Thresholding Approach for Automatic Detection of *Pseudomonas aeruginosa* Biofilms from Fluorescence in situ Hybridization Images

Zonglin Yang, Tatsuya Akiyama, Kerry S. Williamson, Michael J. Franklin, Thiruvarangan Ramaraj

Abstract—*Pseudomonas aeruginosa* is an opportunistic pathogen that forms surface-associated microbial communities (biofilms) on artificial implant devices and on human tissue. Biofilm infections are difficult to treat with antibiotics, in part, because the bacteria in biofilms are physiologically heterogeneous. One measure of biological heterogeneity in a population of cells is to quantify the cellular concentrations of ribosomes, which can be probed with fluorescently labeled nucleic acids. The fluorescent signal intensity following fluorescence in situ hybridization (FISH) analysis correlates to the cellular level of ribosomes. The goals here are to provide computationally and statistically robust approaches to automatically quantify cellular heterogeneity in biofilms from a large library of epifluorescent microscopy FISH images. In this work, the initial steps were developed toward these goals by developing an automated biofilm detection approach for use with FISH images. The approach allows rapid identification of biofilm regions from FISH images that are counterstained with fluorescent dyes. This methodology provides advances over other computational methods, allowing subtraction of spurious signals and non-biological fluorescent substrata. This method will be a robust and user-friendly approach which will enable users to semi-automatically detect biofilm boundaries and extract intensity values from fluorescent images for quantitative analysis of biofilm heterogeneity.

Keywords—Image informatics, *Pseudomonas aeruginosa*, biofilm, FISH, computer vision, data visualization.

I. INTRODUCTION

BIOFILMS are communities of microorganisms and their secreted extracellular polysaccharides and proteins, that are attached to surfaces [1], [2]. Biofilms can be beneficial or harmful for human activities. Beneficial biofilms include those used in wastewater treatment, where biofilm bacteria degrade environmentally impactful wastes. Harmful biofilms may cause fouling of surface such as heat exchange devices and may induce corrosion of metals. Harmful biofilms also include biofilms associated with infectious diseases, such as bacteria that are attached to surfaces of artificial surgical implant devices or human tissue [3]. Infectious biofilms can cause chronic and persistent infections that are difficult and sometimes impossible to clear with antibiotic therapies. One of the key species of bacteria that forms infectious biofilms is *Pseudomonas aeruginosa*, an opportunistic pathogen that forms biofilms on implant devices and on human tissue [4], [5], and the subject of the study presented here.

One reason that infectious biofilms are difficult to treat with antibiotics is that the biofilm bacteria are physiologically heterogeneous [6]. Bacteria growing in biofilms are not all in the same physiological state but may vary depending on their local microenvironment within the biofilm. Since oxygen and nutrients diffuse into the biofilms and are utilized by the bacteria, the physiological state of the biofilm bacteria is dependent on their proximity to the nutrient and oxygen source as well as to the waste products produced by the bacteria. Bacteria away from the nutrient sources may be in a starvation state and may enter dormancy [7]. Since the efficacy of antibiotics is dependent on the physiological status of the bacteria, subpopulations of biofilm bacteria may be tolerant to treatments that would normally kill other members of the biofilm population [7].

One overarching goal of biofilm research is to understand the physiological heterogeneity of microbial biofilms and to determine how biofilm heterogeneity influences treatment and mitigation approaches. One strategy to understand the physiological status of bacteria is to probe their cellular concentration of ribosomes [8]-[10]. Ribosomes are the most abundant macromolecules in cells and are essential for all life forms. The cellular concentration of ribosomes varies depending on the physiological status of the cells, with actively growing cells having high concentrations of ribosomes and dormant cells having a low but a maintenance level of ribosomes [11]. The cellular ribosome concentration in dormant cells is also dependent on the small ribosome-binding protein, hibernation promoting factor (HPF), which protects a subset of ribosomes from degradation when the cells are dormant [8], [12]. Bacteria require a low maintenance level of ribosomes in the dormant cells in order to resuscitate from dormancy when conditions become favorable. Because the cellular concentration of ribosomes varies depending on the physiological status of the bacteria, the concentration of ribosomes within a cell can be used as an indicator of the cellular physiological status.

