Automated Identification and Classification of Rotavirus-A Particle in Digital Microscopic Images

P.S.Hiremath
Dept. of Computer Science
Gulbarga University
Gulbarga. Karnataka. India

Parashuram Bannigidad
Dept. of Computer Science
Gulbarga University
Gulbarga. Karnataka. India

Manjunath Hiremath
Dept. of Computer Science
Gulbarga University
Gulbarga. Karnataka. India

ABSTRACT
Image processing and computer modelling are important tools in most medical imaging domains, and have also drawn the attention of the biological community to biological imaging applications. To date, many of biological data analysis necessitate a considerable degree of human intervention. Manual procedures are based on subjective human interpretation, are prone to large variability between the human experts, are time consuming and are of high cost. Automated tools are, thus, important in achieving objective and repetitive analysis, accurate quantitative measurements and the analysis of increasing data volumes. The objective of the present study is to develop an automatic tool to identify and classify the Rotavirus-A particles in digital microscopic images. Geometric features are used to identify and classify the Rotavirus-A particle. The proposed method yields 98% classification rate using 3σ classifier.

Keywords
Rotavirus-A, image segmentation, classification, image analysis, watershed segmentation, 3σ classifier.

1. INTRODUCTION
A virus is an infectious agent too small to be seen directly with a light microscope. They are not made of cells and can only replicate inside the cells of another organism (the viruses' host). Viruses infect all types of organisms, from animals and plants to bacteria and archaea. Viruses are found in almost every ecosystem on Earth and these minute structures are the most abundant type of biological entity. Viruses consist of two or three parts: all viruses have genes made from either DNA or RNA, long molecules that carry genetic information; all have a protein coat that protects these genes; and some have an envelope of fat that surrounds them when they are outside a cell. Viruses vary from simple helical and icosahedral shapes, to more complex structures. Most viruses are about one hundred times smaller than an average bacterium. Viruses spread in many ways ; plant viruses are often transmitted from plant to plant by insects that feed on sap, such as aphids, while animal viruses can be carried by blood-sucking insects. Influenza viruses are spread by coughing and sneezing.

Rotavirus is a genus of double-stranded RNA virus in the family Reoviridae. It is the leading single cause of severe diarrhoea among infants and young children, and is one of several viruses that cause infections commonly known as stomach flu, despite having no relation to influenza. By the age of five, nearly every child in the world has been infected with rotavirus at least once. However, with each infection, immunity develops, subsequent infections are less severe, and adults are rarely affected. There are seven species of this virus, referred to as A, B, C, D, E, F and G. Humans are primarily infected by species A,B and C, most commonly by species A [4]. All seven species cause disease in other animals.

Within rotavirus-A, there are different strains, called serotypes. As with influenza virus, a dual classification system is used, which is based on two structural proteins on the surface of the virion. The glycoprotein VP7 defines G-types and the protease-sensitive protein VP4 defines P-types. Strains are generally designated by their G serotype specificities (e.g., serotypes G1 to G4 and G9), and the P-type is indicated by a number and a letter for the P-serotype and by a number in square brackets for the corresponding P-genotype. (P-serotypes are difficult to characterize; therefore, molecular methods based on sequence analysis are often used to define the corresponding P-genotype instead. These genotypes correlate well with known P-serotypes). Because the two genes that determine G-types and P-types can be passed on separately to offspring, various combinations occur in any one strain [5].

Automated image analysis of rotavirus particles will play an important role to identify rotavirus particles by using digital image processing techniques. Previously, the segmentation and statistical analysis of individual rotavirus particles is done by Venkataraman, et al. [6]. The automatic classification of bacterial cells in digital microscopic images using simple shape geometric features is studied by Hiremath and Parashuram [9].

Segmentation of cell regions is an important step in computer-aided analysis of rotavirus particle images in microbiology. Accurate and reliable segmentation is an essential step in determining valuable quantitative information on size, shape and texture, which may assist microbiologists in their diagnoses. Parametric algorithms were recently introduced for segmentation of cell images with elliptically shaped cells or cells whose contours were relatively smooth. Nonparametric algorithms can generally be categorized as region-based, edge-based, histogram-based, clustering and neural network based algorithms. The thresholding based on histogram approximation using the step-wise functions can be found in the literature. The method ignores the local correlations in images. Edge or gradient-based segmentation methods use the discontinuity of image intensities or texture at the boundary between different objects.

