4-19-2021

A single-scale fractal feature for classification of color images: A virus case study

Walker Arce

James E. Pierce

Mihaela T. Velcsov

A single-scale fractal feature for classification of color images: A virus case study

Walker Arce [a], James Pierce[a,c], Mihaela Teodora Velcsov [b,*]

[a] *University of Nebraska Medical Center, Omaha, NE 68198, USA*

[b] *University of Nebraska at Omaha, Omaha, NE 68182, USA*

[c] *Institute for Human Neuroscience, Boys Town National Research Hospital, Omaha, NE 68010, USA*

Abstract

Current methods of fractal analysis rely on capturing approximations of an images' fractal dimension by distributing iteratively smaller boxes over the image, counting the set of box and fractal, and using linear regression estimators to estimate the slope of the set count line. To minimize the estimation error in those methods, our aim in this study was to derive a generalized fractal feature that operates without iterative box sizes or any linear regression estimators. To do this, we adapted the Minkowski-Bouligand box counting dimension to a generalized form by fixing the box size to the smallest fundamental unit (the *individual* pixel) and incorporating each pixel's color channels as components of the intensity measurement. The purpose of this study was twofold; to first validate our novel approach, and to then apply that approach to the classification of detailed, organic images of viruses. When validating our method, we a) computed the fractal dimension of known fractal structures to verify accuracy, and b) tested the results of the proposed method against previously published color fractal structures to assess similarity to comparable existing methods. Finally, we performed a case study of twelve virus transmission electron microscope (TEM) images to investigate the effects of fractal features between viruses and across the fac- tors of family (Orthomyxoviridae, Filoviridae, Paramyxoviridae and Coronaviridae) and physical structure (whole cell, capsid and envelope). Our results show that the presented generalized fractal feature is a) accurate when applied to known fractals and b) shows differing trends to comparable existing

methods when performed on color fractals, indicating that the proposed method is indeed a single-scale fractal feature. Finally, results of the analysis of TEM virus images suggest that viruses may be uniquely identified using only their computed fractal features.

Keywords:
Fractal analysis, Virus classification, Color image fractal dimension, Minkowski-Bouligand fractal dimension

Notation glossary

ε	In box-counting algorithms, this is used to determine the box size used across the image.
n_{CH}	The number of color channels in the image, for RGB this would be three.
ρ_A	The number of pixels in a single channel of the image. This is squared since the modified image is in a square space, the height and width are equal.
c_{CH}	The color channel to be selected, in an RGB image, the minimum would be zero and the maximum would be two.
i	The index of the individual pixel, the minimum would be zero and the maximum would be equal to the number of pixels in the image.
	The retrieved pixel intensity value.
c_{CH},i	The normalized, mean pixel intensity of all color channels.
β_C	The standard, log-log plot fitted, fractal dimension. The experimental method for a color, fractal dimension.
$D(S)$ D_C	

1. Introduction

Beginning with Mandelbrot in 1982, fractal geometry has been used to characterize various phenomena in the natural world, from sunspots [1] to the malignancy of tumors [2]. A structure can only be considered a fractal when the fractal dimension is found to be scale invariant, so that the ratio between log $N(\varepsilon)$ and log $1/\varepsilon$ remains constant. Here, $N(\varepsilon)$ represents the minimal number of identical boxes necessary to cover the fractal with square boxes of side length ε. This inherent power-law spatial correlation is what defines a fractal [1,3,4] and by approximating the fractal dimension of a structure, the underlying dynamical processes can be investigated. When considering the most common and accepted methods for approximating

fractal dimension (FD), it was a general requirement that the image be either in grayscale or binarized formats [5–8]. This allows the use of box-counting methods that divide the image into discrete “boxes” of decreasing size and count the number of “on” boxes present in the image. The main drawbacks of this method are that box-counting’s iterative approach can be computationally intensive and/or time-consuming when expanded to higher dimension spaces [9], and a significant portion of the in- formation from the image is removed. Namely, the process of binarization removes most shading information for complex images while eliminating the full red, green, blue (RGB) color space [10]. Within the RGB color space there are sixteen million possible colors while within the binarization color space there are only eight colors; this difference in available information can mean a significant loss of variation depending on the threshold for the binarization algorithm, especially when images have many shades of one color, which can cause critical image details to be ignored [10]. While this could be mitigated by employing more complex binarization algorithms, a fractal dimension calculation compares how detail changes at different scales, providing a statistical index of complexity [11]. So, to lose any detail could wash out these changes from the calculation.

To address the issues surrounding binarization for color image analysis, several techniques have been proposed and explored [12–15]. In general, these methods adapt a standard fractal dimension estimation strategy, such as differential box-counting [12– 14] or simplified Hausdorff Dimension [15] into a five-dimensional *(x, y, r, g, b)* space, and subsequently produce log-log plots within that 5D space for each color channel using boxes of variable size. These methods are, generally speaking, analogous to the analysis of binarized images in terms of their iterative and computationally intensive pipeline; this introduces challenges both in the relative complexity of calculating fractal dimension as well as the comparison of the color fractal dimension to well-known and documented binary fractal patterns (i.e. the Sierpinski Carpet). By extension, this is true

for any n-dimensional fractal of greater complexity than that of a binarized image [16].

It is important to note that, historically, fractals were monochrome, composed of simply ‘on’ and ‘off’ pixels and com- posed of simple geometric shapes, such as lines and curves. While these estimation methods have been used in texture analysis and image segmentation of natural images since 1989, by using the lacunarity, or space-filling, metric [17], the extension to higher order structures has not been fully explored. Consequently, higher order fractals that exist in the color space are a topic of continuing investigation and research. Ivanovici has addressed some of the core issues surrounding color fractal generation and color FD estimation [18,19]. For a structure to be a fractal it must maintain the property of self-similarity [11,18]. This makes color fractals particularly difficult to generate, since natural images cannot be mathematically proven to be infinitely self-similar. Ivanovici addresses this by proposing a color fractal generator and linear regression estimators to compute the fractal dimension of the generated structures [18,19]. These proposed methods still rely on iterative box counting and regression estimators to compute the FD of images but show relative stability at varying scales and adherence to accepted limits.

We propose the use of a novel single-scale fractal feature which can be adapted to n-dimensional space. By adapting the Minkowski-Bouligand fractal dimension equation [6], i.e., box counting, to the individual pixel, a generalized equation can be derived for any image in any color space while simultaneously eliminating linear regression estimators and iterative box size calculations.

2. Background

When calculating the fractal dimension (FD) of any image or structure, the most common methods run in iterations, varying a window size and counting the number of one or zero value windows inside the image, placing results on a log-log plot, and then using the line of best fit to yield an

estimated FD [7]. This is performed to accommodate for the scale variance that occurs when measuring the FD of natural images and by estimating the slope of the line, the scale variance can be compensated. According to Bisoi et al., there are multiple common methods to perform box-counting, including: reticular cell counting, Keller's approach, and differential box-counting [7]. There are also several variations and implementations of these algorithms, including the improved differential box-counting method which, by sliding boxes of various sizes across the image, can solve problems that exist with the differential box-counting method [8]. The main drawback of these methods (improved and original) is that they still generally operate using grayscale images [8,20,21,2].

