Data Discovery in Images Using a Median- Absolute-Deviation Based Filter

Shataneek Banerjee[1], *Amardip* Ghosh[1], *Avijit* Dey[2], and *Prasanta* Pal[3*]

[1]Department of Aerospace Engineering, IIT Kharagpur, Kharagpur 721302, India
[2]Big Bang Software, Kolkata, India
[3]SHIOM LLC, Rhode Island Startup Incubator (RIHUB), Rhode Island, USA

Abstract. The utility of a recently developed, novel, computational digital image curation tool called SOCKS (Statistical Outlier Curation Kernel Software) is demonstrated in this paper. Raw image-data of the James Webb Space Telescope (JWST) is used for the tests. Data is decontaminated through an iterative, parametrically tuned sequence of operations involving the conditional flagging of outliers, followed by the curation of the flagged pixels. Every iteration produces a decontaminated image and an outlier image, till the sequence converges asymptotically producing a final decontaminated image that represents the ground-truth as accurately as possible. The intermediate outlier images are shown to contain a wealth of information about the original image. An iteration-by-iteration approach of deriving insight out of the intermediate outlier images is described in detail. Traditional filtering methods, in contrast, discard outliers causing irreversible loss of information contained therein.

1 Introduction

Noise of various kinds such as thermal noise, shot noise, transit-time noise, and environment-induced noise can affect the clarity of digital images acquired through CCD and CMOS sensors [1-8]. Outliers, in the form of undesired artefacts, could also be classified as 'noise' affecting the visual clarity of the underlying image.

Various filters have been utilized to enhance the clarity in such images. Traditional filters such as the Gaussian filter [9-14] a low pass or smoothing filter de-noises images by inducing the effect of "blur". Wiener filter [15-22], an adaptive linear low pass filter adjusts the filter coefficients according to an optimization algorithm removing high frequency noise while preserving the desirable high frequency components like edges. Median filter [23-30], a non-linear low pass filter preserves desirable high frequency components like fine details and edges while removing unwanted impulse noise. The apparent visual improvement obtained through these filters result from the smoothening of the original image. Two problems arise in such cases. Firstly, every pixel in the image is affected by the filter irrespective of it is a 'good-pixel' or a 'bad-pixel'. Secondly, information about noise and outlier artefacts are not systematically curated since the focus of the effort is often on obtaining a more aesthetic representation of the underlying image.

In certain critical applications, however, the outlier artefacts may contain equal, if not more valuable information as the underlying image. In structural health monitoring, the detection of minute cracks and fracture spots that appear as outliers in structural elements may be required. In early diagnosis of cancer, the detection and classification of exceedingly small calcified nodules that appear as unwanted indiscernible artefacts in human tissue images may be required. In space systems engineering, the detection and classification of minute craters that appear as outliers on the surface of the moon may be required to find possible sites of impact.

In this paper, various properties of such an outlier preserving statistical filter called SOCKS (Statistical Outlier Curation Kernel Software), developed by the authors, is demonstrated through applications to various image processing case studies. The basic difference between the SOCKS filter and traditional filters is that the SOCKS filter preserves raw information by flagging and collecting outliers in an 'outlier image' while strategically curating only affected data points at all reasonable scales and thresholds in the 'source image'. Small perturbations are optionally curated by regression at all scales.

2 SOCKS Filter

The SOCKS algorithm [31] uses a convolution kernel to check whether the absolute value of Modified Z at a particular pixel location is beyond a certain threshold to flag it as an outlier. For a set of samples X, modified Z or $\hat{Z}$ is defined as

