Computers in Biology and Medicine 64 (2015) 101–116

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Automatic detection and morphological delineation of bacteriophages in electron microscopy images

A. Gelzinis [a], A. Verikas [a,b,*], E. Vaiciukynas [a,c], M. Bacauskiene [a], S. Sulcius [d,g], E. Simoliunas [e], J. Staniulis [f], R. Paskauskas [d,h]

[a] Department of Electric Power Systems, Kaunas University of Technology, Studentu 50, LT-51368 Kaunas, Lithuania
[b] Intelligent Systems Laboratory, Centre for Applied Intelligent Systems Research, Halmstad University, Kristian IV:s väg 3, Box 823, S-30118 Halmstad, Sweden
[c] Department Information Systems, Kaunas University of Technology, Studentu 50, LT-51368 Kaunas, Lithuania
[d] Marine Science and Technology Center, Klaipeda University, H. Manto 84, LT-92294 Klaipeda, Lithuania
[e] Department of Molecular Microbiology and Biotechnology, Institute of Biochemistry, Vilnius University, Lithuania
[f] Laboratory of Plant Viruses, Nature Research Centre, Institute of Botany, Mokslininku 12, LT-08662 Vilnius, Lithuania
[g] Open Access Centre for Nature Research, Nature Research Centre, Akademijos str. 2, LT-08412 Vilnius, Lithuania
[h] Laboratory of Algology and Microbial Ecology, Nature Research Centre, Mokslininku 12, LT-08662 Vilnius, Lithuania

ARTICLE INFO

Article history:
Received 20 March 2015
Accepted 18 June 2015

Keywords:
Bacteriophage
Vb-AphaS- CL131
Aphanizomenon flos-aquae
Cyanophage
Cyanobacteria
Electron microscopy
Pattern recognition
Random forest
Open active contours
Bland–Altman

ABSTRACT

Automatic detection, recognition and geometric characterization of bacteriophages in electron microscopy images was the main objective of this work. A novel technique, combining phase congruency-based image enhancement, Hough transform-, Radon transform- and open active contours with free boundary conditions-based object detection was developed to detect and recognize the bacteriophages associated with infection and lysis of cyanobacteria *Aphanizomenon flos-aquae*. A random forest classifier designed to recognize phage capsids provided higher than 99% accuracy, while measurable phage tails were detected and associated with a correct capsid with 81.35% accuracy. Automatically derived morphometric measurements of phage capsids and tails exhibited lower variability than the ones obtained manually. The technique allows performing precise and accurate quantitative (e.g. abundance estimation) and qualitative (e.g. diversity and capsid size) measurements for studying the interactions between host population and different phages that infect the same host.

1. Introduction

Automatic detection and recognition of viruses and other objects in microscopy images is still an active field of research [1–3]. Much work in this area still remains in the form of conventional microscope analysis, which is time consuming and labor intensive, especially when objective quantitative measurements need to be performed. For example, derivation of quantitative abundance estimates requires recognizing and counting objects in hundreds or even thousands of microscopy images. According to Glaeser [4], automatic methods with accuracy higher than 75% compared to a human expert are required. Majority of approaches require human intervention via the choice of reference images for template building [5], deciding on parameter values determining particles of interest [5] or rejecting least confident cases to avoid false positives [6]. Further improvements of automated object detection and recognition techniques are needed to completely eliminate the need for human intervention. It is worth noting that human experts sometimes also disagree [7].

Martin et al. [1] developed tools to detect spherical virus particles based on analysis of pixel intensity levels inside a virus particle and in the background. The technique developed by Yu and Bajaj [8] to detect circular and rectangular particles in electron micrographs is based on geometric features. Features, extracted in frequency (Fourier) domain, were used by Matuszewski and Shark [9] in the iterative Bayesian classifier to discriminate between four types of viruses.

* Corresponding author at: Department of Electric Power Systems, Kaunas University of Technology, Studentu 50, LT-51368 Kaunas, Lithuania. Tel.: +46 35 167140.
E-mail addresses: adas.gelzinis@ktu.lt (A. Gelzinis), antanas.verikas@hh.se (A. Verikas), evaldas.vaiciukynas@ktu.lt (E. Vaiciukynas), marija.bacauskiene@ktu.lt (M. Bacauskiene), sigitas.sulcius@jmtc.ku.lt (S. Sulcius), eugenijus.simoliunas@bchi.vu.lt (E. Simoliunas), juozas.staniulis@botanika.lt (J. Staniulis), ricardas.paskauskas@jmtc.ku.lt (R. Paskauskas).

http://dx.doi.org/10.1016/j.compbiomed.2015.06.015
0010-4825/

Proenca et al. [2] developed a fully automatic technique to detect virus particles in transmission electron microscopy images and categorize found particles into three classes: *Intact*, *Permeated* and *Damaged*. Extraction of regions of interest, selection of candidates using morphological features, credibility test of the candidates, based on features extracted from radial intensity profiles, and an input image specific final validation in a three-parameter space are the main steps of the algorithm. The algorithm was tested using 26 images containing 139 particles and identified 85% of particles correctly (with 15% of false negatives).

To detect cancerous cells in microscopy images, Vigneron et al. [10] applied adaptive filtering and hypothesis testing. Location of cells is formulated as a peak finding task in a view generated by correlating an image of interest with a ring-shaped filter to match circular objects. After applying the watershed transform to split connected cells, the cell detection and counting task is formulated as a hypothesis testing problem. According to the authors, the technique detects $\sim$97% of the cells.