Since all digital images generally have between one (monochrome) and four (ARGB, CMYK) pixel data channels, and all color images have at least three channels to represent color and/or intensity (i.e. RGB, YUV, HSV, etc.), [22] a way to flexibly represent this information for use in the calculation of FD is needed. Structurally, a digital image is comprised of a finite grid of cells represented as pixels that store the color values at a given location. There has been work done by other researchers, most prominently by Ivanovici, to adapt these grayscale algorithms to color images by extending the 2D vector to the RGB space and thereby generating a 5D vector. A general rule proposed in many of their papers is that the minimum color FD for an *(x, y)*, or 2D, image would be $D_{min} = 2$, while for an *(x, y, r, g, b)*, or 5D, image would be $D_{max} = 5$. Here, *(x, y)* represents the pair of coordinates for a point in the xy-plane [19]. Ivanovici et al. did this through the extension of a probability density estimator to the RGB color space, which is then run through various regression estimators to provide a cohesive measure of an image's FD [19]. The validation of their method was achieved by generating novel color fractals through the Hurst parameter and midpoint displacement [18]. These generated fractals act as a type of benchmark that can be used in the characterization of FD estimators. Additionally, [23] proposes an additional set of benchmarks that

include the generation of increasingly random color images, and their 3D histograms, which are used to estimate the fractal dimension, similar to other studies [9].

Various other methods have been proposed to address FD estimation in the color space, intended for differing application cases. For instance, FD has been applied to the authentication of Jackson Pollock paintings [24], whose work has been the subject of interest due to the fractal nature of his paintings and their seemingly intentional application of self-similar structures. There has also been continuing work in the extension of the box-counting dimension to an RGB color space for the purposes of roughness characterization and texture classification [25,26].

3. Methods

For this study, our main interest was in developing a single-scale fractal feature estimation method that does not rely on regression estimators or iteratively sized boxes, while also providing accurate estimates of known FDs. To accomplish this, the parameter ε will be fixed at the size of a single pixel, meaning that $N(\varepsilon)$ will equal the total number of pixels in the image [6,3].

3.1. *Derivation of a generalized single-scale fractal feature equation*

Generally, the total number of pixels in a digital image can be represented by the product of the number of color channels in the image and the height and width of the image. In this notation, nCH is the total number of channels. This is comparable to the calculation of the total number of boxes used by methods such as box- counting. Our goal is to provide an alternative to the FD given by:

$$\boldsymbol{D}(\boldsymbol{S}) = \lim_{\varepsilon \to 0} \frac{\log \boldsymbol{N}(\varepsilon)}{\log \frac{1}{\varepsilon}} \quad (1)$$

Recall that $N(\varepsilon)$ represents the minimal number of identical boxes

necessary to cover the fractal S with square boxes of side length ε. The retrieval of a pixel would be achieved by selecting a color channel (cCH) and an index (i) in the associated image. If an image is single color, then cCH is fixed at zero, for a zero indexed array. The value of a pixel in the image is given by $c_{CH,i}$, representing the luminosity of the pixel selected at a specified index. Since our interest is to create an analogue to binary box counting for color images, we are going to sum the contents of the image pixels, across all channels. Assuming all pixels in the image are used and no filtering step was used, then we can find the average of the pixel intensity by dividing by the number of pixels in the im- age, denoted by ρ_A. Finally, the average can be normalized on a [0, 1] scale by dividing by the maximum possible pixel value, denoted by α_D.

$$\beta_C = \frac{1}{\alpha_D} \cdot \frac{\sum_{c_{CH}=0}^{n_{CH}} \sum_{i=0}^{\rho_A} \ell_{c_{CH},i}}{\rho_A} \quad (2)$$

Eq. (2) is a direct analogue to binarization but performed on the color channels and allowing for the variations to remain. This can be used to determine the minimum value of a cell. An image that has '0's across an entire channel would result in a domain error, so Eq. (3) provides a method to avoid this.

$$\ell_{c_{CH},i_{MIN}} = \frac{\alpha_D}{\rho_A} \quad (3)$$

By using Eq. (3) as a replacement for pixel channels that yield a β_C of '0' the minimum fractal dimension results. Generally, this will not happen in a digital image, since pixels in a color channel are commonly represented as 8-bit or 16-bit values, so they have values that occupy a [0, 255] and [0, 65535] space, respectively. The pixels cannot have fractional values in those spaces and unless the color channel is all '0's, Eq. (3) will not be needed. So, this can be adapted to the Minkowski-Bouligand FD given in Eq. (1). In this approach, we avoid the limiting process so that $N(\varepsilon)$ is

defined as $\rho_A \cdot \beta_C$, applying the mean, normalized pixel value over the entire pixel space, and $1/\varepsilon$ is fixed at the number of pixels in the image raised to the root of the dimension of the image. Plugging the previous derivations into the above equation, the resultant experimental FD equation is obtained by considering:

$$\frac{\log \rho_A \cdot \beta_C}{\log \rho_A^{\frac{1}{D}}} \tag{4}$$

Since the box size is fixed in our implementation, there is no limit to be tested over a continuous range of box sizes. Finally, the equation can be simplified by applying logarithm rules to the numerator and denominator. This is shown in Eq. (5).

$$D_C = \left(\frac{\log \rho_A \cdot \beta_C}{\log \rho_A^{\frac{1}{D}}} \right) = D\left(\frac{\log \beta_C}{\log \rho_A} + 1 \right) \tag{5}$$

3.2. *Implementation of the novel color FD equation*

For simplicity, the proposed algorithm was implemented in the Python language. Considering there are non-square fractal geometries, this implementation also takes into consideration the alpha channel of a PNG image, which provides transparency. By removing the background of non-square fractals, such as the Sierpinski Triangle, the pixels that are part of the background can be subtracted from the calculations to achieve a more accurate result. Additionally, we used MATLAB to compute the Minkowski-Bouligand dimension developed by F. Moisy [27] and to use the estimator provided by Ivanovici [19].

3.3. *Experimental methods*

For this study, three case studies were conducted to validate and evaluate the proposed single-scale fractal feature equation against the results of the Minkowski-Bouligand dimension and the estimator presented in [19]. These studies allow a comparison of the proposed method with those presented in the literature as well as show potential future avenues of

research. To allow the comparison to other methods the proposed estimator will be run on known fractal structures to see its performance in estimating theoretical structures. Then, boundary testing will be performed on the estimator to evaluate its resolution and operation over its entire range of possible input values. These tests will also be accompanied by a timing test to assess its computational performance. Next, the dataset presented by Ivanovici in [18] will be tested to compare the affect the Hurst parameter has on the resultant fractal dimension of the proposed estimator. Finally, a case study looking at the statistical correlations between virus families, virus structures, and a virus' FD will be performed using transmission electron microscope (TEM) images.