$$\hat{Z} = \frac{X - Median}{Median\ Absolute\ Deviation}$$

The filter works on the heuristic that the outliers, which contribute to large-scale noise [32] occur roughly beyond 3 MADs (Median Absolute Deviation) from the median on both sides in a Modified Z distribution analogous to the outliers being roughly beyond 3 SDs (Standard Deviation) from the mean on both sides of the normal distribution [33]. The particular choice of the threshold value in a given SOCKS implementation depends on the nature of the data set as well as specifics of the curation goal in a given context. The outlier positions are contextually curated through interpolation, averaging, or regression techniques. These curated data points account for the missing information of the outliers. As an additional optional step, the non-outlier but presumably perturbed/noisy data points are further regressed with respect to their kernel size specific local neighbourhood. The filter makes use of two key parameters, namely the threshold and the scale. The threshold relates to a percentile measure in the PDF of the Modified Z distribution allowing the detection of outliers. The scale relates to the size of the convolution kernel used in flagging the outliers. The Median and the MAD are robust statistical parameters [34, 35] that are relatively insensitive to extreme values. Since Modified Z uses the Median and the MAD, it performs better than Z in terms of robust statistical performance metrics like breakdown point [36] and influence function [37].

3. Results and Discussions

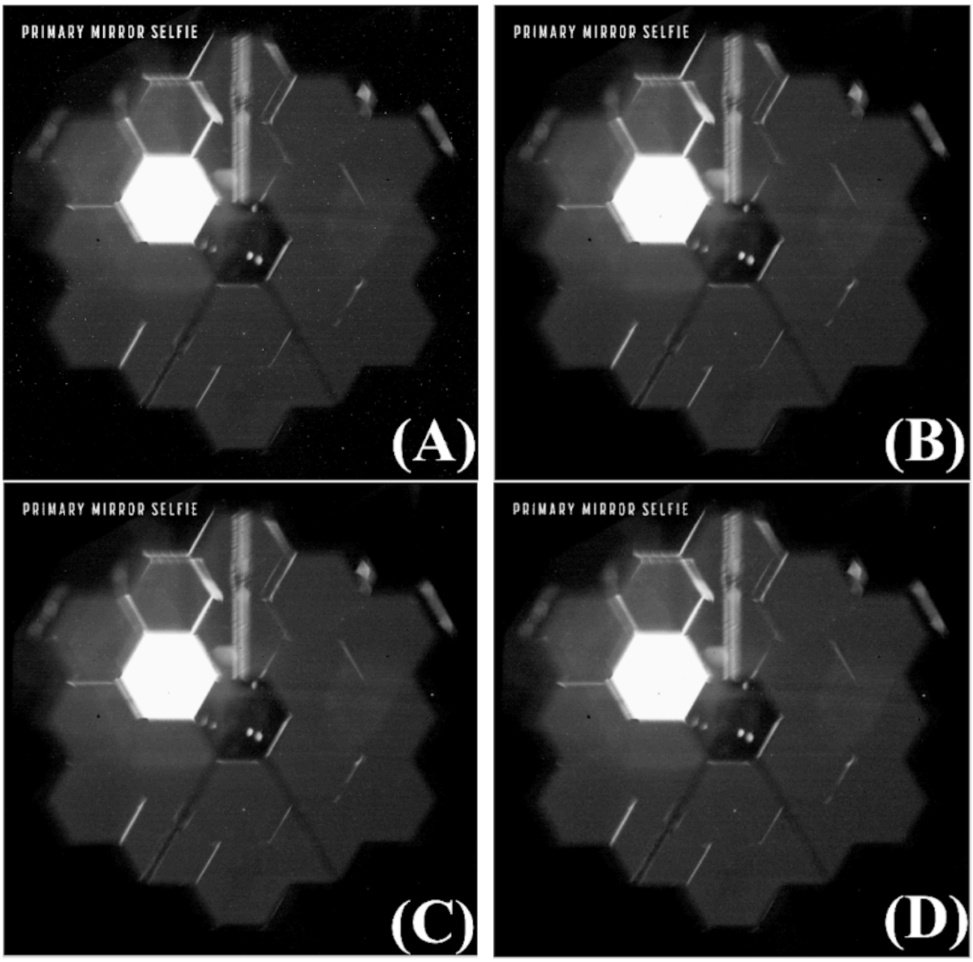

Fig. 1 (A) Original selfie image sent by JWST. (B) Curated image by taking Fig.1A as input and removing the first-order set of outliers (stars and speckles) (C) Curated image by taking Fig.1B as input (D) Curated image by taking Fig.1C as input.