For original noiseless cell images, most algorithms work fine. However, when an image is corrupted by heavy noise, these algorithms may not produce satisfactory segmentation results. The algorithm introduced in this paper deals with the problem of rotavirus-A particle identification and classification from images

corrupted by heavy noise. The following sections discuss the image acquisition, image processing and feature extraction steps in more detail. The data used for development of the algorithm consist of rotavirus-A particle images acquired from child faeces of an infected child [4] by the Transmission Electron Microscope (Figure 1).

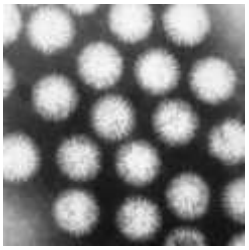


Figure 1. Rotavirus-A TEM image from the faeces of an infected child.

2. MATERIALS AND METHODS

Rotavirus-A particle images on Transmission Electron Microscopy is acquired by using different types of staining methods. Negative staining is an established method, often used in diagnostic microscopy, for contrasting a thin specimen with an optically opaque fluid. For bright field microscopy, negative staining is typically performed using a black ink fluid such as nigrosin. The specimen, such as a wet bacterial culture spread on a glass slide, is mixed with the negative stain and allowed to dry. When viewed with the microscope the rotavirus particles, and perhaps their spores, appear light against the dark surrounding background. An alternative method has been developed using an ordinary waterproof marking pen to deliver the negative stain. In the case of transmission electron microscopy, opaqueness to electrons is related to the atomic number, i.e., the number of electrons. Some suitable negative stains include ammonium molybdate, uranyl acetate, uranyl formate, phosphotungstic acid, osmium tetroxide, osmium ferricyanide and auroglucothionate. These have been chosen because they scatter electrons well and also adsorb to biological matter well. Acquisition of rotavirus particle images is usually done using a CCD camera mounted in the optical path of the microscope. The camera may be full colour or monochrome. Very often, very high resolution cameras are employed to gain as much direct information as possible. Cryogenic cooling is also common, to minimize noise. Often digital cameras used for this application provide pixel intensity data to a resolution of 12-16 bits. We have considered 50 color images of rotavirus-A particles with x82,000 magnification for present study and these are converted into greyscale images [2].

3. IMAGE ANALYSIS

The image processing procedure can be divided into three major phases, namely, *image pre-processing, image segmentation,* and *image post processing*. The pre-processing stage usually includes image enhancement of acquired image and is essentially performed in order to prepare the image for the vital segmentation stage. Individual objects of interest are separated from the background in the segmentation process [7]. This is followed by a labelling operation (post-processing) in which, segmented objects of interest are tagged with unique labels that can be used to count the number of objects in the image. These labels along with spatial information of the segmented objects are used for the subsequent feature extraction procedure. Features are extracted and average statistics are calculated to provide an estimate of average morphology of the specimen under study.

3.1. Image pre-processing :

The procedure for image pre-processing includes the following steps : image registration from raw TEM data files; reconstruction of each image using prior knowledge of the TEM tip radius and applying a simple deconvolution step; estimation of sample background in order to calculate true Line-wise flattening which is then followed by morphological operations based image enhancement. The raw data values for each pixel for individual TEM images were imported into MATLAB and converted to a 500 x 500 matrix with each element representing a pixel of the image.

3.2. Image Segmentation :

To identify individual objects in an image, a segmentation operation is performed. Simple thresholding techniques and edge based segmentation methods like *canny* edge detector were tried and abandoned, since they were unable to identify closely spaced particles as individual objects(data not shown). For this reason, a marker controlled watershed segmentation algorithm was used. A marker controlled watershed algorithm is used to segment the objects of interest (rotavirus particles) from the background. The watershed finds "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low. This works better if foreground and background regions are marked prior to use. The binary image obtained from the above image pre-processing stage is used to mark the foreground. The black pixels are considered as background but the background markers should not be too close to the edges of the objects that are to be segmented. To avoid this, the watershed transform of the distance transform of the binary image obtained from the pre-processing stage is computed to identify the watershed ridge lines of the result. These lines are marked as the background. A gradient image of the labelled image is calculated so that it only has regional maxima at the foreground and background marker locations. This modified gradient image is now segmented using a watershed algorithm [8][9].

The choices of parameters for morphological operations are made in accordance with the resolution of the image. The resolution of an image determines the spatial dimensions of objects in an image. This, in turn, determines the parameters for the structuring element used in the morphological operations and these are defined by the user. Here a simple 'disk' shaped structuring element with a radius I is used to remove unwanted stray pixels in the image segmentation. However, care must be taken when choosing the shape and size of the structuring element in order to avoid the overlapping of closely spaced objects in the final image.

