

Article

Not peer-reviewed version

Histropy: A Computer Program for Quantifications of Histograms of 2D Gray-Scale Images

[Sagarika Menon](#) and [Peter Moeck](#) *

Posted Date: 12 August 2024

doi: 10.20944/preprints202408.0729.v1

Keywords: image entropy; histograms; calculations; computer program



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Histropy: A Computer Program for Quantifications of Histograms of 2D Gray-Scale Images

Sagarika Menon and Peter Moeck *

Nano-Crystallography Group, Department of Physics, Portland State University, Portland, OR 97207-0751, USA

* Correspondence: pmoeck@pdx.edu

Abstract: The computer program "Histropy" is an interactive Python program for the quantification of selected features of two-dimensional (2D) images/patterns (in either JPG/JPEG, PNG, GIF, BMP, or baseline TIF/TIFF formats) by means of calculations based on the pixel intensities in this data, their histograms, and user-selected sections of those histograms. The histograms of these images display pixel-intensity values along the x-axis (of a 2D Cartesian plot), with the frequency of each intensity value within the image represented along the y-axis. The images need to be of 8-bit or 16-bit information depth and can be of arbitrary size. (Up to 1024 pixels maximum on both sides is recommended as larger images tend to significantly slow the program's performance). Histropy generates an image's histogram surrounded by a graphical user interface that allows one to select any range of image-pixel intensity levels, i.e. sections along the histograms' x-axis, using either the computer mouse or numerical text entries. The program subsequently calculates the (so-called Monkey Model) Shannon entropy and root-mean-square contrast for the selected section and displays them as part of what we call a "histogram-workspace-plot." To support the visual identification of small peaks in the histograms, the user can switch between a linear and log-base-10 display scale for the y-axis of the histograms. Pixel intensity data from different images can be overlaid onto the same histogram-workspace-plot for visual comparisons. The visual outputs of the program can be saved as histogram-workspace-plots in the PNG format for future usage. The source code of the program and a brief user manual are published in the supporting materials as well as on GitHub [1]. Instead of taking only 2D images as inputs, the program's functionality could be extended by a few lines of code to other potential uses employing data tables with one or two dimensions in the CSV format.

Keywords: image entropy; histograms; calculations; computer program

1. Introduction

The term histogram was coined around 1892 by statistician Karl Pearson to describe the visual representation of the distribution of data that quantified the frequency of data that fell into "bins" of a certain range. Histograms themselves were "first conceived as a visual aid to statistical approximations" [2]. The visual analysis that a histogram facilitates, combined with the quantitative information that can be extracted from it, gives histograms a wide range of applications, ranging from analyzing the distributions of test scores in a classroom to probabilistically characterizing the behavior of river discharge.

One of the common applications of histograms is in digital image processing. Pixel intensity histograms, which represent the distribution of the intensities of all pixels of an image, provide a measure of the image that can be useful for identifying similar images, compressing the image, and more. The here briefly described computer program "Histropy" generates a pixel intensity histogram for images that are either in grayscale, Red-Green-Blue color (RGB), or a uniform hue with a color-tone-range and allows for the numerical analysis of user-selected bin-ranges within the histogram. (One of its name-giving features is the calculation of the so-called Monkey-Model [3] Shannon entropy of 2D images.)

The computer program's prospective use within our research group is to quantitatively distinguish between symmetries and pseudosymmetries in a series of noisy as well as noise-filtered 2D-periodic images, i.e. crystal patterns. That noise-filtering of experimentally obtained and synthetic crystal

patterns will involve crystallographic image processing ([4] after objective, i.e. information-theory-based [5], crystallographic symmetry classifications [6]. See appendix for an example of a highly pseudosymmetric crystal pattern. As part of such quantitative distinctions between symmetries and pseudosymmetries, both the so-called "Monkey-Model" version [3] of the Shannon entropy and the standard root-mean-square (RMS) contrast in original or converted gray-scale crystal patterns need to be calculated.

For gray-scale patterns such as the one shown in the background of Figure A1a, where there are a few pronounced peaks in the corresponding histogram (Figure A1b), it will be informative to make quantifying calculations for selected ranges of the pixel intensity values. (This may be considered as constituting a very basic form of pattern segmentation.) Competing computer programs, e.g. the histogram routines that are part of the well-known electron crystallography software CRISP [4], do not typically offer the functionality desired for our studies. (Figure A2 illustrates the effect of crystallographic image processing in two non-disjoint plane symmetry groups on a noisy version of the crystal pattern that is shown in the background of Figure A1b).