Ribosomes are complex macromolecules composed of approximately 50 ribosomal proteins (rProteins) and three RNA species (rRNA). In bacteria the rRNAs are the 5S, 16S, and 23S rRNAs. Because ribosomes contain rRNA, the relative

Thiruvarangan Ramaraj* and Zonglin Yang are with School of Computing, DePaul University, Chicago, IL, USA (*corresponding author, e-mail: tramaraj@depaul.edu, zyang43@depaul.edu).
Tatsuya Akiyama is with Emory University, Atlanta, GA, USA (e-mail: tatsy1003@gmail.com).
Kerry S. Williamson and Michael J. Franklin are with the Department of Microbiology & Immunology and Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA (e-mail: ksw@montana.edu, franklin@montana.edu).

concentration of cellular ribosomes may be determined by probing the cells for their rRNA. Nucleotide probes are generated that hybridize to 16S rRNA [13]. The nucleotide probes also contain a fluorescent molecule that can be quantified or imaged with FISH which uses epifluorescence microscopy or confocal scanning laser microscopy. Since different bacterial species have different rRNA sequences, FISH is often used in microbial ecology and biofilm studies to identify species composition within a community [14]. In the study here, FISH is used to probe biological heterogeneity of single species biofilms, based on the fluorescent intensity of 16S rRNA within cells. The primary goals of this research are to: (i) develop an automated pipeline to identify biofilm regions, (ii) quantitatively analyze the abundance and distribution of 16S rRNAs within biofilm, which will ultimately provide information regarding heterogeneity, (iii) determine the effect of starvation on the physiological heterogeneity of biofilms, and (iv) determine the effect of HPFs on biofilm heterogeneity and biofilm survival during dormancy. Here the results of the first goal are presented, developing an automated pipeline to computationally identify biofilm regions from FISH images.

II. BIOLOGICAL DATA BACKGROUND

A. Bacterial Strains and Growth Conditions

P. aeruginosa strain PAO1 [7] was cultured in Tryptic Soy Broth (TSB), diluted in fresh TSB, and adjusted to an optical density at 600 nm (OD600) of 0.5 on a CE2041 Spectrophotometer (Cecil Instruments). To generate biofilms, sterile black polycarbonate filters (0.2-µm pore size, 13-mm diameter, Millipore Sigma) positioned on Tryptic Soy Agar (TSA) were inoculated with 30 µl of the diluted cultures. Plates were incubated at 37 °C, and the membranes containing biofilms were aseptically transferred to fresh TSA plates after 24 h. Following 48 h of growth on TSA, the biofilms were used for cryosectioning and FISH analysis.

B. Sample Preparation and Processing for FISH

The biofilms were cryoembedded by placing the filters containing the biofilms on stainless steel slides that were frozen on dry ice. Tissue-Tek O.C.T. Compound (Sakura Finetechnical Co.) was applied to the biofilms to embed them. Vertical sections (5-10 µm) of the biofilms were obtained by thin-sectioning of the cryo-embedded biofilms using a Leica CM1950 cryostat. The vertical biofilm sections were placed on 0.01% Poly-L-Lysine coated microscope slides and stored at -20 °C until FISH probing, SYTO 9 staining, and epifluorescence imaging. For FISH probing, the thawed biofilm sections on microscope slides were dipped in 1% agarose solution in PBS, pH7.0 and immediately dried with compressed air. Agarose coated slides were then placed in hybridization chambers and sections were fixed by applying 4% paraformaldehyde directly and incubated at 4 °C for 30 min. Slides were washed by dipping in ice-cold PBS pH 7.3. Sections were dehydrated by successive incubation of the slides in 50, 80, and 100% ethanol for 3 min each. Slides were then dried after final dehydration using compressed air. The 16S rRNA probes [13] labeled with the fluorescent tag Cy3 were hybridized to the biofilm thin sections as previously described [8], [15]. Hybridization buffer (0.9M NaCl, 20 mM Tris-HCl, pH 8.0, 0.01% SDS) containing the 16S rRNA oligonucleotide probes was directly placed onto the sections on the microscope slide in a hybridization chamber. After incubating at 46 °C for 2 h, slides were washed in buffer (0.9M NaCl, 20 mM Tris-HCl, pH 8.0) at 48 °C for 10 min. Slides were then dipped in ice-cold water and immediately dried by removing excess water using KimWipe and air drying using compressed air. Biofilm sections were covered with 1 uM SYTO 9 dye and incubated in the dark for 5 min. Slides were then dipped in ice-cold water and immediately dried using a KimWipe and compressed air. Citifluor Antifadent mounting medium, AF-1 (Electron Microscopy Sciences) and a coverslip were placed onto the slides to reduce fading of probes while imaging using an epifluorescent microscope. Filters used are as follows: FITC for SYTO 9, 480/30 excitation, 505LP mirror, 535/40 emission; TRITC for 16S-Cy3, 540/25 excitation, 565LP mirror, 605/55 emission.

