$ is a $m \times r$ matrix, where $\tilde{U}_{ij} = U_{ij}$ if $1 \le i \le m$ and $1 \le j \le r$. $\tilde{D}$ is a $r \times r$ matrix, where $\tilde{D}_{ij} = D_{ij}$ if $1 \le i, j \le r$. $\tilde{V}$ is a $n \times r$ matrix, where $\tilde{V}_{ij} = V_{ij}$ if $1 \le i \le n$ and $1 \le j \le r$. Finally, $r << n$. In the LSA method, the j'th column vector A_{*j} of matrix A is the weight of the j'th column vector, and is represented as the linear combination of the left eigenvectors U_{*j}, which are semantic feature vectors, as shown in Equation (2). That is, the weight of the j'th semantic vector U_{*j} corresponding to the column vector A_{*j} is $\sigma_j V_{ij}$.

$$A_{*j} \approx \sum_{j=1}^{r} \sigma_j \tilde{V}_{ij} \tilde{U}_{*j} \tag{2}$$

3 Proposed method

In this paper, we propose red tide image recognition method consists of training of red tide image by LSA phase and red tide algae recognition phase using roundness and entropy.

3.1 Training of red tide image by LSA

In the training of red tide image phase, we modified Turk's method [8] for training of red tide image. Turk's faces recognition method uses semantic features such as eigenfaces by LSA. His method is fast, relatively simple, and works well in a constrained environment [8].

This phase is described as follows. First, image vector set for recognition of a candidate's image is constructed, where each images are set by 115×95 pixel. The center of image is almost reconciled with the center of image vector. Second, all red tide images are normalized by values of mean and variance of image vectors in order to decrease an error occurring of image background and light. Third, LSA of training image set is calculated using Equation (2). Final, red tide images are recognized having minimum value of Euclidean distance between a query image and the trained image set.

However, the above method has a problem of without matching features between training images and query image, since the training images have different forms regarding 63 species of red tide algae.

3.2 Red tide algae recognition by roundness and entropy

A simple LSA based method or algae image recognition methods have problems in recognizing different image regarding query image according to images of angle and viewpoint. Because there is no matching point between query image and candidate image set. In order to resolve the problem of without matching point, this paper proposes the roundness and entropy based algae image recognition method to enhance the quality of image recognition. The algae images can be recognized to compare roundness of shape between query image and candidate images (i.e. training image set). However, each 9 groups with relation to species of red tide algae have very similar roundness values as shown in the Table 1. (i.e., In the Table 1(e), query image q1.jpg (0.538) has 10 species of algae image regarding a similar roundness values.). Figure 3 shows sample of red tide algae images with regard to roundness in Table 1.

Table 1. Roundness values of training algae images regarding 63 species of algae

(a) 0~0.1		(b) 0.1~0.2		(c) 0.2~0.3		(d) 0.3~0.4		(d) 0.4~0.5	
image	roundness	image	roundness	image	roundness	image	roundness	image	roundness
49	0.0869	16	0.1974	15	0.2949	31	0.3929	10	0.4989
63	0.0623	54	0.1953	55	0.2612	51	0.3816	1	0.4941
50	0.0612	59	0.1412	57	0.2252	52	0.3655	41	0.4832
45	0.0466	56	0.1243	6	0.2234	2	0.3651	47	0.4687
44	0.0395	5	0.1226			11	0.3524	7	0.4528
43	0.0261	62	0.1168			58	0.3432	60	0.4398
		4	0.1106			29	0.3226	53	0.4395
								32	0.4187

(e) 0.5~0.6		(f) 0.6~0.7		(g) 0.7~0.8		(h) 0.8~0.9	
image	roundness	image	roundness	image	roundness	image	roundness
22	0.5955	24	0.6378	19	0.7981	30	0.8816
33	0.5919			25	0.7925	46	0.8667
3	0.5789			14	0.7906	17	0.8546
8	0.5785			18	0.7837	20	0.8302
13	0.574			42	0.7787	38	0.826
48	0.5626			37	0.7419	28	0.8175
40	0.5459			12	0.7406		
39	0.5404			21	0.7283		
9	0.5241			27	0.7272		
23	0.5076			34	0.7234		
				26	0.7178		
				35	0.712		
				36	0.7116		

To solve this limitation, we use entropy of image to recognize red tide algae image with respect to query algae image. The proposed method is described as follows. First, the roundness of algae image objects are calculated by Equation (3). Second, a candidate set for algae image recognition is composed of having top 10 roundness value in connection with query image. The entropy of query and candidate images is calculated by Equation (4). Final, the image having highest entropy value is selected to recognize query algae image.