Since the source code of our program is freely available on GitHub [1], other researchers are invited to download and modify it to support their own studies that do not need to be based on information in two-dimensional crystal patterns or gray-scale images. Other applications of histograms of 2D images may arise over time as well when parts of our code get reused.

2. Description of the Computer Program

As implied by the title of the paper, gray-scale images are the standard input option. Note that Histropy can only handle images of 16-bit depth when they are gray-scale so 16-bit color images must be converted before opening in Histropy. This conversion can be done using the open access program GIMP [7] by going to the "Image" dropdown menu and then selecting "Grayscale" for the "Mode". When the user selects an 8-bit color image to analyze through Histropy, the program converts the image to grayscale using the following equation to compute the intensity value for each individual pixel based on the pixel's RGB value:

$$\text{Grayscale value} = 0.299 * R + 0.587 * G + 0.114 * B, \quad (1)$$

where R stands for red, G for green, and B for blue. The coefficients reflect the fact that humans are most sensitive to green light and least sensitive to blue. Histropy's histogram-workspace plot consists of the histogram of the individual pixel intensities of a user-selected 2D image (or crystal pattern), a toolbar at the bottom of the screen that allows the user to navigate the histogram, four selection spaces to the right of the histogram that allow the user to quantify the histogram and image's information, and a display of the selected image being plotted to the right of the selection spaces, Figure 1. Note that the Histropy window must be viewed in full-screen mode for the layout seen on the next page. When processing 8-bit images, Histropy runs all actions in one to five seconds. Processing 16-bit images is about fifteen times slower on average with most operations taking around thirty seconds. These processing times will increase as the user overlays more images. These timings were found using a MacBook Pro, which has an Apple M3 processor with a 4.05 GHz max CPU clock rate.