III. IMAGE PREPROCESSING

The main preprocessing step was to implement an automated script for visualizing microscope image data created in the Carl Zeiss (CZI) [16] file format. The CZI image file was split into three different channels (Red, Green and Blue) to identify the appropriate channel for performing image analysis.

The images of biofilms were generated using two different fluorescent dyes, SYTO 9 and Cy3. The Cy3-labeled nucleic acid probe hybridizes to the 16S rRNA as an indicator of cellular ribosome content, while SYTO 9 is a DNA stain that stains all cells. SYTO 9 is detected using the FITC filter set and Cy3 using the TRITC filter set. Therefore, the FITC image was used to define the boundary of the biofilm and then results will be transformed to TRITC images to investigate biofilm properties, such as heterogeneity.
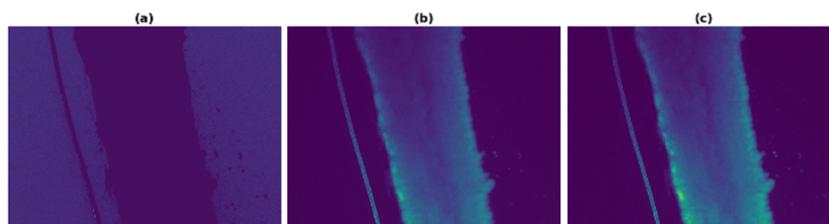

Fig. 1 FITC image. (a) Channel 0, Red (b) Channel 1, Green (c) Channel 2, Blue
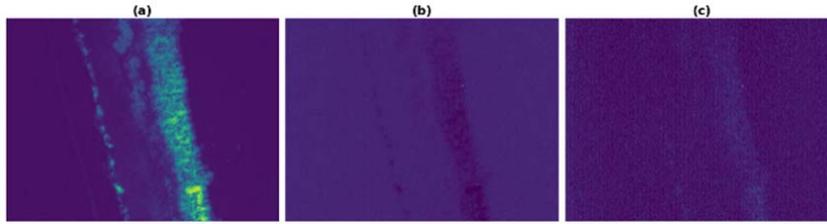
Fig. 2 TRITC image. (a) Channel 0, Red (b) Channel 1, Green (c) Channel 2, Blue

Every biofilm image is in RGB mode. The red channel was used to analyze FITC images because of the ability to clearly distinguish biofilm boundaries (Fig. 1) and because they also contained less information inside the boundaries. Biofilm-related regions, including the biofilm, filter, and outliers such as smearing of the top of the biofilm appear in darker black and with a value of 0; meanwhile, the values of background non-biofilm-related regions vary from ten to 100. The red channel was used for TRITC images because the signals are clearly visualized within this channel (Fig. 2).

IV. IMAGE PROCESSING

The FITC image, red channel (channel 0), was used to generate a 2-dimensional numeric matrix using the Czifile [21] package in Python. Using this matrix, the biofilm region (in addition to the filter and outliers) where the pixel values are zero, and the non-biofilm (background) region where the pixel values vary from ten to 100, were identified. For clearer visualization, the background region was unified by setting the varied values to a fixed value. For this analysis, the fixed value has been set to 50.