In this paper, the roundness scale goes from 0 to 1. (i.e., the roundness of circle shape and horizontal shape is 1 and 0 respectively.). Roundness *R*() is defined as follows [9].

$$R(I) = \frac{4 \times \pi \times area}{perimeter^2} \quad \textbf{(3)}$$

where *I* is 115×95 image matrix, *π* is pi that is the ratio of the circumference of a circle to its diameter (i.e., pi is 3.1416), area is area of image object, perimeter is a path that surrounds an area of image.

In order to remove noise of entropy, algae image converts into a black-and-white image. Entropy of Equation (4) is given as follows [9, 12]:

$$Entropy(I_{bw}) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad \textbf{(4)}$$

where I_{bw} is 115×95 black-and-white image, *p* is number of histogram of I_{bw}.

4 Experiments

This paper uses images of 63 species of red tide algae in the coastal area of South Korea for training of red tide images. The images are obtained from the red tide information on Korea marine environmental information site [10] and Google image search [11]. We used the training data of 3,150 algae images with respect to 63 species of red tide algae. Fifty images are included in the label of one species of red tide algae. We use 350 query images with respect to 7 species of most harmful algae on South Korea.

Average accuracy is used to measure the recognition performance. We have conducted performance evaluation by testing proposed method and comparing it with 7 other recognition methods using same training and query images. We implemented 8 image recognition methods: LSA, NB, NN, SVM, B+SVM, SVDD+SVM, SVM+FCM, and LSA+RE. LSA+RE denotes our proposed method in this paper. LSA denotes Turk's method using LSA [8]. NB denotes image classification method using naïve Bayesian [12]. NN denotes image classification method using back propagation neural network [12]. SVM denotes image classification method using support vector machine [12]. B+SVM denotes Jiang's method using Bayesian and SVM classifiers [3]. SVDD+SVM denotes Jiang's method using support vector data description and SVM [4]. SVM+FCM denotes Lili's method using SVM and fuzzy c-means [5].

The average accuracy of evaluation results of LSA+RE is approximately 15.20% higher than that of LSA, 15.00% higher than that of NN, 12.80% higher than that of NB, 11.19% higher than that of SVM, 8.70% higher than that of B+SVM, 7.10% higher than that of SVDD+SVD, 4.50% higher than that of SVM+FCM.

5 Conclusion

In this paper, we proposed red tide image recognition method, and the architecture of a system to implement it. The proposed method uses semantic features from 63 species of red tide algae in coastal areas of South Korea for training algae images which can recognize various species of red tide algae images. Besides, the method uses roundness of images to apply matching point between query image and trained images. And it uses entropy of red tide images which improves the recognition accuracy of algae images. Our method is simple and works well about various species of red tide algae.

Reference

1. Red Tide Information System, Korea National Fisheries Research & Development Institute, http://portal.nfrdi.re.kr/redtide/index.jsp, (2012)
2. Y. M. Kim, Y. G. Byun, Y. Huh, K. Y. Yu, Detection of Cochlodinium Polykrikoides Red Tide Using MODIS Level 2 Data in Coastal Waters, Journal of Korean Society of Civil Engineers, pp.535-540, (2007)
3. T. Jiang, C. Wang, B. Wang, J. Xie, N. Jiao, T. Luo, Real-time Red Tide Alge Classification Uing Bayes Classifier and SVM, In proceeding of International Conference on Bioinformatics and Biomedical Engineering (ICBBE), pp.2888-2891, (2008)
4. T. Jiang, C. Wang, B. Wang, J. Xie, N. Jia, T. Luo, Real-time Red Tide Algae Recognition using SVM and SVDD, In proceeding of International Conference on Intelligent Computing and Intelligent System (ICIS), pp.602-606, (2010)
5. X. Lili, J. Tao, X. Jiozhen, Z. Shaoping, Red Tide Algae Classification Using SVM-SNP and Semi-supervised FCM, In proceeding of International Conference on Education Technology and Computer (ICETC), pp.389-392, (2010)
6. T. K. Lanauer and S. T. Dumais, A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of Knowledge, Psychological Review, 104, pp. 211-140, (1997)
7. T. K. Landauer, P. W. Foltz, D. Laham, An Introduction to Latent Semantic Analysis, Discourse Processes, 25, pp. 259-284, (1998)
8. M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of cognitive Neuroscience, Vol. 3, No. 1, pp.71-86, (1991)
9. R. C. Gonzalez, R. E. Woods, S. L. Eddins, Digital Image Processing Using MATLAB, New Jersey, Prentice Hall, (2003)
10. Red tide information on Korea marine environmental information site, http://portal.nfrdi.re.kr/redtide/index.jsp, (2012)
11. Google Image Search, http://images.google.com/imghp?hl=EN, (2012)
12. J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann, (2006)