.

Fig. 1 shows a typical use case, demonstrating the systematic noise and outlier removal from a selfie image of the James Webb Space Telescope (JWST). Fig. 1A shows the original selfie image sent by JWST. Fig. 1B shows the curated image after one complete iteration, taking Fig. 1A as input and removing the first order set of outliers, comprising mostly of stars and speckles, with kernel size k=4 and threshold $\tau = 0.86$. Pixels with intensities greater than Median + $3.\tau$.MAD are flagged as outliers. Fig. 1C shows the curated image after one more iteration, taking Fig. 1B as input and using the same parameters (threshold and kernel size) as in the previous iteration. Fig. 1D shows the curated image after one more iteration, taking Fig. 1C as input and using the same parameters as in the previous iteration. Fig. 1

demonstrates that using a systematic iterative process, the visual clarity of an image can be enhanced by removing noise and speckles from the data.

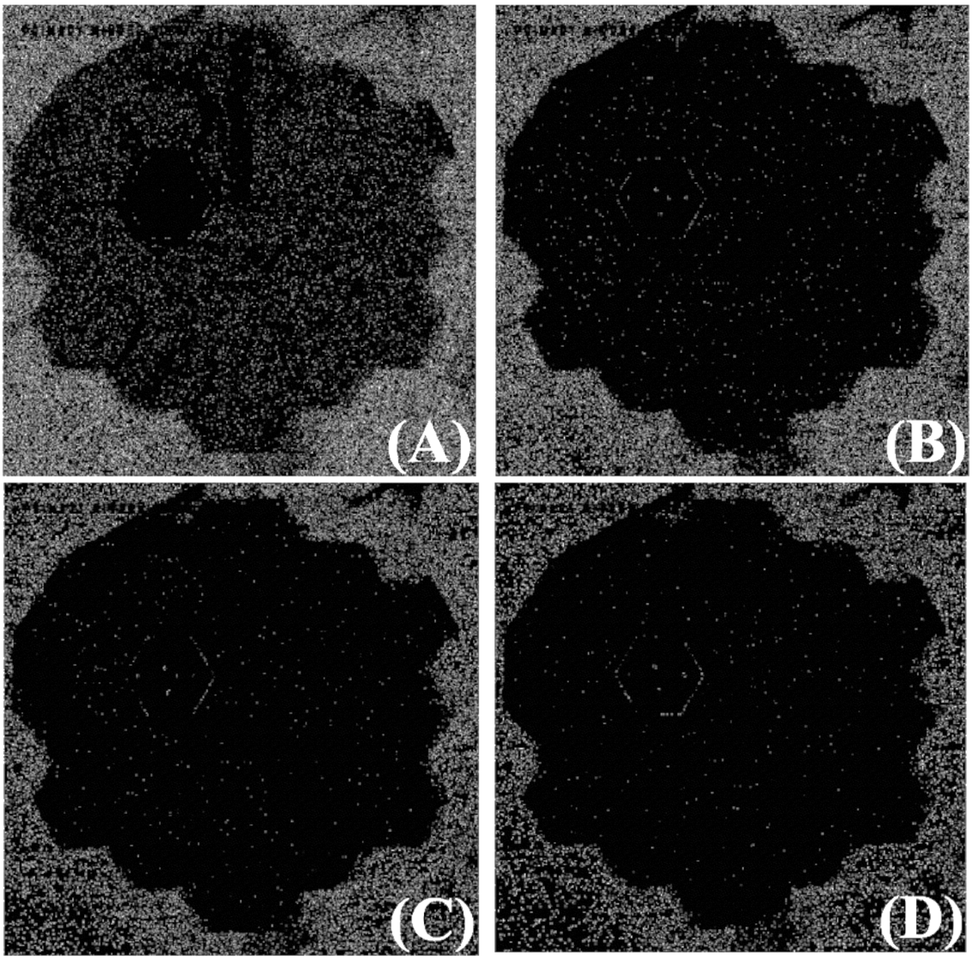

Fig. 2 Demonstration of the spatial structure of outlier (and noise) distribution obtained with Kernel size K=4 from the selfie of JWST as demonstrated in Fig 1. (A) Outlier distribution obtained from Fig. 1A. (B) Outlier distribution obtained from Fig. 1B.(C) Outlier distribution obtained from Fig 1C. (D) Outlier distribution obtained from Fig 1D.