4. Experimental results

Four tests were performed for this study, comparing the proposed method and the box-counting dimension; the first of these was a validation study against known fractals, the second was a boundary test using randomly filled images of increasing size accompanied by a timing test, the third was a comparison study against the color fractals from [18], and the final test examined the three methods (box-counting, [19], and the proposed method) against color images. The fractal dimension estimator provided in [19] by Ivanovici has been made publicly available for research purposes as a MATLAB script and that is what will be used in this experiment, in addition to the script from [27].

The implementation of the Minkowski-Bouligand dimension method from [27] iteratively runs through the image with subsequently smaller squares, plots the results of each iteration, and finds the line of best fit to calculate the FD. By comparison, the proposed method performs a single calculation and has no iterative approach, meaning that its computational requirements are significantly reduced. An ideal fractal image would be one where there are only straight lines and each line is represented by a

sequence of zero intensity pixels, representing the color black. Since this is generally not achieved in non-square fractals, an approximation is used to create the effect of angled lines with smooth edges, as in the Sierpinski Triangle. This will cause the calculated FD value to approximate the true FD, as in Minkowski-Bouligand box-counting. In summary, this means that a higher-resolution im- age yields a more accurate FD for any method, and to achieve high accuracy when compared against the theoretical value, square fractals are optimal.

Table 1
Results for known fractal calculations.

	Fractal name	Theoretical	Composite	Percent error
Experimental	Sierpinski Carpet	1.89278926071437	1.89278926071437	0.0000000%
	Sierpinski Triangle	1.58496250072115	1.57634370604398	0.5437854%
Box Counting	Sierpinski Carpet	1.89278926071437	1.89180000000000	0.0522647%
	Sierpinski Triangle	1.58496250072115	1.59440000000000	0.5954399%

4.1. *Known fractals*

For the first assessment, the FD of two common fractals (the Sierpinski Triangle and Sierpinski Carpet) was computed and com- pared using both the Minkowski-Bouligand box-counting method and the proposed method. These two fractals are well-known and have theoretical FDs of 1.89278 and 1.58496, respectively. They are also monochrome, which in the image is represented as all three- color channels being at either zero intensity (“black”) or full intensity (“white”). This means that, when compared against methods that binarize the image, the results should be similar since they both deal with a theoretical zero and one value, with zero being no intensity for the pixel and one being full intensity for the pixel.

Before testing known fractals, a simple theoretical benchmark is necessary. For a full luminosity image, an all-white image, the FD should be

2. Alternatively, for a no luminosity image, an all-black image, the FD should be 0. This was tested for both methods to see their output and their handling of these boundary cases and both came back consistent. Per Section 3.1, Eq. (2) was used in the pro- posed method to calculate the minimum intensity for each color plane. The results for the fractal analysis of these structures are shown in Tables 1 and 2 for the eleventh iteration of each fractal.

Note that the experimental dimension specifies red, blue, and green color channel calculations, with composite being the com- bination of the three channels, which the box counting dimension does not have. If all four calculations are the same, then the image is grayscale. The percent error is the percent difference between the theoretical FD and the calculated FD using the respective equation. The composite dimension is the proposed method (Table 2).

Table 2
Percent error for known fractal calculations and theoretical values.

	Box counting error	Composite error	Improvement
Sierpinski Carpet	0.0522647%	0.0000000%	100%
Sierpinski Triangle	0.5954399%	0.5437854%	9%

Shown above (Fig. 1) is the output of the MATLAB implementation of the Minkowski-Bouligand dimension, with the input image and the plotted iteration dimension [27]. Shown above (Table 2) are the tabulated results for the testing, including comparisons.

Shown above (Table 2) is the comparison between the two methods for each fractal. The box counting and experimental error is pulled from Table 1. The improvement column shows the increase or decrease in the

FD estimation towards or away from the theoretical value. An ideal case would be when the error goes to 0%, which means it matches the theoretical value. Results for known fractals show a significant improvement in accuracy for square fractals such as the Sierpinski Carpet, with results being identical to theoretical calculations up to fifteen decimal places. Essentially, it gives a perfect calculation of the FD of the ideal fractal structure. For a non-square fractal, the Sierpinski Triangle, there is a loss of accuracy which is due in part to the size of the fractal structure and the amount of approximation that is present to digitally represent an angled line. In the case of the eleventh iteration Sierpinski Triangle, the estimated FD has improved but it is still not exact. The approximation of the angled lines makes it where the true structure of the fractal is obscured from this approximation, leading to a slightly lower FD being estimated.

4.2. *Boundary testing & timing tests*

There are some published experiments that demonstrate the application of a FD estimator to digital color images. The ones used and expanded on here are from Chauveau et al. and Nikolaidis et al., where an image with an x-dimension gradient from [0 – 255] is placed in the red channel, a y-dimension gradient from [255 – 0] is placed in the green channel, and the blue is channel is set to a constant of 128 [23]. This image is proposed to have a fractal dimension of 2. Then, going from the blue to green to red channels, each color plane is randomized, and their fractal dimensions tested. Nikolaidis et al. calculates these to have a fractal dimension of 3, 4, and 5, respectively [9]. In addition to these tests, we are going to also test solid color images of red, green, and blue. Finally, we will test filling an image with solid values across each color channel in increasing increments to perform a boundary test, to see the minimum and maximum fractal dimension estimated by the proposed estimator. These fill bounds are defined as [$cCH, iMIN$, 255] in increments of 0.001, leading to a total of over 250,000 iterations. These results are shown below, (Fig. 2, Table 3, Table 4, Fig. 3).

