

An algorithm selection methodology for automated focusing in optical microscopy

Marina Bonet Sanz  | Dr. Felipe Machado Sánchez  | Dr. Susana Borrromeo 

Rey Juan Carlos University, Móstoles, Madrid, Spain

Correspondence

Marina Bonet Sanz, Rey Juan Carlos University, Móstoles, Madrid 28933, Spain.
Email: marina.bsanz@urjc.es

Funding information

Spanish Department of Science, Innovation and Universities, Grant/Award Number: RTC-2017-6218-1

Review Editor: Alberto Diaspro

Abstract

Autofocus systems are essential in optical microscopy. These systems typically sweep the sample through the focal range and apply an algorithm to determine the contrast value of each image, where the highest value indicates the optimal focus position. As the optimal algorithm may vary according to the images' content, we evaluate the 15 most used algorithms in the field using 150 stacks of images from four different kinds of tissue. We use four measuring criteria and two types of analysis and propose a general methodology to apply to select the best fitting algorithm for any given application. In this paper, we present the results of this evaluation and a detailed discussion of different features: the threshold used for the algorithms, the criteria parameters, the analysis used, the bit depth of the images, their magnification, and the type of tissue, reaching the conclusion that some of these parameters are more relevant to the study than others, and the implementation of the proposed methodology can lead to a fast and reliable autofocus system capable of performing an analysis and selection of algorithms with no supervision required.

KEYWORDS

computer-aided detection and diagnosis, evaluation and performance, image acquisition, microscopy

1 | INTRODUCTION

Automation in the medicine field leads to faster, cheaper, and more accurate results, specifically by using digital imaging in the digital pathology area. Automation and digitalization of diagnostic procedures reduce the acquisition and processing times and improve accuracy, throughput, and reproducibility of the measurements (Saerens et al., 2019; Xu et al., 2017). In addition, it allows sending the acquired digital images to experts all over the world for consults and automatically storing patient records (Liao, 2018). As medical diagnosis is a cognitive process, its automation will as well reduce the physician's workload, setting a better environment to reduce clinical errors and leading to a better health care (Panicker et al., 2016).

Fast and reliable autofocus systems are crucial for microscopy automation, allowing real-time high-resolution image processing of all the possible fields of view in a sample (Bueno-Ibarra et al., 2005; Hosseini et al., 2020). All the subsequent analysis applied on the specimens depend on the quality of this focusing mechanism, as slight deviations from the optimal focus position would generate unreliable results (Hilsenstein, 2005).

There are two main categories for autofocus methods: active and passive (Castillo-Secilla et al., 2017; Israni et al., 2016; Kehtarnavaz & Oh, 2003). Active autofocus implies the emission of an ultrasound or electromagnetic wave, such as an infrared light beam, meant to fall upon the surface of the object to be focused. A sensor captures the reflection of this wave, and the distance from the object to the lens is calculated based on triangulation or the time needed for the signal to

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Microscopy Research and Technique* published by Wiley Periodicals LLC.

travel to the sensor. This method is effective when working in lowlight conditions but not when trying to focus elements through a glass, as it happens in microscopy, due to high reflectivity of this material to the emitted waves.

Passive autofocus methods are not based on emitting any wave but on the content of the received images, and there are two main techniques. The first one, based on sharpness analysis, is called phase-based. It uses a primary mirror to reflect some of the incoming light from the object to a phase detection sensor to evaluate if the image is in-phase, and therefore focused. However, it entails the addition of some optical elements to an already complex optical system.

Conversely, the second technique does not need any additional elements, as it is based on the contrast of the images. This is achieved by sweeping the sample through the focal range to get a stack of images and then measuring the contrast of every image in the stack by applying a contrast calculation algorithm. As a result, all these values define the contrast function of the stack, and the maximum value of this function indicates the location in the stack of the best focused image and thus the optimal focus point as shown in Figure 1.

Although many algorithms are available, not all of them are suited for a given application. In fact, the optimal algorithm for a certain use varies with the sample and its content (Shah et al., 2017; Yousefi et al., 2011). This is why it is expected that different tissues and magnifications will demand the use of different algorithms, as these parameters directly affect the images' content, along with the fact that conventional microscopy and fluorescence microscopy have different features and the best algorithms vary from one to the other.

For this reason, it is necessary to identify the optimal algorithm for each application. Previous work has only focused on solving this issue for the application at hand, and that is why we propose a wider approach: This paper is a review of the methods used in those references, with the aim to broaden the current knowledge by developing a general methodology applicable to any field.

The proposed methodology objectively quantifies the quality of an algorithm and evaluates its performance, comparing the results between different algorithms and selecting the optimal one for any application.

To develop an automated focusing methodology, we have taken a total of 150 stacks of images from four different types of tissue, applying two magnifications and using two bit-depth quality configurations.

Then, we have selected the 15 most used algorithms from the bibliography and developed an automated script in Python to evaluate

their performance over all the stacks. So, in the following sections, we will show the steps followed and propose a methodology suitable for any kind of images.

2 | MATERIALS AND METHODS

2.1 | Data sets and image acquisition

The Faculty of Health Sciences of the Rey Juan Carlos University provided us with four different mouse tissue samples: kidney, stomach, intestine, and adipose tissue, with a 5 μm width and hematoxylin-eosin staining and the means for a manual acquisition of the images with a Zeiss Axioskop 2 microscope equipped with the image analysis software package AxioVision 4.6 in bright-field modality.

We took images with a resolution of 1388 \times 1040 pixels from four different tissue samples. For each tissue sample, we used two magnifications, 5 \times and 10 \times , and each magnification was applied with two bit depths, 8 bits and 16 bits, resulting in a total of 16 categories. Each category contains 10 stacks, except for adipose tissue 10 \times 8 bits and 16 bits, which contain 5 stacks each.

This results in a total of 150 stacks, which were then analyzed by an expert technician in order to determine the optimal focus position for each stack. A sample of these images is shown in Figure 2.

These stacks are the best approximation to a sweep of all the possible fields of view in a sample, taking into account the need to use a finite number of images per stack, fluctuating this number between 10 and 20 images, and the limitations of a manual setting, which made it difficult to establish a fixed number of images per stack, as we were unable to determine a fixed position to begin or end the sweep, or a fixed interval between images.

2.2 | Contrast calculation algorithms

As it has been mentioned, many algorithms for contrast calculation are available. To select the most commonly used in microscopy, we chose five different reviews of autofocus functions from the literature. Those algorithms present in more than one of the references are shown in Table 1. As a result, we selected for this study the 15 most used algorithms.