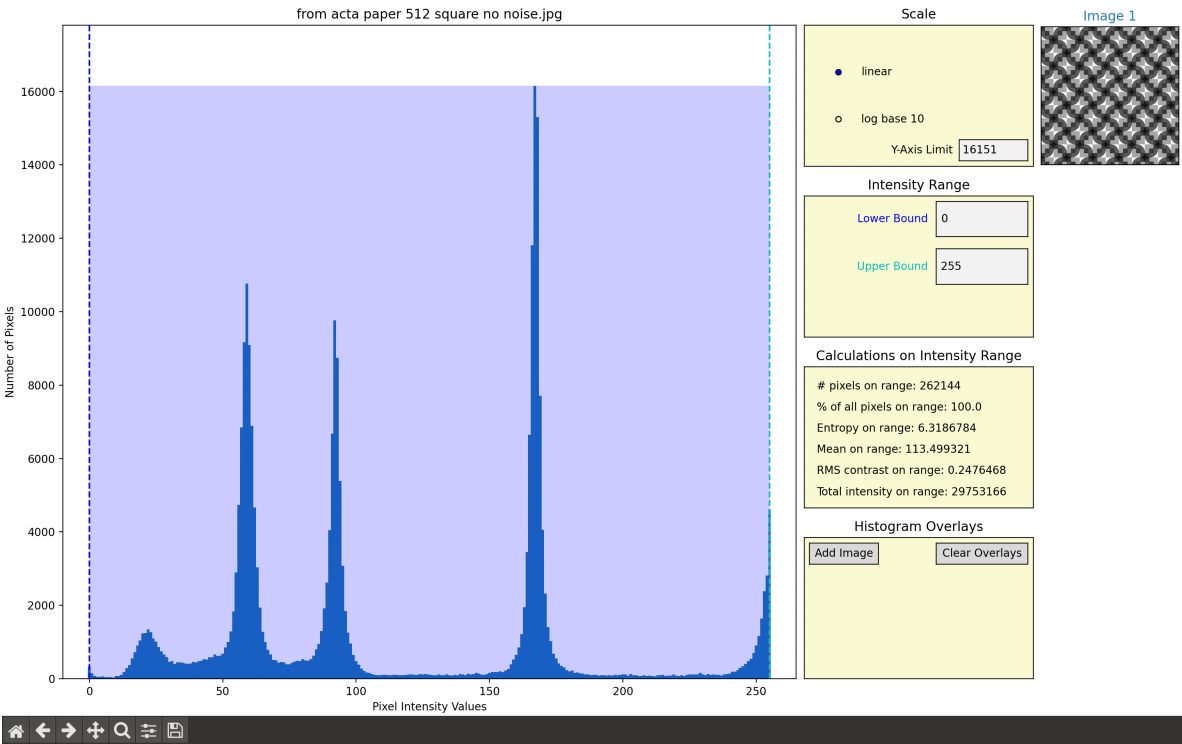


Figure 1. A screenshot of the (initial) Histropy workspace after the user has selected the image displayed in the top right corner. Note that the histogram’s title contains the file name of the selected image. Also note that the title of the image in the top right corner, "Image 1," has a font color corresponding to the color of the histogram. This particular histogram is from a 512 by 512 pixel cutout of the crystal pattern in the background of Figure A1a. (This figure is displayed in color in the online version of this paper.)

2.1. Selection Space 1: Scale

The first selection space, “Scale,” shown in Figure 2, allows the user to click-change between a linear and log base 10 scale, which affects the y-axis (display of number of pixels in a standard bin of unity) scale on the histogram. It also contains an input field for a y-axis limit, which defaults to the maximum y-value in the histogram. This value represents how many pixels have the most common pixel intensity value or the height of the most prominent peak in the histogram.

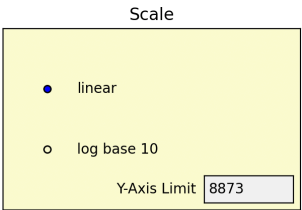


Figure 2. A close up of the scale selection space. The linear y-axis is currently selected and the y-axis limit is set to the default: the max value of the inputted image.

2.2. Selection Space 2: Intensity Range

The text fields in the second selection space, “Intensity Range,” displayed in Figure 3, set the range for calculations performed in the third selection space. These text fields default to the minimum and maximum pixel intensity in the image. They can be set by directly typing into the fields themselves or by clicking directly on the histogram, which will automatically set the range in the selection space text fields to the x-value corresponding to where the user’s mouse is pointing. When the user’s mouse is hovering over the histogram, the text in the bottom right corner of the screen will display

the coordinates that the mouse is over, Figure 4. These coordinates can be used to accurately select the x-values for the range. The range selected is visually represented by the vertical bars and a translucent blue rectangle shown on the histogram, Figure 1. This range selection allows the user to perform a simple form of segmentation by separating out the pixel bins that contribute to a certain histogram peak.

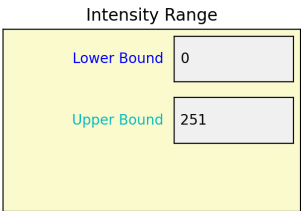


Figure 3. The selection space for the range of pixel intensities (x-values on the histogram) that the calculations are computed over. The text fields shown here are editable directly by clicking on them and typing in new values.