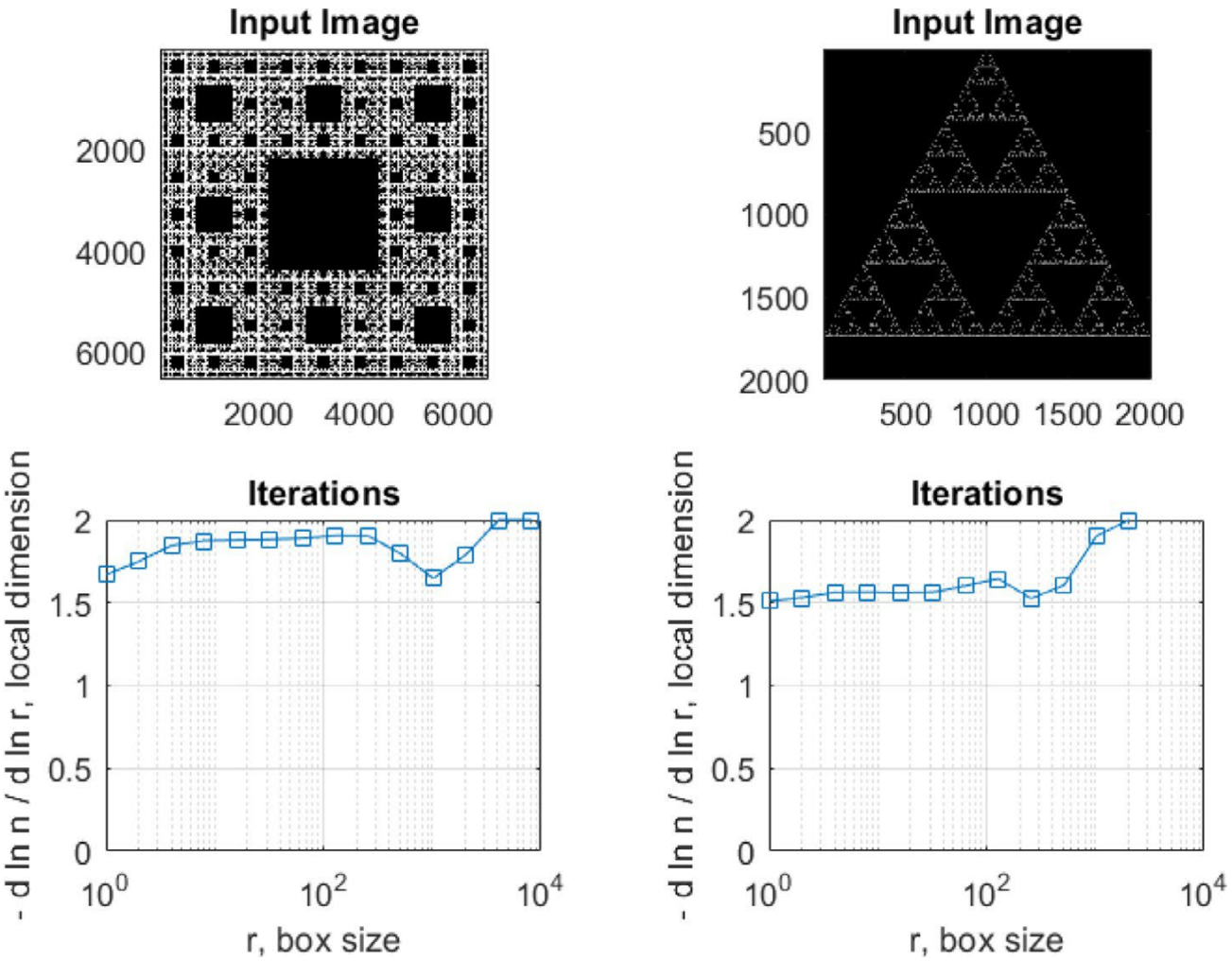

Fig. 1. Plotted output of the MATLAB implementation.

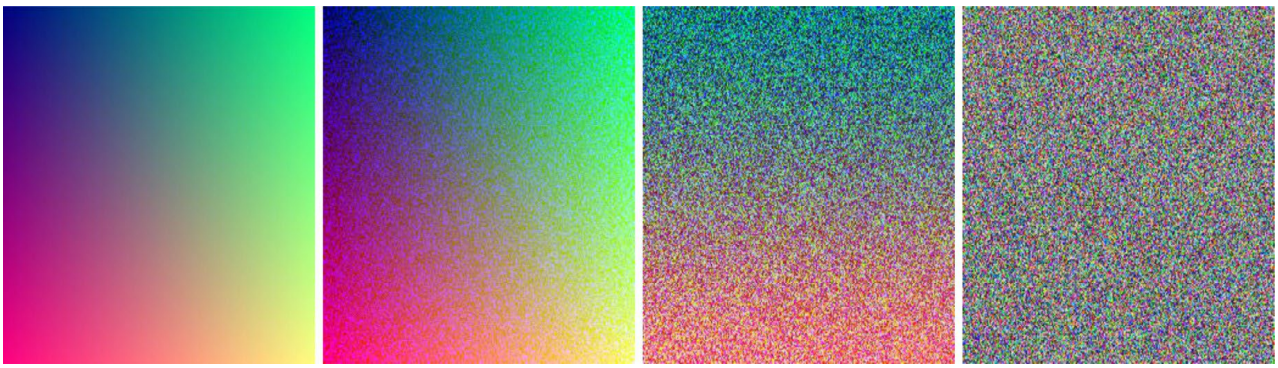

Fig. 2. From left to right, the gradient image has each channel randomized.

While there is a significant difference between the results of the proposed estimator and the estimators presented in [23,9], it is important to note that color images are not fractal structures proper. This is the foundation laid by Ivanovici in his 2010 and 2020 papers where color fractal images with correlated color com- ponents can be generated. These tests provide an interesting way to compare two estimators to see how their outputs may vary and how their measurements may be contrasted or compared. Since the proposed method eschews the use of regression estimators and varying box sizes, it can be expected that the estimator will tend towards higher FD estimations.

Table 3

Color testing for the solid color and gradient images.

	Red	Green	Blue	Gradient
Dimension	4.504712977	4.504712977	4.504712977	4.688088951

Table 4

Color testing results for random images.

	Blue random	Green blue random	Full random
Dimension	4.687716884	4.687644164	4.688462452

Shown above (Fig. 4 and Table 5) are the results of the computation testing, where the estimator was run on randomized im- ages of increasing sizes for 1000 iterations and the execution time was averaged after all computations were completed. Even at the largest image size tested, it still took slightly less than 500 milliseconds to complete the estimation. While the other two estimators, [19] and [27], were provided as MATLAB scripts, they still took several to tens of seconds to perform their estimation.