The 15 algorithms are applied to all the images in a stack to generate its contrast functions, considering that some of them have a threshold value, which is indicated along their formulation.

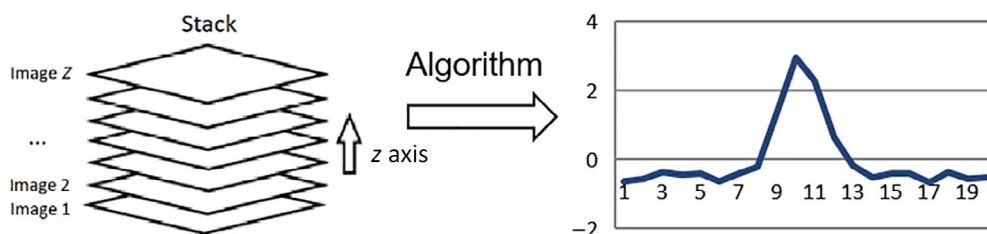


FIGURE 1 Algorithm application example

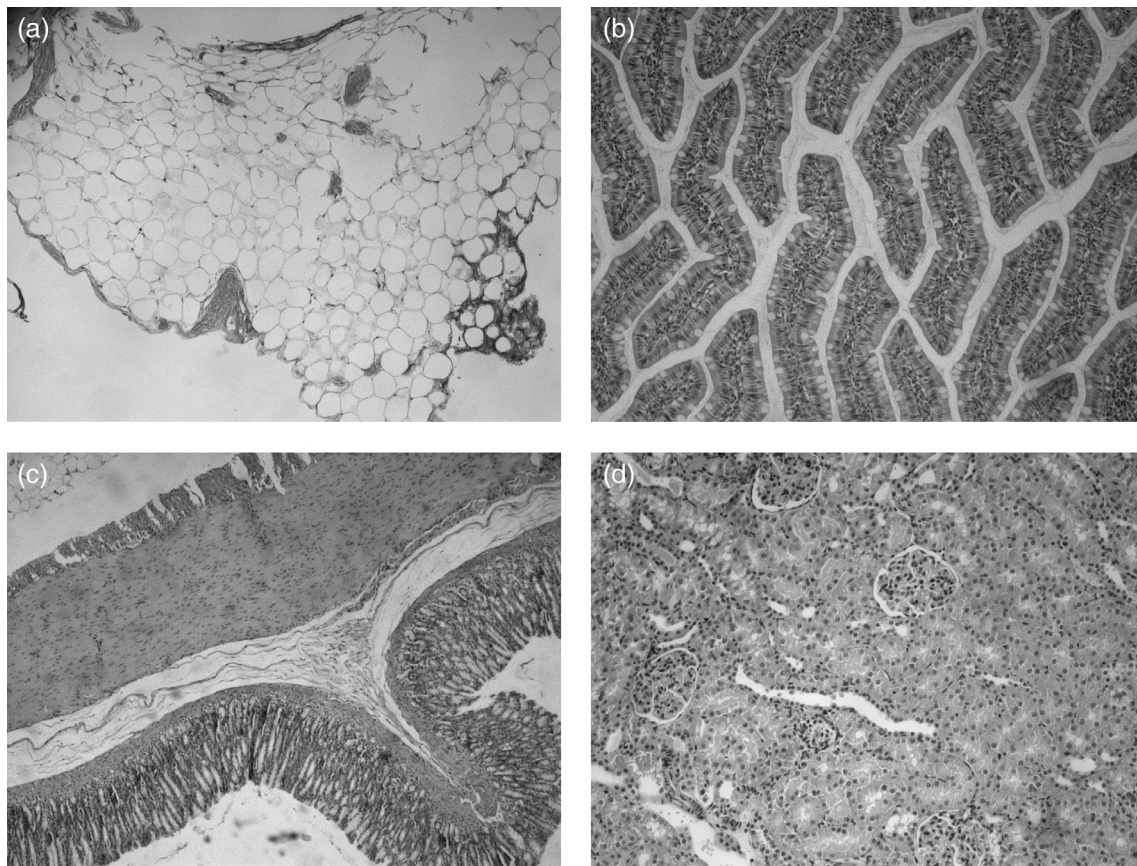


FIGURE 2 Tissue image samples. (a) Adipose 5× 16 bits. (b) Intestine 10× 16 bits. (c) Stomach 5× 8 bits. (d) Kidney 10× 8 bits

TABLE 1 Algorithm selection ranking

Algorithm	Costa Filho & Cost, 2012	Kimura et al., 2010	Santos et al., 1997	Mateos-Pérez et al., 2012	Redondo et al., 2012	Total
1 Vollath's F4	x	x	x	x	x	5
2 Vollath's F5	x	x	x	x	x	5
3 Variance	x	x	x	x	x	5
4 Variance of log histogram	x	x		x	x	4
5 Energy of image Laplacian	x	x		x	x	4
6 Tenengrad gradient	x		x	x	x	4
7 Normalized variance	x	x	x		x	4
8 Image power			x	x	x	3
9 Thresholded pixel count			x	x	x	3
10 Entropy	x	x	x			3
11 Brenner gradient	x	x	x			3
12 First-order Gaussian derivative	x			x	x	3
13 Squared gradient	x		x			2
14 Threshold absolute gradient	x		x			2
15 Absolute Tenengrad				x	x	2

The 15 selected algorithms are presented below with their formulation, where \mathcal{M} is the number of lines of the image, \mathcal{N} the number of pixels per line, $g(i, j)$ is the gray level of the pixel in the (i, j) position, \bar{g} is the mean value of all the pixels in the image, and θ is the threshold value.

2.2.1 | Based on correlation measurement

Vollath's F4

$$F = \sum_{M-1} \sum_N g(i+1, j) * g(i, j) - \sum_{M-2} \sum_N g(i+2, j) * g(i, j). \quad (1)$$

Vollath's F5

$$F = \sum_{M-1} \sum_N g(i+1, j) * g(i, j) - (M-1)N\bar{g}^2. \quad (2)$$

2.2.2 | Based on statistics

Variance

$$F = \frac{1}{MN} \sum_M \sum_N |g(i, j) - \bar{g}|. \quad (3)$$

Normalized variance

$$F = \frac{1}{MN\bar{g}} \sum_M \sum_N |g(i, j) - \bar{g}|. \quad (4)$$

2.2.3 | Based on histogram

Entropy

$$F = - \sum_i p_i \cdot \log_2 p_i, \quad (5)$$

where p_i is the relative frequency of grey level.