Fig. 2A demonstrates the spatial structure of outlier and noise distribution obtained with kernel size K=4 and threshold=0.86 from the selfie image of the JWST as shown in Fig 1A. Fig. 2A shows that there is an abundance of information in the outlier space. The image is binary with white dots representing outliers. The intensity of every outlier is set to 1 while the absence of outliers is set to 0. It is trivial to observe that, the density of stars in the background part of the sky is very dense compared to the same in the original data as shown in Fig 1A. Stars reflected by the primary mirror can also be seen in the outlier space. Although it is apparently surprising, however in the original data we notice only the relatively bright high-intensity speckles as compared to all the marked flags that are visible despite being less intense in the raw data set. Fig. 2B shows the spatial structure of outlier and noise distribution obtained from Fig 1B using a kernel size K=4 and threshold=0.86. Since the outliers shown in Fig. 2B are removed from Fig. 1A to obtain Fig. 1B, the outliers obtained from Fig. 1B are the outliers of lesser intensity compared to the outliers obtained from Fig. 1A. In this process, after every iteration we find outliers of lesser and lesser intensities revealing patterns in the outlier space not visible in the previous iteration due to the presence of more intense outliers in the figures from where they were extracted. Since the intensities of the outliers after every iteration are mapped to a binary scale, they show the same intensity in all the outlier images. Fig. 2C shows the spatial structure of outlier and noise distribution obtained with kernel size K=4 and threshold=0.86 obtained from the image in Fig 1C. Fig. 2D shows the spatial structure of outlier and noise distribution obtained with kernel size K=4 and threshold=0.86 obtained from the image in Fig 1D.

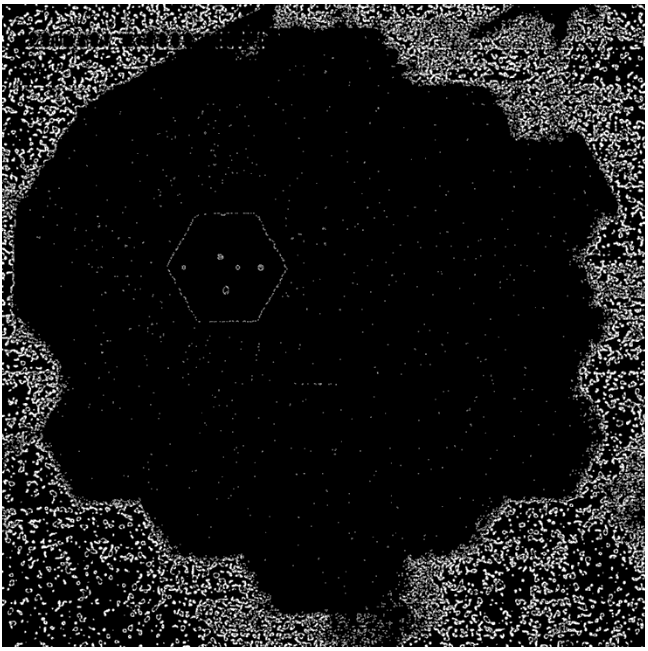

Fig. 3 Outliers after fourth iteration. K = 6, $\tau = 0.86$

Fig. 3 shows outlier distribution after removing outliers from Fig 1D with slightly higher kernel size K=6, but same threshold, $\tau = 0.86$. In Fig. 3, the region corresponding to the illuminated hexagonal region in Fig. 1A, the appearance of 5 circular structures comes as a surprise as visually, the illuminated region in Fig. 1A looks uniform in intensity. Also, the top right portion in Fig. 3 shows a finger-like shadowy region and a possible explanation is the blocking effect of the background stars by the struts surrounding the telescope as described in the design of JWST.