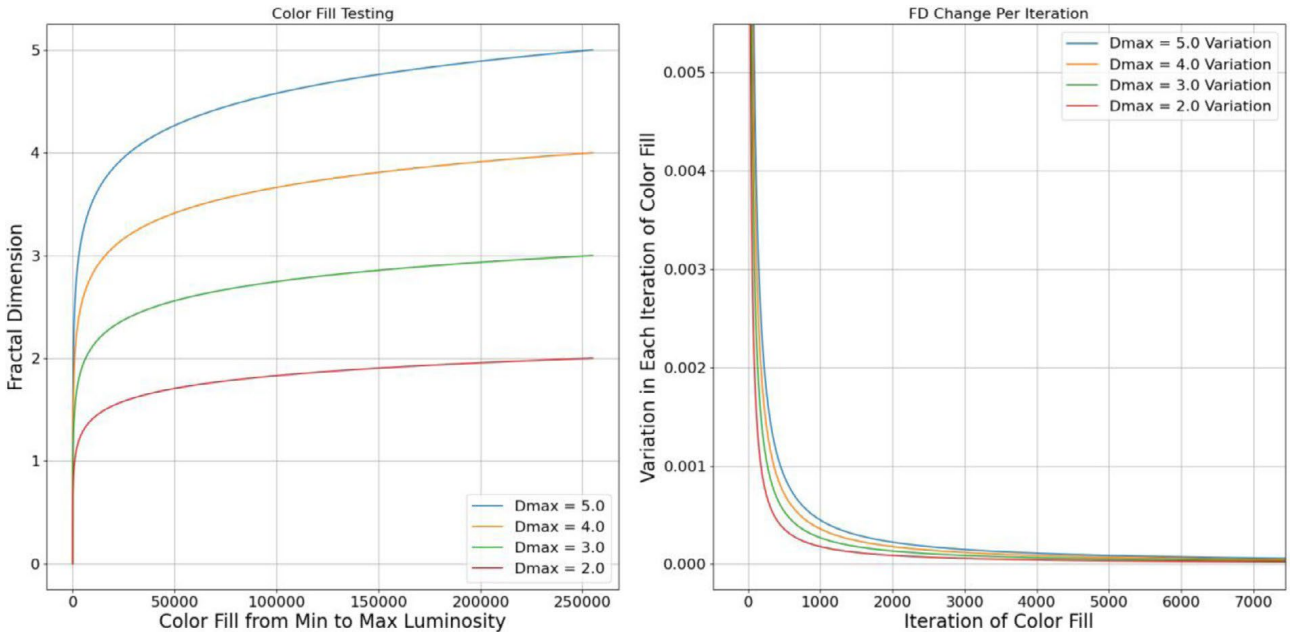

Fig. 3. The left figure is the increase in fractal dimension in each iteration and the right is a zoomed in view of the change in variation between successive fractal dimension estimations.

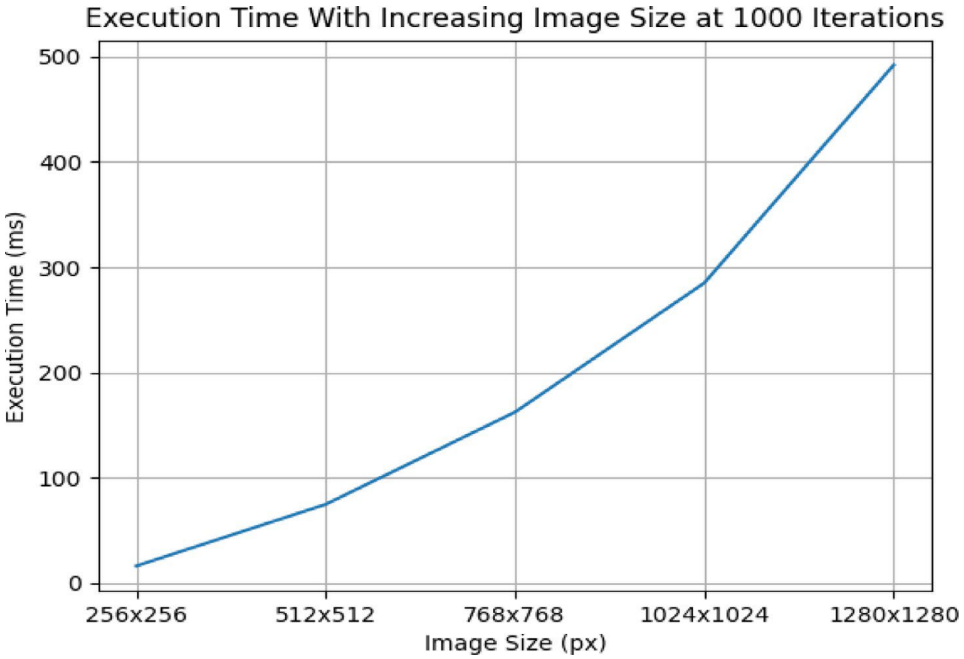

Fig. 4. Plotted results of the timing tests showing a roughly linear increase in the computation time vs. image size.

Table 5

Results of the timing tests with increasing image size and randomly filled images.

Image Size (h x w)	256 × 256	512 × 512	768 × 768	1024 × 1024	1280 × 1280
Image Size (h x w)	16.3731243	74.849884	162.4125342	285.076265	491.9684485

Fig. 5. Examples of the generated color fractals [18], the left being with an H = 0.1, the middle with an H = 0.5, and the right with an H = 0.9.

4.3. *Ivanovici dataset comparison*

Ivanovici made the dataset used in their color fractal study open source for other researchers to use in validation of their own methods for color FD (CFD) estimation [18]. They provide nine color fractal images generated using a midpoint displacement method that was adapted from grayscale to the color domain. By varying the Hurst parameter (H), which is a measure of long-term memory usually associated with a time series, the texture of the structure is made either rougher or smoother on a zero to one scale, respectively. As part of our investigation, we compared their color FD with our own using the same images.

The images used above, and their corresponding computed fractal dimensions (Fig. 5 and Table 6), had their Hurst parameter varied by 0.1 on each iteration, so the first image has an H = 0.1 and the last image has an H = 0.9. The column to the right of each CFD is the difference between the current and the last CFD. This shows the variation that the Hurst parameter introduces on each successive iteration. For the method in [18], the increase of the Hurst parameter reduces the FD significantly on each

iteration, while for our proposed method it increases the FD slightly on each iteration.

Table 6

CFD estimation results with their corresponding difference with changing Hurst parameter.

Image	[18] CFD	[18] Variation	Proposed CFD	Proposed variation
H01	3. 8268		4.6446	
H02	3.9134	0.0866	4.6572	0.0125
H03	3.9113	−0.0021	4.6744	0.0172
H04	3.6373	−0.2740	4.6834	0.0090
H05	3.2692	−0.3681	4.6876	0.0042
H06	2.8623	−0.4069	4.6910	0.0035
H07	2.5673	−0.2950	4.6942	0.0032
H08	2.3573	−0.2100	4.7031	0.0090
H09	2.2372	−0.1201	4.7144	0.0112