Variance of log histogram

$$F = \sum_i (I - E_{\log\{I\}})^2 \cdot \log p_i, \quad (6)$$

$$E_{\log\{I\}} = \sum_i I \cdot \log p_i.$$

2.2.4 | Based on image differentiation

Energy of image Laplacian

$$F = \sum_{M-1} \sum_{N-1} (g(i, j+1) + g(i, j-1) + g(i+1, j) + g(i-1, j) - 4g(i, j))^2. \quad (7)$$

Tenengrad gradient

$$F = \sum_M \sum_N T[g(i, j)],$$

$$T[g(i, j)] = G_x^2(i, j) + G_y^2(i, j), \quad (8)$$

where $G_x(i, j), G_y(i, j)$ is the convolution with Sobel operator.

Brenner gradient

$$F = \sum_M \sum_{N-2} |g(i, j+2) - g(i, j)|^2,$$

$$\text{while } |g(i, j+1) - g(i, j)| \geq \theta. \quad (9)$$

First-order Gaussian derivative

$$F = \frac{1}{MN} \sum_M \sum_N (g(i, j) * G_x(x, y, \sigma))^2 + (g(i, j) * G_y(x, y, \sigma))^2, \quad (10)$$

where $G_x(x, y, \sigma), G_y(x, y, \sigma)$ is the first Gaussian derivatives and $\sigma = \frac{d}{2}\sqrt{3}$ is the standard deviation.

Squared gradient

$$F = \sum_M \sum_{N-1} |g(i, j+1) - g(i, j)|^2,$$

$$\text{while } |g(i, j+1) - g(i, j)| \geq \theta. \quad (11)$$

Threshold absolute gradient

$$F = \sum_M \sum_{N-1} |g(i, j+1) - g(i, j)|,$$

$$\text{while } |g(i, j+1) - g(i, j)| \geq \theta. \quad (12)$$

Absolute Tenengrad

$$F = \sum_M \sum_N |G_x(i, j)| + |G_y(i, j)|, \quad (13)$$

where $G_x(i, j), G_y(i, j)$ is the convolution with Sobel operator.

2.2.5 | Based on depth of peaks and valleys

Image power

$$F = \sum_M \sum_N g(i, j)^2,$$

$$\text{while } g(i, j) \geq \theta. \quad (14)$$

Thresholded pixel count

$$F = \sum_M \sum_N s[g(i,j), \theta],$$

$$s[x, \theta] = \begin{cases} 0, & x \geq \theta \\ 1, & x < \theta \end{cases} \quad (15)$$

2.3 | Evaluation criteria

To quantify the performance of each algorithm in a given application, we need some criteria to objectively evaluate the quality of the resulting contrast functions. Consequently, we chose the four criteria usually used in similar studies.

These criteria have an optimal value equal to zero because they are defined by mimicking the features of an ideal function, which has an unequivocal maximum (Osibote et al., 2010; Qiu et al., 2013). Figure 3 shows an example of each criteria, which are defined as follows.

1. **Accuracy:** Number of images between the maximum of the contrast function and the focus position determined by an expert. A zero value indicates the function prediction of the focus point agrees with the one determined by an expert.
2. **Range:** Difference between the number of images in the stack, and the number of images around the global maximum in which the contrast function is monotonically decreasing. When this criterion equals zero, it means that the range contains all the images in the stack.
3. **False maxima:** Number of local maxima besides the global maximum. If this value equals zero, there is only one maximum in the series, thus making it easier to find.
4. **Full width at 50% maximum:** Width of the global maximum at half of its height. This criterion speaks to the sharpness of the function and the clarity with which it shows the maximum value. Therefore, the smallest the value, the easiest to determine the maximum point.

These criteria quantify the quality of the contrast functions generated by the algorithms. The scores computed in the four criteria for all the contrast functions in the stack define a criteria table. That is, for each stack, we obtain a criteria table with as many rows as resulting functions and as many columns as criteria.

2.4 | Proposed analysis

Once we have calculated the four criteria and obtained the criteria tables, we need an objective method to grade each of the algorithms.

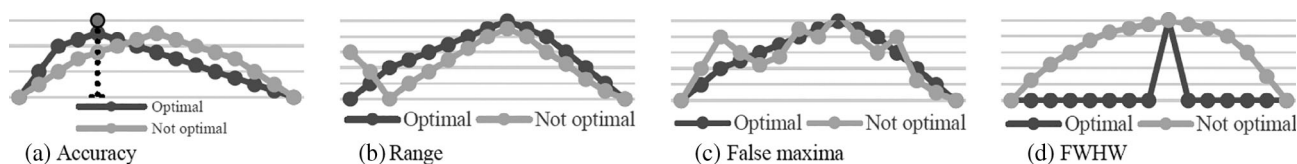


FIGURE 3 Criterion definition. (a) Accuracy. (b) Range. (c) False maxima. (d) Full width at 50% maximum

For this purpose, we propose two different analyses to compare the algorithms: a semiquantitative and a quantitative analysis.

2.4.1 | Semiquantitative

This first analysis is a simpler one, where for a given stack and criterion, the result of an algorithm is compared with the rest and ranked accordingly. The ranking method would be as follows: the algorithm with the best result is given a 1, the second a 2, and so on. If two or more algorithms were to share a score, they all will be ranked the same.

The total score for each criterion is the addition of the scores obtained in all the stacks analyzed, as shown in Figure 4, and the global results for this analysis will be the addition of the total scores of all the criteria.

2.4.2 | Quantitative

As a more extensive procedure, the quantitative analysis compares the results of the algorithms in each criterion not between them but to a theoretical ideal function defined as having a value of zero in all the criteria. Therefore, we normalize the results of all algorithms within a criterion, as well as the value of the ideal function, and compute the distance of each algorithm to the ideal value.

Then, for each criterion along all the stacks, the Euclidean distance is calculated for the 15 algorithms, obtaining the total score for each criterion, as shown in Figure 5. The global results of this analysis are obtained by adding the total scores of each criteria.

2.5 | Methodology

The general methodology we propose summarizes the previous sections, as each of them is one of the steps to follow. This is shown in Figure 6, where the sections of the paper on where each step is explained are included in the labels.

From the tissue samples, we perform the image acquisition, obtaining the set of stacks to analyze. To these stacks, we apply the selected algorithms, resulting in a series of contrast functions, one per each algorithm and stack. These functions are quantified by the evaluation criteria, giving as an output the criteria tables, which are the data used to feed the analyses, finally producing a ranking of the algorithms.