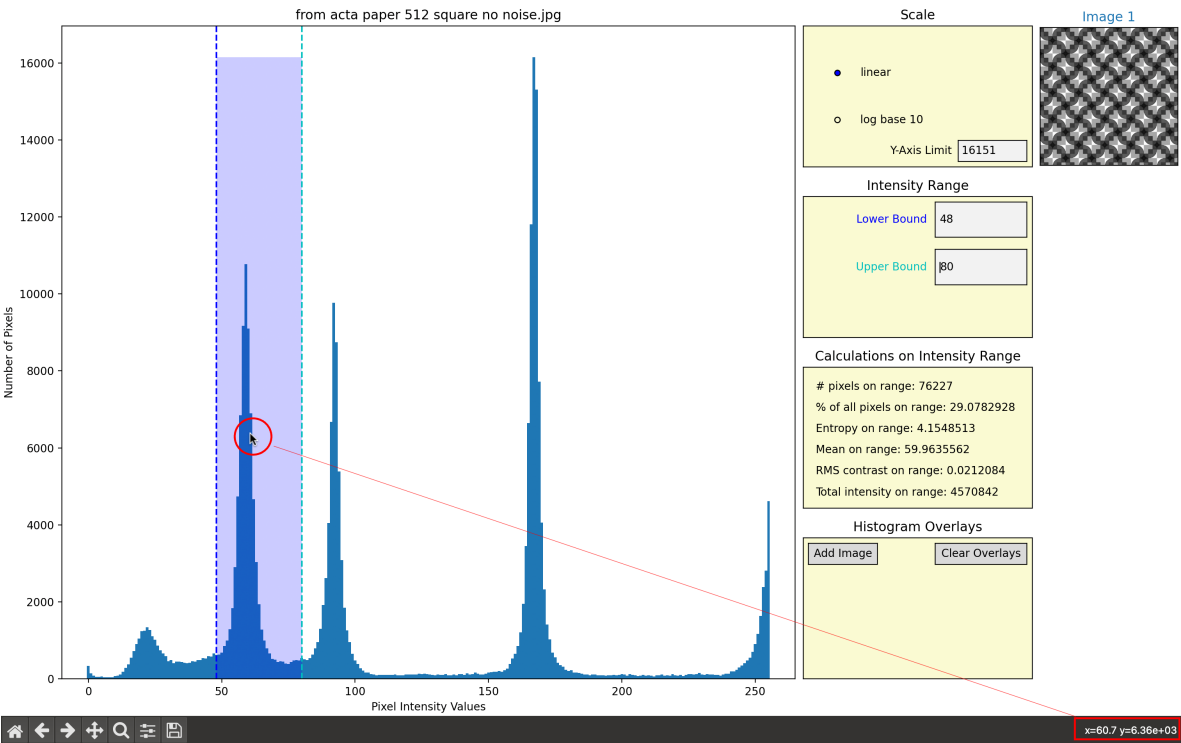


Figure 4. The cursor, circled in red on the histogram here, corresponds to the coordinates shown in the bottom right corner in a red box. The added red boxes and arrows indicate the connection between the cursor placement and the value displayed in the bottom right corner. Upon clicking at this coordinate, the lower bound of the intensity range would update to 61 (where the cursor is pointing), and this change would reflect in the selection space text box. (This figure is displayed in color in the online version of this paper.)

2.3. Selection Space 3: Calculations

The third selection space, “Calculations,” shown in Figure 5, is a display of the following calculations over a user-selected range (which defaults to the entire range of the histogram):

1. The number of pixels on the given range
2. The percent of total pixels that are found in the given range

3. A form of the Shannon entropy on the given range
4. The mean on the given range
5. The RMS contrast on the range
6. The total pixel intensity on the range (a sum)

Calculations on Intensity Range	
# pixels on range:	262144
% of all pixels on range:	100.0
Entropy on range:	6.6837814
Mean on range:	104.1250877
RMS contrast on range:	0.2494588
Total intensity on range:	27295767

Figure 5. The calculation selection space. These text displays are not editable but they will update as the intensity range field is updated, either through directly entering into the fields or by clicking on the histogram.

Note that these calculations are not impacted by y-axis limitings done in the “Scale” selection space. The one-dimensional Shannon entropy is calculated using the following equation:

$$\sum_{i=1}^N \frac{p_i}{N} \cdot \log_2\left(\frac{p_i}{N}\right), \quad (2)$$

where N is the number of pixels on the range and p_i is the intensity of the i^{th} pixel. The mean of the range is the sum of the pixel intensities over a range divided by the number of pixels on the range:

$$\bar{p} = \left(\frac{1}{N}\right) \sum_{i=1}^N p_i. \quad (3)$$

The total intensity is calculated as $N \cdot \bar{p}$. Finally, the RMS Contrast is calculated with the equation

$$\frac{1}{d} \sqrt{\sum_{i=1}^N \frac{(p_i - \bar{p})^2}{N}}, \quad (4)$$

where d is $2^{\text{bit depth}} - 1$ and $\frac{1}{d}$ normalizes the output between 0 and 1.

2.4. Selection Space 4: Histogram Overlay

The final selection space, “Histogram Overlays,” Figure 6, allows the user to add images whose pixel intensity data will be overlaid onto the histogram of the first image. The second image that a user adds will have its quantifying calculation data displayed in the color corresponding to its plot in the “Histogram Overlays” selection space, Figure 7. Currently, Histropy is only capable of displaying these quantifications for one image in addition to the user’s first image. As another image is read into the program using the fourth selection space, the selection space will switch to displaying the file names of the added images in colors corresponding to how they appear on the histogram, Figure 7. The first four images that the user overlays will appear on the right underneath the original image with their title colors corresponding to their appearance on the histogram, Figure 8. Up to 22 images

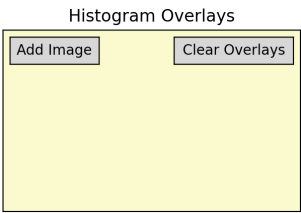


Figure 6. The Histogram Overlays selection space, which contains two buttons, one which opens up a file dialogue to add images to the histogram and another which resets the workspace to its initial state with no overlays.

can be overlaid but it is recommended to stick to 10 overlaid images or less in order to maximize program performance. If the user wishes to remove the overlays, one has to click on the “Clear Overlays” button in the fourth selection space, which will reset the histogram, the selection space, and the image displays to their original states.