3. Conclusions

The use of the SOCKS filter in information separation, image filtering and outlier detection is demonstrated. After every iteration, new patterns in the outlier image sets are detected revealing information buried in the source image at various levels of intensities. This attribute of the filter makes it suitable for the detection of early-stage tumours, calcifications and cysts in human tissues where they may occur as outliers of relatively low intensity which the iterative outlier detection routine can reveal in a systematic stepwise fashion.

References

1. Q. Zhang, R. Ward, *Automatic assessment of signal-to-thermal noise ratio of television images*, IEEE trans. consum. electron., **41**,1, pp. 108–117 (1995)
2. L. Vizioli , S. Moeller, L. Dowdle, M. Akcakaya, F.D. Martino, E. Yacoub, K. Urgurbil *Lowering the thermal noise barrier in functional brain mapping with magnetic resonance imaging*, Nat. Commun.,**12**, 1, p. 5181 (2021)
3. A. Kerr, J. Randa, *Thermal Noise and Noise Measurements—A 2010 Update*, IEEE Microw. Mag., **11**, 6, pp. 40–52 (2010)
4. H. Kuniba, R. S. Berns, *Spectral sensitivity optimization of color image sensors considering photon shot noise*, J. Electron. Imaging, **18**, 2, p. 023002 (2009)
5. D. Zha, T. Qiu, *A new algorithm for shot noise removal in medical ultrasound images based on alpha-stable model*, Int. J. Adapt. Control Signal Process., **20**, 6, pp. 251–263 (2006)
6. P. A. Cheremkhin, N. N. Evtikhiev, V. V. Krasnov, V. G. Rodin, R. S. Starikov, *Shot noise and fixed-pattern noise effects on digital hologram reconstruction*, Opt. Lasers Eng., **139**, 106461, p. 106461 (2021)
7. M. Trippe, G. Bosman, A. Van Der Ziel, *Transit-time effects in the noise of Schottky-barrier diodes*, IEEE Trans. Microw. Theory Tech., **34**, 11, pp. 1183–1192 (1986)
8. I. Sinclair, *Sensors and Transducers*, 3rd ed. London, England: Newnes (2001)
9. A. Dogra, P. Bhalla, *Image sharpening by Gaussian and Butterworth high pass filter*, Biomed. Pharmacol. J., **7**, 2, pp. 707–713 (2014)
10. G. Gomez, *Local Smoothness in terms of Variance: the Adaptive GaussianFilter*, in Procedings of the British Machine Vision Conference 2000 (2000)
11. A. Makandar, B. Halalli, *Image enhancement techniques using highpass and lowpass filters*, Int. J. Comput. Appl., **109**, 14, pp. 21–27 (2015)
12. P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, *Perceptual blur and ringing metrics: application to JPEG2000*, Signal Process. Image Commun., **19**, 2, pp. 163–172 (2004)
13. J. M. Leski, *Robust weighted averaging*, IEEE Trans. Biomed. Eng., **49**, 8, pp. 796–804 (2002)
14. M. Brady, B. K. P. Horn, *Rotationally symmetric operators for surface interpolation*, Comput. Vis. Graph. Image Process., **22**, 1, pp. 70–94 (1983)
15. L. D. Marks, *Wiener-filter enhancement of noisy HREM images*, Ultramicroscopy, **62**, 1–2, pp. 43–52 (1996)
16. J. Chen, J. Benesty, Y. Huang, S. Doclo, *New insights into the noise reduction Wiener filter*, IEEE Trans. Audio Speech Lang. Processing, **14**, 4, pp. 1218–1234 (2006)
17. E. A. Robinson, S. Treitel, *Principles of digital wiener filtering*, Geophys. Prospect., **15**, 3, pp. 311–332 (1967)
18. S. O. Haykin, *Adaptive Filter Theory*, 5th ed. Upper Saddle River, NJ: Pearson (2013)
19. V. S. Frost, J. A. Stiles, K. S. Shanmugam, J. C. Holtzman, S. A. Smith, *An adaptive filter for smoothing noisy radar images*, Proc. IEEE Inst. Electr. Electron. Eng., **69**, 1, pp. 133–135 (1981)
20. R. Saluja, A. Boyat, *Wavelet based image denoising using weighted highpass filtering coefficients and adaptive wiener filter*, in 2015 International Conference on Computer, Communication and Control (IC4) (2015)