FIGURE 4 Total score for semiquantitative analysis

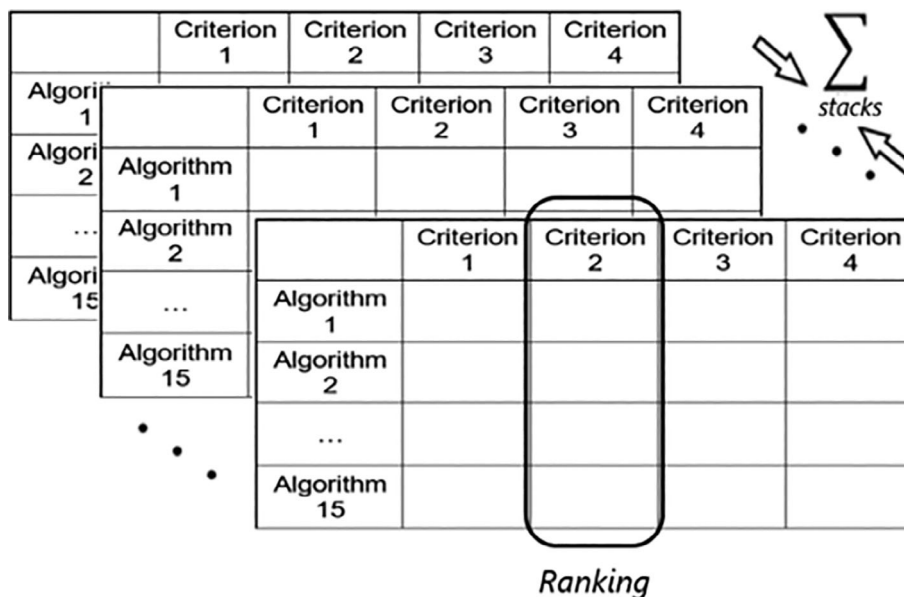
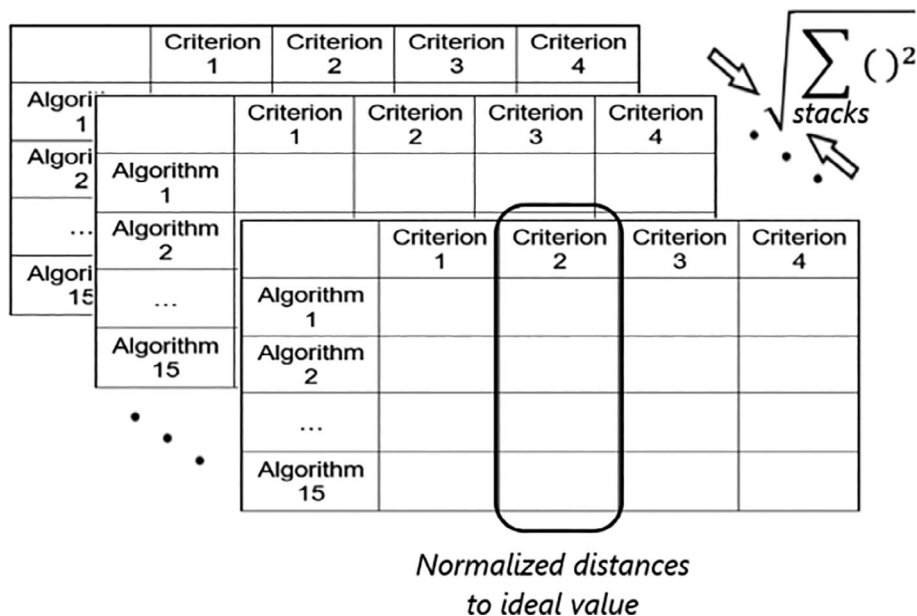


FIGURE 5 Total score for quantitative analysis



2.6 | Implementation

With a view to implement the proposed methodology with an automatic tool, we followed three steps: development, preparation, and execution.

2.6.1 | Development phase

To develop an automatic tool to perform all the steps of the methodology, we have created a Python script that performs the calculation of the contrast functions, followed by the scoring of each of these functions by the evaluation criteria, and then grades this data by performing both analyses.

We selected Python not only because of its popularity and ease of use, as it is a high-level programming language with an enormous background of available libraries, but also because is a language optimized for matrix calculations.

Only a little information is needed for the developed code to work, and it is fed in an array shape with as many elements as the number of stacks we have or number of algorithms to evaluate:

1. The optimal focusing position of each stack, that we need to determine manually.
2. The number of images of each stack.
3. The value of the desired threshold for each algorithm, in case they had one.

Also, we have to determine the path to the folder in which we would like to save the results, as for the one containing the images to analyze.

2.6.2 | Preparation phase

When evaluating new images or algorithms, the following steps need to be followed to prepare the data.

1. Convert the acquired images from RGB into gray scale, significantly reducing the amount of data to process, as the number of pixel matrixes decreases from three to one.
2. Manually examine the images stack per stack to determine the number of images per stack, and the correct focus position, which is required to assess the accuracy.

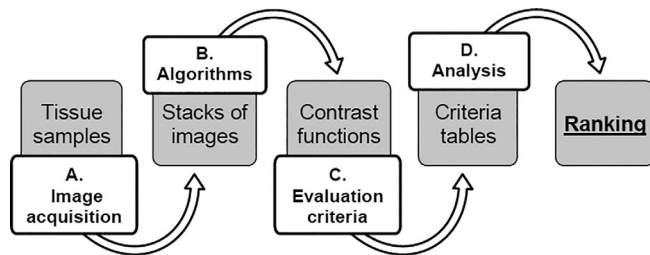


FIGURE 6 General methodology

3. Translate the mathematical formulation of the selected algorithms to Python and include them in the script.

2.6.3 | Execution phase

Once we provide the script with the data, it will run the algorithms and perform the analyses, giving as a global output a spreadsheet document for each tissue and magnification studied.

This document contains a sheet for each stack, with the normalized contrast functions of each algorithm, the criteria table, the scoring of the semiquantitative analysis, and the calculations of the Euclidean distances needed for the quantitative analysis.

There is also another sheet per each bit depth, in which the data of each stack needed for both semiquantitative and quantitative analyses will be featured, together with the total and global scores, and the ranking.

3 | RESULTS

As mentioned in Section 2, some algorithms have a threshold. We visually assessed the results of different threshold values, finding that a threshold of 20% of the maximum gray level in each bit-depth configuration had a better performance.