3.3. Post processing :

Since the image contains multiple objects of interest, the resulting binary image corresponding to the segmented objects of interest (white pixels) is shown in Fig 2(c). As part of the post processing procedure , these segmented objects have to be assigned with unique labels. This procedure is called connected component labelling; in which the grouping is done using a simple neighbourhood principle. The resultant binary image contain

relevant data at specific spatial locations, while the rest of the image is featureless and can be ignored in further processing steps. This is achieved by tabulating spatial coordinates of the various features from the labelled image, thereby enhancing the speed of subsequent feature extraction algorithms.

The proposed method for the segmentation of rotavirus-A particles is given below:

Algorithm 1: ***Training phase***
Step 1 : Input the rotavirus-A colour image.
Step 2 : Convert the colour image into gray scale image.
Step 3 : Perform pre-processing by using morphological operations, namely, erosion , reconstruction and dilation.
Step 4 : Segment the image of step 3 using marker controlled watershed segmentation method and obtain resulting binary image.
Step 5 : Remove the border touching cells obtained in binary image and then perform labelling the segmented binary image.
Step 6 : For each labeled segment, compute geometric shape features x_i, i=1,2,...,7(i.e, Area, Eccentricity, Perimeter, Circularity, Tortuosity, Length/Width ratio, Compactness, respectively), for each particle.
Step 7 : Repeat steps 1 to 6 for all the training images.
Step 8 : Compute minimum $minx_i$ and maximum $maxx_i$ of feature values x_i, i=1,2,...,7, of rotavirus-A particle and store them as knowledge base.

Algorithm 2: ***Testing phase***
Step 1 : Input the rotavirus-A colour image.
Step 2 : Convert the colour image into grayscale image.
Step 3 : Perform pre-processing by using morphological operations, namely, erosion , reconstruction and dilation.
Step 4 : Segment the image of step 3 using marker controlled watershed segmentation method and obtain resulting binary image.
Step 5 : Remove the border touching cells obtained in binary image and then perform labelling the segmented binary image.
Step 6 : For each labeled segment, compute geometric shape features x_i, i=1,2,...,7(i.e, Area, Eccentricity, Perimeter, Circularity, Tortuosity, Length/Width ratio, Compactness, respectively), for each segment.
Step 7 : Apply rule for classification of the Rotavirus-A particles: A segmented region is of Rotavirus-A, if its feature values lie in the range of minimum and maximum values.
Step 8: Repeat the steps 6 and 7 for all labelled segments and output the identified rotavirus-A particles.

3σ classification rule:

The above algorithm for training phase can be modified to compute the mean and standard deviation of feature values x_i, i=1,2,...,7, for known rotavirus-A particles and store them as the knowledge base, $\bar{x}_i$ and σ_i, i=1,2,...,7. The algorithm for classification phase can be modified to apply the following classification rule in the Step 7 :

A segmented region is rotavirus-A particle if its feature values lie in the 3σ interval, i.e., $\bar{x}_i$ lies in $[3\sigma_i-\bar{x}_i, 3\sigma_i+\bar{x}_i]$, i=1,2,...,7.

4. EXPERIMENTAL RESULTS

For the purpose of experimentation, 50 digital images of rotavirus-A particles (non-overlapping) with x82,000 magnification are considered which are taken from transmission electron microscopy. The implementation is done on a Intel core 2 duo processor @ 2.83GHz machine using MATLAB 7.9. The input of rotavirus-A particle image is converted into gray scale image, and the morphological operations such as opening, closing, erosion and dilation, reconstruction are applied. The resulting image is segmented using marker controlled watershed segmentation method to obtain binary image. The segmented image is labelled and for each segmented region (known particles), the geometric features are extracted. The Table 1 presents the geometric feature values computed for the segmented rotavirus-A particle regions of the image in Figure 2. The minimum and maximum values of geometric features and also their mean and standard deviation values are stored in the knowledge base of the rotavirus-A particle and are presented in Table 2. Some sample training images of rotavirus-A are shown in Figure 3.

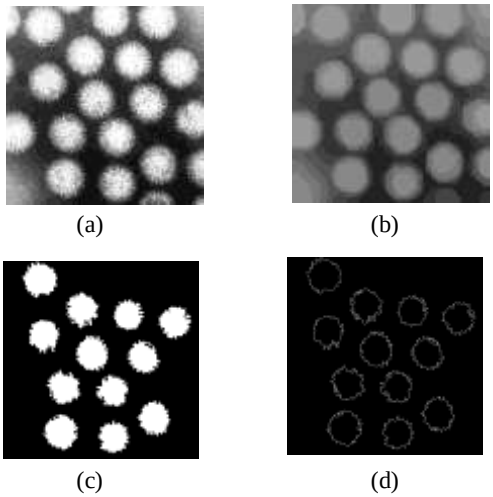


Figure 2. (a) Original Rotavirus-A TEM image.
(b) Morphology operation on grayscale image.(c) regional maxima of enhanced image (d) perimeter of the segmented regions.