21. C. V. Cannistraci, F. M. Montevecchi, M. Alessio, *Median-modified Wiener filter provides efficient denoising, preserving spot edge and morphology in 2-DE image processing*, Proteomics, **9**, 21, pp. 4908–4919 (2009)
22. V. Strela, *Denoising Via Block Wiener Filtering in Wavelet Domain*, in European Congress of Mathematics, Basel: Birkhäuser Basel, pp. 619–625 (2001)
23. T. Sun, Y. Neuvo, *Detail-preserving median based filters in image processing*, Pattern Recognit. Lett., **15**, 4, pp. 341–347 (1994)
24. H.-M. Lin, A. N. Willson, *Median filters with adaptive length*, IEEE Trans. Circuits Syst., **35**, 6, pp. 675–690 (1988)
25. J. C. Church, Y. Chen, S. V. Rice, *A Spatial Median Filter for noise removal in digital images*, in IEEE SoutheastCon 2008 (2008)
26. G. Arce, M. McLoughlin, *Theoretical analysis of the max/Median filter*, IEEE Trans. Acoust., **35**, 1, pp. 60–69 (1987)
27. E. Ritenour, T. Nelson, U. Raff, *Applications of the median filter to digital radiographic images*, in ICASSP, IEEE International Conference on Acoustics, Speech, and Signal Processing (2005)
28. P. S. Windyga, *Fast impulsive noise removal*, IEEE Trans. Image Process., **10**, 1, pp. 173–179 (2001)
29. A. B. Hamza, P. L. Luque-Escamilla, J. Martínez-Aroza, R. Román-Roldán, J. Math. Imaging Vis., **11**, 2, pp. 161–177 (1999)
30. R. H. Chan, C.-W. Ho, M. Nikolova, *Salt-and-Pepper noise removal by median-type noise detectors and detail-preserving regularization*, IEEE Trans. Image Process., **14**, 10, pp. 1479–1485 (2005)
31. P. Pal, R.V. Lutterveld, N. Quirós, V. Taylor, J.A. Brewer*, Statistical Outlier Curation Kernel Software (SOCKS): A Modern, Efficient Outlier Detection and Curation Suite.* TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.15043695.v2 (2021)
32. O. Schall, A. Belyaev, H.-P. Seidel, *Robust filtering of noisy scattered point data*, in Proceedings Eurographics/IEEE VGTC Symposium Point-Based Graphics, 2005 (2005)
33. A. J. Wefald, J. P. Katz, R. G. Downey, K. G. Rust, *Organizational slack and performance: The impact of outliers*, J. Appl. Bus. Res. (JABR), **26**, 1 (2010)
34. P. J. Huber, *Robust Statistics*, 1st ed. Nashville, TN: John Wiley & Sons (2005)
35. P. J. Rousseeuw, M. Hubert, *Robust statistics for outlier detection: Robust statistics for outlier detection*, Wiley Interdiscip. Rev. Data Min. Knowl. Discov., **1**, 1, pp. 73–79, (2011)
36. C. Leys, C. Ley, O. Klein, P. Bernard, L. Licata, *Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median*, J. Exp. Soc. Psychol., **49**, 4, pp. 764–766 (2013)
37. F. R. Hampel, *The influence curve and its role in robust estimation*, J. Am. Stat. Assoc., **69**, 346, pp. 383–393 (1974)