A. Filter, Background & Outlier Removal

Non-biofilm regions, including the filter that was used as a biofilm substratum, and spurious outliers, can hinder the analysis of biofilm specific regions. Hence, it was imperative to identify and remove the non-biofilm regions as much as possible. The row-wise length of the non-biofilm region will be smaller than the row-wise length of the biofilm region. By utilizing this property, a recursive method was proposed to remove the non-biofilm region. In this recursive approach, first, a list of thresholds was set, which can be modified by the user to adjust the accuracy based on the quality of the resulting images. The number of threshold values will be equal to the maximum number of iterations. Second, the non-biofilm row-wise length of all rows with threshold values in the list was iteratively compared. Within each iteration, if the non-biofilm row-wise length is smaller than the threshold, then the non-biofilm region at that row is removed. For example, if the value of the current threshold is 10, after an iteration, any non-biofilm row-wise length smaller than 10 will be removed (please refer to Pseudocode 1). Third, each iteration of removing the non-biofilm region based on comparing the row-wise x coordinate width with the threshold produced a new matrix. After each iteration, this new matrix is compared with the previous matrix, which is a matrix created without entering the current iteration for non-biofilm row-wise removal. If the two matrices are

identical, then the iteration is stopped, and this matrix is the result. Otherwise, the matrix goes to the next iteration until they are identical or until possible thresholds are exhausted.

```
Pseudocode 1: Recursively Removing Filters and Outliers

def func(input_matrix, x_coordinate, number_of_row,
number_of_column, start, end, step):
    # x_coordinate: dictionary, which contains row-wise x coordinate
interval
    threshold_list = [start, start+step, start+2*step, …, end]
    # create a list of thresholds
    matrix_previous = copy(input_matrix )
        # copy the matrix for comparison
    for threshold in threshold_list:
        if [row-wise x coordinate interval] < threshold;
            remove interval
                create a new matrix using the updated x_coordinate as
matrix_current.
                if matrix_current == matrix_previous:
            break # stop the loop
        else:
            matrix_previous = copy(matrix_current)
```

B. Rotation Angle Calculation

The main criterion for finding the rotation angle is that the biofilm left boundary should be parallel to the y-axis. To do this, a linear regression [17] was applied to the biofilm left boundary points. Once the regression line is obtained, the angle between the regression line and the y-axis is calculated. Previously, the background and biofilm regions were set to 50 and 0, respectively. In order to perform rotation for clear visualization, the value of the biofilm region is reset to 100. Then, the angle is calculated to rotate the image. The angle is recorded for the following analysis on TRITC images.

```
Pseudocode 2: Calculate rotation angle

def find_rotation_angle(matrix):
    coordinate = func(get_left_boundary_coordinate(matrix))
    coordinate_df = pandas.DataFrame(coordinate)
    coefficient = func(regression_analysis(coordinate_df))
    angle = func(convert_coeffiecient_to_angle(coefficient))
    return angle
```

```
Pseudocode 3: Performing rotation

def rotation(matrix, angle):
    reset_matrix_value(matrix)
        # reset value of biofilm region from 0 to 100
```

```
    # because performing rotation will add other slice with value of 0
    if angle < 0:
        rotated_matrix = func(rotate(matrix, -(90+angle)))
    else:
        rotated_matrix = func(rotate(matrix, -(90-angle)))
    return rotated_matrix
```

```
Pseudocode 4: Generate tailored matrix

        # This step combines finding boundaries and tailoring.

def tailoring(matrix):
    boundary = dict() # store boundary in dictionary
    LeftRight = func(find_boudary_LeftRight(matrix)) # this function
returns a dictionary
    matrix_intermediate = matrix[:,LeftRight['left']:
LeftRight['right']]
    TopBottom =
func(find_boundary_TopBottom(matrix_intermediate))
    matrix_tailored = matrix_intermediate[TopBottom['top']:
TopBottom['bottom'],:]
```

```
    boundary.update(LeftRight, TopBottom)
    return boundary, matrix_tailored
```

C. Identifying the Biofilm Boundary Site

The last analysis step is to locate the boundary of the biofilm region. For this, the left and right boundaries are identified by composing a vertical line to the x-axis with the farthest point on the left and right of the biofilm. Subsequently, using this left and right boundary, the top and bottom boundaries were identified based on distinction between the biofilm and background regions. Finally, the biofilm only matrix is generated via tailoring the matrix with the boundaries. The boundaries are then recorded for the following analysis on TRITC images.