With this threshold, we applied the algorithms, and for each of the stacks, we obtained 15 contrast functions, having some of them better outputs than the rest. As an example, Figure 7 shows the best

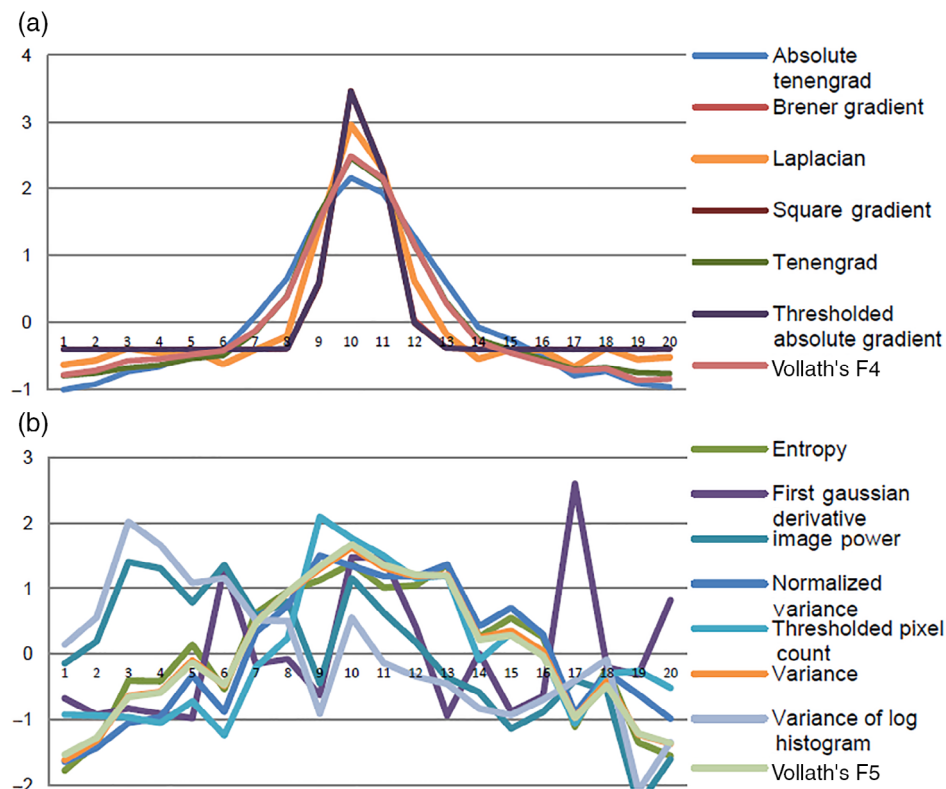


FIGURE 7 (a) Best and (b) worst contrast functions in adipose tissue, 5× magnification

TABLE 2 Criteria table

Algorithm	Criteria 1	Criteria 2	Criteria 3	Criteria 4
Absolute Tenengrad	0	3	1	5.086
Brener gradient	0	0	0	2.014
Entropy	0	14	5	9.862
First Gaussian derivative	7	15	5	0.969
Image power	7	15	4	4.575
Laplacian	0	11	4	2.829
Normalized variance	1	13	4	9.094
Square gradient	0	0	0	1.991
Tenengrad	0	3	1	3.991
Thresholded absolute gradient	0	0	0	2.010
Thresholded pixel count	1	13	4	5.379
Variance	0	13	4	9.096
Variance of log histogram	7	15	3	2.639
Vollath's F4	0	3	1	3.942
Vollath's F5	0	11	3	8.720

(a) and worst (b) contrast functions for a stack of adipose tissue in $5\times$ magnification, where (a) algorithms show a clear maximum whereas (b) fail to meet the evaluation criteria shown in Figure 3.

For each of these curves, we calculate the value for the four criteria; therefore, 150 criteria tables as the one shown in Table 2 are obtained from the evaluated data. These tables are then processed by the two proposed analyses.

For the four different categories in this study—tissue, magnification, bit depth, and analysis—the results of this processing, that is, the global scores and ranking of applying the 15 algorithms to the 150 stacks, are summarized in Tables 3, 4, 5, and 6.

In addition to those tables, Figure 8 shows a summary of the ranking results. In the figure, the results of the 15 algorithms for the four different tissues are highlighted in different colors, where the higher the bar, the better the ranking position; therefore, the best algorithm would be squared gradient. This figure also shows that algorithms perform similarly along different tissues, especially for the top-scoring algorithms, as the different color sections have similar weight. As an example, squared gradient has equal slots for all tissues, whereas entropy clearly fails in adipose tissue. To assess the quality of the results obtained from both analyses and the variations in the different categories, these results are visually compared with a qualitative evaluation of all the contrast functions. To do so, the normalized results for the different curves are weighed against each other and then compared with the ranking generated by the semiquantitative and quantitative analyses.

4 | DISCUSSION

Our goal was to propose a general methodology to automate the algorithm selection method for the focusing systems used in optical microscopy. With this purpose, we evaluated several features and

observed that not all of them are equally relevant, as we comment on the following paragraphs.

4.1 | Threshold

As some of the algorithms have a threshold value, for this study, we have only tested three possible thresholds: 10%, 20%, and 50% of the maximum gray level possible in each bit depth configuration, and the presented results use the 20% value for all the algorithms, as we observed it to be the best featuring one.

If we wanted to add an extra level of accuracy to the analysis, the algorithms with threshold values could be examined to a greater extend by taking an analysis of its own in the first place, where we would test the spectrum of possible thresholds and determine the best performing threshold value. Hence, only the best version of these algorithms would be tested with the algorithms without threshold, resulting in more balanced outcome.

4.2 | Criteria

If we closely examine the 150-criteria tables obtained after the application of the four criteria, as the one showed in Table 2, to all the contrast functions calculated, as the featured in Figure 7, and use them as an example, we will see that the scoring of the algorithms correlates with the qualitative features of the curves: the curves that meet the original criteria of leading to a fast and clear identification of the correct maximum of the function tend to have the lowest scores, while the ones which clearly are not a good fit tend to have higher scores.