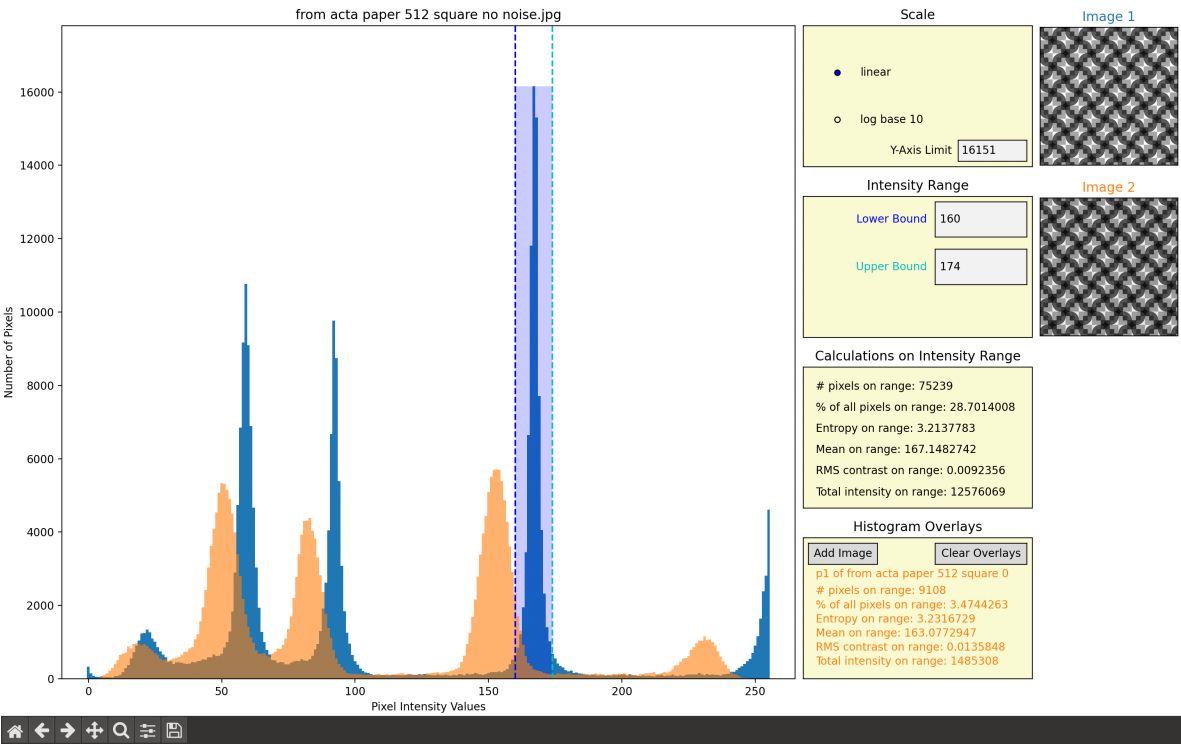


Figure 7. The overlaid histogram, with the calculations in the "Histogram Overlays" selection space. The image corresponding to this histogram is displayed on the right under the title “Image 2,” with the font color matching the color displayed on the histogram and in the “Histogram Overlays” selection space. (This figure is displayed in color in the online version of this paper.)