Once the rotation angle and the biofilm boundaries have been identified using the FITC images, they can be applied to TRITC images for further analysis. The complete workflow is represented pictorially in Fig. 3.
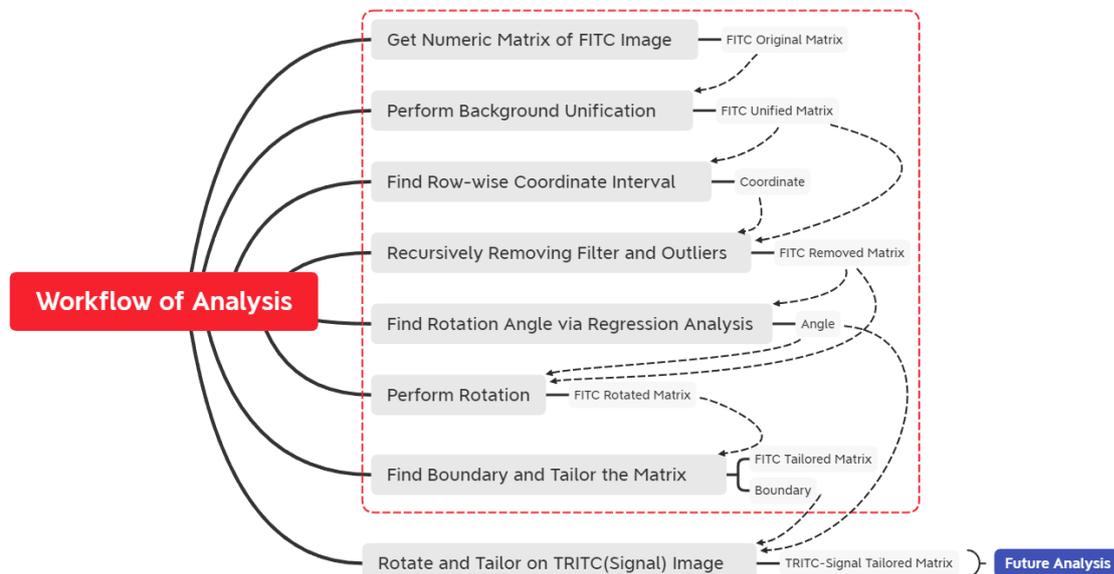

Fig. 3 Image processing workflow: The red dotted line represents analysis on FITC images, and at the last step results are applied to TRITC images for biofilm specific analysis

V. EVALUATION

To evaluate the approach of removing the filter substrata, background, and outliers, this approach was compared to a popular method in computer vision and pattern recognition, the Otsu [18] thresholding method. Otsu's method relies on the image matrix having a bimodal distribution, so that setting a threshold can separate foreground and background. In the proposed algorithm, the matrix does not need to be bimodally distributed. Furthermore, the Otsu method cannot fully remove the filter and outliers.

To evaluate the quality of the biofilm boundary site identification, the results were compared to the results from the Canny edge detection [19] method. ImageJ [20] was used to perform Canny edge detection on FITC images, in the red

channel. The result shows the discontinuous edges of biofilms and extra work is required to manually draw the edge line. Additionally, because of the existence of the filter and outliers, the program also produces edges on those structures, which makes the subsequent analysis more complex.

There are clear advantages to the method developed here. The Canny edge detection method requires trial and error and needs optimization for each image. The Canny method also searches through the entire image (including artifacts) for manual brightness adjustment, which can significantly affect results and add another manual process. Manually drawing lines is straightforward, but difficult to automate. The currently proposed method is easy and fast, but still provides user adjustable features. Fig. 4 shows the efficiency of this