However, this tendency is not always met for all the criteria at once, meaning no single criterion can be excluded from the study as none of them are giving redundant information. And, in the cases

TABLE 3 Global results for adipose tissue

	5 ×						10 ×									
	8 bits			16 bits			8 bits			16 bits						
	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative					
Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking			
Absolute Tenengrad	181	5	7.80	5	181	5	7.80	5	104	6	4.80	6	105	6	4.78	6
Brener gradient	88	2	2.48	2	84	2	2.47	2	27	3	1.28	3	26	3	0.97	3
Entropy	399	13	23.70	13	398	13	23.66	13	180	12	15.90	12	183	12	15.88	12
First Gaussian derivative	449	15	30.33	14	448	15	30.46	14	142	9	6.45	9	145	9	6.45	9
Image power	378	12	21.88	12	378	12	21.82	12	96	4	4.84	7	100	4	4.83	7
Laplacian	248	7	12.86	7	245	7	12.82	7	252	15	25.05	15	252	15	24.98	15
Normalized variance	350	11	18.25	9	350	11	18.24	9	225	13	21.71	13	225	13	21.69	13
Square gradient	76	1	2.46	1	77	1	2.45	1	26	2	1.27	1	24	1	0.97	1
Tenengrad	154	4	6.44	4	154	4	6.44	4	102	5	4.77	5	103	5	4.76	5
Thresholded absolute gradient	89	3	2.49	3	91	3	2.49	3	25	1	1.27	2	25	2	0.97	2
Thresholded pixel count	329	10	16.05	8	331	10	16.36	8	236	14	22.71	14	236	14	22.65	14
Variance	328	9	20.20	11	325	9	20.17	11	173	11	11.77	10	173	11	11.77	10
Variance of log histogram	440	14	32.68	15	441	14	32.69	15	125	8	6.30	8	121	8	6.23	8
Vollath's F4	216	6	9.07	6	216	6	9.06	6	105	7	4.47	4	106	7	4.46	4
Vollath's F5	313	8	19.37	10	312	8	19.33	10	172	10	11.86	11	172	10	11.86	11

TABLE 4 Global results for stomach tissue

	5 ×						10 ×					
	8 bits			16 bits			8 bits			16 bits		
	Semi-quantitative	Quantitative		Semi-quantitative	Quantitative		Semi-quantitative	Quantitative		Semi-quantitative	Quantitative	
Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	
Absolute Tenengrad	175	5	11.43	6	175	5	11.41	6	209	6	8.83	6
Brener gradient	78	2	4.82	2	78	2	4.78	2	57	2	1.96	2
Entropy	302	8	18.33	8	302	8	18.30	8	279	9	14.18	9
First Gaussian derivative	421	13	31.93	15	422	14	31.93	15	272	8	14.25	10
Image power	384	12	28.21	14	383	12	28.02	14	250	7	12.85	7
Laplacian	234	7	14.38	7	233	7	14.37	7	448	15	35.85	15
Normalized variance	425	15	26.57	12	425	15	26.55	12	444	14	34.51	14
Square gradient	70	1	4.80	1	71	1	4.76	1	50	1	1.96	1
Tenengrad	145	4	10.47	4	145	4	10.45	4	170	4	7.90	5
Thresholded absolute gradient	86	3	4.86	3	88	3	4.82	3	64	3	1.97	3
Thresholded pixel count	336	11	19.07	9	337	11	19.06	9	377	13	24.14	13
Variance	311	10	19.42	10	311	10	19.40	10	297	11	14.31	11
Variance of log histogram	422	14	27.84	13	419	13	27.91	13	344	12	18.14	12
Vollath's F4	175	5	11.41	5	175	5	11.40	5	182	5	7.86	4
Vollath's F5	309	9	19.47	11	309	9	19.44	11	290	10	13.95	8

TABLE 5 Global results for intestine tissue

	5 ×						10 ×									
	8 bits			16 bits			8 bits			16 bits						
	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative					
Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking			
Absolute Tenengrad	237	7	19.76	8	236	7	19.70	8	198	6	13.85	6	198	6	13.85	6
Brener gradient	64	2	3.27	1	61	2	3.20	1	58	1	2.41	2	60	2	2.41	2
Entropy	301	8	20.55	11	297	8	20.28	10	248	8	18.11	8	248	8	18.10	8
First Gaussian derivative	392	14	28.74	14	415	14	30.82	14	429	14	34.43	15	430	15	34.43	15
Image power	450	15	34.20	15	446	15	33.16	15	312	11	23.19	12	312	11	23.19	12
Laplacian	213	5	14.46	4	212	5	14.38	4	249	9	19.93	10	249	9	19.93	10
Normalized variance	348	11	20.20	10	347	12	19.99	9	334	12	22.70	11	333	12	22.70	11
Square gradient	61	1	3.28	2	60	1	3.21	2	59	2	2.40	1	56	1	2.41	1
Tenengrad	179	4	15.88	5	178	4	15.84	5	179	4	13.21	5	179	4	13.21	5
Thresholded absolute gradient	68	3	3.30	3	68	3	3.23	3	66	3	2.41	3	68	3	2.41	3
Thresholded pixel count	322	9	16.92	7	319	9	16.64	7	429	14	29.21	14	429	14	29.23	14
Variance	330	10	20.20	9	329	10	20.34	11	234	7	17.53	7	233	7	17.53	7
Variance of log histogram	375	13	26.16	13	374	13	26.05	13	383	13	26.04	13	383	13	25.98	13
Vollath's F4	216	6	16.43	6	214	6	16.36	6	188	5	12.96	4	188	5	12.96	4
Vollath's F5	350	12	21.03	12	346	11	20.84	12	278	10	18.72	9	278	10	18.71	9

TABLE 6 Global results for kidney tissue

	5 ×						10 ×									
	8 bits			16 bits			8 bits			16 bits						
	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative	Quantitative	Semi-quantitative	Quantitative					
Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking	Global score	Ranking					
Absolute Tenengrad	126	6	5.20	7	126	6	4.32	6	119	6	4.33	6	120	6	4.33	6
Brener gradient	77	2	3.34	2	84	2	2.99	2	59	3	1.50	3	58	3	1.50	3
Entropy	309	11	12.86	12	307	11	12.77	12	241	11	7.71	10	241	11	7.71	10
First Gaussian derivative	508	13	35.05	13	508	13	34.90	13	323	12	17.95	13	309	12	15.19	12
Image power	508	13	35.05	13	508	13	34.90	13	188	8	11.81	11	188	8	11.82	11
Laplacian	137	7	4.61	4	135	7	4.42	7	181	7	6.43	7	181	7	6.43	7
Normalized variance	323	12	12.56	11	322	12	12.44	11	518	15	38.86	15	518	15	38.87	15
Square gradient	66	1	3.31	1	72	1	2.95	1	42	1	1.49	1	42	1	1.49	1
Tenengrad	110	5	4.64	6	110	5	3.87	5	88	4	3.53	4	88	4	3.53	4
Thresholded absolute gradient	85	3	3.36	3	90	3	3.00	3	52	2	1.50	2	53	2	1.49	2
Thresholded pixel count	230	10	11.57	10	237	10	11.36	10	475	14	34.66	14	475	14	34.65	14
Variance	182	8	6.37	8	181	8	6.13	8	229	10	7.53	9	229	10	7.53	9
Variance of log histogram	508	13	35.05	13	508	13	34.90	13	442	13	16.96	12	441	13	16.97	13
Vollath's F4	108	4	4.64	5	108	4	3.86	4	101	5	3.63	5	101	5	3.63	5
Vollath's F5	214	9	6.77	9	213	9	6.61	9	194	9	7.38	8	194	9	7.38	8