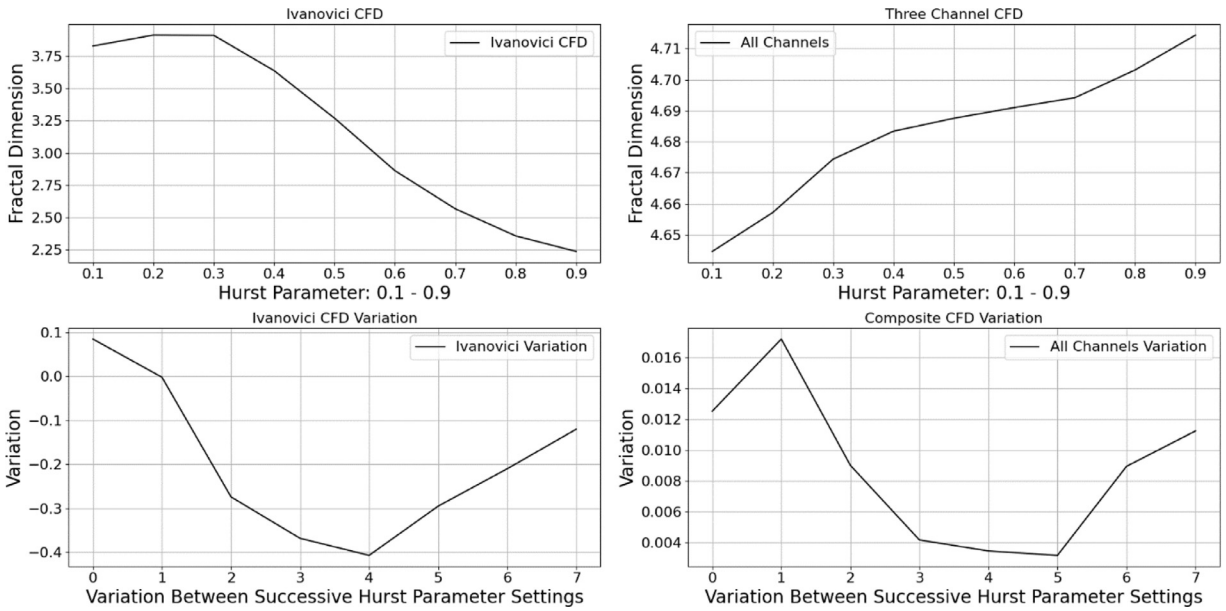

Fig. 6. The plotted variation introduced by the Hurst parameter for both methods.

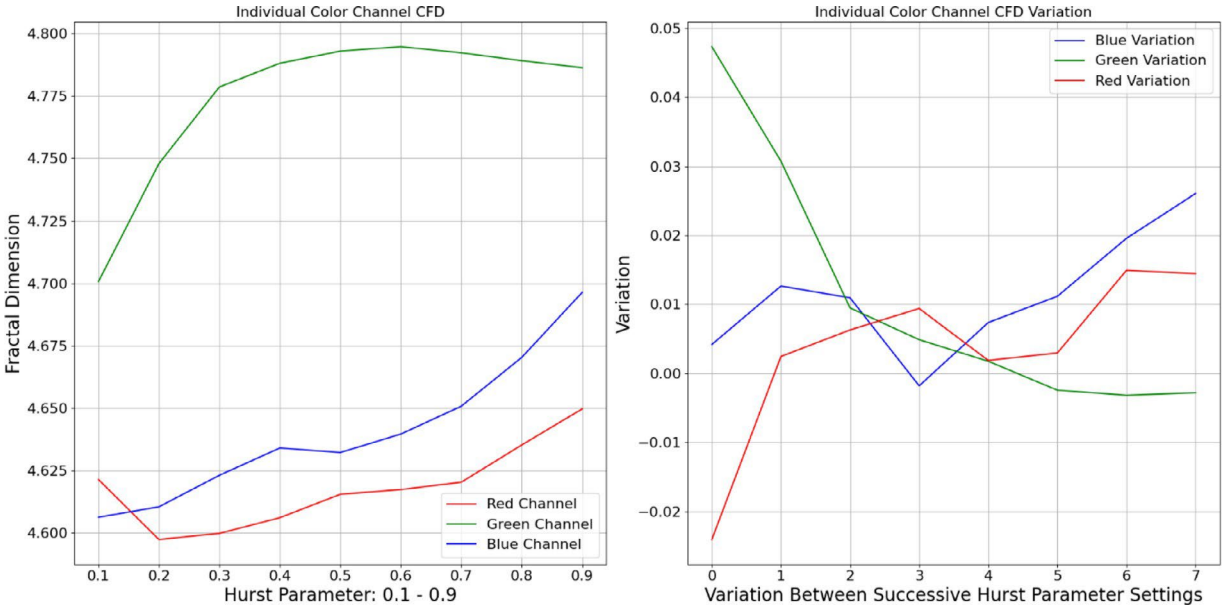

Fig. 7. Plotted CFD and variation in each individual color channel calculated when using the proposed method.

Fig. 6 above show the change in variation over different Hurst scales. The changes in CFD follow a similar curve, where it initially decreases but then increases at higher scales. This shows that as the image gets smoother, the FD shifts approximately in tan- dem. The main difference between the two methods is that for the method in [18], the FD shifts significantly over Hurst scales, but for our proposed method the variation is significantly dampened. With the limitations imposed by a single-scale fractal feature, this indicates the impossibility of an appropriate estimation. Fig. 7 shows the same estimations and variations for each of the color channels. There is no direct comparison available between the individual color channels and the CFD provided by [18], but it does show the significant variation that occurs when accounting for individual color channels, rather than all color channels simultaneously.

4.4. *Virus images and methodology*

While fractal analysis has been applied extensively to a variety of fields, it has relatively little penetration into virus classification and characterization. When considering the visual structure of a virus as seen through a transmission electron microscope (TEM), the envelope of the virus has the most variety. The viral load of a virus is delivered through the glycoproteins on the envelope, which give each family of virus a distinct 2D outline [28]. Additionally, the genome sequence of a virus is a unique representation of a strain of virus and has been used in fractal and entropy analysis to classify strains of viruses [29,30]. For clarification, there are two important sub-structures that are exposed through a TEM im- age: the capsid and envelope. The capsid is the protein shell that encapsulates the viral RNA payload and nucleic acid, and the envelope is the outer shell that has glycoprotein projections, which are used to anchor the virus and deliver the RNA payload [31]. Across different viruses, there is significant variation on the structure, distribution, and size of the envelope and its glycoproteins. Following these studies, fractal analysis was performed across a variety of viruses using Minkowski-Bouligand dimension, the estimator proposed in [19], and the experimental dimension to see fractal dimension estimation differences between the three methods and determine if there are any significant group effects across the dataset variables.

Table 7

Tabulated results for the single variable statistical measures.

	Across family			Across virus			Across structure		
	X2	dF	*p*	X2	dF	*p*	X2	dF	*p*
Red	10.91	3	0.012	29.16	11	0.002	0.194	2	0.908
Green	2.872	3	0.412	24.31	11	0.011	0.176	2	0.916
Blue	3.693	3	0.297	30.99	11	0.001	0.272	2	0.873
Composite	1.951	3	0.583	28.65	11	0.003	0.047	2	0.977
Box Counting	11.49	3	0.009	27.46	11	0.004	4.679	2	0.096
Ivanovici	21.190	3	0.000	29.15	11	0.002	3.182	2	0.204

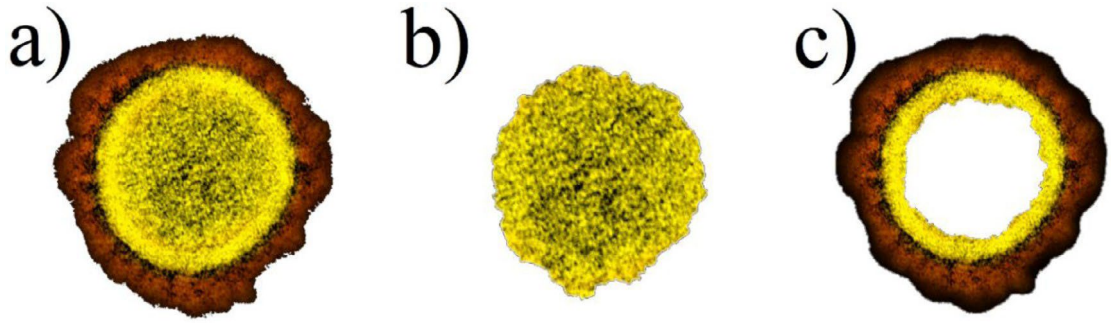

Fig. 8. Example of virus images, specifically SARS-CoV-2. Image "a" is the whole cell, image "b" is the capsid, and image "c" is the envelope. Credit: NIAID Integrated Research Facility (IRF) in Fort Detrick, Maryland.