algorithm in terms of removing the filter and outliers to focus on the biofilm specific region compared to Otsu's and Canny
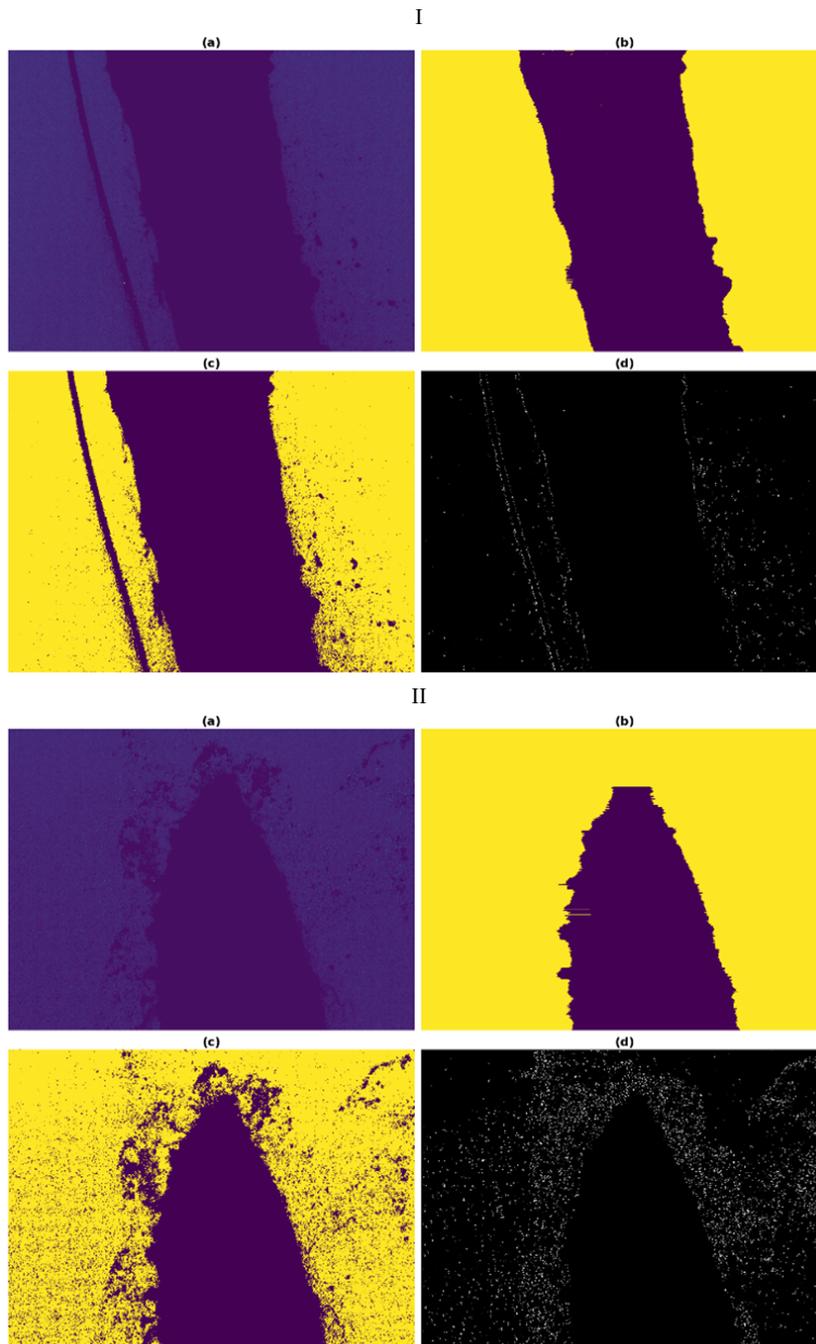
edge detection methods.

Fig. 4 I & II Two examples comparing the algorithm proposed in this work to other methods. For both I & II, (a) an original FITC image in channel 0 (Red), (b) filter and outlier removed using the method described in this work, and (c) background and foreground selection using the Otsu method. (d) Edge identification using the Canny edge detection algorithm

VI. RESULTS

Results from the currently proposed algorithm on two FISH images of biofilms, both stained simultaneously with SYTO 9 and the Cy3 nucleic acid probe, are shown in Fig. 5. One clear biofilm image (Fig. 5, panel I) and one inferior biofilm image (Fig. 5, panel II) were used to demonstrate the developed algorithm's performance. Beginning with the original image (Fig 5 (a)), background unification was applied, resulting in Fig