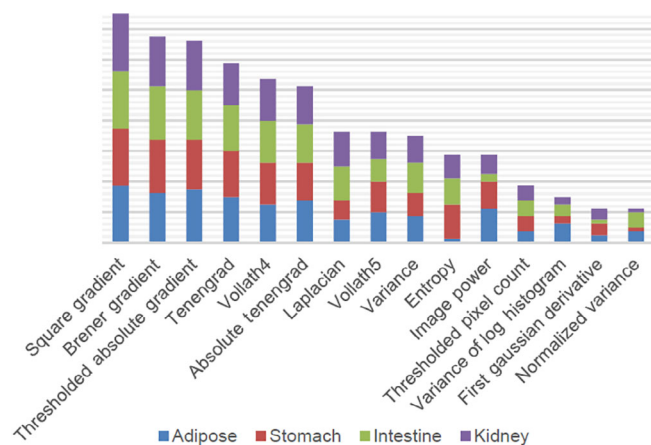


FIGURE 8 Ranking of algorithms summary

where the first three criteria were to be equal, the fourth one would be the tiebreaker, as for (9), (11), and (12). Another consideration is that we deliberately ruled out one criterion also commonly used in other studies: the execution time. We did not evaluate this criterion because our goal was to develop a general methodology to select the best fitting algorithm, prioritizing performance over execution time. Yet, in further studies, it might be an interesting criterion to consider.

4.3 | Analysis

When comparing the results of both analyses with a qualitative evaluation of all the curves obtained, we can appreciate that both analyses are effective and have similar results.

The slight deviations between both analyses can be explained by the fact that for the semiquantitative analysis, it is allowed for two algorithms to share the same score, although this would make no difference in most cases.

Nevertheless, in most of the studied cases, the kind of analysis used was found to not be relevant, as in most scenarios, the results were the same when implementing one or another.

4.4 | Bit depth

Although some variations can be appreciated in the overall ranking for all the categories when switching between 8- or 16-bit depth, those deviations differ only in one ranking place, going up or down one position.

In the results presented, there is only one case not adjusting to this rule, even though it is not relevant enough to indicate a better performance when using 16-bit depth images, turning a fourth position into a seventh.

Consequently, none of the variations observed are found significant enough to justify the use of 16-bit images, that is, double the size images, when the results obtained with 8-bit depth images are almost identical.

4.5 | Magnification

As expected, there are some differences between the 5× and 10× magnifications since this setting directly affects the content of the images and the size of the cells.

However, those differences are not as wide as predicted, switching only a few places, in most cases, the algorithm's ranking and maintaining almost the same algorithms as the most effective ones at the top positions.

But as opposed to the bit depth, this factor is found to be significant enough on account of the differences across the different tissues: In stomach tissue, the ranking is practically identical for both magnifications, but radical changes appear in other tissues, as for (14) in adipose tissue, or (7) in intestine tissue.

Therefore, as seen in the results, there is no assurance in the effectiveness of an algorithm given a specific magnification, even if the performance in a different magnification is known.

4.6 | Tissue

This category has also reported less variations than expected since the cell size, shape, and distribution depend on the tissue, thus directly affecting the content and density of information in the image.

Being that the case, there are three factors to take into account that might have had a part in the results obtained: first, for each of the studied tissues, the different stacks were taken from both the perimeter and the inner part of the sample, balancing the density of information among the stacks.

Secondly, all the images studied were taken in a bright-field microscopy setting, where there are some algorithms with better performance, as there are for fluorescence microscopy (Osibote et al., 2010; Shah et al., 2017). Therefore, it is likely that the same algorithms stand out in all tissues, being always in the top of the ranking.

Finally, a routine dye for microscopy was used, meaning no special areas or cells were highlighted, as in a pathology exam of a tissue would be done, making all the cells present in the sample visible and increasing the level of content in the images.

However, the variations in the ranking between the different tissues cannot be dismissed lightly, as there are many tissues to be studied, and for the purpose of this work, only four have been considered. In any case, these results can be used to distinguish the best algorithms to consider against those that could be directly discarded for bright-field microscopy.

5 | CONCLUSION

Fast and reliable autofocus systems are crucial for microscopy automation. Our goal was to propose a general algorithm selection methodology for automated focusing. In this paper, we have reviewed the methods used in similar studies, developed, and implemented in

Python a general methodology applicable to any field. The proposed methodology allows the user to objectively compare and grade the performance of a set of algorithms through different stacks.

This work has been validated with 150 stacks of images from four different types of tissue, applying two magnifications, and using two bit-depth quality configurations. Thus, it has been shown there are factors more relevant than others, being the type of tissue and the magnification more important than the bit depth or the analysis used, which have been found to give almost the same results whether selecting one feature or the other.

The best performing algorithms can be identified in Figure 8, and a more detailed scoring can be found in Tables 3, 4, 5, and 6. This scoring will show no significant differences in the results between bit depths but slight deviations for different magnifications or tissues. In this way, whenever the type of tissue or the magnification change, a new analysis must be made to select the best algorithm.

However, this process can be now made with a reduced list, having to test only the half of the original 15 algorithms, which are the ones proved to be more effective in bright-field microscopy—that is, (13), (11), (14), (6), (1), and (15). Obviously, this shortening can only be made when working with this illumination technique, and if changed, all 15 algorithms must be tested again.

Additionally, the top performing algorithms have shown to be quite accurate at defining the focus position, that is, the point with a higher contrast value. Therefore, this feature can be used to create a self-contained analysis method, which could independently determine the optimal focus point of a given stack by selecting the most repeated maximum point. And this result could be used as the reference to assess the first criterion: the accuracy.

Thus, if the positioning system of the samples were to be automated, and the number of images taken per stack was always the same, only the threshold value would be needed as an input. This parameter could be fixed, as we used a 20% value for all algorithms, or even included as a variable to study as previously mentioned in the discussion.

Consequently, with this methodology and our implementation, we could provide the system with capability to analyze, select the algorithms, and focus with no supervision required.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Department of Science, Innovation and Universities grant RTC-2017-6218-1 (<http://www.ciencia.gob.es/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DATA AVAILABILITY STATEMENT

This is an open-source project. The data that support the findings of this study are openly available at <https://github.com/aquimb/Auto-focus-URJC>.