Table 1. The geometric feature values of the rotavirs-a particles of the image in fig. 2(c).

Rotavirus-A features	Area	Eccentricity	Perimeter	Circularity	Tortuosity	Length/ Width ratio	Compactness
Particle 1	3215.00	0.52	385.99	0.27	0.18	1.17	3.69
Particle 2	5320.00	0.24	474.76	0.30	0.18	1.03	3.37
Particle 3	4128.00	0.40	447.45	0.26	0.17	1.09	3.86
Particle 4	3754.00	0.28	453.14	0.23	0.16	1.04	4.35
Particle 5	4821.00	0.28	480.07	0.26	0.17	1.04	3.80
Particle 6	3274.00	0.35	425.45	0.23	0.16	1.07	4.40
Particle 7	4485.00	0.36	436.03	0.30	0.18	1.07	3.37
Particle 8	4167.00	0.27	432.86	0.28	0.17	1.04	3.58
Particle 9	4397.00	0.51	418.13	0.32	0.20	1.17	3.16
Particle 10	5276.00	0.43	493.97	0.27	0.18	1.11	3.68
Particle 11	4508.00	0.32	416.37	0.33	0.19	1.05	3.06

In the testing phase, the test image is segmented by the proposed method, and the geometric features are extracted for each segment. The proposed method is computationally less expensive and yet yields comparable classification rates with 96% for rotavirus-A particle. However, the proposed method yields 98% classification rate using 3σ classifier. Figure 4. shows some sample test images used for classification of rotavirus-A particles.

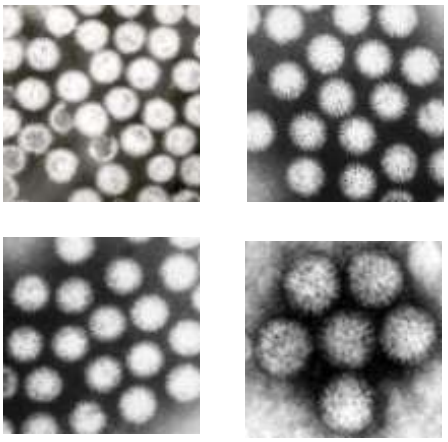


Figure. 3 Sample training images of Rotavirus-A particles.

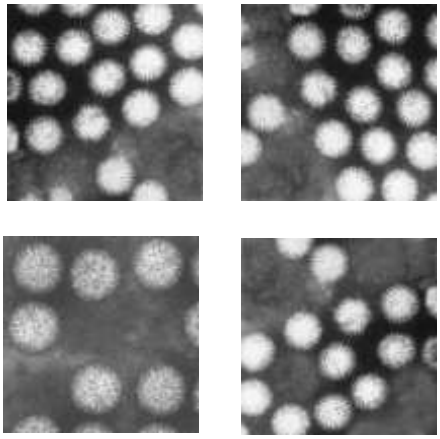
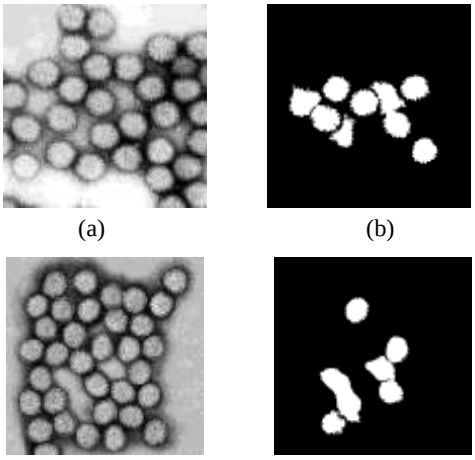


Figure. 4 Sample Test images used for classification of Rotavirus-A particles.

The Figure 5 shows some sample rotavirus-A particle images corresponding to misclassification results. In Fig. 5 (a)&(c), a rotavirus-A particle is not classified (i.e., unknown) due to over segmentation. These problems can be overcome by employing better segmentation methods. Further, the classification results can be improved by using better feature sets and classification techniques. The identification and classification of different species of rotavirus particles and classification of different genera of rotavirus included family, will be considered for our future work.