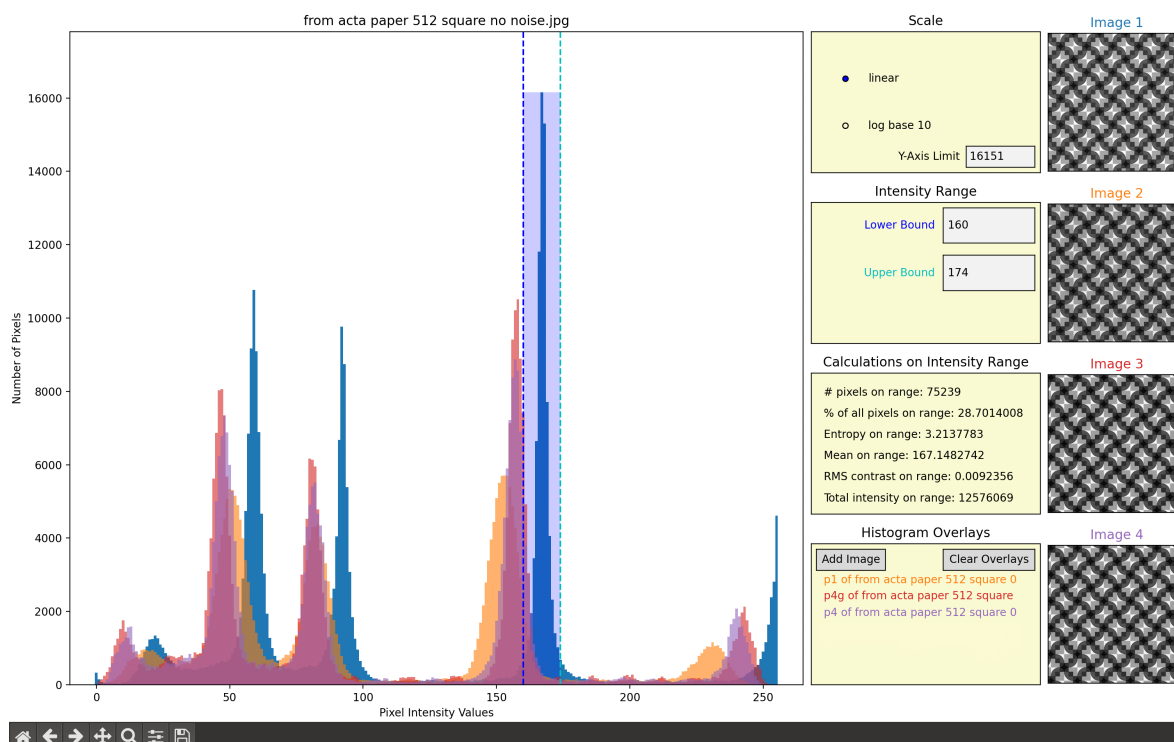


Figure 8. Three overlaid histograms with the truncated file names in the "Histogram Overlays" selection space. Note that the images on the right correspond to the overlaid histograms, with the title color of each image matching the image that appears on the histogram and the filename that appears in the "Histogram Overlays" selection space. (This figure is displayed in color in the online version of this paper.

2.5. Navigating the Histogram

The buttons in the bottom left corner of the Histropy workspace, Figure 1, which appear underneath the histogram, allow the user to zoom in on specific parts of the histogram using the magnifying glass and move around the viewing window along the x- and y-axes with the axes button. When the user wishes to go back to a previous view, they can do so using the back arrow or they can move forward to the next view with the forward arrow. All viewing changes can be reset with the home button and the user can save the full plot image as a PNG with the save button.

3. Concluding Remarks

Histropy is a versatile program that facilitates the visual analysis and quantification of images. The functionality and a potential use of the computer program Histropy has been briefly demonstrated in a few examples based on versions of the crystal pattern in the appendix. An expanded description (handbook) of this program is freely available, together with the source code, in the supporting material for this paper. The authors hope that other researchers will find this paper interesting and eventually make good use of this program or parts of its source code in their own work.

Supplementary Materials: The following supporting information can be downloaded at: <https://drive.google.com/uc?export=download&id=1bkqPXe3lj1wyvU2a8Goa2u6MPQzuCbDp>, Histropy Handbook; Histropy Source Code

Author Contributions: The first author is the creator of the computer program, currently a freshman at the Georgia Institute of Technology. The second (senior) author is an applied crystallographer who has, in recent years, pioneered the application of a geometric form of information theory to crystallographic symmetry classifications of experimental data with two spatial dimensions..

Funding: This project received support from a Faculty Development Grant of Portland State University to Peter Moeck.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RMS Root Mean Square

Appendix A

As an example of a subjective distinction between genuine symmetries and pseudosymmetries, the crystal pattern that forms the background in Figure A1a will likely be classified by most crystallographers at first sight as featuring the plane symmetry group $p4gm$.

