Research on Automatic Analysis Algorithm of Red Tide Algae Image in Offshore

Renying Song

Ningde Normal University, Ningde, Fujian, 352100

Abstract: In order to analyze the types and quantities of harmful algae contained in seawater, early monitoring and forecasting of red tide algae development was carried out, and a computer automatic identification system for red tide algae images was developed. Image processing technology was used to extract algae image morphology and texture features, and genetic algorithm was used to select features. On this basis, a neural network is used to establish a classifier to classify and identify algae images. The results show that the system can effectively improve the learning ability and classification performance, and achieve good classification and recognition of the three main algae that cause red tide. The analysis results are less than the results of manual counting and recognition.

1. Introduction

Red tide is a harmful ecological phenomenon caused by the explosive growth or high concentration of some phytoplankton, protozoa or bacteria in seawater under certain environmental conditions. According to the China Marine Environmental Quality and Disaster Bulletin issued by the State Oceanic Administration, from 2000 to the present, the annual occurrence of red tide disasters exceeds 10,000 square kilometers, which has a serious impact on the aquaculture industry and fishery resources. Strengthening the monitoring of red tide algae is one of the main means to prevent red tides. Most of the traditional methods use manual microscopy to analyze the seawater samples to complete the monitoring of red tides. This method is labor intensive, inefficient, and low precision, and must be completed by professional personnel. Therefore, the establishment of an effective monitoring method for red tide organisms and the development of a red tide algae image automatic identification system are urgent issues to be solved. The red tide biological flow image field monitoring technology combines flow cell technology, microscopic imaging technology and image processing technology to achieve on-site rapid monitoring of target red tide organisms. The technology provides a platform for quantitative detection by flow cytometry. Its rapid and accurate quantification is particularly important for the monitoring of "red tide", an ecological phenomenon directly related to the number of cells; microscopic imaging technology has a taxonomic image, Intuitive advantages; automatic analysis and identification of algae using image processing techniques.

2. Principles and methods for identification of red tide algae

At present, many researchers at home and abroad have studied the classification and recognition algorithms of red tide algae images. The most representative is the FlowCAM developed by FluidImaging Technology of the United States. This system is simpler and less time-consuming than the micro-mirror. The image recognition software has obtained 80%-90% classification accuracy, but the image must still be confirmed by a professional technician. At present, the international research on underwater flow cytometry and imaging systems has been carried out, such as the original flow-through flow meter FlowCy-tobot. FlowCytobot is a laser-based system that "sees" a single cell that is excited by a laser to produce fluorescence. The image is primarily a fluorescent signal. Ji Guangrong et al. [A web-based harmful red tide microscopic image acquisition and diagnosis system was constructed by optical microscopy. Wang Wei and Yu Xinsheng used SVM and neural network to identify plankton and established a monitoring system based on microscope

imaging. The use of micromirrors to capture images has certain flaws in data analysis applications for on-site and navigation. In order to be able to monitor the red tide algae for a long time and visually observe the algae morphology, this paper improved the real-time acquisition system of marine red tide organisms in Xiamen University, and proposed an automatic classification and recognition algorithm for red tide algae flow images.

3. Algae image acquisition, segmentation

The seawater sample flows through the flow chamber from top to bottom at a high velocity, and the flow chamber is a narrow passage with a cross-sectional size of 2 mm x 100 um. The image was acquired using a CCD after being magnified by the microscopy system. According to statistics, the body length and width distribution of common marine red tide algae cells range from 10μm to 200μm. The improved system uses Gigabit CCD camera DMK23G445 with a resolution of 1280×960, which is added between the flow cell and the CCD. A 10x microscope objective is calibrated by a standard fluorescent sphere of 10μm. The actual effective detection area of the CCD on the flow cell is about 400μm × 300μm, which can effectively collect clear red tide algae images. The acquisition subsystem controls the LED to flash at 3 times per second, triggering the CCD to acquire images at 3 frames per second. Similar to camera photography, the CCD exposure is controlled by flickering. When the exposure time is 10μs, the high-speed motion of algae cells produces a smear trajectory of less than 1μm, so that small algae cells with a body length and a width of 10μm can be distinguished. Impurities in the seawater will slowly accumulate in the flow chamber. In order not to affect the analysis and identification of the algae by these deposited impurities, the acquired image needs to be processed, that is, the background information is removed. When the image is collected, the characteristics of the 20-frame image are counted as the starting background information; during the collection process, the background information is periodically and quantitatively modified. The processed image is segmented, and the position of the algae image is obtained by setting a threshold. The basis of the segmentation mainly includes gray scale information, texture information, scale information, and the like. Since the flow cytometer is imaged through the image source through the background light source, the cell has a lower gradation brightness than the pure environmental liquid. The inner part of the algae cell is a structure of fine cytoplasm, which is reflected in the image as a richer form of the texture, while the interior of the impurity such as bubbles is a simple structure, and the texture detail on the image is an important clue to distinguish between cells and impurities. The size of the red tide algae cells is generally larger than 10 μm, and some fine impurities can be distinguished by this feature.