5 (b). Next, the filter, background, and outliers were removed (Fig. 5 (c)). The linear regression analysis was calculated for the points on the left boundary of the image (Fig. 5 (d)) and was used along with the y-axis to calculate the rotation angle and to rotate the image (Fig. 5 (e)). Finally, the located left and right boundaries of the biofilm result in the tailored image (Fig. 5 (f)). Results obtained were transformed onto TRITC-channel FISH images shown in Figs. 5 (g)-(i) for ribosomal abundance-specific analysis. To test the efficiency of the method, the approach was applied to several FISH images that were less clear. One such example with a "corn-shape" biofilm image containing extraneous outliers is shown in Fig. 5, panel II. Based on the results, the method described here performed exceedingly well in identifying biofilm regions in almost all inferior images.
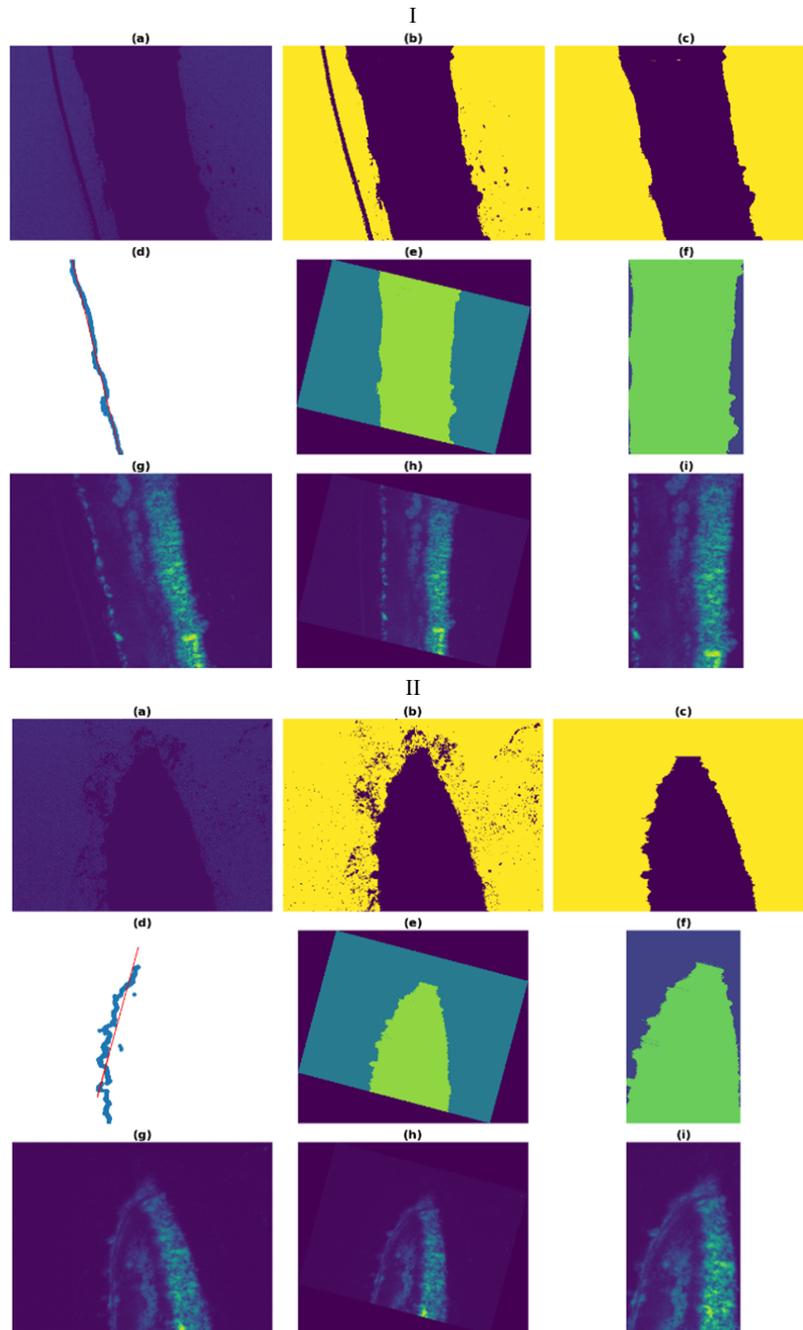

Fig. 5 I and II (a)-(f) for FITC images and (g)-(i) for TRITC images: (a) Original image in red channel, (b) background unification result, (c) after filter and outliers removed, (d) scatter plot with regression line on biofilm left boundary, (e) image after performing rotation, (f) final tailored image; (g) original image in red channel; (h) rotated image; (i) tailored image

VII. DISCUSSION AND CONCLUSION

The proposed algorithm provides a clean and efficient way to identify the biofilm boundaries from SYTO 9 stained FISH images of biofilms. Biofilm boundary detection using this method has generated better results and also provided ease of use compared to ImageJ. Given that this approach is modular, users have the ability to control parameters at different stages, such as setting the threshold for removing filters, background, and outliers, and for the regression line for calculating the rotation angle. Detecting biofilm regions accurately will enable effective quantification of the physiological heterogeneity of biofilms, which will be the main target for a future research study.

