

Distributed system for detection of biological contaminants

V. M. Grishkin^a, K. V. Smirnov^b, N. A. Stepenko^c

Saint Petersburg State University,
7 9 Universitetskaya nab., Saint Petersburg, 199034, Russia

E-mail: ^a valery-grishkin@yandex.ru, ^b kvsmirnov@list.ru, ^c nick_st@mail.ru

The paper proposes a distributed system for detecting the types of biological contaminants existing on objects' surfaces. The system implements biofouling detection method based on an image processing technique. The system processes a series of object images obtained in the visible and near infrared spectral ranges. One image in the series is marked as a base image. All images of the series are converted to one common shooting point and to one common angle. The object of interest is detected in the base image, and then the background is removed from all images. To recognize the type of biological contaminants, we use a pre-trained classifier based on support vector machine method.

The proposed detection method has an obvious parallelism in data processing. Each image in the series, except the base, can be processed independently. Therefore it is quite easy to implement this method on a computing cluster. The central host of the cluster is used to implement non-parallel branches of the image-processing algorithm. These branches are namely interactive segmentation, a search for key points in the base image, and classifier training. The central host also solves the problems of data distribution among cluster nodes and of synchronization of nodes. Other nodes of a cluster are processing other images in a series, except the base image. They implement a search for key points, convert all images to a common shooting point, remove the background, and identify types of biofouling. After processing, we form the map of pollution for each image. This map then gets sent to the storage at the central node of the cluster

Keywords: parallel image processing, pattern recognition, biofouling identification.

The work was supported by Saint Petersburg State University. Research grant 9.37.157.2014

© 2016 Valery M. Grishkin, Konstantin M. Smimov, Nikolai A. Stepenko

Introduction

Biofouling is the result of unwanted accumulation of biological substances on surfaces that are constantly exposed to aggressive environments. A traditional approach to detecting biological destructors is based upon probing surfaces for subsequent laboratory research. This approach results in large effort, is expensive, and requires expert consultations. At the same time, any biological object is able to consume or reflect light rays of different bands of the spectrum. Different kinds of biological objects have different spectral characteristics. This makes it possible to identify the kind of a biological object basing upon its spectral characteristics. In this article we propose a distributed system for biofouling detection, which is based upon processing object images obtained in visible and near infrared spectral bands and allows for parallel processing of large image collections.

The method for detection of biofouling

This method is based upon implicit measurement of spectral characteristics of biological and non-biological substances found on an object's surface, and a subsequent recognition of substance types [Grishkin, Kovshov, ..., 2015]. A proposed method is based upon building a recognition system. It uses the assessed ratios of spectral characteristics to build a vector of informative features. We suggest using feature vector V consisting of 9 components. For recognition of the types of biofouling, a standard classifier is trained against several training sets of the real objects' images. Pictures of the objects were taken in the visible and near infrared bands of the spectrum. For training, a cross-validation technique was used.

For a correct evaluation of feature vectors one needs to convert all used object images to a common shooting point, and to remove background from each image. On the first stage of preprocessing, we pick a base image from the collection. Each image in a collection is converted to a shooting point and an angle of the base image. To achieve that, for each image in the collection, we evaluate a homographic matrix. To evaluate the homographic matrix, we look for key points in basis and current images of a collection. Then, for each key point of any given image, a matching point is found in the base image. With help of these point sets, the algorithm searches for suitable values of homographic transformation parameters. After that, using these suitable values, a homographic transformation is applied to the current image of the series. For the search for key points we use Surf algorithm [Bay, Ess, ..., 2008]. A match between key points of base and current images is found using Ransac method [Fisher, Bolles, 2003].

On a second stage of preprocessing we shall remove the background. For that sake, we shall segment all images of a collection being processed in order to find an object of interest. With help of one of the methods of interactive segmentation (we use a method called GrabCut [Rother, Kolmogorov, ..., 2004], the object of interest is found in the base image. If the segmentation is visually confirmed to be successful, background is removed from the base image, and a binary mask of the base image is built. Then, a logical AND is applied to all converted images of the series and a mask of the base image. This yields a new series of images with segmented objects only, without background [Grishkin, Zhabko, ..., 2014].

Distributed implementation of the method

The proposed method has a clear parallelism by data. Each image in the series, except for the base image, can be processed independently of all others. The method also has two non-parallel branches. One of them is bound to processing the training set of images and to classifier training. Another branch is bound to processing the base image. Data obtained by these branches are used during parallel processing of all images in the series.

The classifier is trained against a pre-selected training set of images of objects having known types of biofouling on their surface. The training set consists of object images taken in visible and near infrared bands of the spectrum. As a result of preprocessing, each infrared image gets converted to a shooting point and an angle of a corresponding visible-band image.