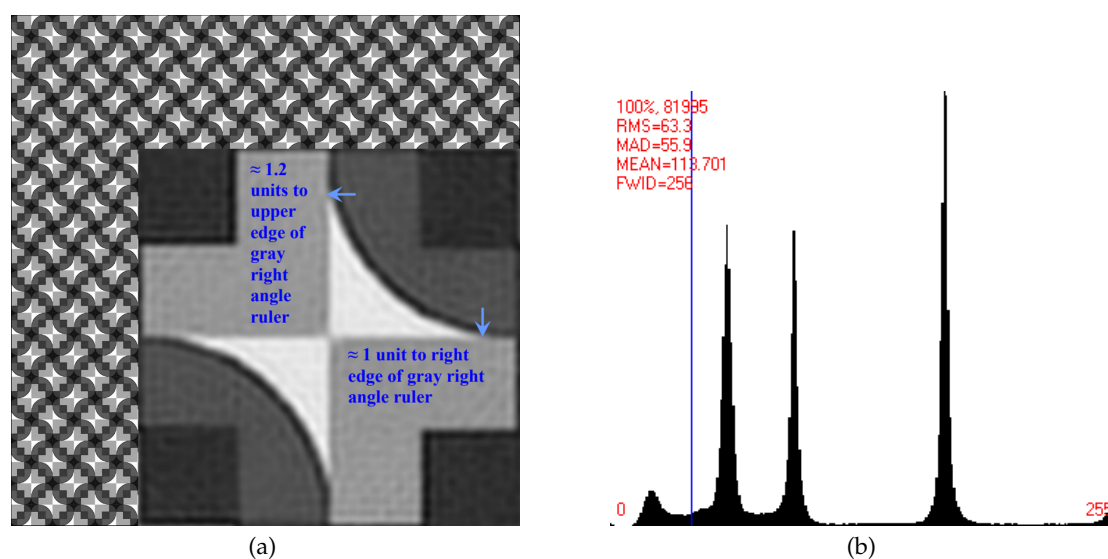


Figure A1. (a) This sub-figure displays a crystal pattern that will, by most people, be subjectively misclassified (at least at first sight) as featuring plane symmetry group $p4gm$. The inset figure, which blows up a small part of the pattern, shows why this plane symmetry group must be dismissed as a pseudosymmetry. Reproduced with permission from the paper and supporting materials of [6]. (b) This sub-figure displays the histogram for the crystal pattern in Figure A1a as obtained by the well-known electron crystallography program CRISP. (The figure is displayed in color in the online version of this paper.)

However, upon closer visual inspection of the inset of Figure A1a, which is a blow-up around site symmetry 2 at position $0, \frac{1}{2}$ of the translation periodic unit cell, it becomes clear that this can only be a pseudosymmetry. Taking into account the intensity values of all of the pattern's pixels, the information-theory-based crystallographic symmetry classification of this crystal pattern [6] revealed that the genuine plane symmetry is indeed only $p4$. In agreement with this objective classification, the "white bow-tie" feature of the inset does not genuinely feature point symmetry $2mm$ by visual inspection because the two diagonal mirror lines are broken to larger extents than the central two-fold rotation point. [6,8] also give brief accounts of the creation of this pattern, from which it becomes clear that all mirror and glide lines can only be strong pseudosymmetries.

Figure A2 illustrates the effect of the crystallographic processing of a noisy version of a 512 by 512 pixel cut-out of the crystal pattern in the background of Figure A1a. Note that both the visual "sharpenings" and relative shifts of the histogram peaks of the two crystallographically processed version of the noisy image that were enforced to display the symmetries of the non-disjoint plane symmetry groups $p2$ and $p4$ are easily explained by the enhanced averagings over progressively smaller asymmetric units of the translation periodic unit cells. Note also that the histogram of the $p4$

(-enforced) version of the first (noisy) image is rather similar to those shown in Figures 1 and A1b. This is a testament to both, the veracity of the image processing method and the correctness of the $p4$ plane symmetry classification of the noisy crystal pattern [8].