ACKNOWLEDGMENT

REFERENCES

[1] Flemming HC, Wingender J. 2010. The biofilm matrix. Nat Rev Microbiol 8:623-33.

[2] Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol 2:95-108.

[3] Costerton JW, Stewart PS, Greenberg EP. 1999. Bacterial biofilms: a common cause of persistent infections. Science 284:1318-1322.

[4] Lyczak JB, Cannon CL, Pier GB. 2002. Lung infections associated with cystic fibrosis. Clin Microbiol Rev 15:194-222.

[5] Malhotra S, Hayes D, Jr., Wozniak DJ. 2019. Cystic Fibrosis and *Pseudomonas aeruginosa*: the Host-Microbe Interface. Clin Microbiol Rev 32.

[6] Stewart PS, Franklin MJ. 2008. Physiological heterogeneity in biofilms. Nat Rev Microbiol 6:199-210.

[7] Williamson KS, Richards LA, Perez-Osorio AC, Pitts B, McInnerney K, Stewart PS, Franklin MJ. 2012. Heterogeneity in *Pseudomonas aeruginosa* biofilms includes expression of ribosome hibernation factors in the antibiotic-tolerant subpopulation and hypoxia-induced stress response in the metabolically active population. J Bacteriol 194:2062-73.

[8] Akiyama T, Williamson KS, Schaefer R, Pratt S, Chang CB, Franklin MJ. 2017. Resuscitation of *Pseudomonas aeruginosa* from dormancy requires hibernation promoting factor (PA4463) for ribosome preservation. Proceedings of the National Academy of Sciences 114:3204-3209.

[9] Franklin MJ, Chang C, Akiyama T, Bothner B. 2015. New Technologies for Studying Biofilms. Microbiol Spectr 3.

[10] Pérez-Osorio AC, Williamson KS, Franklin MJ. 2010. Heterogeneous rpoS and rhlR mRNA levels and 16S rRNA/rDNA (rRNA gene) ratios within *Pseudomonas aeruginosa* biofilms, sampled by laser capture microdissection. J Bacteriol 192:2991-3000.

[11] Deutscher MP. 2003. Degradation of stable RNA in bacteria. J Biol Chem 278:45041-4.

[12] Maki Y, Yoshida H, Wada A. 2000. Two proteins, YfiA and YhbH, associated with resting ribosomes in stationary phase *Escherichia coli*. Genes Cells 5:965-74.

[13] Hogardt M, Trebesius K, Geiger AM, Hornef M, Rosenecker J, Heesemann J. 2000. Specific and rapid detection by fluorescent in situ hybridization of bacteria in clinical samples obtained from cystic fibrosis patients. J Clin Microbiol 38:818-25.

[14] Lee N, Nielsen PH, Andreasen KH, Juretschko S, Nielsen JL, Schleifer KH, Wagner M. 1999. Combination of fluorescent in situ hybridization and microautoradiography-a new tool for structure-function analyses in microbial ecology. Appl Environ Microbiol 65:1289-97.

[15] Brileya KA, Camilleri LB, Fields MW. 2014. 3D-fluorescence in situ hybridization of intact, anaerobic biofilm. Methods Mol Biol 1151:189-97.

[16] Pretorius E, Page MJ, Engelbrecht L, Ellis GC, Kell DB. 2017. Substantial fibrin amyloidogenesis in type 2 diabetes assessed using amyloid-selective fluorescent stains. Cardiovasc Diabetol 16:141.

[17] Montgomery DC, Peck EA, Vining GG. 2012. Introduction to Linear Regression Analysis, 5th ed. Wiley, Hoboken.

[18] Otsu N. 1979. A Threshold Selection Method from Gray-Level Histograms, p 62-66, IEEE Transactions on Systems, Man, and Cybernetics, vol 9. IEEE.

[19] Canny J. 1986. A Computational Approach to Edge Detection, IEEE Transactions on Systems, Man, and Cybernetics, vol 9. IEEE.

[20] Rasband W. 2015. ImageJ, image processing and analysis in Java, p https://imagej.nih.gov/ij/.

[21] https://pypi.org/project/czifile/