4. Discussion of identification technology

The application of shape-based automatic identification methods to the identification of red tide algae requires careful consideration. Red tide algae are living tissues, and the same kind of red tide algae may have different shapes; and different kinds of red tide algae may have the same shape. Therefore, it is not possible to use only shape information to identify red tide algae. However, contour information can still be used as an important condition for identification. Although the contour information cannot distinguish the differences between the species, these species can be divided into different shapes: circular, slender, rectangular, and the like. According to the data obtained by the contour tracking, the following feature quantities can be respectively calculated: circular shape, rectangularity, flatness, invariant moment, Fourier descriptor and the like. The above feature quantity can be obtained by processing the segmented image by using edge tracking, area marking, and chain code. The method of morphological segmentation is used to process the binary image, which can solve the problem of segmentation of connected seaweed. Texture features play an important role in recognition because they contain important information about the arrangement of the surface structure of red tide algae. Compared with other types of features, texture features can better reflect the macroscopic and microstructural properties of red tide algae images. Using the method of co-occurrence matrix, that is, comprehensively applying point gray and regional gray

features, the second-order moment, contrast, correlation, and entropy are used as descriptors of texture information to identify red tide algae.

The three main algae, such as noctiluca, spirulina and spirulina, which cause red tides, are classified and identified. The most effective combination of features should be selected from the various types of algae extracted above to construct a classifier.

The optimal feature subset selection problem is a combinatorial optimization problem. If a set of optimal subsets is selected from the n original features: F op t F= { f1, , f n }, there are 2n-1 feature subset combinations. If you search all the combined spaces of feature vectors, the huge amount of computation is often prohibitive. The genetic algorithm uses the viewpoint of biogenetics, combines the idea of survival of the fittest and the exchange of random information, and realizes the evolution of the population through the mechanisms of natural selection, crossover and mutation. In the process of optimization, GA randomly generates multiple starting points in the high-dimensional solution space and starts searching at the same time. The fitness function guides the search direction. It is a search technology that can quickly find global optimization solutions in complex search spaces. The implementation process of genetic algorithm is basically similar. The difference is that the corresponding coding method, genetic operator, fitness function and GA control parameter are adopted for specific problems. These four processing elements control the convergence speed of the genetic algorithm and the quality of the solution. The fitness function selection is important for the optimization performance and speed of genetic algorithms. We propose the following fitness function:

f itness(C) = Q (F c) + (n-kn)

Where: Q (FC) is the accuracy of the classification of the sample obtained by the selected feature subset FC, using the nearest neighbor classifier and leave-one-ou t method; n is the total number of features; k is the feature in FC Number; for the weight coefficient, take a value between 0 and 0.1. It is ensured that individuals with smaller feature numbers obtain higher fitness values when the classification accuracy is close.

Since the features in the algae feature space are not orthogonal, some features can be nonlinearly divided. If the linear function is used to design the classifier in the statistical pattern recognition method, it is a difficult problem to select an effective decision function, even if different decision functions are selected. Only different feature information is utilized, and the relevant parameters of the decision function are not easily modified in the recognition process. Artificial neural network has the characteristics of high nonlinearity, analog parallelism, high fault tolerance, robustness, self-association, self-learning and self-adaptation. As an adaptive pattern recognition technology, it has a wide range of applications in the field of image recognition. . The most widely used neural network model is the multi-layer forward sensor using the error back propagation training algorithm. The network generally consists of an input layer, an output layer, and a BP hidden layer portion. The hidden layer can be one or more layers. Cybenko points out that when each node uses an S-type function, an implicit layer is sufficient to achieve the problem of arbitrary decision classification. From the point of view of simplicity and practicality, it is also advocated to choose a hidden layer. The number of input layer nodes of the neural network is the sample selection feature dimension; the output layer is the decision classification layer, and the number of nodes is the number of classification categories; the selection of hidden layer nodes is a problem, AJM aren points out the number of optimal hidden layer nodes The geometric mean of the number of input and output nodes is available. The transfer function uses the sigm o id type. The analysis results of the analysis system are less than the results of the manual count recognition, but there are still a small number of lost and false targets, the main reason is the interference of suspended matter. Although the interference of suspended matter is treated in the process of image segmentation, there is still a small amount of suspended matter interference that cannot be eliminated. When there are suspended fragments whose texture and morphological characteristics are similar to those of red tide algae, it will cause false targets. When the suspended debris adheres to the red tide algae, these fragments will affect the texture and morphological characteristics of the target to be identified, resulting in the loss of the target. Solving these problems thoroughly is the focus of future work.

5. Conclusion

This paper combines the knowledge of image processing, artificial intelligence and pattern recognition to describe the characteristics of red tide algae. The experimental results show that the neural network combined with genetic algorithm for feature selection can effectively improve the learning ability and classification performance, and achieve a good classification and recognition of the three main algae that cause red tide. The next step is to push the research results to practical application after improvement and improvement.

Acknowledgements

References

[1] Zheng Shaoping, Xie Jiezhen, Wang Boliang, et al. Research on the identification of dominant species of red tide based on shape feature descriptors [J]. Journal of Xiamen University: Natural Science Edition, 2009, 48(5): 676-679

[2] Jiang Tao, Wang Cheng, Wang Bo Liang, et al. Red tide algae identification based on SVDD and SVM [J]. Journal of Xiamen University: Natural Science Edition, 2010, 49(1): 47-51

[3] Yu Xiaohan, Xie Jiezhen, Zheng Shaoping. The application of Fourier identification in the automatic identification of red tide organisms [J]. Ocean Lake Marsh Bulletin, 2011(4): 11-15

[4] Li Liqin, Gao Huanwen, Zhou Xingxiang. Research on the initial contour selection of Snake model [J]. Computer Engineering and Applications, 2004, 11: 43-45

[5] Xu Xiaojun, Shao Ying, Guo Shangfen. Texture analysis of fire image based on gray degree symbiotic matrix [J]. Computing Technology and Automation, 2007, 26(4): 64-67