To do this, TEM images were collected from the National Institute for Allergies and Infectious Diseases' (NIAID) Flickr account, the National Institutes of Health (NIH) archives, and from the Centers for Disease Control and Prevention (CDC) archives. These im- ages were categorized across virus and family, leading to the collection of twelve viruses equally distributed across four families, with a single TEM image per virus. Then, the images were manually segmented into three separate images: whole cell, capsid, and envelope. While TEM images are intensity based and therefore not true color digital images, their main benefit is that they show a cross sectional view of a complex structure and serve as an ideal subject for a fractal dimension study. Finally, the background of the images was made transparent to take advantage of the experimental method's ability to exclude non-structure pixels from the estimation. The capsid is visually defined as the central, inner, color distribution of the cell, up to a distinct wall where the color dark- ens. The inverse of the capsid is being defined as the envelope, including the glycoproteins. This result is shown in Fig. 8.

4.5. *Virus results and statistical models*

The above procedure was followed for all twelve collected viruses.

Note that there are three grayscale images in this dataset: H7N9, MeV, and SARS-CoV. This is due to the lack of availability of colored TEM images for these viruses. In general, there appears to be a greater degree of variability in measures resulting from the proposed method than with the Minkowski-Bouligand method. Based on the greater degree of accuracy observed with the proposed method in Section 4.1, this would indicate that it more closely estimates the true fractal dimension of the virus structures. To closely analyze these results, Kruskal-Wallis H tests were per- formed across the virus families, structures, and viruses to com- pare the mean differences between their FDs. For all statistical measures, a p-value of 0.05 was considered statistically significant. There are six statistical tests performed on this data: comparisons across virus family only, comparisons across viruses only, comparisons across structures only, comparisons of viruses against structure, comparisons of virus family against each structure, and com- parisons of a virus against the other viruses in the same family.

4.6. *Color virus images statistical results*

Kruskal-Wallis H tests for group mean differences between viral families, viral strains and viral structures were performed to assess the differences present (if any) in FD that may indicate characteristic features of the viruses. The test statistic, degrees of freedom, and significance are tabulated in Table 7 for the first three single variable measures.

These results demonstrate that, when accounting for family as a statistical measure, there are significant mean differences between groups for all three methods used. This implies that the FD of the virus can distinguish between families of viruses. Additionally, each individual virus is distinct in its fractal dimension estimation, with all measures showing significance in their mean FDs. This is con- founded by the *Virus by Structure* statistic, which shows $p = 0.443$ across the board for all measures, implying that using an individual structure, such as *whole cell*, is not sufficient to distinguish be- tween viruses. Finally, there were no significant differences in

FD for viral sub-structure, indicating that segmenting each structure is necessary to isolate features in the virus. The box counting and Ivanovici fractal dimension differs from the proposed method in that it shows a higher (albeit more variable) sensitivity to the seg- mentation.

When accounting for structure and family as statistical measures (Table 8), the differences in mean FD are not statistically significant, yielding no distinct differences, affirming the *Virus by Structure* statistic. This is not the case for the method proposed by Ivanovici for the whole cell and capsid. In addition to affirming the lack of group mean differences between families and structures, it also indicates that the method of segmentation is not necessary to yield a similar FD, and that when using family as the measure, the mean FD for each family is statistically identical.

Table 8

Tabulated results for the family by structure statistic.

	Whole cell			Capsid			Envelope		
	X2	dF	*P*	X2	dF	*p*	X2	dF	*p*
Red	3.718	3	0.294	5	3	0.172	4.128	3	0.248
Green	1.462	3	0.691	1.718	3	0.633	2.077	3	0.557
Blue	1.051	3	0.789	2.077	3	0.557	3.000	3	0.392
Composite	0.744	3	0.863	1.256	3	0.740	1.667	3	0.644
Box Counting	2.897	3	0.408	3.821	3	0.282	5.615	3	0.132
Ivanovici	8.333	3	0.040	8.744	3	0.033	7.205	3	0.066

Table 9

Tabulated results for the virus by family statistic.

	Orthomyxoviridae			Filoviridae			Paramyxoviridae			Coronaviridae		
	X2	dF	*p*	X2	dF	*p*	X2	dF	*p*	X2	dF	*p*
Red	2.400	2	0.301	6.489	2	0.039	5.600	2	0.061	1.156	2	0.561
Green	5.956	2	0.051	7.200	2	0.027	2.756	2	0.252	3.200	2	0.202
Blue	7.200	2	0.027	5.956	2	0.051	7.200	2	0.027	7.200	2	0.027
Composite	6.489	2	0.039	6.489	2	0.039	5.689	2	0.058	5.400	2	0.066
Box Counting	1.422	2	0.491	6.489	2	0.039	5.600	2	0.061	5.067	2	0.079
Ivanovici	5.422	2	0.066	5.600	2	0.061	6.489	2	0.039	1.422	2	0.491

When considering virus and family as statistical measures, all virus families showed significance for some measure of the pro- posed method; this demonstrates that the distribution of the mean FD was not the same across categories of virus. Only Filoviridae was significantly different (p = 0.039) for the box counting fractal dimension measure and only Paramyxoviridae was significantly different (p = 0.039) for the method in [19]. This lends additional weight to the *Across Virus* statistic, which showed that each virus had a significantly different mean FD. That should, and in one case for box counting and [19] does, hold true when the sample size is collapsed to a single family using only three viruses, (Table 9).

5. Discussion

When compared to other methods, such as box-counting, the proposed fractal feature shows greater accuracy for known fractals. Using the proposed method on the Sierpinski Carpet yielded an identical calculation as the theoretical value of the fractal, which the box counting dimension did not. Non-square fractals add in unnecessary pixels to the image that skew the result either up or down, depending on the color and number of pixels. This skew can be significant as well, as shown in the Sierpinski Triangle fractal, but the proposed method still provided a closer approximation of the fractal dimension than the box-counting estimation, regardless of this approximation.