The markup of a preprocessed training set involves participation of an expert. The expert analyzes the converted image taken in the visible band of the spectrum and pinpoints a set of small areas corresponding to one or the other kind of a biological destructor, together with areas of the object not affected by biofouling. For each pinpointed area, averaged values of feature vector V are evaluated and stored into a training set of feature vectors. After processing all training sets, we create a pinpointed training series of feature vectors that will later be used for training the classifier. In our work we use a support vector machine (SVM) [Steinwart, Christmann, 2008] with RBF kernel as a classifier.

Another non-parallel branch of the algorithm processes the base image only. It interactively segments the base image to find an object and creates a binary mask of the segmented object. Besides that, in frames of this branch we search for key points of the base image. The key points and the mask are then used during parallel processing of images in the series.

Non-parallel branches of the algorithm are shown in Fig. 1 and Fig. 2. Both these branches are executed on a local workstation that is not part of the computing cluster.

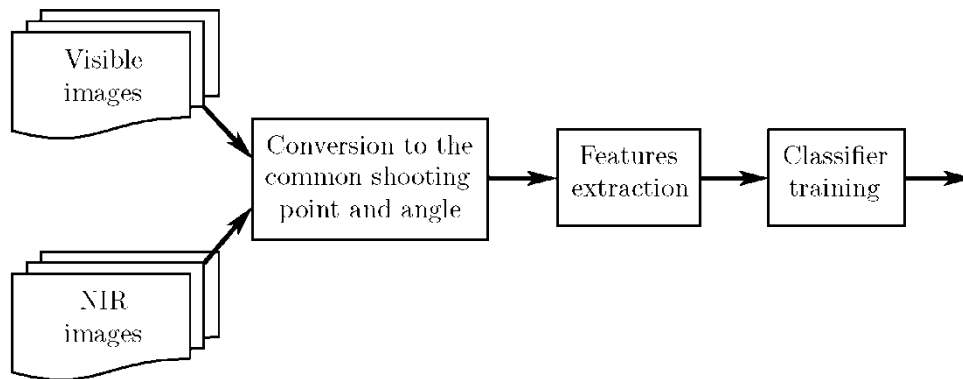


Fig. 1. First non-parallel branch of algorithm

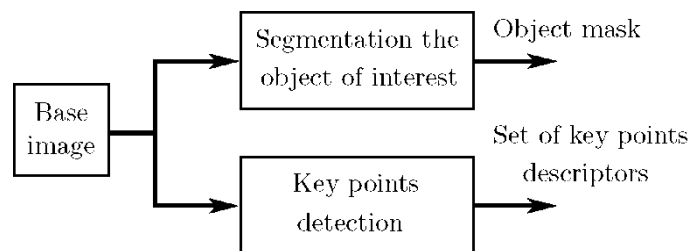


Fig. 2. Second non-parallel branch of algorithm

Within a parallel branch of the algorithm, each pair of images (one visible-band and one infrared) is processed by one cluster node. A processing algorithm run by each cluster node is shown in Fig. 3. While processing one certain series of images, each cluster node loads a set of key points of the base image, a mask of the segmented object, and parameters of the trained classifier. In each pair of images in the series, we find the key points. Using these key points and the key points of the base image, we evaluate a homographic matrix for subsequent conversion of the current image.

Then the converted image gets segmented using an object mask. Results of this processing are images of the object in the visible and infrared bands converted to a common shooting point. Both these images have the background removed. For each pixel that is part of an object, we evaluate a feature vector, which is fed to a trained classifier for recognition of biofouling type. The final result is a map of biofouling on an object's surface.