ORCID

Marina Bonet Sanz  <https://orcid.org/0000-0001-5608-044X>

Felipe Machado Sánchez  <https://orcid.org/0000-0001-7674-0943>

Susana Borromeo  <https://orcid.org/0000-0002-2353-2902>

REFERENCES

- Bueno-Ibarra, M. A., A'lvarez-Borrego, J., Acho, L., & Cha'varez-Sa'anchez, M. A. C. A. (2005). Fast autofocus algorithm for automated microscopes. *Optical Engineering*, 44(6), 063601. <https://doi.org/10.1117/1.1925119>
- Castillo-Secilla, J. M., Saval-Calvo, M., Medina-Valdès, L., Cuenca-Asensi, S., Martínez-Álvarez, A., Sánchez, C., & Cristóbal, G. (2017). Autofocus method for automated microscopy using embedded GPUs. *Biomedical Optics Express*, 8(3), 1731. <https://doi.org/10.1364/BOE.8.001731>
- Costa Filho, C. F. F., & Cost, M. G. F. (2012). Sputum smear microscopy for tuberculosis: Evaluation of autofocus functions and automatic identification of tuberculosis mycobacterium. In P.-J. Cardona (Ed.), *Understanding tuberculosis - global experiences and innovative approaches to the diagnosis*. Janeza Trdine 9, 51000 Rjeka, Croatia: InTech. ISBN 978-953-307-938-7
- V. Hilsenstein, "Robust autofocusing for automated microscopy imaging of fluorescently labelled bacteria [Conference presentation]." Digital image computing: Techniques and applications (DICTA'05), Queensland, Australia, 2005, 15–15, doi: <https://doi.org/10.1109/DICTA.2005.72>.
- Hosseini, M. S., Brawley-Hayes, J. A. Z., Zhang, Y., Chan, L., Plataniotis, K. N., & Damaskinos, S. (2020). Focus quality assessment of high-throughput whole slide imaging in digital pathology. *IEEE Transactions on Medical Imaging*, 39(1), 62–74. <https://doi.org/10.1109/TMI.2019.2919722>
- Israni, D., Patel, S., & Shah, A. (2016). Comparison of different techniques of camera autofocusing. In S. C. Satapathy & S. Das (Eds.), *Proceedings of first international conference on information and communication Technology for Intelligent Systems: Volume 1* (Vol. 50, pp. 125–135). Cham: Springer International Publishing.
- Kehtarnavaz, N., & Oh, H.-J. (2003). Development and real-time implementation of a rule-based auto-focus algorithm. *Real-Time Imaging*, 9(3), 197–203. [https://doi.org/10.1016/S1077-2014\(03\)00037-8](https://doi.org/10.1016/S1077-2014(03)00037-8)
- A. Kimura, M. G. F. Costa, C. F. F. C. Filho, L. B. M. Fujimoto, and J. Salem, "Evaluation of autofocus functions of conventional sputum smear microscopy for tuberculosis [Conference presentation]," 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, 2010, 3041–3044, doi: <https://doi.org/10.1109/IEMBS.2010.5626143>.
- Liao, J. (2018). *Imaging innovations for whole-slide and hyperspectral microscopy* (p. 123). Storrs, CT 06269, United States: University of Connecticut. <https://opencommons.uconn.edu/dissertations/2017/>
- Mateos-Pérez, J. M., Redondo, R., Nava, R., Valdiviezo, J. C., Cristóbal, G., Escalante-Ramírez, B., ... Desco, M. (2012). Comparative evaluation of autofocus algorithms for a real-time system for automatic detection of mycobacterium tuberculosis. *Cytometry Part A*, 81A(3), 213–221. <https://doi.org/10.1002/cyto.a.22020>
- Osibote, O. A., Dendere, R., Krishnan, S., & Douglas, T. S. (2010). Automated focusing in bright-field microscopy for tuberculosis detection: Bright-field autofocusing for tuberculosis detection. *Journal of Microscopy*, 240(2), 155–163. <https://doi.org/10.1111/j.1365-2818.2010.03389.x>
- Panicker, R. O., Soman, B., Saini, G., & Rajan, J. (2016). A review of automatic methods based on image processing techniques for tuberculosis detection from microscopic sputum smear images. *Journal of Medical Systems*, 40(1), 17. <https://doi.org/10.1007/s10916-015-0388-y>
- Qiu, Y., Chen, X., Li, Y., Chen, W. R., Zheng, B., Li, S., & Liu, H. (2013). Evaluations of auto-focusing methods under a microscopic imaging modality for metaphase chromosome image analysis. *Analytical Cellular Pathology*, 36(1–2), 37–44. <https://doi.org/10.1155/2013/412920>
- Redondo, R., Bueno, G., Valdiviezo, J. C., Nava, R., Cristóbal, G., Déniz, O., ... Escalante-Ramírez, B. (2012). Autofocus evaluation for brightfield microscopy pathology. *Journal of Biomedical Optics*, 17(3), 036008. <https://doi.org/10.1117/1.JBO.17.3.036008>

- Saerens, G., Lang, L., Renaut, C., Timpu, F., Vogler-Neuling, V., Durand, C., ... Timofeeva, M. (2019). Image-based autofocusing system for nonlinear optical microscopy with broad spectral tuning. *Optics Express*, 27(14), 19915. <https://doi.org/10.1364/OE.27.019915>
- Santos, A., Ortiz De Solórzano, C., Vaquero, J. J., Peña, J. M., Malpica, N., & Del Pozo, F. (1997). Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of Microscopy*, 188(3), 264–272. <https://doi.org/10.1046/j.1365-2818.1997.2630819.x>
- Shah, M. I., Mishra, S., & Rout, C. (2017). Establishment of hybridized focus measure functions as a universal method for autofocusing. *Journal of Biomedical Optics*, 22(12), 126004. <https://doi.org/10.1117/1.JBO.22.12.126004>
- Xu, J., Tian, X., Meng, X., Kong, Y., Gao, S., Cui, H. Y., ... Wang, S. (2017). Wavefront-sensing-based autofocusing in microscopy. *Journal of Biomedical Optics*, 22(08), 1. <https://doi.org/10.1117/1.JBO.22.8.086012>
- Yousefi, S., Rahman, M., & Kehtarnavaz, N. (2011). A new auto-focus sharpness function for digital and smart-phone cameras. *IEEE Transactions on Consumer Electronics*, 57(3), 1003–1009. <https://doi.org/10.1109/TCE.2011.6018848>

How to cite this article: Bonet Sanz, M., Machado Sánchez, F., & Borromeo, S. (2022). An algorithm selection methodology for automated focusing in optical microscopy. *Microscopy Research and Technique*, 85(5), 1742–1756. <https://doi.org/10.1002/jemt.24035>