Table 2. The knowledge base for rotavirus-a particles

Features	Area	Eccentricity	Perimeter	Circularity	Tortuosity	Length/ Width ratio	Compactness
Minimum Value	2999	0.14	261.28	0.16	0.13	1.01	1.73
Maximum Value	12315	0.63	967.90	0.58	0.25	1.29	6.42
Mean $\bar{x}_i$	4480.79	0.34	427.48	0.32	0.19	1.07	3.35
STD σ_i	988.49	0.09	69.39	0.09	0.03	0.04	0.92



(a) (b) (c) (d)

Figure. 5 some sample rotavirus-A particle images corresponding to misclassification results, (a)&(c) original rotavirus-A particles, (b)&(d) segmented images corresponding to rotavirus-A in (a)&(c).

5. CONCLUSION

In this paper, we have proposed an automated rotavirus-A particle image segmentation and classification of transmission electron microscope images and extracting geometric features of rotavirus-A particles. The experimental results are compared with the manual results obtained by microbiological experts. The proposed method is more reliable and computationally less expensive. It yields a classification rate of 96% using min-max classification rule and 98% using 3σ classifier for rotavirus-A particles. It could be improved further by better pre-processing methods, feature sets and classifiers. The classification of other rotavirus particles also will be considered in future work.

6. ACKNOWLEDGEMENTS

The authors are indebted to the referees for their helpful comments and suggestions. Further, the authors are grateful to Dr. A Dayanand, Professor of Microbiology, Gulbarga University, Gulbarga and Dr. Ramakrishna, Department of Microbiology, Government Degree College, Gulbarga for providing rotavirus-A particle images and manual results of the rotavirus-A images by visual inspection.

7. REFERENCES

[1] Rafel C. Gonzalez and Richard E. Woods, "Digital Image Processing", Pearson Education Asia (2002)

[2] M. A. Hayat (2000). *Principles and techniques of electron microscopy: biological applications*. Cambridge University Press. pp. 45-61. ISBN 0521632870.

[3] Bozzola, John J.; Russell, Lonnie D. (1999). "Specimen Preparation for Transmission Electron Microscopy". *Electron microscopy : principles and techniques for biologists*. Sudbury, Mass.: Jones and Bartlett. pp. 21–31. ISBN 9780763701925.

[4] Velázquez FR, Matson DO, Calva JJ, Guerrero L, Morrow AL, Carter-Campbell S, Glass RI, Estes MK, Pickering LK, Ruiz-Palacios GM (1996). "Rotavirus infections in infants as protection against subsequent infections". *N. Engl. J. Med.* 335 (14): 1022–8. doi:10.1056/NEJM199610033351404. PMID 8793926.

[5] Arnoldi F, Campagna M, Eichwald C, Desselberger U, Burrone OR (2007). "Interaction of rotavirus polymerase VP1 with nonstructural protein NSP5 is stronger than that with NSP2". *J. Virol.* 81 (5): 2128–37. doi:10.1128/JVI.01494-06. PMID 17182692. http://jvi.asm.org/ cgi/content/full/81/5/2128.

[6] S. Venkataraman, D.P. Allison, H. Qi, J.L. Morrell-Falvey, N.L. Kallewaard, J.E. Crowe Jr. and M.J. Doktycz. "Automated image analysis of atomic force microscopy images of rotavirus particles." *Ultramicroscopy, Volume 106, Issues 8-9, June-July 2006, Pages 829-837.*

[7] Gerald J.F. Banon, Junior Barrera, Ulisses M. Braga-Neto (Eds.) "*Mathematical Morphology and its Applications to Signal and Image Processing",* proceedings of the 8th international symposium on mathematical morphology (ISMM'07), ISBN 978-85-17-00032-4 (2007)

[8] K. Parvati,1 B. S. Prakasa Rao,2 and M. Mariya Das3, "Image Segmentation Using Gray-Scale Morphology and Marker-Controlled Watershed Transformation", Hindawi Publishing Corporation. Volume 2008 (2008), Article ID 384346, 8 pages

[9] P.S. Hiremath and Parashuram Bannigidad , Automatic Classification of Bacterial Cells in Digital Microscopic Images, ICDIP 2010, February 26-28, 2010, Singapore.

[10] Dennis Kunkel Microscopy, Inc, Science Stock Photography, [http://www.denniskunkel.com].

[11] Qiang Wu, Fatima Merchant, Kenneth R. Castlem, "Microscope Image Processing", Elsevier,2008.

[12] Al Bovik, "Essential guide to Image processing", Elsevier, ISBN: 978-0-12-374457-9 (2009).