Several machine learning-based approaches have been developed for particle selection from electron microscopy images. Ogura and Chikara [11] trained a multilayer perceptron and achieved particle recognition accuracy of 98%. Sorzano et al. [12] applied an ensemble of Naive Bayes classifiers trained using a set of rotationally invariant features. To distinguish between seven species of *Eimeria*, a protozoan parasite of domestic fowl, Castanon et al. [13] trained a Gaussian classifier using 13 features reflecting multiscale curvature, geometry and texture of oocysts of the species extracted from microscopy images. An overall correct classification accuracy of 87.75% was achieved.

Detection and segmentation of cell nuclei is a widely studied task, which closely resembles the task of our study. Active contours [14] and gradient flow tracking [15] are two examples of techniques developed to achieve robust localization and segmentation of cell nuclei.

Segmentation and classification of nanoparticles is a task similar to cell segmentation and categorization. The technique developed by Park et al. [16] focuses on automated morphology analysis of partially overlapping nano-particles in electron microscopy images. Ultimate erosion for overlapping convex sets, extraction and association of contour evidences, and expectation maximization-based contour evolution with multiple reference shapes are the main steps of the technique. A contour is modeled as a uniform periodic B-spline curve.

Aquatic bacteriophages (viruses that infect bacteria) are now recognized to be the most abundant biological entities in the global ocean with estimated total counts of between 10^4 and 10^8 viruses in 1 ml of water sample [17]. Viruses are the major source of bacterial mortality and significantly contribute to the control of host population dynamics, and, consequently, to overall functioning of the aquatic ecosystem [18,19]. Viruses are also extremely diverse with morphologies ranging from a variety of polyhedral tailed phages to long filamentous, spindle or lemon-shaped viruses [20]. The vast majority of all observed viruses (*sim*96%) belong to order Caudovirales [21,20] and are divided into three families (Myoviridae, Siphoviridae and Podoviridae) according to their relative proportion of the capsid and tail structures [22]. It has been suggested that the manner and degree to which virus affects host population dynamics is intrinsically related to virus morphology and size [23,24]. For example, by using viruses, belonging to the three different above mentioned families, capable to infect the same host, Holmfeldt et al. [25] have found that the success of host infection by a virus varied between distinct morphotypes. Bacteriophage morphology can also reflect virus life cycle and host range. It has been experimentally demonstrated that members of Podoviridae are extremely host specific compared to viruses that belong to the family Myoviridae and are capable of infecting many different host strains [26,27]. However, despite differences in host range, both podo- and myo-viruses more often exhibit a lytic infection cycle compared to the members of Siphoviridae that frequently turn into lysogenic interactions with its host [23]. In addition, the size of the virus is also an important parameter of virus–host interactions, since it strongly determines virus progeny (burst size) formation within the cell [28]. For example, a larger virus capsid diameter results in a smaller burst size, which in turn requires higher infection and host mortality rates to maintain the abundance of virus population similar to that of smaller capsid size viruses. Therefore, such differences between morphologically distinct phages that infect the same host have significant implications for the ecology of virus–host interactions. On the other hand, clonal composition of host population can also result in production of different morphological types of viruses as it has been shown for some cyanobacteria [29]. Therefore, to better understand the interactions between host population and different phages that infect the same host, precise and accurate tools for quantitative (e.g. abundance estimation) and qualitative (e.g. diversity and capsid size) measurements are necessary.

The main purpose of this work is automated detection and recognition of specific types of co-occurring bacteriophages in electron microscopy images. Quantitative characterization of detected bacteriophages is another important task of this study. For modeling purposes, we use cyanobacteria *Aphanizomenon flos-aquae* and two different *A. flos-aquae* infecting bacteriophages. Cyanobacteria *A. flos-aquae* are globally distributed species in temperate aquatic ecosystems that regularly produce harmful blooms. Bacteriophages was recently shown to play an important role in controlling the structure and dynamics of *A. flos-aquae* population [30], suggesting that they can significantly contribute to the development and decline of cyanobacterial blooms. Thus the study of a *A. flos-aquae*-bacteriophages system has not only fundamental (e.g. study of evolution of virus–host interactions) but also considerable ecological importance. The technique combines phase congruency-based image preprocessing [31], detection of circular objects using the Hough transform, Radon transform- and open active contours with free boundary conditions-based detection of linear objects, classification of objects, and RF-based quality assessment of virus tails allowing to associate a phage tail with a relevant capsid.

2. Data

For the experiments we used a mixture of three bacteriophages (Figs. 1 and 2): bacteriophage *Vb-AphaS-CL131* (hereafter virus 1; Fig. 1a), bacteriophage *T4* (hereafter virus 2; Fig. 1b) and bacteriophage *Vb-AphaM-CL132* (hereafter virus 3; Fig. 1c), all representing morphologically different types of tailed phages belonging to the order *Caudovirales* (Table 1). Both bacteriophages *Vb-AphaS-CL131* and *Vb-AphaM-CL132* are associated with infection and lysis of cyanobacteria *A. flos-aquae*, and can be found together in the purified lysates of *A. flos-aquae*. These bacteriophages represent different morphological families of viruses (Table 1) that were shown to have the different infection strategies. Bacteriophage *T4* was added to the *A. flos-aquae* lysate containing *Vb-AphaS-CL131* and *Vb-AphaM-CL132* bacteriophages as a control bacteriophage to ensure accurate and precise measurements of *A. flos-aquae* viruses. Bacteriophage *T4* has been studied extensively by using a variety of techniques, and the structures of the head and tail are well morphologically characterized [32], making T4 a reference bacteriophage in electron microscopy studies.

Morphological and structural characteristics of all three bacteriophages used in this study are given in Table 1. Bacteriophage *Vb-*