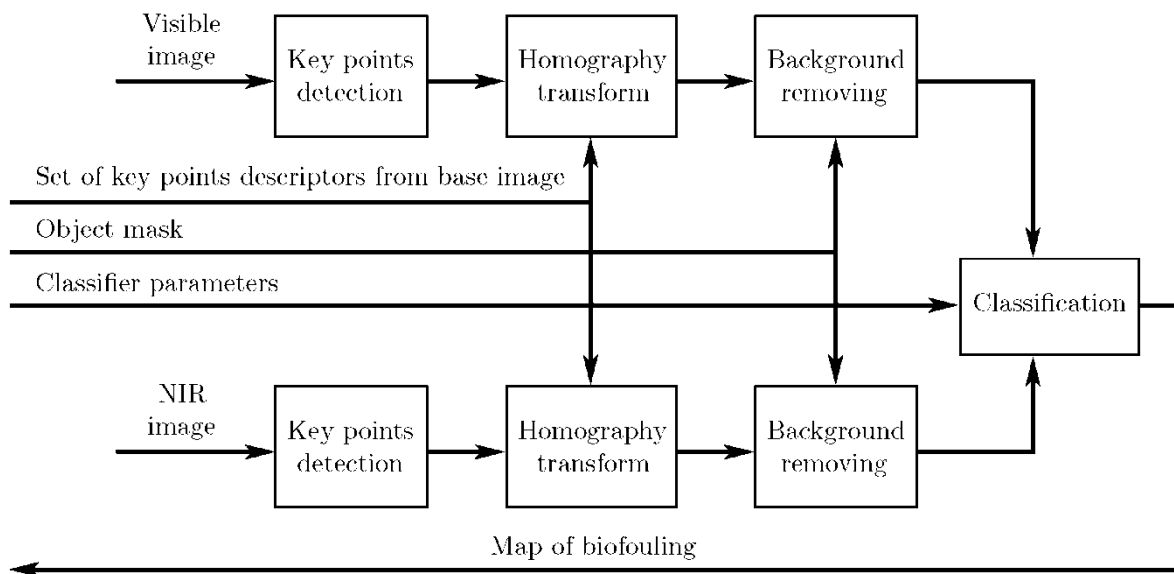


Fig. 3. Parallel branch of algorithm

An implementation of the system

Some components of a distributed system run on a computing cluster; others run on a local computer. The local computer hosts non-parallel branches of the image processing algorithm, namely interactive segmentation and the search for key points in the base image. Processing results from non-parallel branches of the algorithm are loaded into the central host of the cluster.

The computing cluster hosts parallel branches of the algorithm using standard parallel computing libraries such as MPI [Jina, Jespersena, ..., 2011]. The central host of the cluster solves the problems of data distribution among cluster nodes, synchronization of nodes, and is also used to store results. Other nodes of a cluster process all images in a series, except for the base image. They implement the search for key points, convert all images to a common shooting point, remove background, and identify types of biofouling. After processing, the map of pollution is formed for each image. This map is sent to the storage located at the central node of the cluster.

Experimental results

For training the classifier, we used a series of 95 images of biological destructors taken by photographing cultural heritage monuments made of stone. Prior to our experimenting, an expert classified this series into 9 classes of most dangerous biological destructors found at monuments, and 6 kinds of materials the monuments have been made of.

After the training was done, at the control set of 630 images, our system was able to properly classify 92% of biological destructors and building materials. An example of the monument's source image and its corresponding map of distribution of biological destructors are shown in Fig. 4. Execution of a distributed algorithm on a computing cluster of eight nodes allowed for 7.5-times reduction in processing time compared to serial processing on a local workstation.



Fig. 4. Source image of monument with natural background and map of biofouling: ■ — monument material 25%, ■ — fungi 46%, ■ — colonies of moss 11%, ■ — lichens 18%.

Conclusion

In this paper we propose the distributed system for processing a series of object images photographed in a natural environment. The system converts all the images in one series to a common shooting point and angle, segments the image of the object itself, and removes the background. The system recognizes types of biological pollutants found on the object's surface and builds pollution distribution maps for all images in the series. Experimental results show the effectiveness of the proposed recognition system in assessment of biofouling. Use of parallel processing technique allows for substantial reduction in processing time for large series of images.

References

- Grishkin V. M., Kovshov A. M., Schigorec S. B., Vlasov D. Yu., Zhabko A. P., Iakushkin O. O. A system for the recognition of biofouling on the surface of the monuments of cultural heritage // in 2015 International Conference on "Stability and Control Processes" in Memory of V. I. Zubov. // SCP 2015 – Proceedings. — 2015. — P. 630–633. — doi:10.1109/SCP.2015.7342244
- Rother C., Kolmogorov V., Blake A. GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts. // ACM Transactions on Graphics (SIGGRAPH'04). — 2004.
- Bay H., Ess A., Tuytelaars T., Van Gool L. SURF: Speeded Up Robust Features // Computer Vision and Image Understanding (CVIU). — 2008. — Vol. 110, No. 3. — P. 346–359.
- Fischler M., Bolles R. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography // Comm. of the ACM 24. — 2003. — P. 381–395.
- Steinwart I., Christmann A. Support Vector Machines. — New York, USA: Springer-Verlag. 2008. — 610 p.
- Jina H., Jespersena D., Mehrotraa P., Biswasa R., Huangb L., Chapmanb B. High performance computing using MPI and OpenMP on multi-core parallel systems // Par. Computing. — 2011. — Vol. 37, Issue 9. — P. 562–575.
- Grishkin V. M., Zhabko A. P., Vlasov D. Yu., Schigorec S. B. Multiple segmentation of the image series // in Int. Conf. Computer Technologies Physical and Engineering Applications (ICCTPEA), 2014. — doi:10.1109/ICCTPEA.2014.6893275