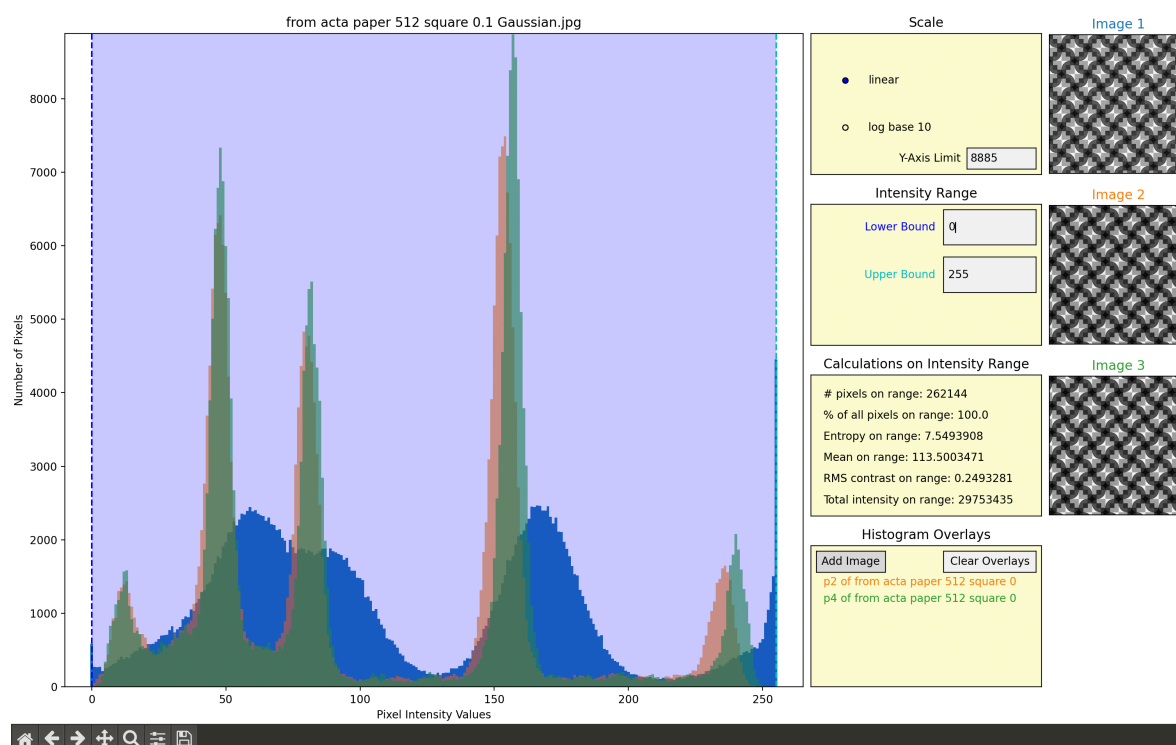


Figure A2. Histograms of a noisy version of the crystal pattern in the background of Figure A1a (displayed in blue) overlaid with the histograms of the $p2$ (displayed in orange) and $p4$ (displayed in green) image versions that were obtained by crystallographic image processing of that noisy image. (This figure is displayed in color in the online version of this paper.) The visual results, i.e. both peak sharpenings and shifts, in the histograms of the symmetry enforced versions of Image 1 are as expected.

References

1. GitHub SMenon-14/Histropy. Available online: <https://github.com/SMenon-14/Histropy> (accessed on 30 July 2024).
2. Ioniddis, Y. The History of Histograms (abridged). In Proceedings of the 29th International Conference on Very Large Data Bases, Berlin, Germany, (9th September 2003); 19–30.
3. Razlighi, Q.R.; Nasser, K. A comparison study of image spatial entropy. In Proceedings of IS&T/SPIE Electronic Imaging, San Jose, California, United States, (18th January 2009); 72571X1 – 72571X-10.
4. Xiaodong, Z.; Sven, H.; Oleynikov, P. *Electron crystallography : electron microscopy and electron diffraction.*, Illustrated ed.; Oxford University Press: Oxford, New York, United States, 2016;
5. Kenichi, K. Geometric Information Criterion for Model Selection. *IJCV* **1998**, *26*, 171–189.
6. Moeck, P. Objective crystallographic symmetry classifications of a noisy crystal pattern with strong Fedorov-type pseudosymmetries and its optimal image-quality enhancement. *Acta Crystallogr. A* **2022**, *78*, 172–199.
7. GIMP. Freely available online: <https://www.gimp.org/> (accessed on 30th July 2024).
8. Moeck, P. Genuine Plane Symmetries versus Pseudosymmetries in Two Crystal Patterns of Graphic Artwork. arXiv:2304.03915v3 (<https://doi.org/10.48550/arXiv.2304.03915>) and EasyChair Preprint № 10089 (<https://easychair.org/publications/preprint/Cj97>). (preprint) 2023.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.