The derived fractal feature not only retains the color space present in digital images but accomplishes this without the use of an iterative process or linear estimators. The accuracy of this feature was shown by using known fractal structures and comparing it against the most common box-counting dimension, which the feature was based from. Additionally, color fractals created by Ivanovici in [18] were evaluated and the feature was shown to exhibit a similar fractal dimension shift as the theoretical dimensions for the fractals. The main issue when showing the use of the proposed estimation method against color images is that there are no known

fractal dimensions for color images other than the work done by Ivanovici in [18] and [19]. Regardless, the analysis of the sub-structures of viruses yielded promising results which showed statistical phenomena when comparing the virus' fractal features against the other virus' mean fractal features in the dataset. Even with a relatively small number of data points available, which is restricted by limited availability of virus TEM images, there was consistent significance in the comparison of mean fractal features across all studied viruses, suggesting a possible correlation be- tween a virus' fractal features and its identification, indicating that viruses, such as 2009 H1N1, can be identified by its fractal features when segmented. By pursuing a broader study with greater sample size and control over TEM image consistency these correlations can be explored with a broader collection of viruses.

Through this study, a new, generalized method for the calculation of fractal features for images of arbitrary size was proposed and was validated using well-known fractals. The application of this fractal feature was explored using various color benchmarks, a color fractal dataset, and a virus case study that showed both that the proposed method more closely estimates the true structure of the processed images but also that viruses can be identified using the fractal features of their sub-structures alone, all while maintaining a low computational overhead.

Author statement

Walker Arce: Conceptualization of fractal dimension equation, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, original draft, review, editing

James Pierce III: Conceptualization of fractal dimension equation, formal analysis, validation, review

Mihaela Teodora Velcsov: Conceptualization of virus case study, formal analysis, validation, review, editing, methodology

Data statement

The data and the results used to compose this text has been made available on Mendeley Data.

Declaration of Competing Interest

This work was performed without any financial interest and without any financial support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chaos.2021.110849.

References

[1] Mandelbrot Benoit B, Mandelbrot Benoit B. The fractal geometry of nature, 1. New York: WH freeman; 1982.

[2] Chan Alan, Tuszynski Jack A. Automatic prediction of tumour malignancy in breast cancer with fractal dimension. Royal Society open science 2016;3(12):160558.

[3] Schroeder Manfred, Fractals Chaos. Power Laws: Minutes from an infinite paradise; 1991. p. 429.

[4] Bak P, Chen K. The physics of fractals. Physica D 1989;38(1):5–12.

[5] Chen Wen-Shiung, et al. Algorithms to estimating fractal dimension of textured images. In: 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221), 3. IEEE; 2001.

[6] Backes, André Ricardo, and Odemir Martinez Bruno. "Fractal and Multi-Scale Fractal Dimension analysis: a comparative study of Bouligand-Minkowski method." arXiv preprint arXiv:1201.3153 (2012)

[7] Bisoi Ajay Kumar, Mishra Jibitesh. On calculation of fractal dimension of images. Pattern Recognition Letters 2001;22(6-7):631–7.

[8] Liu YU, et al. An improved differential box-counting method to

estimate fractal dimensions of gray-level images. Journal of visual communication and Image Representation 2014;25(5):1102–11.
[9] Nikolaidis Nikolaos S, Nikolaidis Ioannis N. The box-merging implementation of the box-counting algorithm. Journal of the Mechanical Behavior of Materials 2016;25(1-2):61–7.
[10] Sheikh Liaquat Majeed, et al. An adaptive multi-thresholding technique for binarization of color images. WSEAS Transactions on Information Science and Applications 2005;2(8):1202–7.
[11] Falconer Kenneth. Fractal geometry: mathematical foundations and applications. John Wiley and Sons; 2004.
[12] Nayak Soumya Ranjan, et al. Fractal dimension of RGB color images. Optik 2018;162:196–205.
[13] Nayak Soumya, Khandual Asimananda, Mishra Jibitesh. Ground truth study on fractal dimension of color images of similar texture. The Journal of the Textile Institute 2018;109(9):1159–67.
[14] Nayak Soumya Ranjan, Ranganath Abadhan, Mishra Jibitesh. Analysing fractal dimension of color images. In: 2015 International Conference on Computational Intelligence and Networks. IEEE; 2015.
[15] Panigrahy Chinmaya, Seal Ayan, Mahato Nihar Kumar. Fractal dimension of synthesized and natural color images in lab space. Pattern Analysis and Ap- plications 2020;23(2):819–36.
[16] Schafer A, Muller B. Bounds for the fractal dimension of space. Journal of Physics A: Mathematical and General 1986;19(18):3891.
[17] Keller James M, Chen Susan, Crownover Richard M. Texture description and segmentation through fractal geometry.. Computer Vision, Graphics, and image processing 1989;45(2):150–66.
[18] Ivanovici Mihai. Fractal dimension of color fractal images with correlated color components. IEEE Transactions on Image Processing 2020;29:8069–82.
[19] Ivanovici Mihai, Richard Noël. Fractal dimension of color fractal images. IEEE Transactions on Image Processing 2010;20(1):227–35.

[20] Souza Paulo, Victor S, Alves RL, Balthazar WF. A Tool to Study Fractals in an Interdisciplinary Perspective. The Physics Teacher 2019;57(7):467–9.

[21] Chamorro-Posada Pedro. A simple method for estimating the fractal dimension from digital images: The compression dimension. Chaos, Solitons and Fractals 2016;91:562–72.

[22] Schwarz Michael W, Cowan William B, Beatty John C. An experimental com- parison of RGB, YIQ, LAB, HSV, and opponent color models. ACM Transactions on Graphics (TOG) 1987;6(2):123–58.

[23] Chauveau Julien, Rousseau David, Chapeau-Blondeau François. Fractal capacity dimension of three-dimensional histogram from color images. Multidimensional Systems and Signal Processing 2010;21(2):197–211.

[24] Taylor Robert P, et al. Authenticating Pollock paintings using fractal geometry. Pattern Recognition Letters 2007;28(6):695–702.

[25] Nayak Soumya Ranjan, Mishra Jibitesh. An improved method to estimate the fractal dimension of colour images. Perspectives in Science 2016;8:412–16.

[26] Nayak Soumya Ranjan, Mishra Jibitesh. On calculation of fractal dimension of color images. International Journal of Image, Graphics and Signal Processing 2017;9(3):33.

[27] Moisy, F., Computing a fractal dimension with matlab: 1D, 2D and 3D box- counting, 9 July 2008. [Online]. Available: :http://www.fast.u-psud.fr/~moisy/ml/boxcount/html/demo.html [Accessed 6 November 2018].

[28] He Ji-Huan. Fatalness of virus depends upon its cell fractal geometry. Chaos, Solitons and Fractals 2008;38(5):1390–3.

[29] George Tremberger Jr, et al. Archaeon and archaeal virus diversity classification via sequence entropy and fractal dimension. Instruments, Methods, and Missions for Astrobiology XIII, 7819. International Society for Optics and Pho- tonics; 2010.

[30] Holden Todd, et al. Nipah virus classification via fractal dimension and shannon entropy. In: 2010 4th International Conference on Bioinformatics and Biomedical Engineering. IEEE; 2010.

[31] Horne RW, Wildy P. Virus structure revealed by negative staining. Advances in virus research 1964;10:101–70.